The Journal of Chemical Physics

Limitations of field-theory simulation for exploring phase separation: The role of repulsion in a lattice protein model

Cite as: J. Chem. Phys. **156**, 015101 (2022); https://doi.org/10.1063/5.0070412
Submitted: 06 September 2021 • Accepted: 19 December 2021 • Published Online: 04 January 2022

Daniel Nilsson, Behruz Bozorg, Sandipan Mohanty, et al.

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

Limitations of field-theory simulation for exploring phase separation: The role of repulsion in a lattice protein model

Cite as: J. Chem. Phys. 156, 015101 (2022); doi: 10.1063/5.0070412
Submitted: 6 September 2021 • Accepted: 19 December 2021 •
Published Online: 4 January 2022

Daniel Nilsson,[1,a] Behruz Bozorg,[1,b] Sandipan Mohanty,[2,c] Bo Söderberg,[1,d] and Anders Irbäck[1,e]

AFFILIATIONS

[1] Computational Biology and Biological Physics, Department of Astronomy and Theoretical Physics, Lund University, Sölvegatan 14A, SE-22362 Lund, Sweden
[2] Institute for Advanced Simulation, Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52425 Jülich, Germany

[a] Electronic mail: daniel.nilsson@thep.lu.se
[b] Electronic mail: behruz.bozorg@thep.lu.se
[c] Electronic mail: s.mohanty@fz-juelich.de
[d] Electronic mail: bo.soderberg@thep.lu.se
[e] Author to whom correspondence should be addressed: anders.irback@thep.lu.se

ABSTRACT

Field-theory simulation by the complex Langevin method offers an alternative to conventional sampling techniques for exploring the forces driving biomolecular liquid–liquid phase separation. Such simulations have recently been used to study several polyampholyte systems. Here, we formulate a field theory corresponding to the hydrophobic/polar (HP) lattice protein model, with finite same-site repulsion and nearest-neighbor attraction between HH bead pairs. By direct comparison with particle-based Monte Carlo simulations, we show that complex Langevin sampling of the field theory reproduces the thermodynamic properties of the HP model only if the same-site repulsion is not too strong. Unfortunately, the repulsion has to be taken weaker than what is needed to prevent condensed droplets from assuming an artificially compact shape. Analysis of a minimal and analytically solvable toy model hints that the sampling problems caused by repulsive interaction may stem from loss of ergodicity.

I. INTRODUCTION

Advances over the past 15 years have identified liquid–liquid phase separation (LLPS) as a driver of compartmentalization in living cells.[1,2] Through LLPS, membraneless droplets are formed, with high concentrations of proteins and nucleic acids. In this process, it has been found that intrinsically disordered proteins (IDPs) often play a key role, and several such IDPs have been shown to phase separate on their own.[3–5]

To gain insight into the forces driving IDP LLPS, a broad set of theoretical and computational methods has been employed. The Flory–Huggins[6,7] and Voorn–Overbeek [8] mean-field methods provide useful analytical estimates, which, however, are insensitive to the ordering of the amino acids along the protein chains. By using the random phase approximation,[9,10] the sequence dependence of polyampholytes can be explored without resorting to extensive simulations, at the price of assuming Gaussian chains. To be able to avoid approximations made in the above methods, there have also been many studies of biomolecular LLPS based on explicit-chain simulation.[11–18] In particular, using various coarse-grained models, the sequence determinants of polyampholyte LLPS were elucidated.[11–15] However, particle-based simulation (PBS), with explicit chains, becomes computationally expensive for large systems, even with coarse-grained models.

Another approach to dense polymer systems is to use field-theory simulation (FTS),[19,20] which has recently been applied for the first time to biomolecular LLPS.[21] Here, by means of a Hubbard–Stratonovich transformation, the original polymer system

is reformulated as a statistical field theory, which can be investigated by simulation.[19,21] This approach has the advantage of removing direct interchain interactions, which makes it, at least formally, easy to increase the number of chains in the simulations. A disadvantage is that the effective energy of the field theory is complex-valued, which renders standard sampling techniques inadequate. A potential solution to this problem is offered by the complex Langevin method.[22–24] Indeed, using this method, several investigations of biomolecular LLPS in both one- and two-component systems have been reported.[21,25–29] In all these systems, phase separation was driven by Coulomb interactions, which are well suited for field-theoretic treatment.

In this article, we test the FTS approach on a hydrophobic/polar (HP) protein model, where phase separation is driven by short-range hydrophobic attraction rather than electrostatics. Since the FTS method requires the introduction of an auxiliary spatial grid, we deliberately consider a lattice-based protein model. In this way, it becomes possible to compare FTS and PBS results in a direct fashion, without having to extrapolate FTS results to the limit of vanishing lattice spacing. Specifically, we consider a variant of the well-known HP lattice model for protein folding,[30] with a finite same-site repulsion strength, Λ. As will be shown below, this model can be mapped onto a field theory with a simple structure. Note that with a lattice-based protein model, particle densities are, by construction, smeared. With this implicit smearing present, there is no need for the explicit Gaussian smearing typically used in continuous models.

The strength of the same-site repulsion, Λ, is a critically important parameter. On physical grounds, Λ has to be sufficiently large to prevent condensed clusters from collapsing to an artificially compact shape. If, on the other hand, Λ is taken too large, it turns out that the complex Langevin method breaks down. To elucidate these two conflicting requirements, we simulate and analyze in some detail a lattice gas, consisting of H particles rather than HP chains. Comparing FTS and PBS data, we find that the two Λ regions, where the respective requirements are met, unfortunately, do not overlap. To determine whether the sampling problems that we observe at large Λ are a peculiarity of our particular model or a more general problem associated with repulsive interactions, we construct a minimal toy model, which can be solved analytically. In the presence of strong repulsive interactions, we find that the complex Langevin method fails in this toy model as well. Finally, we present some examples of FTS results for systems of HP chains, which, again, are compared with PBS data for the same systems.

II. METHODS

A. Biophysical model

We consider a system of N linear chains with M beads each, on a simple cubic lattice with volume V and periodic boundaries in all three directions. The beads can be either hydrophobic (H) or polar (P). For simplicity, we assume that all N chains share the same sequence, which we write as $\sigma = (\sigma_1, \ldots, \sigma_M)$, where $\sigma_m = 1$ for an H bead and $\sigma_m = 0$ for a P bead. Throughout the paper, we use dimensionless values for energy and length, with the lattice spacing set to unity.

The interaction potential is pairwise additive, $U = \sum_{i<j} u_{ij}$, where the sum runs over all pairs of beads, both intrachain and interchain pairs. The pair potential u_{ij} has a repulsive part, which assigns an energy penalty $\Lambda > 0$ to any pair of beads residing on the same lattice site. In addition, there is an attractive nearest-neighbor interaction, which is felt only by HH pairs. In total, thus, the pair potential is given by

$$u_{ij} = \begin{cases} \Lambda & \text{if beads } i \text{ and } j \text{ are on the same site,} \\ -\sigma_i\sigma_j & \text{if beads } i \text{ and } j \text{ are nearest neighbors,} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

The full potential may, thus, be written as

$$U = U_r + U_a, \quad (2)$$

where $U_r = \Lambda \times \{\text{number of same} - \text{site} - \text{pairs}\}$ and $U_a = -\{\text{number of nearest-neighbor HH pairs}\}$.

The thermodynamic behavior of the system at inverse temperature β is determined by the partition function

$$Z = \sum_C e^{-\beta U}, \quad (3)$$

where the sum runs over all possible configurations C of the N-chain system.

To obtain a field theory representation of this particle-based system, we first express the potential U in terms of bead counts rather than bead positions. To this end, we consider the ansatz

$$U_{bc} = \frac{\Lambda}{2}\sum_{\mathbf{r}} n(\mathbf{r})^2 + \frac{1}{2}\sum_{\mathbf{r},\hat{\mu}} \tilde{n}_H(\mathbf{r},\hat{\mu})^2,$$

with

$$\tilde{n}_H = \alpha n_H(\mathbf{r}+\hat{\mu}) - \alpha^* n_H(\mathbf{r}), \quad (4)$$

where $\mathbf{r}$ denotes a lattice site, $\hat{\mu}$ is one of the three lattice unit vectors, α is a complex parameter, $n(\mathbf{r})$ is the total number of beads at site $\mathbf{r}$, and $n_H(\mathbf{r})$ is the number of H beads at site $\mathbf{r}$. The first sum is over lattice sites, while the second is over links. Note that the second sum does not have the common form $\sum_{\mathbf{r}}[\sum_{\mathbf{r}'}\Gamma(\mathbf{r},\mathbf{r}')n_H(\mathbf{r}')]^2$, where Γ is a smearing matrix and the outer sum is over sites rather than links. The link-based form in Eq. (4) makes it possible to avoid interactions beyond the nearest-neighbor distance.

The bead counts n and n_H can be written as

$$n(\mathbf{r}) = \sum_i \delta(\mathbf{r},\mathbf{r}_i) \quad \text{and} \quad n_H(\mathbf{r}) = \sum_i \sigma_i \delta(\mathbf{r},\mathbf{r}_i), \quad (5)$$

where $\mathbf{r}_i$ and σ_i denote, respectively, the location and type of bead i and δ is the Kronecker delta. Using Eq. (5), U_{bc} [Eq. (4)] can be rewritten as a sum over bead pairs. One finds that the n-dependent part of U_{bc} is equal to U_r plus a constant self-energy term, given by $NM\Lambda/2$. All the results presented below are based on simulations in the canonical ensemble, for which this shift is unimportant. The $\tilde{n}_H$-dependent part of U_{bc} generally contains both same-site and nearest-neighbor interactions. However, this mixing can be avoided by choosing the parameter $\alpha = e^{i\pi/4}$. With this α, this part of U_{bc} becomes equal to U_a, which means that

$$U_{bc} = U + \frac{NM\Lambda}{2}. \quad (6)$$

All numerical results presented below were obtained using $\alpha = e^{i\pi/4}$.

B. Field theory

It follows from above that, adding the constant self-energy term to U, the partition function [Eq. (3)] can be expressed as

$$Z = \sum_C e^{-\beta U_{\rm bc}}, \tag{7}$$

where $U_{\rm bc}$ [Eq. (4)] depends quadratically on both the site variables $n(\mathbf{r})$ and the link variables $\tilde{n}_{\rm H}(\mathbf{r},\hat{\mu})$. These quadratic dependencies can be linearized by introducing auxiliary fields, $w(\mathbf{r})$ and $\varphi(\mathbf{r},\hat{\mu})$, by means of the Hubbard–Stratonovich method. Note that since $\tilde{n}_{\rm H}(\mathbf{r},\hat{\mu})$ is associated with links, the corresponding field $\varphi(\mathbf{r},\hat{\mu})$ can be seen as a discrete version of a vector field, living on the links of the lattice.

The fields are introduced through the relations

$$\exp\left(-\frac{\beta\Lambda}{2}n(\mathbf{r})^2\right) \propto \int dw(\mathbf{r}) \exp\left(-\frac{1}{2\beta\Lambda}w(\mathbf{r})^2 - iw(\mathbf{r})n(\mathbf{r})\right), \tag{8}$$

$$\exp\left(-\frac{\beta}{2}\tilde{n}_{\rm H}(\mathbf{r},\hat{\mu})^2\right) \propto \int d\varphi(\mathbf{r},\hat{\mu}) \exp\left(-\frac{1}{2\beta}\varphi(\mathbf{r},\hat{\mu})^2 - i\varphi(\mathbf{r},\hat{\mu})\tilde{n}_{\rm H}(\mathbf{r},\hat{\mu})\right).$$

An alternative approach for the attractive term, which has proven useful for other interactions, would be to choose a version of the Hubbard–Stratonovich transform with a site field $\varphi(\mathbf{r})$ interacting directly with the local H count $n_{\rm H}(\mathbf{r})$. Then, the quadratic self-interaction for the φ field would involve the inverse of the nearest-neighbor stencil, which can be dealt with, e.g., by a diagonalizing discrete Fourier transform (DFT). However, as can be seen from the DFT, the eigenvalues are given by a sum of three cosines, which results in an abundance of zero and near-zero modes. The zero modes make the nearest-neighbor stencil non-invertible but could be handled, e.g., by clamping the corresponding field combinations to zero. The near-zero modes, however, will severely slow down the Langevin dynamics, by necessitating a small time step. Hence, for the particular form of the attractive term used here, we see no better choice than to use our link field approach.

The partition function for the fields then becomes

$$Z \propto Z_{\rm FT} = \int \prod_{\mathbf{r}} dw(\mathbf{r}) \prod_{\mathbf{r},\hat{\mu}} d\varphi(\mathbf{r},\hat{\mu}) e^{-H[w,\varphi]}, \tag{9}$$

where the effective energy $H[w,\varphi]$ is given by

$$H[w,\varphi] = \frac{1}{2\nu}\sum_{\mathbf{r}} w(\mathbf{r})^2 + \frac{1}{2\eta}\sum_{\mathbf{r},\hat{\mu}} \varphi(\mathbf{r},\hat{\mu})^2 - N \ln Q[w,\varphi], \tag{10}$$

with $\nu = \beta\Lambda$ and $\eta = \beta$. Here, $Q[w,\varphi]$ is a conditional single-chain partition function, given by

$$Q[w,\varphi] = \sum_{C_1} \exp\left[-i\sum_{m=1}^{M}\left(w(\mathbf{r}_m) + \sigma_m \sum_{\hat{\mu}}\left[\alpha\varphi(\mathbf{r}_m - \hat{\mu},\hat{\mu}) - \alpha^*\varphi(\mathbf{r}_m,\hat{\mu})\right]\right)\right], \tag{11}$$

where the outer sum is over single-chain configurations $C_1 = (\mathbf{r}_1,\ldots,\mathbf{r}_M)$, corresponding to a random walk on the cubic lattice with $M-1$ unit steps. The evaluation of Q, given w and φ, can be conveniently organized by rewriting Eq. (11) in the form

$$\begin{aligned} Q[w,\varphi] &= \sum_{C_1}\prod_{m=1}^{M}\chi_{\sigma_m}(\mathbf{r}_m) \\ &= \sum_{\mathbf{r}_1,\ldots,\mathbf{r}_M} \chi_{\sigma_M}(\mathbf{r}_M)\cdots T(\mathbf{r}_3,\mathbf{r}_2)\chi_{\sigma_2}(\mathbf{r}_2)T(\mathbf{r}_2,\mathbf{r}_1)\chi_{\sigma_1}(\mathbf{r}_1), \end{aligned} \tag{12}$$

where $T(\mathbf{r},\mathbf{r}') = 1$ if $\mathbf{r}$ and $\mathbf{r}'$ are nearest neighbors, $T(\mathbf{r},\mathbf{r}') = 0$ otherwise, and

$$\chi_\sigma(\mathbf{r}) = \begin{cases} e^{-iw(\mathbf{r})} & \text{if } \sigma = 0, \\ e^{-iw(\mathbf{r}) - i\sum_{\hat{\mu}}\left[\alpha\varphi(\mathbf{r}-\hat{\mu},\hat{\mu}) - \alpha^*\varphi(\mathbf{r},\hat{\mu})\right]} & \text{if } \sigma = 1. \end{cases} \tag{13}$$

C. Extracting polymer properties from the fields

In the field representation, the original bead count variables are not readily available, but hidden in the conditional partition function Q. However, it is possible to derive useful identities between bead count and field correlations.[19] A whole series of such identities can be derived by noting that the fields $w(\mathbf{r})$ and $\varphi(\mathbf{r},\hat{\mu})$ [Eq. (8)] can be expressed as

$$\begin{aligned} w(\mathbf{r}) &= u(\mathbf{r}) - i\nu n(\mathbf{r}), \\ \varphi(\mathbf{r},\hat{\mu}) &= u_{\rm H}(\mathbf{r},\hat{\mu}) - i\eta\tilde{n}_{\rm H}(\mathbf{r},\hat{\mu}), \end{aligned} \tag{14}$$

where $u(\mathbf{r})$ and $u_{\rm H}(\mathbf{r},\hat{\mu})$ are auxiliary zero-mean Gaussian fields with $\langle u(\mathbf{r})u(\mathbf{r}')\rangle = \nu\delta(\mathbf{r},\mathbf{r}')$ and $\langle u_{\rm H}(\mathbf{r},\hat{\mu})u_{\rm H}(\mathbf{r}',\hat{\mu}')\rangle = \eta\,\delta(\mathbf{r},\mathbf{r}')\delta(\hat{\mu},\hat{\mu}')$. At the one- and two-point levels, one finds the identities

$$\begin{aligned} \langle w(\mathbf{r})\rangle &= -i\nu\langle n(\mathbf{r})\rangle \qquad (= -i\nu NM/V), \\ \langle \varphi(\mathbf{r},\hat{\mu})\rangle &= -i\eta\langle\tilde{n}_{\rm H}(\mathbf{r},\hat{\mu})\rangle \qquad (= 2\eta\,{\rm Im}(\alpha)NM_{\rm H}/V), \\ \langle w(\mathbf{r})w(\mathbf{r}')\rangle &= \nu\delta(\mathbf{r},\mathbf{r}') - \nu^2\langle n(\mathbf{r})n(\mathbf{r}')\rangle, \\ \langle \varphi(\mathbf{r},\hat{\mu})\varphi(\mathbf{r}',\hat{\mu}')\rangle &= \eta\delta(\mathbf{r},\mathbf{r}')\delta(\hat{\mu},\hat{\mu}') - \eta^2\langle\tilde{n}_{\rm H}(\mathbf{r},\hat{\mu})\tilde{n}_{\rm H}(\mathbf{r}',\hat{\mu}')\rangle, \end{aligned} \tag{15}$$

where $M_{\rm H}$ denotes the number of H beads per chain. From the last two of these identities, consequently,

$$\begin{aligned} \langle U_{\rm r}\rangle + \frac{NM\Lambda}{2} &= \frac{V}{2\beta} - \frac{1}{2\beta^2\Lambda}\sum_{\mathbf{r}}\langle w(\mathbf{r})^2\rangle, \\ \langle U_{\rm a}\rangle &= \frac{3V}{2\beta} - \frac{1}{2\beta^2}\sum_{\mathbf{r},\hat{\mu}}\langle\varphi(\mathbf{r},\hat{\mu})^2\rangle. \end{aligned} \tag{16}$$

The average particle-based total energy $U = U_{\rm r} + U_{\rm a}$ [Eq. (3)], therefore, can be obtained as the field-theory average of the estimator,

$$U_{\rm FT} = \frac{2V}{\beta} - \frac{NM\Lambda}{2} - \frac{1}{2\beta^2}\left[\frac{1}{\Lambda}\sum_{\mathbf{r}} w(\mathbf{r})^2 + \sum_{\mathbf{r},\hat{\mu}}\varphi(\mathbf{r},\hat{\mu})^2\right]. \tag{17}$$

When studying phase separation, a common choice is to use elongated simulation boxes, with volume $V = L_z L^2$, in which droplets tend to be slab-like rather than spherical. Droplets

can then be detected by determining the density profile $\rho(z) = L^{-2}\sum_{x,y} n(x,y,z)$. The simultaneous presence of two bulk phases with different densities leads to a large spatial variance of $\rho(z)$, defined as

$$\sigma_\rho^2 = \frac{1}{L_z-1}\sum_{z=1}^{L_z}(\rho(z)-\overline{\rho})^2 = \frac{1}{L_z-1}\left(\sum_{z=1}^{L_z}\rho(z)^2 - L_z\overline{\rho}^2\right), \tag{18}$$

where $\overline{\rho} = NM/V$ denotes the total density. Using Eq. (15), it can be easily verified that the ensemble average of this quantity, $\langle\sigma_\rho^2\rangle$, can be determined by using the field-theoretic estimator,

$$\sigma_{\rho,\mathrm{FT}}^2 = \frac{1}{L_z-1}\left(\frac{L_z}{L^2\beta\Lambda} - L_z\overline{\rho}^2 - \frac{1}{(\beta\Lambda)^2}\sum_z \rho_w(z)^2\right), \tag{19}$$

where $\rho_w(z) = L^{-2}\sum_{x,y} w(x,y,z)$.

D. Complex Langevin sampling

The statistical field theory defined by Eq. (9) has a complex-valued effective energy $H[w,\varphi]$ and, therefore, a complex weight function e^{-H}, which renders sampling techniques such as Markov chain Monte Carlo inadequate. In principle, this problem can be overcome by sampling the distribution $e^{-\mathrm{Re}H}$ and using reweighting methods. However, this approach typically requires estimating rapidly fluctuating observables, which makes it inefficient. A potentially useful alternative is to use Langevin dynamics,[22–24] defined by

$$\begin{aligned}\dot{w}(\mathbf{r}) &= -\frac{\partial H}{\partial w(\mathbf{r})} + \sqrt{2}\,\Xi_w(\mathbf{r},t),\\ \dot{\varphi}(\mathbf{r},\hat{\mu}) &= -\frac{\partial H}{\partial \varphi(\mathbf{r},\hat{\mu})} + \sqrt{2}\,\Xi_\varphi(\mathbf{r},\hat{\mu},t),\end{aligned} \tag{20}$$

where t is the Langevin time, a dot indicates the time derivative, Ξ_w is the standard Gaussian noise with zero mean and correlations given by $\langle\Xi_w(\mathbf{r},t)\Xi_w(\mathbf{r}',t')\rangle = \delta(\mathbf{r},\mathbf{r}')\delta(t-t')$, and similarly for Ξ_φ. In a simulation, these continuous-time equations have to be discretized. A simple discrete form is

$$\begin{aligned} w(\mathbf{r})_{k+1} &= w(\mathbf{r})_k - dt\,\frac{\partial H}{\partial w(\mathbf{r})}\bigg|_k + \sqrt{2dt}\,\xi_w(\mathbf{r},t_k)\\ &= (1-\nu\,dt)\,w(\mathbf{r})_k + dt\,\frac{N}{Q}\frac{\partial Q}{\partial w(\mathbf{r})}\bigg|_k + \sqrt{2dt}\,\xi_w(\mathbf{r},t_k),\end{aligned} \tag{21}$$

$$\begin{aligned}\varphi(\mathbf{r},\hat{\mu})_{k+1} &= \varphi(\mathbf{r},\hat{\mu})_k - dt\,\frac{\partial H}{\partial \varphi(\mathbf{r},\hat{\mu})}\bigg|_k + \sqrt{2dt}\,\xi_\varphi(\mathbf{r},\hat{\mu},t_k)\\ &= (1-\eta\,dt)\,\varphi(\mathbf{r},\hat{\mu})_k + dt\,\frac{N}{Q}\frac{\partial Q}{\partial \varphi(\mathbf{r},\hat{\mu})}\bigg|_k\\ &\quad + \sqrt{2dt}\,\xi_\varphi(\mathbf{r},\hat{\mu},t_k),\end{aligned}$$

where dt is the time step and k is the time index, while $\xi_w(\mathbf{r},t_k)$ and $\xi_\varphi(\mathbf{r},\hat{\mu},t_k)$ are two sets of independent Gaussian random variables with zero mean and unit variance. In Eq. (21), it is possible and potentially advantageous to use different time steps for the w and φ fields, dt_w and dt_φ, depending on ν and η.[31,32] However, throughout this paper, we use the same dt for all degrees of freedom.

Due to the complex nature of H, the fields will not be restricted to real values when evolving according to Eq. (21), but will wander off into the complex plane. Thus, we will have a probability distribution over complex-valued fields or, equivalently, a joint probability distribution over their real and imaginary parts. Under fairly general conditions, the Langevin dynamics allows for this distribution over complex fields to converge to one that mimics the formal, complex-valued Boltzmann distribution over real fields, e^{-H}, in the sense that expectation values of *analytic* functions of the fields will converge to the correct values. However, it is well-known that the success of the method is system-dependent.[33,34]

E. Simulation details

We test the FTS method on systems consisting of H particles or multiple copies of one of two different ten-bead HP chains. For comparison, we apply PBS techniques to the same systems, to generate reference data.

The FTS results are time averages over Langevin trajectories, generated using Eq. (21) with a fixed step size in the range $5\times10^{-6} \le dt \le 10^{-4}$. The simulations are started from randomly perturbed uniform field configurations. Each run covers a total Langevin time of 4×10^3 (H particles) or 5×10^3 (HP chains), the first 20% of which is discarded for thermalization.

The PBS results are obtained using Monte Carlo methods. For H particles, a single type of move is employed, namely, displacement of individual particles to nearest-neighbor sites on the lattice. A majority of the results are from fixed-temperature simulations with the Metropolis algorithm. All these runs are started from random initial states, and 20% of each run is discarded for thermalization. However, near the condensation/evaporation transition, this sampling method becomes inefficient, because transitions between states with and without a droplet are rare. To overcome this problem, some of our simulations use the Wang–Landau algorithm,[35–37] which, in particular, facilitates the determination of the condensation/evaporation temperature.

The PBS results for HP chains are based on a set of three elementary moves. The first move alters the internal structure of a random chain, by rotating one of its $M-1$ bond vectors. The second move is a rigid-body translation or rotation of an individual chain. The third and final move is a rigid-body translation of a cluster of chains. The construction of the cluster to be moved is stochastic, following a Swendsen–Wang type procedure.[38,39] All three moves are subject to a Metropolis accept/reject test.

III. RESULTS

Above, we gave a field-theoretic representation (Sec. II B) of the HP lattice protein model with finite same-site repulsion (Sec. II A). In this section, we evaluate to what extent simulation of this field theory by the complex Langevin method (Sec. II D) reproduces the thermodynamic properties of the HP model, using reference data obtained by conventional particle-based Monte Carlo simulation. First, we investigate in some detail the case of a lattice gas, where the system consists of (one-bead) H particles. To shed some light upon the findings for the lattice gas, we then introduce a minimal, analytically solvable toy model, whose behavior under Langevin dynamics can be analyzed and understood. Finally, we present some results from simulations with ten-bead HP chains.

A. H particles

Throughout this subsection, we consider systems consisting of 64 H particles at fixed density $\rho = N/V = 0.125$. The lattice used is either cubic (8^3) or elongated in one direction (32×4^2). In the latter case, condensed droplets assume a slab-like rather than a spherical shape. We study how the ability of the FTS method to reproduce results obtained using conventional PBS techniques depends on the two parameters of the model, the repulsion strength Λ, and the inverse temperature β. Another important issue is how large Λ has to be taken in order to prevent condensed droplets from becoming artificially compact.

Figures 1(a) and 1(c) compare FTS and PBS data for the repulsive and attractive energies U_r and U_a, respectively, for different Λ at fixed $\beta = 0.3$, on a cubic lattice. At this β, the system is in an uncondensed gas state for all Λ values considered. The FTS results are in agreement with the PBS data for $\Lambda \lesssim 3$, but deviations develop as Λ is increased. That the FTS method suffers from sampling errors at large Λ is underscored by the fact that the quantity U_r, which is positive by construction, turns negative.

The corresponding data at $\beta = 0.4$ follow a similar pattern [Figs. 1(b) and 1(d)], although the accuracy of the FTS method starts deteriorating at a lower Λ in this case. At $\beta = 0.4$, we omitted the data obtained for $\Lambda = 0.5$. The reason for this is that a condensation transition takes place as Λ is reduced from 1 to 0.5, which leads to U_r and U_a values far outside the plotted ranges.

For fixed $\Lambda = 0.5$, the above results imply that a temperature-induced condensation transition occurs as β is increased from 0.3 to 0.4. This transition is illustrated in Fig. 2(a), which shows both PBS and FTS data for the total energy $U = U_r + U_a$. The curve representing the PBS data is computed by using the Wang–Landau algorithm, along with reweighting techniques. The condensation transition occurs at $\beta_t \approx 0.375$, with β_t defined by having the maximum heat capacity. For $\Lambda = 0.5$, the Wang–Landau method is much more efficient than standard constant-temperature Monte Carlo, which becomes slow in the vicinity of the condensation transition, due to a strongly bimodal energy distribution, $P(U)$ [Fig. 2(b)]. The two peaks in $P(U)$ correspond to states with and without a droplet, respectively. At the transition temperature, the valley between the two peaks is statistically suppressed by about eight orders of magnitude, despite the modest size of the system.

Figure 2(a) also shows the FTS results for the same Λ, for four different β, two on each side of the condensation transition. The FTS data points fall close to the PBS curve, which shows that the FTS method is able to describe the transition quite well. It is worth

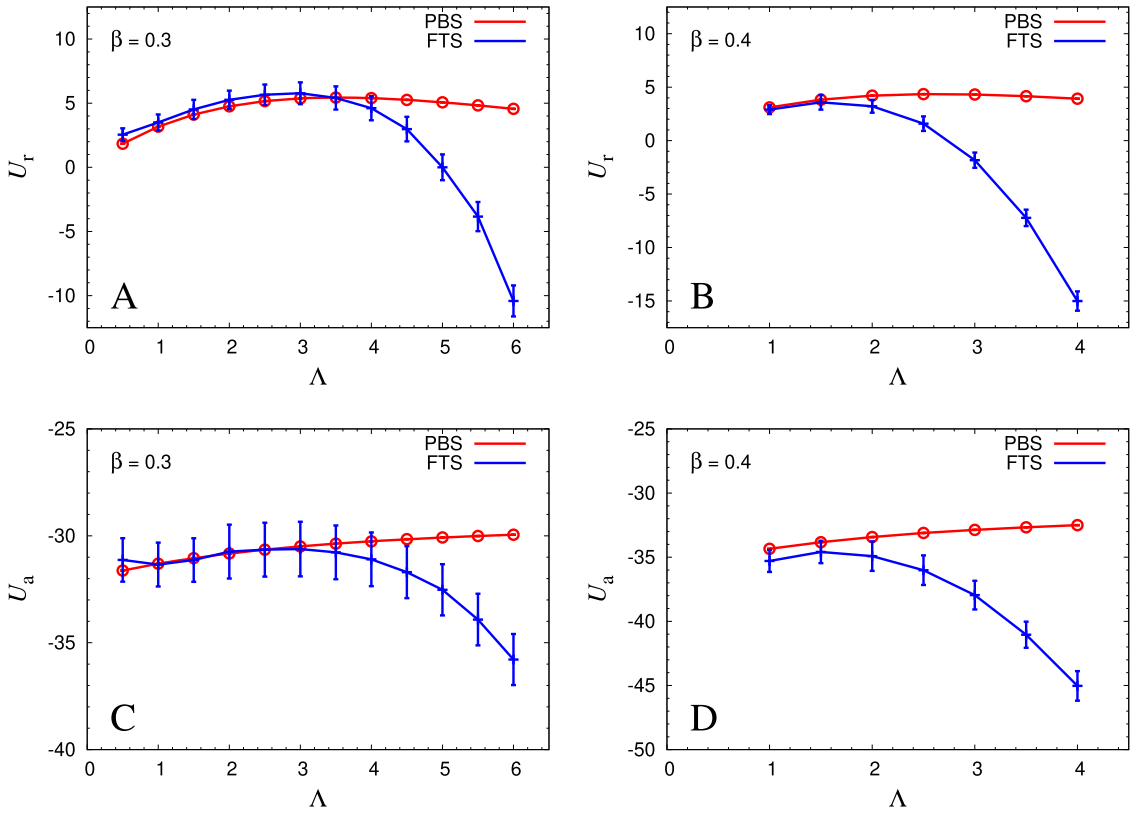

FIG. 1. Λ-dependence of the repulsion and attraction energies U_r and U_a at $\beta = 0.3$ and $\beta = 0.4$, for a system of 64 H particles on an 8^3 lattice, as obtained using PBS (red) and FTS (blue). Lines are drawn to guide the eye. (a) U_r at $\beta = 0.3$, (b) U_r at $\beta = 0.4$, (c) U_a at $\beta = 0.3$, and (d) U_a at $\beta = 0.4$.

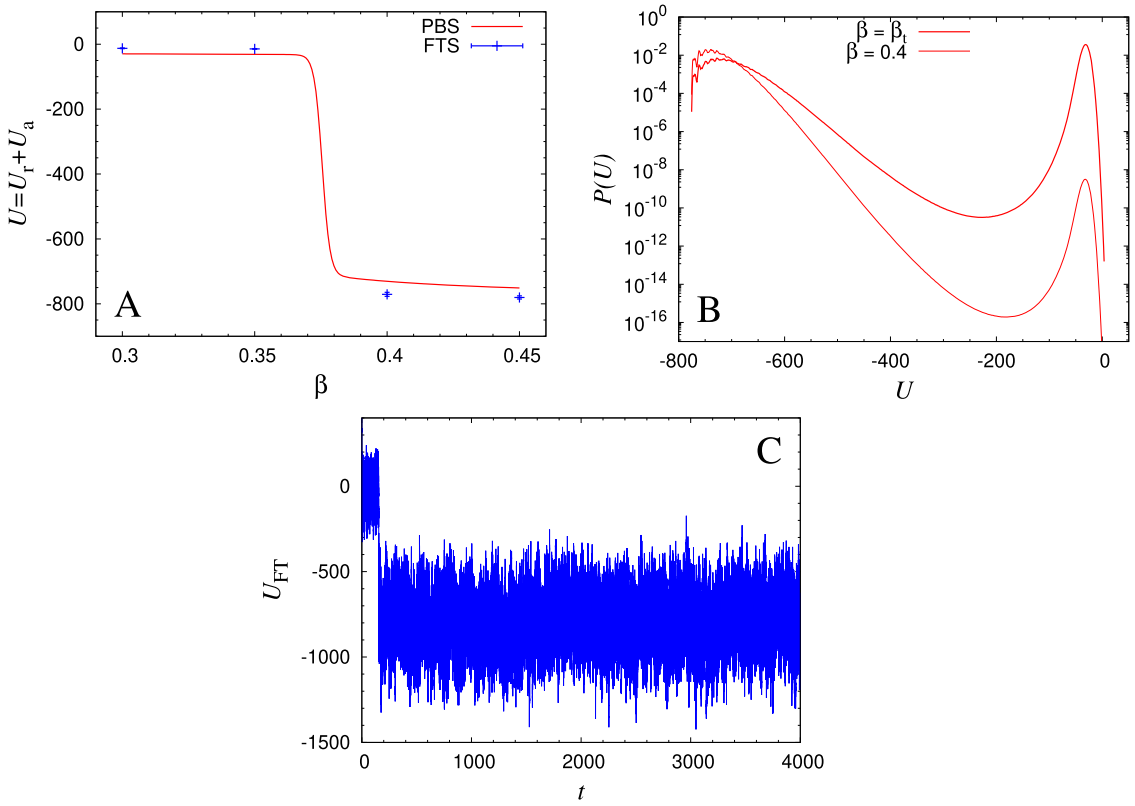

FIG. 2. Droplet condensation in a system of 64 H particles on an 8^3 lattice for $\Lambda = 0.5$. (a) β-dependence of the total energy $U = U_r + U_a$. The red curve represents PBS data, obtained with the Wang–Landau algorithm.[35] Blue symbols indicate the FTS results. (b) Energy distribution, $P(U)$, using a logarithmic scale at $\beta = 0.375 \approx \beta_t$ (thick line) and $\beta = 0.4$ (thin line), based on the PBS data obtained with the Wang–Landau method. (c) Time evolution of the field-theoretic estimator U_{FT} of U [Eq. (17)] in an FTS run at $\beta = 0.4$.

noting that all the FTS runs are started from random initial field configurations. Therefore, for $\beta > \beta_t$, the Langevin dynamics has to bring the system from a random state to field configurations corresponding to a droplet-containing state. Figure 2(c) shows the time evolution of the field-theoretic energy estimator U_{FT} [Eq. (17)] in a run at $\beta = 0.4 \approx 1.07\beta_t$. At this β, the energy distribution $P(U)$ remains bimodal [Fig. 2(b)]; the low-energy peak dominates, but a high-energy peak is still present. Consistent with this, the FTS run is initially trapped in a high-U_{FT} state [Fig. 2(c)]. However, the system is able to escape to a state in which U_{FT} matches well with the position of the low-energy peak in $P(U)$.

Next, we investigate the droplet condensation transition in some more detail for three values of Λ (0.5, 2.0, and 5.0), using an elongated simulation box (32×4^2). To this end, we first consider the longitudinal distribution of particles, $\rho(z) = L^{-2} \sum_{x,y} n(x,y,z)$. In particular, we compute the spatial variance of this distribution, σ_ρ^2 [Eq. (18)], and its field-theoretic estimator, $\sigma_{\rho,FT}^2$ [Eq. (19)]. The formation of a dense droplet in a dilute background leads to an increased spatial variance σ_ρ^2. Figure 3(a) shows PBS and FTS data for σ_ρ^2 and $\sigma_{\rho,FT}^2$, respectively, in the vicinity of the inverse transition temperature, β_t, for all three choices of Λ. As in Fig. 2(a), the FTS data agree quite well with the PBS data for $\Lambda = 0.5$. By contrast, but not surprisingly given the data in Fig. 1, the FTS method fails to properly describe the condensation transition for $\Lambda = 2.0$ and $\Lambda = 5.0$.

The change in σ_ρ^2 near β_t is abrupt and large for $\Lambda = 0.5$, while becoming less drastic as Λ is increased [Fig. 3(a)]. The abruptness of the transition for small Λ is linked to the collapse of condensed droplets, which leads to artificially low energies for droplet-containing configurations. Figure 3(b) shows the total number of lattice sites hosting at least one of the 64 particles in the system. For $\Lambda = 0.5$, it can be seen that this number, n_s, drops from ≈60 to ≈10 upon droplet condensation. By contrast, for $\Lambda = 5.0$, n_s stays above 63 throughout the β range studied, $0.93 \lesssim \beta/\beta_t \lesssim 1.07$.

In summary, in the lattice gas studied here, in order for the FTS sampling errors to stay small, the repulsion strength Λ must not be too large. At the same time, in order to prevent the formation of artificially compact droplets, Λ must not be too small. Unfortunately, at least with the standard Langevin scheme used here, the Λ regions where these two requirements are met do not overlap, as is illustrated by the results for $\Lambda = 2.0$ in Fig. 3. This Λ is too large to avoid large sampling errors [Fig. 3(a)],

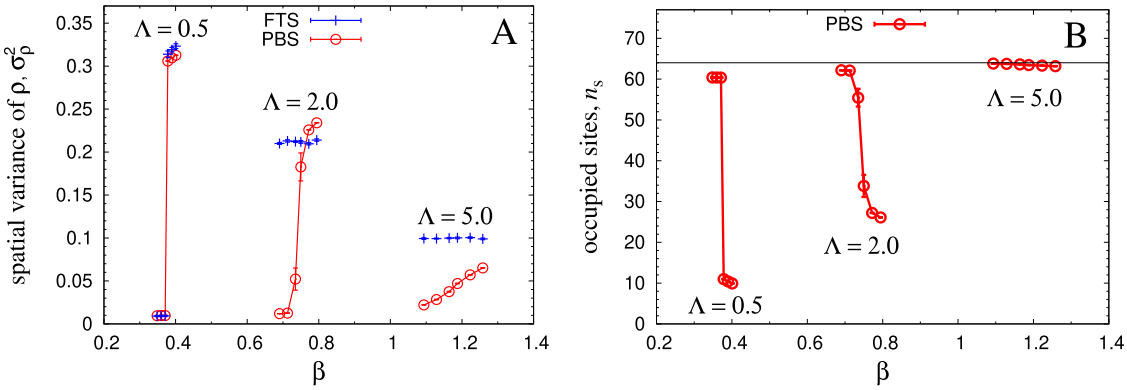

FIG. 3. Droplet condensation in a system of 64 H particles on a 32×4^2 lattice for three values of Λ (0.5, 2.0, and 5.0), studied using PBS (red) and FTS (blue). For each Λ, data were acquired for six values of β/β_t (0.93, 0.96, 0.99, 1.01, 1.04, and 1.07), where β_t is the inverse transition temperature, defined as the heat capacity maximum. Lines are drawn to guide the eye. (a) Spatial variance of the density $\rho(z)$, calculated using Eq. (18) (PBS) or Eq. (19) (FTS). (b) The number of lattice sites hosting at least one particle, n_s, with its maximal value (64) indicated by the horizontal line.

but still too small to prevent condensed droplets from collapsing [Fig. 3(b)].

B. Toy model

In this subsection, we turn to a minimal toy model to elucidate how increased repulsion strength can cause sampling problems in the FTS approach.

Thus, we consider a single particle (or a gas of N identical ones) on a lattice with only two sites, labeled 1 and 2, respectively, with two possible types of same-site pair interactions, of either repulsive or attractive nature. The repulsive interaction gives a penalty of $v \geq 0$ for each same-site pair, while the attractive one instead gives a reward $\eta \geq 0$, as expressed by the respective interaction energies,

$$\begin{aligned} \beta U_r &= \frac{v}{2}\left(n_1^2 + n_2^2\right), \\ \beta U_a &= -\frac{\eta}{2}\left(n_1^2 + n_2^2\right). \end{aligned} \tag{22}$$

In a similar way as for the main model, these systems can be transformed into field theories, with the respective effective energies,

$$\begin{aligned} H_r(w_1, w_2) &= \frac{1}{2v}\left(w_1^2 + w_2^2\right) - N \log Q_r, \\ H_a(\varphi_1, \varphi_2) &= \frac{1}{2\eta}\left(\varphi_1^2 + \varphi_2^2\right) - N \log Q_a. \end{aligned} \tag{23}$$

Here, w_1, w_2 and φ_1, φ_2 are the site fields for the repulsive and attractive cases, respectively, while Q_r and Q_a are the conditional, single particle partition functions, given by $Q_r = e^{-iw_1} + e^{-iw_2}$ and $Q_a = e^{-\varphi_1} + e^{-\varphi_2}$.

Conveniently, H in each case separates in terms of the sum and difference of the fields on the two sites, given by $W = w_1 + w_2$, $w = w_1 - w_2$ for the repulsive case, and $\Phi = \varphi_1 + \varphi_2$, $\varphi = \varphi_1 - \varphi_2$ for the attractive one. The quadratic terms then become $\left(W^2 + w^2\right)/4v$ and $\left(\Phi^2 + \varphi^2\right)/4\eta$, respectively, while the conditional partition functions factorize as, respectively, $Q_r = e^{-iW/2}\left(e^{-iw/2} + e^{iw/2}\right)$ and $Q_a = e^{-\Phi/2}\left(e^{-\varphi/2} + e^{\varphi/2}\right)$.

As a result, the summed fields, W or Φ, have quadratic effective energies,

$$\begin{aligned} h_r(W) &= \frac{W^2}{4v} + iN\frac{W}{2}, \\ h_a(\Phi) &= \frac{\Phi^2}{4\eta} + N\Phi, \end{aligned} \tag{24}$$

and become simple Gaussian variables with rather trivial Langevin dynamics. Neglecting these, we can focus on the non-trivial difference fields, w or φ, with the effective energies,

$$\begin{aligned} H_r(w) &= \frac{w^2}{4v} - N \log \cos\left(\frac{w}{2}\right), \\ H_a(\varphi) &= \frac{\varphi^2}{4\eta} - N \log \cosh\left(\frac{\varphi}{2}\right). \end{aligned} \tag{25}$$

Applying conventional (complex) Langevin dynamics to the original fields leads to the following dynamics for the difference fields:

$$\begin{aligned} \dot{w} &= -\frac{w}{v} - N \tan\left(\frac{w}{2}\right) + 2\,\Xi_r, \\ \dot{\varphi} &= -\frac{\varphi}{\eta} + N \tanh\left(\frac{\varphi}{2}\right) + 2\,\Xi_a, \end{aligned} \tag{26}$$

where Ξ_r is a standard Gaussian noise with zero mean and $\langle \Xi_r(t)\Xi_r(t')\rangle = \delta(t - t')$, and similarly for Ξ_a. As before, the continuous-time evolution in Eq. (26) has to be approximated by discrete time equations.

As it turns out, the dynamics differs significantly between the two cases, and we will, therefore, consider them separately.

1. Repulsive case

For the repulsive case, the target distribution on the real w line, determined by H_r [Eq. (25)], reads

$$P(w) \propto e^{-w^2/2\nu} \cos^N(w/2), \tag{27}$$

which is real on the real line, but with a varying sign, at least for odd N, due to the cosine factor. Henceforth, we will assume $N = 1$. Figure 4 illustrates the drift in the complex w plane for two values of ν.

The zeros of the cosine at odd multiples of π define poles of H, at which the drift term in the Langevin equation for w diverges [Eq. (26)]. This leads to wild behavior, unless regulated, e.g., with a dynamical time step. Between the poles, the drift is smooth and leaves the real line an invariant manifold that attracts the motion. On the real w line, the poles at odd multiples of π are repulsive under the drift and alternate with attracting fixed points. The noise term, however, spreads out the trajectories.

Thus, it is clear that the real line acts as an attractor for the Langevin dynamics. In computer simulation with a finite time step, trajectories will be trapped on the real line, in the intervals between consecutive poles, and only occasionally pass to a neighboring interval. Within each interval, the resulting distribution will be proportional to $|P(w)|$, but with different random normalization constants in the different intervals, in a manner that depends on the particular simulation details.

Due to the Gaussian factor in P, for small enough repulsion strength, $\nu \ll 1$, the distribution is dominated by the central peak around the fixed point at $w = 0$, and the error in the Gaussian tail can be neglected. Hence, we would expect essentially correct long-term averages from computer simulations of the Langevin dynamics for small enough ν, while they would deteriorate for larger ν.

2. Attractive case

For the attractive case, on the other hand, the target distribution on the real φ line, determined by H_a [Eq. (25)], reads

$$P(\varphi) \propto e^{-\varphi^2/2\eta} \cosh^N(\varphi/2), \tag{28}$$

which is real and positive on the entire real line. For simplicity, we again focus on the case $N = 1$. Figure 5 illustrates the drift in the complex φ plane for two values of η.

In Eq. (28), the cosine of Eq. (27) is replaced by a cosh, which means that $P(\varphi)$ instead has zeroes on the imaginary axis. These zeroes again correspond to poles of H, but are less disturbing, being away from the real φ line. As in the w case, the real line is invariant, but the drift term now is smooth there [Eq. (26)]. However, the dynamics close to the real line depends on the size of the attraction strength η.

For $\eta < 2$, the real line is everywhere attracting, and the drift has a single attractive fixed point there, $\varphi = 0$, with a basin of attraction containing the whole real line. This indicates that a numerical simulation of the Langevin dynamics [Eq. (26)] will result in long-term averages consistent with $P(\varphi)$.

At $\eta = 2$, the system undergoes a pitchfork bifurcation, where the central fixed point at $\varphi = 0$ turns unstable, while a previously repelling pair of fixed points on the imaginary line have closed in on the origin and instead becomes a pair of attracting fixed points on the real line, on either side of the origin.

For $\eta > 2$, the real line locally attracts only outside a pair of points lying inside the new attracting fixed points; however, a strip around the real line, $|\mathrm{Im}\,\varphi| < \pi$, is attracting. Within this strip, there are no sampling barriers in the real direction, indicating that Langevin sampling may not suffer from the same problems as in the repulsive case.

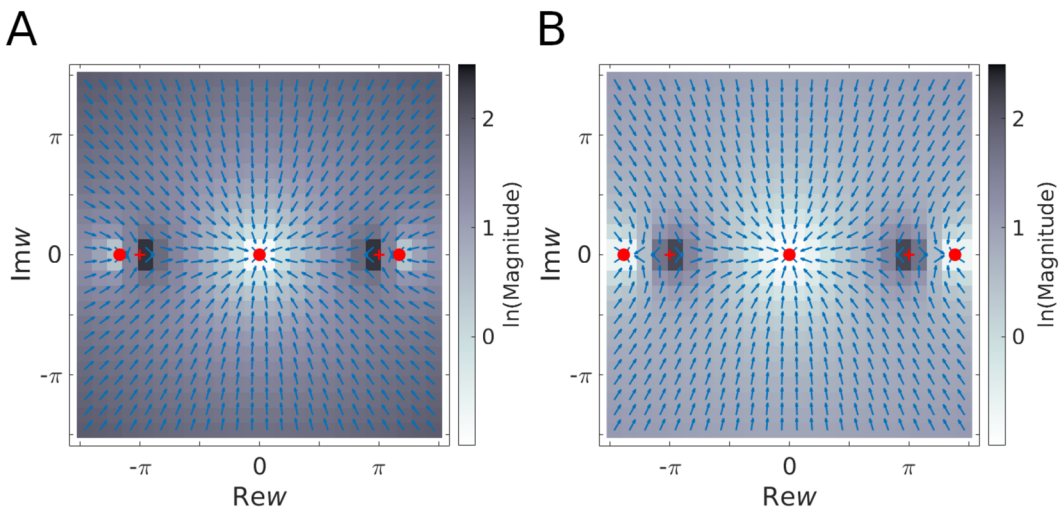

FIG. 4. Drift in the complex w plane of the repulsive toy model with $N = 1$, for (a) $\nu = 1$ and (b) $\nu = 3$. The arrows are normalized and indicate only the direction of the drift. The magnitude of the drift is indicated by the background color. Red symbols indicate attractive fixed points (filled circles) and poles (plus signs). The latter are repelling/attracting in the real/imaginary direction, respectively.

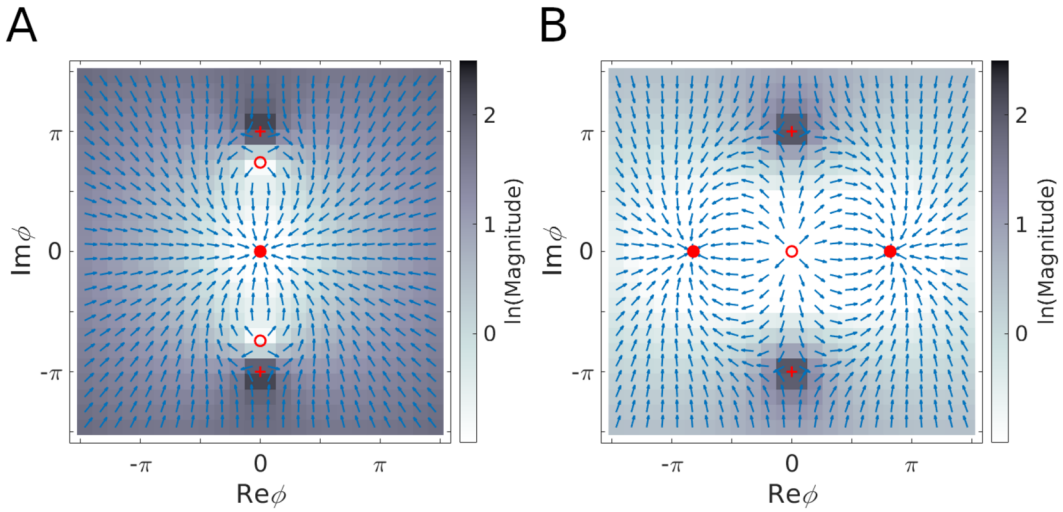

FIG. 5. Drift in the complex φ plane of the attractive toy model with $N = 1$, for (a) $\eta = 1$ and (b) $\eta = 3$. The arrows are normalized and indicate only the direction of the drift. The magnitude of the drift is indicated by the background color. Red symbols indicate attractive fixed points (filled circles), repulsive fixed points (open circles), and poles (plus signs). The latter are repelling/attracting in the real/imaginary direction, respectively.

3. Numerical results and implications

We have performed a set of simulations to probe the performance of the complex Langevin method for both the repulsive and the attractive toy model, using $N = 1$. Figure 6(a) shows the second moment of w for the repulsive model, as compared to the correct value $\langle w^2 \rangle = 2\nu - \nu^2$. Likewise, Fig. 6(b) shows the second moment of φ for the attractive model, as compared to the correct value $\langle \varphi^2 \rangle = 2\eta + \eta^2$. The simulations, indeed, confirm that the method significantly deteriorates in the repulsive case for $\nu \gtrsim 0.5$, while no notable deviation from the correct values is seen in the attractive case.

This toy model illustrates how the Langevin dynamics yields correct results for an attractive pair interaction, but deteriorates for a strong enough repulsive one, due to loss of ergodicity, in this case caused by poles on the real line. This behavior is qualitatively similar to what we observe in the larger model, where the complex Langevin dynamics fails to yield correct results when the repulsive part of the interaction is too large as compared to the attractive part.

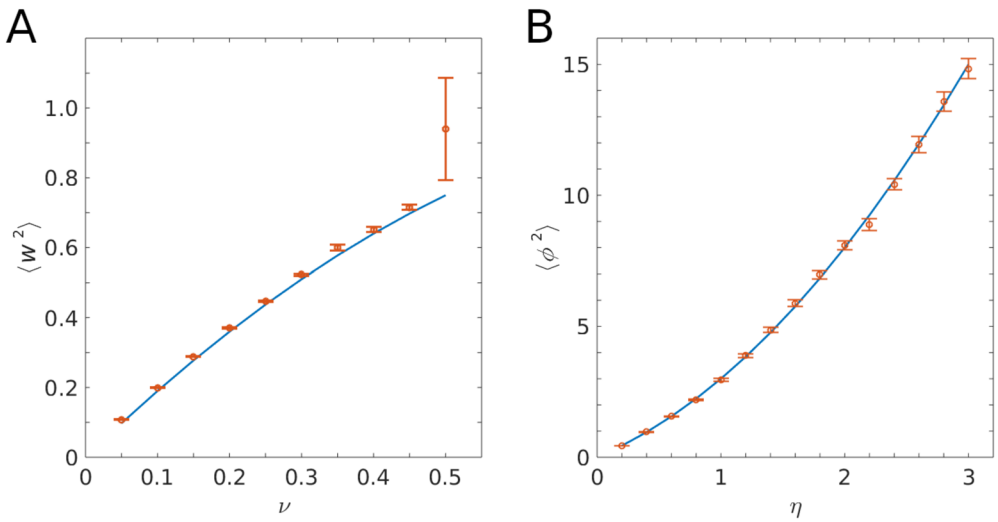

FIG. 6. Simulation data (red symbols) vs theoretical results (blue lines) for the toy model with $N = 1$. (a) Repulsive case: the second moment of w as a function of ν. Simulation data are well behaved for small ν, but deteriorate at $\nu \approx 0.5$. This erratic behavior will only become worse for higher ν values and is conjectured to be due to loss of ergodicity. (b) Attractive case: the second moment of φ as a function of η. Simulation data follow the theoretical curve over the whole range.

The similarity in behavior suggests that also for the larger model, the problems might be due to loss of ergodicity. Note that the zeros of Q form pole manifolds of H with a complex codimension one, corresponding a real codimension two. Normally, this should not jeopardize ergodicity. However, if there exists an attractor in the field space of codimension one or more – such as the real line in the repulsive toy model—ergodicity could be destroyed.

We speculate that the failure of the complex Langevin algorithm for strong repulsion might be due to a bifurcation to a situation with a codimension-one attractor, inside which the pole manifold may have a real codimension one. This would be enough to block trajectories and destroy ergodicity for the exact continuum version of the complex Langevin dynamics. In computer simulations with a finite time step, trajectories may actually jump over the pole blockage, but in a way leading to erroneous probabilities.

C. HP chains

We now return to the HP lattice model, with repulsion strength Λ. In the lattice gas (Sec. III A), we saw that $\Lambda = 2$ was too small to prevent condensed droplets from collapsing, whereas this problem was significantly alleviated when using $\Lambda = 5$. In this subsection, we present some simulation results for two ten-bead HP sequences, obtained with $\Lambda = 5$. We wish to explore how the FTS method performs when applied to chain systems at this Λ.

The two HP sequences considered are the alternating sequence $(HP)_5$, called A, and the block sequence H_5P_5, called B, which share the same composition. These two sequences have previously been studied using a coarse-grained continuous model,[40,41] where cluster formation was found to set in at a higher temperature for sequence B than for sequence A. However, the clusters formed by sequence B were micelle-like and, therefore, did not represent a bulk phase. By contrast, sequence A did phase separate.[40,41] These findings are consistent with the results obtained using a related model.[42,43]

Here, we consider systems consisting of 64 copies of either the A or the B sequence, on an elongated 36×12^2 grid. Figure 7 shows the β-dependence of the longitudinal bead density distribution, $\rho(z)$, using PBS data obtained with the Wang–Landau algorithm. Here, before averaging over snapshots, the distribution $\rho(z)$ in a given snapshot is shifted, in such a way that if a single droplet is present, then its center of mass ends up close to the center of the box (in the z direction). From Fig. 7, it can be seen that cluster formation, indeed, sets in at a lower β for sequence B than for sequence A, as in previous work.[40,41] We estimate that $\beta_t \approx 0.77 \pm 0.01$ for sequence B and $\beta_t \approx 1.31 \pm 0.03$ for sequence A.

TABLE I. Estimates of the energy U obtained with FTS and PBS for the ten-bead HP sequences A and B at $\beta = 0.5$ and $\beta = 1.0$, for $\Lambda = 5$ and 64 chains on a 36×12^2 lattice.

	Sequence A		Sequence B	
	$\beta = 1.0$	$\beta = 0.5$	$\beta = 1.0$	$\beta = 0.5$
FTS	−4558 ± 16	−2992 ± 15	−7858 ± 14	−6253 ± 12
PBS	− 102 ± 15	− 12 ± 9	− 549 ± 15	− 22 ± 15

We test the FTS method using $\beta = 0.5$ and $\beta = 1.0$. Table I compares the FTS data for the energy estimator U_{FT} [Eq. (17)] with the PBS data for the energy U. The β_t estimates above imply that a large cluster is present in only one of the four systems studied, namely, for sequence B at $\beta = 1.0$. Therefore, the energy U is much lower in this system than in the other three. In all four cases, we find that the FTS method severely underestimates U, which is not unexpected given that $\Lambda = 5$ (cf. Fig. 1). Now, it may be argued that the energy is a model-dependent quantity and, therefore, less interesting than basic structural properties, such as the presence or absence of large clusters.

In Fig. 8, we, therefore, also compare bead density profiles, $\rho(z)$, obtained using FTS and PBS, respectively, at the same two β values. As expected, the PBS profiles show that a large droplet is present for sequence B at $\beta = 1.0$, but not in any of the other three systems studied. In sharp contrast, the FTS data erroneously indicate that a large cluster is present in all four systems. Thus, in the systems studied, there is clear tendency for FTS sampling errors to cause a bias toward cluster formation. We also note that the maximal (averaged aligned) densities $\rho(z)$ from the FTS runs tend to be high, with

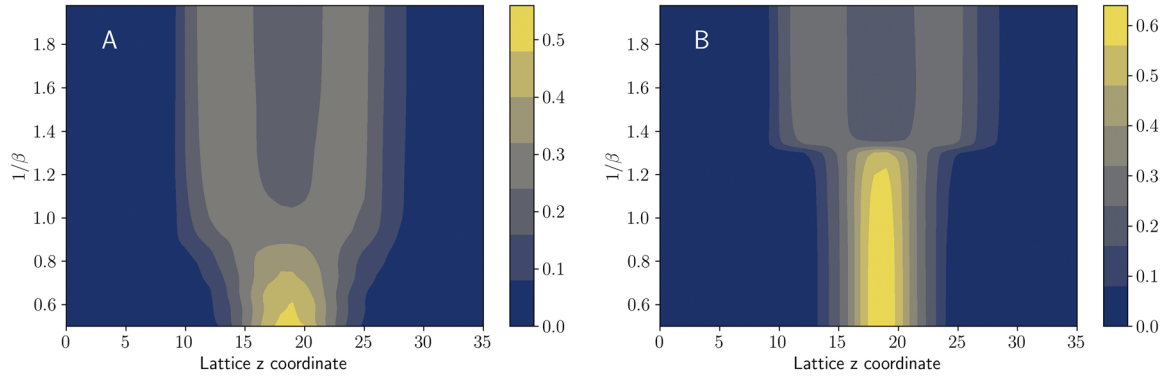

FIG. 7. Heat maps showing the temperature $(1/\beta)$-dependence of the bead density profile $\rho(z)$, for (a) sequence A and (b) sequence B. At a given β, $\rho(z)$ is either clearly unimodal, indicating the presence of a single dominant droplet, or weakly bimodal. In the latter case, the system tends to exhibit two main clusters. The data are from PBS simulations with the Wang–Landau algorithm.[35] The simulated systems consist of 64 chains on a 36×12^2 lattice, for $\Lambda = 5$.

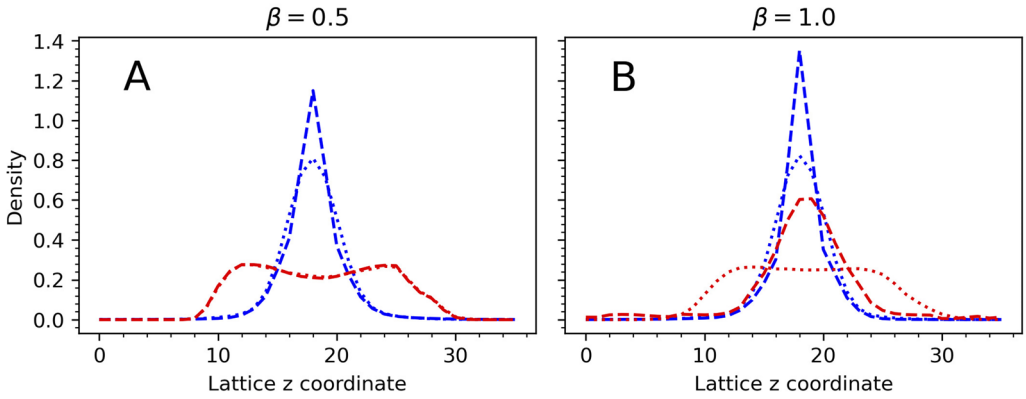

FIG. 8. Bead density profiles, $\rho(z)$, calculated using FTS (blue) and PBS (red) for sequences A (dotted) and B (dashed), for (a) $\beta = 0.5$ and (b) $\beta = 1.0$. The profiles are either clearly unimodal, indicating the presence of a single dominant droplet, or weakly bimodal. The latter systems tend to exhibit two main clusters. The FTS results for $\rho(z)$ are obtained using a field-theoretic estimator derived from Eq. (15). The simulated systems consist of 64 chains on a 36×12^2 lattice, for $\Lambda = 5$.

values exceeding unity for sequence B. A value of unity corresponds to one bead per site.

IV. DISCUSSION

FTS offers a new tool for investigating the mechanisms of biomolecular LLPS, with potential advantages over traditional PBS. The FTS approach has previously been used to investigate various systems where phase separation is driven by electrostatics.[21,25–29] In this paper, we have studied systems where phase separation is driven by short-range hydrophobic attraction. In preliminary work, we considered a continuous protein model similar to those in previous FTS studies,[21,25–29] but with Coulomb interaction replaced by an effective attraction between hydrophobic beads. For simplicity, we decided, however, to focus on a lattice-based HP protein model, with finite same-site repulsion and nearest-neighbor attraction between HH pairs. We showed that this model can be mapped onto a field theory with a simple structure by using an unconventional link-based form for the HH attraction. One advantage of choosing this lattice-based protein model is that FTS results can then be directly compared with PBS data, without having to extrapolate the FTS results to the continuum limit.

For a protein model to be amenable to standard FTS techniques, its excluded-volume repulsion has to be soft. However, if this repulsion is made too soft, one risks affecting the phase behavior.[14,21,28] A previous FTS study found that the excluded-volume strength affected the phase separation propensity in one-component polyampholyte systems,[21] in line with theoretical results.[44] Using both FTS and PBS, another study found that the excluded-volume strength affected demixing in two-component polyampholyte systems.[28]

In this paper, we have investigated the ability of the FTS method to accurately capture the thermodynamic behavior of the HP lattice protein model, with same-site repulsion strength Λ. We examined in some detail the special case of a lattice gas, consisting of (one-bead) H particles. We asked whether, at a given Λ, FTS remains accurate at sufficiently large β to permit the study of droplet condensation. Unfortunately, we find that this holds only for Λ values too small to prevent the formation of artificially compact droplets (Fig. 3). For small Λ, the condensation transition is sharp, with a strongly bimodal energy distribution (Fig. 2), which renders standard constant-temperature PBS inefficient. It is therefore interesting to note from a sampling perspective that FTS data describe the condensation transition quite well for $\Lambda = 0.5$ (Fig. 2).

To get an idea of the origin and generality of the FTS sampling problems observed at large Λ in the lattice gas, we introduced a minimal two-site toy model with either attractive or repulsive interaction, which can be solved analytically. We find that this model can be simulated using the complex Langevin method if the interaction is attractive, whereas sampling problems arise if the interaction is repulsive and strong. We, thus, observe the same trends as in our lattice gas simulations. This similarity hints that the FTS sampling problems might be of the same nature for the lattice gas as in the toy model, where they can be linked to the loss of ergodicity.

Finally, we also presented results from some simulations of HP chains. Here, we wanted to explore the size and nature of the FTS sampling errors in chain systems with significant same-site repulsion. We studied 64-chain systems for two ten-bead HP sequences and observed a clear tendency for FTS sampling errors to cause a bias toward droplet formation. In particular, using a temperature at which both systems should be free from large clusters, FTS data instead indicated the presence of a high-density droplet in both the cases.

In summary, from our investigations of the HP lattice and toy models and our preliminary studies of the above-mentioned off-lattice model, we believe that the observed sampling problems caused by too strong repulsion have some generality. At the same time, we have certainly not fully explored alternative forms for the Langevin dynamics[24] or alternative time integration schemes.[45] Furthermore, in the HP model, the gap is not huge between the two regions with acceptable FTS sampling errors and physically acceptable repulsion strength, respectively. To determine whether this

gap can be bridged by, for example, fine-tuning the FTS approach, further investigation is needed.

ACKNOWLEDGMENTS

This work was, in part, supported by the Swedish Research Council (Grant Nos. 621-2014-4522 and 621-2018-04976) and the Swedish strategic research program eSSENCE. Part of the simulations was performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at LUNARC, Lund University, Sweden.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, and A. A. Hyman, "Germline P granules are liquid droplets that localize by controlled dissolution/condensation," Science **324**, 1729–1732 (2009).

[2]S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, "Biomolecular condensates: Organizers of cellular biochemistry," Nat. Rev. Mol. Cell Biol. **18**, 285–298 (2017).

[3]T. J. Nott, E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, T. D. Craggs, D. P. Bazett-Jones, T. Pawson, J. D. Forman-Kay, and A. J. Baldwin, "Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles," Mol. Cell **57**, 936–947 (2015).

[4]A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, and J. P. Taylor, "Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization," Cell **163**, 123–133 (2015).

[5]K. A. Burke, A. M. Janke, C. L. Rhine, and N. L. Fawzi, "Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II," Mol. Cell **60**, 231–241 (2015).

[6]M. L. Huggins, "Solutions of long chain compounds," J. Chem. Phys. **9**, 440 (1941).

[7]P. J. Flory, "Thermodynamics of high polymer solutions," J. Chem. Phys. **10**, 51–61 (1942).

[8]J. T. G. Overbeek and M. J. Voorn, "Phase separation in polyelectrolyte solutions. Theory of complex coacervation," J. Cell Comp. Physiol. **49**, 7–26 (1957).

[9]J. Wittmer, A. Johner, and J. F. Joanny, "Random and alternating polyampholytes," Europhys. Lett. **24**, 263–268 (1993).

[10]Y.-H. Lin, J. D. Forman-Kay, and H. S. Chan, "Sequence-specific polyampholyte phase separation in membraneless organelles," Phys. Rev. Lett. **117**, 178101 (2016).

[11]G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, "Sequence determinants of protein phase behavior from a coarse-grained model," PLoS Comput. Biol. **14**, e1005941 (2018).

[12]G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, "Relation between single-molecule properties and phase behavior of intrinsically disordered proteins," Proc. Natl. Acad. Sci. U. S. A. **115**, 9929–9934 (2018).

[13]S. Das, A. Eisen, Y.-H. Lin, and H. S. Chan, "A lattice model of charge-pattern-dependent polyampholyte phase separation," J. Phys. Chem. B **122**, 5418–5431 (2018).

[14]S. Das, A. N. Amin, Y.-H. Lin, and H. S. Chan, "Coarse-grained residue-based models of disordered protein condensates: Utility and limitations of simple charge pattern parameters," Phys. Chem. Chem. Phys. **20**, 28558–28574 (2018).

[15]N. A. S. Robichaud, I. Saika-Voivod, and S. Wallin, "Phase behavior of blocky charge lattice polymers: Crystals, liquids, sheets, filaments, and clusters," Phys. Rev. E **100**, 052404 (2019).

[16]T. S. Harmon, A. S. Holehouse, M. K. Rosen, and R. V. Pappu, "Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins," eLife **6**, e30294 (2017).

[17]T. S. Harmon, A. S. Holehouse, and R. V. Pappu, "Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins," New J. Phys. **20**, 045002 (2018).

[18]S. Qin and H.-X. Zhou, "Fast method for computing chemical potentials and liquid–liquid phase equilibria of macromolecular solutions," J. Phys. Chem. B **120**, 8164–8174 (2016).

[19]G. H. Fredrickson, V. Ganesan, and F. Drolet, "Field-theoretic computer simulation methods for polymers and complex fluids," Macromolecules **35**, 16–39 (2002).

[20]K. T. Delaney and G. H. Fredrickson, "Recent developments in fully fluctuating field-theoretic simulations of polymer melts and solutions," J. Phys. Chem. B **120**, 7615–7634 (2016).

[21]J. McCarty, K. T. Delaney, S. P. O. Danielsen, G. H. Fredrickson, and J.-E. Shea, "Complete phase diagram for liquid–liquid phase separation of intrinsically disordered proteins," J. Phys. Chem. Lett. **10**, 1644–1652 (2019).

[22]G. Parisi, "On complex probabilities," Phys. Lett. B **131**, 393–395 (1983).

[23]J. R. Klauder, "A Langevin approach to fermion and quantum spin correlation functions," J. Phys. A: Math. Gen. **16**, L317–L319 (1983).

[24]B. Söderberg, "On the complex Langevin equation," Nucl. Phys. B **295**, 396–408 (1988).

[25]Y. Lin, J. McCarty, J. N. Rauch, K. T. Delaney, K. S. Kosik, G. H. Fredrickson, J. E. Shea, and S. Han, "Narrow equilibrium window for complex coacervation of tau and RNA under cellular conditions," eLife **8**, e42571 (2019).

[26]S. P. O. Danielsen, J. McCarty, J.-E. Shea, K. T. Delaney, and G. H. Fredrickson, "Molecular design of self-coacervation phenomena in block polyampholytes," Proc. Natl. Acad. Sci. U. S. A. **116**, 8224–8232 (2019).

[27]S. Najafi, Y. Lin, A. P. Longhini, X. Zhang, K. T. Delaney, K. S. Kosik, G. H. Fredrickson, J. E. Shea, and S. Han, "Liquid–liquid phase separation of Tau by self and complex coacervation," Protein Sci. **30**, 1393–1407 (2021).

[28]T. Pal, J. Wessén, S. Das, and H. S. Chan, "Subcompartmentalization of polyampholyte species in organelle-like condensates is promoted by charge-pattern mismatch and strong excluded-volume interaction," Phys. Rev. E **103**, 042406 (2021).

[29]J. Wessén, T. Pal, S. Das, Y.-H. Lin, and H. S. Chan, "A simple explicit-solvent model of polyampholyte phase behaviors and its ramifications for dielectric effects in biomolecular condensates," J. Phys. Chem. B **125**, 4337–4358 (2021).

[30]K. F. Lau and K. A. Dill, "A lattice statistical mechanics model of the conformational and sequence spaces of proteins," Macromolecules **22**, 3986–3997 (1989).

[31]G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, "Langevin simulations of lattice field theories," Phys. Rev. D **32**, 2736–2747 (1985).

[32]A. Irbäck, "Hybrid Monte Carlo simulation of polymer chains," J. Chem. Phys. **101**, 1661–1667 (1994).

[33]J. Ambjørn, M. Flensburg, and C. Peterson, "The complex Langevin equation and Monte Carlo simulations of actions with static charges," Nucl. Phys. B **275**, 375–397 (1986).

[34]E. Seiler, "Status of complex Langevin," EPJ Web Conf. **175**, 01019 (2018).

[35]F. Wang and D. P. Landau, "Efficient, multiple-range random walk algorithm to calculate density of states," Phys. Rev. Lett. **86**, 2050–2053 (2001).

[36]S. Æ. Jónsson, S. Mohanty, and A. Irbäck, "Accelerating atomic-level protein simulations by flat-histogram techniques," J. Chem. Phys. **135**, 125102 (2011).

[37]A. C. K. Farris, G. Shi, T. Wüst, and D. P. Landau, "The role of chain-stiffness in lattice protein models: A replica-exchange Wang-Landau study," J. Chem. Phys. **149**, 125101 (2018).
[38]R. H. Swendsen and J.-S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," Phys. Rev. Lett. **58**, 86–88 (1987).
[39]A. Irbäck, S. Æ. Jónsson, N. Linnemann, B. Linse, and S. Wallin, "Aggregate geometry in amyloid fibril nucleation," Phys. Rev. Lett. **110**, 058101 (2013).
[40]D. Nilsson and A. Irbäck, "Finite-size scaling analysis of protein droplet formation," Phys. Rev. E **101**, 022413 (2020).
[41]D. Nilsson and A. Irbäck, "Finite-size shifts in simulated protein droplet phase diagrams," J. Chem. Phys. **154**, 235101 (2021).
[42]A. Statt, H. Casademunt, C. P. Brangwynne, and A. Z. Panagiotopoulos, "Model for disordered proteins with strongly sequence-dependent liquid phase behavior," J. Chem. Phys. **152**, 075101 (2020).
[43]U. Rana, C. P. Brangwynne, and A. Z. Panagiotopoulos, "Phase separation vs aggregation behavior for model disordered proteins," J. Chem. Phys. **155**, 125101 (2021).
[44]S. L. Perry and C. E. Sing, "PRISM-based theory of complex coacervation: Excluded volume versus chain correlation," Macromolecules **48**, 5040–5053 (2015).
[45]G. Aarts, F. A. James, E. Seiler, and I.-O. Stamatescu, "Adaptive stepsize and instabilities in complex Langevin dynamics," Phys. Lett. B **687**, 154–159 (2010).