# Reconstructing the Free-Energy Landscape of a Mechanically Unfolded Model Protein

Alberto Imparato*

*Dipartimento di Fisica, INFN Sezione di Torino, CNISM-Sezione di Torino, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*

Stefano Luccioli and Alessandro Torcini

*Istituto dei Sistemi Complessi, CNR, via Madonna del Piano 10, 50019 Sesto Fiorentino, Italy*
*INFN Sezione di Firenze, via Sansone, 1 - 50019 Sesto Fiorentino, Italy*

(Received 11 May 2007; published 15 October 2007)

The equilibrium free-energy landscape of an off-lattice model protein as a function of an internal (reaction) coordinate is reconstructed from out-of-equilibrium mechanical unfolding manipulations. This task is accomplished via two independent methods: by employing an extended version of the Jarzynski equality (EJE) and the protein inherent structures (ISs). In a range of temperatures around the ''folding transition'' we find a good quantitative agreement between the free energies obtained via EJE and IS approaches. This indicates that the two methodologies are consistent and able to reproduce equilibrium properties of the examined system. Moreover, for the studied model the structural transitions induced by pulling can be related to thermodynamical aspects of folding.

DOI: 10.1103/PhysRevLett.99.168101 PACS numbers: 87.15.Aa, 05.90.+m, 82.37.Rs

The properties of the (free) energy landscape can heavily influence the dynamical and thermodynamical features of a large class of systems: supercooled liquids, glasses, atomic clusters, and biomolecules [1]. In particular, the shape of the landscape plays a major role in determining the folding properties of proteins [2]. A fruitful approach to the analysis of the landscape relies on the identification of the local minima of the potential energy, i.e., the ''inherent structures'' (ISs) of the system [3]. The investigation of the ISs has led to the identification of the structural-arrest temperature in glasses [4] and supercooled liquids [5]. Furthermore, this analysis has been extended also to the study of proteins [1,6–10].

Mechanical unfolding of single biomolecules represents a powerful technique to extract information on their internal structure as well as on their unfolding and refolding pathways [11]. However, mechanical unfolding of biomolecules is an out-of-equilibrium process: unfolding events occur on time scales much shorter than the typical relaxation time of the molecule towards equilibrium. Nonetheless, by using the equality introduced by Jarzynski [12], the free energy of mechanically manipulated biomolecules can be recovered as a function of an externally controlled parameter [13].

In this Letter, we reconstruct the equilibrium free-energy landscape (FEL) associated to a mesoscopic off-lattice protein model as a function of an internal coordinate of the system (namely, the end-to-end distance $\zeta$). At variance with previous studies [14–16], here we exploit two independent methods: one based on an extended version of the Jarzynski equality (EJE) and the other on thermodynamical averages over ISs. Moreover, the agreement of the results obtained with the two approaches indicates that these two methodologies can be fruitfully integrated to provide complementary information on the protein landscape. In particular, the investigation of the ISs allows us to give an estimate of the (free) energetic and entropic barriers.

The model studied in this Letter is a modified version of the 3D off-lattice model introduced in Ref. [17] and successively generalized to include a harmonic interaction between next-neighboring beads instead of rigid bonds [6,18]. The model consists of a chain of 46 pointlike monomers mimicking the residues of a polypeptidic chain, where each residue is of one of the three types: hydrophobic ($B$), polar ($P$), and neutral ($N$) ones.

The residues within the protein interact via an off-lattice coarse-grained potential composed of four terms: a stiff nearest-neighbor harmonic potential intended to maintain the bond distance almost constant, a three-body bending interaction associated to the bond angles, a four-body interaction mimicking the torsion effects, and a long-range Lennard-Jones potential reproducing in an effective way the solvent mediated interactions between pairs of residues noncovalently bonded [19]. The 46-mer sequence $B_9N_3(PB)_4N_3B_9N_3(PB)_5P$, which exhibits a four stranded $\beta$-barrel native configuration (NC), is here analyzed with the same potential and parameter set reported in Ref. [18]. This sequence has been previously studied, for different choices of the potential parameters, in the context of spontaneous folding [6,17,18,20,21] as well as of mechanical unfolding and refolding [22,23]. The NC is stabilized by the attractive hydrophobic interactions among the $B$ residues; in particular, the first and third $B_9$ strands, forming the core of the NC, are parallel to each other and antiparallel to the second and fourth strand, namely, $(PB)_4$ and $(PB)_5P$. The latter strands are exposed towards the exterior due to the presence of polar residues.

The main thermodynamic features can be summarized with reference to three different transition temperatures [1,7,10,24]: the $\theta$ temperature $T_\theta$ discriminating between

 

phases dominated by random-coil configurations rather than collapsed ones, the folding temperature $T_f$, below which the protein stays predominantly in the native valley, and the glassy temperature $T_g$ indicating the freezing of large conformational rearrangements [9]. Following the procedures reported in Ref. [24], we have determined these temperatures and obtained $T_\theta = 0.65(1)$, $T_f = 0.26(1)$, and $T_g = 0.12(2)$. These values are in good agreement with those reported in [7,10], where $T_f$ and $T_g$ have been identified via different protocols.

In order to mimic the mechanical pulling of the protein attached to a cantilever of an atomic force microscope, or trapped in optical tweezers, while one extremum of the chain was kept fixed, the last bead was attached to a pulling device with a spring of elastic constant $\kappa$. The external force is applied at time $t = 0$ by moving the device along a fixed direction with a constant velocity protocol $z(t) = z(0) + v_p t$. The protein is initially rotated to have the first and last bead aligned along the pulling direction; therefore, the external potential reads $U_{z(t)}(\zeta) = \kappa[z(t) - \zeta]^2/2$. Moreover, to reproduce the experimental conditions, the thermalization procedure consists of two steps: a first stage when the protein evolves freely starting from the NC, followed by a second one in presence of the pulling apparatus. The resulting configuration is employed as the initial state for the forced unfolding performed at constant temperature via low friction Langevin molecular dynamics (MD).

Following Ref. [14], we briefly review how to reconstruct the equilibrium FEL as a function of the collective coordinate $\zeta$ starting from out-of-equilibrium measurements. Let the system (unperturbed) Hamiltonian $H_0(x)$ be a function of the positions and momenta of the residues $x = \{\boldsymbol{r}_i, \boldsymbol{p}_i\}$; the free energy of the *constrained* ensemble, characterized by a given value $\zeta$ of the macroscopic observable $\zeta(x)$, reads $\beta f(\zeta) = -\ln \int \mathrm{d}x \delta[\zeta - \zeta(x)] \times \exp[-\beta H_0(x)]$. The system is driven out of equilibrium by the external potential, $U_{z(t)}(\zeta)$, and the work *done* on the system by the external force associated to $U_{z(t)}(\zeta)$ is $W_t = \int_0^t \mathrm{d}\tau v_p \kappa[z(\tau) - \zeta(x(\tau))]$. Because of thermal fluctuations the trajectory $x(t)$ followed by the system, and therefore $W_t$, varies between one realization of the manipulation process and the other. As discussed in Ref. [25], an extended version of the Jarzynski equality relates $f(\zeta)$ to the work done on the system, for arbitrary external potential. Such a relation reads

$$\langle \delta(\zeta - \zeta(x)) \mathrm{e}^{-\beta W} \rangle_t = \mathrm{e}^{-\beta[f(\zeta) + U_{z(t)}(\zeta)]}/Z_0, \qquad (1)$$

where $Z_0 = \int \mathrm{d}x \exp[-\beta H_0(x)]$ and the average $\langle \cdot \rangle_t$ is performed over different trajectories with fixed time length $t$; for implementation details, see [14,15].

As shown in Fig. 1, the FEL estimated by using Eq. (1) collapses into an asymptotic curve as the pulling velocity decreases in agreement with the results reported in [14,16]. Let us now discuss, by referring to Fig. 1, the structural transitions (STs) induced by the pulling. As shown in the inset, the asymptotic $f(\zeta)$ profile exhibits a clear minimum in correspondence of the end-to-end distance of the NC (namely, $\zeta_0 \sim 1.9$). Moreover, up to $\zeta \sim 6$, the protein remains in nativelike configurations characterized by a $\beta$-barrel made up of 4 strands, while the escape from the native valley is signaled by the small dip at $\zeta \sim 6$, and it is indicated in the inset of Fig. 1 as ST1. This ST has been recently analyzed in [23] in terms of the potential energy of ISs. For higher $\zeta$ the configurations are characterized by an almost intact core (made of 3 strands) plus a stretched tail corresponding to the pulled fourth strand [see Figs. 2(b) and 2(c)]. The second ST amounts to pull the strand $(PB)_5P$ out of the barrel. In order to do this, it is necessary to break 22 hydrophobic links [26], amounting to an energy cost ~21. The corresponding free-energy barrier height is instead quite lower [=11(1), as estimated from Fig. 1]. Since the potential energy barrier is essentially due to the hydrophobic interactions, this implies that a non-negligible entropic cost is associated to ST2. Instead, in the range $13 < \zeta < 18.5$ the curve $f(\zeta)$ appears as essentially flat, thus indicating that almost no work is needed to completely stretch the tail once detached from the barrel. The pulling of the third strand (that is part of the core of the NC) leads to a definitive destabilization of the $\beta$ barrel and to the breakdown of the remaining 36 $BB$ links with an energetic cost ~35. A finite entropic barrier should be associated also to this final stage of the unfolding (termed ST3), because the energy increase due to the hydrophobic terms is much higher than the free-energy barrier [=26(2), see ST3 in Fig. 1]. The second plateau in $f(\zeta)$ corresponds to protein structures made up of a single strand [similar to

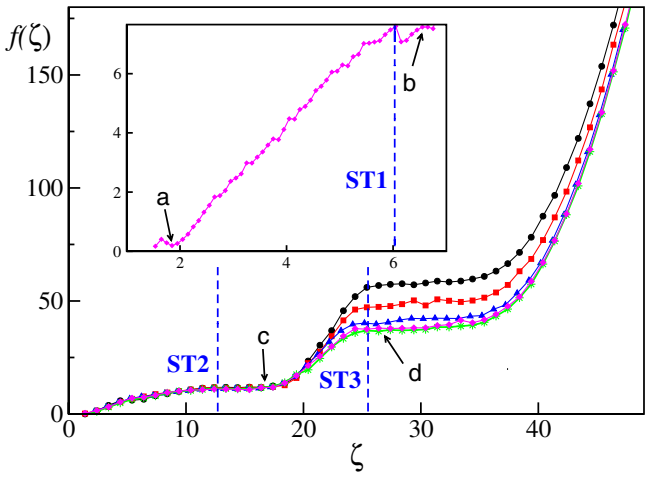


FIG. 1 (color online). Free-energy profile $f$ as a function of the end-to-end distance $\zeta$, obtained by Eq. (1) for various pulling velocities: from top to bottom $v_p = 5 \times 10^{-2}$, $1 \times 10^{-2}$, $5 \times 10^{-3}$, $5 \times 10^{-4}$, and $2 \times 10^{-4}$. In the inset, an enlargement of the curve for $v_p = 5 \times 10^{-4}$ at low $\zeta$ is reported. Each curve has been obtained by averaging over 160–240 repetitions of the same pulling protocol at $T = 0.3$. The letters refer to the $f(\zeta)$ values for the configurations shown in Fig. 2 and the (blue) dashed lines to the STs' locations.

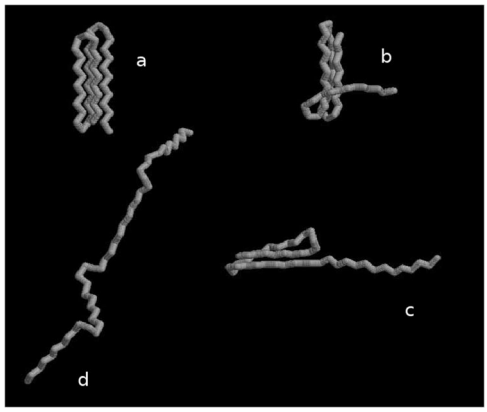


FIG. 2. Pulled configurations at $T = 0.3$: the NC (a) has $\zeta_0 \sim 1.9$; the others are characterized by $\zeta = 6.8$ (b), $\zeta = 16.8$ (c), and $\zeta = 27.1$ (d).

Fig. 2(d)]. The final quadratic rise of $f(\zeta)$ for $\zeta \geq 36$ is associated to the stretching of bond angles and distances beyond their equilibrium values.

As shown in Fig. 3, the FEL is strongly affected by temperature variations. In particular, for temperatures around $T_f$ one still observes a clear minimum around $\zeta_0$ and a FEL resembling the one found for $T = 0.3$. A native-like minimum is still observable for $T = 0.5 < T_\theta$; however, its position $\zeta > \zeta_0$ indicates that the NC is no longer the most favorite configuration. Furthermore, the dip around $\zeta \sim 6$–$7$ disappears and the heights of the two other barriers reduce. By approaching $T_\theta$, the minimum broadens noticeably and the first barrier almost disappears, thus suggesting that 4-stranded $\beta$-barrel configurations coexist with partially unfolded ones. Above $T_\theta$ only one barrier remains, and the absolute minimum is now associated to extended conformations similar to type (b) or (c) in Fig. 2 with some residual barrel structure.

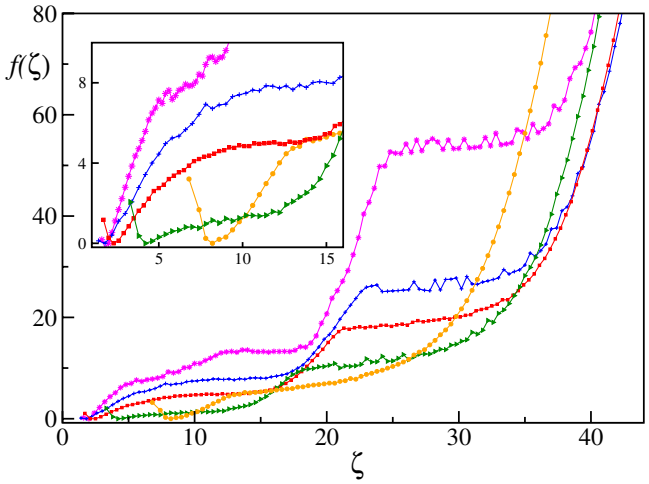


FIG. 3 (color online). Free-energy profile $f(\zeta)$ as obtained by Eq. (1) for various temperatures: $T = 0.2$ (magenta stars), 0.4 (blue plus signs), 0.5 (red squares), 0.6 (green triangles), and 0.7 (orange circles). In the inset an enlargement is reported at small $\zeta$. The data refer to $v_p = 5 \times 10^{-4}$.

Let us now introduce the reconstruction of the free energy in terms of the inherent states (ISs). ISs correspond to local minima of the potential energy; in particular, the phase space visited by the protein during its dynamical evolution can be decomposed in disjoint attraction basins, each corresponding to a specific IS [1,3]. In the context of the superposition approximation [1], the free energy can be expressed as a sum over the basins of attraction:

$$e^{-\beta f_{IS}} = \sum_a e^{-\beta(V_a + R_a)} \simeq \sum_a e^{-\beta V_a} \prod_{j=1}^{3N-6} (T/\omega_a^j), \quad (2)$$

where $a$ labels distinct IS and $V_a$ (respectively $R_a$) is the corresponding potential (respectively vibrational free) energy. $R_a$ represents an entropic contribution due to the fluctuations around the considered minimum and is analytically estimated by assuming a harmonic basin of attraction in terms of the $3N - 6$ nonzero frequencies $\{\omega_a^j\}$ of the vibrational modes [1]. For the examined model, the harmonic superposition approximation works reasonably well up to $T < 0.6$, as shown in [10]. We have built up two data banks of ISs: the thermal data bank (TDB) obtained by performing equilibrium canonical simulations and the pulling data bank (PDB) by mechanically unfolding the protein [27].

In order to estimate the FEL $f_{IS}(\zeta)$ as a function of the variable $\zeta$ characterizing different ISs, the sum in (2) should be restricted to ISs with an end-to-end distance within a narrow interval $[\zeta; \zeta + d\zeta]$ [9]. As shown in Fig. 4, the comparison between $f_{IS}(\zeta)$ and the $f(\zeta)$ obtained via the EJE reconstruction in proximity of $T_f$ reveals an almost complete coincidence up to $\zeta \sim 5$, while for

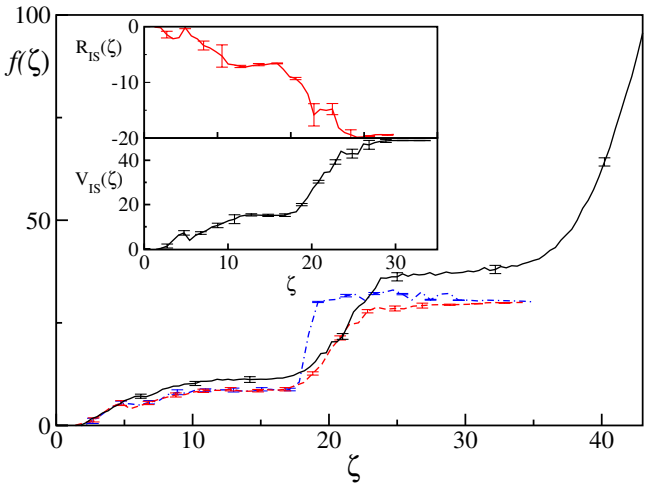


FIG. 4 (color online). Free-energy profiles $f(\zeta)$ and $f_{IS}(\zeta)$ as a function of the elongation $\zeta$ for $T = 0.3$. The black solid line refers to the reconstruction in terms of the EJE, while the red dashed one corresponds to $f_{IS}$ for a set of pulling experiments with $v_p = 2 \times 10^{-4}$. The blue dot-dashed line is the $f_{IS}(\zeta)$ obtained in terms of the ISs of the TDB. In the insets are reported the reconstructed $V_{IS}(\zeta)$ (lower panel) and $R_{IS}(\zeta)$ (upper panel) by employing ISs in the PDB.

larger $\zeta$, $f_{\mathrm{IS}}(\zeta)$ slightly underestimates the free energy. We think that, at least for $T < 0.6$, this discrepancy could be noticeably reduced by including in the IS analysis the saddles of the potential. A further comparison among the IS reconstructions obtained via the TDB and PDB reveals an almost perfect coincidence up to $\zeta \sim 17$. The two $f_{\mathrm{IS}}$ differ only during the last stage of the unfolding: the TDB FEL is steeper with respect to the PDB one, thus suggesting that the protein can reach lower energy states with large $\zeta$ during mechanical unfolding, states that have a low probability to be visited during the dynamics at thermal equilibrium. However, the value of the barrier to overcome and that of the final plateau are essentially the same. The IS conformation with the maximal end-to-end distance is the *all-trans* configuration [23] corresponding to $\zeta_{\mathrm{trans}} = 35.70$; therefore, the IS approach does not allow to evaluate the FEL for $\zeta > \zeta_{\mathrm{trans}}$. However, the IS analysis provides us an estimate of the profiles of the potential and vibrational free energies $V_{\mathrm{IS}}(\zeta)$ and $R_{\mathrm{IS}}(\zeta)$, respectively. From the latter quantity, the entropic costs associated to the unfolding stages can be estimated. As shown in the inset of Fig. 4 for $T = 0.3$ the unfolding stages previously described correspond to clear ''entropic'' barriers. In particular, in order to stretch the protein from the NC to the *all-trans* configuration the decrease of $R_{\mathrm{IS}}(\zeta)$ is 19(1), in agreement with the previous estimate obtained by considering the EJE reconstruction of the FEL.

Finally, one can try to put in correspondence the three unfolding stages previously discussed with thermodynamical aspects of the protein folding. In particular, by considering the energy profile $V_{\mathrm{IS}}(\zeta)$, an energy barrier $\Delta V_{\mathrm{IS}}$ and a typical transition temperature $T_t = (2\Delta V_{\mathrm{IS}})/(3N)$ [28] can be associated to each of the STs. The first transition ST1 corresponds to a barrier $\Delta V_{\mathrm{IS}} = 8(1)$ and therefore to $T_t = 0.11(1)$, that, within error bars, essentially coincide with $T_g$. For the ST2 transition the barrier to overcome is $\Delta V_{\mathrm{IS}} = 16(1)$ and this is associated to a temperature $T_t = 0.23(2)$ (slightly below $T_f$). The energetic cost to completely stretch the protein is 50(1) with a transition temperature $T_t = 0.72(1)$ that is not too far from the $\theta$ temperature. At least for this specific model, our results indicate that the observed STs induced by pulling can be put in direct relationship with the thermal transitions usually identified for the folding or unfolding process.

We conclude by noticing that the information on the equilibrium FEL obtained with both the EJE and the IS methodologies is consistent and gives substantiated hints about the thermal unfolding. Moreover, the publication of the first experimental FEL reconstruction obtained via the EJE for a titin I27 domain [29] paves the way to a verification of our findings for a real biomolecule by complementing experimental data with steered MD simulations [30].

Useful discussions with L. Delfini, S. Lepri, and L. Peliti are acknowledged, as well as partial support by the European Contract No. 12835 - EMBIO.

*alberto.imparato@polito.it

[1] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, England, 2003).
[2] P. E. Leopold *et al.*, Proc. Natl. Acad. Sci. U.S.A. **89**, 8721 (1992).
[3] F. H. Stillinger and T. A. Weber, Science **225**, 983 (1984).
[4] S. Sastry *et al.*, Nature (London) **393**, 554 (1998).
[5] L. Angelani *et al.*, Phys. Rev. Lett. **85**, 5356 (2000).
[6] R. S. Berry *et al.*, Proc. Natl. Acad. Sci. U.S.A. **94**, 9520 (1997).
[7] D. A. Evans and D. J. Wales, J. Chem. Phys. **118**, 3891 (2003).
[8] A. Baumketner *et al.*, Phys. Rev. E **67**, 011912 (2003).
[9] N. Nakagawa and M. Peyrard, Proc. Natl. Acad. Sci. U.S.A. **103**, 5279 (2006); Phys. Rev. E **74**, 041916 (2006).
[10] J. Kim and T. Keyes, J. Phys. Chem. B **111**, 2647 (2007).
[11] M. Carrion-Vazquez *et al.*, Proc. Natl. Acad. Sci. U.S.A. **96**, 3694 (1999); B. Onoa *et al.*, Science **299**, 1892 (2003).
[12] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997); G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).
[13] D. Collin *et al.*, Nature (London) **437**, 231 (2005).
[14] A. Imparato and L. Peliti, J. Stat. Mech. (2006) P03005.
[15] O. Braun *et al.*, Phys. Rev. Lett. **93**, 158105 (2004).
[16] A. Imparato *et al.*, Phys. Rev. Lett. **98**, 148102 (2007).
[17] J. D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **87**, 3526 (1990).
[18] T. Veitshans *et al.*, Folding Des. **2**, 1 (1997).
[19] All quantities are here expressed in adimensional units for a comparison with physical units; see [18].
[20] Z. Guo and C. L. Brooks III, Biopolymers **42**, 745 (1997).
[21] J. Kim *et al.*, Phys. Rev. Lett. **97**, 050601 (2006).
[22] F.-Y. Li *et al.*, Phys. Rev. E **63**, 021905 (2001).
[23] D. J. Lacks, Biophys. J. **88**, 3494 (2005).
[24] A. Torcini *et al.*, J. Biol. Phys. **27**, 181 (2001); L. Bongini *et al.*, Phys. Rev. E **68**, 061111 (2003).
[25] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).
[26] We assume that a hydrophobic contact is formed between two $B$ residues if their distance is smaller than 1.25.
[27] In order to find the different ISs the equilibrium (respectively out-of-equilibrium) Langevin trajectory is sampled at constant time intervals $\delta t = 5$ (respectively at constant elongation increments $\delta\zeta = 0.1$) to pinpoint a series of configurations, which afterwards are relaxed via a steepest descent dynamics. For mechanical unfolding, the protein is unblocked and the pulling apparatus removed before the relaxation stage. The TDB contains ~600 000 distinct ISs collected via equilibrium simulations at various temperatures in the range [0.3–2.0]. The PDB contains 5000–50 000 ISs for each examined temperature. The implementation of the EJE method requires at each $T$, for 100 repetitions of the protocol at $v_p = 2 \times 10^{-4}$, roughly 20 days of CPU on a Opteron AMD64 at 2.0 Ghz and the realization of the corresponding PDB additional 5 days of CPU, while the CPU cost to obtain the whole TDB has been of 80 days.
[28] Each transition temperature $T_t$ has been estimated by equating the kinetic energy to the barrier value $\Delta V_{\mathrm{IS}}$.
[29] N. C. Harris *et al.*, Phys. Rev. Lett. **99**, 068101 (2007).
[30] H. Lu *et al.*, Biophys. J. **75**, 662 (1998).