A Smoluchowski model of crystallization dynamics of small colloidal clusters

Daniel J. Beltran-Villegas, Ray M. Sehgal, Dimitrios Maroudas, David M. Ford, and Michael A. Bevan

Related Articles

Additional information on J. Chem. Phys.

A Smoluchowski model of crystallization dynamics of small colloidal clusters

Daniel J. Beltran-Villegas,[1] Ray M. Sehgal,[2] Dimitrios Maroudas,[2] David M. Ford,[2,a)] and Michael A. Bevan[1,a)]

[1]*Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA*
[2]*Department of Chemical Engineering, University of Massachusetts, Amherst, Massachusetts 01003, USA*

We investigate the dynamics of colloidal crystallization in a 32-particle system at a fixed value of interparticle depletion attraction that produces coexisting fluid and solid phases. Free energy landscapes (FELs) and diffusivity landscapes (DLs) are obtained as coefficients of 1D Smoluchowski equations using as order parameters either the radius of gyration or the average crystallinity. FELs and DLs are estimated by fitting the Smoluchowski equations to Brownian dynamics (BD) simulations using either linear fits to locally initiated trajectories or global fits to unbiased trajectories using Bayesian inference. The resulting FELs are compared to Monte Carlo Umbrella Sampling results. The accuracy of the FELs and DLs for modeling colloidal crystallization dynamics is evaluated by comparing mean first-passage times from BD simulations with analytical predictions using the FEL and DL models. While the 1D models accurately capture dynamics near the free energy minimum fluid and crystal configurations, predictions near the transition region are not quantitatively accurate. A preliminary investigation of ensemble averaged 2D order parameter trajectories suggests that 2D models are required to capture crystallization dynamics in the transition region. © *2011 American Institute of Physics*. [doi:10.1063/1.3652967]

I. INTRODUCTION

Small ordered clusters, or crystals, of atoms[1] (e.g., nanoparticles), molecules[2] (e.g., micelles), macromolecules[3] (e.g., protein crystals), nanoparticles[4] (e.g., metamaterials), and colloids[5] (e.g., photonic materials) are important in numerous scientific and technological problems. The relative free energies of different cluster configurations, and the probability with which they are observed under equilibrium conditions, are determined by the rules of equilibrium statistical mechanics.[6] In contrast, the temporal evolution of cluster configurations is a non-equilibrium kinetic problem.[7] Designing schemes to assemble ordered clusters with specific configurations requires identification of the conditions to produce the desired stable cluster (i.e., free energy minimum) and a kinetic pathway to obtain the desired state over reasonable time scales. In brief, understanding both the thermodynamics and kinetics of ordered cluster formation is necessary for the rational assembly of crystalline materials.

Energy landscapes have emerged as a powerful tool over several decades to understand thermodynamic aspects of cluster formation (see, e.g., the book by Wales[1]). To provide a context for the present study, we briefly review the use of order parameters in conjunction with energy landscape models applied to ordered clusters. Order parameters provide scalar representations of cluster configurations that serve as a basis for low-dimensional models of the cluster state. Order parameters are usually chosen as metrics of some coarse-grained physics of interest, and there is no general guarantee of uniqueness. Examples of order parameters for clusters include the radius of gyration, R_g, which can be used to monitor condensation, or the average degree of sixfold order, $\langle C_6 \rangle$, which can be used to monitor crystallinity. After identifying order parameters suitable for capturing different cluster configurations, a method such as Monte Carlo Umbrella Sampling (MC-US) (Refs. 8 and 9) can be used to determine the relative free energy of the cluster at different values of the order parameter. The relative free energy as a function of the order parameter is referred to as a free energy landscape (FEL). With regard to ordered particle clusters, FELs have primarily been constructed in simulation studies, with a few notable exceptions involving optical microscopy of colloidal clusters.[10,11]

The kinetics of cluster configurational changes can be modeled as diffusion on the FEL where the order parameter plays the role of a *reaction* coordinate (strictly speaking, the order parameter is the *migration* coordinate of the diffusion process). For example, crystal nucleation studies (e.g., those reviewed in Ref. 12) use a Kramers-type model of diffusive barrier crossing[13] to predict nucleation rates based on the FEL barrier height and the diffusion coefficient at the barrier. This type of analysis is limited to FEL barrier heights $\gg kT$. Such high barriers result in slow kinetics and may be of less practical interest to assembling ordered clusters in applications. For barrier heights $\sim kT$, the diffusion coefficient along the entire reaction coordinate determines the transition rate between different configurations.[14] In the case of $\sim kT$ barriers, methods have been developed to extract both the FEL and the coordinate dependent diffusivity (i.e., the diffusivity

---

landscape (DL)) by fitting coefficients of the Smoluchowski equation to order parameter trajectories. Two examples of methods for fitting Smoluchowski equation coefficients include linear fits to locally initiated order parameter trajectories[2] or global fits to unbiased order parameter trajectories using Bayesian inference.[15] To date, these methods have primarily been used to extract FELs and DLs in simulations of protein folding (see, e.g., references cited within Refs. 15 and 16), although a few studies have explored disordered clusters.[2,17] We are not aware of any studies using such an approach to model the dynamics of crystallization in simulations or experiments.

In this work, we report results for extracting FELs and DLs from Smoluchowski equation coefficients fit to Brownian dynamics (BD) simulations of colloidal crystallization mediated by interparticle depletion attraction. The model system is one in which the potentials and 2D phase behavior have been measured in optical microscopy experiments.[5,18] This same system was also investigated using MC-US simulations to generate FELs for different levels of depletion attraction.[19] Based on the findings of these previous studies, the present study focuses on a single value of depletion attraction that produces coexisting fluid and crystalline configurations, which allows us to investigate crystallization dynamics. With direct connections to prior experiments, successful modeling of simulated experiments in this work can eventually be validated in microscopy experiments.

In our analysis, we explore both the local linear fitting and global Bayesian inference methods to obtain FELs and DLs from Smoluchowski equation coefficients. We explore the use of two different order parameters to produce 1D FELs and DLs, where "1D" refers to FELs and DLs constructed in one order parameter (or reaction coordinate). The order parameters investigated are R_g and $\langle C_6 \rangle$ (as described earlier and defined below). These analyses produce four possible 1D models using combinations of the two order parameters and the two methods of analysis. To test the resulting models' abilities to capture particle-scale dynamics, we compare mean first-passage times measured directly from BD simulations with analytical predictions using the extracted FELs and DLs. Finally, we present some preliminary "2D" results that provide some insights into limitations of the 1D models and motivate the future development of 2D models to accurately capture the crystallization dynamics of small colloidal clusters.

II. THEORY

A. Finding coefficients in the Smoluchowski equation

We model the dynamic evolution of the system in terms of order parameters by using the Smoluchowski equation,[15]

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x} \left\{ D(x) e^{-W(x)/kT} \frac{\partial}{\partial x} [e^{W(x)/kT} p(x,t)] \right\}, \tag{1}$$

where x is an order parameter or coarse system variable, $p(x,t)$ is the probability density to find the system in a state characterized by an order parameter of value x at time t, $D(x)$ is the diffusivity landscape, and $W(x)$ is the free energy landscape

for an arbitrary coordinate, x. The Smoluchowski equation can also be written as[2]

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x} \left[-v(x) p(x,t) + \frac{\partial}{\partial x} D(x) p(x,t) \right], \tag{2}$$

where $v(x)$ is the corresponding drift velocity.

In an approach we refer to as the linear fitting method,[2] the Smoluchowski equation coefficients in Eq. (2) can be estimated from ensemble averages over sets of short trajectories as

$$v(x_i) = \lim_{\tau \to 0} \frac{1}{\tau} \langle \xi(t+\tau) - \xi(t) \rangle_{\xi(t)=x_i}, \tag{3}$$

$$D(x_i) = \frac{1}{2} \lim_{\tau \to 0} \frac{1}{\tau} [\langle [\xi(t+\tau) - \xi(t)]^2 \rangle]_{\xi(t)=x_i}, \tag{4}$$

which directly yields $D(x)$ and can be used to obtain $W(x)$ as

$$W(x_i)/kT = \int_{x_0}^{x_i} \frac{dD(x')/dx - v(x')}{D(x')} dx'. \tag{5}$$

In another approach we refer to as the Bayesian inference method[15] (based on the application of Bayes theorem), the Smoluchowski equation in Eq. (1) is discretized in one dimension as a system of rate equations

$$\dot{p}_i(t) = R_{i,i-1} p_{i-1}(t) - (R_{i-1,i} + R_{i+1,i}) p_i(t) + R_{i,i+1} p_{i+1}(t), \tag{6}$$

where $p_i(t)$ is the probability of being in interval i around x_i at time t, and $\mathbf{R} = [R_{ij}]$ is a rate matrix with matrix elements $R_{ij} \geq 0$ for $i \neq j$, $R_{ij} < 0$ for $i = j$, and $\Sigma R_{ij} = 0$. The matrix $\mathbf{R}$ can be estimated from a set of dynamic observations by estimating the variability of the likelihood function

$$L \equiv \prod_\alpha p(x_{i_\alpha}, t_{k_\alpha} | x_{j_\alpha}, 0), \tag{7}$$

where α refers to a realization of the system. This variability is sampled by means of the Metropolis Monte Carlo algorithm using the negative log-likelihood as the energy function as

$$-\ln L = -\sum_k \sum_j \sum_i N(x_i, t_k | x_j, 0) \ln p(x_i, t_k | x_j, 0), \tag{8}$$

where N is called the jump matrix, which is obtained by collecting the number of transitions from x_j to x_i in a time interval t_k, and $p(x_i,t_k|x_j,0) = [\exp(\mathbf{R}t_k)]_{ij}$ are transition probabilities estimated for given values of $\mathbf{R}$ and $\mathbf{P}$ following the procedure of Bicout and Szabo.[20] The vector of equilibrium probabilities, $\mathbf{P}$, of being in interval i around x_i is an eigenvector of $\mathbf{R}$ with eigenvalue 0. The Smoluchowski equation coefficients can be obtained from $\mathbf{P}$ and $\mathbf{R}$ as

$$W(x_i) = -kT \ln \left(\frac{P_i}{x_{i+1} - x_i} \right), \tag{9}$$

$$D \left(\frac{x_i + x_{i+1}}{2} \right) = (x_{i+1} - x_i)^2 R_{i+1,i} \left(\frac{P_i}{P_{i+1}} \right)^{1/2}. \tag{10}$$

B. Order parameters

The radius of gyration, R_g*, for N particles of radius a with position vector **r** is given by

$$Rg* \equiv \frac{1}{aN}\left[\frac{1}{2}\sum_{i,j}|\mathbf{r}_i - \mathbf{r}_j|^2\right]^{1/2}, \tag{11}$$

where the asterisk (*) denotes normalization by the particle radius.

The crystallinity order parameter, $\langle C_6 \rangle$, is defined as the average number of crystalline nearest neighbors around each particle in an ensemble. The number of coordinated neighbors, $N_C(i)$, to particle i are all particles j within a coordination radius, r_C. Identification of crystalline nearest neighbors is based on a local sixfold orientational order parameter for the particle i given by[21]

$$\overline{q}_{6m}(i) \equiv \frac{1}{N_C(i)}\sum_{j=1}^{N_C(i)} Y_{6m}(\hat{\mathbf{r}}_{ij}), \tag{12}$$

where $\hat{r}_{ij}$ is a unit vector specifying the orientation of the vector $\mathbf{r}_{ij}$ joining neighboring particles i and j, and $Y_{6m}(\hat{\mathbf{r}}_{ij})$ are spherical harmonic components. Equation (12) can be normalized as

$$\tilde{q}_{6m}(i) = \frac{\overline{q}_{6m}(i)}{\left[\sum_{m=-6}^{6}|\overline{q}_{6m}(i)|^2\right]^{1/2}}, \tag{13}$$

which is used to determine the crystalline connectivity, $\chi_6(ij)$, between particle i and neighboring particles j as

$$\chi_6(ij) = \sum_{m=-6}^{6} \tilde{q}_{6m}(i)\tilde{q}_{6m}(j)^*, \tag{14}$$

where $\tilde{q}_{6m}(j)^*$ is the complex conjugate of $\tilde{q}_{6m}(j)$. The number of crystalline near neighbors, $C_6(i)$, for particle i is[21]

$$C_6(i) = \sum_{j=1}^{N_C(i)}\left[\begin{vmatrix} 1 & \chi_6(ij) \geq 0.5 \\ 0 & \chi_6(ij) < 0.5 \end{vmatrix}\right], \tag{15}$$

which gives the value of $\langle C_6 \rangle$ as the average over all particles in an ensemble as

$$\langle C_6 \rangle = \frac{1}{N}\sum_{i=1}^{N} C_6(i). \tag{16}$$

III. METHODS

All BD simulations used numerical methods described in our previous papers.[22–24] Simulations were performed for 20 000s with an integration time of 0.01 ms. Periodic boundary conditions were applied to the simulation cell, which contained 32 particles corresponding to a volume fraction of $\phi = 0.0695$. A separation- and configuration-independent particle diffusivity was used in all BD simulations. Conservative forces between colloids were modeled as the superposition of a DLVO repulsive electrostatic potential and an AO attractive depletion potential using parameters from our previous measurements[5,18] reported in Table I.

TABLE I. Parameters (using notations in Ref. 18) for potentials in all BD and MC-US simulations: (a) colloid diameter, $2a$, (b) depletant diameter, $2L$, (c) colloid surface potential, $-\psi_p$, (d) solvent dielectric constant, ε, (e) Debye length, κ^{-1}, (f) depletant osmotic pressure, Π/kT, and (g) colloid volume fraction, ϕ.

Parameter	Value
$2a$ (nm)	2200
$2L$ (nm)	220
$-\psi_p$ (mV)	50
ε	78
κ^{-1} (nm)	8
Π/kT ($10^{-7}\mathrm{nm}^{-3}$)	1.11
ϕ	0.0695

BD simulations used in the linear fitting and Bayesian inference analyses were started from 40 different initial configurations with particle coordinates stored every 100 ms. After BD simulations were performed, the resulting configurations were analyzed to calculate order-parameter trajectories for use in the linear fitting [Eqs. (3)–(5)] and Bayesian inference [Eqs. (6)–(10)] analyses. The resolution for each order-parameter-method combination was set as: 0.25 for $\langle C_6 \rangle$-BI, 0.0625 for $\langle C_6 \rangle$-LF, 0.08 for R_g*-BI, and 0.01 for R_g*-LF; in the above abbreviated models, BI and LF denote Bayesian inference and linear fitting, respectively.

BD simulations for mean first-passage time analyses included 50 simulations starting from different initial configurations at fixed order-parameter values. Each simulation was stopped after the order parameter reached the desired end state. Analytical predictions of mean first-passage times were performed using[14]

$$\tau(x, x_f) = \int_x^{x_f} dx' \frac{\exp[W(x')/kT]}{D(x')}$$
$$\times \int_{x_0}^{x'} dx'' \exp[-W(x'')/kT], \tag{17}$$

where x_0 corresponds to a reflective boundary for the Smoluchowski equation, taken as $\langle C_6 \rangle = 0$ and $R_g* = 4.5$ when starting from the fluid basin or $\langle C_6 \rangle = 7.5$ and $R_g* = 2.7$ when starting from the crystal basin.

MC-US simulations were performed using the Metropolis-Monte Carlo algorithm for 1.5×10^8 steps in each window and calculating order parameters every 10 MC steps. The resolution for the R_g* FEL was set to 0.01 with a window size of 0.16 and a window-window overlap of 0.08. The resolution for the $\langle C_6 \rangle$ FEL was set to 0.0625 with a window size of 0.625 and a window-window overlap of 0.3125. Hard walls were used for biasing potentials in each window.[25]

IV. RESULTS AND DISCUSSION

A. Stochastic dynamics in the order parameters

Figure 1 shows three sample trajectories from particle-scale BD simulations in terms of the $\langle C_6 \rangle$ and R_g* order parameters as functions of time over a period of 10^4 s. The order
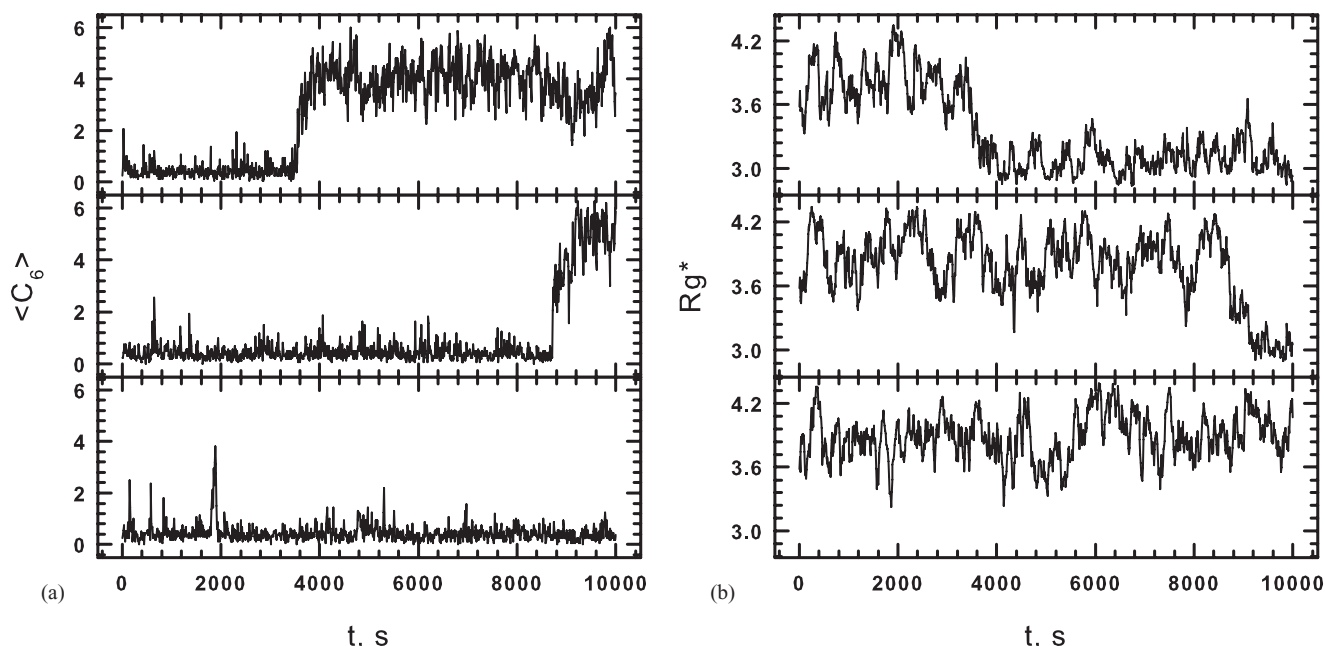
FIG. 1. Sample trajectories from BD simulations in terms of the evolution of order parameters (a) $\langle C_6 \rangle$ and (b) R_g^*. All three trajectories start from the same initial configuration ($\langle C_6 \rangle = 0.3$, $R_g^* = 3.7$) but represent different stochastic realizations.

parameters used in this study were chosen based on an empirical search of more than a dozen candidates in our previous study.[19] The trajectories in Fig. 1 all start from the same initial configuration ($\langle C_6 \rangle = 0.3$, $R_g^* = 3.7$) but show different realizations of the system evolution in time.

Figure 1 shows that the system may undergo a distinct change from fluid configurations at $\langle C_6 \rangle < 2$ and $R_g^* > 3.6$ to crystalline configurations at $\langle C_6 \rangle > 2$ and $R_g^* < 3.5$. This transition does not happen at a specified time but occurs stochastically; in fact, for the third sample trajectory (bottom row) the transition does not occur during the observation period. These results are consistent with our previous findings,[19] where it was shown that this system has to overcome a free energy barrier to undergo crystallization. In the following, we develop quantitative stochastic models in each

of these order parameters to capture the phase change dynamics based on FELs and DLs constructed using the linear fitting [Eqs. (3)–(5)] and Bayesian inference [Eqs. (6)–(10)] analyses.

B. Analyses of stochastic dynamics

Figure 2 shows dynamic quantities, as extracted from multiple BD simulations, that are used in the Bayesian inference and linear fitting analyses. These are not all the data but a representative subset for illustration purposes. For the Bayesian inference analysis, Fig. 2(a) depicts elements of the jump matrix **N** that is the input quantity in Eq. (8). Figure 2(a) shows the number of jumps from an initial value of $\langle C_6 \rangle = 0.3$ to other values for different times ranging from
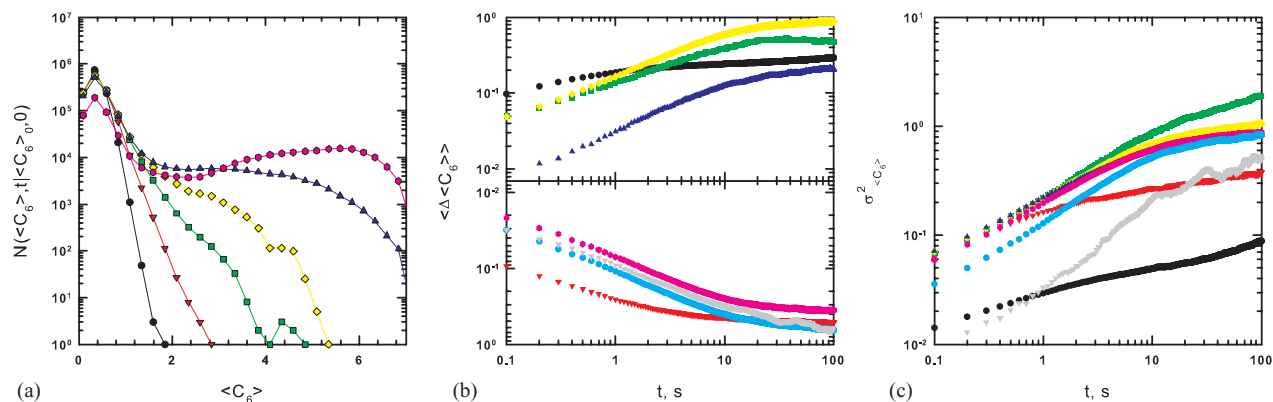

FIG. 2. Sample dynamic information obtained from BD simulations using $\langle C_6 \rangle$ as the order parameter: (a) Number of jumps from an initial $\langle C_6 \rangle \approx 0.3$ to other values of the order parameter as functions of time (black circles, $t = 0.1$ s; red inverted triangles, $t = 1$ s; green squares, $t = 10$ s; yellow diamonds, $t = 10^2$ s; blue triangles, $t = 10^3$ s; pink hexagons, $t = 10^4$ s). (b) Evolution of ensemble-averaged displacement for different initial values (black circles, $\langle C_6 \rangle = 0.1$; red inverted triangles, $\langle C_6 \rangle = 1.1$; green squares, $\langle C_6 \rangle = 2.1$; yellow diamonds, $\langle C_6 \rangle = 3.1$; blue triangles, $\langle C_6 \rangle = 4.1$; pink hexagons, $\langle C_6 \rangle = 5.1$; cyan circles, $\langle C_6 \rangle = 6.1$; and grey inverted triangles, $\langle C_6 \rangle = 7.1$). Positive displacements in the upper plot and negative displacements in the lower plot. (c) Evolution of ensemble averaged displacement variances for different initial values (same color scheme as in Fig. 2(b)).
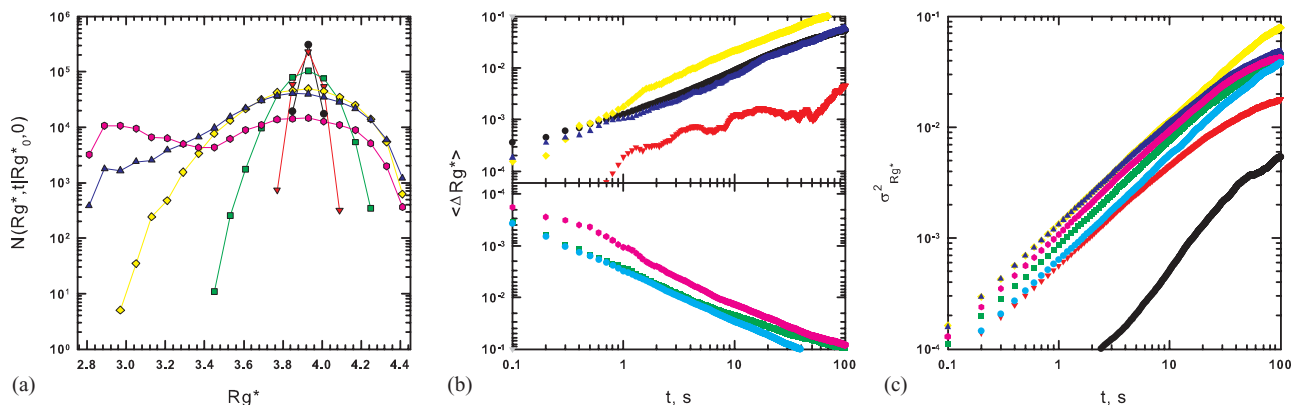
FIG. 3. Sample dynamic information obtained from BD simulations using R_g^* as the order parameter: (a) Evolution of number of jumps from an initial $R_g^* \approx$ 3.7 to other values of the order parameter with same color-symbol scheme as in Fig. 2(a). (b) Evolution of ensemble averaged displacement for different initial R_g^* values (black circles, $R_g^* = 2.81$; red inverted triangles, $R_g^* = 3.05$; green squares, $R_g^* = 3.29$; yellow diamonds, $R_g^* = 3.53$; blue triangles, $R_g^* = 4.77$; pink hexagons, $R_g^* = 4.01$; cyan circles, $R_g^* = 4.25$). Positive displacements in the upper plot and negative displacements in the lower plot. (c) Evolution of ensemble averaged displacement variances for different initial values (same color scheme as in Fig. 3(b)).

0.1 to 10^4 s. These data show that at short times, the system is localized at conditions corresponding to fluid configurations but at longer times it is able to cross the barrier to produce crystalline configurations. After 10^3 s, the jump matrix captures characteristics of the system across the entire range of $\langle C_6 \rangle$, including a peak at high $\langle C_6 \rangle$ values that represents crystalline configurations.

For the linear fitting analysis, Figs. 2(b) and 2(c) show mean $\langle C_6 \rangle$ displacements and displacement variances from different $\langle C_6 \rangle$ starting positions as functions of time. These are the input quantities used to estimate the drift and diffusion coefficients in Eqs. (3) and (4). Different starting positions yield either positive displacements toward higher $\langle C_6 \rangle$ values or negative displacements toward lower $\langle C_6 \rangle$ values, which depends on local features of the underlying landscapes. The variances are always positive and represent the spread in the statistical distribution of displacements.

The existence of a linear region at some intermediate time scale is necessary for the linear fitting analysis of these data. The plots in Figs. 2(b) and 2(c) are shown on log-log scales to assist in the identification of linear regions, which should appear as straight lines with a slope of one. Detailed inspection of the curves shows that they all have slopes that are significantly less than one over the entire time period, thus calling into question the validity of a linear fitting analysis of the $\langle C_6 \rangle$ data; this issue will be discussed further below in the context of the landscape results.

The same dynamic quantities reported in Fig. 2 are also shown in Fig. 3 using the R_g^* order parameter instead of $\langle C_6 \rangle$. Figure 3(a) shows the number of jumps from an initial value of $R_g^* = 4.0$ to other values for the same times as in Fig. 2(a). As with $\langle C_6 \rangle$, the system remains localized at conditions corresponding to fluid configurations at short times but makes excursions across the barrier at longer times. After 10^3 s, the jump matrix captures characteristics of the system across the entire range of R_g^* values, including a peak at $R_g^* \approx 3$ that corresponds to crystalline configurations. Figures 3(b) and 3(c) show average displacements and displacement variances for different starting R_g^* values. In contrast to the plots for $\langle C_6 \rangle$, the log-log plots in Figs. 3(b) and

3(c) clearly show linear regions of slope equal to one for all values of initial R_g^* (except perhaps for 3.05, which will be discussed below), which is a positive indicator for the validity of the linear fitting analysis.

C. Free energy and diffusivity landscapes

Figure 4 shows the FEL [i.e., $W(\langle C_6 \rangle)$] and DL [i.e., $D(\langle C_6 \rangle)$] in $\langle C_6 \rangle$ from the Bayesian inference and linear fitting analyses of the dynamic quantities reported in Fig. 2. Figure 4(a) shows $W(\langle C_6 \rangle)$ from a Bayesian inference analysis of the jump matrix data in Fig. 2(a) and from linear fitting analyses of the mean displacements and displacement variances in Figs. 2(b) and 2(c). For comparison, Fig. 4(a) also reports $W(\langle C_6 \rangle)$ obtained from MC-US; we consider this MC-US result as the benchmark. Figure 4(b) shows $D(\langle C_6 \rangle)$ results obtained from the Bayesian inference and linear fitting analyses of the quantities in Fig. 2. For the linear fitting analyses, the results in Fig. 4 are shown in different colors to represent different time intervals for which the linear slopes were extracted via least squares fits.

Figure 4(a) shows good agreement between the MC-US and Bayesian inference estimates of $W(\langle C_6 \rangle)$. While the linear fitting analysis yields $W(\langle C_6 \rangle)$ with qualitatively correct shapes, quantitative agreement with the MC-US results is not obtained, indicating that the linear fitting analysis is not adequate in this case. The degree of agreement in the linear fitting results in Fig. 4(a) does not show a monotonic trend with increasing or decreasing time intervals employed for the analysis. For example, blue hexagons ($t = [4.0, 10.0]$ s) are better than red squares ($t = [0.1, 1.0]$ s) but yellow triangles ($t = [2.0, 4.0]$ s) are worse. Of course, failure of the linear fitting analysis to capture the correct $W(\langle C_6 \rangle)$ is not surprising given the lack of any evident linear region in the displacement and variance data in Figs. 2(b) and 2(c). The Bayesian inference analysis is not constrained by any such linearity requirement and yields a quantitatively accurate $W(\langle C_6 \rangle)$.
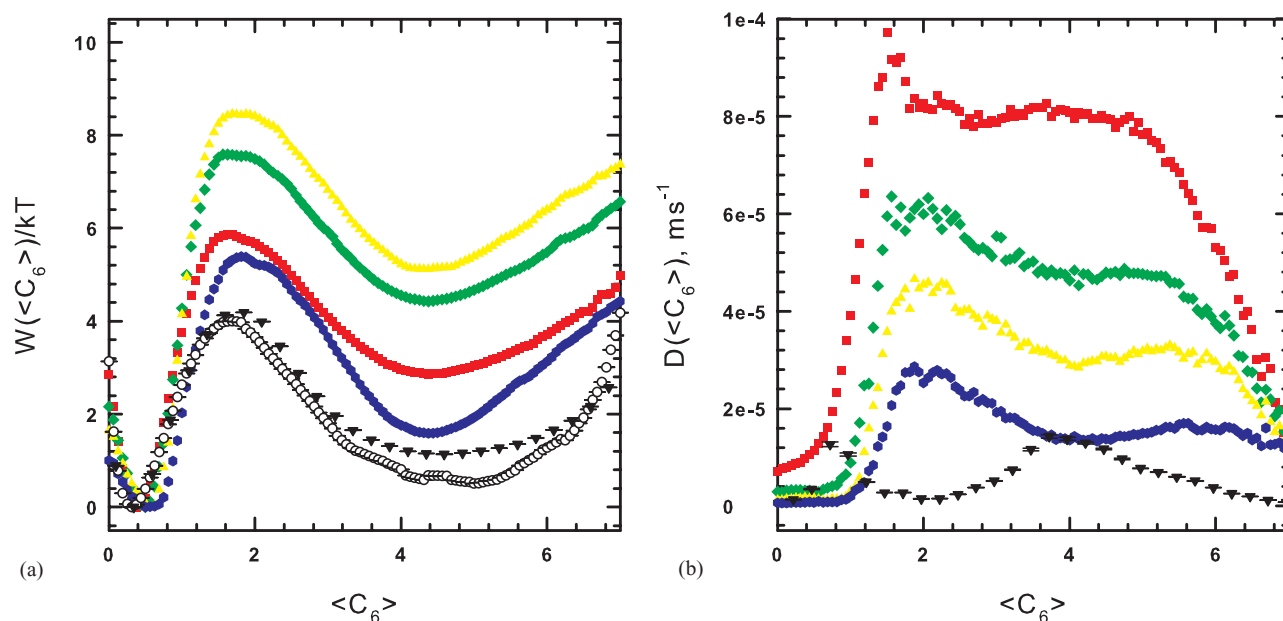
FIG. 4. Results for $\langle C_6 \rangle$ as the order parameter: (a) FEL and (b) DL. Open circles show MC-US estimates; black inverted triangles show Bayesian inference estimates; linear fitting estimates are shown for different time windows (red squares, $t = [0.1,1.0]$ s; green diamonds, $t = [1.0,2.0]$ s; yellow triangles, $t = [2.0,4.0]$ s; blue hexagons, $t = [4.0,10.0]$ s).

Interestingly, Fig. 4(b) shows that the linear fitting and Bayesian inference results for $D(\langle C_6 \rangle)$ do not agree even qualitatively. The Bayesian inference result has the overall lowest magnitude with peaks near the $\langle C_6 \rangle$ values of 0.8 and 4.0. The locations of these peaks in $D(\langle C_6 \rangle)$ do not appear to correspond to any notable features in the $W(\langle C_6 \rangle)$ landscape of Fig. 4(a), so it is difficult to develop a physical explanation for them. Note that the magnitude of the linear fitting result depends on the time interval used for the analysis, with uniformly decreasing D values for increasing time scale (consistent with the trends in the displacement variance plots of Fig. 2(c)), and the shape of the linear fitting curves is quite different from that of the Bayesian inference curve. None of these linear fitting results, which are shown for the sake of completeness, are expected to be reliable due to the lack of linear behavior in the displacement and variance data in Figs. 2(b) and 2(c). At this point, one might hypothesize that $\langle C_6 \rangle$ is a reasonable choice for a dynamical order parameter, as long as the more general (nonlinear) Bayesian inference technique is used to generate the $W(\langle C_6 \rangle)$ and $D(\langle C_6 \rangle)$. Before testing this hypothesis against actual dynamical data, we first evaluate R_g^* as a possible order parameter.

Figure 5 shows FELs [i.e., $W(R_g^*)$] and DLs [i.e., $D(R_g^*)$] in R_g^* from the Bayesian inference and linear fitting analyses. Figure 5(a) shows good agreement between the $W(R_g^*)$ obtained by both the analyses employed. Furthermore, the linear fitting results are now quite insensitive to the time interval used to obtain the drift velocities and diffusion coefficients. These findings are consistent with the observations of long linear regions in Figs. 3(b) and 3(c). However, both the Bayesian inference and linear fitting results disagree with the $W(R_g^*)$ benchmark landscape from MC-US. While the shape of the fluid and crystal basins is the same in all cases, there is quantitative disagreement in the vicinity of the

transition region over the range $R_g^* = [3.4,3.7]$. This disagreement between the Smoluchowski equation results and the MC-US estimates of $W(R_g^*)$ indicates that the thermodynamics is not being correctly captured and raises doubts about the appropriateness of R_g^* as an order parameter for describing transitions between the fluid and crystal phases.

Figure 5(b) shows the corresponding $D(R_g^*)$ results. As expected, for the linear fitting analysis, the results are fairly insensitive to the time interval. Interestingly, the linear fitting estimates of $D(R_g^*)$ also agree with the Bayesian inference estimate at most values except for an interval around the transition region $R_g^* = [3.2,3.6]$. This disagreement between the linear fitting and Bayesian inference estimates of $D(R_g^*)$ in Fig. 5(b) most likely stems from the fact that linear fitting analysis is based on a local fit to measured dynamics at each value of the order parameter, whereas the Bayesian inference analysis is based on a global fit to the dynamics at all values of the order parameter. As a result, even though the linear fitting and Bayesian inference analysis produce nearly identical $W(R_g^*)$ in Fig. 5(a), the Bayesian inference analysis produces a different $D(R_g^*)$ in the vicinity of the free energy barrier to minimize the global error when fitting the coefficients to particle-scale dynamics. However, arguments regarding the relative merits of the two $D(R_g^*)$ predictions are somewhat moot because neither analysis predicts an accurate $W(R_g^*)$ (as compared to MC-US).

D. Mean first-passage times from free energy and diffusivity landscapes

We can now evaluate the ability of the 1D models presented in Figs. 4 and 5 to predict a key dynamical quantity, the mean first-passage time. For completeness, we compare BD simulations results with mean first-passage time predictions
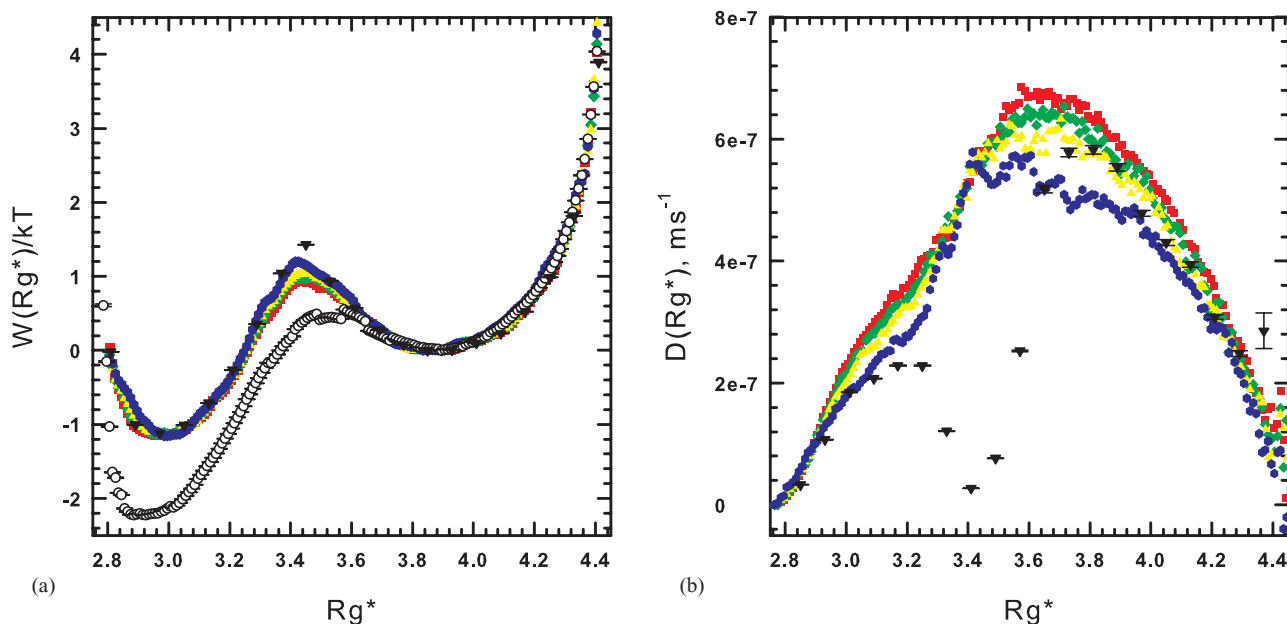
FIG. 5. Results for R_g* as the order parameter: (a) FEL and (b) DL. Same symbol-color scheme as in Fig. 4.

from Eq. (17) using all four models (i.e., $\langle C_6\rangle$-BI, $\langle C_6\rangle$-LF, R_g*-BI, and R_g*-LF). Both plots in Fig. 6 report mean first-passage times with starting coordinates from: (1) the fluid basin minimum (i.e., $\langle C_6\rangle = 0.35$ and $R_g* = 3.9$) for trajectories moving toward more condensed and ordered states and (2) the crystal basin minimum (i.e., $\langle C_6\rangle = 4.5$ and $R_g* = 2.95$) for trajectories moving toward more expanded and disordered states. Figure 6(a) shows results based on $\langle C_6\rangle$, and Fig. 6(b) shows results based on R_g*. In both plots in Fig. 6, mean first-passage times from BD simulations are shown by open symbols, Bayesian inference predictions are shown by a red line, and linear fitting predictions are shown by a green line. Different symbol and line types are used for trajectories starting in the fluid and crystal minima. The mean first-passage time results from the BD simulations are essentially exact and provide a benchmark for the predictions based on the four models.

All four landscape models in Fig. 6 display the correct qualitative trends and, furthermore, they are all (with the exception of $\langle C_6\rangle$-LF) quantitatively accurate within the free energy basin associated with their starting coordinate. However, none of the models quantitatively predicts mean first-passage times up to and beyond the free energy barrier from both basins. The Bayesian inference model, in either order parameter, is quantitatively accurate into the barrier region when starting from the crystal basin but significantly over predicts the mean first-passage time when starting from the fluid basin.

The accuracy, or lack thereof, of the models in predicting the mean first-passage time is consistent with expectations based on the results in Figs. 2–5. Since none of the $W(R_g*)$ landscapes in Fig. 5 agree with the MC-US results across the transition region, models based on R_g* alone were not expected to capture particle scale crystallization dynamics. All of the linear-fitting-based results for $W(\langle C_6\rangle)$ were suspect because of the lack of a linear region in the underlying displacement and variance data in Fig. 2 and, in

fact, none of these results agreed with those from MC-US in Fig. 4. Only the $\langle C_6\rangle$-BI analysis produced a FEL that agreed with the MC-US result and therefore, this analysis had the best chance of succeeding in the mean first-passage time predictions. Although the $\langle C_6\rangle$-BI breaks down as the free energy barrier is approached from the fluid basin ($\langle C_6\rangle > 1$), it is quantitatively accurate for all other conditions.

E. Motivation for 2D landscape models

Based on the limitations of the mean first-passage time predictions in Fig. 6, we present preliminary results that motivate the development of 2D landscape models, which show promise for accurately capturing colloidal crystallization dynamics. Figure 7 shows plots of averaged trajectory traces based on both $\langle C_6\rangle$ and R_g* coordinates. Each spectrum-colored trajectory represents the average path of at least 50 independent BD trajectories passing through the same starting coordinate. The colors within a trace represent linear time progression, with violet for the beginning and red for the end of each trajectory. Traces in Figs. 7(a)–7(c) depict 100 s of elapsed time while those in Fig. 7(d) show 10 s. Trajectories are overlaid on a contour plot of $W(\langle C_6\rangle, R_g*)$ from our previous MC-US study.[19]] In the following, we discuss the results of Fig. 7(a) in terms of three characteristic regions: (I) the fluid basin at low $\langle C_6\rangle$ and high R_g* coordinates (magnified region in Fig. 7(b)), (II) the crystal basin at high $\langle C_6\rangle$ and low R_g* coordinates (magnified region in Fig. 7(c)), and (III) a transition region containing an energy barrier at intermediate values of $\langle C_6\rangle$ and R_g* (magnified region in Fig. 7(d)).

In the fluid basin in region I, there is a 1D manifold that runs nearly parallel to the R_g* axis. The traces indicate a separation of time scales with a rapid approach to the manifold (i.e., still purple, ~10% of elapsed trajectory) followed by motion along it. A local free energy minimum near $R_g* \approx 3.85$ can be detected by observing which direction the
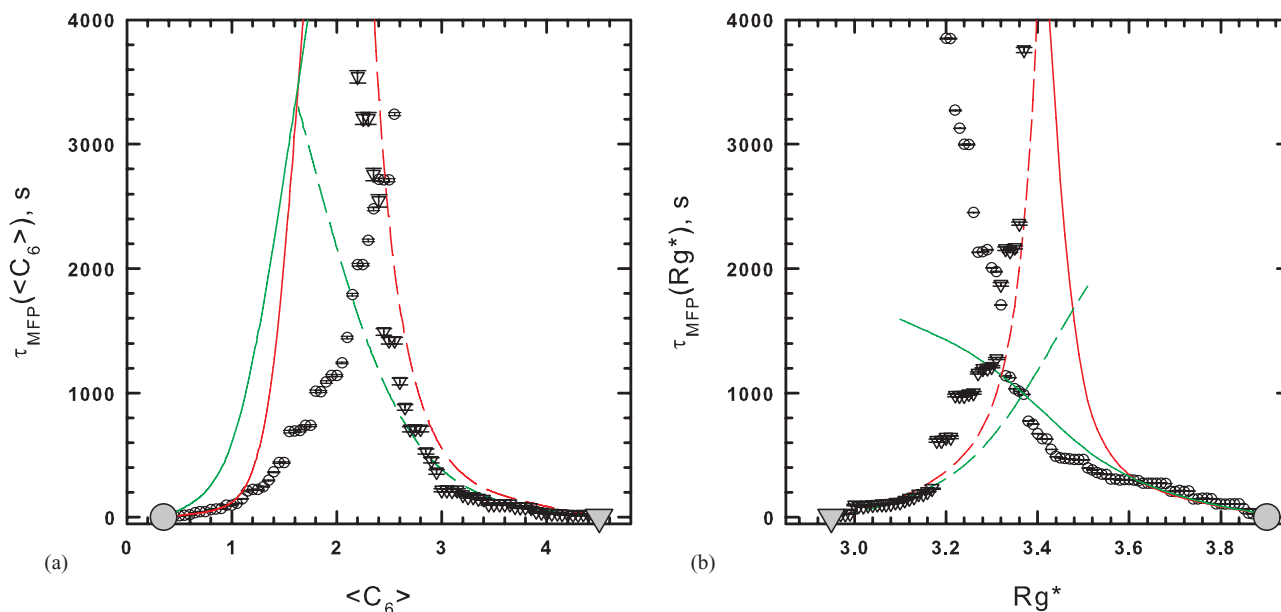
FIG. 6. Mean first-passage times: (a) using $\langle C_6 \rangle$ as order parameter, starting from fluid ($\langle C_6 \rangle = 0.35$, grey circle) and crystal ($\langle C_6 \rangle = 4.5$, grey triangle) vs. $\langle C_6 \rangle$ end value; (b) using R_g^* as order parameter, starting from fluid ($R_g^* = 3.9$, grey circle) and crystal ($R_g^* = 2.95$, grey triangle) configurations vs. R_g^* end value. Open symbols show averages from BD simulations with error bars. Lines show analytical results using Eq. (17). Solid lines starting from fluid configurations, dashed starting from crystal configurations, red for Bayesian inference estimates, green for linear fitting estimates for $t = [4.0, 10.0]$ s.

traces move along the manifold after reaching it. The underlying $W(\langle C_6 \rangle, R_g^*)$ shows a valley that encloses these trajectories and the location of the fluid basin minimum coinciding with the position to which the trajectories converge.

These 2D trajectories in region I explain some of the 1D results in Figs. 2 and 3. For example, the plots in Fig. 3(b) for the mean displacement evolution ($\langle \Delta R_g^* \rangle$ vs. t) with initial positions $R_g^* > 3.4$ are nonlinear at short times, representing the approach to the manifold, and linear at longer times, representing motion along the manifold. In contrast, in Fig. 2(b), all mean displacement evolution curves ($\langle C_6 \rangle$ vs. t) with initial positions $\langle C_6 \rangle < 1.1$ never reach a linear region because the system dynamics is dictated by a manifold nearly orthogonal to the $\langle C_6 \rangle$ axis. The biggest changes in $\langle C_6 \rangle$ happen at the shortest times as trajectories approach the manifold in a nonlinear manner. After that, things become complicated because Fig. 7 shows trajectories having the same initial $\langle C_6 \rangle$ value may actually move in different directions along the manifold depending on their initial R_g^* value. In any case, the values of $\langle C_6 \rangle$ change very little along the manifold because of its almost-perpendicular orientation to the manifold. It is actually surprising that $\langle C_6 \rangle$ is even somewhat adequate as a 1D order parameter in the fluid region.

In the crystal basin, in region II, there is no identifiable 1D manifold but the trajectories do terminate at points in a common 2D region that has a strong correspondence with the underlying basin in $W(\langle C_6 \rangle, R_g^*)$ from MC-US. In Fig. 7(c), 100-s trajectories starting at the periphery of this region display a separation of time scales; they proceed rapidly (i.e., still purple, first $\sim 10\%$ of elapsed trajectory) toward the interior and then spend the rest of the time there. Observations of traces in region II explain some of the corresponding results in the 1D average displacement plots of Figs. 2 and 3. For example, consider all of the traces that start at $R_g^* 3.05$. These

traces have quite different time histories and ending points depending on their starting $\langle C_6 \rangle$ values, which span from range 2 to 6. The noise and roughness in the plot for mean displacement evolution ($\langle \Delta R_g^* \rangle$ vs. t) with initial positions $R_{g,0} = 3.05$, in Fig. 3(b), is a consequence of averaging over this disparate set of trajectories. This particular example highlights a general problem that all of the 1D analyses will have in this region because of the lack of a 1D manifold.

While the underlying FEL and the direction of trajectories display a strong correspondence in region II (i.e., Figs. 7(a) and 7(c)), the free energy minimum from 2D MC-US (marked with a "circled W") does not exactly coincide with the apparent region of localization of trajectories within the crystal basin. However, the trajectories are not only dictated by the underlying $W(\langle C_6 \rangle, R_g^*)$ alone, but also depend on contributions from a DL. In fact, Eq. (5) shows that the drift velocity v, or the apparent velocity of the trajectories in Fig. 7(c), depends on the local gradient of W, as well as the absolute value and gradient of D. As the gradient of W vanishes at the minimum, the local value of D and its gradient become increasingly important. Although we are not yet prepared to compute a 2D DL as part of the preliminary analysis in Fig. 7, we mark a "circled D" as a point of local maximum diffusivity based on the 1D DLs from the $\langle C_6 \rangle$-BI and R_g^*-BI analyses in Figs. 4(b) and 5(b). We also do not make any attempt to approximate gradients in 1D DLs, but the "circled D" suggests that DL may be responsible for the drift of the trajectories away from "circled W" (drift should occur for positive gradients in D toward local maxima). In any case, the lack of an obvious 1D manifold in region II suggests the need for 2D models to capture the dynamics within the crystal basin.

The transition region connecting regions I and II (i.e., region III) is shown in Fig. 7(d); note the change in scale
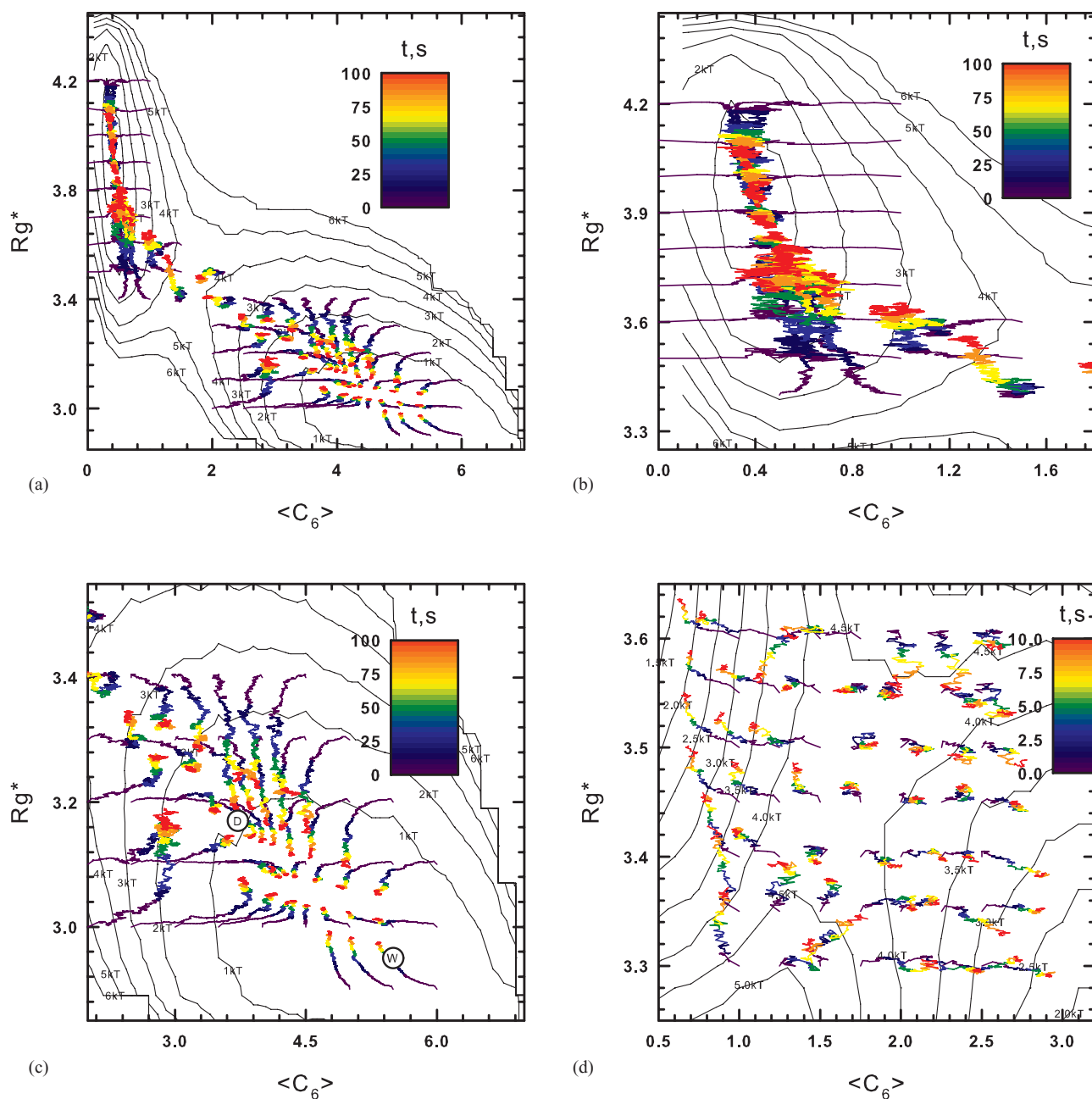
FIG. 7. Average trajectories in 2D order parameter space ($\langle C_6 \rangle$, R_g^*) from different starting positions overlaid on 2D MC-US FELs.[19] (a) Total trajectory time is 100 s represented by linear color scheme from violet to red. (b) Detail in fluid basin. (c) Detail in crystal basin. Crystal free energy minimum from 2D MC-US FEL shown by "W." "D" corresponds to maxima from 1D-DL from Bayesian inference analyses in Figs. 4(b) and 5(b). (d) Detail around transition region for 10-s trajectories with a similar coloring scheme as that in Fig. 7(a).

of elapsed time to 10 s. The underlying $W(\langle C_6 \rangle, R_g^*)$ from MC-US strongly corresponds to the trajectory traces. The traces indicate orthogonal movement away from the transition region as if it were an elevated ridge. Interestingly, the traces that move toward the fluid basin converge into the lower end of the 1D manifold within region I. The transition state is observable as a saddle in $W(\langle C_6 \rangle, R_g^*)$ near ($\langle C_6 \rangle = 1.8$, $R_g^* = 3.45$), where a single trace stays highly localized throughout its 10 s lifetime. The strong correspondence between the trajectories and the $W(\langle C_6 \rangle, R_g^*)$ from 2D MC-US suggests a dominant role for the FEL (and a minimal role for the DL) in determining dynamics through the transition state in region

III. As in the crystal basin, there is not an obvious 1D manifold in the transition region.

Ultimately, the 2D results in Fig. 7 are consistent not only with the 1D results reported in Figs. 1–6, but also capture additional aspects in local 2D trajectories. This finding motivates the development of 2D models for capturing mean first-passage times where the 1D models fail in Fig. 6. As a result, ongoing and future work will focus on implementing 2D models using $\langle C_6 \rangle$ and R_g^* as order parameters to capture particle-scale mean first-passage time results. In addition, ongoing work will use the Bayesian inference method (rather than the linear fitting method) based on its success in the 1D

cases in this work without the need to search for linear temporal behavior. Implementation of such 2D models follows directly from the 1D models in the present study without any obvious conceptual issues, although a number of practical analytical and numerical modifications are required.

As already noted, the order parameters used in this study were chosen based on precedent and an empirical search of more than a dozen candidates. Although the results in Fig. 7 appear quite promising for developing 2D models based on $\langle C_6 \rangle$ and R_g^*, ongoing efforts are also focused on finding new order parameters via more objective approaches (such as principal component analysis, diffusion maps, etc.). With these modifications, it is anticipated that a 2D stochastic model capable of capturing colloidal crystallization dynamics will be developed.

V. CONCLUSIONS

We have reported the development of coarse-grained models to capture both dynamic and equilibrium aspects of the formation of ordered clusters in a 32-particle colloidal system interacting via a short-ranged depletion attraction. We investigated a single value of interparticle attraction that results in equilibrium between disordered fluid configurations and ordered configurations as demonstrated in our previous work using MC-US. The linear fitting and Bayesian inference methods were used to extract FELs and DLs from 1D models based on Smoluchowski equations with either $\langle C_6 \rangle$ or R_g^* as the order parameter.

FELs from the linear fitting and Bayesian inference analyses were compared with those generated from MC-US, which are taken as the benchmark equilibrium results. Only the $\langle C_6 \rangle$-BI model produced a free energy landscape in quantitative agreement with MC-US. Analytical predictions of mean first-passage times based on FELs and DLs using the four possible models [i.e., models based on the two order parameters ($\langle C_6 \rangle$, R_g^*) with two methods of analysis (linear fitting, Bayesian inference)] were compared with mean first-passage times measured directly from BD simulations, which are taken as exact dynamic results. The $\langle C_6 \rangle$-BI model performed better than the other models by accurately capturing mean first-passage times within the free energy basin that corresponds to ordered clusters and in the vicinity of the free energy barrier, as well as showing the correct qualitative (although not quantitative) behavior within the fluid basin. The failure of any of the 1D models to accurately predict the dynamics across the entire range of interest prompted an preliminary analysis of 2D $\langle C_6 \rangle$, R_g^* trajectories, which provided some insight into the limitations of the 1D models and motivates future work to develop 2D $\langle C_6 \rangle$, R_g^*-BI models to capture equilibrium and dynamic aspects of ordered cluster for-

mation. Ultimately, our findings demonstrate a coarse-grained dynamic modeling approach that appears promising for a fundamental understanding of the dynamic assembly of ordered clusters and crystalline materials.

[1] D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses* (Cambridge University Press, Cambridge, 2004).

[2] D. I. Kopelevich, A. Z. Panagiotopoulos, and I. G. Kevrekidis, J. Chem. Phys. **122**, 044908 (2005).

[3] P. R. ten Wolde and D. Frenkel, Science **277**, 1975 (1997).

[4] A. M. Kalsin, M. Fialkowski, M. Paszewski, S. K. Smoukov, K. J. M. Bishop, and B. A. Grzybowski, Science **312**, 420 (2006).

[5] G. E. Fernandes, D. J. Beltran-Villegas, and M. A. Bevan, J. Chem. Phys. **131**, 134705 (2009).

[6] T. L. Hill, *Thermodynamics of Small Systems* (Dover Phoenix, Mineola, New York, 2002).

[7] D. Reguera, J. M. Rubí, and J. M. G. Vilar, J. Phys. Chem. B **109**, 21502 (2005).

[8] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).

[9] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. (Academic, San Diego, CA, 2002).

[10] J. R. Savage and A. D. Dinsmore, Phys. Rev. Lett. **102**, 198302 (2009).

[11] G. Meng, N. Arkus, M. P. Brenner, and V. N. Manoharan, Science **327**, 560 (2010).

[12] S. Auer and D. Frenkel, Annu. Rev. Phys. Chem. **55**, 333 (2004).

[13] P. Hanggi, and P. Talkner, Rev. Mod. Phys. **62**, 251 (1990).

[14] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, 2nd ed., Springer Series in Synergetics, Vol. 18 (Springer, Berlin, 1996).

[15] G. Hummer, New J. Phys. **7**, 34 (2005).

[16] R. B. Best and G. Hummer, Proc. Natl. Acad. Sci. U.S.A. **107**, 34 (2010).

[17] J. Wedekind and D. Reguera, J. Phys. Chem. B **112**, 11060 (2008).

[18] G. E. Fernandes, D. J. Beltran-Villegas, and M. A. Bevan, Langmuir **24**, 10776 (2008).

[19] D. J. Beltran-Villegas and M. A. Bevan, Soft Matter **7**, 3280 (2011).

[20] D. J. Bicout and A. Szabo, J. Chem. Phys. **109**, 2325 (1998).

[21] P. R. ten Wolde, M. J. RuizMontero, and D. Frenkel, J. Chem. Phys. **104**, 9932 (1996).

[22] S. Anekal and M. A. Bevan, J. Chem. Phys. **122**, 034903 (2005).

[23] S. Anekal and M. A. Bevan, J. Chem. Phys. **125**, 034906 (2006).

[24] S. G. Anekal, P. Bahukudumbi, and M. A. Bevan, Phys. Rev. E **73**, 020403 (2006).

[25] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, 2nd ed. (Springer, Berlin, 1996).

[26] E. E. Borrero and F. A. Escobedo, J. Phys. Chem. B **113**, 6434 (2009).