Checking Markovity

Although the purpose of the work introduced in the manuscript was to illustrate a way to handle the Conformational Markov Network (CMN) in order to obtain the basins of attraction and characterize the Free Energy Landscape, the lector must be advise about the necessity of checking the markovian character of the model (network in our case).

With regard to the CMNs that appear in the manuscript:

(i) The particle in the *funnel like potential* is simulated using an overdamped dynamics. In this case, the continuous trajectory integrated is inherently Markovian [1, 2]. When we discretize the coordinate space to lump the trajectory into micro-states, the Markovity of the new description can be defied [3, 4]. Therefore, one has to care about the time step to describe the process since it must be larger than the longest equilibration time among the micro-states [5]. In ref. [6] three approaches are mentioned to evaluate the degree of Markovity of a stochastic model. Among these approaches, we have taken the criterion presented by Park and Pande in [7] based on Shannon's entropy. This method provides a unique magnitude not as sensitive to statistical and numerical noise as other methods. Another reason to make this choice is that a necessary and sufficient condition for Markovity is checked (while the observation of the eigenvalues of the transition matrix is not a sufficient condition). The analysis is based on the comparison of a first-order conditional entropy $H_\tau(X_n|X_{n-1})$ and the second-order one $H_\tau(X_n|X_{n-1}, X_{n-2})$ (see [7] for further details and notation). The magnitude R_τ quantifies, given X_{n-1}, what fraction of the information in X_n is the mutual information between X_n and X_{n-2}. Although the method is computationally expensive, our models are small enough to carry out this analysis. The Figure S1 reveals that with the lag time used in our analysis ($\tau = 1$) the memory effect (non-Markovity) is less than a 5%.

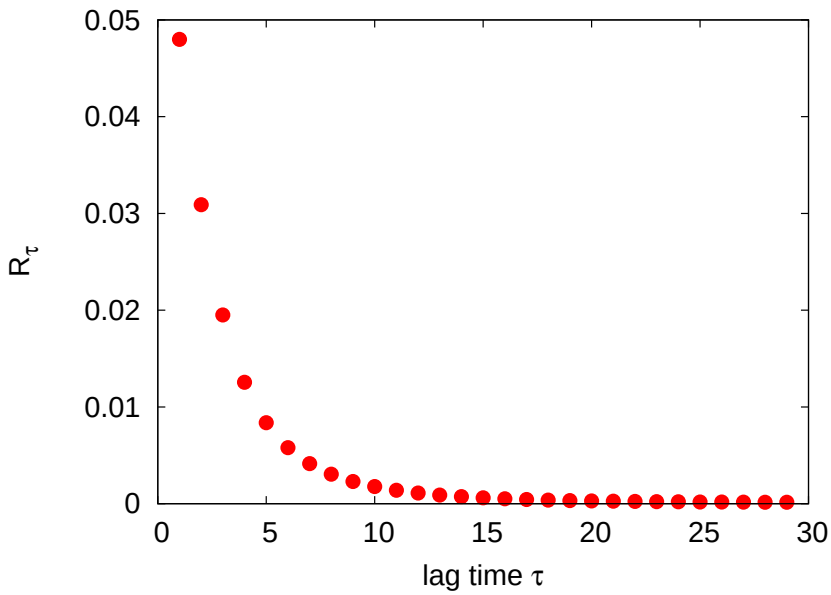

Figure S1. Checking Markovity: The relative mutual information R_τ quantifies the degree of non-Markovity of our stochastic model. The lag time used to construct the CMN -funnel like potential-, $\tau = 1$, reveals a memory effect lower than a 5%.

We have to remark, as it is discussed in ref. [7], that there's probably no answer to the question: when does the model behaves Markovian? However the appropriate question should be: Given an observable, what is the degree of Markovity needed to have a correct measure? Regarding to our work, note that the analysis of the basins detection depends on the detailed balance condition, but not on the relaxation times. On the other hand, a small deviation such as the 5%, only affects at the analysis of the temporal hierarchy of basins, where Chapman-Kolmogorov is used explicitly. In order to distinguish whether the error is propagated (and increased) or not when the transition matrix is raised to the power of n, we compute the Kullback-Leibler divergence between the transition matrix $S(\tau)^n$ and the matrix computed from the trajectory $S(n\tau)$:

$$D_{KL}(S(n\tau)||S(\tau)^n) = \sum_{ij} S(n\tau)_{ij} P_j \ln \frac{S(n\tau)_{ij}}{[S(\tau)^n]_{ij}}$$

The Figure S2 reveals that the error in the transitions computed with Chapman-Kolmogorov are far from the experimental values during a short range of $n\tau$ close to the original lag time, but the divergence decrease with time after few steps. In the limit $n \to \infty$, the matrix obtained must be equal to the experimental one (having the stationary probability distribution in each column of the transition matrix).

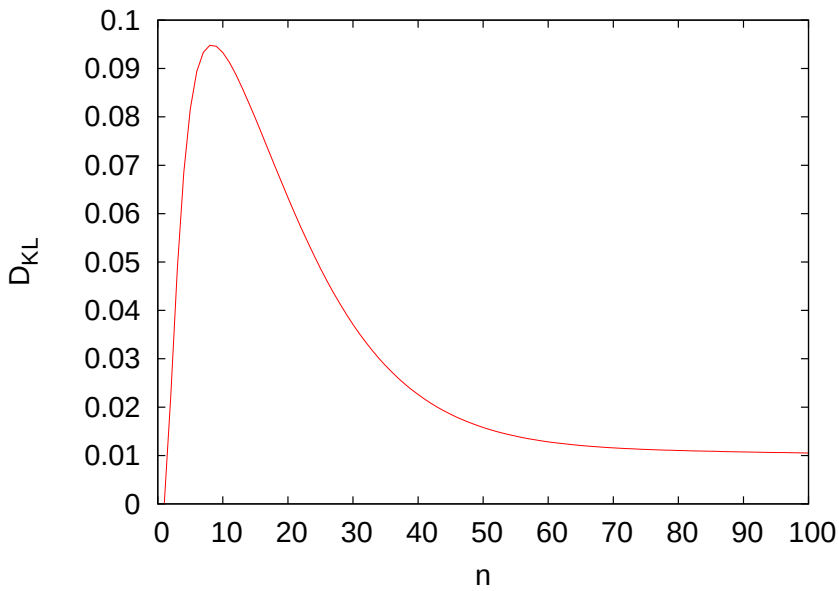

Figure S2. The Kullback-Leibler divergence between the "experimental" transition matrix $S(n\tau)$ and $S(\tau)^n$ decreases after few time steps (τ).

(ii) The *alanine dipeptide* is simulated with the Langevin formalism (see details in the manuscript and in the Text S3). The continuum trajectory in the space of coordinates and momenta is also inherently Markovian [1, 2]. On the other hand, we are integrating momenta and discretizing the coordinates space in our description, which can result in a non-Markovian chain depending on the time step used. In our case the same analysis applied before provides a value of 0.8% of memory effect.

References

1. van Kampen NG (1998) Remarks on non-Markov processes. Braz J Phys 28: 90–96.

2. Zwanzig R (2001) Nonequilibrium Statistical Mechanics. Oxford University Press, USA, 222 pp.

3. Huisinga W, Schütte C, Stuart AM (2003) Extracting macroscopic stochastic dynamics: Model problems. Commun Pure Appl Math 56: 234–269.

4. Jernigan RW, Baran RH (2003) Testing lumpability in Markov chains. Statist Probab Lett 64: 17–23.

5. Swope WC, Pitera JW, Suits F (2004) Describing protein folding kinetics by molecular dynamics simulations. 1. theory. J Phys Chem B 108: 6571–6581.

6. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC (2007) Automatic discovery of metastable states for the construction of markov models of macromolecular conformational dynamics. J Chem Phys 126: 155101-1–155101-17.

7. Park S, Pande VS (2006) Validation of Markov state models using Shannon's entropy. J Chem Phys 124: 054118-1–054118-5.