Non-Markovian Dynamic Models of Protein Conformational Changes

Xuhui Huang
Department of Chemistry
University of Wisconsin-Madison
Dynamics of Conformational Changes are Crucial for Protein Function

Dynamic and localized transitions between pairs of conformational states:

**RNA Polymerase translocation**


**DNA repair protein translocates on dsDNA**

*PNAS*, 117, 21889, (2020)

**Protein-ligand recognition**

*PLOS. Comp. Bio.*, 7, e1002054, (2011)

**DNA loading into RNA Polymerase**

*PNAS*, 118(17), e2024324118, (2021)
Projection Operator Approach of Protein Dynamics

Evolution of density in phase space satisfies Liouville’s Equation:

$$\frac{\partial \rho(\Gamma, t)}{\partial t} = \mathcal{L}\rho(\Gamma, t)$$

\(\Gamma = (x; p)\)

Mori-Zwanzig projection operator: \(\Gamma \rightarrow \chi\)

\(\mathcal{P}\): slow variables \(\chi\)

\(\mathcal{Q}\): fast variables \(\Gamma \cap \bar{\chi}\)

Phase space \(\Gamma\)

Zwanzig, Mori, Berne
Projecting Kinetics onto Coarse-grained States Introduces Memory

Hummer-Szabo projection operator:

\[ \mathcal{P} := \sum_{j=1}^{n} |\rho(R; \text{eq})\chi_j(x)\rangle \cdot \pi_j^{-1}\langle \chi_j(x) | \]

We choose \( |\chi\rangle \) to be the state indicator function \( \chi_i(x) = 1 \): conformation \( x \) belongs to state \( i \)

\[ \mathcal{P} : \text{Slow transitions between states} \]
\[ \mathcal{Q} : \text{Fast transitions within state} \]

Separation of timescales

The projected kinetics satisfy a **Generalized Master Equation**:

\[ \frac{\partial}{\partial t} \mathbf{T}(t) = \dot{\mathbf{T}}(0)\mathbf{T}(t) + \int_{0}^{t} \mathbf{K}(t')\mathbf{T}(t - t')dt' \]

Memory kernel: \( \mathbf{K}(t) = \langle \mathcal{L}_e^{QLt}QL \rangle_{\rho,\pi^{-1}} \)

Projecting Kinetics onto Coarse-grained States Introduces Memory

Hummer-Szabo projection operator:

\[ \mathbb{P} := \sum_{j=1}^{n} |\rho(R; \text{eq}) \chi_j(x)\rangle \cdot \pi_j^{-1} \langle \chi_j(x) | \]

We choose \( |\chi\rangle \) to be the state indicator function \( \chi_i(x) = 1 \): conformation \( x \) belongs to state \( i \)

\[ \mathbb{P} : \text{Slow transitions between states} \]
\[ \mathbb{Q} : \text{Fast transitions within state} \]

Discretion of time (lag time: \( \tau \)) is sufficiently long so that: \( \mathbb{P} e^{\mathbb{Q} \tau} \mathbb{Q} \approx 0 \)

We obtain a Markov State Model (MSM):

\[ p(t + \tau) = T \cdot p(t) \]

MSMs are often non-Markovian due to Limited Length of MD Simulations

NTL9 folding:

- **2000-state MSM**: lag time = 10ns. 
- **14-state model**: Not Markovian
  

RNA Polymerase backtracking:

- **800-state MSM**: lag time = 8ns (480 100-ns MD simulations)
- **4-state model**: Not Markovian
  
Quasi-Markov State Model (qMSM) Theory

**Key Insights:** Due to separation of timescales, the memory kernel (mainly reflecting intra-state transition) decays faster than the Markovian lag time.

- **Our approach:** Propagate dynamics using a Generalized Master Equation with memory kernel.

\[
\tau_K \ll \tau_M
\]

\[
\tau_K < \tau_M < t_0
\]

Computing Memory Kernel for Protein Dynamics

Direct discretization of GME:

\[
\dot{T}(n\Delta t) = \dot{T}(0) T(n\Delta t) + \Delta t \sum_{m=1}^{n} \mathcal{H}(m\Delta t) T((n-m)\Delta t)
\]

GME:

\[
\dot{T}(t) = \dot{T}(0) T(t) - \int_{0}^{\min[\tau_K,t]} d\tau \mathcal{H}(\tau) T(t-\tau)
\]

A Simple Kinetic Model

3-state

Good lumping

Memory Kernel Integral

Root mean square errors of TPM

qMSM is more accurate than MSM.

WW domain Folding

4-State

$qMSM$ saves $>10$ times in MD length than MSM!

$qMSM \tau_K = 15\text{ns}$

$MSM \tau_M = 200\text{ns}$

Dynamics of RNAP Gate Opening is Crucial for DNA Promoter Loading

β-lobes
Clamp
Clamp Open
PDB ID: 5TGY

RNAP
Clamp Closed
PDB ID: 4XLN

Our Recipe for Constructing qMSMs to Study Functional Conformational Changes

**A** Initial pathway generation and optimization (TMD, Climber, String Method, TAPs)

**B** MD sampling

**C** Feature selection (Spectral oASIS, Feature importance selection)

**D** Dimensionality reduction (tICA and deep learning, i.e.: SRV)

**E** Clustering (K-means, K-centers, APLoD)

**F** Macrostate qMSM

*JACS Au*, 1, 1330-1341, (2021)
Our Recipe for Constructing qMSMs to Study Functional Conformational Changes

**JACS Au**, 1, 1330-1341, (2021)
Coarse-Grained MD Simulations to Generate Initial Path

Shoji Takada
Kyoto U

CafeMol
Back-map coarse–grained conformations to all-atom conformations.

Our Recipe for Constructing qMSMs to Study Functional Conformational Changes

306 200-ns MD simulations
System size: 543K atoms
Automatic Selection of Features that can Describe Protein Conformational Changes

microRNA target recognition by RNA induced silencing complex


Spectral oASIS to choose 1,000 residue-residue distance features

*Communications Biology* 4 (1), 1345, (2021)
We selected 1770 residue-residue pairwise distances that can best describe the RNAP gate opening.

Our Recipe for Constructing qMSMs to Study Functional Conformational Changes

RNAP gate dynamics: a 4-state qMSM model
Clamp Closing is Rate-Limiting and Occurs at Milliseconds

Two intermediate states with different conformations of Switch 2.
Clamp Closing is Dynamically Correlated with the Switch-2 Region

Switch-2 conformation:
S1-S2: alpha-helix
S3-S4: pi-helix

RNAP Clamp-Switch 2 has highest dynamic correlation.
Unfolded Switch-2 Conformations are Spontaneously Sampled by the Partially Closed Intermediate State
Unfolded Switch-2 Conformations Allow the Binding of Antibiotics Myx

Partially unfolded and detached Switch-2 conformations allow sufficient space for the binding of antibiotics Myx.
F614A Mutant Causes Hypersensitivity to Myx

F614A allows for more favourable binding with Myx by providing more space in the binding pocket.

F614A allows for more favourable binding with Myx by providing more space in the binding pocket.


qMSMs Greatly Outperform MSMs

For MSM, it requires a lag time as long as 2000 ns in order to build a Markovian model (estimated by GME).

Overcoming the Challenges that still Face GME Models
Numerical Instability of Memory Kernels for Complex Systems

Failed for most $\tau_k$!
A Smoothing Scheme Improves Numerical Stability of Memory Kernels, but Not Enough!

We fit time evolutions of TPM’s eigenvalues to multiexponential functions and re-assemble TPMs based on smoothed eigenvalues.
Integrative-GME (IGME) Method

Taylor expansion of the memory term in GME:

$$\frac{d}{dt} \ln T(t) = \dot{T}(0) - M_0(t) - \left( T(t)^{-1} \sum_{m=1}^{n} \frac{(-1)^m}{m!} \frac{d^m}{dt^m} T(t) \right) M_m(t)$$

Integrals of memory:

$$M_m(t) = \int_0^t K(s) s^m ds$$

We solve this equation self-consistently.

IGME substantially outperforms qMSM in yielding stable estimations of dynamics!

“Integrative Generalized Master Equation”, Chemrxiv: DOI: 10.26434/chemrxiv-2022-0n9ld-v2
Integrative-GME (IGME) Models are Stable

IGME provides a numerically robust way to propagate the GME because it only considers the integrals of the memory kernels.

“Integrative Generalized Master Equation”, Chemrxiv: DOI: 10.26434/chemrxiv-2022-0n9ld-v2
Integrative Generalized Master Equation: A Theory to Study Long-timescale Biomolecular Dynamics via the Integrals of Memory Kernels

Sigin Cao University of Wisconsin–Madison,
Yunrui Qiu University of Wisconsin–Madison,
Michael Kalin University of Wisconsin–Madison,
Xuhui Huang University of Wisconsin–Madison

Abstract

The generalized master equation (GME) provides a powerful approach to study biomolecular dynamics via non-Markovian dynamic models built from molecular dynamics (MD) simulations. Previously, we have implemented the GME for biomolecular dynamics, namely the quasi Markov State Model (qMSM), where we explicitly calculate the memory kernel and propagate protein dynamics using a discretized GME. qMSM can be constructed with much shorter MD simulation trajectories than the Markov State Model (MSM). However, since qMSM needs to explicitly compute the time-dependent memory kernels, it is heavily affected by the

IGME: https://github.com/xuhuihuang/IGME
Acknowledgement

Group members

Siqin Cao
Ilona Unarta
Bojun Liu
Yunrui Qiu
Michael Kalin

Collaborators

Tom Markland (Stanford U.)
Andres Montoya-Castollo (CU Boulder)
Shoji Takada (Kyoto U.)
Xin Gao (KAUST)
Lizhe Zhu (CUHK)

Hirschfelder Endowed Professorship Fund