Resolving the apparent gap in complexity between simulated and measured kinetics of biomolecules

Bettina G. Keller, Jan-Hendrik Prinz, and Frank Noé

Freie Universität Berlin,
Arnimallee 6, 14195 Berlin, Germany
E-mail: bettina.keller@fu-berlin.de, jan-hendrik.prinz@fu-berlin.de, frank.noe@fu-berlin.de

Molecular simulations of biomolecules often reveal a complex picture of the their kinetics, whereas kinetic experiments typically seem to indicate considerably simpler two- or three-state kinetics. Markov state models (MSM) provide a tool to link between simulation and experiment, and to resolve this apparent contradiction.

1 Introduction

Molecular simulations of large biomolecules typically reveal a complex picture of the free-energy surface with many kinetically relevant states[1–3]. Complementary to this, advanced experimental techniques allow probing the equilibrium kinetics of biomolecules directly. This can either be done by perturbation techniques, such as temperature or pressure jump, or by equilibrium experiments in which the measured signal is autocorrelated to obtain kinetic information, as e.g. in fluorescence correlation spectroscopy. These measurements can be described by dynamical fingerprints, i.e. densities of relaxation timescales where each peak corresponds to an exponential relaxation process. In many cases, single- or double-peaked fingerprints are found, suggesting that a two- or three-state model may provide a satisfactory description of the biomolecule studied[4,5].

We sketch an approach combining Markov state models (MSM)[6,3] of the simulated dynamics with dynamical fingerprints, which allows addressing the following questions: (i) Is the largest relaxation timescales observed always due to the folding process? (ii) Can a given experiment detect all relaxation processes that are present in the dynamics of the molecule? (iii) Are the processes observed in perturbation experiments the same as those observed in equilibrium experiments? We illustrate our findings using a four-state model of a protein folding equilibrium.

2 Theory

In MSMs of conformational kinetics, the conformational space of the molecule is discretized into N states. The kinetics between the states is described by transition probability t_{ij} of going from a state i to a state j within a time step τ, which are summarized in a transition matrix $\mathbf{T}(\tau)$. The entire information of the kinetics of the system is contained in this matrix. In particular, each (left) eigenvector $\mathbf{l}_i$ of the transition matrix represents one of the kinetic processes of the system. For a detailed review of the theory of MSM and their application to the simulation of biomolecules see Ref. 6 and 3.

Kinetic experiments yield time series $\mu(t)$, either of the observed signal a directly (perturbation experiments), or of the autocorrelation of the signal (equilibrium experiments).

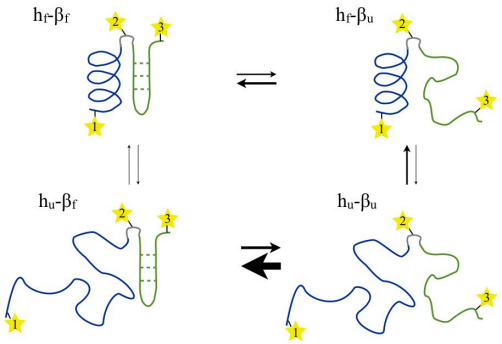

Figure 1. Sketch of a protein folding equilibrium. The arrows represent possible transitions between conformational states. Their thickness corresponds to the transition probability. The yellow stars represent possible chromophore attachment points.

From physical principles, we expect that in both cases $\mu(t)$ is a noisy realization of a sum of multiple exponential functions

$$\mu(t) = \int_{t'} dt' \gamma(t') \exp\left(-\frac{t}{t'}\right) \approx \sum_{i}^{N} \gamma_i \exp\left(-\frac{t}{t'_i}\right) . \tag{1}$$

$\gamma(t')$ is inverse the Laplace transform of $\mu(t)$ and is called the *dynamical fingerprint*[7]. The approximation in Eq. 1 results from the fact that any experimental signal is time-discretized, and that in an MSM-representation of the kinetics the number of processes which contribute to the dynamical fingerprint is limited to the number of states N. The relaxation timescales t_i of the system are linked to the eigenvalues λ_i of the transition matrix by $t'_i = -\tau / \ln \lambda_i$ One can derive analytical expressions[7,8] for the amplitudes γ_i of equilibrium experiments $\gamma_i^{\boldsymbol{\pi},a} = \langle \mathbf{a}, \mathbf{l}_i \rangle^2$ and of perturbation experiments $\gamma_i^{\mathbf{p}(0),a} = \langle \mathbf{a}, \mathbf{l}_i \rangle \langle \mathbf{p}'(0), \mathbf{l}_i \rangle$. $\mathbf{p}'(0)$ is the excess probability density with $p'_j(0) = p_j(0)/\pi_j$, where $\boldsymbol{\pi}$ is the equilibrium density. $\mathbf{a}$ is the observable vector which associates each states in the MSM with a signal strength of the observable a.

3 Model system and results

Each of the secondary structure elements of the protein folding model in Fig. 1 can fold and unfold in a single distinct step. This leads to a MSM of the conformational equilibrium with $N = 4$ metastable states. The transition probabilities are represented by the thickness of the arrows in Fig. 1. None of the eigenvectors of the MSM (Fig. 2.a) reflects our notion of folding. Hence, the folding process may not be contained in a kinetic process and can thus not always be associated with a single folding rate.

The yellow stars in Fig. 1 mark possible chromophore attachment points. We choose our observable vectors to resemble FRET constructs, i.e. two chromophores attached to the molecule with strong signal if they are close to each other, weak signal if they are further apart (Fig. 2.c).

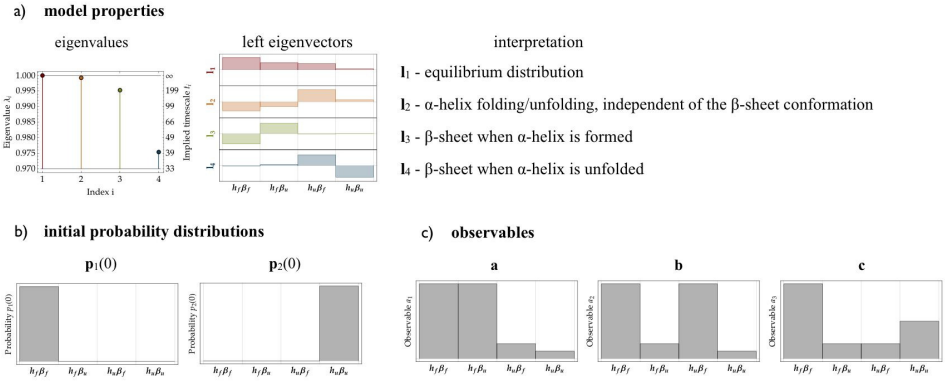

Figure 2. Markov model and experimental setup for the protein folding model.

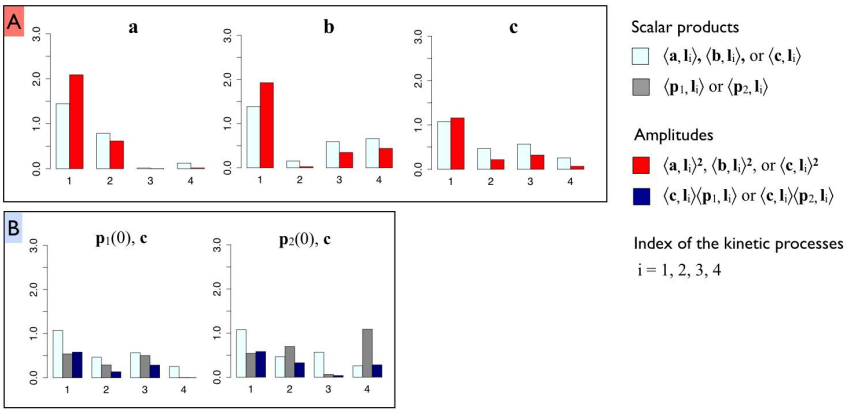

Figure 3. Amplitudes of the dynamical fingerprints of a variety of experimental setups for the protein folding model. (A) Equilibrium experiment, dynamical fingerprint of the autocorrelation function of observables $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$. (B) Perturbation experiment, dynamical fingerprint of the decay signal of the observable $\mathbf{c}$ combined with initial distributions $\mathbf{p}_1(0)$ and $\mathbf{p}_2(0)$, respectively.

Fig. 3.A show the dynamical-fingerprint amplitudes of equilibrium experiments using each of the three possible observable vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$. The scalar product of observable vector $\mathbf{a}$ with eigenvectors $\mathbf{l}_3$ and $\mathbf{l}_4$, is close to zero and therefore the corresponding experiment is insensitive to these processes. This is in line with the intuition that an experiment in which the chromophores are attached at site 1 and 2 will be most sensitive to conformational changes of the α-helix. Conversely, observable $\mathbf{b}$ has little overlap with eigenvectors $\mathbf{l}_2$, and the corresponding experiment is insensitive to this kinetic process representing the α-helix folding. Only with observable $\mathbf{c}$ all relevant kinetic processes can be observed. This shows that a single experiment will typically be insensitive to some of the processes present in the system and explains why kinetics appear often simpler in experiments than in simulations.

A perturbation experiment cannot reveal a kinetic process which would be invisible in an equilibrium experiment with the same observable. This is easily seen by considering that the amplitude is given as $\gamma_i^{\mathbf{p}(0),a} = \langle \mathbf{a}, \mathbf{l}_i \rangle \langle \mathbf{p}'(0), \mathbf{l}_i \rangle$. If the first factor is zero, i.e. the process is not detectable in an equilibirum experiment, the corresponding amplitude in the fingerprint of the perturbation experiment will also be zero - independent of the initial distribution $\mathbf{p}(0)$. However, not all kinetic processes are involved in relaxing a particular initial distribution to the equilibrium distribution and therefore the second factor can become zero, too. This is shown in Fig. 3.B, where $\langle \mathbf{p}_1'(0), \mathbf{l}_3 \rangle \approx 0$ and $\langle \mathbf{p}_2'(0), \mathbf{l}_4 \rangle \approx 0$, respectively (see also Fig. 2.a and see Fig. 2.c). Consequently, these perturbation experiments are insensitive to the third and fourth kinetic process respectively, even though they are conducted with an observable which sensitive to all processes in the system.

Given a MSM of the conformational kinetics of biomolecule, our approach can also be used to suggest optimal attachment points for the chromophores[7,8].

Acknowledgments

Funding from the German Science Foundation (DFG) through grant number NO 825/2 and through research center MATHEON is gratefully acknowledged.

References

1. S. Muff and A. Caflisch, *Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a beta-sheet miniprotein*, Proteins-Struct., Funct., Bioinf., **70**, no. 4, 1185–1195, 2008.
2. F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, *Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations*, Proc. Natl. Acad. Sci. USA, **106**, no. 45, 19011–19016, 2009.
3. G. R. Bowman, K. A. Beauchamp, G. Boxer, and V. S. Pande, *Progress and challenges in the automated construction of Markov state models for full protein systems*, J. Chem. Phys., **131**, no. 12, 124101, 2009.
4. Marcus Jäger, Houbi Nguyen, Jason C. Crane, Jeffery W. Kelly, and Martin Gruebele, *The folding mechanism of a beta-sheet: the WW domain*, J. Mol. Biol., **311**, no. 2, 373–393, 2001.
5. Hannes Neuweiler, Marc Löllmann, Sören Doose, and M. Sauer, *Dynamics of Unfolded Polypeptide Chains in Crowded Environment Studied by Fluorescence Correlation Spectroscopy*, J. Mol. Biol., **365**, 856–869, 2007.
6. J-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Fischbach, M. Held, J. D. Chodera, C. Schütte, and F. Noé, *Markov model and molecular kinetics: generation and validation*, J. Chem. Phys., **134**, 174105, 2011.
7. F. Noé, S. Doose, I. Daidone, M. Lollmann, M. Sauer, J. D. Chodera, and J. C. Smith, *Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments*, Proc. Natl. Acad. Sci. USA, **108**, no. 12, 4822–4827, 2011.
8. B. G. Keller, J.-H. Prinz, and F. Noé, *Markov models and dynamical fingerprints: unraveling the complexity of molecular kinetics.*, Chem. Phys., submitted 2011.