Free Energy Estimation in Dissipative Particle Dynamics

Subin Bang, Chanwoo Noh†, and YounJoon Jung*

Department of Chemistry, Seoul National University, Seoul 151-747, Korea.
Received Date; Accepted Date (to be inserted by the publisher after your manuscript is accepted)

Abstract: The methods for estimating the change of free energy in dissipative particle dynamics (DPD) are discussed on the basis of fluctuation theorems. Fluctuation theorems are tactics to evaluate free energy changes from non-equilibrium work distributions and have several forms, as proposed by Jarzynski, Crooks, and Bennett. The validity of these methods however, has been shown merely with the molecular dynamics or Langevin dynamics. In this study, the appropriate forms of fluctuation theorems for dissipative particle dynamics, which has similar structure to that of Langevin dynamics, are suggested using Liouville's theorem, and they are proved equivalent to original fluctuation theorems. Work distribution functions, which are probability distribution functions of works exerted on the system within the systematic change, are the basics of fluctuation theorems and their shapes are turned out to be dependent on the phase space trajectory of the change of the system. The reliability of Jarzynski and Crooks methods is highly dependent on the number of simulations to measure works and the shapes of the work distribution functions. Bennett method, however, can evaluate free energy changes even when Jarzynski and Crooks methods fail to do so.

Keywords: Dissipative Particle Dynamics, Free Energy, Fluctuation Theorem, Liouville's Theorem.

Introduction

In statistical mechanics, microscopic dynamics of small particles results in overall macroscopic phenomena. Therefore macroscopic properties of certain system such as pressure and chemical potential can be described by computational modeling of microscopic interactions. Classical molecular dynamics is one of the modeling techniques and evaluates dynamics of small molecules by time-integration of equations of Newtonian mechanics. Since the numerical integration of forces generated by hard potentials such as Lennard-Jones, and Coulomb potentials are very sensitive to integration tactics, the system evolves relatively slowly. In addition, due to the large number of particles of the system, classical molecular dynamics simulation is very expensive computation.

Coarse graining is developed to overcome such difficulties by raising the order of scope of the system, from microscopic to mesoscopic dynamics. Generalized equations of motion of many-body coarse grained system were proposed with three major forces[1], which are conservative, random, and friction forces. A special result of this generalized equation is Langevin dynamics which also contains three parts of forces[2]. However, several stochastic dynamics, including Langevin dynamics, have inborn weaknesses. For instance, though equations of motion should be Galilean invariant (time symmetry) and conserve total momentum of system in real-world dynamics, those in Langevin dynamics cannot guarantee all of these important conditions [3].

Dissipative particle dynamics (DPD) is proposed to cure such difficulties in order to explain real-world dynamics[3]. DPD is a kind of stochastic dynamics consisting of three aforementioned major parts, conservative, random, and friction forces. To be more specific, the particles in DPD are coarse-grained clusters of molecules being governed by soft potentials and pairwise dissipations. A reliable form of the three forces has been suggested[4]. In addition DPD has been reviewed as one of the good coarse grained approaches with dissipative thermostat[5].

According to several articles and computer simulations, system of DPD clusters show reasonable estimation of macroscopic properties[6,7] within relatively low time cost. Compressibility, viscosity, and configurational temperature are well known examples. However, free energy of certain systematic reaction has not been evaluated within DPD even if ways to estimate it in classical molecular dynamics and Monte Carlo simulation have been developed, such as Bennett acceptance ratio method[8], Crooks distributional method[9], and Jarzynski method[10]. They are introduced as *fluctuation theorems* and of great preference due to their direct relation with probability distribution functions of works done on the system, which can be measured by repeated simulations. In recent days there have been analytical researches that relate free energy change with work distribution functions in stochastic dynamics[11,12]. For

†Email: eric8617@snu.ac.kr
***Email: yjjung@snu.ac.kr**

stochastic coarse grained model, this relation is formulated on the basis of Langevin dynamics, and the form of the relation is found equivalent to original ones proposed by Jarzynski and Crooks. However, because of cumbersome features of Liouville operators in DPD, it is difficult to deal with trajectories in phase space, which are directly related with the work distribution functions. It is the one of the main reasons that fluctuation theorems in DPD have not been discussed even if DPD and Langevin dynamics have similar forms and share similar stochastic approach.

In this work, the form of the relation between work distribution functions and free energy, that is fluctuation theorem, is shown. At first step the structure of the partition function of the system in DPD is reviewed and developed to a form of canonical partition function. Then, the free energy relation is proven to be equivalent to that of original researches, proposed by Crooks[9]. Corresponding Jarzynski and Bennett methods in DPD can be derived from Crooks relation. Using these formulae, free energy of unfolding of a linear polymer is estimated. The dynamics of monomers is investigated through the bead-spring model[13] with Lennard-Jones potentials at the ends of the polymer branch. Folded with Lennard-Jones potential[14], the ends of the polymer are pulled with constant velocity. The effectiveness of Jarzynski, Crooks, and Bennett estimation methods is shown to vary with pulling velocities. As a result, in agreement with recent researches[15], Bennett method is the best approach to calculate free energy change of the given system, even if the number of simulation samples is small.

Theory and Computational Method

A. *Structure of DPD*

The equations of motion in DPD follow classical Newtonian mechanics scheme

$$\frac{d\boldsymbol{r_i}}{dt} = \boldsymbol{v}_i, \qquad \frac{d\boldsymbol{v_i}}{dt} = \boldsymbol{f}_i,$$

and DPD forces, $\boldsymbol{f_i}$, consist of three major parts[4]

$$\boldsymbol{f_i} = \sum_{i \neq j} \left(\boldsymbol{F_{ij}^C} + \boldsymbol{F_{ij}^D} + \boldsymbol{F_{ij}^R} \right),$$

where C, D, R refer to conservative, dissipative, and random forces respectively. There are proposed momentum-conserving forms of $\boldsymbol{F_{ij}^D}$, $\boldsymbol{F_{ij}^R}$ that are used typically

$$\boldsymbol{F_{ij}^D} = -\gamma \omega^D(r_{ij}) (\hat{\boldsymbol{r}}_{\boldsymbol{ij}} \cdot \boldsymbol{v_{ij}}) \hat{\boldsymbol{r}}_{\boldsymbol{ij}},$$

$$\boldsymbol{F_{ij}^R} = \sigma \omega^R(r_{ij}) \xi_{ij} (\Delta t)^{-1/2} \hat{\boldsymbol{r}}_{\boldsymbol{ij}},$$

where ω^D, ω^R are r_{ij}-dependent weighted functions[3] vanishing for $r_{ij} > r_c$, $\boldsymbol{r_{ij}} = \boldsymbol{r_i} - \boldsymbol{r_j}$, $\boldsymbol{v_{ij}} = \boldsymbol{v_i} - \boldsymbol{v_j}$, $\hat{\boldsymbol{r}}_{\boldsymbol{ij}} = \boldsymbol{r_{ij}}/|\boldsymbol{r_{ij}}|$, and ξ_{ij} is a normally-distributed random number chosen independently for each pair of interacting clusters for each time step. r_c is dynamical cut-off radius. γ and σ are corresponding force constants. The form of conservative force $\boldsymbol{F_{ij}^C}$, is

$$\boldsymbol{F_{ij}^C} = -\nabla U,$$

where U is systematic potential, but since DPD is coarse grained model of clusters with soft potential, an useful form has been proposed[4],

$$\boldsymbol{F_{ij}^C} = \begin{cases} A_{ij}\left(1 - \frac{r_{ij}}{r_c}\right) & (r_{ij} < r_c) \\ 0 & (r_{ij} \geq r_c) \end{cases},$$

where A_{ij} represents maximum magnitude of soft repulsion. In addition, fluctuation-dissipation theorem requires that $\sigma^2/2k_BT = \gamma$ and $\omega^D = (\omega^R)^2$. Often it is convenient to set[16] $\omega^R = (1 - r_{ij}/r_c)$. Reduced units and their meanings are shown in table A1 and A2.

B. *Fluctuation Theorem*

Hamiltonian of a classical system can be viewed as

$$H(\Gamma; \lambda) = \mathrm{H_k}(\boldsymbol{p_i}) + H_r(\Gamma; \lambda),$$

where k and r represents kinetic and rest (potential) part of system Hamiltonian, Γ is a point in phase space and λ is a parameter that can be manipulated externally. If λ varies (protocol), work W done on the system by surroundings is[17]

$$\delta W \equiv d\lambda \frac{\partial H}{\partial \lambda}(\Gamma; \lambda) = dt \frac{d\lambda}{dt} \frac{\partial H_r}{\partial \lambda}(\Gamma; \lambda),$$

$$W = \int \delta W = \int dt \frac{d\lambda}{dt} \frac{\partial H_r}{\partial \lambda}\big(\Gamma(t); \lambda(t)\big).$$

Here protocols refer to the procedures which the system in thermal equilibrium evolves along, and the evolution of the system is determined by the change of certain parameter, λ. If we choose the change $\lambda_0 \rightarrow \lambda_t$ to be *forward* protocol that the system in thermal equilibrium with λ_0 advances toward λ_t state, *backward* protocol is the process that the system in thermal equilibrium with λ_t evolves to λ_0. The final position of each protocol needs not to be in equilibrium state since there is no additional work done on the system during its re-equilibrium process (figure 1).

Meanwhile configurational probability distribution function (PDF), alongside the partition function of a *NVT*-system in canonical ensemble is

$$p_{\lambda,T}^{eq}(\Gamma) = \frac{1}{Z_{\lambda,T}} e^{-H(\Gamma;\lambda)/k_BT},$$

$$Z_{\lambda,T} = \int e^{-H(\Gamma;\lambda)/k_BT} \delta\Gamma,$$

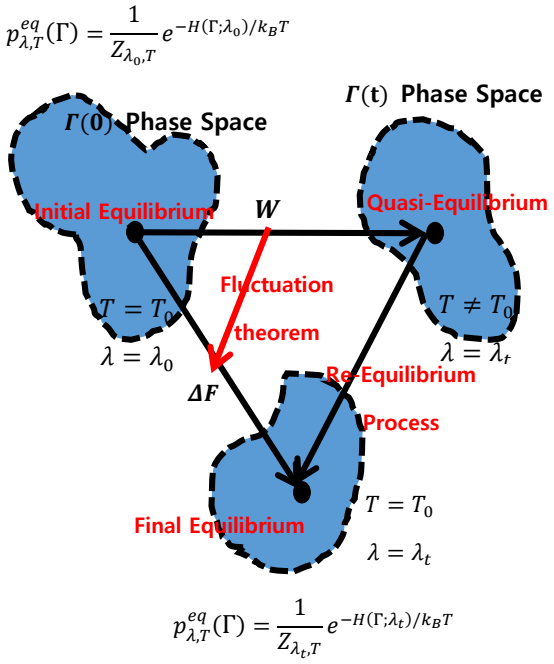

Figure 1. Diagram of evolution of the system in the phase space. $\boldsymbol{W}$ is the measured work done on the system and $\boldsymbol{p^{eq}_{\lambda,T}}$ refers to canonical ensemble distribution. Probability distribution function cannot be defined in the quasi-equilibrium region.

where $Z_{\lambda,T}$ is the canonical partition function. With Hamiltonian formulation of plain Newtonian dynamics and the relation between free energy and canonical partition function, $F_{\lambda,T} = -k_BT \ln Z_{\lambda,T}$, Jarzynski[10], Crooks[9], and Bennett[8] have shown that with the works measured repeatedly through the same protocol,

$$\langle e^{-W/k_BT} \rangle_F = e^{-\Delta F/k_BT},$$

$$\frac{P_F[W;\gamma_F]}{P_R[-W;\gamma_R]} = e^{(W-\Delta F)/k_BT},$$

$$l(\Delta F) = \sum_{i=1}^{K} \ln \frac{1}{1+\exp(-\beta(W=W_F)+\beta\Delta F)} + \sum_{i=1}^{K} \ln \frac{1}{1+\exp\left(-\beta\,(\widetilde{W}=W_R)-\beta\Delta F\right)},$$

where $\langle A \rangle$ denotes average of physical quantity A over the work distribution function from the measured works, P is probability distribution to observe specific phase space trajectory γ, K is the number of measurements of the works, $l(\Delta F)$ is logarithmic likelihood to measure ΔF as the free energy change of the system, K is the number of work measures, and F, R represent forward (original) and backward (reverse) protocols. From several simulations one can get a set of measured W. By exponentially averaging, solving the intersection between P_F and P_R, or finding ΔF that maximizes $l(\Delta F)$ with this set, free energy change of the protocol can be evaluated. Often above equations are referred to Jarzynski equality, Crooks fluctuation theorem, and Bennett acceptance ratio method named after their constructors, respectively.

C. *Simulation Scheme*

In this computational experiment aforementioned free energy estimation techniques are illustrated by considering unfolding processes of linear polymers. There are two types of linear polymers in the experiments. One is a *spring-bead polymer* which consists of N_1 monomer-beads with harmonic potentials U_1,

$$U_{total} = U_1 = \frac{k}{2} \sum_{i=1}^{N_1-1} (\boldsymbol{r_i} - \boldsymbol{r_{i+1}})^2,$$

where $\boldsymbol{r_i}$ indicates the position vector of the ith bead.

The other is a *hairpin polymer* that monomers at the ends of the polymer have additional Lennard-Jones potentials U_2,

$$U_{total} = U_1 + U_2 = \frac{k}{2} \sum_{i=1}^{N_1-1} (\boldsymbol{r_i} - \boldsymbol{r_{i+1}})^2 + \sum_{\substack{\boldsymbol{ij} \\ \text{attractives}}}^{\boldsymbol{n}} 4\epsilon \left[\left(\frac{l}{r_{ij}}\right)^{12} - \left(\frac{l}{r_{ij}}\right)^{6} \right],$$

where $n < N_1$ is the number of monomers with Lennard-Jones potential which is main cause of the folded, hairpin-like linear polymer[15].

Both types of the polymers containing N_1 monomer-beads are solvated with N_2 solvent particles. Every particle exerts conventional DPD forces,

$$\boldsymbol{F^C_{ij}} = \begin{cases} A_{ij}\left(1 - \frac{r_{ij}}{r_c}\right) & (r_{ij} < 1) \\ 0 & (r_{ij} \geq 1) \end{cases},$$

$$\boldsymbol{F^D_{ij}} = -\gamma \omega^D(r_{ij})(\hat{\boldsymbol{r}}_{\boldsymbol{ij}} \cdot \boldsymbol{v_{ij}})\hat{\boldsymbol{r}}_{\boldsymbol{ij}},$$

$$\boldsymbol{F^R_{ij}} = \sigma \omega^R(r_{ij})\xi_{ij}(\Delta t)^{-1/2}\hat{\boldsymbol{r}}_{\boldsymbol{ij}},$$

where i, j refers to $(N_1+N_2)(N_1+N_2-1)$ monomer-solvent pairs with $\sigma^2/2k_BT = \gamma$ and $\omega^D = \left(\omega^R = 1 - r_{ij}\right)^2$.

The linear polymers are prepared to reach their thermal equilibrium, with fixed positions of the first and last monomers at $\boldsymbol{r_1} = L(0.1, 0.5, 0.1)$ and $\boldsymbol{r_{N_1}} = L(x, 0.5, 0.1)$ given that L is diameter of cubic simulation box. Then the last monomers are pulled with constant speed until they reaches the point $\boldsymbol{r_{N_1}} = L(y, 0.5, 0.1)$, $y > x$ which are

unfolding processes or, forward protocols in fluctuation theorem. If the systems reach thermal equilibrium with their unfolded states, the backward protocols start, which are the pulling-back procedures of the last monomers, from $\boldsymbol{r_{N_1}} = L(y, 0.5, 0.1)$ to their original positions $\boldsymbol{r_{N_1}} = L(x, 0.5, 0.1)$ with same absolute speed of forward protocols. These processes are depicted in figure 2. During both protocols, works done on the system are calculated by analyzing trajectories of particles. Specific values of constants used in experiments of Results and Discussion section are shown in tables in appendix.

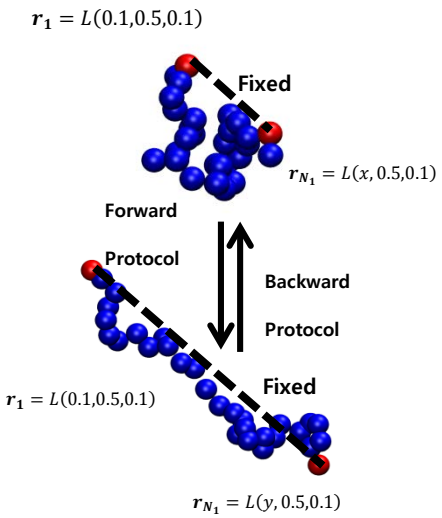

Figure 2. Forward and backward protocols of the linear polymer with $\boldsymbol{N_1 = 30}$.

Results and Discussion

Before the simulation, we should prove that partition function of DPD system is canonical partition function and should find appropriate form of the fluctuations theorems in DPD. This procedure is shown in Algebraic Results section.

A. Algebraic Results

First, we begin our discussion by proving that DPD forms NVT-system. The time derivative of PDF p can be expressed by Liouville's equation with Poisson bracket $\{\cdot, H\}$,

$$\frac{dp(\Gamma, t)}{dt} = \frac{\partial p}{\partial t} - \{p, H\} = \frac{\partial p}{\partial t} + \hat{L}p.$$

Since distribution function is constant along any trajectory in phase space (Liouville's theorem),

$$\frac{\partial p}{\partial t} - \{p, H\} = \frac{\partial p}{\partial t} + \hat{L}p = 0.$$

Corresponding Liouville operator $\hat{L}$ for DPD is already derived as[18]

$$\hat{L} = \hat{L}_C + \hat{L}_D + \hat{L}_R,$$

$$\hat{L}_C = -\sum_i \left(\boldsymbol{v_i} \cdot \frac{\partial}{\partial \boldsymbol{r_i}} + \frac{\boldsymbol{F_i}}{m} \cdot \frac{\partial}{\partial \boldsymbol{v_i}} \right),$$

$$\hat{L}_D = \sum_{i \neq j} \gamma \omega_D(r_{ij}) \left(\hat{\boldsymbol{r}}_{\boldsymbol{ij}} \cdot \frac{\partial}{\partial \boldsymbol{v_i}} \right) (\hat{\boldsymbol{r}}_{\boldsymbol{ij}} \cdot \boldsymbol{v_{ij}}),$$

$$\hat{L}_R = \sum_{i \neq j} \frac{\sigma^2}{2} \omega_R^2(r_{ij}) \left(\hat{\boldsymbol{r}}_{\boldsymbol{ij}} \cdot \frac{\partial}{\partial \boldsymbol{v_i}} \right) \hat{\boldsymbol{r}}_{\boldsymbol{ij}} \cdot \left(\frac{\partial}{\partial \boldsymbol{v_i}} - \frac{\partial}{\partial \boldsymbol{v_j}} \right).$$

However this does not mean $\hat{L}_{\mathrm{DPD}} = -\{\cdot, H\}$ but $\left(\frac{\partial}{\partial t} + \hat{L}_{DPD}\right) p = 0$ and $\hat{L}_{\mathrm{C}} = -\{\cdot, H\}$, because $\partial p/\partial t$ contains dissipative terms. If the system reaches its equilibrium state, $p = p^{eq}$ and $\partial p/\partial t = 0$. Then

$$\hat{L}p = (\hat{L}_C + \hat{L}_D + \hat{L}_R)p = 0.$$

At this step we assume quasi-solution of Liouville's equation from Maxwell-Boltzmann approach,

$$p^{eq}(\Gamma) = A(r_1, \cdots, r_N) \exp\left(-\frac{1}{2\theta_0} \sum_i m|\boldsymbol{v_i}|^2 \right),$$

where θ_0 is equilibrium temperature and A is position-dependent term. Inserting this quasi-solution into Liouville's equation gives (proving $\hat{L}_D p = \hat{L}_{\mathrm{R}} p = 0$ is straightforward due to phase space symmetry),

$$(\hat{L}_C + \hat{L}_D + \hat{L}_R)p = \hat{L}_C p,$$

$$\hat{L}_C p = -\sum_i \left[\left(\boldsymbol{v_i} \cdot \frac{\partial}{\partial \boldsymbol{r_i}} A(\boldsymbol{q}) \right) \exp\left(-\frac{1}{2\theta_0} \sum_j m|\boldsymbol{v_j}|^2 \right) + A(\boldsymbol{q}) \frac{\boldsymbol{F_i}}{m} \cdot \frac{\partial}{\partial \boldsymbol{v_i}} \exp\left(-\frac{1}{2\theta_0} \sum_j m|\boldsymbol{v_j}|^2 \right) \right].$$

To be $\hat{L}_C p = 0$ for all $\boldsymbol{v_j}$'s,

$$\left(\boldsymbol{v_i} \cdot \frac{\partial}{\partial \boldsymbol{r_i}} + \frac{\partial U}{\partial \boldsymbol{r_i}} \cdot \frac{\boldsymbol{v_i}}{\theta_0} \right) A(\boldsymbol{q}) = \boldsymbol{v_i} \cdot \left(\frac{\partial A(\boldsymbol{q})}{\partial \boldsymbol{r_i}} + A(\boldsymbol{q}) \frac{\partial \left(\frac{U}{\theta_0} \right)}{\partial \boldsymbol{r_i}} \right) = 0,$$

and this gives

$$A(r_1, \cdots, r_N) = 1/Z \exp(-U/\theta_0),$$

where Z is normalization constant. Using this equation, $p^{eq}(\Gamma)$ becomes[19]

$$p^{eq}(\Gamma) = \frac{1}{Z} \exp\left(-\frac{H(\Gamma)}{\theta_0} \right),$$

which is Gibbs (canonical) distribution, PDF of NVT-ensemble. Therefore the system of DPD clusters has constant temperature at theoretical level. Further study shows that $\sigma^2/2k_BT = \gamma$, which can be derived from fluctuation-dissipation theorem toward p^{eq}. $\sigma^2/2k_BT = \gamma$ acts like thermostat[3].

Next we prove that NVT-ensemble of DPD establishes fluctuation theorem that has equivalent forms of original researches. When the hydrodynamic system relaxes toward local equilibrium during its journey of $\lambda_0 \to \lambda_t$ (forward protocol)[19],

$$\partial_t n = \nabla \cdot (n\boldsymbol{u}),$$

(from the condition of balanced system) where

$$n(\boldsymbol{r},t) = \langle \sum_i \delta(\boldsymbol{r}-\boldsymbol{r_i}) \rangle = \int d\boldsymbol{v} f(\boldsymbol{v},\boldsymbol{r},t),$$

$$n\boldsymbol{u}(\boldsymbol{r},t) = \langle \sum_i \boldsymbol{v_i}\delta(\boldsymbol{r}-\boldsymbol{r_i}) \rangle = \int d\boldsymbol{v} f(\boldsymbol{v},\boldsymbol{r},t)\,\boldsymbol{v},$$

$$f(\Gamma,t) = f(\boldsymbol{r},\boldsymbol{v},t) = \langle \sum_i \delta(\Gamma-\Gamma_i) \rangle,$$

and $\langle A \rangle$ denotes macroscopic quantity of A, averaged over canonical ensemble PDF. Often f is called single particle probability distribution and $\boldsymbol{u}$ is referred to the phase space velocity corresponding to Hamiltonian flow. If the protocol is sufficiently slow $\partial_t n \ll 1$ and then

$$\nabla \cdot (n\boldsymbol{u}) = \nabla \cdot \int d\boldsymbol{v} f(\boldsymbol{v},\boldsymbol{r},t)\boldsymbol{v} = \int d\boldsymbol{v} \nabla \cdot (f(\boldsymbol{v},\boldsymbol{r},t)\boldsymbol{v}) \cong 0,$$

$$\int d\boldsymbol{v} \nabla f(\boldsymbol{v},\boldsymbol{r},t) \cdot \boldsymbol{v} = 0.$$

Here again we propose a quasi-solution,

$$f(\Gamma,t) \cong f_0 = n(\boldsymbol{r},t)\left(\frac{m}{2\pi\theta}\right)^{\frac{d}{2}} \exp\left(-\frac{m}{2\theta_0}\boldsymbol{v}^2\right),$$

at local equilibrium, i.e. if the system stays at the global equilibrium, $n(r,t) \to N/V$, which is number density. Note that proposed forms of f and p^{eq} have similar structures. Here again, we need to confirm whether quasi-f is simultaneous solution of both $\partial_t n = \nabla \cdot (n\boldsymbol{u})$ and Liouville equation, since f is a single-particle version of entire PDF. Inserting quasi-f into

$$\int d\boldsymbol{v} \nabla f(\boldsymbol{v},\boldsymbol{r},t) \cdot \boldsymbol{v} = 0,$$

then

$$\int d\boldsymbol{v} \left(\hat{L}_C + \sum_i \frac{\boldsymbol{F_i}}{m} \cdot \frac{\partial}{\partial \boldsymbol{v_i}} \right) f = 0.$$

Here we can verify $\hat{L}_C f = 0$ instantaneously from $\hat{L}_C p = 0$. Then

$$\sum_i \int d\boldsymbol{v} \frac{\nabla_i U}{m} \cdot \frac{\partial f}{\partial \boldsymbol{v_i}} = 0,$$

$$\sum_i \int d\boldsymbol{v} \frac{n(\boldsymbol{r},t)}{\phi} \nabla_i U \cdot \left(\frac{\boldsymbol{v_i}}{\theta_0}\right) \exp\left(-\frac{m}{2\theta_0}\boldsymbol{v_i^2}\right) = 0,$$

where ϕ is a normalization constant. Proving this equation is straightforward assuming that overall potential on the i-th particle is symmetric over the space. Therefore proposed form of f is plausible single-particle distribution function at local equilibrium, f_{local}^{eq}.

Again Liouville's theorem for $f = f_{local}^{eq}$ tells that

$$\partial_t f = \left(\hat{L}_C + \hat{L}_D + \hat{L}_R\right) f = 0,$$

and treating $n(\boldsymbol{r},t) \to A(r_1,\cdots,r_N)$ during showing Liouville's theorem for $p^{eq}(\Gamma)$ gives

$$n = \frac{1}{\zeta(t)} \exp\left(-\frac{U(\boldsymbol{r},t)}{\theta_0}\right),$$

with normalization constant, $\zeta(t)$.

Suppose that sequences $\{\Gamma_0,\Gamma_1,\cdots,\Gamma_t\}$ and $\{n_0,n_1,\cdots,n_t\}$ represent evolution of the system during the forward protocol, $\lambda_0 \to \lambda_t$. And

$$\lim_{\tau\to 0} \frac{n_{i+1}(\tau)}{n_i(\tau)} = \lim_{\tau\to 0} \frac{\exp(-U_{i+1}/\theta_0)\,\zeta_i}{\exp(-U_i/\theta_0)\,\zeta_{i+1}} = \frac{\zeta_i}{\zeta_{i+1}} e^{-\Delta H_i/\theta_0},$$

since kinetic part of Hamiltonian of the system will be nearly invariant under DPD thermostat in overall. Thus[17,20]

$$\frac{P_F[\gamma_F]}{P_R[\gamma_R]} = \prod_i \left(\lim_{\tau\to 0} \frac{n_{i+1}(\tau)}{n_i(\tau)} \right) = \prod_i \left(\frac{\zeta_i}{\zeta_{i+1}} e^{-\Delta H_i/\theta_0} \right) = \frac{\zeta_{initial}}{\zeta_{final}} e^{-\Delta H/\theta_0} = e^{(W-\Delta F)/k_BT}.$$

Therefore Crooks fluctuation theorem in DPD has same form with that in general NVT-ensemble system, for Jarzynski and Bennett methods can be deduced from this relation. This derivation is given in Comparison of Estimators section.

B. Properties of DPD Polymer

Generally the dynamics of the system with DPD is governed by thermostatting parts (random and dissipative forces), and conservative (attractive; Lennard-Jones and harmonic forces and repulsive; DPD repulsion) parts. What is worth to remark is that since the form of DPD repulsion force is suggested without any strict physical reasoning, it is necessary to investigate the best way to describe real world dynamics with this soft potential[1].

Alongside this disadvantage of DPD, there are discussions about how to choose appropriate repulsion parameter A_{ij} to cure such a problem[1,21,22]. Especially, for DPD is the easy tool to consider the effects of hydrodynamics, we need to

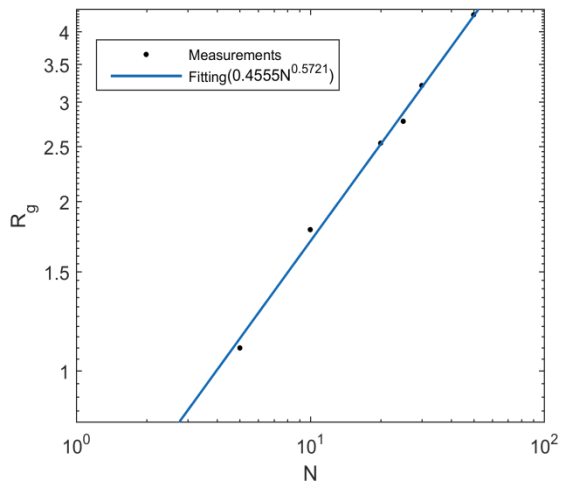

Figure 3. Change of radius of gyration over the number of monomers. The quality of the fitting curve is RMSE=0.06844 and $\boldsymbol{R^2 = 0.997}$.

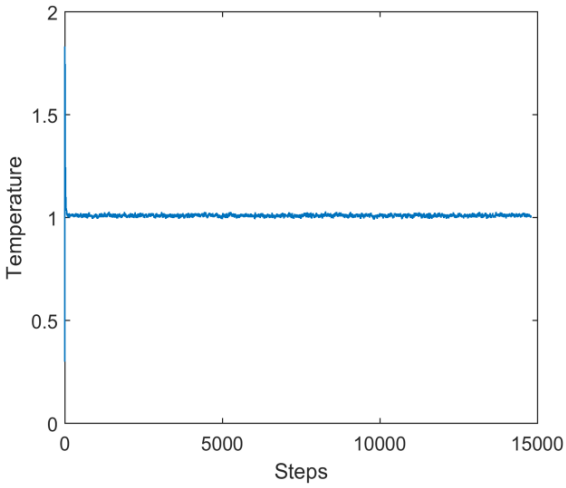

Figure 4. Kinetic temperature of solvated linear polymer system with $\boldsymbol{N = 20}$, $\boldsymbol{k_BT = 1.0}$ and $\boldsymbol{\sigma = 3.35}$.

consider the general behavior of the *polymers in solution* with DPD. In several researches the suitable value of A_{ij} is $25k_BT$ with number density 3[1,22]. In figure 3, the plot of radius of gyration of the spring-bead type linear polymer versus the number of monomers is shown. The constants and parameters used in this experiment are given in table A3. Similar to Flory theory of polymers[23] radius of gyration, R_g, grows as the number of beads, N, in the polymer increases with the conventional rule, $R_g \sim N^v, v = 0.5721$, which is close to so-called good solvent condition, $v = 0.588$. It is possible to approximate the real-world polymers with simplified potentials in DPD.

Free Energy Estimation

With the thermostat parameters $k_BT = 1.0$ and $\sigma = 3.35$, the kinetic temperature of given linear polymer system remains nearly constant as desired, in figure 4. It may be plausible to set $\sigma \cong 3$ under this environment. If σ is too small, rise of cluster velocities due to large repulsion, $A_{ij} = 25k_BT$, cannot be controlled by the nearest solvent clusters via dissipation. On the contrary, if sigma is very large, the dynamical effect of dissipation is dominant over the conservative forces so that we cannot get any meaningful trajectories of clusters. On this basis, in free energy estimating section, the systems with $A_{ij} = 25k_BT$, $k_BT = 1.0$, and $\sigma = 3$ are studied.

C. Free Energy of Linear Polymer

Since radius of gyration of the solvated, spring-bead type polymer in the given solution obeys $R_g \sim N^v$ with $v = 0.5721$, which is close to the condition of good solvents, free energy F of the linear polymer would follow[23,24]

$$F = k_BTV\frac{N^2}{R^3} + k_BT\frac{R^2}{Nb^2},$$

where V is excluded volume, b is a positive number and R is the length of end-to-end vector of the linear polymer. Above equation is proposed by Flory. Regarding unfolding process of the linear polymer as a forward protocol in the free energy estimation methods, the size of end-to-end vector is enforced to vary as the protocol proceeds. Thus one can compare the changes in free energy measured by the protocols, manipulating the final position of the last bead of the linear polymer with Flory theory.

To find proper values of free energy change and use fluctuation theorem, repeated experiments are required since we need to evaluate work distribution function. Since the work W is defined as the change in Hamiltonian H,

$$W = \int \delta W = \int dt \frac{d\lambda}{dt}\frac{\partial H_r}{\partial \lambda}\big(\Gamma(t);\lambda(t)\big),$$

the form of H should be found. Dissipative and random forces make the process to find the analytical form cumbersome, but within some approximation[25],

$$H \cong H_k + H_p + \sum C_v(T - T_0),$$

where k and p refer to kinetic (momenta) and potential part of Hamiltonian respectively and C_v is heat capacity of the system. In other words energy that arise from dissipative and random forces can be tied up to single, temperature dependent term. As seen in figure 4, the system temperature T is well-thermostated, then

$$H \cong H_k + H_p \text{ with } T \cong T_0,$$

This means that the net work done by dissipative and random forces can be neglected. However this does not imply that the shape of the work distribution would not be influenced by random and dissipative forces at all, because they distort trajectory $\Gamma(t)$ through the protocol.

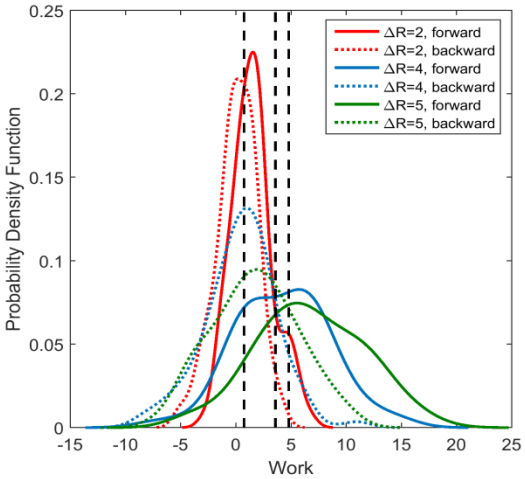

Figure 5. Work distribution function of the spring-bead type linear polymer with initial $\boldsymbol{R = 3}$, sampled from $\mathbf{10^2}$ simulations. For example $\boldsymbol{\Delta R = 2}$, backward curve is the distribution function of work done on the system during the protocol $\boldsymbol{R = 5 \rightarrow 3}$. Dotted black lines refer to intersections of forward and backward distributions.

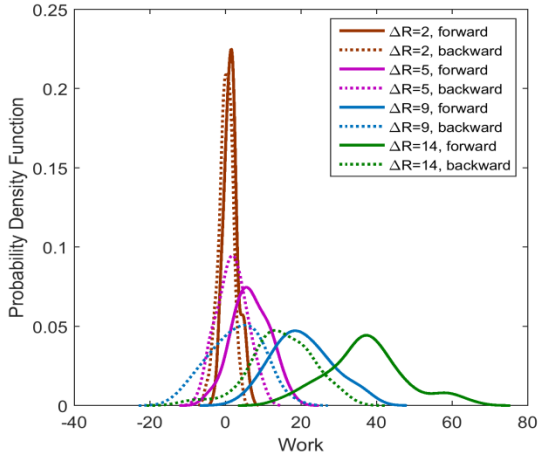

Figure 6. Work distribution function of the linear polymer with initial $\boldsymbol{R = 3}$, sampled from $\mathbf{10^2}$ simulations. As $\boldsymbol{\Delta R}$ increases the deviations of the distributions also rise.

On this basis, work distribution functions of unfolding spring-bead type linear polymer with initial length of end-to-end vector = 3.0 is shown in figure 5 ($\Delta R = 2\sim5$) and 6 ($\Delta R = 2\sim14$). The constants and parameters used in this experiment are given in table A4. Since the shapes of work distributions and the reliability of estimating methods heavily dependent on the speed of the protocols, the pulling velocities $|\boldsymbol{v}|_{pull}$ are fixed as

$$|\boldsymbol{v}|_{pull} = \frac{\Delta R}{\text{steps}} = 1.8 \times 10^{-3}.$$

In spite of equivalent speed of protocols the deviations of work distributions become larger as the length of end-to-end vector increases. Specific reason of this phenomenon originates from the fact that the deviation relies on the phase space trajectories during the protocol[26]. Long journey through the phase space represents large influence of solvent or other particles on the monomers, resulting final quasi-equilibrium phase space broader than that of short trip. In molecular dynamics, once the initial configuration is specified, its final destination in the phase space is settled into narrow region. However in stochastic dynamics the longer the trip is, the larger lapse from original trajectory is.

Another useful tool to test the degree of deviation of work distribution functions is to plot Bennett function, $B(\Delta F) = -B_f(\Delta F) + B_b(\Delta F)$. As mentioned before Bennett suggested the logarithmic likelihood of free energy change $l(\Delta F)$ with given forward and backward work distributions W_i, $\widetilde{W}_i$ as

$$l(\Delta F) = \sum_{i=1}^{K} \ln \frac{1}{1+\exp(-\beta W_i + \beta \Delta F)} + \sum_{i=1}^{K} \ln \frac{1}{1+\exp\left(-\beta \widetilde{W}_i - \beta \Delta F\right)}.$$

We can plot $l(\Delta F)$ for $-\infty < \Delta F < \infty$ to find the point $\Delta F = \Delta F_0$ that maximizes $l(\Delta F)$. If $l(\Delta F)$ is maximized at $\Delta F = \Delta F_0$,

$$\left.\frac{\partial l(\Delta F)}{\partial \Delta F}\right|_{\Delta F_0} = 0 = \sum_{i=1}^{K} \frac{-\beta \exp(-\beta W_i + \beta \Delta F_0)}{1+\exp(-\beta W_i + \beta \Delta F_0)} + \sum_{i=1}^{K} \frac{\beta \exp\left(-\beta \widetilde{W}_i - \beta \Delta F_0\right)}{1+\exp\left(-\beta \widetilde{W}_i - \beta \Delta F_0\right)},$$

$$\sum_{i=1}^{K} \frac{\beta \exp(-\beta W_i + \beta \Delta F_0)}{1+\exp(-\beta W_i + \beta \Delta F_0)} = \sum_{i=1}^{K} \frac{\beta \exp\left(-\beta \widetilde{W}_i - \beta \Delta F_0\right)}{1+\exp\left(-\beta \widetilde{W}_i - \beta \Delta F_0\right)},$$

$$\frac{1}{K}\sum_{i=1}^{K}\frac{1}{1+\exp(\beta W_i-\beta\Delta F_0)} = \frac{1}{K}\sum_{i=1}^{K}\frac{1}{1+\exp\left(\beta\widetilde{W}_i+\beta\Delta F_0\right)},$$

and introducing forward and backward Bennett function,

$$B_f(\Delta F) \overset{\text{def}}{=} \frac{1}{K}\sum_{i=1}^{K}\frac{1}{1+\exp(\beta W_i-\beta\Delta F)},$$

$$B_b(\Delta F) \overset{\text{def}}{=} \frac{1}{K}\sum_{i=1}^{K}\frac{1}{1+\exp\left(\beta\widetilde{W}_i+\beta\Delta F\right)},$$

$$\frac{\partial l(\Delta F)}{\partial \Delta F} = KB(\Delta F) = K\left(-B_f(\Delta F)+B_b(\Delta F)\right),$$

and

$$\left.\frac{\partial l(\Delta F)}{\partial \Delta F}\right|_{\Delta F_0} = KB(\Delta F_0) = K\left(-B_f(\Delta F_0)+B_b(\Delta F_0)\right) = 0,$$

is satisfied[27] at $\Delta F = \Delta F_0$. At this step we can ponder what shape Bennett function would have if two distribution sets $\{W_i\}$ and $\{\widetilde{W}_i\}$ are quite similar and overlapped. With $\Delta F = \Delta F_0$ as a center, $B_f(\Delta F)$ is increasing function from 0 to 1 and $B_b(\Delta F)$ is decreasing function from 1 to 0. Furthermore since $\{W_i\}$ and $\{\widetilde{W}_i\}$ are overlapped, these two functions have point symmetry at $\Delta F = \Delta F_0$. Thus Bennett function $B(\Delta F) = -B_f(\Delta F) + B_b(\Delta F)$ would be Heaviside-like function varying from 1 to -1 and become zero when $\Delta F = \Delta F_0$. Conversely if the overlap between two distributions is small or the shapes of them are different, $B(\Delta F)$ is still varying from 1 to -1 but has no steep curve in the region near $\Delta F = \Delta F_0$. This means that the probability to measure free energy is almost same along that region making precise estimation of change in free energy difficult.

For each protocol figure 7 is the plots of corresponding Bennett functions and figure 8 is the result of Flory-like fitting of free energy measurements,

$$F = k_B TV\frac{N^2}{R^3} + k_B T\frac{R^2}{Nb^2} = \frac{2.209}{R^3} + 0.1283R^2,$$

where F is estimated with Bennett method, $B(\Delta F_0) = 0$. We can confirm that both V and b are positive values. As a result of fitting, the linear polymer obeys Flory's theory quite well. Since the value of excluded volume is turned to be positive, we can go with the deduction that $A_{ij} = 25k_BT$ solvent acts as good solvent. And as ΔR increases, the trajectories of each simulation diverge, causing smoothed-down Bennett functions (purple, green and light blue lines) seen in figure 7.

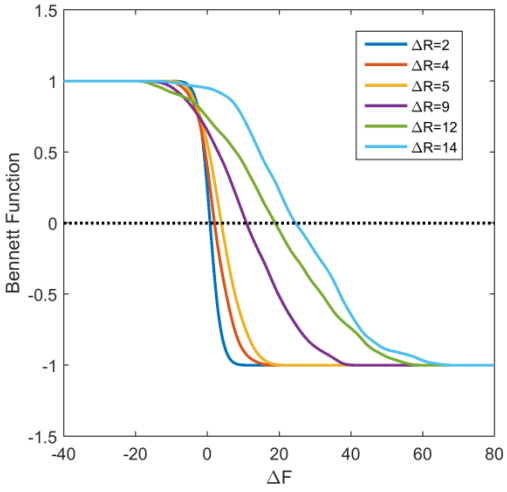

Figure 7. Bennett function of the linear polymer with initial $\boldsymbol{R = 3}$, sampled from $\mathbf{10^2}$ simulations. As $\boldsymbol{\Delta R}$ increases Bennett function becomes deliberate.

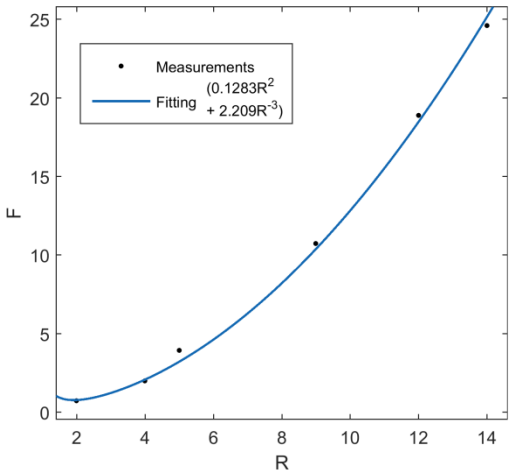

Figure 8. Curve fitting of relationship between end-to-end vector and corresponding free energy of the linear polymer. The quality of the fitting is RMSE=0.5304 and $\boldsymbol{R^2 = 0.9977}$.

D. Comparison of Free Energy Estimators

From the definition of the work in the protocol $\lambda_0 \to \lambda_t$,

$$W = \int \delta W = \int dt\,\frac{d\lambda}{dt}\frac{\partial H_r}{\partial \lambda}\left(\Gamma(t);\lambda(t)\right),$$

work W is dependent on three values, phase space trajectory $\Gamma(0) \rightarrow \Gamma(t)$, initial configuration $\Gamma(0)$, and final configuration $\Gamma(t)$. As proven earlier the distribution of initial configurations is canonical distribution

$$p^{eq}(\Gamma(0)) = \frac{1}{Z_0}\exp\left(-\frac{H(\Gamma(0))}{\theta_0}\right),$$

so that most of the protocols will begin from the phase space points with low $H(\Gamma(0))$. However with figure 1, the final points of the protocols are in quasi-equilibrium and they do not have any distributional rules. The final points are merely determined from the trajectories that they have undergone. If the distances between the trajectories in the phase space are far from each other among the repeated simulations, certainly the final destinations of the protocols diverge even if they start with similar molecular configuration[26,28,29].

In this reason, the shapes of work distribution functions become different if the rates of protocols are not equivalent, though they evolve along the same protocols. If a chain procedure $\lambda_0 \rightarrow \lambda_1 \rightarrow \cdots \rightarrow \lambda_i \rightarrow \lambda_{i+1} \rightarrow \cdots \rightarrow \lambda_t$ emerges fast, relaxation between quasi-equilibriums λ_i is hindered[30]. Thus $\lambda_i \rightarrow \lambda_{i+1}$ process becomes more independent of $\lambda_{i-1} \rightarrow \lambda_i$ making phase space trajectory divagate and so do the shapes of work distribution functions. Aforementioned divergence of trajectories due to the lengths of the protocols (with same initial states and rates of protocols) can be explained in the same manner (figure 7).

For the reliability of free energy estimating methods is low with fast protocols (Bennett function becomes flat near $\Delta F = \Delta F_0$), sufficiently slow procedure is indispensable for evaluating the energy change. However three estimators, Jarzynski, Crooks and Bennett methods have different sensitivity toward the rate of the protocols. With Crooks theorem,

$$P_F[W;\gamma_F] = e^{(W-\Delta F_0)/k_BT}P_R[-W;\gamma_R],$$

dividing both sides of the equation by $1 + e^{(W-\Delta F_0)/k_BT}$ and integrate with W gives

$$\int \frac{P_F[W;\gamma_F]}{1+e^{(W-\Delta F_0)/k_BT}}dW = \int \frac{P_R[W;\gamma_R]}{1+e^{(W+\Delta F_0)/k_BT}}dW,$$

and by discretizing these integrands with discontinuous K simulations results in[15]

$$\sum_{i=1}^{K}\frac{1}{1+\exp(\beta W_i - \beta\Delta F_0)} = \sum_{i=1}^{K}\frac{1}{1+\exp\left(\beta\widetilde{W_i} + \beta\Delta F_0\right)},$$

which is Bennett method. To appraise free energy change with Crooks theorem, one needs to solve $P_F[W;\gamma_F]$ and $P_R[-W;\gamma_R]$ requiring large K. Also if $P_F[W;\gamma_F]$ and $P_R[-W;\gamma_R]$ are so divergent that they do not have intersection point at all, then the free energy estimation fails. In comparison Bennett estimator does not use work distributions directly and can be applied even if distributions do not meet.

On the other hand, by integrating the both sides of Crooks theorem,

$$\begin{aligned}\int P_F[W;\gamma_F]dW &= 1 \\ &= e^{(-\Delta F_0)/k_BT}\int e^{(-W)/k_BT}P_R[W;\gamma_R]dW \\ &= e^{(-\Delta F_0)/k_BT}\langle e^{(-W)/k_BT}\rangle_R,\end{aligned}$$

or

$$\begin{aligned}\int P_R[-W;\gamma_R]dW &= 1 \\ &= e^{(\Delta F_0)/k_BT}\int e^{(-W)/k_BT}P_F[W;\gamma_F]dW \\ &= e^{(\Delta F_0)/k_BT}\langle e^{(-W)/k_BT}\rangle_F,\end{aligned}$$

we can get Jarzynski equality. Jarzynski method also requires $P_F[W;\gamma_F]$ and $P_R[-W;\gamma_R]$ and since it utilizes only one of two distributions, estimation value ΔF_0 is under the control of low W values[30]. For instance if P_F is so dispersed that it has both $W_F < 0$ and $W_F > 0$ samples, and K is not large, the value of $\langle e^{(-W)/k_BT}\rangle_F$ will be significantly dependent on $W_F < 0$ samples due to exponential decaying of $W_F > 0$ samples during averaging process.

Thus with large K and compact work distributions, Jarzynski, Crooks, and Bennett methods have plausible performances. But in opposite circumstances, where work distributions diverge quickly Jarzynski and Crooks methods seem to have limitations.

Meanwhile, again, with Crooks theorem we can prove that

$$\begin{aligned}\frac{P_F[\gamma_F]}{P_R[\gamma_R]} = e^{(W-\Delta F_0)/k_BT} &= \prod_{i=1}^{t}\frac{P\left[\gamma_{i-1}\overset{\lambda_i}{\rightarrow}\gamma_i\right]}{P\left[\gamma_i\overset{\lambda_i}{\rightarrow}\gamma_{i-1}\right]} \\ &= \prod_{i=1}^{t}\frac{\exp\left(-\beta H(\gamma_i,\lambda_i)\right)}{\exp\left(-\beta H(\gamma_{i-1},\lambda_i)\right)} = \exp(-\beta Q),\end{aligned}$$

where Q is total heat supplied by a reservoir[31]. Since $-\beta Q$ is the amount of increased entropy (S_{res}/k_B) of the reservoir,

$$W_{diss} \equiv W - \Delta F_0 = -\beta Q = \frac{S_{res}}{k_B},$$

Free Energy Estimation

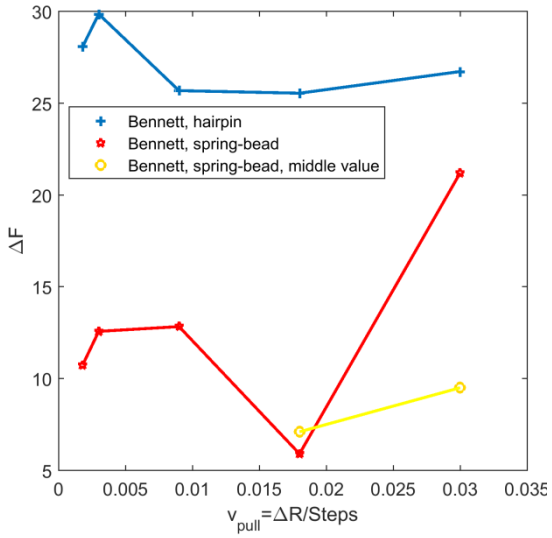

Figure 9. Estimated free energy changes with Bennett method. *Middle value* means that $\boldsymbol{\Delta F_0}$ is evaluated from $\frac{\boldsymbol{\Delta F_1}+\boldsymbol{\Delta F_2}}{\mathbf{2}}$ where $\boldsymbol{B(\Delta F_1) = 0.01, B(\Delta F_2) = -0.01}$.

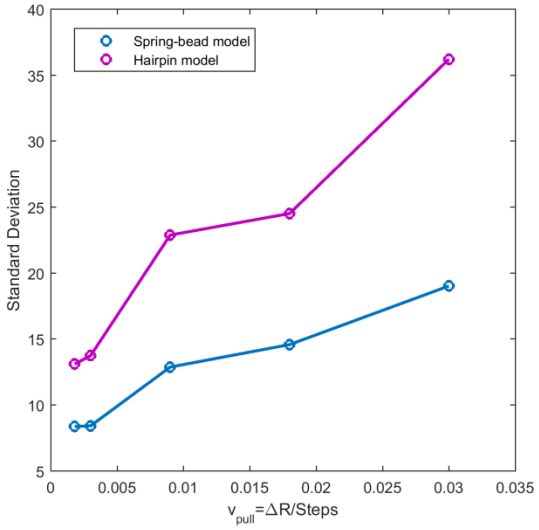

Figure 10. Change in standard deviations of the forward work distributions of the linear polymers. As the rate of the protocols increases, deviations of distributions rise.

where W_{diss} refers to the degree of dissipated works, related with the entropy production[17]. Thus if the distances between ΔF_0 point and work distributions are large, we can think of the protocol to be abruptly proceeded, in other word, irreversible.

To test the reliabilities of three estimators, free energy changes of unfolding-folding of the two types (spring-bead and hairpin types) of linear polymers via same protocols are estimated with these estimators. The positions of the last monomers are varied from $\boldsymbol{r_{N_1}} = L(3,0.5,0.1)$ to $\boldsymbol{r_{N_1}} = L(12,0.5,0.1)$ with different speeds. With $dt = 0.01$ very slow, slow, intermediate (denoted as *inter.*), fast, very fast protocols refer to change of $\Delta x = 9$ during 5000, 3000, 1000, 500 and 300 integration steps respectively. Each protocol is sampled from 2×10^2 simulations. Other parameters used in this experiment are shown in table A5 and A6.

Figure A1 ~ A3 are plots of work distribution and Bennett function of the linear polymer consisting of only harmonic and DPD repulsion forces (spring-bead model) and table 1 is corresponding estimated free energy changes of the unfolding protocols. Figure 12 ~ 14 are plots of work distribution and Bennett function of the linear polymer consisting of not only harmonic and DPD repulsion forces but Lennard-Jones forces at the ends of the polymer (hairpin model) and table 2 is corresponding evaluated free energy changes of unfolding.

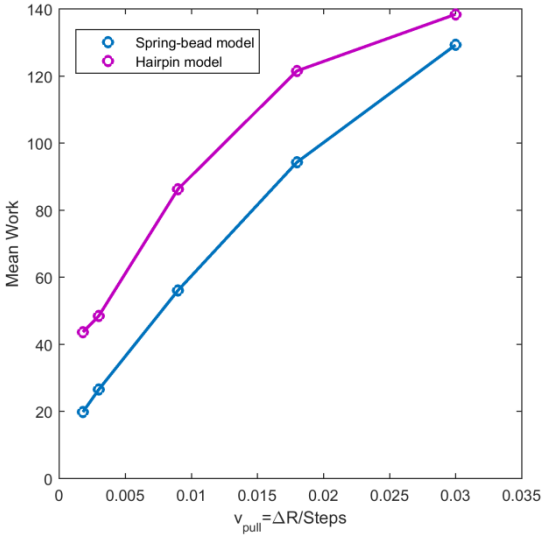

Figure 11. Arithmetic mean values of the forward work samples versus the pulling speeds. Large mean work implies that forward and backward distributions are apart from each other.

On the basis of Crooks and Bennett estimators along the very slow protocol estimated free energy change of unfolding is about 10.73~10.91 (table 1) without Lennard-Jones potential and 26.78~28.73 (table 2) with Lennard-Jones potential. In overall figure 9 shows free energy changes of the both types of the polymers estimated by

Bennett method. Bennett-estimated free energies of the hairpin polymers are definitely larger than those of the spring-bead polymers regardless of pulling velocities. Along the additional attractive forces at the ends, the fact that energy required for unfolding processes becomes larger is closely related with the folding nature of hairpin proteins[32].

As discussed earlier, increasing rates of protocols make work distributions broad and apart from each other (figure 10 and 11; for each work distributions, see figure 12 and A1). Increasing the speed of the protocols, figure 10 shows the degree of deviation of each distribution and figure 11 refers to, roughly, half of the distance between the forward and backward distributions. This increasing phenomenon originates from the fact that the fast protocols scatter phase space trajectories though they emerge from similar initial configurations.

Bennett functions in figure 14 and A3 also agree with aforementioned analytical approaches, showing Heaviside-like property when the rate of the protocol is small and vice versa meaning that at fast protocols the probability to estimate desirable free energy change falls rapidly (flat line near the point $\Delta F = \Delta F_0$). One thing worth to notice is that every Bennett function is intersected at one point making $B(\Delta F_0) = -B_f(\Delta F_0) + B_b(\Delta F_0) = 0$ regardless of the speeds of the protocols. This implies the stability of Bennett estimator.

Comparing figure 12 with A1, the work distributions become more divergent when the works are measured with Lennard-Jones potential. It can be seen in figure 10, for the hairpin polymers have larger standard deviations in the work distributions than the spring-bead polymers. Considering each protocol step $\lambda_i \to \lambda_{i+1}$, it is plausible to expect that the system is able to choose more various paths with the additional force fields so that the volume of final quasi-equilibrium destinations of the protocols increases. Seeing figure 13, though in the limit of $K \to \infty$, work distributions converge toward Gaussian distributions and figure A2 (without Lennard-Jones forces) fits well, the work distributions with Lennard-Jones forces have slight asymmetry toward their tails near $\Delta F = \Delta F_0$. This establishes instability of the trajectories.

In addition the entropy production of the protocol, that is, dissipated works, also increases as the rates go up, since the distributions become apart from ΔF_0 points meaning that $W_{diss} \equiv W - \Delta F_0 \gg 1$ (figure 11; comparing figure 12 and A1). In figure 15 we can see the change of radius of gyration of the polymer with Lennard-Jones attraction over the protocols. In the very fast protocol, radius of gyration rises instantaneously after the protocol begins. This implies the quick breaking of Lennard-Jones bonding between the ends of the polymer. On the other hand, with the very slow protocol, radius of gyration remains still at the very start of the protocol. After some pulling process, with additional 3000 steps, it starts to soak. In this protocol, the attractions are not broken right away but weaken step by step. Thus the slow protocol, which is closer to reversible process than the fast protocol, gives the distributions with small dissipated works.

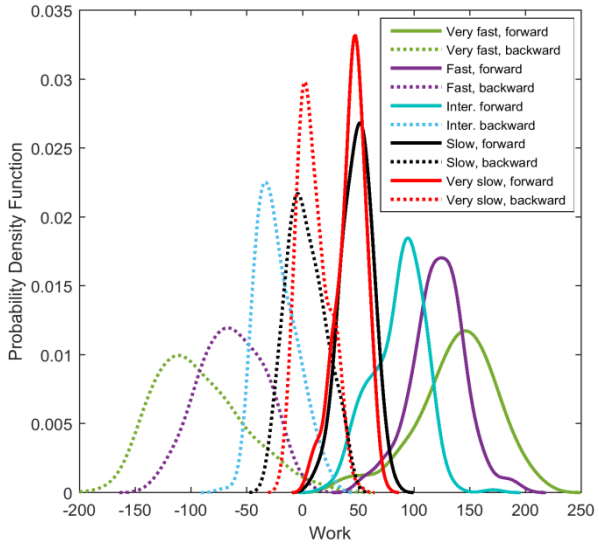

Figure 12. Work distribution of unfolding-folding protocol of the hairpin type polymer (Lennard-Jones forces at the ends).

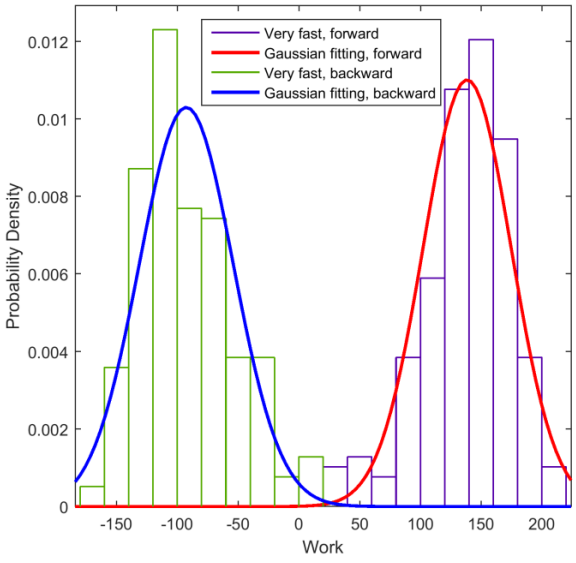

Figure 13. Histogram of measured works in the very fast protocol and its Gaussian fitting of the hairpin type polymer.

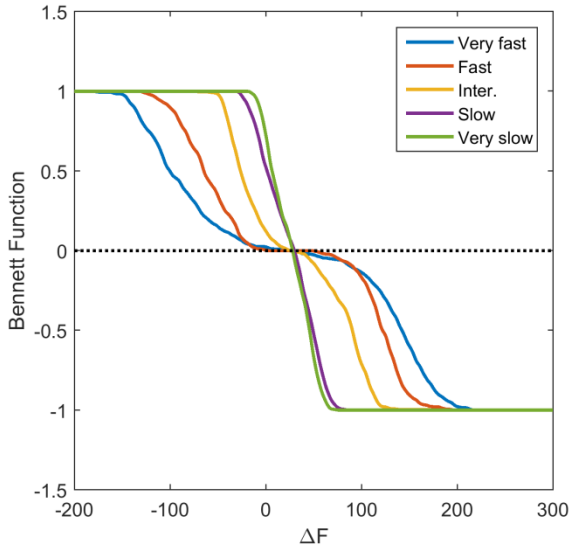

Figure 14. Bennett Function of the hairpin-type polymer.

The estimation values of Jarzynski method has very instable behavior, shown in table 1 and 2. For this phenomenon the stability of the work distributions should be thoroughly specified, since, as already mentioned, the value of Jarzynski estimator $\langle e^{(-W)/k_BT}\rangle_F$ is highly dependent on low W values[33]. In figure A1, the left tail of forward work distribution of the linear polymer over very slow protocol (black line) is stretched over negative work region. These negative work samples cause

$$\langle e^{(-W)/k_BT}\rangle_F \cong \frac{1}{n}\sum_{\substack{i=1\\ W_i<0}}^{n} e^{\frac{(-W_i)}{k_BT}} \gg 1,$$

resulting in

$$e^{(-\Delta F_0)/k_BT} = \langle e^{(-W)/k_BT}\rangle_F \gg 1 \Leftrightarrow \Delta F_0 \ll 0,$$

where n refers to the number of negative work samples. For this reason, Jarzynski estimator gives $\Delta F_0 = -1.12$ in the very slow protocol in table 1. Given that total number of work samples K is small, these equations can be generalized as

$$\langle e^{(-W)/k_BT}\rangle_F \cong \frac{1}{n}\sum_{\substack{i=1\\ W_i\ll\langle W\rangle_a}}^{n} e^{\frac{(-W_i)}{k_BT}},$$

where $\langle W\rangle_a$ is arithmetic average of W. Thus the estimated values of Jarzynski method merely show the whereabouts of the tails of distributions, rather than that of free energies. To cure such a disaster, $n \ll K$ simulations are needed, since then the effect of low W values is canceled out with very small P_F.

Compared to Jarzynski estimator, Crooks method gives highly reliable values of estimated free energy, in table 1 and 2 for this method uses information from both forward and backward processes. Statistical errors of one protocol can be compensated by the opposite protocol. However with fast protocols the estimation performance abruptly shrinks, due to instability of tails of work distributions[30] (which is very crucial for deciding the whereabouts of intersections of distributions - estimated free energy values) or, even nonexistence of intersections of distributions. In figure 13 and A2, for example, the histograms of measured works do not meet resulting in failure to estimate free energy change via Crooks method. In this situation, if we cannot do $K \gg 1$ simulations, we just need to rely on statistical interpolating techniques such as Gaussian and kernel density fitting, dealing with discontinuous data sets. In table 1 and 2 data with parentheses are the estimated values merely on the basis of statistical interpolations, though work histograms do not have any intersections. The values with parentheses are quite reasonable but we cannot guarantee thermodynamic legitimacy of those data since we do not know whether the distributions would meet each other or not when $K \to \infty$. Furthermore in table 1, even the interpolating tactics cannot give estimated free energy for the distance between forward and backward distributions is too large. See corresponding Bennet functions (red line in figure A3.). Free energies in the region $-100 < \Delta F < 100$ have nearly equivalent likelihood to be measured.

However, Bennett estimator can give estimated free energy values regardless of the rates of the protocols. In slow protocol regime, Bennett estimator shows steady performance, in agreement with Crooks estimator. Even with fast protocols the evaluated values do not deviate so much from those with slow protocols, since it uses discretized approach (applicable when $K \ll \infty$) and forward-backward compensation method (canceling out erratic behavior of tails of forward distributions with backward distributions). The main disadvantage of Bennett method is that when the protocol is too fast, the broad region with

$$\left.\frac{\partial B(\Delta F)}{\partial \Delta F}\right|_{\Delta F\sim\Delta F_0}, B(\Delta F)|_{\Delta F\sim\Delta F_0} \cong 0,$$

makes it hard to find ΔF_0 satisfying $B(\Delta F_0) = 0$ (see, broad flatness of green, and red lines in figure A3).

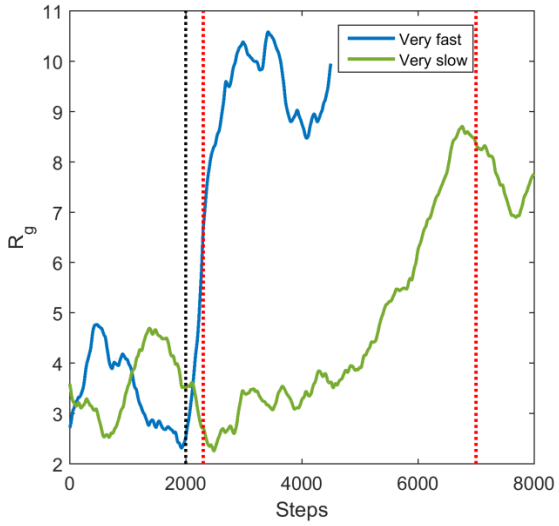

Figure 15. Change of radius of gyration ($\boldsymbol{R_g}$) during the very fast and slow protocols versus additional steps after the equilibrium. Black dotted line indicates the starting step of both protocols and red dotted lines refer to the end steps of them.

Table 1. Estimated Free Energy Values of Unfolding Process in the Spring-Bead Type Polymer**.

	Jarzynski	Crooks -Kernel	Crooks -Gaussian	Bennett
Very Slow	-1.12	10.73	10.91	10.75
Slow	7.73	11.09	12.41	12.58
Inter.	25.23	Impossible (12.88)	Impossible (12.58)	12.84
Fast	61.78	Impossible	Impossible	5.91 7.11***
Very Fast	88.98	Impossible	Impossible	21.21 9.51***

**Kernel* and *Gaussian* refer to probability density fitting using kernel density estimation and normal distribution respectively. Data with parentheses are measured from interpolating curves. ***Since Bennett functions of the fast and very fast protocols are too flat to estimate $\boldsymbol{\Delta F_0}$, $\boldsymbol{\Delta F_0}$ is evaluated from $\frac{\boldsymbol{\Delta F_1+\Delta F_2}}{\boldsymbol{2}}$ where $\boldsymbol{B(\Delta F_1)=0.01, B(\Delta F_2)=-0.01}$.

Table 2. Estimated Free Energy Values of Unfolding Process with the Hairpin-Type Polymer**.

	Jarzynski	Crooks -Kernel	Crooks -Gaussian	Bennett
Very Slow	9.16	28.73	26.78	28.09
Slow	11.91	27.47	27.25	29.86
Inter.	26.82	28.15	25.77	25.69
Fast	56.17	Impossible (30.10)	Impossible (34.89)	25.55
Very Fast	38.69	Impossible (27.47)	Impossible (27.25)	26.73

**Kernel* and *Gaussian* refer to probability density fitting using kernel density estimation and normal distribution respectively. Data with parentheses are measured from interpolating curves.

Conclusion

We derived the applicable form of fluctuation theorems for dissipative particle dynamics (DPD). Probability density distribution of the ensemble of the system within DPD turned out to be that of NVT-system. Furthermore, the ratio between the forward and backward work distribution functions in DPD follows conventional Crooks relation, $P_F[W;\gamma_F] = e^{(W-\Delta F_0)/k_BT}P_R[-W;\gamma_R]$. Corresponding other forms of fluctuation theorem such as Jarzynski and Bennett-forms could be deduced from this result.

In order to match real-world dynamics with soft potentials of DPD, constants and parameters from existing researches concerned with hydrodynamics are used and within these values the system of the solvated linear polymer behaves like Flory theory of the linear polymers with good solvent.

From fluctuation theorems, in order to estimate free energy change of unfolding process of the linear polymer, work distribution functions were plotted via repeated simulations. Since the work distributions severely rely on the trajectories from which they have been taken, they become broad and divergent when the unfolding processes (forward protocols) evolve fast or take long journey over the phase space.

Though Jarzynski, Crooks, and Bennett estimation methods used the same work distribution data, each had its own inborn weakness. With small number of simulations the performance of Jarzynski estimator was very poor since it highly depends on the degree of stability of tails in the distributions. Compensating this erratic behavior with backward protocols, Crooks estimator gave a lot more reasonable estimation values. However, when the speed of the protocol is sufficiently rapid, the reliability of this method was severely diminished. Among three estimators

Bennett estimator showed the best estimating stability since it uses discretized approach (applicable when $K \ll \infty$) and forward-backward compensation method (canceling out erratic behavior of tails of forward distributions with backward distributions).

Acknowledgments. This work has been supported by the project EDISON (EDucation-research Integration through Simulation On the Net), Chemistry. Subin Bang acknowledges a lot of advices from Computational Nano-Bio Chemistry Laboratory in Department of Chemistry, Seoul National University.

References

1. Kinjo, T., & Hyodo, S. A., Equation of motion for coarse-grained simulation based on microscopic description, Physical Review E, 75(5), 051109, 2007.
2. Pastor, R. W., Brooks, B. R., & Szabo, A., An analysis of the accuracy of Langevin and molecular dynamics algorithms, Molecular Physics, 65(6), 1409-1419, 1988.
3. Espanol, P., & Warren, P., Statistical mechanics of dissipative particle dynamics, Europhysics Letters, 30(4), 191, 1995.
4. Groot, R. D., & Warren, P. B., Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation, Journal of Chemical Physics, 107(11), 4423, 1997.
5. Espanol, P., & Revenga, M., Smoothed dissipative particle dynamics, Physical Review E, 67(2), 026705, 2003.
6. Spenley, N. A, Scaling laws for polymers in dissipative particle dynamics. Europhysics Letters, 49(4), 534, 2000.
7. Lei, H., Caswell, B., & Karniadakis, G. E., Direct construction of mesoscopic models from microscopic simulations, Physical Review E, 81(2), 026704, 2010.
8. Bennett, C. H., Efficient estimation of free energy differences from Monte Carlo data, Journal of Computational Physics, 22(2), 245-268, 1976.
9. Crooks, G. E., Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences, Physical Review E, 60(3), 2721, 1999.
10. Jarzynski, C., Nonequilibrium equality for free energy differences, Physical Review Letters, 78(14), 2690, 1997.
11. Kurchan, J., Fluctuation theorem for stochastic dynamics, Journal of Physics A: Mathematical and General, 31(16), 3719, 1998.
12. Seifert, U., Stochastic thermodynamics, fluctuation theorems and molecular machines, Reports on Progress in Physics, 75(12), 126001, 2012.
13. Strobl, G., The physics of polymers: concepts for understanding their structure and behavior, 2nd ed., Springer-Verlag, Berlin, 1997.
14. Irbäck, A., Peterson, C., Potthast, F., & Sommelius, O., Local interactions and protein folding: A three-dimensional off-lattice approach, The Journal of Chemical Physics, 107(1), 273-282, 1997.
15. Kim, S., Kim, Y. W., Talkner, P., & Yi, J., Comparison of free-energy estimators and their dependence on dissipated work, Physical Review E, 86(4), 041130, 2012.
16. Moeendarbary, E., Ng, T. Y., & Zangeneh, M., Dissipative particle dynamics: introduction, methodology and complex fluid applications—a review, International Journal of Applied Mechanics, 1(04), 737-763, 2009.
17. Jarzynski, C., Equalities and inequalities: irreversibility and the second law of thermodynamics at the nanoscale, Annual Review of Condensed Matter Physics, 2(1), 329-351, 2011.
18. Marsh, C. A., & Coveney, P. V., Detailed balance and H-theorems for dissipative particle dynamics, Journal of Physics A: Mathematical and General, 31(31), 6561, 1998.
19. Marsh, C. A., Backx, G., & Ernst, M. H., Static and dynamic properties of dissipative particle dynamics, Physical Review E, 56(2), 1676, 1997.
20. Pérez-Madrid, A., & Santamaría-Holek, I., Fluctuation theorems for systems under Fokker-Planck dynamics, Physical Review E, 79(1), 011101, 2009.
21. Spenley, N. A. Scaling laws for polymers in dissipative particle dynamics, Europhysics Letters, 49(4), 534, 2000.
22. Jamali, S., Boromand, A., Khani, S., Wagner, J., Yamanoi, M., & Maia, J., Generalized mapping of multi-body dissipative particle dynamics onto fluid compressibility and the Flory-Huggins theory, The Journal of chemical physics, 142(16), 164902, 2015.
23. Paul J. Flory, Principles of polymer chemistry, Cornell University Press, 2015.
24. Doi, M., Soft Matter Physics, Oxford University Press, 2013.
25. Li, Z., Tang, Y. H., Lei, H., Caswell, B., & Karniadakis, G. E., Energy-conserving dissipative particle dynamics with temperature-dependent properties, Journal of Computational Physics, 265, 113-127, 2014.
26. Crooks, G. E., Path-ensemble averages in systems driven far from equilibrium, Physical review E, 61(3), 2361, 2000.
27. Shirts, M. R., Bair, E., Hooker, G., & Pande, V. S., Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods, Physical review letters, 91(14), 140601, 2003.
28. Crooks, G. E., Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences, Physical Review E, 60(3), 2721, 1999.
29. Mazonka, O., & Jarzynski, C., Exactly solvable model illustrating far-from-equilibrium predictions, arXiv preprint cond-mat/9912121, 1999.
30. Liphardt, J., Dumont, S., Smith, S. B., Tinoco, I., & Bustamante, C., Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality, Science, 296(5574), 1832-1835, 2002.
31. Crooks, G. E., Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems., Journal of Statistical Physics, 90(5-6), 1481-1487, 1998.
32. Clementi, C., Vendruscolo, M., Maritan, A., & Domany, E., Folding Lennard-Jones proteins by a contact potential, Proteins: Structure, Function, and Bioinformatics, 37(4), 544-553, 1999.
33. Jarzynski, C., Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach, Physical Review E, 56(5), 5018.1, 2007.

Appendix

Table A1. Reduced Units Used in this Article[**].

Quantity	Unit
Energy	k_BT=1
Distance	r_c=1
Mass	$m_{solvent}$=1
Time	$\sqrt{m_{solvent}r_c^2/kT}$=1

[**]For specific meaning of symbols, see table A2.

Table A2. Meaning of Parameters and Constants Controlled over the Simulations.

Parameters /Constants	Meaning
Steps	Number of integrations of equations of motion
$N_{solvent}$	Number of solvent particles
N_{pol}	Number of monomers (beads) in the polymer
N_{phil}	Total number of monomers with Lennard Jones potential (LJ)
$N_{phil,side}$	Number of monomers with LJ at one end of the polymer
$N_{phil} - N_{phil,side}$	Number of monomers with LJ at the other end of the polymer
$m_{solvent}$	Mass of solvent particles
m_{pol}	Mass of monomers
L	Diameter of simulation box with periodic boundary conditions
ΔR	Length of unfolded end-to-end vector
r_c	Cut-off radius of $\omega^{C,D,R}$
dt	Time advance of equations of motion within 1 integration
σ	Magnitude of random force
T	Kinetic temperature
A_{ij}	Magnitude of DPD repulsion
K	Magnitude of harmonic force
ϵ	Magnitude of LJ force

Table A3. Values of Parameters and Constants Used in the Simulations Measuring Radius of Gyration of the Spring-Bead Type Polymer.

Parameters/Constants	Values
Steps	70000
$N_{solvent}$	24000
N_{pol}	5, 10, 20, 25, 30, 50
N_{phil}	0
$N_{phil,side}$	0
$N_{phil} - N_{phil,side}$	0
$m_{solvent}$	1.0
m_{pol}	1.0
L	20
ΔR	-
r_c	1.0
dt	0.05
σ	3.35
T	1.0
A_{ij}	25.0
K	2.0
ϵ	0.0

Table A4. Values of Parameters and Constants Used in the Simulations Measuring Fee Energy Changes over the Lengths of End-To-End Vectors of the Spring-Bead Type Polymer.

Parameters/Constants	Values
Steps	-
$N_{solvent}$	16000
N_{pol}	30
N_{phil}	0
$N_{phil,side}$	0
$N_{phil} - N_{phil,side}$	0
$m_{solvent}$	1.0
m_{pol}	5.0
L	30
ΔR	2, 4, 5, 9 , 12, 14
r_c	1.0
dt	0.01
σ	3.0
T	1.0
A_{ij}	25.0
K	4.0
ϵ	0.0

Table A5. Values of Parameters and Constants Used in the Simulations Measuring Free Energy of the Spring-Bead Type Linear Polymer over the Rates of the Protocols.

Parameters/Constants	Values
Steps	-
$N_{solvent}$	16000
N_{pol}	30
N_{phil}	0
$N_{phil,side}$	0
$N_{phil} - N_{phil,side}$	0
$m_{solvent}$	1.0
m_{pol}	5.0
L	30
ΔR	9 (3 to 12)
r_c	1.0
dt	0.01
σ	3.0
T	1.0
A_{ij}	25.0
K	4.0
ϵ	0.0

Table A6. Values of Parameters and Constants Used in the Simulations Measuring Free Energy of the Hairpin Type Polymer over the Rates of the Protocols.

Parameters/Constants	Values
Steps	-
$N_{solvent}$	16000
N_{pol}	30
N_{phil}	20
$N_{phil,side}$	10
$N_{phil} - N_{phil,side}$	10
$m_{solvent}$	1.0
m_{pol}	5.0
L	30
ΔR	9 (3 to 12)
r_c	1.0
dt	0.01
σ	3.0
T	1.0
A_{ij}	25.0
K	4.0
ϵ	2.0

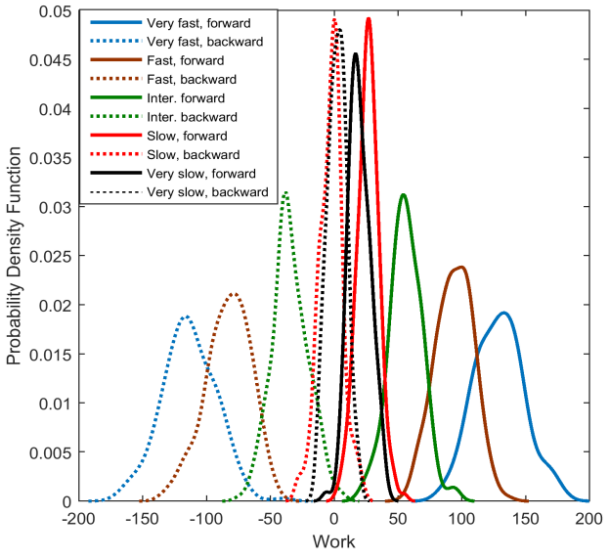

Figure A1. Work distribution of unfolding-folding protocol of the spring-bead type polymer.

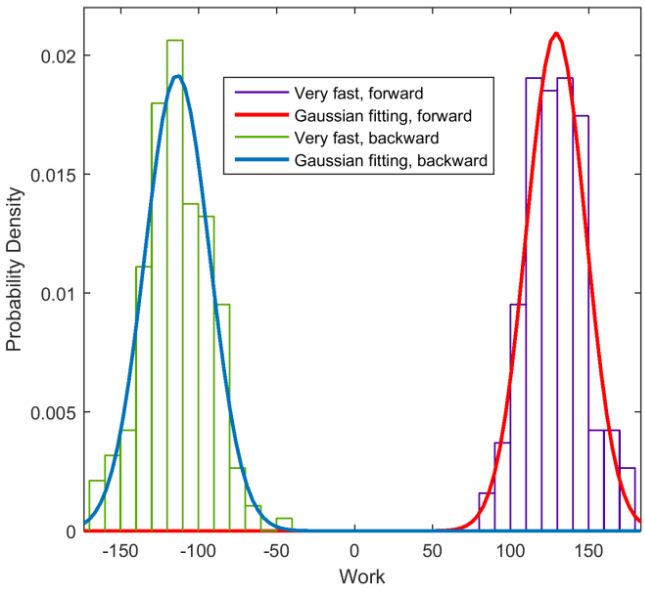

Figure A2. Histogram of measured works and its Gaussian fitting of the spring-bead type polymer.

Free Energy Estimation

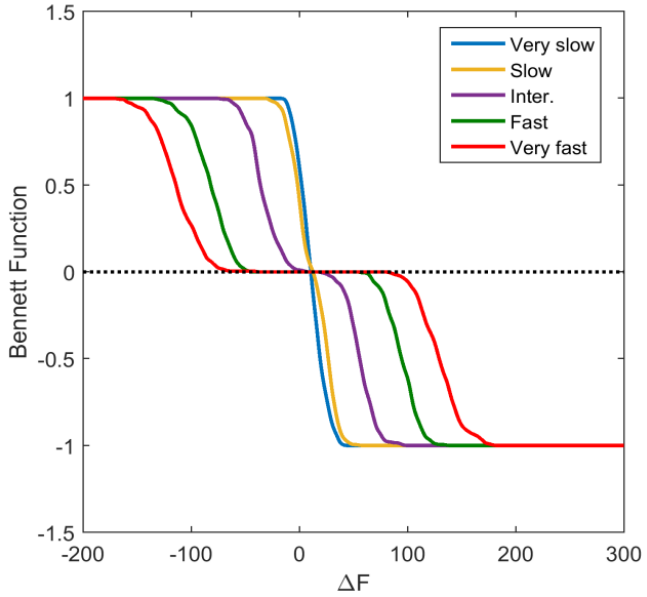

Figure A3. Bennett Function of the spring-bead type polymer.