Statistical Biophysics Blog: Statistical mechanics in biology and biocomputation

FAQ on Trajectory Ensembles

Q: What is a trajectory?
A trajectory is the time-ordered sequence of system configurations which occur as all the coordinates evolve in time following some rules - hopefully rules embodying reasonable physical dynamics, such as Newton's laws or constant-temperature molecular dynamics.

Q: What is a trajectory ensemble?
It's a set of *independent* trajectories that *together* characterize a particular condition such as equilibrium or a non-equilibrium steady state. That is, the trajectories do not interact in any way, but statistically they describe some condition because of how they have been initiated - and when they are observed relative to their initialization … see below.

Q: What are the types of trajectory ensemble?
There are three types of trajectory ensembles. Assume each to consist of a very large number of trajectories. In the figures, individual trajectories are shown in blue and average behavior is shown schematically with solid lines.
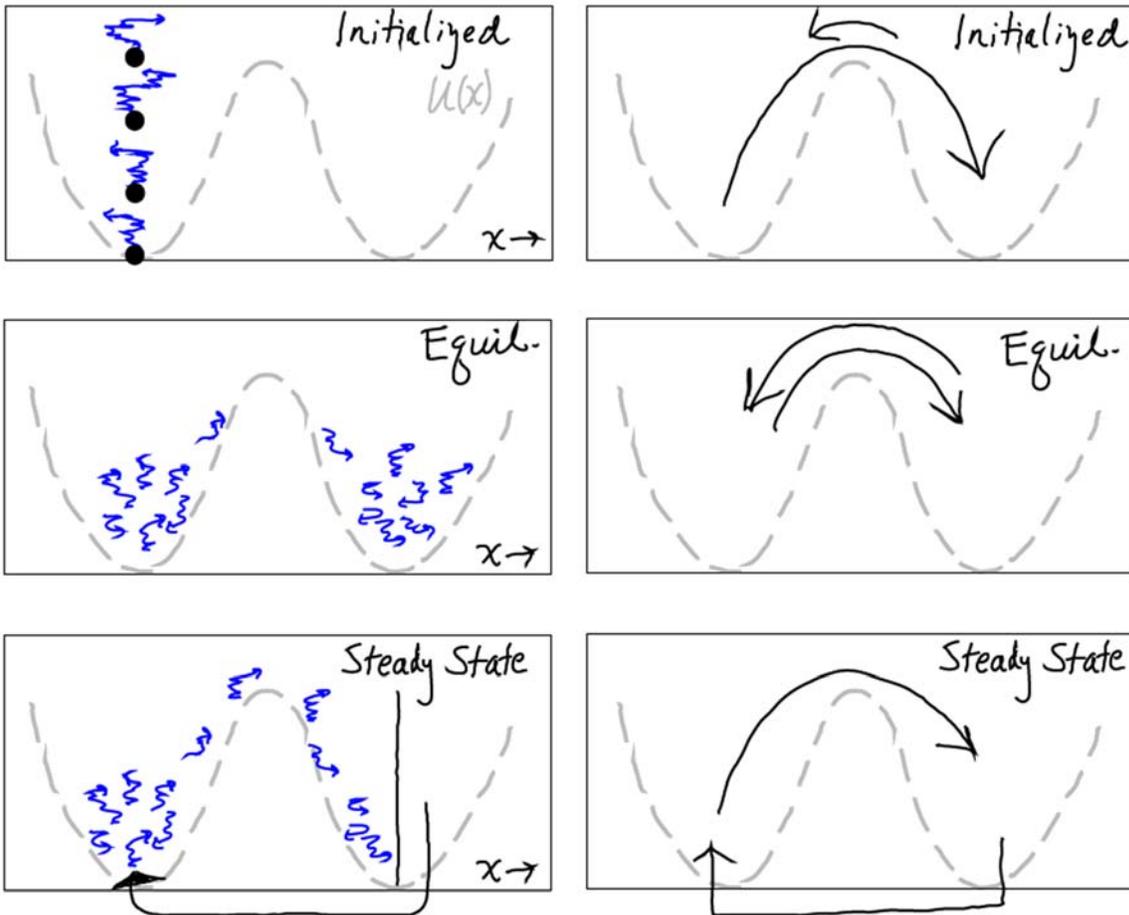
1. <u>Equilibrium</u>. This is a set of trajectories that obeys detailed balance, which in turn implies what might be called "arbitrary balance": for any two arbitrary regions of configuration or phase space, the number of trajectories flowing from one to the other in a fixed time interval will exactly match the reverse flow. The equilibrium trajectory ensemble yields the equilibrium distribution (Boltzmann factor distribution of configurations for fixed temperature) when trajectories are observed at any fixed time. That is, taking a set of configurations x - one from each trajectory - at any time point t will yield the Boltzmann-factor distribution of *configurations*. We can write $p(x; t) = p^{eq}(x)$, independent of t. Important properties of the equilibrium ensemble are discussed in <u>another post</u>. *Making the ensemble.* In principle, one can imagine generating an equilibrium trajectory ensemble in a couple of ways … by somehow knowing $p^{eq}(x)$ in advance, generating configurations x accordingly, and assigning each coordinate a velocity chosen from the Maxwell-Boltzmann equilibrium velocity distribution … alternatively, *any* initial distribution will ultimately relax to equilibrium after enough time so long as there are no external sources or sinks; that is, $p(x; t > t^{eq}) = p^{eq}(x)$ where t^{eq} is the time required for the system to relax to equilibrium. See the figure below.

2. <u>Non-equilibrium steady state</u>. A non-equilibrium steady-state is a trajectory ensemble in which the distribution of configurations is unchanging in time, but which does *not* obey the detailed balance condition. We can write $p(x; t) = p^{ss}(x) \neq p^{eq}(x)$. Hence, there are

net flows occurring in the phase space ... from some source of trajectories (where they are initiated) to a sink (where they are absorbed). The steady state ensemble is critically important because it enables estimation of the rate or first-passage time. *Making the ensemble.* As in the case of the equilibrium distribution, an *arbitrary* initial distribution will relax to a non-equilibrium steady state so long as trajectories encountering sink points are re-initiated at the source points. Thus, $p(x; t > t^{ss}) = p^{ss}(x)$ where t^{ss} is the time required for the system to relax to the steady state. Both the source and sink can be rather arbitrary and not simply single configurations. For example, in the system's configuration space, trajectories could be initiated according to an arbitrary distribution on an arbitrary subset of configurations (e.g., 80% in one region and 20% in another) - and be absorbed on another arbitrary subset.

3. <u>Initialized</u>. The initialized trajectory ensemble evolves in time, having started from some specified distribution of configurations (possibly with velocities). This ensemble has not yet relaxed into equilibrium or a steady state. Thus, by definition, it is still in its transient phase, so the distribution of configurations $p(x; t)$, obtained by taking all configurations at time time t, changes in time. See left panels in the figure below. *Making the ensemble.* Such an ensemble can be constructed from an arbitrary initial distribution of configurations - just by letting each evolve in time according to the natural system dynamics. The simplest example would be starting many trajectories from the same configuration (with different velocities and/or random seeds) which corresponds to the time evolution of an initial delta function distribution. However, any initial distribution can be used.
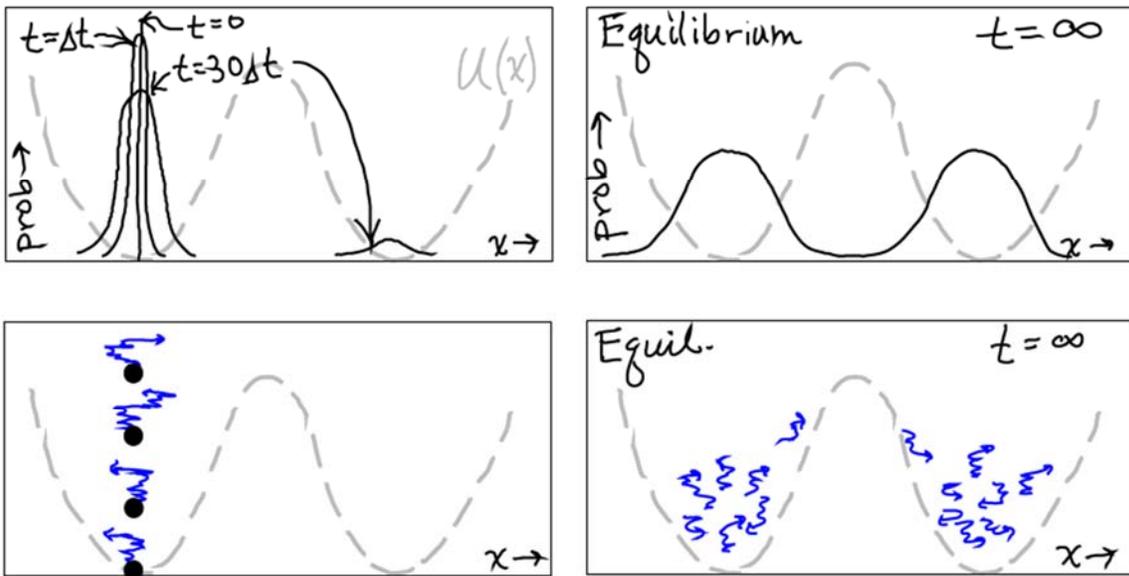
Q: What is an "equilibrium trajectory"?

Because equilibrium is an ensemble property that characterizes a set of configurations or trajectories, it is somewhat oxymoronic to describe a single trajectory as equilibrium (or non-equilibrium) in nature. However, informally, the phrase "equilibrium trajectory" is used when a

trajectory is not subject to any artificial biasing forces; sometimes there is the implicit notion that the trajectory is visiting high-probability parts of configuration space. Of course, it is difficult to know whether a trajectory is typical or whether it has visited the important parts of configuration space without examining many trajectories or a very long trajectory. If a trajectory were extremely long - in the sense of exhibiting multiple transitions among all important states of a system - then it would be more reasonable to describe it as a representation of equilibrium. Unfortunately, for many of the most interesting biomolecular systems, this is currently impractical and is likely to remain so for years to come. One notable exception would appear to be the folding/unfolding trajectories generated by the DE Shaw group for small proteins - but note these were generated at elevated temperatures to enable transitions. On a related point, how would one know whether equilibrium-like sampling has been obtained in any given simulation? This question has been addressed by my group and others' using the notion of "effective sample size" - see references below.

Q: What's the value of a trajectory ensemble compared to a string-like optimal path?
Methods for finding optimal paths and statistical average "strings" have become increasingly popular, and they can be very valuable in principle for describing a dominant mechanism. It is useful to be aware of the pros and cons of such approaches as compared to the trajectory ensemble picture. At least in principle, the trajectory ensemble has several advantages: (i) Steady-state trajectory ensembles directly provide unbiased estimation of the rate via the Hill relation, whereas optimal-path approaches have to make inferences based on possibly strong assumptions regarding the connection between high-dimensional landscape and kinetics. I hope to comment on those assumptions in a future post. (ii) Although a single path or string provides a reassuringly concrete picture, a true description of mechanism must account for all the heterogeneity that may be present, such as alternative pathways and/or path fluctuations. The path ensemble automatically includes all the latter. (iii) Most single-path methods are subject to the well-known "trapping" phenomenon in which the calculation fails to visit a channel which is most important to the process at hand. Ensemble methods may also suffer from trapping, but some variants may be more robust. I hope to expand on this point in a future post.

References
- "Heterogeneous Path Ensembles for Conformational Transitions in Semiatomistic Models of Adenylate Kinase," Bhatt, D. & Zuckerman, D. M. Journal of Chemical Theory and Computation, 2010, 6, 3527-3539
- "Reaction path study of conformational transitions and helix formation in a tetrapeptide," Czerminski, R. & Elber, R., Proc. Nat. Acad. Sci., 1989, 86, 6963-6967
- "Finite Temperature String Method for the Study of Rare Events," Weinan E, Weiqing Ren, Eric Vanden-Eijnden, J. Phys. Chem. B, 2005, 109 (14), pp 6688–6693
- Lindorff-Larsen, K.; Piana, S.; Dror, R. O. & Shaw, D. E. "How fast-folding proteins fold," Science, 2011, 334, 517-520
- "On the structural convergence of biomolecular simulations by determination of the effective sample size," E. Lyman and D. M. Zuckerman, *J. Phys. Chem. B.* 111:12876-12882 (2007).

- *Statistical Physics of Biomolecules: An Introduction*, Zuckerman, D. M., CRC Press, 2010