Pacific Symposium on Biocomputing 15:260-268(2010)

"CROSS-GRAINING:" EFFICIENT MULTI-SCALE SIMULATION VIA MARKOV STATE MODELS

PETER M. KASSON
Departments of Chemistry and Structural Biology , Stanford University
Stanford, CA 94305 USA

VIJAY S. PANDE
Departments of Chemistry and Structural Biology , Stanford University
Stanford, CA 94305 USA

Accurate and efficient methods to simulate biomolecular systems at multiple levels of detail simultaneously are an ongoing challenge for the simulation community. Here we present a new method for multi-scale simulation where a complex system can be partitioned into two loosely-coupled sub-systems, one coarse-grained and one atomistic. If the coupling between the coarse-grained and atomistic systems can be encoded into discrete states that interconvert slowly, we can construct a Markov state model where we approximate any given transition $P[(s_i,t_j)\text{->}(s_k,t_l)]$ in the joint space of the coarse-grained and atomistic systems as the product of two orthogonal transitions $P(s_i\text{->}s_k \mid t_j)$ and $P(t_j\text{->}t_l \mid s_j)$. We provide a formalism for constructing such models and describe how they may be applied to multi-scale simulation of membrane proteins. This "cross-graining" methodology may provide a general means to efficiently simulate mixed-scale systems.

1. Introduction

Many biological systems have the following canonical challenge: outcomes depend on fine degrees of freedom, yet they involve large systems and large timescales. Indeed, an underlying principle behind biological signal transduction is that fine sensitivity to subtle molecular changes controls large-scale effects. This poses a challenge for mechanistic simulation; it is often computationally intractable to capture all the fine-scale details of a system while simultaneously reaching the time and length scales required to address emergent behavior.

Driven by this challenge, "coarse-grained" approaches have been developed to use a reduced degree-of-freedom representation for more efficient computation. However, many systems have important transitions governed by fine-grained degrees of freedom such that it is challenging to derive a coarse-grained representation that is computationally efficient yet sufficiently predictive. Full mixed-scale simulation, in which fine-grained and coarse-grained representations can be combined freely, is an ongoing challenge for the field. In this work, we derive an alternate multi-scaling technique where fine-grained and coarse-grained representations are combined at the level of a statistical reaction model, a Markov state model. We show that under certain conditions this method, which we term "cross-graining," performs at least as well as the ideal mixed-scale simulation that has yet to be fully realized.

Whether atomistic or coarse-grained representations are used, Markov state models have recently gained traction as a means to analyze and interpret molecular simulation data [1-4]. They draw on the notion that at some level molecular reactions are Markov chains—stochastic processes where, given a suitable encoding of state information at time t and a minimum "lag time" τ, the state of the system at time $t + \tau$ depends only on the state at t. Methods have recently been developed to determine such state encodings from a set of simulation trajectories, assess model quality, and use the resulting statistical reaction model to predict ensemble properties and long-timescale behavior of the system [5-7].

Here we address an important subset of "multi-scale" problems where one can partition a system into two processes of interest, each with associated degrees of freedom. For instance, one process may depend primarily on atomistic degrees of freedom $\{x_i\}$, while another process depends primarily on coarse-grained degrees of freedom $\{y_i\}$. An example is given in Figure 1. As we will discuss, these processes can be correlated. We show how Markov state models yield a means to obtain multi-scale dynamics in this case and further derive a scheme for "cross-graining" where each set of transitions can be computed using only the relevant degrees of freedom. We

provide an example of how this approach yields efficient sampling compared even to a full "ideal" multi-scale simulation.

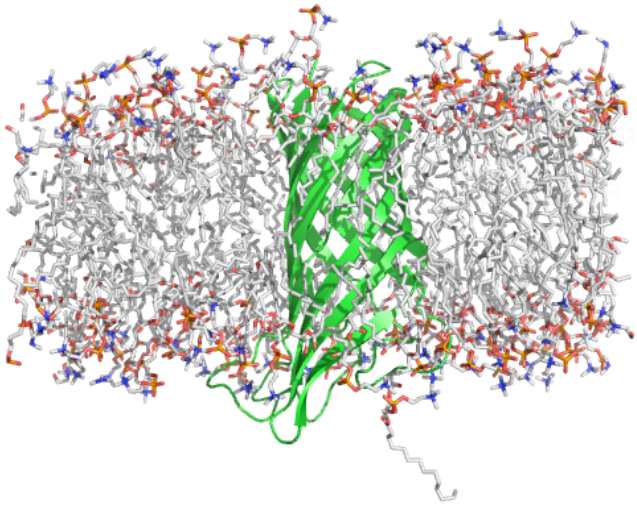

Figure 1. Sample membrane-protein system. Shown here is the *E. coli* outer membrane protein A (OmpA) inserted in a lipid bilayer [8]. OmpA is unfolded or aggregated in solution but assumes a beta-strand conformation in membranes; these strands then assemble to form a beta-barrel pore. To simulate this in a multi-scale manner, we consider the protein and its immediate environment atomistically as degrees of freedom $\{x_i\}$ and the larger membrane-water system as degrees of freedom $\{y_i\}$.

We motivate our cross-graining approach by considering the insertion and assembly of membrane pores. Contemporary coarse-gained force fields such as MARTINI [9] have had good success in simulating lipid membranes [10-16], and they have also been used to simulate protein assembly in membranes, given a constrained protein conformation [8, 17-19]. Robust simulation of protein refolding has proven more challenging. We seek an approach that can use atomistic simulation to determine the protein conformational dynamics within each coarse-grained "state" of the membrane system and also use the protein conformation in each state to inform a coarse-grained simulation of large-scale protein-membrane interactions—in this case by providing data for the coarse-grained protein torsions and conformational restraints. Grossly, one could imagine the fine-grained instantiations of each coarse-grained state as simulating the protein in solution, the protein in an encounter complex with the membrane, and a variety of membrane-inserted states. The coarse-grained state would determine the composition and macroscopic restraints on the protein surroundings for the fine-grained simulation; the fine-grained state would yield torsional restraints to constrain the protein for the coarse-grained simulations. Under the proper conditions, such an approach can provide both fine-scale conformational dynamics and efficient simulation of large systems. The cross-graining method we have developed addresses these challenges, although it also proves more general as well.

2. Theory

Let us consider a molecular system where the degrees of freedom can be partitioned into two sets, $\{x_i\}$ and $\{y_i\}$. Let us further consider two sets of processes, one of which primarily depends on $\{x_i\}$ (our fine-grained degrees of freedom) and one of which primarily depends on $\{y_i\}$ (our coarse-grained degrees of freedom). These two processes can be inter-dependent; we will both state this formally below and show a simple 2D example to make this intuitive. The fundamental approximation, however, is that at some level of granularity we can treat the joint free energy landscape $(\{x_i\},\{y_i\})$ as the direct product of the individual free energy landscapes $\{x_i\} \otimes \{y_i\}$.

Any classical molecular system can be represented as a Markov chain for some set of states and some time resolution [1]. A trivial proof of this is to consider the set of states corresponding to all points in phase space and

continuous time (or for a simulation, time resolution corresponding to the integrator time step). The time evolution of this system is a history-independent stochastic process. Therefore suppose we construct a Markov state model to represent the behavior of our system on degrees of freedom $\{x_i\}$ and a second Markov state model to represent the behavior on degrees of freedom $\{y_i\}$. For the moment, we will take the partitioning of degrees of freedom as given; determining the optimal partitioning of phase space into states and the optimal partitioning of degrees of freedom into $\{x_i\}$ and $\{y_i\}$ are separate problems, each non-trivial.

Since these degrees of freedom $\{x_i\}$ and $\{y_i\}$ may be coupled, any transition a -> b on $\{x_i\}$ will be Markovian on $\{x_i\}$ but may depend on $\{y_i\}$. Stated more formally, given a sequence of states $(u_1..u_n)$ on $\{x_i\}$ and a sequence of states $(v_1..v_n)$ on $\{y_i\}$, the mutual information $I(u_{n+1} ; u_n \mid u_1..u_{n-1}) = 0$ [20] but $I(u_{n+1} ; u_n \mid v_n) \geq 0$. We now partition state space $\{x_i\}$ into states $\{s_i\}$ and $\{y_i\}$ into "macrostates" $\{t_i\}$. We choose a set of macrostate definitions for S and T such that knowledge of (s,t) is sufficient information to make $P(a \rightarrow b)$ Markovian by the above criterion. In the limit of completely uncoupled degrees of freedom, there is no dependence on the state Y, while in the opposite limit of completely coupled degrees of freedom one could have a system where each point in phase space $\{y_i\}$ corresponds to a distinct state.

To make this intuitive, we consider the case where x and y each consist of one degree of freedom. A free energy diagram in this 2D space is schematized in Figure 2. Cross-graining applies well when each set of $(s_i,t_j) \rightarrow (s_k,t_l)$ transitions occurs on well-separated timescales. In one such case we may have $s_i \rightarrow s_k$ transitions followed sequentially by $t_j \rightarrow t_l$ transitions (Fig. 2a); in another the $s_i \rightarrow s_k$ transitions relax quickly enough compared to $t_j \rightarrow t_l$ (Fig. 2b) that we can make the following approximation:

$$P\left(\left(\{s\},t_j\right)\rightarrow\left(\{s\},t_l\right)\right)\approx\sum_{s_i\in\{s\}}P\left(s_i\right)\cdot P\left(\left(t_j\mid s_i\right)\rightarrow\left(t_l\mid s_i\right)\right) \qquad (1)$$

Cross-graining performs poorly where fast or highly cooperative transitions exist on both $\{x_i\}$ and $\{y_i\}$ (Fig. 2c); this can in principle be solved by using a shorter Markov timescale τ, but this comes at the cost of a greatly increased sampling requirement and thus can become prohibitively expensive.

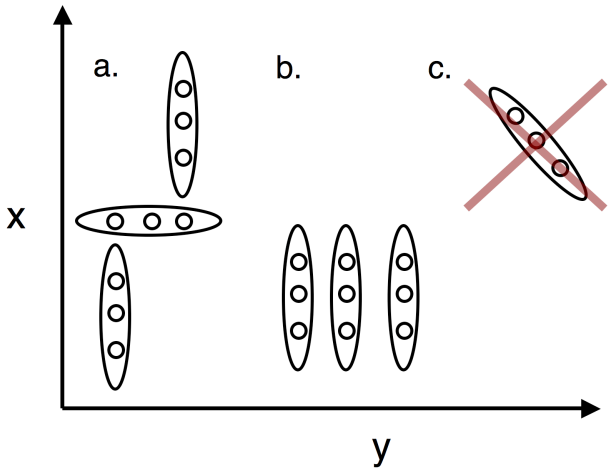

Figure 2. Two-dimensional free energy diagram. This schematic contour plot shows state configurations well suited (a and b) and poorly suited (c) for cross-graining.

Let us leave the problem of state selection aside for the moment and assume that we can select states, either using prior experimental knowledge or by applying a clustering algorithm to an initial sampling of state space [6, 7]. To specify the matrix of transition probabilities $P(s_i \rightarrow s_j \mid t \in T, \tau_1)$ we can sample trajectories in $\{x_i\}$ using a sampling algorithm of our choice while holding t constant and repeating this for each t in T. We perform the same operation on $\{y_i\}$ while holding s constant for each state s in S. This sampling (where one might use molecular

dynamics simulation in a molecular system) illustrates the "cross-grained" nature of our approach—when sampling on $\{x_i\}$, the information encoded in the state variable t provides sufficient information regarding $\{y_i\}$, the other degrees of freedom are effectively coarse-grained away. The converse applies to sampling $\{y_i\}$. By treating each subspace separately in this manner, we can substantially ease the task of sampling phase space.

To use the example of protein insertion and assembly with the MARTINI force field, the atomistic representation would include the protein and, depending on the coarse-grained state, the surrounding water or membrane. These surroundings encode the coarse-grained state information. The coarse-grained representation contains the full system and encodes the atomistic state by using the protein secondary structure to determine torsional potentials (as is standard in MARTINI) and optionally some additional conformational restrains (c.f. the work of Marrink, Sansom and co-workers [17, 21]).

Given the transition probabilities $P(s_i \rightarrow s_j \mid t, \tau_1)$, $P(t_i \rightarrow t_j \mid s, \tau_2)$, we can then construct a single first-order Markov state model on the joint space (s_i,t_j). If we formulate the transition probabilities on $\tau_3 >= \max(\tau_1, \tau_2)$, $P[(s_i,t_j)\rightarrow(s_k,t_l) \mid \tau_3] = P(s_i \rightarrow s_k \mid t_j, \tau_3) P(t_j \rightarrow t_l \mid s_i, \tau_3)$. This approximation of conditional independence is a key one; transitions that are highly cooperative between coarse-grained and fine-grained degrees of freedom will result in a poor approximation here. This can be partially addressed with a finer sampling of state space, but at a cost of increased sampling. Thus, highly cooperative transitions are an area for further methods development. For most cases, however, this approximation provides an MSM describing the full state of the system, which we can query for equilibrium probabilities, kinetic properties, ensemble experimental observables, etc. much as any single-resolution Markov state model [1, 5, 16].

3. A simple model for the insertion and assembly of a membrane pore

Let us consider a simple system for which phase space can be partitioned into degrees of freedom $\{x_i\}$ and $\{y_i\}$ and that has the following transitions:

$$\begin{aligned} \text{on } \{x_i\}: \quad & A \underset{k_4}{\overset{k_1}{\rightleftharpoons}} I_1 \underset{k_5}{\overset{k_2}{\rightleftharpoons}} I_2 \underset{k_6}{\overset{k_3}{\rightleftharpoons}} B \\ \text{on } \{y_i\}: \quad & C \underset{k_{10}}{\overset{k_7}{\rightleftharpoons}} I_3 \underset{k_{11}}{\overset{k_8}{\rightleftharpoons}} I_4 \underset{k_{12}}{\overset{k_9}{\rightleftharpoons}} D \end{aligned} \qquad (2)$$

and where the states $\{s_1..s_4\}$ = {A,I_1,I_2,B} and $\{t_1..t_4\}$={C,I_3,I_4,D} are Markovian given the joint state of the system (s_i,t_j). These states and the rates given below were chosen to loosely resemble a simple model for the insertion and assembly of a membrane pore, similar to the rough model proposed for the assembly of hemolysin E [22]. The fine-grained state A corresponds to the solution protein conformation, I1 to a hydrophobic-exposed protein conformation in contact with the membrane, I2 to a membrane-inserted and refolded conformation, and B to the protein conformation in the assembled pore. The coarse-grained state C corresponds to the protein distant from the membrane, I3 to the protein at the membrane interface, I4 to the protein inserted in the membrane, and D to the fully assembled pore. A Markov model in full molecular detail for this process would likely have many more states, but this provides an experimentally-motived simple model for our proof of principle. If we have knowledge of the macrostate definitions i.e. the mappings X->S and Y->T but not the transition rates $\{k\}$, we can set up the following sampling approach to estimate the rates:

1. Start n simulation trajectories on $\{x_i\}$ for each macrostate t on $\{y_i\}$; these can be from a single start state (e.g. A) or distributed among start states from a random seeding approach or a more sophisticated sampling scheme such as that described in [23]. Similarly start m simulation trajectories on $\{y_i\}$ for each macrostate s on $\{x_i\}$. Using the state mappings given, this will yield a set of trajectories each of the form $(u_1..u_n \mid t_i)$ where the macrostate ti remains fixed and $(v_1..v_n \mid s_i)$ where s_i remains fixed. Macrostate mappings can also be iteratively refined at this stage.
2. Estimate the transition probabilities for each of these Markovian sub-graphs:

$$P\left(s_i \rightarrow s_j \mid t_k\right) = \frac{Count\left(u_n \in s_i, u_{n+1} \in s_j \mid t_k\right)}{Count\left(u_n \in s_i \mid t_k\right)} \quad (3)$$

3. Construct the transition probabilities for the combined Markovian state model $P[(s_i,t_j)\text{->}(s_k,t_l) \mid \tau_3] = P(s_i\text{->}s_k \mid t_j, \tau_3)P(t_j\text{->}t_l \mid si, \tau_3)$.

We demonstrate this approach by sampling n random trajectories on $\{x_i\}$ for each macrostate $t \in T$ using a transition probability matrix constructed from the rates $\{k\}$ listed in Table 1 for the reaction scheme given above. We similarly take n random trajectories on $\{y_i\}$ for each macrostate $s \in S$. Starting states are randomly selected from a uniform distribution over macrostates. We construct the transition probabilities for the combined Markovian state model as specified above using our "observed" trajectories and compare to the same number of $8n$ total trajectories sampled directly from the combined transition probability matrix. We let n vary between 20 and 100, perform 100 random sampling procedures of this nature for each value of n and compare the convergence of the transition probability matrix eigenvalues as performed previously [23]. Transition probability eigenvalues yield the implied timescales for the system and are hence an important target for validation. Results are plotted in Figure 3.

Table 1. Rate constants used for simple system. First-order rate constants are given in reduced units of time τ.

		Macrostate on $\{y_i\}$			
		C (solution)	I3 (interfacial)	I4 (inserted)	D (pore)
Rate	k_1	1.00E-09	1.00E-05	1.00E-05	1.00E-05
	k_2	1.00E-09	1.00E-05	1.00E-05	1.00E-05
	k_3	1	1.00E-07	1.00E-05	1.00E-05
	k_4	1	1.00E+00	1.00E-05	1.00E-09
	k_5	1	1.00E-05	1.00E-07	1.00E-09
	k_6	1	1.00E-05	1.00E-09	1.00E-09

		Macrostate on $\{x_i\}$			
		A (solution conformation)	I1 (hydrophobic exposure)	I2 (partially refolded)	B (membrane fold)
Rate	k_7	1.00E-05	1.00E-03	1.00E-03	1.00E-03
	k_8	1.00E-09	1.00E-05	1.00E-03	1.00E-03
	k_9	1.00E-09	1.00E-09	1.00E-05	1.00E-05
	k_{10}	1	1.00E-05	1.00E-05	1.00E-05
	k_{11}	1	1.00E-05	1.00E-07	1.00E-09
	k_{12}	1.00E-05	1.00E-09	1.00E-09	1.00E-09

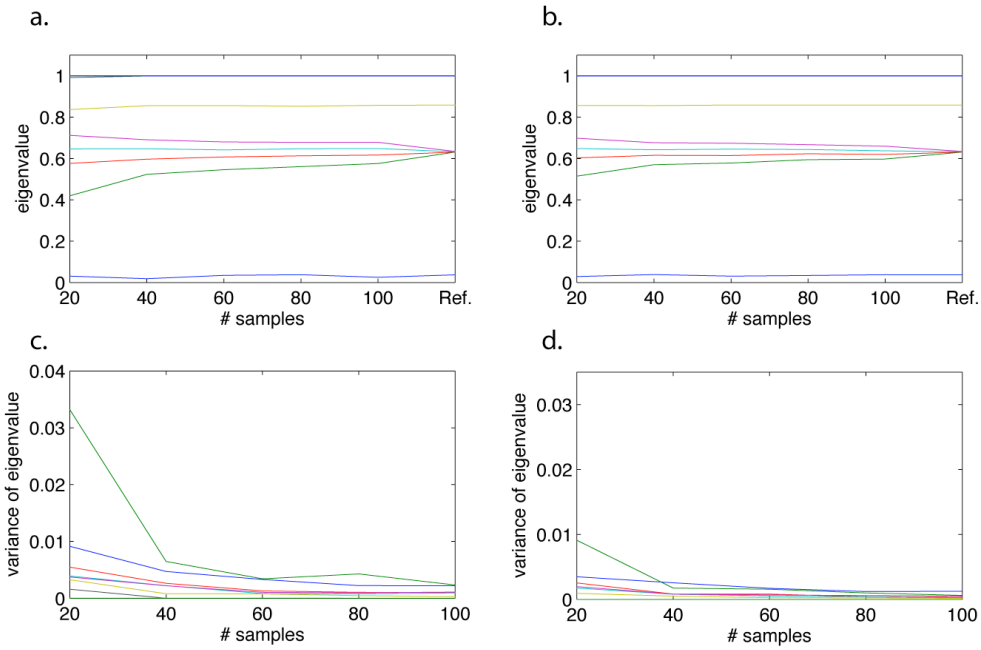

Figure 3. Convergence of eigenvalues for cross-graining. Convergence is compared between cross-grained sampling and direct sampling on the combined transition probability matrix. Panels (a) and (b) show convergence of the mean of each eigenvalue across 100 random samples, while panels (c) and (d) show convergence of the variance.

Convergence of the relaxation timescales yielded by sampling with our "cross-graining" strategy are compared to convergence with an "ideal" multi-scale simulation that can seamlessly mix coarse-grained and atomistic degrees of freedom (mixed CG/AA simulation). Both systems show good convergence over the range 40-100 trajectories of length 1000 τ in each of 8 start state combinations. The clear advantage for cross-graining becomes evident, however, when one considers the compute time required.

We examine computational efficiency in two respects, first with regard to the efficiency of the individual simulations and with regard to the sampling techniques employed. In each of these, cross-graining performs at least as well as the optimal mixed CG/AA simulation. Within the individual simulations, cross-graining allows the coarse-grained and atomistic simulations to be effectively decoupled and run in parallel. This results in optimal scaling across processors and performance equivalent to the hypothetical perfect multi-scale integrator, in which force calculations and coordinate updates can be performed at different timesteps for different parts of the system with no overhead.

Cross-graining also performs at least as well as optimal mixed CG/AA simulation with regard to sampling efficiency. If we have a naive sampling approach where we uniformly sample transitions in the macrostate space (s_i, t_j), cross-graining and mixed CG/AA simulation are equivalent. This equivalence also assumes that for every (s_i, t_j) start state we require even sampling of the s_i -> s_k transitions and the t_j -> t_l transitions. We have taken such an even-sampling approach in the example above, but recent work has shown a substantial advantage for "adaptive sampling" methods that weight the distribution of starting states according to their contribution to the variance of transition matrix eigenvalues [23]. In the case of villin headpiece folding, such a strategy improved convergence by a factor of over 1000. For a given (s_i,t_j) pair, unless we require precisely the same degree of sampling from $s_i \mid t_j$ and $t_j \mid s_i$, we "waste" the sampling across the other degrees of freedom. Since multi-scale simulation typically targets systems where coarse-grained degrees of freedom relate to long-timescale phenomena and atomistic degrees of freedom relate to shorter-timescale phenomena, we end up performing atomistic simulations over the coarse-grained timescales. Using the hemolysin example discussed above, the protein alone comprises 100,000 atoms; if we consider the solution case, simulate water molecules within 1 nm of the protein, and take τ = 1ns, then every

unnecessary trajectory of the resulting 375,000 atom system sampled at length 1000τ would take several months of simulation time using 128 Intel Cloverton cores (benchmarked with Gromacs 4.0 [24]).

4. Outlook

The "cross-graining" approach that we describe is particularly amenable to problems such as insertion, refolding, and assembly of membrane protein oligomers. Membranes and membrane proteins have long been a target for coarse-grained simulation [17, 25, 26], and assembly of folded protein subunits has been studied by several researchers, as has the insertion of pre-formed helices. However, the process of insertion and folding of peptides that have different conformations in solution and in the membrane is more challenging for coarse-grained simulation [27]. Even more difficult are large systems such as E. coli hemolysin (Figure 4) that are are thought to undergo partial unfolding and refolding at the membrane interface prior to assembly in the membrane [22]. One approach to cross-graining such a system is to coarse-grain the hemolysin-membrane-water system with states representing protein in solution, protein at the membrane interface, inserted protein, and various intermediates in assembly. The hemolysin monomer and its immediate surroundings (solution or membrane) are represented atomistically. Sampling the atomistic system will yield Markov state models for the conformational dynamics of the protein in each of the coarse-grained states—data that are difficult to compute accurately in the coarse-grained system alone. Sampling the coarse-grained system will yield the long-timescale behavior of protein insertion and pore formation, data that are not computationally tractable via atomistic systems. Experimental data from crystal structures and biochemical studies can guide the initial state definitions; this approach also provides a means to computationally evaluate structural models for refolding, insertion, and assembly such as those recently proposed for hemolysin [22].

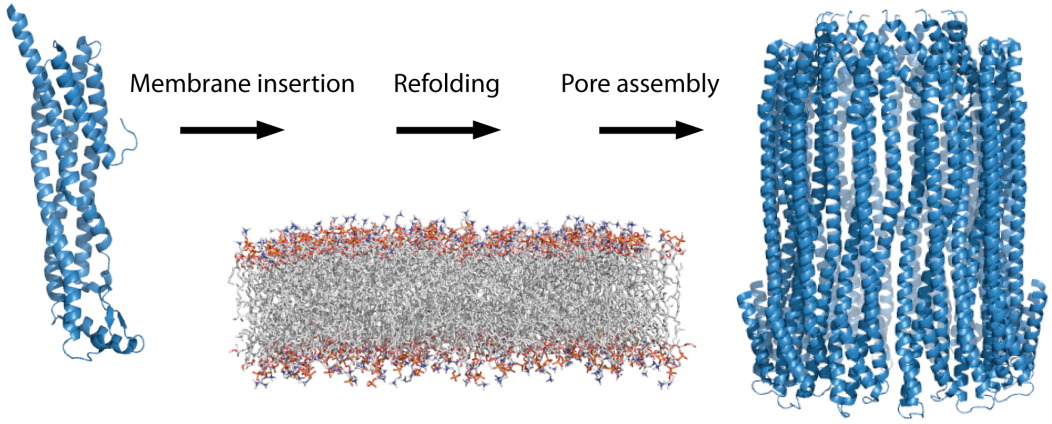

Figure 4. Solution-state and assembled conformations of E. coli hemolysin E.

5. Conclusions

We have presented a means to construct unified multi-scale kinetic models by performing individual simulations at a single scale and combining the simulations and scales in a Markov state model. This is accomplished by partitioning a system into two sets of degrees of freedom that are loosely coupled. This loose coupling and the specification of a discrete number of Markovian "macrostates" allows the approximation of any given transition $P[(s_i,t_j)\text{->}(s_k,t_l)]$ as the product of two orthogonal transitions $P(s_i\text{->}s_k \mid t_j)$ and $P(t_j\text{->}t_l \mid s_j)$. We outline how this method applies to membrane protein simulations, but we anticipate it will prove a much more general method for performing multi-scale simulation, helping to overcome the challenges of simultaneous mixed-scale simulation.

Acknowledgments

PK was supported by a fellowship from the Berry Foundation. The authors would like to thank S. Bacallado, G. Bowman, and D. Ensign for many helpful discussions.

References

1. Swope, W.C., J.W. Pitera, and F. Suits, *Describing protein folding kinetics by molecular dynamics simulations. 1. Theory.* Journal of Physical Chemistry B, 2004. **108**(21): p. 6571-6581.
2. Buchete, N.V. and G. Hummer, *Peptide folding kinetics from replica exchange molecular dynamics.* Phys Rev E Stat Nonlin Soft Matter Phys, 2008. **77**(3 Pt 1): p. 030902.
3. Sriraman, S., I.G. Kevrekidis, and G. Hummer, *Coarse master equation from Bayesian analysis of replica molecular dynamics simulations.* J Phys Chem B, 2005. **109**(14): p. 6479-84.
4. Noe, F. and S. Fischer, *Transition networks for modeling the kinetics of conformational change in macromolecules.* Curr Opin Struct Biol, 2008. **18**(2): p. 154-62.
5. Singhal, N., C.D. Snow, and V.S. Pande, *Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin.* Journal of Chemical Physics, 2004. **121**(1): p. 415-425.
6. Bowman, G.R., X. Huang, and V.S. Pande, *Using generalized ensemble simulations and Markov state models to identify conformational states.* Methods, 2009.
7. Chodera, J.D., et al., *Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics.* J Chem Phys, 2007. **126**(15): p. 155101.
8. Sansom, M.S., K.A. Scott, and P.J. Bond, *Coarse-grained simulation: a high-throughput computational approach to membrane proteins.* Biochem Soc Trans, 2008. **36**(Pt 1): p. 27-32.
9. Marrink, S.J., et al., *The MARTINI force field: coarse grained model for biomolecular simulations.* J Phys Chem B, 2007. **111**(27): p. 7812-24.
10. Marrink, S.J. and A.E. Mark, *Effect of undulations on surface tension in simulated bilayers.* Journal of Physical Chemistry B, 2001. **105**(26): p. 6122-6127.
11. Faller, R. and S.J. Marrink, *Simulation of domain formation in DLPC-DSPC mixed bilayers.* Langmuir, 2004. **20**(18): p. 7686-93.
12. Marrink, S.J., J. Risselada, and A.E. Mark, *Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model.* Chemistry and Physics of Lipids, 2005.
13. Izvekov, S. and G.A. Voth, *Multiscale coarse graining of liquid-state systems.* J Chem Phys, 2005. **123**(13): p. 134105.
14. Shi, Q. and G.A. Voth, *Multi-scale modeling of phase separation in mixed lipid bilayers.* Biophys J, 2005. **89**(4): p. 2385-94.
15. Ayton, G.S. and G.A. Voth, *Mesoscopic lateral diffusion in lipid bilayers.* Biophys J, 2004. **87**(5): p. 3299-311.
16. Kasson, P.M., et al., *Ensemble molecular dynamics yields sub-millisecond kinetics and intermediates of membrane fusion.* . Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11916-21.
17. Bond, P.J. and M.S. Sansom, *Insertion and assembly of membrane proteins via simulation.* J Am Chem Soc, 2006. **128**(8): p. 2697-704.
18. Carpenter, T., et al., *Self-assembly of a simple membrane protein: coarse-grained molecular dynamics simulations of the influenza M2 channel.* Biophys J, 2008. **95**(8): p. 3790-801.
19. Ayton, G.S., P.D. Blood, and G.A. Voth, *Membrane remodeling from N-BAR domain interactions: insights from multi-scale simulation.* Biophys J, 2007. **92**(10): p. 3595-602.
20. Park, S. and V.S. Pande, *Validation of Markov state models using Shannon's entropy.* J Chem Phys, 2006. **124**(5): p. 54118.
21. Yefimov, S., et al., *Mechanosensitive membrane channels in action.* Biophys J, 2008. **94**(8): p. 2994-3002.
22. Mueller, M., et al., *The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism.* Nature, 2009. **459**(7247): p. 726-30.
23. Hinrichs, N.S. and V.S. Pande, *Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics.* The Journal of chemical physics, 2007. **126**(24): p. 244101.
24. Hess, B., et al., *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation.* Journal of Chemical Theory and Computation, 2008. **4**: p. 435-447.
25. Shih, A.Y., et al., *Coarse grained protein-lipid model with application to lipoprotein particles.* J Phys Chem B, 2006. **110**(8): p. 3674-84.

26. Marrink, S.J., A.H. de Vries, and A.E. Mark, *Coarse grained model for semiquantitative lipid simulations.* Journal of Physical Chemistry B, 2004. **108**(2): p. 750-760.
27. Marrink, S.J., A.H. de Vries, and D.P. Tieleman, *Lipids on the move: simulations of membrane pores, domains, stalks and curves.* Biochim Biophys Acta, 2009. **1788**(1): p. 149-68.