Multiscale molecular kinetics by coupling Markov state models and reaction-diffusion dynamics

del Razo, M.J.; Dibak, M.; Schuette, C.; Noé, F.

DOI
10.1063/5.0060314

Publication date
2021

Document Version
Final published version

Published in
Journal of Chemical Physics

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Multiscale molecular kinetics by coupling Markov state models and reaction-diffusion dynamics

Cite as: J. Chem. Phys. 155, 124109 (2021); doi: 10.1063/5.0060314
Submitted: 17 June 2021 • Accepted: 2 September 2021 •
Published Online: 27 September 2021

Mauricio J. del Razo,[1,2,3,4,a)] Manuel Dibak,[4] Christof Schütte,[5] and Frank Noé[4,6,7,a)]

AFFILIATIONS

[1] Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, The Netherlands
[2] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, The Netherlands
[3] Dutch Institute for Emergent Phenomena, Amsterdam, The Netherlands
[4] Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany
[5] Zuse Institute Berlin, Berlin, Germany
[6] Department of Physics, Freie Universität Berlin, Berlin, Germany
[7] Department of Chemistry, Rice University, Houston, Texas 77005, USA

[a)] **Authors to whom correspondence should be addressed:** maojrs@gmail.com and frank.noe@fu-berlin.de

ABSTRACT

A novel approach to simulate simple protein–ligand systems at large time and length scales is to couple Markov state models (MSMs) of molecular kinetics with particle-based reaction-diffusion (RD) simulations, MSM/RD. Currently, MSM/RD lacks a mathematical framework to derive coupling schemes, is limited to isotropic ligands in a single conformational state, and lacks multiparticle extensions. In this work, we address these needs by developing a general MSM/RD framework by coarse-graining molecular dynamics into hybrid switching diffusion processes. Given enough data to parameterize the model, it is capable of modeling protein–protein interactions over large time and length scales, and it can be extended to handle multiple molecules. We derive the MSM/RD framework, and we implement and verify it for two protein–protein benchmark systems and one multiparticle implementation to model the formation of pentameric ring molecules. To enable reproducibility, we have published our code in the MSM/RD software package.

I. INTRODUCTION

Molecular dynamics (MD) simulations have allowed the study of a broad range of biological systems from small molecules, such as anesthetics or small peptides, to large protein complexes, such as ribosomes or even virus capsids.[1] One of the main challenges faced by MD simulations is their high computational cost, which can lead to inadequate sampling of conformational states. While there is a large body of research focused on sampling of long-time-scale dynamics of individual macromolecules, there has been less attention to sampling and simulating the interactions of many macromolecules on larger length scales. This is a more complex problem since it not only involves the long-time dynamics but can also involve several orders of magnitude in length scale. One landmark example is cellular signaling, where relevant processes happen across 6 orders of magnitude in length scales (0.1 nm to 100 μm) and 18 orders of magnitude in time scales (femtoseconds to hours).[2–4] Two of the most successful approaches to model biomolecular processes at larger time or length scales are the following:

- Markov state models (MSMs) of molecular kinetics are one of the most well-known techniques to mitigate the MD sampling problem.[5–10] They approximate the long-time dynamics of MD systems by Markov chains on discrete partitions of configuration space. This allows us to extract the long-time kinetics from short MD trajectories and to calculate molecular observables. State-of-the-art developments have pipelined the MSM approach into deep learning frameworks.[11,12] However, larger and more complex systems require sampling an exponentially growing number of states, constraining its applicability to small domains with one or a few macromolecules.
- Particle-based reaction-diffusion (RD) simulations are orders of magnitude more efficient than MD since they

model each molecule as one particle undergoing Brownian diffusion. The solvent effects are implicitly modeled through the Brownian noise term,[13–15] and the reactions are regulated by reaction rates. For reactions involving two or more particles, if a pair of reactive particles is close enough to each other, they can react with a certain reaction rate. They are ideal to model multiparticle processes at large length scales but lack atomic detail. There is a large amount of particle-based RD literature,[16–26] as well as several software packages and simulation schemes.[27–36]

By coupling MSMs of molecular kinetics with particle-based RD we can combine the best of both worlds and perform multiscale molecular simulations across large time and length scales; we call this coupling MSM/reaction-diffusion (MSM/RD). However, this coupling is not trivial. The existing implementation of MSM/RD in Ref. 37 suffered from several limitations: there was no underlying mathematical theory to justify and derive the coupling scheme, it was limited to simple ligand–protein systems, the protein was assumed fixed in the frame of reference, the ligand orientation and possible conformation switching was not taken into account, and multiparticle extensions were not implemented. However, the aim of MSM/RD remains the same, to produce efficient multiscale simulations that reproduce the essential statistical behavior of a practically unaffordable large-scale MD simulation by employing only statistics obtained from simulations of the constituent molecules in small solvent boxes.

In this work, we develop a general framework for MSM/RD that overcomes the previous shortcomings. It is derived by coarse-graining molecular kinetics into hybrid switching diffusion processes,[38,39] also known as diffusion processes with Markovian switching. These correspond to a class of stochastic hybrid systems, called "hybrid" due to the coexistence of continuous dynamics and discrete events in the same process. The diffusion of molecules corresponds to the continuous part, while their conformation switching corresponds to the discrete part. By discretizing the framework, we derive MSM/RD schemes; we implement and validate them for two protein–protein benchmark systems and one multiparticle implementation to model the formation of pentameric ring molecules. Implementations in more realistic systems are left for future work. However, given enough data to parameterize the model, the framework is ideal for applications on protein–ligand and protein–protein dynamics, as well as self-assembly of structures composed of several copies of the same or a small set of molecules, such as virus capsids.

Previous relevant works[40–42] introduced spatially dependent reaction rates, a fundamental concept in our framework, and[43–47] have modeled fluctuations on the reactivity of the species using Markovian gates, a special case of our framework. A related multiscale method[34,35] couples MD with Green's function reaction dynamics, including anisotropic interactions. Although still constrained by MD computations, this method could potentially be combined with our approach to accelerate both MD and particle-based RD simulations. The work[48] provides an excellent review on several multiscale methods for protein–ligand binding, including Refs. 49 and 50 where multiscale simulations are used to estimate kinetic rates. The ideas presented in this work could help enhance these methods. Moreover, Refs. 51–55 have focused on implementing several stochastic hybrid models in different fields in biology, which emphasizes the relevance of stochastic hybrid models in biological settings. Note that hybrid switching diffusions are a general coarse-grained model of MD, and they can thus be applied to many other applications beyond MSM/RD, such as the diffusion and conformation switching of molecules under a concentration or temperature gradient.

II. MOLECULAR KINETICS AS HYBRID SWITCHING DIFFUSIONS

A. One molecule

Consider a molecule A. If we fix the position and orientation of the molecule, the position of its atoms only changes due to conformational changes. We can then coarse-grain the all-atom dynamics in configuration space into an MSM.[8,10,56] Let us assume that our molecule A can be described by switching between two MSM states $A_1 \rightleftharpoons A_2$.

If molecule A is now diffusing instead of being fixed in space, we would expect different diffusion coefficients in different conformations. The diffusion and the conformation switching can be modeled together,

$$\frac{\partial}{\partial t}\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} D_1\nabla^2 p_1 \\ D_2\nabla^2 p_2 \end{bmatrix} + \begin{pmatrix} -r_{12} & r_{21} \\ r_{12} & -r_{21} \end{pmatrix}\begin{bmatrix} p_1 \\ p_2 \end{bmatrix},$$

where $p(x,t)$ is the vector of probability densities $(p_1, p_2)^T$ of being in conformation A_1 or A_2 at position x, and r_{ij} are the transition rates from conformation A_i to A_j that form the corresponding transition rate matrix. Note that the first term of the right-hand side corresponds to the Fokker–Planck equations of the diffusion processes, while the second term corresponds to a continuous-time MSM or Master-equation model.[56] We would like to incorporate rotational diffusion and generalize it to N different conformations. The resulting generalization yields

$$\frac{\partial p(x,t)}{\partial t} = \underbrace{\mathcal{D}p(x,t)}_{\text{Diffusion}} + \underbrace{\mathbb{Q}p(x,t)}_{\text{MSM}}, \tag{1}$$

where $p(x,t) = (p_1, \ldots, p_N)^T$ is the vector of probability densities of being in the corresponding conformations at x and time t, with $x = (r, \theta)$ denoting the position and orientation coordinates of the molecule. The operator $\mathcal{D}$ describes the translational and rotational diffusion of the molecule in each of its conformations. The matrix $\mathbb{Q}$ is a $N \times N$ transition rate matrix describing the conformation switching; its diagonal entries are all negative and its non-diagonal ones positive; its columns sum to zero. In this way, the equation models simultaneously the molecule's diffusion and the switching of conformation. Equation (1) is an example of a hybrid switching diffusion process, and one could also write a stochastic differential equation (SDE) for the individual stochastic trajectories. A detailed derivation of this theory is presented in the supplementary material, Appendix A. The diffusion operator and the transition rate matrix can be a function of x, $\mathcal{D}(x)$, and $\mathbb{Q}(x)$, which provides a robust framework for several interesting applications. In this work, we are interested in the interaction between two molecules, so we generalize this result for two interacting molecules.

B. Two interacting molecules

Consider two molecules A and B. If they are far enough from each other, they will not interact. Each molecule has a state vector assigned, p_A and p_B, with sizes N_A and N_B corresponding to their respective number of conformations. The conformations are denoted by A_i and B_j, with $i = 1, \ldots, N_A$ and $j = 1, \ldots, N_B$. The diffusion operators $\mathcal{D}_A$ and $\mathcal{D}_B$ encode the rotational and translational diffusion, which in the simplest case will correspond to Laplacian operators with diffusion coefficients for the different conformations, D_{A_i} and D_{B_j}. The rate matrices $\mathbb{Q}_A$ and $\mathbb{Q}_B$ encode the rates at which they switch conformation. Each molecule will satisfy its own version of Eq. (1),

$$\frac{\partial p_A}{\partial t} = \mathcal{D}_A p_A + \mathbb{Q}_A p_A, \quad \frac{\partial p_B}{\partial t} = \mathcal{D}_B p_B + \mathbb{Q}_B p_B; \tag{2}$$

see Fig. 1(a) for a graphical reference. The state of the system p_{AB} is given by all the possible combinations of states of A and states of B. This corresponds to the tensor product of all the states of A with all the states of B, i.e., $p_{AB} = p_A \otimes p_B$.[57] For instance, if A and B have two states each, A_1, A_2 and B_1, B_2, respectively, the full system given by the tensor product has four possible states: A_1B_1, A_1B_2, A_2B_1, and A_2B_2. Taking the time derivative of $p_A \otimes p_B$ and using Eq. (2), we obtain

$$\frac{\partial p_{AB}(x)}{\partial t} = \mathcal{D} p_{AB}(x) + (\mathbb{Q}_A \oplus \mathbb{Q}_B) p_{AB}(x), \tag{3}$$

where the diffusion operator is applied independently before taking the tensor product $\mathcal{D} p_{AB} = \mathcal{D}_A p_A \otimes p_B + p_A \otimes \mathcal{D}_B p_B$ and $\mathbb{Q}_A \oplus \mathbb{Q}_B = (\mathbb{Q}_A \otimes \mathbb{I}_{N_B}) + (\mathbb{I}_{N_A} \otimes \mathbb{Q}_B)$ is the Kronecker sum with $\mathbb{I}_K$ the identity matrix of order K. The appearance of the Kronecker sum results evident when computing the solution of the full system as the tensor product of the individual solutions of Eq. (2), $p_{AB}(x,t) = e^{t\mathcal{D}_A} \otimes e^{t\mathcal{D}_B} \otimes e^{t(\mathbb{Q}_A \oplus \mathbb{Q}_B)} p_{AB}(x,0)$.[57] This means that the rate matrix of the full system is given by the transition rate matrix $\mathbb{Q}_A \oplus \mathbb{Q}_B$ [Fig. 1(b)]. Note that if we were using a discrete-time MSM, the transition probability matrix of the full system would simply be the tensor product of the independent transition probability matrices.

Let us assume now that molecules A and B are close to each other and are interacting such that they can be considered as a complex C that diffuses as a single entity. The state vector is p_C with dimension N_C, corresponding to the bound conformations, C_k, with $k = 1, \ldots, N_C$. We can thus write it in the form of Eq. (2),

$$\frac{\partial p_C}{\partial t} = \mathcal{D}_C p_C + \mathbb{Q}_c p_C. \tag{4}$$

We would like to switch smoothly between the non-interacting regime [Eq. (3)] and the bound regime [Eq. (4)], so we introduce a transition regime, where the molecules are still dissociated but interacting, and the transition rates strongly depend on the relative position and orientation between the molecules. The dynamics of the system in the three regimes can be written in terms of the probability of being in any of the dissociated states (AB) [Eq. (3)] and any of the bound states (C) (4), namely, $p(x,t) = (p_{AB}, p_C)^T$, and a transition rate matrix $\mathbb{Q}(x)$ that depends on the phase space coordinates x, more specifically on the relative position and orientation between the molecules. We can write the dynamics of $p(x,t)$ as a hybrid switching diffusion process

$$\frac{\partial p(x)}{\partial t} = \mathcal{D} p(x) + \mathbb{Q}(x) p(x), \quad \mathbb{Q}(x) = \left(\begin{array}{c|c} \mathbb{Q}_{AB} & \mathbb{Q}_{C\to AB} \\ \hline \mathbb{Q}_{AB\to C} & \mathbb{Q}_C \end{array}\right). \tag{5}$$

The matrix $\mathbb{Q}_{AB\to C}$ contains the transition rates from dissociated states (AB) to bound states (C) and vice versa for the matrix $\mathbb{Q}_{C\to AB}$.

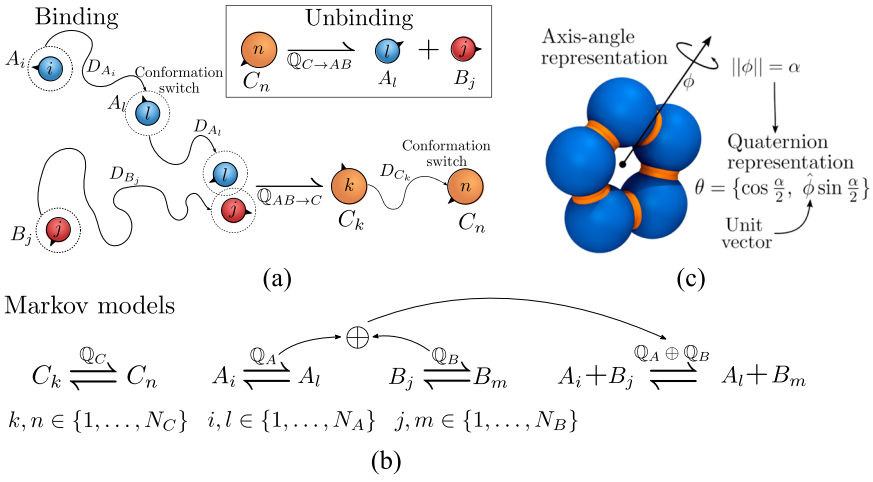

FIG. 1. Diagrams to illustrate MSM/RD theory and general rotations. (a) Diagram of the binding and unbinding of two reactive molecules, $A + B \rightleftharpoons C$, when modeling their kinetics as hybrid switching diffusions. Molecules are represented by particles with position and orientation (black pointer). The three molecules have a conformation-dependent diffusion, and the conformations are denoted by a subindex. If molecules A and B are close enough to each other, they transition to a bound compound C with a configuration-dependent rate given by $\mathbb{Q}_{AB\to C}$. The compound C can also unbind into molecules A and B with a configuration-dependent rate given by $\mathbb{Q}_{C\to AB}$. (b) Diagram showing the individual Markov models for C, A, and B and the Markov model for the joint system of molecules A and B when not interacting. (c) Orientation of a pentameric ring molecule using the axis-angle representation with the molecule's center as reference. The direction of the ϕ vector, $\hat{\phi}$, represents the axis of rotation, and its magnitude $\|\phi\| = \alpha$ represents the radians to be rotated. We can translate this to its quaternion representation.

If the initial relative distance between the molecules is large enough, the molecules are in the non-interacting regime; $\mathbb{Q}_{AB\to C}$ is zero; the system can only reach the states accessible by $\mathbb{Q}_A \oplus \mathbb{Q}_B$—so $\mathbb{Q}_{AB} = \mathbb{Q}_A \oplus \mathbb{Q}_B$—and the dynamics given by Eq. (3) are recovered. However, diffusion can bring the molecules together into the transition regime, making $\mathbb{Q}_{AB\to C}$ nonzero and allowing the system to transition to the bound regime. The system can then transition to other bound states through $\mathbb{Q}_C$, and it can transition out of the bound regime into the transition regime (dissociated) through $\mathbb{Q}_{C\to AB}$. Note that columns of $\mathbb{Q}(x)$ should sum to zero for any given x, and $\mathbb{Q}_C$ on Eq. (5) is a renormalized version of $\mathbb{Q}_c$ in Eq. (4).

Equation (5) constitutes the general MSM/RD framework, and its dynamics are represented in Figs. 1(a) and 1(b). Appendix A of the supplementary material shows a more detailed derivation of this theory. Discretizations of this model are used to generate MSM/RD schemes. In Sec. III, we derive the MSM/RD schemes used throughout this work by doing a piecewise constant discretization of $\mathbb{Q}(x)$. Their parameterization and explicit algorithms are given in the supplementary material, Appendices B and C.

C. Quaternions

The MSM/RD framework requires a representation for the orientation or rotation θ of a rigid body, such as Euler angles, rotation matrices, or unit quaternions among others. Some of these have severe disadvantages, such as the gimbal lock in Euler angles, while unit quaternions have proved to be the most simple, robust, and numerically efficient.[35,58–60] A quaternion, $\theta = \{s, p\}$, consists of a real part s and a three-dimensional vector part p. If normalized to one, $s^2 + p \cdot p = 1$, it can be used to represent a three-dimensional rotation. Let us consider first a more physically intuitive representation of rotations, the axis-angle representation, where an arbitrary rotation is represented by a three-dimensional vector ϕ [Fig. 1(c)]. Its direction $\hat{\phi} = \phi/\|\phi\|$ corresponds to the axis of rotation following the right-hand rule, and the length of the vector $\|\phi\|$ corresponds to the magnitude of rotation [Fig. 1(c)]. The corresponding quaternion associated with this rotation is

$$\theta = \{\cos(\|\phi\|/2),\ \sin(\|\phi\|/2)\hat{\phi}\}. \tag{6}$$

Similar to complex numbers, quaternions are further endowed with an algebraic structure such that the resulting rotation of consecutive rotations, θ_1 and θ_2, is obtained by an algebraic multiplication, $\theta = \theta_2\theta_1$,

$$\theta = \{s_2 s_1 - p_2 p_1,\ s_2 p_1 + s_2 p_1 + p_2 \times p_1\},$$

where the cross product makes the multiplication non-commutative, as expected for rotations. The unit quaternion $\theta^{-1} = \{s, -p\}$ is the inverse quaternion of θ representing the inverse rotation, such that $\theta\theta^{-1} = I$ is the identity rotation. Note the quaternion $-\theta$ corresponds to the same rotation as θ; therefore, it is enough to use half of the surface of the four-dimensional unit sphere to describe all possible rotations in three-dimensional space. If a one-to-one relation is desired, the simplest choice is to restrict to $s \geq 0$. More detailed accounts of quaternions can be found in the literature.[35,58,60]

III. METHODS

A. A general MSM/RD scheme

The general MSM/RD framework for two interacting molecules is condensed in Eq. (5). In most cases, we will not know the rate functions constituting Eq. (5). However, we can discretize the equation and obtain a specific MSM/RD coupling scheme, which can be parameterized with MD trajectories. Equation (5) thus provides a robust theoretical foundation from which different MSM/RD schemes can be derived by applying different discretizations; it serves as a guideline to derive different and better suited schemes for the situation at hand.

The MSM/RD schemes used throughout this work originate from piecewise constant discretizations of the transition rate matrix $\mathbb{Q}(x)$ from Eq. (5). We first divide the phase space into the three main regions/regimes: non-interacting, transition, and bound regimes [Fig. 2(a)]. The definition of these regimes will be system dependent and based on the relative position between the two molecules. As a rule of thumb, the interaction between molecules must be weak in the transition regime and effectively zero in the non-interacting regime. MSM/RD requires parameterizing two MSMs, one for the non-interacting regime, $\mathbb{Q}_A \oplus \mathbb{Q}_B$, and the other for the transition and bound regimes together. In Secs. III B, III C, and III D we will cover how the MSM/RD dynamics are constructed in each of these regions. In the supplementary material, Appendices C and B, we further show the corresponding MSM/RD algorithm and how to discretize the MD trajectories to parameterize the MSM/RD scheme.

B. Non-interacting regime

We consider two molecules A and B as rigid bodies with relative position $r_{AB} = r_B - r_A$ and relative orientation $\theta_{AB} = \theta_B\theta_A^{-1}$, where θ_A and θ_B are quaternions representing orientations. Following Eq. (2), if the two molecules are far enough apart, $\|r_{AB}\| \geq R$, they diffuse and change conformation independently. Thus, the rates of the transition matrix $\mathbb{Q}(x)$ do not depend on r_{AB} or θ_{AB}, and the dynamics of the individual molecules are discretized into individual MSMs using standard methods,[10] yielding $\mathbb{Q}_{AB} = \mathbb{Q}_A \oplus \mathbb{Q}_B$. For the sake of simplicity and without loss of generality, we assume that the particles are modeled with overdamped Langevin dynamics. The corresponding SDE is (see the supplementary material, Appendix A 1)

$$\frac{dX_k(\eta_k, t)}{dt} = \sqrt{2k_BT}M_k^{\frac{1}{2}}(\eta_k)\xi(t), \tag{7}$$

where k denotes the molecule A or B; $dX_k = [dr_k, d\Phi_k]$, with dr_k being the change in position of molecule k and $d\Phi_k$ its change of orientation in the axis-angle representation; M_k is the mobility matrix of molecule k, which depends on its conformation η_k; and $\xi(t)$ corresponds to six-dimensional Gaussian white noise. The conformation η_k of each molecule changes following a discrete- or continuous-time MSM with the constant rates from the transition matrix $\mathbb{Q}_A \oplus \mathbb{Q}_B$. Thus, η_k can be propagated by simply sampling transition probabilities in the discrete case or by using a Gillespie-type algorithm[61,62] in the continuous case. This description corresponds to the trajectory representation of the stochastic process described by Eq. (2). If the translational and rotational motions are weakly coupled and both are isotropic, we can approximate

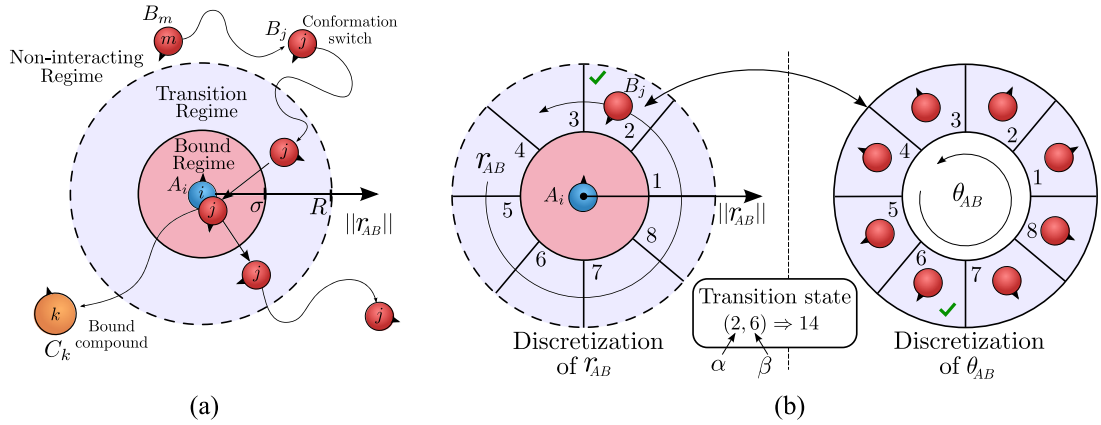

FIG. 2. Discretization diagrams for the MSM/RD scheme. (a) Diagram of the three different regimes in the MSM/RD scheme defined by $\sigma < \|r_{AB}\| < R$. The bound regime is shaded in red, the transition regime is shaded in blue, and the non-interacting regime is in white. To define these regions, we fix the frame of reference to molecule *A*. In the non-interacting regime, they both diffuse and change conformation freely. In the transition regime, they can transition to a bound compound state. From the bound state, they can unbind and switch to a specific configuration in the transition regime. From the transition regime, they can diffuse away into the non-interacting regime. Note the orientation is specified by a small black pointer attached to each particle. (b) Definition of transition states (or unbound transition states) within the transition regime. To define the transition states, we perform two discretizations: one for the relative position r_{AB} and the other for the relative orientation θ_{AB}. In this illustration, the relative position is simply a two dimensional vector, so we simply partition the blue shaded ring ($\sigma < \|r_{AB}\| < R$) from 0 to 2π into eight parts. For each discrete value of the relative position, the molecules can still have an arbitrary relative orientation, so, we also need to discretize the relative orientation. The relative orientation is represented with one degree of freedom, so we discretize it by partitioning the circle into eight parts. If α is the discretization state of r_{AB} and β the one of θ_{AB}, the transition state number is given by $(\alpha - 1)8 + \beta$. This discretization yields a total of $8 \times 8 = 64$ transition states. In each transition state, the rates are approximated by a constant value, yielding a piecewise constant approximation of the rates in $\mathbb{Q}$. In our MSM/RD implementation, an analogous discretization is done in three dimensions (six degrees of freedom).

Eq. (7) by

$$\begin{aligned} \frac{dr_k(\eta_k, t)}{dt} &= \sqrt{2D_k(\eta_k)}\xi(t), \\ \frac{d\Phi_k(\eta_k, t)}{dt} &= \sqrt{2D_k^{\text{rot}}(\eta_k)}\xi_{\text{rot}}(t), \end{aligned} \tag{8}$$

where D_k and D_k^{rot} are the translation and rotational diffusion coefficients of molecule k, and in these equations, $\xi(t)$ and $\xi_{\text{rot}}(t)$ each correspond to three-dimensional Gaussian white noise. Note that in this region, the *C* states are not accessible, so only $\mathbb{Q}_{AB}$ is relevant. The numerical discretization of this equation has the same form as Eq. (11) but with zero force and torque terms. The diffusion coefficients (or matrices in the general case) should also be estimated from MD trajectories. There are several works focused on this topic;[59,63,64] we also added a small section about it in the supplementary material, Appendix D.

C. Transition regime

The transition regime is defined by the region between the non-interacting and the bound regime, $\sigma < \|r_{AB}\| < R$. In this regime, the transition rates depend continuously on the relative position and orientation of the molecules. As we plan to infer these rates from MD simulations, it is convenient to discretize $\mathbb{Q}(x)$ into a relatively small number of transition regions/states where the rates are approximated by constant values, yielding a piecewise constant approximation of $\mathbb{Q}(x)$ that is easier to infer from MD data.

Figure 2 shows an illustration of the different regions/states and the discretization of the transition regime for a simplified lower dimensional case. For each transition state, given by the combination of a discrete value of the relative position and of the relative orientation, we approximate the rates in $\mathbb{Q}(x)$ by a constant value, yielding a piecewise constant approximation of $\mathbb{Q}_{AB}$ and $\mathbb{Q}_{AB\to C}$ in the transition regime. In this regime, the particles are always dissociated, so $\mathbb{Q}_{C\to AB}$ and $\mathbb{Q}_C$ are not relevant.

In general, the relative position and the relative orientation account for a total of six degrees of freedom, so the discretization of the transition regime is much more complex than in Fig. 2, but it still follows the same principle. The first step is to provide an equal area partition of the surface of the sphere following Ref. 65, yielding a discretization of the relative position in the transition region. Then, we need to discretize the relative orientation, which is given in terms of a unit quaternion. As unit quaternions can be projected into the top half three-dimensional unit sphere, we use the same equal area partition sphere with a few additional cuts along the radial direction, yielding an effective discretization of all the possible relative orientations. It is important to keep the number of divisions in these partitions as small as possible to avoid an exploding number of transition states.

Note that conformation switching within the transition regime is naturally incorporated in the framework. The transition matrix $\mathbb{Q}(x)$ acts on $p(x, t) = (p_{AB}, p_C)^T$, where p_{AB} includes one entry for every possible conformation combination between the two molecules. Thus, the discretization of $\mathbb{Q}(x)$ includes the rates corresponding to conformation changes within the transition regime. Alternatively, by collapsing all conformations into one state in the parameterization, one can obtain averaged rates over all conformations for all the transitions within the transition regime.

In the transition regime, the diffusion of the particles—approximated by Eq. (7)—and the propagation of the MSM—following $\mathbb{Q}_{AB}$ and $\mathbb{Q}_{AB\to C}$—are run in parallel. If $\|r_{AB}\|$ becomes larger than R due to diffusion, the MSM is ignored and the dynamics switch to the non-interacting regime. If a binding event occurs, the diffusion of the binding particle is ignored and the dynamics switch to the bound regime.

D. Bound regime

If molecules A and B are close enough to each other, $\|r_{AB}\| \le \sigma$, they are strongly interacting and can be considered as a bound compound C with several metastable configurations. In this case, their diffusion and conformation switching are no longer independent, and the transition rates do not depend on r_{AB} and θ_{AB}, so they are assumed constant. The transitions in the bound regime can be between metastable states (following $\mathbb{Q}_C$) or toward an unbound state in the transition regime (following $\mathbb{Q}_{C\to AB}$). Analogously to the previous example, we assume without loss of generality that the dynamics of the compound follow overdamped Langevin dynamics,

$$\frac{dX_C(\eta_C,t)}{dt} = \sqrt{2k_BT}M_C^{\frac{1}{2}}(\eta_C)\xi(t), \tag{9}$$

with $dX_C = [dr_C, d\Phi_C]$. The conformation η_C is propagated using the rates from $\mathbb{Q}_C$ and $\mathbb{Q}_{C\to AB}$. The dynamics are propagated in the same way as Eq. (7). If the translational and rotational motion are weakly coupled and isotropic, we can obtain analogous results to that of Eq. (8) with $k = C$ and with analogous numerical discretization and diffusion coefficient estimation. If a transition toward a dissociated state in the transition regime happens, the dynamics switch to the transition regime. In the bound regime, particles are always bound, so $\mathbb{Q}_{AB\to C}$ and $\mathbb{Q}_{AB}$ are not relevant. Note that when parameterizing $\mathbb{Q}(x)$ from MD data, we obtain one MSM at once for both the transition and bound regimes (see the supplementary material, Appendix B), which describes all states in which A and B are interacting, including strongly and weakly bound states, intermediates between unbound and bound states, and even dissociated states in which A and B are sufficiently close to induce a force upon each other. Such MSMs have, for example, been computed for protein–ligand and protein–protein association in the past few years.[66–69]

E. Benchmark MD model: Patchy particles

To validate MSM/RD schemes, we require an inexpensive model of molecules capable of representing complex behavior observed in realistic MD systems, such as translational and rotational diffusion, position and orientation-dependent pair interactions, orientation-dependent binding, multiple binding sites, and conformation switching. We can construct such a model based on patchy particles.[70–72] We model molecules as diffusive spherical particles with an isotropic repulsive potential $U_{\text{isotropic}}$ to avoid overlapping; an attractive isotropic part can also be incorporated. Patches are then placed on the surface of the particles, and each patch produces a short-range attractive potential with patches from other particles, generating translational and rotational motion.[35] The potential energy between patch i of particle A and patch j of particle B can be decomposed into two parts. The first part U_r^{ij} depends only on the relative distance between the patches, r_{ij}, and the types of the patches. It corresponds to an attractive force that pulls patches together. The second part U_θ^{ij} depends on the relative orientation, θ_{AB}, of the particles, and it is activated if two patches are close enough to each other. This will favor specific relative orientations for the different bindings between patches. In all the models used for this work, the overall interaction potential between two particles, A and B, can be written in the following form:

$$U_{AB} = U_{\text{isotropic}}(r_{AB}) + \sum_{i,j=1}^{N_A,N_B}\left(U_r^{ij}(r_{ij}) + U_\theta^{ij}(r_{ij},\theta_{AB})\right),$$

where i runs over the patches of particle A and j runs over the patches of particle B. N_A and N_B are the total number of patches of A and B, respectively. In general, particles A and B can both have conformational changes; each combination of conformations is allowed to have a completely different potential energy. In this work, conformation changes will correspond to turning on and off specific patches. Figure 3 shows the potential between a pair of patchy particles with one patch, as well as examples of orientation-dependent potentials.

The position of the particles is simply given by the coordinates of the center of the sphere $r(t)$, and their orientation $\theta(t)$ is given in terms of quaternions. In order to model the translational and orientational diffusion of one particle, we use overdamped Langevin dynamics. We assume that the translational and rotational diffusions are independent and both isotropic, so we obtain

$$\begin{aligned}\frac{dr(t,\eta)}{dt} &= \frac{1}{\gamma}F(r,\eta) + \sqrt{2D(\eta)}\xi(t),\\ \frac{d\Phi(\eta,t)}{dt} &= \frac{1}{\gamma_{\text{rot}}}\mathrm{T}(\theta,\eta) + \sqrt{2D^{\text{rot}}(\eta)}\xi_{\text{rot}}(t),\end{aligned} \tag{10}$$

where Φ is the orientation in the axis-angle representation, γ and γ_{rot} are the translational and rotational damping coefficients, F and T the force and torque due to pair interactions and external fields, η is the conformation of the particle, and $\xi(t)$ and $\xi_{\text{rot}}(t)$ each correspond to three-dimensional Gaussian white noise. The force can be rewritten in terms of the potential as $F = -\nabla U$; the torque is convenient to leave explicitly since it is not trivial to write a potential in terms of axis-angle variables or quaternions. These two equations can be discretized using the Euler–Maruyama scheme[73] using a time step δt,

$$\begin{aligned}r(t+\delta t,\eta) &= r(t,\eta) - \frac{\delta t}{\gamma}\nabla U(r,\eta) + \sqrt{2D(\eta)\delta t}\mathcal{N}(0,1),\\ d\Phi(t,\eta) &= \frac{\delta t}{\gamma_{\text{rot}}}\mathrm{T}(\theta,\eta) + \sqrt{2D^{\text{rot}}(\eta)\delta t}\mathcal{N}(0,1).\end{aligned} \tag{11}$$

The rotation represented by the change in axis-angle $d\Phi(t)$ can be rewritten as a quaternion $d\theta(t)$ using Eq. (6). The new orientation is simply given by the quaternion product $\theta(t+\delta t) = d\theta(t)\theta(t)$. In each case, $\mathcal{N}(0,1)$ represents an independent three-dimensional vector with each entry a normal random variable with mean zero and variance 1. As the diffusion coefficients depend on the conformation, it is convenient to assume that the switching of conformation η is modeled with an MSM using a fixed lag-time τ that is a multiple of $\tau = n\delta t$, with n being a positive integer. This is not strictly

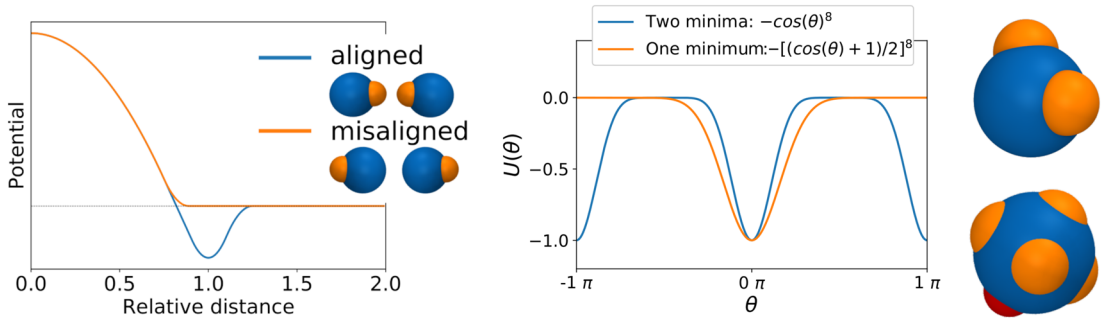

FIG. 3. Illustration of the patchy particle potential. (a) This plot shows the patchy particle potential between two particles with a diameter of one, each with one patch. The potential is plotted as a function of the relative distance between the two particles for orientations corresponding to aligned or misaligned patches. If aligned, we observe a stable minimum in the potential corresponding to particles binding. If misaligned, there is no stable minimum, and the isotropic repulsion prevents overlapping. See Ref. 35 for the specific form of the potential. (b) Two examples of angular potentials used in this work, corresponding to one and two metastable orientations. As the bindings between patches already fix two orientational degrees of freedom, we only require a one dimensional angular potential to completely fix the orientation. (c) Examples of patchy particles with two and six patches. The patches can also be of different types corresponding to different interaction potentials, and they can be turned on and off depending on the current conformation.

required, but it simplifies the implementation since the conformation change occurs always at the end of a time step. The forces and torques are calculated directly from potential energies like the ones shown in Fig. 3.

This model satisfies all the requirements we established at the beginning of this section. It can be generalized to non-isotropic and coupled rotational and translational dynamics,[58,72] and molecules can even be modeled by multiple overlapping beads with reaction patches.[72]

IV. RESULTS

To test and verify MSM/RD, we construct an MD benchmark model of molecules that is simple enough such that we can produce a large amount of data but complex enough so it models complex behavior observed in realistic MD systems. This model is based on patchy particles;[70–72] we model molecules as spherical particles with isotropic diffusion and an isotropic repulsive potential to avoid overlapping. Patches are then placed on the surface of the particles, and each patch produces a short-range configuration-dependent attractive potential with patches from other particles, generating translational and rotational motion, see Sec. III E for details. This model is the basis for all the benchmark models in this section.

In the following examples, it is not necessary to parameterize the diffusion operator since it is the same for both the benchmark and the MSM/RD simulation. This serves to isolate the dynamics of $\mathbb{Q}(x)$ and validate the coupling mechanism. For general protein–protein systems, one needs to extract the diffusion coefficients/matrices from the MD data using well documented methods[59,63] (see the supplementary material, Appendix D).

A. MSM/RD for protein–protein systems

The benchmark model consists of two molecules, A and B [Fig. 4(a)], represented by different patchy particles. Particle A has only one conformation and six binding patches: five of them have the same attraction potential (yellow) and the other one has a stronger attraction potential (red). Particle B has two conformations. In one conformation (B), it has one binding patch (red), and in the other one (B^*), the patch is turned off and it cannot bind. Each binding allows only one metastable relative orientation, yielding a total of six possible bound states. The diffusion of B depends on its conformation, and it is visualized as a three-dimensional asterisk to distinguish its orientation. We illustrate an MSM/RD trajectory of the system on Fig. 4(a).

To parameterize the MSM/RD scheme, we simulate the benchmark MD model with specific settings to mimic a common MD simulation. We assume that both molecules have a diameter of 5 nm, which is a typical size for a real protein. We simulate using periodic boundary conditions and a cube with an edge length of 25 nm as the unit cell. Each simulation runs for 6×10^6 time steps of 1×10^{-5} μs each, yielding a total simulation time of 60 μs. We run 600 of these simulations independently, and we use them to parameterize the MSM/RD scheme following the steps illustrated in the supplementary material, Appendix B.

In Fig. 4, we compare the MSM/RD results against the MD benchmark. We calculate the first passage times (FPTs) of a given transition by running the benchmark and the MSM/RD simulations in equal conditions, and we run the same number of FPT samples for each model. Figure 4(b) compares the FPT distribution from the unbound state to any bound states and vice versa. The left panel of Fig. 4(c) compares the FPT distributions for all the possible transitions between bound states. Note that these transitions include pathways that start at a bound state, unbind completely, and end in another bound state, so it is ideal to evaluate if the MSM/RD produces an accurate coupling. The right panel of Fig. 4(c) shows the MSM/RD scheme percent error for the transition rates between bound states. Overall, MSM/RD can reproduce the dynamics of the MD benchmark with good accuracy.

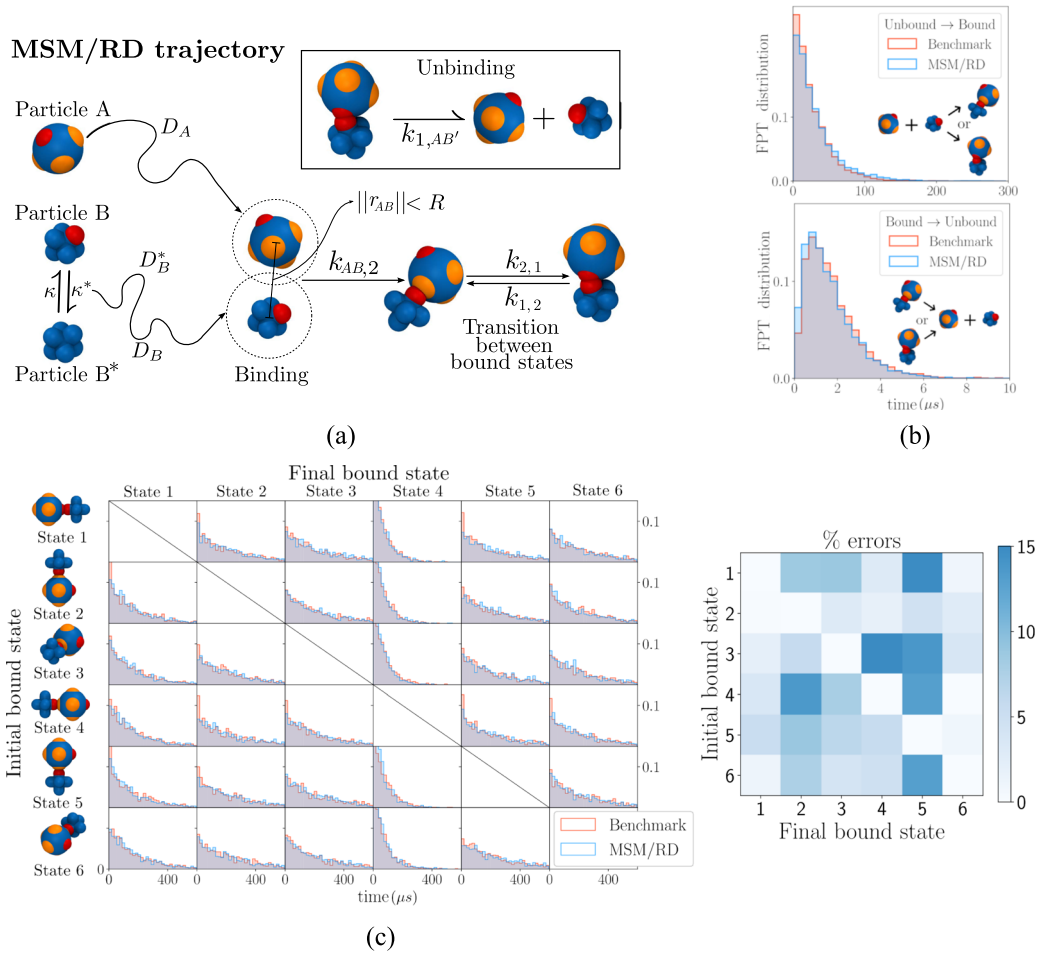

FIG. 4. Illustrations and results of the MSM/RD scheme for the protein–protein system. (a) MSM/RD sample trajectory. Particle A diffuses with coefficient D_A and particle B with coefficient D_B or D_B^* depending on its conformation. If the relative distance satisfies $\|r_{AB}\| < R$, we switch from the non-interacting to the transition regime. Here, the particles can transition to one of the six bound states with rate $k_{AB,n}$ that depends on their relative configuration $x_{AB} = (r_{AB}, \theta_{AB})$ and the final bound state $n = 1, \ldots, 6$. From a bound state n, the compound can transition to another bound state m with rate $\kappa_{n,m}$, or it can unbind to another relative configuration x'_{AB} with rate $k_{n,AB'}$. MSM/RD provides a piecewise constant approximation of all the configuration-dependent rates. (b) Comparisons of first passage time (FPT) distributions between the benchmark and MSM/RD from an unbound state to any bound state and vice versa. Each distribution was calculated with 5000 simulations. (c) Comparison of FPT distributions for all transitions between the six possible bound states, each calculated over 1000 simulations. The blue grid shows the relative error of the corresponding rates, $\kappa_{n,m}$, calculated as the inverse mean first passage time (MFPT). The average percentage error is of 5%, while the maximum is of 16%.

B. MSM/RD for dimer of two-patch particle

The benchmark model consists of two identical molecules, each with two equally strong binding patches. The molecules can bind together through any of their two binding sites. Unlike the previous example, we allow for two metastable relative orientations per patch binding, allowing for a conformation change in the bound configuration.

As the molecules have two patches each, they can bind in four different ways; each of these has two stable relative orientations, so this system has a total of eight metastable bound states. However, as they are all identical, many of these eight states are indistinguishable from each other and can all be collapsed into two functional states, A and B. This is depicted graphically in Fig. 5(a). Nonetheless, note that each of these metastable states corresponds to a different relative position and orientation between the molecules. We parameterize the MSM/RD scheme with the same setup as in the protein–protein system example (see the supplementary material, Appendix B).

In Fig. 5, we compare MSM/RD results against the MD benchmark. We calculate FPTs for both the MD benchmark and the MSM/RD simulations in equal conditions. In Fig. 5(b), we show the binding rates as a function of concentration, where each binding rate is calculated as the inverse of the mean first passage time (MFPT). The concentration is adjusted by changing the edge length of the simulation box, starting at 30 nm and increasing 5 nm for each point. In Fig. 5(c), we compare the FPT distributions for several relevant transitions. Even when the original simulations to parameterize the

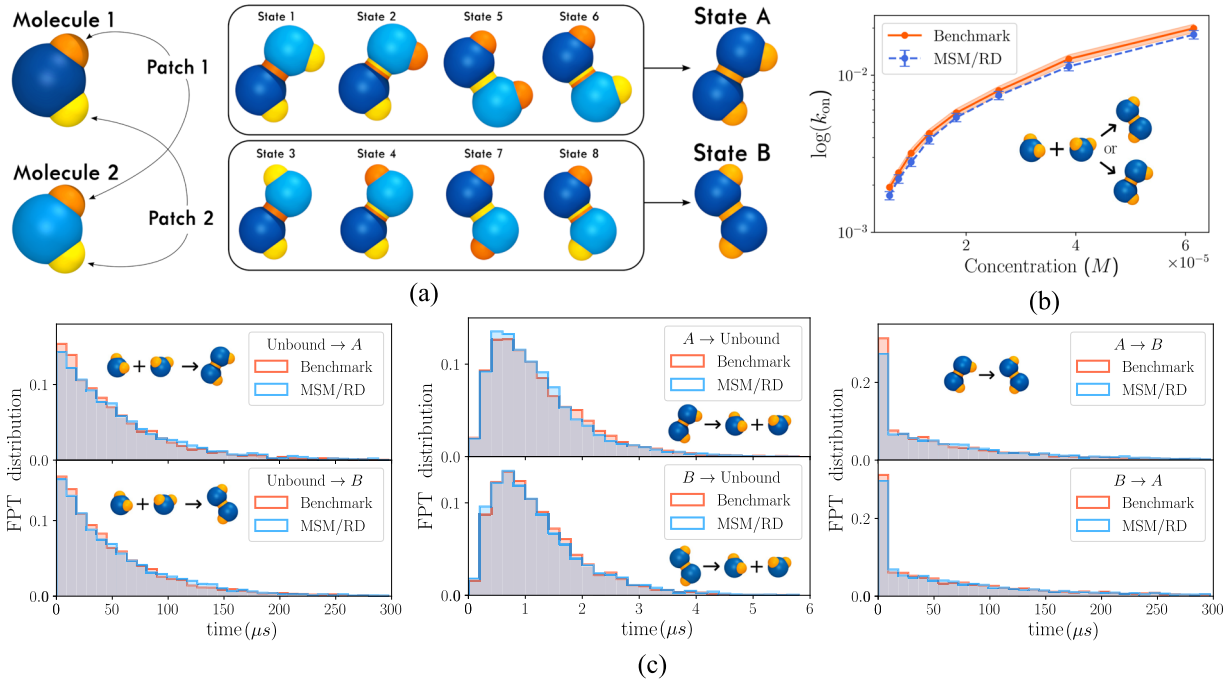

FIG. 5. Illustrations and results of the MSM/RD implementation for two identical interacting molecules with two interacting patches each and two stable angular configurations. (a) The two molecules can bind in eight different ways (eight bound states). For illustration purposes, particle one is shown in dark blue and particle two in light blue; the first patch is shown in orange and the second one in yellow. All these states collapse into two bound states: *A* and *B*. (b) Comparison of the on-rate, transition from unbound to either A or B state, for different molar concentrations. Each point was calculated as the inverse of the MFPT obtained from 5000 simulations; the error bars represent the standard deviation over 2000 bootstrapped samples. Note that in the generation of the MSM faster time scales are neglected; therefore, it is expected that MSM/RD produces slightly slower results than the benchmark. (c) Comparisons of the FPT distributions obtained with MSM/RD and the benchmark for six cases: from the unbound state to the two bound states and vice versa and between the bound states. Each distribution was computed using 5000 simulations. These are shown next to each graph and they are all in μs. Note that in the last two histograms, there is a time-scale separation. This corresponds to the difference between direct transitions between the bound states and transitions that first unbound and later rebound in a different bound state.

scheme ran for only 30 μs, the MSM/RD scheme produces excellent results for transitions with higher MFPTs.

C. Multiparticle MSM/RD: Formation of pentameric ring

We develop and implement the first multiparticle MSM/RD scheme to study the formation of pentameric ring molecules (inspired by Ref. 70). The benchmark MD model is a modified version of the two-patch dimer model. It consists again of two identical molecules, each with two equally strong binding sites. Unlike the previous example, we only allow one metastable relative orientation per patch binding. We further increase the binding strength such that unbinding events are very rare and not observed in the time scales of interest. The particles can bind with each other forming chains, which eventually can close forming trimeric, tetrameric, or pentameric ring structures [Fig. 6(a)].

We parameterize the MSM/RD scheme with the same setup as in the protein–protein system example (see the supplementary material, Appendix B). The multiparticle MSM/RD scheme requires modifications to the two-particle MSM/RD algorithm. These modifications are shown in the supplementary material, Appendix C. We further need to estimate the diffusion coefficients of the multiparticle chains. We employ standard methods to estimate them (see the supplementary material, Appendix D).

In Fig. 6(b), we compare MSM/RD results against the MD benchmark. We calculate FPTs for the formation of all the three ring molecules, using both the MD benchmark and the MSM/RD simulations in equal conditions: five particles with random positions and orientations placed in a simulation box of edge length of 30 nm with periodic boundaries. In Fig. 5(c), we show the rate of formation of pentameric rings for different concentrations by changing the simulation box size. In Table I, we show the exact values and relative errors of the rates plotted in Fig. 6(c).

Note that MSM/RD is not as good at approximating the formation of trimeric and tetrameric rings [Fig. 6(b)]. This is due to MSM/RD modeling the particle chains in steps 3 and 4 of Fig. 6(a) as a fixed structure, while in the MD benchmark the chain is flexible, allowing for patches to get closer together, which increases the rate at which the ring is closed. This could be fixed by using a new MSM to describe the dynamics between the three or four particle chains and an additional particle. Nonetheless, note that the rates of formation of pentameric rings are not affected by these problem since they are conditioned on not having trimeric or tetrameric rings forming beforehand. In its current formulation, MSM/RD

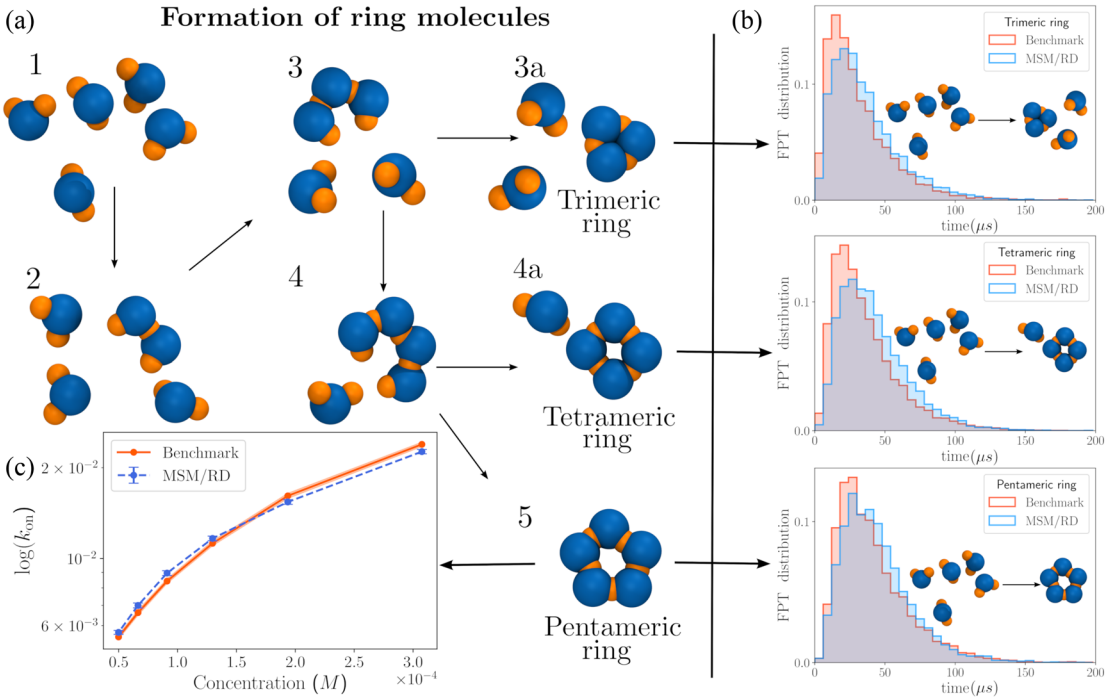

FIG. 6. Illustrations and multiparticle MSM/RD results for the formation of ring molecules. (a) Diagram showing the formation of the trimeric, tetrameric, and pentameric ring molecules. (b) Comparisons of the FPT distributions obtained with multiparticle MSM/RD and the benchmark for the formation of the trimeric, tetrameric, and pentameric ring molecules. The results were obtained from 5000 simulations for each case. (c) Comparison of the rate at which a pentameric ring is generated for different molar concentrations. Each point was calculated as the inverse of the MFPT obtained from 1000 simulations; the error bars represent the standard deviation over 500 bootstrapped samples.

TABLE I. Comparison between the MD benchmark and MSM/RD of the rates of formation of pentameric rings for different concentrations. The rates were calculated as the inverse of the MFPTs averaged over 1000 simulations; the uncertainties represent the standard deviation over 100 bootstrapped samples.

Rate (ms^{-1})	Concentration	MD benchmark	MSM/RD	Error (%)
k_{on}	$3.08 \cdot 10^{-4} M$	23.96 ± 0.44	22.67 ± 0.41	5.4
	$1.94 \cdot 10^{-4} M$	16.16 ± 0.33	15.43 ± 0.31	4.5
	$1.30 \cdot 10^{-4} M$	11.27 ± 0.23	11.72 ± 0.21	3.4
	$9.11 \cdot 10^{-5} M$	8.42 ± 0.15	8.95 ± 0.16	6.3
	$6.64 \cdot 10^{-5} M$	6.62 ± 0.13	6.99 ± 0.13	5.7
	$4.99 \cdot 10^{-5} M$	5.50 ± 0.09	5.57 ± 0.08	3.6

multiparticle implementations are limited to non-crowded environments since only pair interactions are parameterized. It is important to take these issues into account when implementing MSM/RD applications.

V. DISCUSSION

We presented a coarse-grained model of molecular kinetics based on hybrid switching diffusions. With this model, we developed a robust framework for coupling Markov models of molecular kinetics with particle-based reaction diffusion (MSM/RD). Based on this framework, we derived one possible MSM/RD scheme by discretizing the underlying equation [Eq. (5)], generalizing previous approaches.[37] We implemented and verified it for three benchmark systems: the first two involve two protein–protein systems, while the third one is a multiparticle system to model the formation of pentameric molecules. We obtained an excellent agreement between the FPT distributions and reaction rates of relevant transitions.

The framework is well-suited to model protein–ligand binding in large domains and time scales as in the previous work.[37] Given enough data for the parameterization, it is also suited to model protein–protein dynamics since it incorporates arbitrary orientations, conformation switching, and multiple binding sites. To parameterize the MSM/RD scheme for protein–protein systems, we

would need the MD data of the two proteins interacting and individually, both in small simulation boxes. The interacting proteins data would serve to parameterize the scheme in the bound and transition regime, similar to the works[66–69] with the addition of the transition states, which might require a slightly larger box. The individual molecules data would serve to parameterize the scheme in the non-interacting regime. The resulting MSM/RD scheme could run simulations at much larger time and length scales than those allowed by MD.

The multiparticle implementation of MSM/RD has promising applications to the study of self-assembly of structures composed of several copies of the same molecule (or a small set of molecules), such as virus capsids[74] or soft matter self-assembly. This setting is ideal since we only need MD data of one pair (or a few pairs) of molecules in a small simulation box to parameterize an MSM/RD multiparticle simulation, which could potentially model the formation of the full capsid.

The main caveat of MSM/RD is that the parameterization requires a large amount of MD data, which is not yet possible to obtain for most systems of interest. However, given the increasing computational power, more and more systems will soon be within reach of MSM/RD. The scheme might also become less effective in the presence of long-ranged interactions, though it might be possible to incorporate them into the dynamics of the non-interacting-regime using coarse-grained potentials.[75,76] Finally, in its current form, the MSM/RD multiparticle implementation only takes into account pair interactions, and thus the scheme is not yet adequate for crowded multi-molecular environments.

Although application-dependent, one can expect MSM/RD to reduce computational cost by several orders of magnitude in comparison to MD. MD simulations propagate the position and velocity of every atom, which corresponds to several thousands of degrees of freedom in an average protein–protein simulation. MSM/RD only propagates two independent Brownian bodies together with an MSM. This corresponds to, at most, 14 degrees of freedom, six for the position/orientation and one for the MSM (per molecule). Finally, considering that MSM/RD can operate in larger domains with larger time steps, equivalent MD simulations would need to increase dramatically the number of solvent molecules yielding an exploding number of degrees of freedom, while still limited to small time steps.

A. Software

To enable reproducibility and implementation of this work, we developed the MSM/RD software package, a C++/python package. All the code and software developed for this work are open source and available under an MIT license in github.com/markovmodel/msmrd2 and Zenodo.[77] The data used in this work was produced using the MSM/RD software.

SUPPLEMENTARY MATERIAL

See the supplementary material for detailed derivations and explanations of the theory and the parameterization of MSM/RD schemes, as well as the specific algorithms that are referred to in the main text.

AUTHORS' CONTRIBUTIONS

M.J.R., C.S., and F.N. designed the research; M.J.R. and M.D. performed the research; and M.J.R. and F.N. wrote the paper.

ACKNOWLEDGMENTS

We acknowledge support from the European Commission (Grant No. ERC CoG 772230), the German Ministry for Education and Research (Berlin Institute for the Foundations of Learning and Data BIFOLD), the Deutsche Forschungsgemeinschaft (Grant Nos. SFB1114/C03 and TRR186/A12), the Berlin Mathematics Research Center Math+ (Project No. AA1-6), and the Dutch Institute for Emergent Phenomena at the University of Amsterdam. M.J.d.R. thanks Hong Qian for helpful discussions over the course of this work. We also thank two anonymous reviewers who greatly improved the presentation of this work.

The authors declare no conflict of interest.

DATA AVAILABILITY

The data and scripts to produce the plots in this work are available in Zenodo.[78] The complete datasets that support the findings of this study are available from the corresponding authors upon reasonable request.

REFERENCES

[1] R. C. Bernardi, M. C. R. Melo, and K. Schulten, "Enhanced sampling techniques in molecular dynamics simulations of biological systems," Biochim. Biophys. Acta, Gen. Subj. **1850**, 872–877 (2015).
[2] R. A. Bradshaw and E. A. Dennis, *Handbook of Cell Signaling* (Academic Press, 2009).
[3] J. T. Hancock, *Cell Signalling* (Oxford University Press, 2017).
[4] W. Lim, B. Mayer, and T. Pawson, *Cell Signaling* (Taylor & Francis, 2014).
[5] "An introduction to Markov state models and their application to long timescale molecular simulation," in *Advances in Experimental Medicine and Biology*, edited by G. R. Bowman, V. S. Pande, and F. Noé (Springer Heidelberg, 2014), Vol. 797.
[6] J. D. Chodera and F. Noé, "Markov state models of biomolecular conformational dynamics," Curr. Opin. Struct. Biol. **25**, 135–144 (2014).
[7] S. Doerr, M. J. Harvey, F. Noé, and G. De Fabritiis, "HTMD: High-throughput molecular dynamics for molecular discovery," J. Chem. Theory Comput. **12**, 1845–1852 (2016).
[8] B. E. Husic and V. S. Pande, "Markov state models: From an art to a science," J. Am. Chem. Soc. **140**, 2386–2396 (2018).
[9] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, "Constructing the full ensemble of folding pathways from short off-equilibrium simulations," Proc. Natl. Acad. Sci. U. S. A. **106**, 19011–19016 (2009).
[10] J.-H. Prinz *et al.*, "Markov models of molecular kinetics: Generation and validation," J. Chem. Phys. **134**, 174105 (2011).
[11] A. Mardt, L. Pasquali, H. Wu, and F. Noé, "VAMPnets for deep learning of molecular kinetics," Nat. Commun. **9**, 5 (2018).
[12] H. Wu, A. Mardt, L. Pasquali, and F. Noe, "Deep generative Markov state models," in Proceedings of the 32nd International Conference on Neural Information Processing Systems, December 2018 (2018), pp. 3979–3988.
[13] P. Hänggi, P. Talkner, and M. Borkovec, "Reaction-rate theory: Fifty years after Kramers," Rev. Mod. Phys. **62**, 251 (1990).
[14] P. Mereghetti, D. Kokh, J. A. McCammon, and R. C. Wade, "Diffusion and association processes in biological systems: Theory, computation and experiment," BMC Biophys. **4**, 2 (2011).
[15] J. Schöneberg, A. Ullrich, and F. Noé, "Simulation tools for particle-based reaction-diffusion dynamics in continuous space," BMC Biophys. **7**, 11 (2014).

[16]N. Agmon and A. Szabo, "Theory of reversible diffusion-influenced reactions," J. Chem. Phys. **92**, 5270–5284 (1990).
[17]M. J. del Razo, W. Pan, H. Qian, and G. Lin, "Fluorescence correlation spectroscopy and nonlinear stochastic reaction–diffusion," J. Phys. Chem. B **118**, 7037–7046 (2014).
[18]M. J. del Razo and H. Qian, "A discrete stochastic formulation for reversible bimolecular reactions via diffusion encounter," Commun. Math. Sci. **14**, 1741–1772 (2016).
[19]M. J. del Razo, H. Qian, and F. Noé, "Grand canonical diffusion-influenced reactions: A stochastic theory with applications to multiscale reaction-diffusion simulations," J. Chem. Phys. **149**, 044102 (2018).
[20]M. Dibak, C. Fröhner, F. Noé, and F. Höfling, "Diffusion-influenced reaction rates in the presence of pair interactions," J. Chem. Phys. **151**, 164105 (2019).
[21]C. Fröhner and F. Noé, "Reversible interacting-particle reaction dynamics," J. Phys. Chem. B **122**, 11240–11250 (2018).
[22]M. Kostré, C. Schütte, F. Noé, and M. J. del Razo, "Coupling particle-based reaction-diffusion simulations with reservoirs mediated by reaction-diffusion PDES," SIAM Multiscale Model. Simul. (to be published0 (2021).
[23]A. Szabo, K. Schulten, and Z. Schulten, "First passage time approach to diffusion controlled reactions," J. Chem. Phys. **72**, 4350–4357 (1980).
[24]D. Shoup and A. Szabo, "Role of diffusion in ligand binding to macromolecules and cell-bound receptors," Biophys. J. **40**, 33–39 (1982).
[25]Z. Schuss, A. Singer, and D. Holcman, "The narrow escape problem for diffusion in cellular microdomains," Proc. Natl. Acad. Sci. U. S. A. **104**, 16098–16103 (2007).
[26]A. Szabo, "Theory of diffusion-influenced fluorescence quenching," J. Phys. Chem. **93**, 6929–6939 (1989).
[27]S. S. Andrews and D. Bray, "Stochastic simulation of chemical reactions with spatial resolution and single molecule detail," Phys. Biol. **1**, 137 (2004).
[28]A. Donev *et al.*, "A first-passage kinetic Monte Carlo algorithm for complex diffusion–reaction systems," J. Comput. Phys. **229**, 3214–3236 (2010).
[29]B. Drawert, S. Engblom, and A. Hellander, "URDME: A modular framework for stochastic simulation of reaction-transport processes in complex geometries," BMC Syst. Biol. **6**, 76 (2012).
[30]M. Hoffmann, C. Fröhner, and F. Noé, "ReaDDy 2: Fast and flexible software framework for interacting-particle reaction dynamics," PLoS Comput. Biol. **15**, e1006830 (2019).
[31]K. Takahashi, S. Tanase-Nicola, and P. R. ten Wolde, "Spatio-temporal correlations can drastically change the response of a MAPK pathway," Proc. Natl. Acad. Sci. U. S. A. **107**, 2473–2478 (2010).
[32]S. Wils and E. De Schutter, "STEPS: Modeling and simulating complex reaction-diffusion systems with Python," Front. Neuroinf. **3**, 15 (2009).
[33]J. S. van Zon and P. R. ten Wolde, "Green's-function reaction dynamics: A particle-based approach for simulating biochemical networks in time and space," J. Chem. Phys. **123**, 234910 (2005).
[34]A. Vijaykumar, P. G. Bolhuis, and P. R. ten Wolde, "Combining molecular dynamics with mesoscopic Green's function reaction dynamics simulations," J. Chem. Phys. **143**, 214102 (2015).
[35]A. Vijaykumar, T. E. Ouldridge, P. R. ten Wolde, and P. G. Bolhuis, "Multiscale simulations of anisotropic particles combining molecular dynamics and Green's function reaction dynamics," J. Chem. Phys. **146**, 114106 (2017).
[36]J. S. van Zon and P. R. ten Wolde, "Simulating biochemical networks at the particle level in time and space: Green's function reaction dynamics," Phys. Rev. Lett. **94**, 128103 (2005).
[37]M. Dibak, M. J. del Razo, D. De Sancho, C. Schütte, and F. Noé, "MSM/RD: Coupling Markov state models of molecular kinetics with reaction-diffusion simulations," J. Chem. Phys. **148**, 214107 (2018).
[38]X. Mao and C. Yuan, *Stochastic Differential Equations with Markovian Switching* (Imperial College Press, 2006).
[39]G. Yin and C. Zhu, *Hybrid Switching Diffusions: Properties and Applications* (Springer New York, 2010), Vol. 63.
[40]J. Keizer, "Diffusion effects on rapid bimolecular chemical reactions," Chem. Rev. **87**, 167–180 (1987).
[41]J. Keizer, "Nonequilibrium statistical thermodynamics and the effect of diffusion on chemical reaction rates," J. Phys. Chem. **86**, 5052–5067 (1982).
[42]J. Keizer, "Theory of rapid bimolecular reactions in solution and membranes," Acc. Chem. Res. **18**, 235–241 (1985).
[43]I. V. Gopich, K. M. Solntsev, and N. Agmon, "Excited-state reversible geminate reaction. I. Two different lifetimes," J. Chem. Phys. **110**, 2164–2174 (1999).
[44]I. V. Gopich and A. Szabo, "Kinetics of reversible diffusion influenced reactions: The self-consistent relaxation time approximation," J. Chem. Phys. **117**, 507–517 (2002).
[45]I. V. Gopich and A. Szabo, "Reversible stochastically gated diffusion-influenced reactions," J. Phys. Chem. B **120**, 8080–8089 (2016).
[46]A. V. Popov, N. Agmon, I. V. Gopich, and A. Szabo, "Influence of diffusion on the kinetics of excited-state association–dissociation reactions: Comparison of theory and simulation," J. Chem. Phys. **120**, 6111–6116 (2004).
[47]A. Szabo, D. Shoup, S. H. Northrup, and J. A. McCammon, "Stochastically gated diffusion-influenced reactions," J. Chem. Phys. **77**, 4484–4493 (1982).
[48]B. R. Jagger, S. E. Kochanek, S. Haldar, R. E. Amaro, and A. J. Mulholland, "Multiscale simulation approaches to modeling drug–protein binding," Curr. Opin. Struct. Biol. **61**, 213–221 (2020).
[49]B. R. Jagger, C. T. Lee, and R. E. Amaro, "Quantitative ranking of ligand binding kinetics with a multiscale milestoning simulation approach," J. Phys. Chem. Lett. **9**, 4941–4948 (2018).
[50]L. W. Votapka, B. R. Jagger, A. L. Heyneman, and R. E. Amaro, "SEEKR: Simulation enabled estimation of kinetic rates, a computational tool to estimate molecular kinetics and its application to trypsin–benzamidine binding," J. Phys. Chem. B **121**, 3597–3606 (2017).
[51]P. C. Bressloff and S. D. Lawley, "Stochastically gated diffusion-limited reactions for a small target in a bounded domain," Phys. Rev. E **92**, 062117 (2015).
[52]P. C. Bressloff, "Stochastic switching in biology: From genotype to phenotype," J. Phys. A: Math. Theor. **50**, 133001 (2017).
[53]P. C. Bressloff and S. D. Lawley, "Hybrid colored noise process with space-dependent switching rates," Phys. Rev. E **96**, 012129 (2017).
[54]P. C. Bressloff and J. N. Maclaurin, "Stochastic hybrid systems in cellular neuroscience," J. Math. Neurosci. **8**, 12 (2018).
[55]P. C. Bressloff, S. D. Lawley, and P. Murphy, "Protein concentration gradients and switching diffusions," Phys. Rev. E **99**, 032409 (2019).
[56]N.-V. Buchete and G. Hummer, "Coarse master equations for peptide folding dynamics," J. Phys. Chem. B **112**, 6057–6069 (2008).
[57]T. Hempel *et al.*, "Independent Markov decomposition: Towards modeling kinetics of biomolecular complexes," Proc. Natl. Acad. Sci. U. S. A. **118**, e2105230118 (2021).
[58]S. Delong, F. Balboa Usabiaga, and A. Donev, "Brownian dynamics of confined rigid bodies," J. Chem. Phys. **143**, 144107 (2015).
[59]M. Linke, J. Köfinger, and G. Hummer, "Fully anisotropic rotational diffusion tensor from molecular dynamics simulations," J. Phys. Chem. B **122**, 5630–5639 (2018).
[60]D. C. Rapaport, "Molecular dynamics simulation using quaternions," J. Comput. Phys. **60**, 306–314 (1985).
[61]D. F. Anderson and T. G. Kurtz, *Stochastic Analysis of Biochemical Systems* (Springer, 2015), Vol. 1.
[62]D. T. Gillespie, "Stochastic simulation of chemical kinetics," Annu. Rev. Phys. Chem. **58**, 35–55 (2007).
[63]J. T. Bullerjahn, S. von Bülow, and G. Hummer, "Optimal estimates of self-diffusion coefficients from molecular dynamics simulations," J. Chem. Phys. **153**, 024116 (2020).
[64]H. Qian, M. P. Sheetz, and E. L. Elson, "Single particle tracking. Analysis of diffusion and flow in two-dimensional systems," Biophys. J. **60**, 910–921 (1991).
[65]P. Leopardi, "A partition of the unit sphere into regions of equal area and small diameter," Electron. Trans. Numer. Anal. **25**, 309–327 (2006).
[66]I. Buch, T. Giorgino, and G. De Fabritiis, "Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations," Proc. Natl. Acad. Sci. U. S. A. **108**, 10184–10189 (2011).
[67]D.-A. Silva, G. R. Bowman, A. Sosa-Peinado, and X. Huang, "A role for both conformational selection and induced fit in ligand binding by the Lao protein," PLoS Comput. Biol. **7**, e1002054 (2011).
[68]N. Plattner and F. Noé, "Protein conformational plasticity and complex ligand binding kinetics explored by atomistic simulations and Markov models," Nat. Commun. **6**, 7653 (2015).

[69]N. Plattner, S. Doerr, G. De Fabritiis, and F. Noé, "Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling," Nat. Chem. **9**, 1005–1011 (2017).
[70]H. C. Klein and U. S. Schwarz, "Studying protein assembly with reversible Brownian dynamics of patchy particles," J. Chem. Phys. **140**, 184112 (2014).
[71]A. C. Newton, J. Groenewold, W. K. Kegel, and P. G. Bolhuis, "Rotational diffusion affects the dynamical self-assembly pathways of patchy particles," Proc. Natl. Acad. Sci. U. S. A. **112**, 15308–15313 (2015).
[72]J. Schluttig, D. Alamanova, V. Helms, and U. S. Schwarz, "Dynamics of protein-protein encounter: A Langevin equation approach with reaction patches," J. Chem. Phys. **129**, 155106 (2008).
[73]D. J. Higham, "An algorithmic introduction to numerical simulation of stochastic differential equations," SIAM Rev. **43**, 525–546 (2001).
[74]A. Arkhipov, P. L. Freddolino, and K. Schulten, "Stability and dynamics of virus capsids described by coarse-grained modeling," Structure **14**, 1767–1777 (2006).
[75]A. Davtyan *et al.*, "AWSEM-MD: Protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing," J. Phys. Chem. B **116**, 8494–8503 (2012).
[76]J. Wang *et al.*, "Multi-body effects in a coarse-grained protein force field," J. Chem. Phys. **154**, 164113 (2021).
[77]M. J. del Razo, M. Dibak, C. Schütte, and F. Noé (2021), MSM/RD software http://dx.doi.org/10.5281/zenodo.4596416.
[78]M. J. del Razo, M. Dibak, C. Schütte, and F. Noé (2021), MSM/RD generated data/plots http://dx.doi.org/10.5281/zenodo.5236640.