

BY GREGORY R. BOWMAN

## Understanding Molecular Kinetics with Markov State Models



Atomistic simulations have the potential to elucidate the molecular basis of biological processes like protein misfolding in Alzheimer's disease or the conformational changes that drive transcription or translation. However, most simulations can only capture the nanosecond to microsecond timescale, whereas most biological processes of interest occur on millisecond and longer timescales. Also, even with an infinitely fast computer, extracting meaningful insight from simulations is difficult because of the complexity of the underlying free energy landscapes. Fortunately, Markov State Models (MSMs) can help overcome these limitations.

MSMs may be used to model any random process where the next state depends solely on the current state. For example, imagine exploring New York City by

rolling a die to randomly select which direction to go in each time you came to an intersection. Such a process could be described by an MSM with a state for each intersection. Each state might have a probability of 1/8 of going to each of the four neighboring intersections, a probability of 1/2 of getting stuck at a red light (e.g., staying at the current state), and a probability of 0 of going directly to any other intersection. Drawing such a model would result in something resembling a road map with speed limits replaced by probabilities. Of course, the probabilities of going North, West, East, or South at a given intersection don't have to be the same; they just have to sum to one—because you have to go somewhere.

MSMs for molecular kinetics are conceptually similar to our road-map example but instead of intersections, the states now correspond to basins in the free energy landscape governing the dynamics of the molecule. These states are referred to as metastable states because a molecule is more likely to stay in a particular state than to transition to a new one. Each state may also have many more connections than a typical intersection because of the enormous number of degrees of freedom in most biomolecules.

In our road-map example, one can easily imagine defining states and their connectivity by referring to satellite images and road signs. MSMs for molecular kinetics, however, must be inferred from simulation trajectories (like molecular dynamics trajectories). It's like being asked to draw a map of New York City from GPS coordinates taken at regular intervals by a few drivers. Fortunately, we can make great headway by recognizing

MSMs for molecular kinetics must be inferred from simulation trajectories...

It's like being asked to draw a map of New York City from GPS coordinates taken at regular intervals by a few drivers.

that it should be possible to quickly transition between conformations in the same free energy basin (or metastable state) while transitions between different basins will be slow because they are separated by significant free energy barriers. Thus, we can build MSMs by grouping conformations that can reach one another quickly. These groups of conformations become the states of our model and we can simply count transitions between states in our simulations to determine the probabilities of going from one to another.

MSMs for molecular kinetics have many advantages over other approaches. Simply inspecting an MSM can provide an intuition for the dynamics of the system and calculations performed with the matrix representation of MSMs, plus a few representative conformations from each state, make it possible to quantitatively compare with experimental measurements, like fluorescence relaxation curves. MSMs also provide a means of aggregating the data from many simulations into a single model. Moreover, just as it is possible to build up a road map by assembling information representing different locations, a larger, more comprehensive molecular kinetics map can be assembled from many shorter simulations. While MSMs have mostly been used to understand conformational changes on the microsecond to millisecond timescale with atomic resolution, an exciting future direction will be to use them to address ever larger, slower, and more biologically relevant systems. □

MSMs for molecular kinetics have many advantages over other approaches. Simply inspecting an MSM can provide an intuition for the dynamics of the system and calculations performed with the matrix representation of MSMs, plus a few representative conformations from each state, make it possible to quantitatively compare with experimental measurements, like fluorescence relaxation curves. MSMs also provide a means of aggregating the data from many simulations into a single model. Moreover, just as it is possible to build up a road map by assembling information representing different locations, a larger, more comprehensive molecular kinetics map can be assembled from many shorter simulations. While MSMs have mostly been used to understand conformational changes on the microsecond to millisecond timescale with atomic resolution, an exciting future direction will be to use them to address ever larger, slower, and more biologically relevant systems. □

### DETAILS

Gregory R. Bowman is a PhD student in Vijay Pande's lab at Stanford University. He is the primary developer of MSMBuilder, a freely available tool for the automated construction and analysis of MSMs (<https://simtk.org/home/msmbuilder>), and is currently using it to understand protein and RNA folding.