Memory unlocks the future of biomolecular dynamics: Transformative tools to uncover physical insights accurately and efficiently

Anthony J. Dominic III,[1] Siqin Cao,[2] Andrés Montoya-Castillo,[1, a)] and Xuhui Huang[2, b)]
[1)] *Department of Chemistry, University of Colorado Boulder, Boulder, CO 80309, USA*
[2)] *Department of Chemistry, Theoretical Chemistry Institute, University of Wisconsin-Madison, Madison, WI 53706, USA*

(Dated: 28 January 2023)

Conformational changes underpin function and encode complex biomolecular mechanisms. Gaining atomic-level detail of how such changes occur has the potential to reveal these mechanisms and is of critical importance in identifying drug targets, facilitating rational drug design, and enabling bioengineering applications. While the past two decades have brought Markov State Model techniques to the point where practitioners can regularly use them to glimpse the long-time dynamics of slow conformations in complex systems, many systems are still beyond their reach. In this perspective, we discuss how memory can reduce the computational cost to predict the long-time dynamics in these complex systems by orders of magnitude and with greater accuracy and resolution than state-of-the-art Markov State Models. We illustrate how memory lies at the heart of successful and promising techniques, ranging from the Fokker-Planck and generalized Langevin equations to deep learning recurrent neural networks and generalized master equations. We delineate how these techniques work, identify insights that they can offer in biomolecular systems, and discuss their advantages and disadvantages in practical settings. We show how generalized master equations can enable the investigation of, for example, the gate-opening process in RNA polymerase II and demonstrate how our recent advances tame the deleterious influence of statistical underconvergence of the molecular dynamics simulations used to parameterize these techniques. This represents a significant leap forward that will enable our memory-based techniques to interrogate systems that are currently beyond the reach of even the best Markov State Models. We conclude by discussing some current challenges and future prospects for how exploiting memory will open the door to many exciting opportunities.

I. INTRODUCTION

Biological macromolecules often need to dynamically change their shapes or conformations to perform their functions. Conformational dynamics thus play an important role in many biological processes such as protein mis-folding and aggregation, protein-ligand recognition, and numerous other functional conformational changes. Investigating functional conformational changes is thus essential for elucidating molecular mechanisms of numerous fundamental biological processes[1,2] and facilitating rational drug design.[3–5]

Molecular dynamics (MD) simulations provide a powerful tool to study biomolecular dynamics in complement to experimental techniques. However, the timescales for biologically relevant processes (often milliseconds or longer) of biomolecular complexes remain challenging for all-atom MD simulations (microseconds or shorter for complexes like RNA polymerases) to routinely reach. Markov State Models (MSMs) have become a popular approach to bridge this timescale gap by modeling long timescale dynamics based on many short MD simulations.[6–16] MSMs have been widely applied to study protein folding,[11,17–20] protein-ligand binding,[21–24] and functional conformational changes of biomolecules.[12,25–34]

In MSMs, conformational changes are modeled as a series of Markovian, or memoryless, transitions among metastable conformational states (or free energy minima) at discrete time intervals: i.e., lag times. The key requirement for the validity and success of an MSM is that the simulation time used to parameterize these models, or lag time, is sufficiently long to allow transitions among states to become memoryless. The memory of these transitions is mainly determined by dynamic relaxation within each state. In practice, this is challenging as the lag time is bound by the length of short MD simulations available to estimate the frequency of these transitions. To achieve the Markovian property, one often needs to construct MSMs containing a large number of states, so that each state is sufficiently small and has relatively fast relaxation dynamics to allow affordable lag times. For example, an MSM containing 2,000 states (with a lag time of 12 ns) proved necessary to model the millisecond folding of the NTL9 peptide.[35] Furthermore, our work on the RNA polymerase II (RNAP) backtracking also showed that MSMs consisting of 800 states are needed to reach Markovianity.[36] MSMs containing thousands of states are useful to make quantitative predictions to be tested against experiments, but often hinder biological insight.

To address the above challenge, we recently developed a method to go beyond the Markovian models (namely

[a)] Electronic mail: Andres.MontoyaCastillo@colorado.edu
[b)] Electronic mail: xhuang@chem.wisc.edu

quasi-MSM or qMSM), where we explicitly calculate the memory kernel and propagate protein dynamics using a discretized Generalized Master Equation (GME).[37] Protein dynamics span a wide range of timescales. There often exist separations of these timescales: e.g., transitions between metastable states (e.g., protein backbone folding) are much slower than relaxations within a state (e.g., side-chain rotations). To study the long-timescale dynamic modes, we can apply the projection operator scheme to project the kinetics onto slow degrees of freedom (i.e., inter-state transitions). In our previous work, we followed Hummer and Szabo to define the projector operator,[38] in which MD conformations are assigned to metastable states via a state indicator function. When there exist separations of timescales, dynamics reach equilibrium quickly within each state, and we can build on it to obtain a GME containing fast-decaying memory kernels.[7,37] We have demonstrated that our GME method greatly improves upon MSMs by accurately predicting long-timescale dynamics while requiring significantly shorter MD simulations for the WW domain folding[39,40] and the gate opening motion of a bacterial RNA Polymerase.[36]

In this perspective, we explore the salient problems that beset MSMs and lay out a vision for how GMEs can be used to radically expand the robustness and applicability of MSMs so that we can tackle complex conformational changes underlying biomolecular functions.

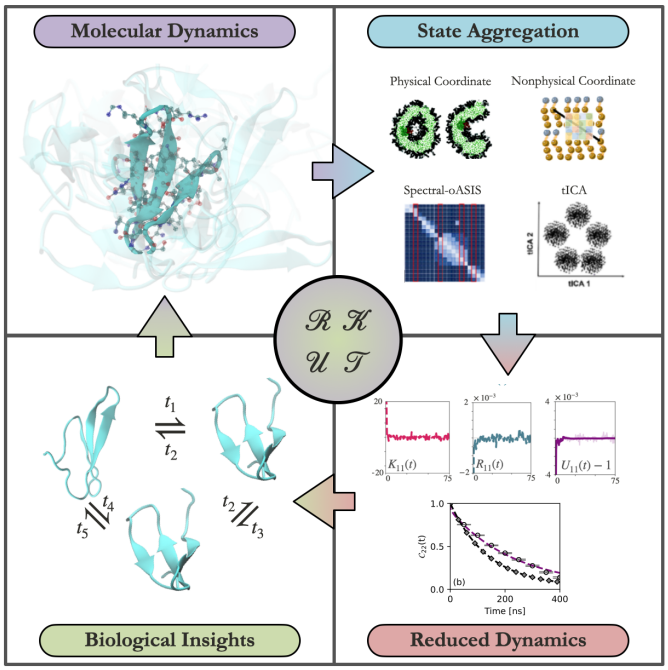

FIG. 1. Schematic figure of the non-Makovian dynamic models incorporating memory kernels for biomolecular dynamics.

II. GENERALIZED MASTER EQUATIONS: THE ADVANTAGES OF MEMORY

The core idea underlying MSMs is that of reducing the dimension of the chemical dynamics problem from one that explicitly tracks the concerted motions of thousands, even millions, of atoms to one that predicts the slow interconversion between a handful of macroscopic states, $\{A_j\}$, that offer a simpler interpretable framework for a complex chemical problem. To construct this simple, intuitive, and low dimensional framework, however, one first needs to *identify* the handful of slow coordinates — a formidable problem in general. While work in the last few decades has produced a robust set of tools to exploit collections of short-time all-atom MD simulations of complex biomolecular systems to identify these elusive and important slow coordinates, this problem of identifying slow degrees of freedom remains an open question of fundamental importance to many fields.[11,41–49] Once these slow coordinates are (perfectly or approximately) identified, one can employ a variety of tools to estimate the transition probabilities[11,42] and construct the simple equation of motion — which takes the form of a simple chemical kinetics problem — for the interconversion of these slow macrostates at the heart of the MSM:[10,42]

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{C}(t) = \mathbf{M}\mathbf{C}(t), \tag{1}$$

where $[\mathbf{C}(t)]_{n,m}$ is the transition probability matrix (TPM) which corresponds to the time-dependent conditional probability of finding the protein in the n^{th} configuration given that the protein started in the m^{th} configuration. Clearly, from this TPM, one can obtain the time-dependence of the conditional probabilities for *any* initial preparation of the protein evolving under thermal equilibrium. As in a chemical kinetics problem, the matrix $\mathbf{M}$, contains a collection of the rates, $\mathbf{M}_{j,k}$, that determine how fast configuration j turns into configuration k. *This framework can provide an accurate approximation to the dynamics of the macrostates $\{A_j\}$ at sufficiently long times, but degrades at short times.* In fact, the fundamental limit of temporal resolution for an MSM is dictated by its *lag time*, the minimum time step required for the dynamics to be properly described as a simple kinetic scheme. In other words, the lag time corresponds to the amount of simulation time required for one to obtain a converged estimate for the rates, $\mathbf{M}_{j,k}$. Hence, this quantity is also directly proportional to the computational cost associated with building an MSM as it determines the minimum timescale of the MD simulations required to identify and construct the set of slow coordinates $\{A_j\}$ and their kinetic description.

However, dimensionality reduction comes at a cost: the equations of motion that dictate the true evolution of $\{A_j\}$ generally depend not just on the state of these configurations at the previous snapshot in time but also on the *history* of how these configurations have interconverted. Using the celebrated projection operator

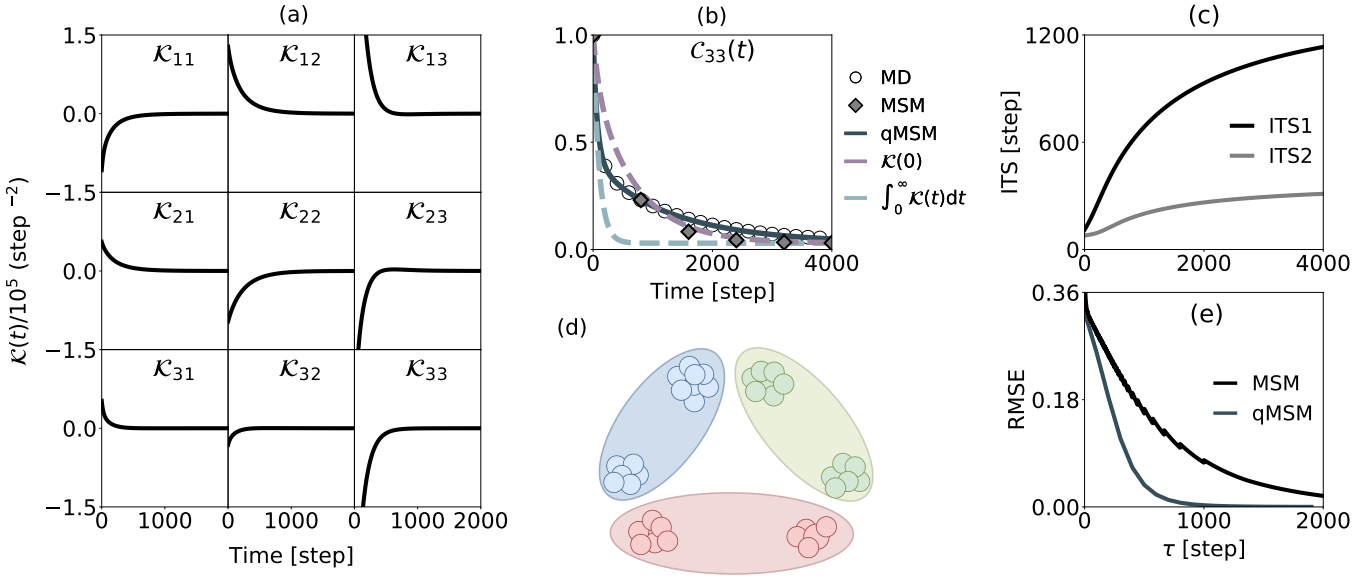

FIG. 2. Demonstrating the strict advantage of memory-based approaches in cases where a sub-optimal projector is employed. (a) The full (3×3) memory kernel shown as a function of time. (b) Chapman-Kolmogorov test: state 3 residence probabilities obtained from the original MD of the 6 node model compared to those obtained from the MSM (τ_L = 800 steps), qMSM (τ_K = 800 steps), short-memory approximation, and the Markovian approximation. (c) The implied timescales of the full TPM as a function of lag time τ_L. (d) Mapping the six node model to a three node model. (e) Root mean square error curves for the MSM and qMSM calculated with respect to the benchmark MD data.

techniques,[50] which take as input the slow macrostates identified above, one can derive an equation of motion, called the GME, that *exactly* encodes the dynamics for the probability of populating each of the macrostates,[51–53]

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{C}(t) = \mathbf{\Omega}\mathbf{C}(t) - \int_0^t \mathrm{d}\tau\ \boldsymbol{\mathcal{K}}(t-\tau)\mathbf{C}(\tau). \qquad (2)$$

Here $\mathbf{\Omega} = \dot{\mathbf{C}}(0)$ is the memory-less component of the evolution of the slow macrostates $\{A_j\}$ and $\boldsymbol{\mathcal{K}}(t)$ is known as the memory kernel. This memory kernel plays the role of a time-dependent friction constant and encodes the non-instantaneous response of the atomic motions in the full biomolecular problem treated at a microscopic level.

GMEs provide a powerful theoretical infrastructure with impact in various fields and have served as a starting point for a number of approximate treatments to dynamical problems ranging from the classical simulation of liquids[54–58] and glasses,[59–61] to structural relaxation in polymers,[62,63] coarse-graining of biological simulations and polymers,[64,65] and quantum dynamics of charge and energy transfer and transport in condensed phase systems.[66–68] In addition, the GME has served as the basis for leveraging single-molecule force spectroscopy data to characterize RNA unfolding processes by parameterizing the memory kernel using maximum entropy methods.[69,70] Inspired by the recent work in quantum dynamics that has exploited the self-consistent expansion of the memory kernel,[71–79] we have recently introduced quasi-Markov State Models (qMSMs), which describe the few-macrostate dynamics with a GME whose memory kernel is calculated with data that would otherwise have been used to create an MSM.

In practice, identifying the slow coordinates in a complex many-body system is a formidable challenge.[43,47,48,80–82] In turn, projectors constructed for complex systems generally do not exactly identify the optimal state delineation that leads to the greatest timescale separation between intra- and interstate transitions.[9,42] The difficulty with optimal state selection increases the amount of simulation time (τ_L) required to parameterize a valid MSM, which can render their construction prohibitively expensive. In contrast, by explicitly including memory, the GME can be parameterized with significantly less data, thereby offering a more efficient way to capture the reduced dynamics, even when using a suboptimal projector. We illustrate this point in Fig. 2 by comparing the ability of the MSM and qMSM approaches to recapitulate the reference dynamics (MD) in a 3-state model that arises from the downfolding of a 6-state kinetic model [Fig. 2(d)] that we introduced in Ref. 37. To emphasize the differences between the performance of these two approaches, we employ a suboptimal projector that lumps together states with low (rather than high) transition rates.

Central to the construction of the qMSM is the memory kernel, $\boldsymbol{\mathcal{K}}(t)$. In Fig. 2(a), we show the full memory kernel in the reduced space — a 3×3 time-dependent matrix that encodes the historically weighted transitions in the reduced model. We define the timescale required for $\boldsymbol{\mathcal{K}}$ to decay to zero as the memory kernel lifetime

(τ_K). This enables the replacement of the upper integral limit in Eq. (2) with $\min\{t, \tau_K\}$, bounding the computational expense of evolving the reduced dynamics for arbitrarily long times. In this bad lumping model, $\tau_K = 800$ steps [Fig. 2(a)]. Indeed, Fig. 2(b) shows that the qMSM, when parameterized with $\tau_K = 800$ ns, accurately captures the reference dynamics MD data. In contrast, the MSM parameterized with the same amount of data ($\tau_L = 800$ steps) deviates from the reference dynamics and results in overly fast equilibration of the reduced dynamics.

In practice, one identifies the Markovian lag time (τ_L) for the MSM construction with the onset of a plateau in the implied timescales (ITS) as a function of trial lag time (τ), given by

$$\mathrm{ITS}_i(\tau) = -\frac{\tau}{\ln \lambda_i(\tau)}, \tag{3}$$

where the $\{\lambda_i(\tau)\}$ are the eigenvalues of $\mathbf{C}(t)$.[10] The ITS curves describe the equilibration timescales for the intrastate degrees of freedom within the selected macrostates. In addition, the TPM always contains an eigenvalue equal to unity for all time, corresponding to the stationary distribution of the model.[83] Figure 2(c) shows the non-unity ITS curves. The fact that the ITS curves have not plateaued by the trial lag time of $\tau_L = 800$ steps [Fig. 2(c)] explains why the MSM parameterized with this lag time fails to capture the exact dynamics [Fig. 2(b)]: the macrostates have not yet achieved intrastate equilibration. Importantly, these ITS curves do not plateau at any time within the interval $[0, 4000]$ steps, implying that a valid MSM cannot be constructed without additional computational cost (and lower temporal resolution) compared to the qMSM which only requires $\tau_K = 800$ steps to accurately capture the dynamics.

The GME [Eq. (2)] also provides a measure of the appropriateness of the Markovian approximation in MSMs. In the limit where the timescales associated with the interconversion between the chosen handful of configurations $\{A_j\}$ is slow in comparison to all other dynamics in the system (e.g., fast intramolecular vibrations, small fluctuations of the beta sheets, etc.), then one may employ the 'short-memory' approximation where the memory kernel is replaced by a delta function in time, $\boldsymbol{\mathcal{K}}(t) = \bar{\boldsymbol{\mathcal{K}}}\delta(t)$, allowing one to identify the rate matrix, $\mathbf{M} = \boldsymbol{\Omega} - \boldsymbol{\mathcal{K}}(0)$ with the sum of the memory-free component of the GME evolution $\boldsymbol{\Omega}$ and the magnitude of the instantaneously decaying memory kernel, $\boldsymbol{\mathcal{K}}$. Since the early work of Zwanzig,[84] this has served as the explicit connection between MSMs and the more general GMEs. Figure 2(b) illustrates that the dynamics that arise from making the 'short-memory' approximation can underestimate the equilibration rate. Alternatively, one may invoke the 'Markovian' approximation, which uses the integrated value of the memory kernel to construct the rate matrix, $\mathbf{M} = \boldsymbol{\Omega} - \int_0^\infty \mathrm{d}t\, \boldsymbol{\mathcal{K}}(t)$.[85] Here, whether $[\boldsymbol{\mathcal{K}}(t)]_{ij}$ approaches zero from above or below the x-axis is fully determined by the sign of $[\boldsymbol{\mathcal{K}}(0)]_{ij}$ and thus $|[\boldsymbol{\mathcal{K}}(0)]_{ij}| < \left|\int_0^\infty \mathrm{d}t\, [\boldsymbol{\mathcal{K}}(t)]_{ij}\right|$. This causes the resulting dynamics to relax overly fast and overestimate the equilibration rate of the reduced dynamics [Fig. 2(b)]. In the limit of extreme timescale separation, i.e., where the memory kernel approximates a δ-function, both approximations lead to the same result, but differ when the δ-function approximation is inappropriate.

Whether using the MSM or qMSM, the resulting dynamics require choosing the smallest appropriate value for τ_L or τ_K, respectively. We have recently employed the root mean square error (RMSE) to identify these parameters.[36,37,86,87] Specifically, we define the RMSE between the reference (MD) dynamics $\mathbf{E}(t)$ and the predicted dynamics $\mathbf{P}(t;\tau_x)$ to be

$$\mathrm{RMSE}(\mathbf{E},\mathbf{P}) = \sqrt{\sum_{t=1}^{t_{\max}} \sum_{ij}^{N} \frac{[\mathbf{E}(t) - \mathbf{P}(t;\tau_x)]_{ij}^2}{t_{\max} N^2}}, \tag{4}$$

where $\tau_x \in \{\tau_L, \tau_K\}$, N is the dimension of the reduced subspace, and $t_{\max}$ is the simulation time of the reference dynamics. Figure 2(e) illustrates the RMSE curves for both the MSM and qMSM. The RMSE curve for the qMSM decays faster than that of the MSM as a function of increasing τ. To decide an appropriate value of τ_L and τ_K, one only needs to define an acceptable threshold for the error. To fall below a threshold of 5 % error, one can choose $\tau_K = 800$ steps while $\tau_L = 1200$ steps — a factor of 1.5 bigger, even in this highly simplified model. Thus, the qMSM always offers an improvement over the MSM, especially when using a suboptimal projector, regardless of the amount of data used to parameterize the model.

Importantly, we have found that the lifetime of the memory kernel, τ_K, is shorter than the lag time of the analogous MSM required to describe the same few-state dynamics, suggesting that the GME provides a more economical way to construct and investigate long-time biomolecular dynamics. Moreover, unlike the MSM, whose temporal resolution is limited by the lag time, the GME provides full temporal resolution over the period containing transient dynamics and which is of interest in experiments such as two-dimensional spectroscopies that aim to elucidate the complex dynamics of these systems. We have, for instance, demonstrated that the qMSM can reduce the simulation time required to construct the conditional probability dynamics by up to an order of magnitude compared the state-of-the-art MSMs in systems like alanine dipeptide and the FiP35 WW domain.[37] In a subsequent study, when applied to study the gate opening motion of RNAP, our 4-state qMSM model (built from MD simulation segments amounting to tens of nanoseconds) significantly outperforms MSMs by predicting the correct time evolution of state residence probabilities and yielding converged mean first passage times at several millisecond timescales.[36]

III. ALTERNATIVE TREATMENTS OF MEMORY

Here, we discuss approaches that have emerged to ameliorate the difficulties associated with the memoryless approximation in MSMs. We discuss these techniques through the lens of the GME, which provides an exact treatment of memory in reduced dimensions.

A. Transfer Tensor Method

The most directly analogous method to the non-Markovian qMSM presented in Sec. II is the Transfer Tensor Method (TTM).[88] The TTM, which was recently developed and applied to problems in quantum relaxation,[89,90] is simply a discretization of the integrated form of the GME in Eq. (2), enabling the construction of the TPM at timestep n by summing over the weighted history of its evolution,

$$\mathbf{C}(n) = \sum_{k=0}^{n-1} \mathbf{T}(n-k)\mathbf{C}(k), \tag{5}$$

where $\mathbf{T}$ is the transfer tensor. One can obtain an expression for $\mathbf{T}$ in terms of the memory kernel and memory-less component of the evolution by discretizing the integrated form of Eq. (2),

$$\mathbf{T}(n,k) \approx (1 - \mathbf{\Omega} dt)\delta_{k,n-1} + \boldsymbol{\mathcal{K}}(t_n - t_k)dt^2, \tag{6}$$

as shown in Ref. 91. The TTM does not require taking time derivatives of the reference data or their subsequent integration, while our original implementation of the qMSM[37] requires the first time-derivative and the subsequent integration of the GME. However, the TTM has not been implemented in the context of biomolecular dynamics, so its potential advantages remain to be established in this context.

B. Fokker-Planck & Generalized Langevin equations

Returning to the continuous-time treatment, one can employ projection operator techniques to obtain an equation of motion for the *probability density of a continuous rather than a discrete variable*. In such cases, one obtains the celebrated Fokker-Planck equation.[92] One can then perform an ensemble average against the Fokker-Planck equation to obtain an equation of motion for the *expectation value of a continuous variable*, such as the average position of a heavy particle immersed in a fluid of light solvent, thus obtaining the generalized Langevin equation (GLE).[92,93] GLEs have also recently garnered much attention as a useful tool to address the non-Markovian nature of biomolecular configurational dynamics. As the continuous variable version of GMEs, GLEs can also be derived from the projection operator technique and describe the dynamics of a generalized coordinate, q. In recent work, this coordinate has taken the form of a curved conformational coordinate for hexapeptide neurotensin,[94] dihedral angle-based principal components in alanine peptides,[95] and α-helical conformation markers (i.e., summed distances between the H-bond donor nitrogen of residue n and the acceptor oxygen of residue $n+4$) in FiP35 WW domain.[96,97] For simplicity, in one dimension, the GLE takes the form

$$m\ddot{q}(t) = -\nabla U[q(t)] - \int_0^t \mathrm{d}\tau\, \Gamma(t-\tau)\dot{q}(\tau) + F_R(t) \tag{7}$$

where $U[q(t)]$ is the potential of mean force and $\Gamma(t)$ is the memory kernel. These terms are analogous to analogous to the memory-less component of the subspace evolution $\mathbf{\Omega}$ and the memory kernel, $\boldsymbol{\mathcal{K}}(t)$, in Eq. (2). The only term that differs is the random force, $F_R(t)$, which is related to the memory kernel via the fluctuation-dissipation theorem.[93] The GME analog of the random force term in the GLE is the inhomogeneous term, which disappears when one is interested in the TPM — or the time-dependent conditional probability of transitioning from one state to another — for a system at thermal equilibrium.[68,98] Indeed, recent work demonstrating the applicability of GLEs to protein folding has focused on demonstrating how to construct the potential of mean force, memory kernel, and random force terms.[94,96,97,99,100] However, the protocols for finding these quantities are still being developed, can be complex, and ultimately lead to a stochastic differential equation. In contrast, in our GME-based approaches, we provide a simple and theoretically transparent protocol to exactly parameterize the memory kernel based on reference MD data, which offers a deterministic rather than a stochastic equation of motion for the TPM. Yet, the methods we and others have developed to obtain the memory kernel and random force (or inhomogeneous term) contributions[101–103] are fully compatible with the GLE and can be directly applied in such contexts. Excitingly, current work has only recently set the stage to determine the distinct advantages that the discrete GME versus the continuous GLE offer for the configurational dynamics of biomolecules.

C. Hidden Markov Models

Although MSMs have been successfully applied to systems with clean timescale separations,[9] the construction of low-dimensional (macrostate) MSMs becomes formidably challenging in large biological systems, where such separations may not be clear. Additionally, the quality of an MSM depends sensitively on the state selection algorithm employed. Thus a suboptimal state aggregation can qualitatively alter the mechanisms that MSMs reveal.[11,42,104–107] To remedy this, one can alternatively construct a macrostate MSM by employing Hidden Markov Model (HMM) techniques,[108] where one as-

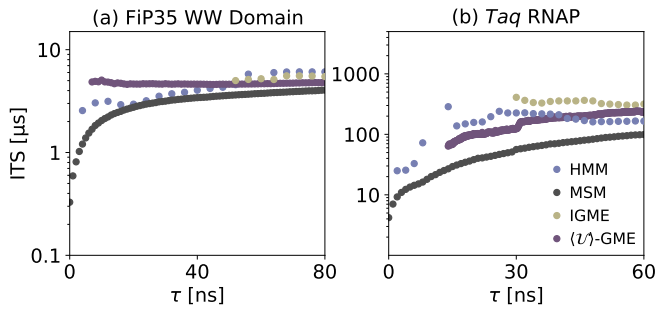

FIG. 3. Comparing the ability of the HMM, MSM, IGME, and $\langle\mathcal{U}\rangle$-GME to accurately and efficiently capture the slowest implied timescale (ITS) for (a) FiP35 WW Domain and (b) RNAP. The 4-state HMMs for FiP35 and RNAP were constructed from 400-state and 20-state MSMs, respectively. To obtain MSM, IGME, and $\langle\mathcal{U}\rangle$-GME results, we utilized the 4-state decompositions as described in previous work.

sumes that a time series of physical observables is encoded into a *hidden* TPM corresponding to an unknown collection of metastable regions in the system's configurational space. One can use the dynamics of a many-state (microstate) TPM as the input to the HMM, as well as the number of states one would like to consider in the few-state (macrostate) model to be obtained from the HMM. This means that one must have a preliminary discretization of configuration space to construct the macrostate TPM. One can then employ sophisticated maximum likelihood algorithms[108–110] to obtain the most probable composition of the macrostate model that optimizes for Markovian evolution. Importantly, this approach allows for the composition of the macrostates to arise from a "fuzzy" or non-orthogonal distribution of the microstates in the many-state MSM. This removes the disadvantage of having to compose the macrostates from potentially suboptimal microstates. HMMs have opened the door to probing biomolecular systems that were once beyond the reach of the original MSM methods, such as the analysis of Ubiquitin and the c-Src Tyrosine Kinase.[111] Additionally, the HMM is compatible with single molecule Förster resonance energy transfer (sm-FRET) experiments,[112] where the HMM has enabled the analysis of sm-FRET data by giving access to the conformational dynamics that lead to the folding of the RNA ribozyme Diels-Alderase.[112]

HMMs describe Markovian dynamics but contain "fuzzy" state boundaries, while the GME models contain crisp state boundaries but consider non-Markovian dynamics. Nevertheless, both approaches can effectively model biomolecular dynamics using models containing only a few macrostates built from MD simulations that are significantly shorter than those required to build MSMs. We have further compared the performance of HMM with IGME and $\langle\mathcal{U}\rangle$-GME methods (see Secs. IV A and IV B, respectively) in predicting the slowest implied timescale [see Fig. 3]. For the WW domain, we use PyEMMA86 to build HMMs with 4 hidden states to approximate the dynamics of 800-microstate models. For this thoroughly sampled system, the HMM and IGME display comparable performance, reaching the expected implied timescale around 50 ns (see Fig. 3(a)), while the $\langle\mathcal{U}\rangle$-GME reaches the expected implied timescale within the first 20 ns. However, for the large RNAP with less sufficient sampling (306 individual 200 ns trajectories), we show that both the IGME and $\langle\mathcal{U}\rangle$-GME outperform the HMM and yield a longer slowest implied timescale when approaching the variational bound (see Fig.3(b)). We anticipate that both the IGME and $\langle\mathcal{U}\rangle$-GME are numerically more robust than HMM. This is because the current implementation of HMM maximizes the likelihood function defined by the entire input trajectories, while the IGME is fitted to several TPMs at lag times longer than τ_k and the $\langle\mathcal{U}\rangle$-GME employs a simple averaging procedure that tames the statistical noise of underconverged TPMs. Additionally, the HMM framework is beset by (i) ambiguity in selecting the likelihood function and (ii) parameterization with an adequate number of states.

D. Other models considering a fraction of memory

Several methods have been developed to partially account for memory effects. For example, the history-augmented MSM (haMSM)[113] yields a history-dependent non-Markovian model for the steady-state dynamics of weighted ensemble simulations. The haMSM adopts a transition rate matrix with expanded dimensionality to compute the mean first passage time (MFPT) between two sets of states, where the expanded dimensionality incorporates the history of the dynamics. The haMSM easily yields the flux and MFPT and can be built using reweighted data from the weighted ensemble simulations. It can also be built with a smaller number of microbins, which can accelerate the estimation of steady-state dynamics. However, the haMSM only considers one step in the memory and can therefore be inaccurate if memory decays slowly and multiple steps of memory need to be considered.

Another method leverages Bayesian theory to construct high-order reversible Markov models[114] that account for non-Markovian dynamics. This approach can rigorously and accurately consider r steps of history with an r-order Markov chain as it explicitly computes all potential historical paths. However, the number of paths grows exponentially with the number of steps, which limits the maximum number of memory steps one can afford to consider. Yet, not all historical paths need to be considered. According to the Mori-Nakajima-Zwanzig equation, the history-dependent dynamics is a linear combination of dynamics in the same space, suggesting that most of the expanded dimensionality in the high-order Markov model is not necessary. In short, the Mori-Nakajima-Zwanzig equation, which leads to the qMSM, can largely

reduce the complexity of incorporating history and enable its exact inclusion.

E. Deep learning recurrent neural network models

Notably, a recently developed method introduces an algorithm based on the long short-term memory (LSTM) model to consider the memory functions of protein conformational dynamics.[115] This approach exploits a recurrent neural network architecture that can retain the memory of the past states in a temporal sequence via gating nodes that capture lags between long-timescale events. The success of this deep learning approach hinges on the connection of the loss function with the path entropy, which enables the LSTM method to accurately predict equilibrium distributions and kinetics for alanine dipeptide and experimental single-molecular FRET data. More recently, it was shown that subsampling the LSTM-predicted paths subject to physical constraints based on the Maximum Caliber principle and re-training the LSTM model on this subset of paths could lead to improved quality in obtaining thermodynamic and kinetic properties of biomolecular systems.[116]

As the recurrent neural network approach was originally developed for one-dimensional natural language processing, we expect that this LSTM approach alone may perform optimally on one-dimensional data.[117] Nevertheless, the LSTM architecture can be incorporated into a larger framework to perform complex multidimensional tasks. For example, LSTM lies at the core of AlphaStar,[118] which combines with other network architectures, such as transformers,[119] to processes complex inputs. Unlike the sequential LSTM model that only considers the direct connection between two adjacent MD conformations, transformer-based models adopt the self-attention mechanism that directly measures the connections between pairs of MD conformations at any two distinct time points. As a result, the transformer-based approach can capture the long-term and complex information within the time series data in a more effective way than LSTM. Thus, this framework can be expected to have great potential in handling, upon its generalization, multidimensional MD trajectories of functional conformational changes in the future.

IV. RECENT WORK & FUTURE PROMISES

While the qMSM has already demonstrated the ability to significantly reduce the amount of MD data required to predict the long-time conformational dynamics of systems ranging from alanine dipeptide and the FiP35 WW domain,[37] to the human argonaute complex[86] and RNA polymerase,[36] construction of the memory kernel can be challenging,[87,120] its truncation can be susceptible to statistical noise associated with limited simulation data, and its interpretation eludes the simple states-and-rates framework that renders MSMs conceptually transparent and straightforward. Below we introduce two recently developed methods to address these problems which demonstrate great potential in leveraging memory to capture the long-timescale biomolecular dynamics.

A. Recent Work: Integrative generalized master equation: IGME

Our initial GME-based approach, the qMSM,[37] requires sufficient MD data to fully parameterize the memory kernel up to its lifetime τ_K. Recently, we have shown that constructing $\boldsymbol{\mathcal{K}}(t)$ additionally requires first-order time derivatives (as in Ref. 37) or both first- and second-order time derivatives (as in Ref. 87) of the TPMs and we demonstrated that these approaches accurately predict the intermediate- and long-time dynamics of alanine dipeptide, a system small enough to affordably converge the underlying reference dynamics. However, obtaining converged TPMs becomes formidably challenging as we tackle larger biological systems such as FiP35 WW Domain, RNAP, and the human argonaute complex. Importantly, we established that the stability of $\boldsymbol{\mathcal{K}}(t)$ is crucially dependent on the statistical convergence of the underlying MD simulation.[87,120] In particular, we have shown that the qMSM dynamics predicted at various kernel cut-off times lead to noticeable fluctuations in the prediction of the slowest ITS of FiP35 WW Domain, although the TPMs were constructed from the state-of-the-art MD simulation data set provided by D. E. Shaw.[39] This issue is further amplified when probing the gate opening mechanism of RNAP, where atomistic MD simulations are prohibitively expensive (over 540,000 atoms in the simulation box), ultimately preventing sufficient sampling, and thus result in noisy memory kernels as shown in Fig. 4(b). Moreover, the error of the qMSM does not monotonically decrease with increasing τ_K values, as one would expect [Fig. 4(a)]. Importantly, the qMSM error decays faster than that of the MSM, suggesting that the qMSM still offers a strict improvement over the MSM. Indeed, the state 1 dynamics predicted utilizing the qMSM and MSM ($\tau_L = \tau_K = 30$ ns) show that the qMSM offers a slight improvement over the MSM, but is unable to accurately recapitulate the reference dynamics (open circles) within the error bars [Fig. 4(d)]. We have recently employed a data smoothing procedure to reduce numerical fluctuations, which enabled us to implement the qMSM to accurately capture the RNAP gate opening mechanism.[36]

For protein dynamics, the key incentive of GMEs lies in the fast decay of memory kernels when there exists a sufficiently good separation of timescales. Motivated by the insight that we are mainly interested in the slow modes underlying conformational changes, it is advantageous to focus only on the integrals of these memory kernels (which become constants when $\tau > \tau_K$) rather than the time-dependent kernel functions themselves ($\boldsymbol{\mathcal{K}}(t)$).

Therefore, we have recently developed an integrative GME (IGME)[120] that considers only the integrals of the memory kernels. To achieve this, we have reformulated the GME via a Taylor expansion of the memory term

$$\frac{\mathrm{d}}{\mathrm{d}t}\ln \mathbf{C}(t) = \dot{\mathbf{C}}(0) - \mathbf{M}_0(t) - \sum_{n=1}^{\infty}\left[\mathbf{C}^{-1}(t)\frac{(-1)^n}{n!}\frac{\mathrm{d}^n}{\mathrm{d}t^n}\mathbf{C}(t)\right]\mathbf{M}_n(t), \tag{8}$$

where $\mathbf{M}_n(t)$ correspond to the time integrals of memory kernels: $\mathbf{M}_n(t) = \int_0^t \boldsymbol{\mathcal{K}}(s)s^n\,\mathrm{d}s$. We obtained the analytical solution of the IGME [Eq. (8)] from a self-consistent approach.[120] This solution can be further parameterized with the MD simulation trajectories via a least square fitting scheme, providing accurate and numerically stable predictions of the long-time dynamics. As shown in Fig. 4(a), our IGME yields smaller and more stable RMSE curves of TPMs in comparison with qMSM over a wide range of τ_K for the gate opening dynamics of RNAP. Furthermore, we have previously shown that the qMSM has failed to provide a stable estimation of the slowest ITS for this system, while IGME yields consistent slowest ITS for different values of τ_K.[120] In addition, the IGME framework provides a convenient way of computing integrals of memory kernels [i.e., the zero-order term $M_0(t)$ in Eq. (8)]. We have previously shown that the memory kernel integrals computed from IGME are more stable than in the qMSM.[120] However, the implementation of IGME is computationally more expensive than qMSM, as one needs to scan two hyperparameters to choose the best IGME models. It is interesting to consider our approach in the context of the Integrated Variational Approach to conformational dynamics (IVAC).[121] IVAC integrates over multiple lag times when applying the variational approach to MSMs and can reduce numerical fluctuations and yield more robust results than the variational approach.

B. Recent Work: Noise-resilient generalized master equation: $\langle\mathcal{U}\rangle$-GME

To address the sensitivity to statistical noise and offer a simple and conceptually transparent framework to interpret non-Markovian conformational dynamics using the states-and-rates language of chemical kinetics, we have recently developed and demonstrated the validity of a time-convolutionless GME (TCL-GME) approach to biomolecular dynamics.[87] The TCL-GME offers an exact formulation of non-Markovian evolution of the TPM in terms of a time-dependent rate matrix, $\boldsymbol{\mathcal{R}}(t)$

$$\dot{\mathbf{C}}(t) = \boldsymbol{\mathcal{R}}(t)\mathbf{C}(t), \tag{9}$$

which becomes constant after some time τ_R associated with the duration of the non-Markovian evolution of the TPM. We have proved analytically that, at worst, τ_R is as long as the MSM lag time, τ_L, and that it is often much less, $\tau_R \leq \tau_L$. In the case of the argonaute complex, we showed that $\tau_L/\tau_R \sim 10^3$, indicating that our time-local non-Markovian framework is up to 1000 times more data-efficient than the state-of-the-art MSMs. Moreover, a simple comparison of Eq. (9) with the MSM equation of motion in Eq. (1) allows one identify the rate matrix in MSMs to be the same one obtained from the long-time average of the time-dependent rate matrix.

While a direct application of the qMSM and TCL-GME to biomolecular TPM data can suffer from numerical instabilities associated with low-resolution and statistically noisy reference data, our work addresses both problems by introducing our $\langle\mathcal{U}\rangle$-GME approach,[87] which combines our recently developed discrete-time TCL-GME[122] with a uniformly convergent and systematically improvable averaging procedure that tames the deleterious effects of statistical noise. Our discrete-time TCL-GME is easy to interpret as it avoids the convolutional nature of the qMSM [Eq. (2)] and the TTM [Eq. (5), and its discrete-time formulation eliminates potential sensitivity to the temporal resolution of the reference TPM dynamics by removing differentiation and integration discretization errors. These methodological advances have enabled us to predict, for example, the long-time dynamics of large and complex systems like the human argonaute complex. In contrast, MSMs fail to predict these dynamics because of insufficient reference data, while the qMSM struggles to overcome statistical noise in the reference data.

Importantly, the dynamical propagator $\mathcal{U}(t)$ shares the same plateau time τ_R as the time-dependent rate matrix $\mathcal{R}(t)$. In the case of the highly converged FiP35 WW Domain data set, we have shown that the $\langle\mathcal{U}\rangle$-GME is able to accurately predict the slowest ITS using only 10 ns, whereas the MSM requires 100 ns. In other words, $\tau_L/\tau_R \sim 10$, which is consistent with the inequality we derived in Ref. 87. In addition, the slowest ITS prediction by the $\langle\mathcal{U}\rangle$-GMEremains remarkably stable when the model is parameterized with increasing amounts of reference dynamics (increasing τ). Furthermore, our $\langle\mathcal{U}\rangle$-GME approach can even predict experimentally sensitive observables, such as the folding time of the FiP35 WW Domain with 92% accuracy using only 50 ns of the reference data, whereas a similarly accurate MSM would have required ∼500 ns. In RNAP, where obtaining converged reference MD data for sufficiently long times is computationally expensive, the $\langle\mathcal{U}\rangle$-GMEapproaches the slowest ITS from below [Fig. 3(b)]. In fact, both the IGME and $\langle\mathcal{U}\rangle$-GMEapproach the same long-time limit, lending further credence to their joint prediction of $\sim 120\mu$s for the slowest ITS.

In addition to its ability to accurately and efficiently predict conformational dynamics of biomolecular systems where other methods, such as MSMs and qMSMs, struggle, an attractive feature of our $\langle\mathcal{U}\rangle$-GME method is its ease of use and construction. The $\langle\mathcal{U}\rangle$-GME scheme leverages the fact that, upon integrating the TCL-GME

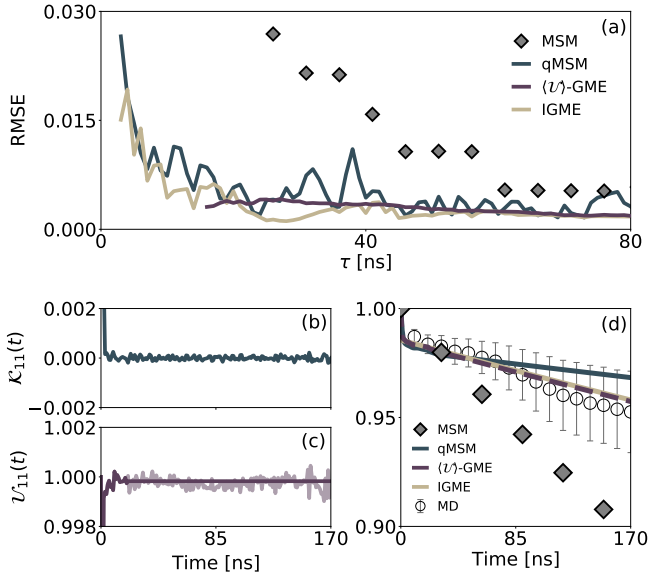

FIG. 4. Illustrating the strict advantage of noise-resilient approaches in the case of RNAP. (a) Root mean square error (RMSE) curves for the MSM, qMSM, $\langle \mathcal{U} \rangle$-GME ($t_r = 15$ ns), and IGME calculated with respect to the full 4-state reference MD data set (open circles). (b) State 1 memory kernel shown as a function of time. (c) The replacement of the state 1 $\mathcal{U}$ matrix (transparent line) with its time-average $\langle \mathcal{U}_\infty(\delta t) \rangle$ (dark line). (d) Chapman-Kolmogorov test: State 1 residence probabilities obtained from MD of the 4-state RNA polymerase model compared to those obtained from the MSM ($\tau_L = 30$ ns), qMSM ($\tau_K = 30$ ns), $\langle \mathcal{U} \rangle$-GME ($t_r = 20$ ns & $\tau_R = 30$ ns), and IGME ($\tau_K = 30$ ns).

in Eq. (9), one obtains an expression for the propagator in the reduced space that becomes constant beyond the Markovian onset, τ_R,

$$\boldsymbol{C}(\tau_R + n\delta t) = [\boldsymbol{\mathcal{U}}_\infty(\delta t)]^n \boldsymbol{C}(\tau_R), \tag{10}$$

where $\boldsymbol{\mathcal{U}}_\infty(\delta t)$ is the Markovian propagator. Generally, the propagator that evolves the TPM from time t to $t + \delta t$ can be obtained via simple matrix inversion and multiplication: $\boldsymbol{\mathcal{U}}(t + \delta t, t) = \mathbf{C}(t + \delta t)[\mathbf{C}(t)]^{-1}$. For $t \geq \tau_R$, $\boldsymbol{\mathcal{U}}(t + \delta t, t) = \boldsymbol{\mathcal{U}}_\infty(\delta t)$ is a constant rotation matrix. When simulation data is limited, statistical noise can mask the constant nature of $\boldsymbol{\mathcal{U}}_\infty(\delta t)$. To address this, our denoising procedure averages over individual $\boldsymbol{\mathcal{U}}_\infty(\delta t)$ obtained times greater than τ_R. We have further developed error measures to unambiguously identify the correct average long-time propagator, $\langle \boldsymbol{\mathcal{U}}_\infty(\delta t) \rangle$, through a variational procedure that also identifies τ_R (the onset of averaging) and the final averaging time.[87]

To highlight some of the advantages of the $\langle \mathcal{U} \rangle$-GME in predicting stable dynamics based on statistically underconverged data, we revisit the 4-state model used to investigate the gate opening dynamics of RNAP. Figure 4(c) shows the replacement of the state 1 propagator $\mathcal{U}_{11}(t + \delta t, t)$ with its time-averaged long-time limit $\langle \mathcal{U}_\infty(\delta t) \rangle$ after it has plateaued at $\tau_R = 30$ ns. Despite the apparent noisy fluctuations in $\mathcal{U}_{11}$, the averaging procedure leads to remarkable stability clear from the nearly monotonically decreasing RMSE with increasing values of τ_R in Fig. 4a. Note that increasing τ_R corresponds to increasing the size of the averaging window with the onset of averaging fixed at $t_r = 15$ ns. We utilize the methods we developed in Ref. 87 to identify $(t_r, \tau_R) = (20, 30)$ ns to parameterize our $\langle \mathcal{U} \rangle$-GME, which yields dynamics that faithfully recapitulate the reference MD dynamics [Fig. 4(d)] and outperforms both the qMSM and MSM approaches, despite being parameterized with the same quantity of reference dynamics (30 ns).

Hence, like the MSM, the $\langle \mathcal{U} \rangle$-GME is both easy to implement and provides a chemically intuitive kinetic description of the dynamics of biological configurations (e.g., folded, unfolded, bound, unbound). Unlike the MSM, the $\langle \mathcal{U} \rangle$-GME is formally exact, requires significantly less data to parameterize, and offers a resolution equal to that of the underlying reference MD simulations, whereas MSM resolution is limited by their lag time. Thus the $\langle \mathcal{U} \rangle$-GME can be expected to enable the investigation of biological mechanisms in complex systems with greater accuracy, stability, and resolution than the standard approaches.

C. Future Promises

The Hummer-Szabo projection operator[38] and its extension[7] provide a simple way to project all-atom biomolecular dynamics into transitions among a discrete set of metastable states. The success of this projection operator as a means to describe reduced dynamics hinges on the existence of timescale separation, which is often the case for biomolecular dynamics. In the future, we anticipate that variations of this projection operator could be developed to further improve its performance. For example, one could introduce a "fuzzy" projection operator ($\mathbb{P}$, where each MD conformation has certain probabilities to be assigned to multiple states), as the advantage of the fuzzy state boundaries has already been demonstrated by HMMs.[108] In addition, a hierarchical form of the Hummer-Szabo projection operator could be established where each metastable state could be further divided into smaller ones hierarchically.

The variational principle has been widely applied to find optimal collective variables that can faithfully describe the slowest timescales of biomolecular dynamics of interest. For example, the loss functions at the heart of popular deep-learning approaches, including the variational approach for Markov processes VAMPnets[81] and State-free Reversible VAMPnets (SRVnets),[123] rely on the variational principle of Markovian processes. More recently, Schütte and co-workers have proved that the projection of high-dimensional stochastic dynamics onto optimal collective variables is Markovian or memoryless, under the condition that there are timescale separations.[124] Therefore, besides the variational prin-

ciple of Markovian dynamics, one could envision finding optimal collective variables by minimizing memory kernels lifetimes, offering a potentially promising way to design the loss functions for deep neural network approaches.

Finally, while memory kernels have captured the imagination for many decades, it is only recent advances that have made it possible to extract their *exact* forms based on reference dynamics for complex many-body systems. Thus, investigations of memory kernels may provide new insights into biomolecular mechanisms. For example, recent work has leveraged persistent homology analysis to identify the existence of flux vortices during the conformational transition of alanine dipeptide.[125,126] We expect that such vortices will leave an imprint on the memory kernel. Yet, the extent to which such features can be simply gleaned and interpreted from the structure of the non-Markovian generators — whether these are memory kernels, their integrated moments, or their time-local analogs — remains an open question. Hence, the examination of such generators could facilitate our understanding of complex behaviors exhibited by the non-Markovian dynamics of biomolecular systems.

V. OUTLOOK

A central goal in biophysics is to elucidate and ultimately develop strategies to control the mechanisms of large biomolecular complexes whose conformational dynamics encode their function and span processes ranging from signal transduction, to gene regulation, and allostery. While MSMs serve as a major workhorse for investigating these conformational changes, their accuracy and efficiency depend on correctly identifying the slowest degrees of freedom in complex systems embedded in similarly complex environments — a formidable open question. When identifying these collective variables is difficult, when such separation of timescales does not exist, or when the processes of interest are not the slowest (e.g., dihedral flips in proteins[9]), constructing MSMs can become computationally expensive and, because of their eponymous Markovian approximation, yield inaccurate results.

In this perspective, we have argued that memory can provide a transformative tool to go beyond these limitations, and has even allowed us to interrogate the long-time dynamics of large biomolecular systems and their complex processes, such as the gate-opening of the RNAP II and the sequence-specific recognition of mRNA by the human argonaute complex 2, both of which are crucial to human gene regulation. We have demonstrated how memory is the persistent theme that a wide variety of cutting-edge schemes — ranging from GLEs, to history-augmented MSMs, and methods leveraging recurrent neural networks — employ to achieve greater accuracy while minimizing computational cost. GMEs constructed from equilibrium MD data employing a configurational, macrostate-based projection operator of the Zwanzig[84] or Hummer-Szabo[38] types allow one to encode the influence of the intrastate dynamics into the memory kernel in a time convolutional form. Enabled by recent advances, including the self-consistent expansion of the memory kernel in terms of $\boldsymbol{\mathcal{C}}(t)$,[74,76,87] its direct construction via discretization of the convolution integral,[37] and the TTM,[88] we are now in a position to directly obtain an exactly parameterized memory kernel based on short-time reference dynamics, as we have demonstrated in systems including alanine dipeptide,[37] FiP35 WW Domain,[37] RNAP,[36] and the human argonaute complex.[86]

The future of biomolecular dynamics, especially in large and complex systems where collecting sufficient statistics becomes prohibitively expensive and which present a challenge to state-of-the-art MSMs and the direct application of the time-convolution GME (i.e., qMSM), looks particularly bright given recent advances like the integrative GME (or IGME) and the noise-resilient $\langle\mathcal{U}\rangle$-GME approaches. These methods leverage the infrastructure laid for MSMs over the past two decades, from state clustering to TPM estimation, and have the potential to inspire novel approaches to finding slow degrees of freedom and collective variables through a variational formulation of the exact non-Markovian dynamics. Moreover, these methods offer distinct advantages: they can tame the deleterious effect of statistical noise in reference dynamics, require only minimal reference data compared with MSMs resulting in significant computational savings, and predict accurate long-timescale dynamics of the chosen degrees of freedom (even when these are not the slowest), correctly identify the slowest implied timescales, and faithfully capture sensitive experimental observables, such as protein folding times. Thus, by offering stability, accuracy, and efficiency, we anticipate that advances leveraging memory open exciting possibilities to unravel the mechanistic mysteries of biomolecular systems previously considered beyond the reach of existing methods.

VI. ACKNOWLEDGEMENT

X.H. acknowledges the support from the Office of the Vice-Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation. X.H. also acknowledges the support from the Hirschfelder Professorship Fund. A.M.C. acknowledges the start-up funds from the University of Colorado, Boulder.

[1] J. D. Chodera and F. Noé, Current Opinion in Structural Biology **25**, 135 (2014).
[2] L. Zhang, F. Pardo-Avila, I. C. Unarta, P. P.-H. Cheung, G. Wang, D. Wang, and X. Huang, Accounts of Chemical Research **49**, 687 (2016).
[3] G. R. Bowman, E. R. Bolin, K. M. Hart, B. C. Maguire, and S. Marqusee, PNAS **112**, 2734 (2015).

[4] J. R. Wagner, C. T. Lee, J. D. Durrant, R. D. Malmstrom, V. A. Feher, and R. E. Amaro, Chemical Reviews **116**, 6370 (2016).
[5] D. Shukla, Y. Meng, B. Roux, and V. S. Pande, Nature Communications **5** (2014).
[6] C. R. Schwantes, R. T. McGibbon, and V. S. Pande, Journal of Chemical Physics **141**, 090902 (2014).
[7] W. Wang, S. Cao, L. Zhu, and X. Huang, WIREs Computational Molecular Science **8**, e1343 (2018).
[8] V. S. Pande, K. Beauchamp, and G. R. Bowman, Methods **52**, 99 (2010).
[9] B. E. Husic and V. S. Pande, Journal of the American Chemical Society **140**, 2386 (2018).
[10] G. R. Bowman, V. S. Pande, and F. Noé, *Advances in Experimental Medicine and Biology*, Vol. 797 (2013).
[11] N. V. Buchete and G. Hummer, Journal of Physical Chemistry B **112**, 6057 (2008).
[12] R. D. Malmstrom, C. T. Lee, A. T. Van Wart, and R. E. Amaro, Journal of Chemical Theory and Computation **10**, 2648 (2014).
[13] X. Huang, G. R. Bowman, S. Bacallado, and V. S. Pande, PNAS **106**, 19765 (2009).
[14] F. Morcos, S. Chatterjee, C. L. McClendon, P. R. Brenner, R. López-Rendón, J. Zintsmaster, M. Ercsey-Ravasz, C. R. Sweet, M. P. Jacobson, J. W. Peng, and J. A. Izaguirre, PLoS Computational Biology **6**, e1001015 (2010).
[15] B. W. Zhang, W. Dai, E. Gallicchio, P. He, J. Xia, Z. Tan, and R. M. Levy, Journal of Physical Chemistry B **120**, 8289 (2016).
[16] A. C. Pan and B. Roux, Journal of Chemical Physics **129**, 064107 (2008).
[17] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, PNAS **106**, 19011 (2009).
[18] G. R. Bowman, V. A. Voelz, and V. S. Pande, Current Opinion in Structural Biology **21**, 4 (2011).
[19] N.-J. Deng, W. Dai, and R. M. Levy, Journal of Physical Chemistry B **117**, 12787 (2013).
[20] H. Wan, Y. Ge, A. Razavi, and V. A. Voelz, Journal of Chemical Theory and Computation **16**, 1333 (2020).
[21] I. Buch, T. Giorgino, and G. De Fabritiis, PNAS **108**, 10184 (2011).
[22] M. Lawrenz, D. Shukla, and V. S. Pande, Scientific Reports **5**, 7918 (2015).
[23] D.-A. Silva, G. R. Bowman, A. Sosa-Peinado, and X. Huang, PLoS Computational Biology **7**, e1002054 (2011).
[24] N. Plattner and F. Noé, Nature Communications **6**, 7653 (2015).
[25] D.-A. Silva, D. R. Weiss, F. P. Avila, L.-T. Da, M. Levitt, D. Wang, and X. Huang, Proceedings of the National Academy of Sciences of the United States of America **111**, 7665 (2014).
[26] K. J. Kohlhoff, D. Shukla, M. Lawrenz, G. R. Bowman, D. E. Konerding, D. Belov, R. B. Altman, and V. S. Pande, Nature Chemistry **6**, 15 (2014).
[27] L. T. Da, F. Pardo-Avila, L. Xu, D. A. Silva, L. Zhang, X. Gao, D. Wang, and X. Huang, Nature Communications **7**, 11244 (2016).
[28] L.-T. Da, D. Wang, and X. Huang, Journal of the American Chemical Society **134**, 2399 (2012).
[29] R. D. Malmstrom, A. P. Kornev, S. S. Taylor, and R. E. Amaro, Nature Communications **6**, 7588 (2015).
[30] B. Wang, R. E. Sexton, and M. Feig, Biochimica et Biophysica Acta - Gene Regulatory Mechanisms **1860**, 482 (2017).
[31] M. Khaled, A. Gorfe, and A. Sayyed-Ahmad, Journal of Physical Chemistry B **123**, 7667 (2019).
[32] E. P. Barros, Demir, J. Soto, M. J. Cocco, and R. E. Amaro, Chemical Science **12**, 1891 (2021).
[33] J. Feng, B. Selvam, and D. Shukla, Structure **29**, 922 (2021).
[34] C. Y. Son, A. Yethiraj, and Q. Cui, PNAS **114**, E8830 (2017).
[35] V. A. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande, Journal of the American Chemical Society **132**, 1526 (2010).
[36] I. Christy Unarta, S. Cao, S. Kubo, W. Wang, P. Pak-Hang Cheung, X. Gao, S. Takada, and X. Huang, PNAS **118**, e2024324118 (2021).
[37] S. Cao, A. Montoya-Castillo, W. Wang, T. E. Markland, and X. Huang, The Journal of Chemical Physics **153**, 014105 (2020).
[38] G. Hummer and A. Szabo, Journal of Physical Chemistry B **119**, 9029 (2015).
[39] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, Science **330**, 341 (2010).
[40] F. Liu, D. Du, A. A. Fuller, J. E. Davoren, P. Wipf, J. W. Kelly, and M. Gruebele, PNAS **105**, 2369 (2007).
[41] K. Röder and D. J. Wales, Frontiers in Molecular Biosciences **9** (2022), 10.3389/fmolb.2022.820792.
[42] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, The Journal of Chemical Physics **134**, 174105 (2011).
[43] F. Nüske, B. G. Keller, G. Pérez-Hernández, A. S. Mey, and F. Noé, Journal of Chemical Theory and Computation **10**, 1739 (2014).
[44] K. A. Konovalov, I. C. Unarta, S. Cao, E. C. Goonetilleke, and X. Huang, JACS Au **1**, 1330 (2021).
[45] J. Lu and E. Vanden-Eijnden, Journal of Chemical Physics **141**, 044109 (2014).
[46] A. Kai-Hei Yik, Y. Qiu, I. C. Unarta, S. Cao, and X. Huang, ChemRxiv (2022).
[47] Z. F. Brotzakis and M. Parrinello, Journal of Chemical Theory and Computation **15**, 1393 (2019).
[48] J. Rogal, E. Schneider, and M. E. Tuckerman, Physical Review Letters **123**, 245701 (2019).
[49] H. Klem, G. M. Hocky, and M. McCullagh, Journal of Chemical Theory and Computation **18**, 3218 (2021).
[50] H. Grabert, *Projection Operator Techniques in Nonequilibrium Statistical Mechanics*, Vol. 95 (Springer Berlin, Heidelberg, 1982).
[51] S. Nakajima, Progress of Theoretical Physics **20**, 948 (1958).
[52] H. Mori, The Physical Review **112**, 1829 (1958).
[53] R. Zwanzig, The Journal of Chemical Physics **33**, 1338 (1960).
[54] P. C. Martin and S. Yip, Physical Review **170**, 151 (1968).
[55] G. D. Harp and B. J. Berne, Physical Review A **2** (1970).
[56] D. Levesque and L. Verle, Physical Review A **2**, 2514 (1970).
[57] J.-P. Boon and S. A. Rice, The Journal of Chemical Physics **47**, 2480 (1967).
[58] T.-W. Nee and R. Zwanzig, The Journal of Chemical Physics **52**, 6353 (1970).
[59] E. Rabani, K. Miyazaki, and D. R. Reichman, Journal of Chemical Physics **122**, 034502 (2005).
[60] E. Rabani and D. R. Reichman, Annual Review of Physical Chemistry **56**, 157 (2005).
[61] T. E. Markland, J. A. Morrone, K. Miyazaki, B. J. Berne, D. R. Reichman, and E. Rabani, Journal of Chemical Physics **136**, 074511 (2012).
[62] B. U. Felderhof, J. M. Deutch, and U. M. Titulaer, The Journal of Chemical Physics **63**, 740 (1975).
[63] K. S. Schweizer, The Journal of Chemical Physics **91**, 5802 (1989).
[64] A. Davtyan, G. A. Voth, and H. C. Andersen, Journal of Chemical Physics **145** (2016).
[65] P. Español and P. B. Warren, Journal of Chemical Physics **146**, 150901 (2017).
[66] A. Nitzan, *Chemical Dynamics in Condensed Phases* (Oxford University Press, 2006).
[67] V. May and O. Kühn, *Charge and Energy Transfer Dynamics in Molecular Systems*, 3rd ed. (John Wiley & Sons, 2011).
[68] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, 1985) pp. 444–447.
[69] S. Pressé, J. Lee, and K. A. Dill, Journal of Physical Chemistry B **117**, 495 (2013).
[70] S. Pressé, J. Peterson, J. Lee, P. Elms, J. L. Maccallum, S. Marqusee, C. Bustamante, and K. Dill, Journal of Physical Chemistry B **118**, 6597 (2014).
[71] Q. Shi and E. Geva, Journal of Chemical Physics **119**, 12063 (2003).

[72] M. L. Zhang, B. J. Ka, and E. Geva, Journal of Chemical Physics **125** (2006).
[73] G. Cohen, E. Y. Wilner, and E. Rabani, New Journal of Physics **15** (2013).
[74] A. Montoya-Castillo and D. R. Reichman, Journal of Chemical Physics **144**, 184104 (2016).
[75] A. Montoya-Castillo and D. R. Reichman, Journal of Chemical Physics **146**, 084110 (2017).
[76] A. Kelly, A. Montoya-Castillo, L. Wang, and T. E. Markland, Journal of Chemical Physics **144**, 184105 (2016).
[77] N. Ng, D. T. Limmer, and E. Rabani, Journal of Chemical Physics **155**, 156101 (2021).
[78] E. Mulvihill and E. Geva, Journal of Physical Chemistry B **125**, 9834 (2021).
[79] G. Amati, M. A. C. Saller, A. Kelly, and J. O. Richardson, Journal of Chemical Physics **157**, 1 (2022).
[80] R. T. McGibbon, B. E. Husic, and V. S. Pande, Journal of Chemical Physics **146**, 044109 (2017).
[81] A. Mardt, L. Pasquali, H. Wu, and F. Noé, Nature Communications **9**, 5 (2018).
[82] F. Sittel and G. Stock, Journal of Chemical Physics **149**, 150901 (2018).
[83] W. E, T. Li, and E. Vanden-Eijnden, *Applied Stochastic Analysis*, Vol. 199 (American Mathematical Society, Providence, RI, 2019).
[84] R. Zwanzig, Journal of Statistical Physics **30**, 255 (1983).
[85] K. Blum, *Density Matrix Theory and Applications*, 3rd ed., Vol. 64 (Springer Berlin Heidelberg, 2012) pp. 280–281.
[86] L. Zhu, H. Jiang, S. Cao, I. C. Unarta, X. Gao, and X. Huang, Communications Biology **4**, 1345 (2021).
[87] A. J. Dominic III, T. Sayer, S. Cao, T. E. Markland, X. Huang, and A. Montoya-Castillo, bioRxiv 2022.10.17.512620 (2022).
[88] J. Cerrillo and J. Cao, Physical Review Letters **112**, 110401 (2014).
[89] A. A. Kananenka, C.-Y. Hsieh, J. Cao, and E. Geva, Journal of Physical Chemistry Letters **7**, 4809 (2016).
[90] R. Rosenbach, J. Cerrillo, S. F. Huelga, J. Cao, and M. B. Plenio, New Journal of Physics **18**, 023035 (2016).
[91] F. A. Pollock and K. Modi, Quantum **2** (2018), 10.22331/q-2018-07-11-76.
[92] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, 3rd ed. (Elsevier Science & Technology Books, 1981).
[93] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, 2001).
[94] O. F. Lange and H. Grubmüller, Journal of Chemical Physics **124**, 214903 (2006).
[95] R. Hegger and G. Stock, Journal of Chemical Physics **130**, 034106 (2009).
[96] C. Ayaz, L. Tepper, F. N. Brünig, J. Kappler, J. O. Daldrop, and R. R. Netz, PNAS **118**, e2023856118 (2021).
[97] I. Horenko, C. Hartmann, C. Schütte, and F. Noe, Physical Review E **76**, 016706 (2007).
[98] E. Fick and G. Sauermann, *The Quantum Statistics of Dynamic Processes* (Springer-Verlag, Berlin, 1990).
[99] H. Vroylandt, L. Goudenège, P. Monmarché, F. Pietrucci, and B. Rotenberg, PNAS **119**, e117586119 (2022).
[100] C. Ayaz, L. Scalfi, B. A. Dalton, and R. R. Netz, Physical Review E **105**, 054138 (2022).
[101] G. Jung, M. Hanke, and F. Schmid, Journal of Chemical Theory and Computation **13**, 2481 (2017).
[102] R. Satija and D. E. Makarov, Journal of Physical Chemistry B **123**, 802 (2019).
[103] V. Klippenstein, M. Tripathy, G. Jung, F. Schmid, and N. F. Van Der Vegt, Journal of Physical Chemistry B **125**, 4931 (2021).
[104] W. C. Swope, J. W. Pitera, and F. Suits, Journal of Physical Chemistry B **108**, 6571 (2004).
[105] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, Journal of Chemical Physics **126**, 155101 (2007).
[106] C. R. Schwantes and V. S. Pande, Journal of Chemical Theory and Computation **9**, 2000 (2013).
[107] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, Journal of Chemical Physics **139**, 015102 (2013).
[108] F. Noé, H. Wu, J. H. Prinz, and N. Plattner, Journal of Chemical Physics **139**, 184114 (2013).
[109] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, The Annals of Mathematical Statistics **41**, 164 (1970).
[110] L. R. Welch, IEEE Information Theory Society Newsletter **53**, 1 (2003).
[111] R. T. McGibbon, B. Ramsundar, M. M. Sultan, G. Kiss, and V. S. Pande, Proceedings of the 31st International Conference on Machine Learning **32**, 1197 (2014).
[112] B. G. Keller, A. Kobitski, A. Jäschke, G. U. Nienhaus, and F. Noé, Journal of the American Chemical Society **136**, 4534 (2014).
[113] J. Copperman and D. M. Zuckerman, Journal of Chemical Theory and Computation **16**, 6763 (2020).
[114] S. Bacallado, Annals of Statistics **39**, 838 (2011).
[115] S.-T. Tsai, E.-J. Kuo, and P. Tiwary, Nature Communications **11** (2020).
[116] S.-T. Tsai, E. Fields, Y. Xu, E.-J. Kuo, and P. Tiwary, Nature Communications **13**, 1 (2022).
[117] W. Zeng, S. Cao, X. Huang, and Y. Yao, arXiv preprint (2021), https://doi.org/10.48550/arXiv.2107.06573.
[118] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, and D. Silver, Nature **575**, 350 (2019).
[119] A. Vaswani, G. Brain, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin, Advances in neural information processing systems **30** (2017).
[120] S. Cao, Y. Qiu, M. Kalin, and X. Huang, ChemRxiv 10.26434/chemrxiv-2022-0n9ld (2022), 10.26434/chemrxiv-2022-0n9ld.
[121] C. Lorpaiboon, E. H. Thiede, R. J. Webber, J. Weare, and A. R. Dinner, Journal of Physical Chemistry B **124**, 9354 (2020).
[122] T. Sayer and A. Montoya-Castillo, Journal of Chemical Physics **158**, 014105 (2023).
[123] H. Sidky, W. Chen, and A. L. Ferguson, Journal of Physical Chemistry B **123**, 7999 (2019).
[124] W. Zhang, C. Hartmann, and C. Schütte, Faraday Discussions **195**, 365 (2016).
[125] F. Manuchehrfar, H. Li, W. Tian, A. Ma, and J. Liang, Journal of Physical Chemistry B **125**, 4667 (2021).
[126] F. Manuchehrfar, H. Li, A. Ma, and J. Liang, The Journal of Physical Chemistry B (2022).