Chemical Physics 375 (2010) 8–15

Contents lists available at ScienceDirect

Chemical Physics

journal homepage: www.elsevier.com/locate/chemphys

Energy dissipation asymmetry in the non equilibrium folding/unfolding of the single molecule alanine decapeptide

Piero Procacci [a,*], Simone Marsili [b]

[a] Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, I-50019 Sesto Fiorentino, Italy
[b] Centro Interdipartimentale per lo Studio delle Dinamiche Complesse(CSDC), Via Sansone 1, I-50019 Sesto Fiorentino, Italy

ARTICLE INFO

Article history:
Received 6 May 2010
In final form 9 July 2010
Available online 22 July 2010

Keywords:
Non equilibrium processes
Crooks theorem
Steered molecular dynamics
Decaalanine
Time asymmetry

ABSTRACT

We present non equilibrium molecular dynamics experiments of the unfolding and refolding of a single molecule alanine decapeptide *in vacuo* subject to a Nosé thermostat. Forward (unfolding) and reverse (refolding) work distribution are numerically calculated for various duration times of the non equilibrium experiments. Crooks theorem is accurately verified for all non equilibrium regimes and the time asymmetry of the process is measured using the recently proposed Jensen–Shannon divergence [E.H. Feng, G. Crooks, Phys. Rev. Lett. 101, 090602 (2008)]. Results on the alanine decapeptide are found similar to recent experimental data on m-RNA molecule in solution, thus evidencing the universal character of the Jensen–Shannon divergence. The patent non Markovianity of the process is rationalized by assuming that the observed forward and reverse distributions can be each described by a combination of two normal distributions satisfying the Crooks theorem, representative of two mutually exclusive linear events. Such bimodal approach reproduces with surprising accuracy the observed non Markovian work distributions.

1. Introduction

Some time ago Crooks [1] derived, in the context of Monte Carlo simulations, an exact formula involving the dissipative work of a system driven out of equilibrium through a time dependent external potential and in contact with a thermal bath at temperature $T = 1/k_B\beta$. This formula, ever since known as the Crooks theorem (CT), reads:

$$\frac{P(x,\Lambda)}{P(\hat{x},\widehat{\Lambda})} = e^{\beta(W-\Delta F)} \tag{1}$$

where $P(x,\ \Lambda)$, $P(\hat{x},\ \widehat{\Lambda})$ are the probabilities of observing a forward trajectory x, giving the time schedule (or protocol) Λ, and of observing its conjugate trajectory $\hat{x}$ with inverted transformation protocol $\widehat{\Lambda}$, respectively; $\Delta F \equiv F_B - F_A$ is the free energy difference between the initial and final canonical ensembles and W is the work done in the *forward* driven non equilibrium experiment. The Crooks formula has been later recognized of much broader validity, and it was shown to hold for deterministic systems in the context classical molecular dynamics simulations [2–5], Langevin dynamics [6,7], quantum systems [8,9] and verified in real [10] and computer [11,3,12] experiments.

The essential points for Eq. (1) to hold is that the driven forward and reverse experiments ought to be started from equilibrium distributions and that the transformation protocols of the forward and reverse process (that can involve mechanical and thermodynamic variables [5] as well) must be related by a time-reversal transformation. As the work done in the non equilibrium trajectory inverts sign by time-reversal, the trajectories and their time-reversal counterpart can be labeled using the work such that Eq. (1) can be also written as:

$$\frac{P(W|F)}{P(-W|R)} = e^{\beta(W-\Delta F)} \tag{2}$$

where $P(W|F)$, $P(-W|R)$ are the probability of observing a work W in the forward and reverse experiment. Eq. (2) says that trajectories that are highly dissipative (i.e. $W - \Delta F \gg 0$) in the forward sense are difficult to observe in the reverse sense since for such trajectories the dissipation of its time-reversal counterpart would be negative, thus transiently violating the second law. In the functional form of Eq. (2), the Crooks theorems applies, with some provisions [13] related to the form of the external driving agent, to the controlled mechanical manipulation of a single molecule through optical tweezers [10] or atomic force microscopy [14]. We conclude this introductory remarks by stating that Eq. (2), one of the very few *exact* equations in non equilibrium thermodynamics, holds for any regime: for instantaneous pulling we have that $W = H_B - H_A$ and, by averaging over all trajectories, one recovers the Zwanzig [15]

* Corresponding author. Tel.: +39 0554573086; fax: +39 0554573077.
E-mail address: procacci@unifi.it (P. Procacci).

0301-0104/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemphys.2010.07.009

formula $< e^{-\beta(H_B-H_A)} >_A = e^{\beta\Delta F}$. For infinitely slow pulling, i.e. for quasi-static reversible transformations, $W = \Delta F$ and the forward and backward distribution are indistinguishable and $P(W|F) = P(-W|R) = \delta(W - \Delta F)$.

Recently there has been considerable progress in the interpretation of non equilibrium experiments coming both from measurements on single molecules using AFM or optical traps [16] and from deterministic or stochastic simulations [17]. Feng and Crooks proposed to use the Jensen–Shannon divergence [18,19] (JSD) between the probability of a trajectory and its time-reversal conjugate as a definition and a measure of the *time asymmetry* in a thermodynamic system. If we use the work W (which changes sign by time-reversal) as a label for trajectories, then the JSD can be written in terms of work distributions as:

$$\mathrm{JSD} = \frac{1}{2}\int P(W|F)\ln\frac{2P(W|F)}{P(W|F)+P(-W|R)}dW + \frac{1}{2}\int P(-W|R)\ln\frac{2P(-W|R)}{P(W|F)+P(-W|R)}dW \quad (3)$$

The JSD can be shown [16] to be equal to the average gain of information about the orientation of time's arrow from one single realization of the experiment. This quantity, plotted against the average dissipation obtained in the forward and reverse driven experiments, goes to zero for reversible processes, and to one full nats of information ln 2 (i.e. 1 bit) when the two distributions do not overlap (i.e. for large average dissipation). In this latter case, it is easy to assign an observed trajectory (taken from the pool of forward and reverse non equilibrium experiments) to one of two distributions, or, stated in other words, it is easy to guess, from the analysis of one single random trajectory, in which direction the time is flowing. On such basis, when plotted against the average mean dissipation, the JSD may then give indication on the energetic cost (i.e. the dissipation needed) to ensure that a molecular process (e.g. a molecular motor) advances in time. For Markovian (linear) systems, the work distributions are always Gaussian [11,3] with variance twice the average dissipation. In this case, JSD vs dissipation is analytic and identical for all Markovian system. Therefore, Eq. (3) can also be used as a measure of the non linearity of the system.

In the context of non equilibrium thermodynamics, similar concepts were put forward recently by Kawai et al. [17] These authors interestingly restated the CT, Eq. (2), in the form

$$< W > -\Delta F = k_B T \int dW P(W|F)\log\left(\frac{P(W|F)}{P(-W|R)}\right) \quad (4)$$

$$= k_B T D[P(W|F)|P(-W|R)] \quad (5)$$

The integral in Eq. (4) defines $D[P(W|F)|P(-W|R)]$, the Kullback–Leibler divergence (KLD), [19] a strictly positive quantity measuring, in information theory, the expected extra message-length per datum that must be communicated if a code that is optimal for a given (wrong) distribution $P(-W|R)$ is used, compared to using a code based on the (true) distribution $P(W|F)$. In general the KLD is not symmetric, i.e. if q, p are two non identical distributions, $D(p|q) \neq D(q|p)$. For Markovian systems, however, the KLD is always symmetric. Moreover, for such systems, k_BT times the KLD can be calculated analytically yielding the dissipation $\beta\sigma^2/2$, with σ^2 being the variance of the Gaussian distribution. KLD between the forward and reverse distributions has the same characteristics of the JSD divergence, being the former like the latter both a measure of the time asymmetry (i.e. of the possibility for distinguish in which sense the time is flowing) and of non linearity. However KLD, as suggested by Kawai et al., could be effectively used as a tool for obtaining a better upper bound of the free energy than the average work W. This is so since, according to the chain rule, [19] the relative entropy (or KLD) decreases upon coarse graining. An extremely simple scheme could be that of approximating coarse grain histograms of the forward and backward work distribution with the best linear model satisfying the Crooks theorem. This approach has been advocated recently by Forney et al. [20] in the context of steered molecular dynamics of alanine decapeptide *in vacuo* along the end-to-end distance. These authors, in their so-called *FR* method [21,20], produce a coarse grain histogram with few work measurements in both directions that are then fitted using a linear (Markovian) model, retrieving the free energy difference between initial and final states and the dissipation of the system. However, when the driven coordinates exhibit clear non linear effects (i.e. the noise due to all other "solvent" coordinates is not white or Gaussian), as is the case of folding and refolding of small proteins along the end-to-end distance, then other less simplistic coarse grain schemes could and should be adopted.

In this paper we further develop the concepts of time asymmetry and coarse graining introduced in Refs. [17,16] by presenting extensive non equilibrium molecular dynamics simulation data of unfolding and refolding process of alanine decapeptide *in vacuo* performed with the deterministic Nosé–Hoover thermostat at 300 K. In spite of the fact that alanine decapeptide *in vacuo* has been extensively studied in the recent past by non equilibrium computational techniques [11,3,20], the rationalization and interpretation of the observed data is still a matter of debate. α-helix formation/disruption is also important *per se* and as a paradigm for an elementary folding/unfolding process.

Our results on alanine decapeptide are interpreted by means of the JSD and KLD quantities above introduced. We further present a simple coarse grain and totally general model satisfying the CT based on the assumption of the occurrence, in the refolding process, of two mutually exclusive events. Such a simple dual model explains many features of the observed work distributions and can be rationalized with the existence of two competing minima for low values of the end-to-end distance in alanine decapeptide, i.e. one of enthalpic nature (the helix), easily accessible, in the refolding process, at low dissipation regimes, and the other of entropic origin corresponding to a manifold or misfolded coil structures which emerges at large dissipation when trying to rapidly refold alanine decapeptide from extended structures. This view appears to be quite general and is fully consistent with the rugged funnel picture of the folding process, in the sense that escaping the rugged funnel from below is a much tamer process than reentering the funnel from above.

The present paper is organized as follows. Section 2 is dedicated to the description of the systems and of the methods used in the non equilibrium simulations. In Section 3 we present the computer experiment results of the unfolding/refolding of a single molecule of alanine decapeptide along with a discussion focusing on the thermodynamic and microscopic aspects of the process. Conclusive remarks and futures perspective regarding the applicability of the presented methodology to real experiments are presented in Section 4.

2. Methods

In this section we provide the technical details on the steered molecular dynamics simulations of the alanine decapeptide (A_{10}) *in vacuo*. The unperturbed system is described with the all-atom force field CHARMM whose parameters are given in Ref. [22]. A constant temperature of 300 K is imposed through a Nosé–Hoover thermostat [23]. The resulting deterministic equations of motions are efficiently integrated using a reference system propagator algorithm [24] at three time steps, 3.0 fs for medium and long range non-bonded interactions (no cut-off is imposed), 1.5 fs for torsional potential involving hydrogen atoms and for short-ranged (14) non-bonded interactions, and 0.5 fs for stretching and bending