John von Neumann Institute for Computing

A Network-Based Approach to Biomolecular Dynamics

F. Noé, J. C. Smith, Ch. Schütte

published in

From Computational Biophysics to Systems Biology (CBSB07),
Proceedings of the NIC Workshop 2007,
Ulrich H. E. Hansmann, Jan Meinke, Sandipan Mohanty,
Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **36**, ISBN 978-3-9810843-2-0, pp. 247-250, 2007.

http://www.fz-juelich.de/nic-series/volume36

A Network-Based Approach to Biomolecular Dynamics

Frank Noé[1]**, Jeremy C. Smith**[2]**, and Christof Schütte**[1]

[1] DFG Research Center Matheon,
FU Berlin, Arnimallee 6, 14195 Berlin, Germany
E-mail: {noe, schuette}@math.fu-berlin.de

[2] Interdisciplinary Center for Scientific Computing,
University of Heidelberg, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany
E-mail: biocomputing@iwr.uni-heidelberg.de

Molecular dynamics simulation generates large quantities of data that must be interpreted using physically meaningful analysis. A common approach is to describe the system dynamics in terms of transitions between coarse partitions of conformational space. In contrast to previous work that partitions the space according to geometric proximity, we examine here clustering based on kinetics, merging configurational microstates together so as to identify long-lived, *i.e.* dynamically metastable, states. The method is applied to microsecond molecular dynamics simulations of Ala_{12}. The system clearly exhibit metastability, with some kinetically distinct metastable states being geometrically very similar.

1 Introduction

Given the large amount of molecular dynamics (MD) simulation data available for macromolecules, models are needed to analyze the data that capture the essential dynamical features and can be physically interpreted. Here we use transition networks which encode the transitions of the system between a set of discrete states. In a transition network, a vertex corresponds to a discrete conformational state, and an edge corresponds to a transition between two states. Each edge may be weighted by the associated transition rate or probability.

A central question when generating discrete state models is how best to partition state space into discrete states. It is often desirable that the partition reflects the dynamical behavior of the system. In particular, biomolecular function often depends on the ability to undergo transitions between long-lived, or "metastable" states. These inter-state transitions are rare events in which a molecular system stays for long periods of time within one state, before rapidly switching to another.

Geometric clustering approaches do generally not fulfill these claims. Therefore, here we use the Perron Cluster Cluster Analysis (PCCA)[2,3], a dynamical data-based kinetic clustering method to partition the state space. The method is applied to a 4-microsecond molecular dynamics simulation of a 12-mer polyalanine. Using different numbers of metastable states, a hierarchical picture for the kinetics is provided.

2 Kinetic Clustering Method

The state space of the system is first partitioned into fine discrete sets, called "micro-states" (typically on the order of 10^3 to 10^4). This may be done by characterizing the torsion

rotamer state or by using a geometric clustering method with a large number of clusters on the Cartesian coordinates of the molecule with rotation and translation removed. See Ref.[4] for details.

A metastable state is a set of micro-states in which the system stays for a long time before leaving it. Following this idea, a straightforward definition for a partition into metastable states would be the following: Given a discrete trajectory $[i(0), i(\tau), i(2\tau), ...]$ in space of micro-states $\{1, ..., m\}$, find a partition of micro-states into C metastable sets, such that the number of transitions within sets is maximized. This formulation of the problem is known in graph theory as a C-min-cut problem but is, unfortunately, NP hard. A similar partition, however, can be obtained very efficiently with the Perron cluster cluster analysis (PCCA) method which was first introduced in[1]. Its basic idea is the following: The Markov time evolution of the probability distribution of states, $\mathbf{p}(t)$, is given by $\mathbf{p}(t+\tau) = \mathbf{p}(t)\mathbf{T}(\tau)$. This equation can also be expressed in terms of an expansion in left eigenvectors of $\mathbf{T}$, $\mathbf{q}_i$, and corresponding eigenvalues, λ_i, and for longer time increments, $n\tau$:

$$\mathbf{p}(t + n\tau) = \sum_{i=1..m} c_i \lambda_i^n \mathbf{q}_i. \quad (1)$$

For systems fulfilling detailed balance there is only a single eigenvalue $\lambda_1 = 1$, while $\lambda_i < 1$ for all $i > 1$. Thus, for $n \to \infty$ (after infinity time), all terms but the first one vanish and Eq. (1) converges to: $\mathbf{p}(\infty) = c_1\mathbf{q}_1$, where c_1 is simply a normalization factor. Thus, the left eigenvector with the largest eigenvalue, $\lambda_1 = 1$, specifies the stationary distribution (and has therefore only nonnegative entries). All other eigenvectors, $\mathbf{q}_2...\mathbf{q}_m$, are associated with decaying processes, their speed of decay being determined by the magnitudes of the associated eigenvalues. As apparent from Eq. (1), eigenvectors associated with eigenvalues close to 1 correspond to processes which decay very slowly and thus are, by definition, related to transitions between metastable states.

The improved PCCA method[2,3] used here identifies C representative micro-states with maximum pairwise distances in the coordinates of the first C eigenvectors. Each microstate can then be expressed as a convex combination of these representative micro-states and thus be assigned a degree of similarity with each representative micro-state according to the convex coordinates. Therefore, clustering is made by assigning each microstate to the cluster containing the representant with which it has the strongest similarity.

3 Results

A complex network of metastable states is identified. Use of only a few metastable states (*e.g.* 2, 3, ...) leads to the distinction between the "main state" (the basin with the globally-minimal energy and several "kinetic traps" which are thermally accessible and have half-life times similar to, or greater than, that of the main state. By increasing the number of metastable states the main state is further decomposed into conformational subsets. An interesting feature of the networks is that some states which are structurally very similar are not kinetically contiguous, but rather interchange *via* structurally very different intermediates. This clearly demonstrates the usefulness of the kinetic clustering method in biomolecular simulation studies.

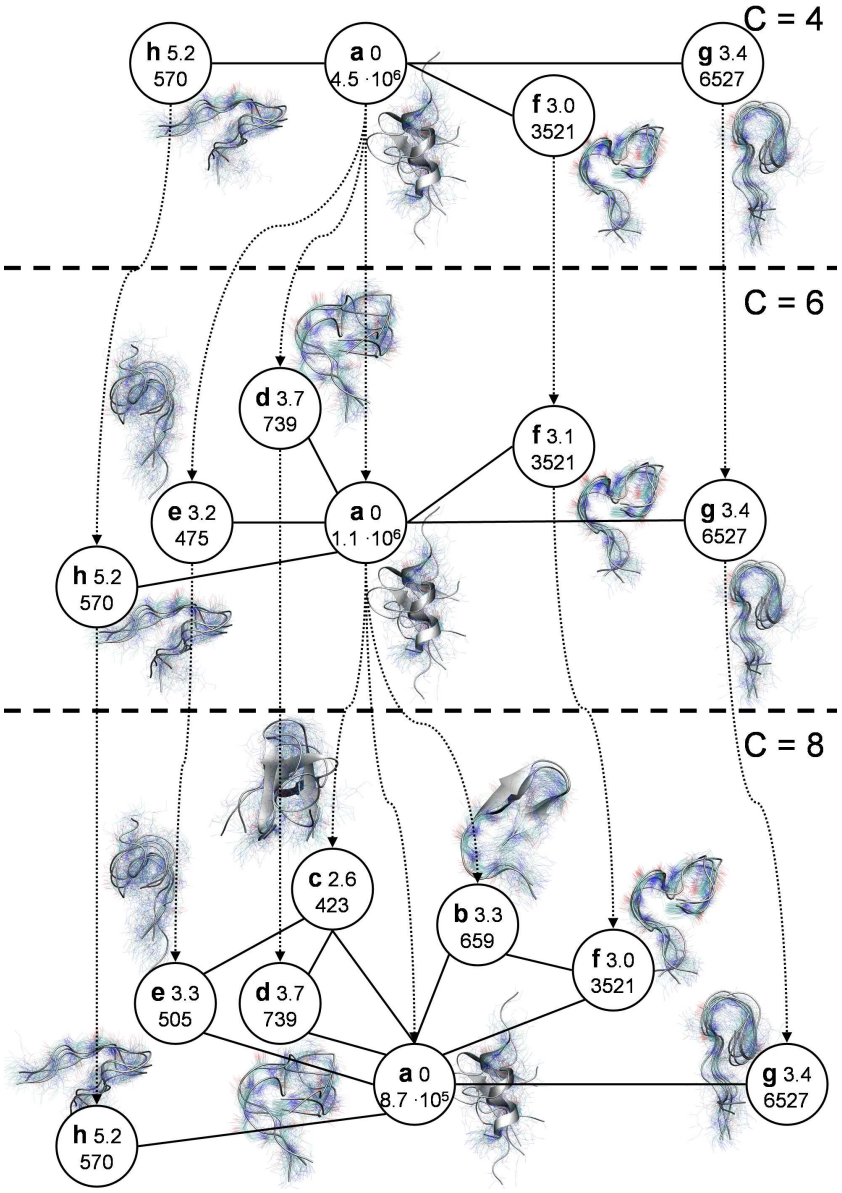

Figure 1. Hierarchical Transition Network analysis for the Ala_{12} peptide for 4, 6 and 8 metastable sets. Each bullet and the structure next to it corresponds to one metastable set. The bullets contain the state name (a letter), the free energy in kcal/mol (upper number) and the mean lifetime in picoseconds (lower number). Each structure is shown by a few representative tubes and an overlay of 100 examples randomly drawn from the ensemble of structures of each state, shown as line representations of the backbone. A pair of states is connected if at least one transition between these states was observed in the trajectory. The hierarchical relationship between the three networks is indicated by the dotted arrows. Each arrow starts at the metastable state in the higher-order network which contains the majority of micro-states in the state the arrow points to. For example, the micro-states of state a in the $C = 4$ network are split into three sub-states, a, d and e, in the $C = 6$ network.

References

1. C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, "A direct approach to conformational dynamics based on hybrid monte carlo," *J. Comp. Phys.*, vol. 151, pp. 146–168, 1999.
2. M. Weber, "Improved perron cluster analysis," *ZIB Report*, vol. 03-04, 2003.
3. P. Deuflhard and M. Weber, "Robust perron cluster analysis in conformation dynamics," *ZIB Report*, vol. 03-09, 2003.
4. F. Noé, I. Horenko, C. Schütte, J.C. Smith, "Hierarchical Analysis of Conformational Dynamics in Biomolecules: Transition Networks of Metastable States" *J. Chem. Phys.*, vol. 126, p. 155102, 2007.