Extremely rare ultra-fast non-equilibrium processes close to equilibrium: RNA unfolding and refolding

Peter Werner and Alexander K. Hartmann
Institut für Physik, Universität Oldenburg, 26111 Oldenburg, Germany

We study numerically the behavior of RNA secondary structures (or single-stranded DNA) under influence of a varying external force. This allows to measure the work W during the resulting fast unfolding and refolding processes. Here, we investigate a medium-size hairpin structure. Using a sophisticated large-deviation algorithm, we are able to measure work distributions with high precision down to probabilities as small as 10^{-46}. Due to this precision and by comparison with exact free-energy calculations we are able to verify the theorems of Crooks and Jarzynski. Furthermore, we analyze force-extension curves and the configurations of the secondary structures during unfolding and refolding for typical equilibrium processes and non-equilibrium processes, conditioned to selected values of the measured work W, typical and rare ones. We find that the non-equilibrium processes where the work values are close to those which are most relevant for applying Crooks and Jarzynski theorems, respectively, are most and quite similar to the equilibrium processes. Thus, a similarity of equilibrium and non-equilibrium behavior with respect to a mere scalar variable, which occurs with a very small probability but can be generated in a controlled but non-targeted way, is related to a high similarity for the set of configurations sampled along the full dynamical trajectory.

I. INTRODUCTION

In Statistical Physics, the cleanest and so far best-justified description is obtained for systems in equilibrium. Nevertheless, due to open system boundaries and lack of infinite time to perform experiments or simulations, most real and simulated model systems are constantly in non-equilibrium. A bridge between both worlds is provided. e.g., by the theorems of Jarzynski [1] and Crooks [2], where the distribution $P(W)$ of work W is measured for arbitrary fast non-equilibrium processes obtained from sampling equilibrium initial configurations and possibly stochastic non-equilibrium trajectories. Correspondingly $P_{\mathrm{rev}}(W)$ is the distribution for the reverse process. For a system coupled to a heat bath, Crooks theorem reads $P(W) = P_{\mathrm{rev}}(-W)\exp(-(\Delta F - W)/T)$. This can be used to reconstruct the true free energy difference ΔF between initial and final state, because $P(W)$ and $P_{\mathrm{rev}}(-W)$ cross at $W = \Delta F$. Correspondingly the equation of Jarzynski reads $\langle e^{-W/T}\rangle = e^{-\Delta F/T}$. These and related theorems have led to many applications and extensions relating equilibrium and non-equilibrium processes [3–10]. A fruitful field of applications is biophysics, where these theorems are used to measure properties of small molecules like RNA.

One major goal of stochastic thermodynamics is to extract equilibrium information from non-equilibrium measurements or simulations [11]. The fluctuation theorems concern specific measurable scalar quantities like work [1, 2, 12], entropy [13–21], or a quantity measuring the volume of the phase space [22]. However, beyond statistics of particular scalar quantities, the fluctuation theorems do not provide information about the behavior along a corresponding equilibrium trajectory, i.e., with respect to arbitrary measurable quantities. Standard derivations of the fluctuations theorems only involve terms which include energies and probabilities of the initial and final state. What may we expect when we analyze the full trajectory of a non-equilibrium process? First, a *typical*, i.e., highly probable sample of a non-equilibrium trajectory will look very different from a corresponding trajectory sampled during an equilibrium process. Second, it is known that when reweighting trajectories suitably in a time-dependent way, they also carry some information about the intermediate states which would occur for the same values of the control parameter [21, 23, 24] which allows for the reconstruction of full free-energy profiles beyond initial and final state. Third, it is somehow intuitive to believe that the *rare* non-equilibrium processes which contribute most to the estimation of ΔF are in a comprehensive way, without reweighting, similar or even equal to the corresponding equilibrium processes. For the case of the theorems of Crooks and Jarzynski, the statistics of the work distributions are most relevant for particular work values $W = \Delta F$ and $W = W_{\mathrm{J}}^*$, where the latter one is the value where the integrand $e^{-W/T}P(W)$ exhibits a maximum. Note that these values are highly improbable to occur for large system sizes. On the other hand, beyond this intuition, there is no solid reason that these rare possibly very fast processes completely resemble true equilibrium processes: A non-equilibrium process always depends on the history, i.e., on many configurations encountered so far, while each equilibrium state in a process does not depend at all on the history. In particular, non-equilibrium processes depend on the speed of performance, while the equilibrium is for infinitely low speed.

This motivates our present work: We investigate in a comprehensive way the dynamics of fast non-equilibrium processes conditioned to various non-equilibrium work values W, typical and rare ones, and compare with the equilibrium process behavior. For this purpose we want to study a model which exhibits a complex low temperature landscape, includes the possibility of performing external work and, at the same time, allows for an efficient

sampling of configurations in equilibrium. Therefore, we have chosen unfolding and refolding of RNA secondary structures subject to an external force [25]. The former one, denoted as *forward* process, involves stretching an RNA by subjecting it to an external force f which is increased from starting at zero. For the latter one, denoted as *reverse* process, one starts with a large force and reduces it to zero. For small RNAs consisting of few dozens of bases, Crooks theorem has been confirmed in experiments [26, 27] and simulations [12, 28] for slow unfolding and refolding processes. For such small RNA and slow processes, the resulting work distributions are rather broad and the distribution for forward and reverse processes are close to each other such that they cross at high-probability values which are easily accessible. For larger RNA molecules, the crossing points will move to smaller probabilities, such that the crossing can not be observed in experiments or standard simulations. To go beyond such limiting system sizes, we applied for our study sophisticated large-deviation algorithms [29, 30], which allowed us to measure probability distributions numerically down to extremely small probabilities. These algorithms have also applied successfully to non-equilibrium processes like the transition-path sampling approach to study protein folding [31, 32], population-based approaches to study asymmetric exclusion processes [33, 34] or Markov-chain Monte Carlo methods to investigate, e.g., traffic models [35] and the Kardar-Parisi-Zhang equation [36]. In particular such an algorithm has also been applied to measure with high precision the work distribution of an Ising model subject to a varying external field [37], providing the first confirmation of the theorems of Jarzynski and Crooks for a large system with many thousands of particles.

Thus, here we will provide similar evidence for RNA secondary structure unfolding and refolding by applying such a rare-event algorithm, allowing us to obtain the work distributions of intermediate-sized RNAs down to probabilities as small as 10^{-46}. Furthermore, we will analyze the temporal structure of the non-equilibrium processes, conditioned to the occurring work values W. We will compare this to the corresponding equilibrium process, which can be sampled exactly [38–40] and efficiently, i.e., in polynomial time, for RNA secondary structures without pseudo-knots. Beyond confirming the theorems of Jarzynski and Crooks we find in particular that the non-equilibrium processes can be very similar in their development to the equilibrium ones. The highest similarity is reached for processes which exhibit a work value W *between* the values $W = \Delta F$ and $W = W_{\mathrm{J}}^*$ which are most relevant for the Crooks and Jarzynski theorem, respectively.

We will next present our model and the used simulation methods. Then we show our results and finish by a discussion.

II. MODEL

Each RNA molecule is a linear chain $\mathcal{R} = (r_i)_{i=1,\dots,L}$ of bases, also called residues, with $r_i \in \{\mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{U}\}$ and L is the length of the sequence. For a given sequence $\mathcal{R}$ of bases, a *secondary structure* is a set of pairs of bases, such that for the present simple model only *complementary* (Watson-Crick) base pairs A-U and C-G are allowed. This can be described by a set $\mathcal{S}$ of pairs (i, j) (with the convention $1 \le i < j \le L$), meaning that bases r_i and r_j are paired. For convenience, we also use $s(i) = j$ if i is paired to j, which implies $s(j) = i$, and $s(i) = 0$ if i is not paired. Our restriction to Watson-Crick pairs means for A-U either r_i =A and r_j =U or vice versa, correspondingly for the C-G pair.

Two restrictions are used: (i) We exclude so called *pseudo-knots*, that means, for any $(i, j), (i', j') \in \mathcal{S}$ with $i < i'$, either $i < j < i' < j'$ or $i < i' < j' < j$ must hold. In the first case, pair (i, j) is located entirely before (i', j') in the sequence. In the second case, pair (i', j') is called to be *inside* of (i, j). If a bond is not inside of any other bonds, we say it occurs on the *first level*, i.e., it is the topmost pair of the structure enclosing all other pairs. Neglecting pseudo knots follows the notion of them being more an element of the tertiary structure [41]. It also means that it is always possible to draw the molecule as a single line and connect all pairs by lines such that no intersections occur. (ii) Du to the bending rigidity of the molecule, between two paired bases a minimum distance is required, i.e. $|j - i| > s$.

Every secondary structure $\mathcal{S}$ is assigned a certain energy $E(\mathcal{S})$, where the dependency on the sequence $\mathcal{R}$is not explicitly indicated. This energy is defined by assigning each pair (i, j) a certain energy $e(r_i, r_j)$ depending only on the kind of bases.

Furthermore there is a contribution arising from the external force f which stretches the chain to its *extension* $n = n(\mathcal{S})$, as introduced previously [25]. The *extension* of the structure is the part of the RNA which is outside any paired base, plus length 2 for any paired base on the first level. Hence, any globule in the chain contributes two length units. This is illustrated in Fig. 1. This interaction with the external force f gives rise to an energy contribution $-f \times n$.

The total energy for the most basic model is the sum over all pairs plus the interaction with the external force

$$E(\mathcal{S}) = \sum_{(i,j)\in\mathcal{S}} e(r_i, r_j) - n(\mathcal{S})\, f\,. \tag{1}$$

By choosing $e(r, r') = +\infty$ for non-complementary bases r and r', pairings of this kind are suppressed. Here we use the most simple energy model, i.e., $e(r, r') = -1$ for complementary bases A-U and C-G. Note that in this simple form our model can also be seen as a model for single-stranded DNA. When aiming at modelling real RNA as realistically as possible, one would have to use a more sophisticated model. For such cases, stacking energies

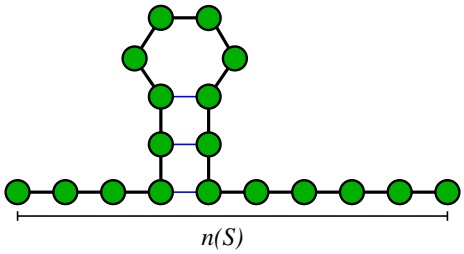

FIG. 1. (color online) An example for a RNA secondary structure with one globule and a line indicating the extension $n(\mathcal{S})$ of the folded RNA. Circles denote bases, thick black lines links between consecutive bases, and thin blue lines hydrogen bonds between complementary bases.

have to be included and also, when it comes to free energy calculations, entropic effects related to steric constraints that are originating from different types of loops and multiloops. Software packages like *mfold* [42], *Sfold* [43], *RNAstructure* [44] or the *Vienna RNA Package* [40] exist for such detailed modelling in general. For our present work, where our main aim is to investigate the relationship between equilibrium and rare non-equilibrium behavior, restricted to qualitative relevance to real systems, the present simple model is sufficient.

III. ALGORITHMS

In the following, we discuss the algorithms we have used. First, we show how RNA secondary structures can be sampled directly in equilibrium. For the non-equilibrium folding and unfolding simulations, our dynamics consist of local steps performed within Markov chain Monte Carlo (MC) simulations of RNA structures, as presented in the second subsection. Thus, one time unit of the simulation is one MC sweep. To study rare unfolding and folding trajectories, we used a second type of MC simulations, which is wrapped around the structure MC simulations, as presented in the third subsection.

A. Sampling secondary structures

For RNA secondary structures it is possible to sample them directly in equilibrium for finite temperatures T in time $O(L^3)$. We used an extension of the approach for the zero-force case [38]. For this purpose, one needs also to calculate partition functions for some sub sequences, which is possible using *dynamic programming* in polynomial time. These approaches [25, 45] are also extensions of the zero-force case method [46].

The partition function $Z_{i,j}$ $(i \le j)$ for sub sequence $r_i \dots r_j$ at inverse temperature $\beta = 1/T$ without external force and without length constraints, obeying the minimum distance s between two paired bases, is given by

$$\begin{aligned} Z_{i,j} &= 1 \qquad \text{for} \quad j-i \le s \\ Z_{i,j} &= Z_{i,j-1} \\ &\quad + \sum_{k=i}^{j-s-1} Z_{i,k-1} e^{-\beta e(r_k,r_j)} Z_{k+1,j-1} \qquad \text{else} \end{aligned} \tag{2}$$

All $O(L^2)$ values of $Z_{i,j}$ can be conveniently calculated [46] by a *dynamic programming approach*, i.e. starting with $Z_{i,i}$ and continuing with increasing values of $j-i$. Since most contributions involve a sum of $O(L)$ terms, the algorithm has a running time of $O(L^3)$.

In order to include the interaction with the external force, one needs additionally the partition function $Q_{1,j,n}$ of the sub sequence $r_1 \dots r_j$ such that the extension is fixed to the value n, with $n \le j$. We include the fixed index 1 for matching with the notation for $Z_{i,j}$,

Our approach follows the lines of corresponding methods [25, 45] for calculation of partition functions and ground state energies of RNA secondary structures subject to an external force. The partition function reads :

$$\begin{aligned} Q_{1,1,1} &= 1 \\ Q_{1,j,1} &= 0 \qquad \text{for} \quad j > 1, \\ Q_{1,2,2} &= Z_{1,2} \\ Q_{1,j,2} &= 0 \qquad \text{for} \quad 2 < j \le s+1 \\ Q_{1,j,2} &= e^{-\beta e(r_1,r_j)} Z_{2,j-1} \qquad \text{for} \quad j > s+1 \\ Q_{1,j,n} &= Q_{1,j-1,n-1} + \qquad \text{for} \quad n > 2, j \ge n \\ &\quad \sum_{k=n-1}^{j-s-1} Q_{1,k-1,n-2} e^{-\beta e(r_k,r_j)} Z_{k+1,j-1} \,. \end{aligned} \tag{3}$$

Also all these partition functions can be conveniently calculated by dynamic programming in time $O(L^3)$.

This allows us to calculate the partition function with force for sub sequence $r_1, \dots, r_j$ by

$$\tilde{Z}_{1,j}(f) = \sum_{n=1}^{j} Q_{1,j,n} e^{\beta n f} \,. \tag{4}$$

Note that the case $n = 0$ can not occur and the case $n = 1$ corresponds only to one single base.

The availability of the above partition functions allows us to sample secondary structures in the presence of an external force directly, i.e. rejection free, also in polynomial time. The approach is an extension of the zero-force algorithm [38] to the case $f \ge 0$.

For sampling a structure, the following probabilities are needed. The probability $p^p_{i,j,k}$ that for sub sequence $r_i, \dots, r_j$, without the presence or influence of a force, base j is paired to base k with $i \le k < j$ and $j-k > s$ is given by

$$p^p_{i,j,k} = \frac{Z_{i,k-1} e^{-\beta e(r_k,r_j)} Z_{k+1,j-1}}{Z_{i,j}} \,. \tag{5}$$

For $j-k \leq s$, this probability is zero. The probability that base j is not paired is given by

$$p^u_{i,j} = \frac{Z_{i,j-1}}{Z_{i,j}} \,. \tag{6}$$

The probability $\tilde{p}^p_{1,j,k}(f)$ that for sub sequence $r_1, \ldots, r_j$, with the presence of a force f, base j is paired to base k with $1 \leq k < j$ and $j-k > s$ is given by

$$\tilde{p}^p_{1,j,k}(f) = \frac{\tilde{Z}_{1,k-1}(f) e^{-\beta e(r_k,r_j)+\beta 2f} Z_{k+1,j-1}}{\tilde{Z}_{1,j}(f)} \tag{7}$$

For $j-k \leq s$, this probability is zero. The probability that base j is not paired is given by

$$\tilde{p}^u_{1,j}(f) = \frac{\tilde{Z}_{1,j-1}(f) e^{\beta f}}{\tilde{Z}_{1,j}(f)}. \tag{8}$$

The sampling of a structure is now performed as follows. Each time one starts for the full sequence $r_1, \ldots, r_L$ by considering the *case with force* f:

- *Case with force* f for sub sequence $r_1, \ldots, r_j$

  Base j is paired to one of the bases $k = 1, \ldots, j-s-1$ with probability $\tilde{p}^p_{1,j,k}(f)$, respectively, and remains unpaired with probability $\tilde{p}^u_{1,j}(f)$.

  Now, if base j has been paired to base k, recursively the sequence $r_1, \ldots, r_{k-1}$ is treated in the same way (*case with force* f) and the sub sequence $r_{k+1}, \ldots, r_{j-1}$ is treated as described in the *case without force*.

  If base j has not been paired, the sequence $r_1, \ldots, r_{j-1}$ is treated in the same way (*case with force* f).

- *Case without force* for sub sequence $r_i, \ldots, r_j$

  Base j is paired to one of the bases $k = i, \ldots, j-s-1$ with probability $p^p_{i,j,k}$, respectively, and remains unpaired with probability $p^u_{i,j}$.

  Now, if base j has been paired to base k, recursively the sequence $r_i, \ldots, r_{k-1}$ and $r_{k+1}, \ldots, r_{j-1}$ are treated in the same way (*case without force*).

  If base j has not been paired, the sequence $r_i, \ldots, r_{j-1}$ is treated in the same way (*case without force*).

In this way, each time a structure is independently drawn according to the Boltzmann distribution, i.e., the algorithm constitutes ideal sampling.

When we sample a folding or an unfolding trajectory in equilibrium, i.e., for a sequence of force values $f_k = f_0 + k\Delta f$ $(k = 0, 1, 2, \ldots)$, we just sample an equilibrium structure $\mathcal{S}_k$ for all force values $\{f_k\}$ encountered. Here we use $f_k \in [0, 2]$ and 400 different force values, i.e. $\Delta f = \pm 0.005$. To each trajectory a work of $W = -\sum_k n(\mathcal{S}_k)\Delta f$ is associated, corresponding to a small force increment and a subsequent imaginary (infinite) long waiting time until the next equilibrium structure is encountered. Since we use very small force increments, the work we measure is actually numerically very close to ΔF, as we have verified.

B. Folding and Unfolding Algorithm

The algorithm for performing an unfolding or refolding process in non-equilibrium , and to measure the performed work W for a given sequence $\mathcal{R}$, works as follows: First, an initial secondary structure $\mathcal{S}$ is drawn in equilibrium at some given initial value f_0 of the force and for RNA temperature T. Then a Monte Carlo (MC) simulation allowing to change the secondary structure with total of $n_{\rm MC}$ sweeps is performed while the force parameter f is increased or reduced depending on Δf. For the unfolding process we used $f_0 = 0$ and increased the work until $f = 2$ was reached, while for the refolding process we started at $f_0 = 2$ and decreased the force to $f = 0$. During the MC simulation, $n_{\rm force}$ times the force is increased by Δf. Each time the force is changed, we obtained a contribution $\Delta W = -n(\mathcal{S})\Delta f$ to the work, where $n(\mathcal{S})$ is the current extension.

The MC sweeps allow for the influence of thermal fluctuations. Since the number of possible Watson-Crick pairs is $O(L^2)$, we define one sweep as $L^2/2$ Monte Carlo steps. For the individual MC steps, each time two random residues i and j are selected. If these are already paired to each other in the current structure $\mathcal{S}$, a trial configuration $\mathcal{S}'$ is made by removing the pair, i.e., the bond is broken. In case of two non-bonded bases, they will be paired in the trial configuration $\mathcal{S}$' if they are complementary, and if they have a distance larger than s, and if no pseudo-knots would be created. The configuration is not changed when just one of the selected bases is already bounded, since a base can only connect to a single other one. For these cases, the trial configuration is accepted, i.e. becomes the current one, with the usual Metropolis probability $p_{\rm Metr} = \min\{1, \exp(-\beta\Delta E)\}$ determined by the energy change $\Delta E = E(\mathcal{S}') - E(\mathcal{S})$. The random numbers which are used during the MC simulation are generated before a call to the subroutine and stored in a vector ξ. In this way, all the randomness is removed outside this subroutine [47], for a reason we will present in the next section. Note that all other parameters like $\mathcal{R}$, T etc. remain the same during a simulation, thus the work obtained during unfolding or refolding is a deterministic function of ξ:

algorithm $W(\xi)$

```
begin
  draw for R an equilibrium structure S at
    initial force f_0 and RNA temperature T
  f = f_0
  W = 0
  for j = 0, ··· , n_force
  begin
    perform L^2 n_MC/(2 n_force) MC-Steps:
    begin
      select two random residues l, m ∈ {1, ..., L}
      if (l, m) ∈ S, remove pair with prob. p_Metr.
      else if (l, m) is allowed set S = S ∪ {(l, m)}
              with prob. p_Metr.
    end
    f = f + Δf
    W = W − n(S)Δf
  end
  return(W)
end
```

The vector $\xi = (\xi_1, \xi_2, \ldots, \xi_K)$ contains $K = L - 1 + 3L^2 n_{\mathrm{MC}}/2$ random numbers which are uniformly distributed in $[0, 1]$. These are all random numbers that are needed to perform one full unfolding or refolding simulation. Each random number has a specific fixed purpose. The first $L-1$ entries are required to sample an configuration from the partition function, where an individual random number is utilized to determine if base $j \in [2, \ldots, L]$ is either connected to base $k \in [1, \ldots, j-s-1]$ or unconnected. Not all of these $L-1$ random numbers are necessarily used during a specific sampling process, e.g., if for base j the remaining sub sequence for a potential pairing partner is too small. In this case, the corresponding random number is just ignored, The subsequent MC steps need three random numbers each, two for selecting a pair and potentially one more, if the Metropolis criterion is used. If not, the third random number is also ignored, respectively. This results in a number of $3L^2 n_{\mathrm{MC}}/2$ additional entries in ξ.

Note that more efficient Monte Carlo algorithms for RNA secondary structures exists [48, 49], which are event-driven Gillespie algorithms. Also they take as possible Monte Carlo moves only allowed moves into account, i.e., either pairs are removed, or only allowed pairs are proposed, avoiding non-complementary base pairs or pseudo knots. This requires keeping track of the allowed moves, which also generates quite some overhead in computation and it also involves the calculation of necessary corrections factors due to the varying number of accessible neighboring secondary structure configurations, in order to guarantee detailed balance. Also, the Gillespie nature of these algorithms make the use of random numbers dependent on the history of previous events. Nevertheless, for the present application, the work process is embedded into another higher-level Monte-Carlo simulation, see below. For a good performance of the higher-level MC simulation this requires that for each entry of the vector a specific purpose is assigned, as presented above. If this requirement is met, small changes to ξ yield typically small, i.e., not too "chaotic" changes in the resulting work $W = W(\xi)$. This is the case with the present algorithm.

C. Large-deviation approach

By repeating an unfolding or refolding simulation many times, one can measure approximately the work distributions $P(W)$ and $P_{\mathrm{rev}}(W)$, respectively. Nevertheless, this *simple sampling approach* allows one only to obtain the work distributions down to rather large probabilities, like 10^{-9}. To obtain the work distributions down to much smaller probabilities, we applied sophisticated large-deviation algorithms [29, 30]. Our approach has already been used to measure work distributions for large Ising systems [37]. The basic idea is to drive the forward and reverse processes, respectively, by vectors ξ of random numbers and control the composition of the vectors with a Markov chain Monte Carlo simulation, with a known, i.e. removable, bias depending on the measured work.

As mentioned in the previous section, for a given sequence $\mathcal{R}$, temperature T and the other parameters, which are all kept fixed for a set of simulations, the outcome of the unfolding or refolding process is solely determined by the random values contained in the vector ξ. Thus, to perform a standard *simple sampling* simulation, each time a random vector ξ is drawn with all its entries being a pseudo random number uniformly distributed in $[0, 1]$. This results in one work value W which is sampled from the true distribution. Thus, if one repeats the simple sampling many times, one can collect many work values and calculate a histogram to approximate the full distribution. Nevertheless, running the simple sampling K times, will one only allow to resolve probabilities larger or equal to $1/K$ in the histogram.

In order to access the work distribution down to very small probabilities, the following is done: A Markov chain Monte Carlo (MCMC) simulation is employed, where the states of the simulation are represented by samples $\xi^{(t)}$ of the random vectors that drive the RNA unfolding or folding simulations. Thus, each state of the Markov chain corresponds to exactly one instance of a full process consisting of starting with an initial state in equilibrium and performing a, typically fast, non-equilibrium process during which the force is changed. The system has a bit of time to relax between two force changes. In the end, a work value $W = W(\xi^{(t)})$ is obtained. Therefore, the MCMC simulation takes place on a higher level than the unfolding or refolding simulations. Now, the main idea is to include a bias in the MCMC simulation, which involves a Metropolis acceptance depending on the change in the resulting work.

To be more precise, say we have the current state $\xi^{(t)}$ with work $W^{(t)} = W(\xi^{(t)})$ in the MCMC simulation. First, a *trial state* ξ' is generated, by copying $\xi^{(t)}$ and then redrawing a number $n_\xi < K$ of randomly se-

lected entries from the K entries of ξ'. Next, a complete work process is performed for ξ', which results in the measured work $W' = W(\xi')$. The trial state is then accepted, i.e., $\xi^{(t+1)} = \xi'$ with Metropolis probability $\tilde{p}_{\mathrm{Metr}} = \min\{1, \exp(-\Delta W/\Theta)\}$, where $\Delta W = W' - W^{(t)}$ is the change in work and Θ is a temperature-like control parameter. Otherwise, the trial state is rejected, i.e., $\xi^{(t+1)} = \xi^{(t)}$. Note that an empirical acceptance rate of around 0.5 is aimed for, such that n_ξ is typically small for small values of Θ and larger for larger values of Θ. Actual values are given below.

Since the setup of the MCMC simulation is like any standard MCMC approach for a system coupled to a heat bath, only that the energy is replaced by the work and Θ is used for the temperature, it is obvious that our approach will sample the true work distribution but including a bias which is exactly the Boltzmann factor $\sim \exp(-W/\Theta)$. As usual, the initial phase of the Markov chain, i.e., the equilibration phase, is discarded and sample values are drawn only at suitable large time intervals. Thus, one can in principle perform simulations for a given value of Θ, measure a histogram approximating the biased distribution $P_\Theta(W) \sim P(W)\exp(-W/\Theta)$ and obtain an estimate for the true distribution $P(W)$ by multiplication with $\exp(+W/\Theta)$, up to a normalization constant. Note that, technically, to resolve the distribution over a large range of the support, one needs to perform simulations at several suitably chosen values of the control temperature Θ, get the normlization constants for all measured histograms and combine them into one single finally normalized histogram [29]. Details, in particular for the case of the work distribution of on Ising model in an external field, can be found elsewhere [37]. This approach has already been applied to other non-equilibrium processes like the Kardar-Parisi-Zhang model [36] or traffic flows [35].

Note finally, that if one aims at using a more realistic (free) energy model for the RNA calculations, it would not be possible to just use existing packages like *mfold* [42], *Sfold* [43], *RNAstructure* [44], or the *Vienna RNA Package* [40] because they use random number generators internally and do not allow for feeding in vectors ξ of numbers to be used.

IV. RESULTS

An RNA sequence is considered, which is not too small, such that differences between equilibrium and non-equilibrium secondary structure configurations can be observed with suitable resolution. Concretely, we studied a hairpin structure of length $L = 100$ which has the sequence $(\mathrm{AC})^{25}(\mathrm{UG})^{25}$, resulting in a ground state of one large stack with a small loop. This sequence is choosen because hairpins are common secondary structure elements of RNA, which have been studied frequently experimentally not only in thermal equilibrium [51], but also in direct use to verify Crook's Theorem and the Jarzynski Relation [26, 27] by folding and unfolding hairpins by mechanical force. Hairpins are likewise used in simulations that aim to reproduce experimental results [28, 52]. Also numerically, an exponential increase of the unfoldingtime with the hairpin stem length was observed, while the folding time is almost independent of stem and loop length [53]. This shows that the thermodynamic behavior can be complex, making it an ideal candidate for our study. One the other hand, the fluctuations in force-extension curves decrease with the number of hairpins in the overall secondary structure, due to a compensation effect [45]. Finally and interestingly, due to its simple structure, there even exists an analytical solution of the partition function in the limit of large L for a hairpin structure [54].

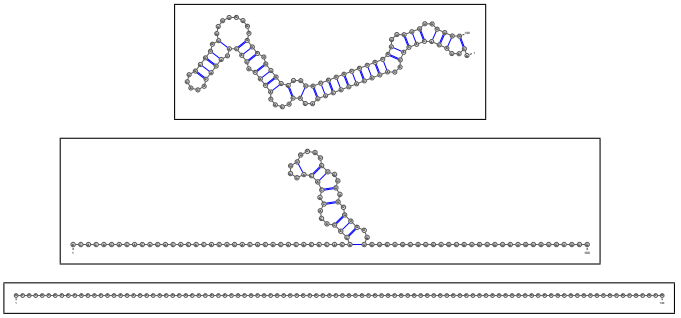

FIG. 2. Exemplary equilibrium secondary structures at $T = 1$ for different forces. Top: $f = 0$. Middle: $f = 0.805$. Bottom: $f = 2$. Drawn with the VARNA package [50].

For the studied RNA size of $L = 100$, the application of large-deviation algorithm is necessary to measure the work distribution with suitable accuracy such that the theorems of Jarzynski and Crooks can be applied and the unfolding and refolding histories can be captured.

We considered the RNA to be coupled to a heat bath at temperatures $T = 0.3$ and $T = 1$, respectively. These are low enough temperatures, such that in the force-free case, the RNA is basically folded, but exhibits thermal fluctuations. Example equilibrium secondary structures are shown in Fig. 2. It becomes apparent how the extension increases with the force parameter f.

For all unfolding and refolding processes, the force was increased from $f_0 = 0$ to $f_{\max} = 2$ and vice versa, with 400 steps each. Thus, the change of the force was $\Delta f = \pm 0.005$. Two different speeds of the processes were simulated, i.e. two different numbers n_{MC} of sweeps performed during the process, here $n_{\mathrm{MC}} = 8$ and $n_{\mathrm{MC}} = 16$. This can be compared to typical time scales needed to equilibrate RNA secondary structures of the hairpin using the MC approach, while beginning from an empty structure. We performed some tests and found that for $T = 1$ and $f = 0$ the hairpin structure finds typical configurations (as measured by the overlap, see Sec. IV C) in about 1000 sweeps while for $T = 0.3$ it takes roughly 40000 sweeps. Thus, our non-equilibrium processes are fast compared to the equilibration time. Table I shows

the other simulation parameters we have used.

T	$n_{\rm MC}$	f	n_Θ	$\Theta_{\rm min}$	$\Theta_{\rm max}$	$n_{\xi,\rm min}$	$n_{\xi,\rm max}$	$t_{\rm ld}/10^8$
0.3	8	$0\to 2$	17	0.6	7	938	9×10^4	5.44
0.3	8	$2\to 0$	10	0.4	2	1587	6×10^4	4.05
0.3	16	$0\to 2$	18	0.6	8	1407	12×10^4	2.50
0.3	16	$2\to 0$	10	0.457	2	2557	9×10^4	1.82
1	8	$0\to 2$	11	0.8	10	354	6×10^4	6.95
1	8	$2\to 0$	13	1	5	1350	6×10^4	2.81
1	16	$0\to 2$	11	0.8	10	938	75×10^2	4.41
1	16	$2\to 0$	10	0.8	5	2344	12×10^4	2.35

TABLE I. Simulation parameters for different temperatures T, for different process speeds $n_{\rm MC}$ and unfolding ($f=0\to2$) and refolding ($f=2\to0$) processes. For the large-deviation MCMC simulation n_Θ different values of the temperature-like parameter $\Theta\in[\Theta_{\rm min},\Theta_{\rm max}]$ were considered. In each MCMC step a number $n_\xi\in[n_{\xi,\rm min},n_{\xi,\rm max}]$ of entries from the vectors ξ of random numbers are changed. For the lowest value of Θ we have $n_\xi=n_{\xi,\rm min}$, for the largest $n_\xi=n_{\xi,\rm max}$, for the others in between. The total number of MCMC steps in the large-deviation simulation was always larger than the given values $t_{\rm ld}$, the actual values depending on the value of Θ and on the available computing time on the computing cluster, respectively. The longest running time occurred for the unfolding (forward) process $T=1, n_{\rm MC}=8$ and took $t_{\rm ld}=14.5\times10^8$ steps.

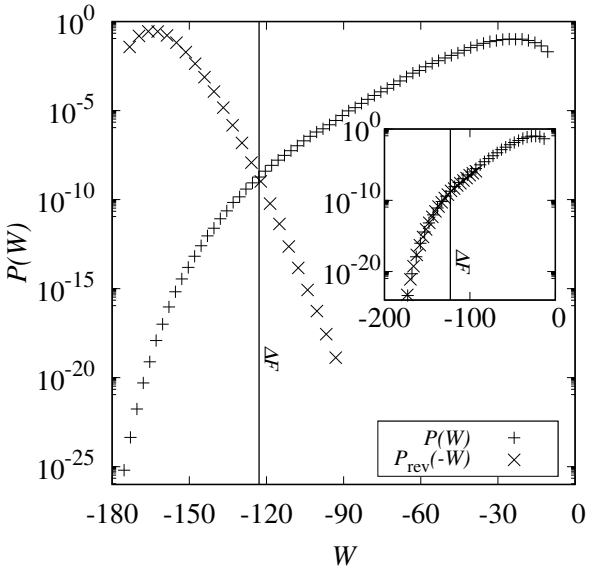

FIG. 3. Plain and mirrored work distributions for $T=1$ and 16 sweeps of the forward and reverse process, respectively. They intersect near $W=\Delta F$, which is the exact value and indicated by the vertical line. The inset shows the same plot but with the distribution for the reverse process (cross symbols) rescaled as $P_{\rm rev}(-W)\exp(-(\Delta F-W)/T)$, according to the equation of Crooks, yielding a good agreement with $P(W)$.

A. Work distributions

In Fig. 3 the work distributions $P(W)$ of the forward and $P_{\rm rev}(-W)$ of the reverse processes are shown for the case $T=1$ and $n_{\rm MC}=16$. With the application of the large-deviation scheme, very small probabilities down to 10^{-26} could be resolved, i.e., over 26 orders in magnitude. The crossing of the distributions at a work value $W=\Delta F$ predicted by the theorem of Crooks [2] can be well observed. For the present model, because the partition function can be calculated exactly numerically, we are able to obtain $\Delta F=1/T\log\{Z(f=f_0)/Z(f=f_{\rm max})\}$. Apparently, the data matches the expectations from Crooks theorem with high precision.

Crooks relation means that when $P_{\rm rev}(-W)$ is rescaled according the exponential $\exp(-(\Delta F-W)/T)$, it equals $P(W)$. This is also confirmed very convincingly by our data over up to 15 decades, as shown in the inset of Fig. 3. This in particular shows that our higher-level MCMC simulation is well equilibrated [37]. Similar results were obtained for the faster $n_{\rm MC}=8$ process (not shown).

In Fig. 4 the corresponding results for the lower temperature $T=0.3$ ($n_{\rm MC}=16$) are shown. Again, Crooks theorem is confirmed with high precision. For the case $n_{\rm MC}=8$ (not shown) the distribution even reaches probabilities as small as 10^{-46}.

B. Jarzynski Integrand

The integrand of $\langle e^{-W/T}\rangle=\int dW P(W)e^{-W/T}$ is shown in Fig. 5, for $T=1$, $n_{\rm MC}=8$ and forward and reverse work processes, respectively. The point where the integrand peaks is exponentially relevant and can be used to approximate the integral. Here this is the value $W_{\rm J}^*\approx-172$ This, together with its probability, determines according to Jarzynski's equation the free energy difference via $P(W_{\rm J}^*)e^{-W_{\rm J}^*/T}\approx e^{-\Delta F/T}$, i.e., $\Delta F=W_{\rm J}^*-T\log P(W_{\rm J}^*)$, which explaines the notable difference of $W_{\rm J}^*$ from ΔF.

C. Similarity to equilibrium

Our results allow us to go beyond calculation of distributions and study the actual dynamic processes. This is in particular possible when conditioning to any value of W. We concentrate now on $T=1$, the results for $T=0.3$ are similar. During a forced process, we sampled structures, one for each considered value of f, in equilibrium and in non-equilibrium ensembles, respectively. To compare two sampled structures $\mathcal{S}$ and $\mathcal{S}'$ from either of both ensembles, we define an *overlap* σ, which runs over all bases of the sequence, and counts $1/L$ if for both struc-

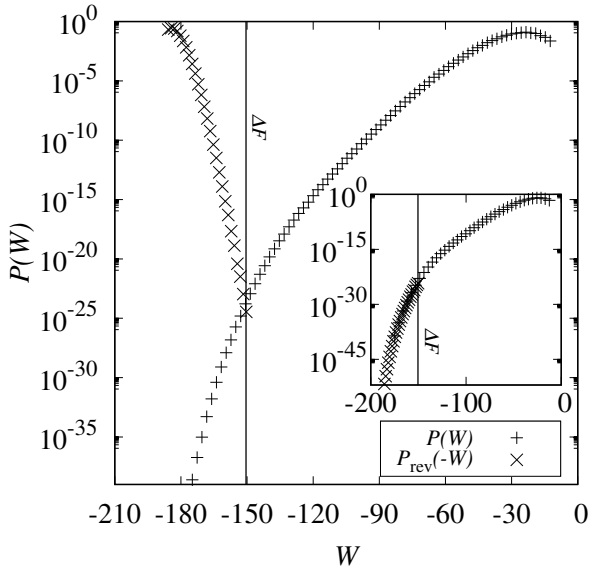

FIG. 4. Plain and mirrored work distributions for $T = 0.3$ and 16 sweeps of the forward and reverse process, respectively. They intersect near $W = \Delta F$, which is the exact value and indicated by the vertical line. The inset shows the same plot but with the distribution for the reverse process (cross symbols) rescaled as $P_{\rm rev}(-W)\exp(-(\Delta F - W)/T)$, according to the equation of Crooks, yielding a good agreement with $P(W)$.

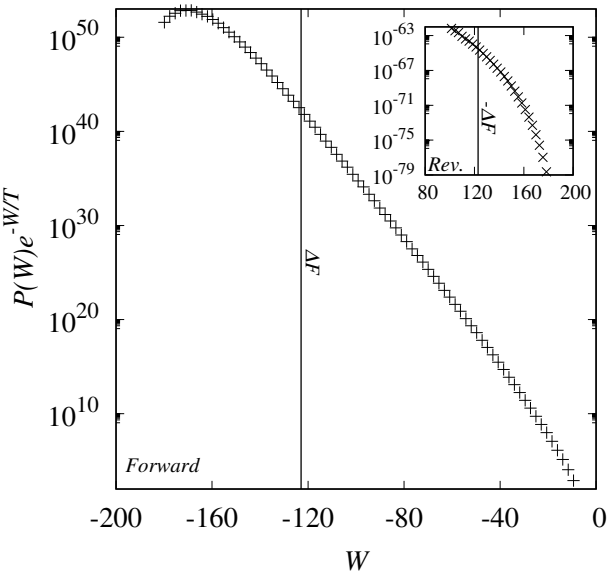

FIG. 5. Jarzynski integrand of the forward process for 8 sweeps at $T = 1$. Inset: Same for the reverse process, in which the maximum is not reached. Error bars are smaller than symbol sizes.

tures the base is not paired or if for both structures it is paired with the same base. Otherwise zero is counted. By using the equivalent notations $\{s(i)\}$ and $\{s'(i)\}$ for the pairing partners of the residues (0 if not paired), the overlap is given by

$$\sigma(\mathcal{S},\mathcal{S}') = \frac{1}{L}\sum_{i=1}^{L}\delta_{s(i),s'(i)} \qquad (9)$$

where the Kronecker delta is given by $\delta_{k,l} = 1$ if $k = l$ and $\delta_{k,l} = 0$ else. Thus, the overlap equals one when $\mathcal{S},\mathcal{S}'$ denote the same secondary structure, and zero when they are completely different. Overlap quantities are used frequently to determine order in complex systems, e.g., spin glass [55].

Fig. 6 shows average *non-equilibrium profiles* $\overline{\sigma}(f)$, i.e., averaged overlaps σ as function of f, where in the calculation of the overlaps one structure is a given non-equilibrium sample of a forward or a reverse process and the other structure is a sampled equilibrium structure. The average is always taken over many equilibrium structures. Thus $\overline{\sigma}(f)$ also reflects the fluctuations of the configurations, both those arising from the ensemble where $\mathcal{S}$ and $\mathcal{S}'$ are taken from. For comparison in all plots the average *equilibrium* profile is shown, where both structures are sampled from equilibrium. Our results show that folded structures at low force value f are characterized by a variety of secondary structures, while at high values of f, where the RNA is basically stretched, the secondary structures are very similar to each other. Note that for the reverse trajectories in the non-equilibrium case, the overlap is for $f \to 0$ slightly higher than the equilibrium value. This is probably due to the fact that the non-equilibrium trajectories contributing the structures $\mathcal{S}'$ to $\overline{\sigma}$ are selected with respect to the work W, i.e., they represent a sub ensemble with an almost fixed work value. Thus, they overall fluctuate less as compared to the equilibrium structures, where the work fluctuates more. We see that for *typical* work values, i.e., where $P(W)$ and $P_{\rm rev}(W)$ peak, in particular for the forward process, large differences for non-equilibrium profiles compared to the average equilibrium profile occur. For work values *near* the region $W = \Delta F$ and $W = W_{\rm J}^*$ on the other hand, a higher similarity is observed, i.e., these very rare non-equilibrium processes enroll close to the equilibrium ones. Note that in Fig. 6 we actually show the non-equilibrium trajectories in between $W = \Delta F$ and $W = W_{\rm J}^*$ for $W \approx -139$, which exhibit the largest overal similarity to the equilibrium profiles, which is defined next.

We quantify the similarity I^σ of the non-equilibrium processes to the equilibrium case by integrating over all force values f the absolute difference of $\overline{\sigma}(f)$ between the equilibrium and non-equilibrium case, and average this integral over close-by values of W, i.e., obtaining $I^\sigma(W)$. The result is shown in Fig. 7, in which rather larger differences for typical values of W are observed. This is in particular true for the forward process, see also Fig. 6.

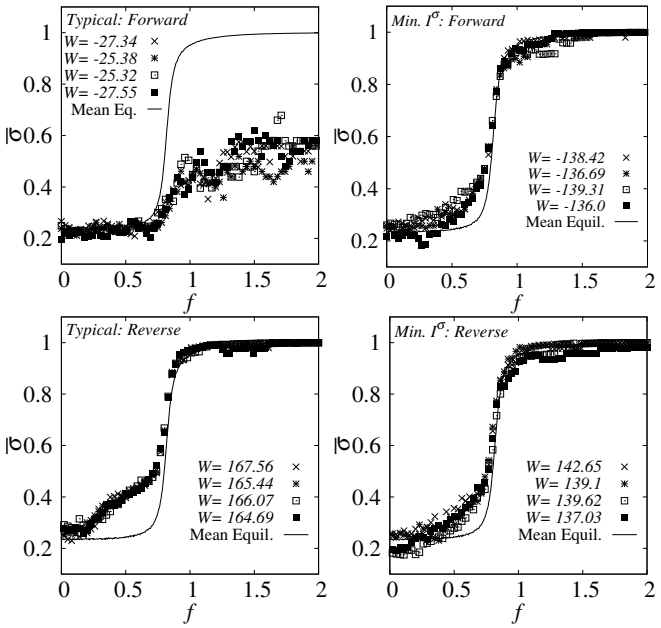

FIG. 6. Average non-equilibrium overlap profiles $\overline{\sigma}(f)$, i.e. between one non-equilibrium and one equilibrium structure, for some sample processes at $T = 1$ and 16 sweeps, with mentioned non-equilibrium work values W. The solid line is the averaged equilibrium overlap, i.e., between two equilibrium structures, respectively. Top row: forward process for typical (left) and very rare (right) values of W. Bottom row: the same for the reverse process. Error bars are smaller than symbol size.

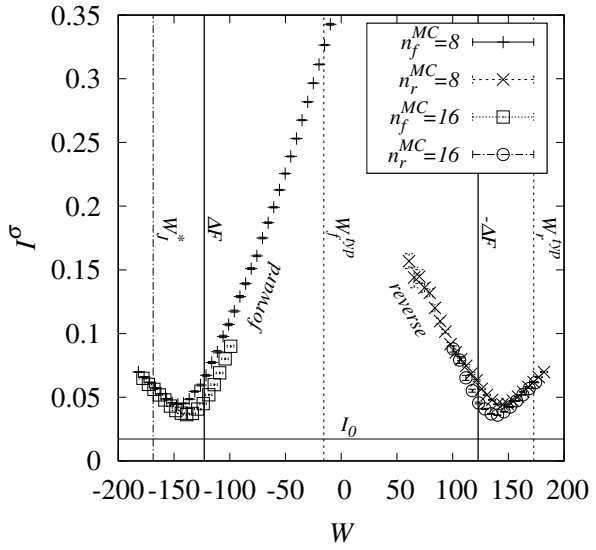

FIG. 7. Integrated difference I^σ between equilibrium and non-equilibrium overlap profiles at $T = 1$, for forward (left) and backward (right) processes. For 16 sweeps the data is only partially shown, for better visibility. The horizontal line indicates I_0, the value of I^σ in equilibrium. Vertical lines indicate work values at (from left to right) the maximum W_J^* of the Jarzynski integrand, the free energy difference ΔF, the maximum W_f^{typ} of the forward process work distribution, the negative free energy difference $-\Delta F$ and the maximum point W_r^{typ} of the reverse process work distribution.

The reason is that for each sequence of nested pairs only the first-level pair is subject to the force. Thus, when increasing the force, opening pairs below the first level pair it is energetically not favorable. Thefore, basically one pair must be openend after the other, while the reminder of the structure will not change much. Thus, the structure is opened like a zipper, but due to the random update order this takes a while. On the other hand when $f \to 0$ is decreased, any formed pair will decouple a certain subsequence from the force, which creates many potential pairs which are energetically favorable. This leads to faster folding as compared to unfolding in non-equilibrium.

Next, near $W \approx \pm\Delta F$ the similarity is of the order of the similarity I_0 obtained by averaging I^σ over many independent equilibrium processes, which represents the equilibrium fluctuations. Also the forward processes sampled for work values near the value $W_J^* \approx -170$ where the Jarzynski integrand $P(W)e^{-W/T}$ peaks exhibit a high similarity to the equilibrium case. Note that for the reverse process, the value of W_J^* occurs outside our sampled region, thus we do not have processes for this case. For the slower case of $n_\mathrm{MC} = 16$ sweeps, i.e., a bit nearer to equilibrium, the location minimum moves closer to ΔF and even decreases in height towards the equilibrium value I_0.

Thus, our results show that the rare processes near $W = \pm\Delta F$ do not only have similar work values like the equilibrium processes, they exhibit also very similar sequences, as function of the force f, of sampled structures. We obtained a similar result when considering force-extension curves, see SM.

D. Force-extension curves

In addition to the overlap profiles $\overline{\sigma}(f)$ presented before, we also used force extension curves (FECs) $n(f)$ to compare processes for equilibrium and non-equilibrium situations. Note that the extension $n(\mathcal{S})$ of a structure can be very much influenced by single base pairs. Thus two processes, which look very similar on the level of secondary structures, can be very different with respect to force-extension curves.

Samples for equilibrium and non-equilibrium FECs for forward processes, along with corresponding averages, are shown in Fig. 8. For the equilibrium case, a sigmoidal form can be observed, with some fluctuations, and a strong change near the critical force value, where the folding-unfolding transition takes place [25]. For the non-equilibrium case, the typical FECs, i.e., with typical work values W far from ΔF, agree only for small values

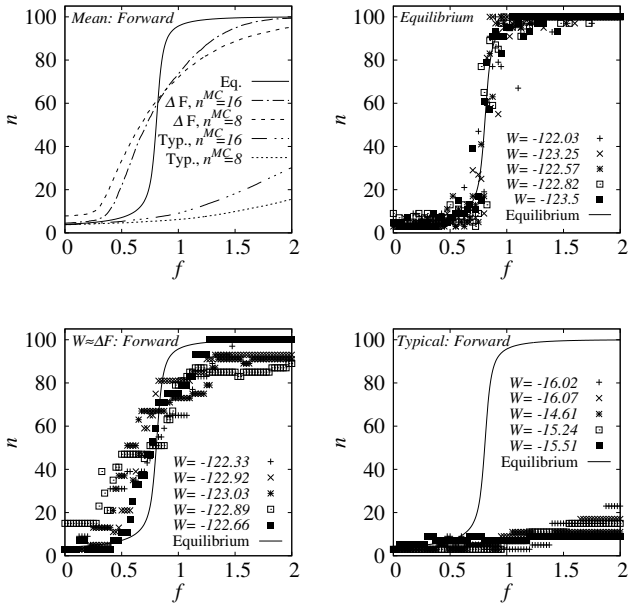

FIG. 8. Top left: Mean FECs, in equilibrium as well as in non-equilibrium for typical forward processes and for work values near ΔF, for two different numbers $n_{\rm MC}$ of sweeps at $T = 1$. Top right: Samples of such single FECs in equilibrium. Bottom left: samples of non-equilibrium FECs with $n_{\rm MC} = 8$ for W near ΔF. Bottom right: samples of typical non-equilibrium FECs, i.e., where $W \gg \Delta F$, with $n_{\rm MC} = 8$. The solid line represents always the mean equilibrium FEC.

of f, i.e., in the initial phase of the process. On the other hand, the rare processes with W close to ΔF, where five different examples are shown here, are much more similar to the equilibrium FECs. Here, differences appear mainly near the critical folding-unfolding force.

Samples for equilibrium and non-equilibrium FECs for backward processes, along with corresponding averages, are shown in Fig. 9. The results correspond to the forward case, but the processes with *typical* values of W agree well with the average equilibrium FEC only for large values of f but not for small values of f. This means they also agree in the initial phase of the process, before the critical folding-unfolding force value is reached. The FECs for work values $W \approx \Delta F$ are also for reverse processes much more similar to the equilibrium case than typical reverse processes.

These results are confirmed by averaging the absolute value of the differences between one FEC $n(f)$ and the mean equilibrium FEC $\overline{n}_{Eq}(f)$ over all available values of the force f, i.e., calculating $I^n = [\frac{1}{n_{\rm f}} \sum_f |n(f) - \overline{n}_{Eq}(f)|]$ where the average $[\ldots]$ is over different realisations of $n(f)$. Even when considering equilibrium FECs for $n(f)$, respectively, there is some variation reflected by a non-zero average value I_0. When using non-equilibrium FECs, with a specified binned value of W, one sees stronger differences, as visible in Fig. 10. Similar to I^σ, the closest agreements between non-equilibrium and equilibrium are seen near $W \approx \Delta F$. In contrast to I^σ the level of the equilibrium fluctuations is not reached for the measurable quantity FEC.

V. DISCUSSION

RNA unfolding and refolding in equilibrium and in non-equilibrium for one particular RNA sequence have been studied. For the non-equilibrium case, by using sophisticated large-deviation algorithms, we could access a large range of the support of the probability distribution for the work. This allowed us to confirm the theorems of Crooks and Jarzynski over several dozends decades in probability. Furthermore, we analyzed the trajectories in force-extension as well as in secondary-structure space conditioned to various values of W. It was observed that near the most relevant, but very improbable, values of W, the sampled trajectories reach a high similarity with true equilibrium. Thus, the study here does not depend on assigning a time-dependent weight to the trajectories as, e.g., in Refs. [21, 23, 24], the selection is solely by the total work performed during the process and suitably evaluating fluctuation theorems. Also no other particular similarity to equilibrium is enforced explicitly by our procedure. Our approach and results may open a pathway to

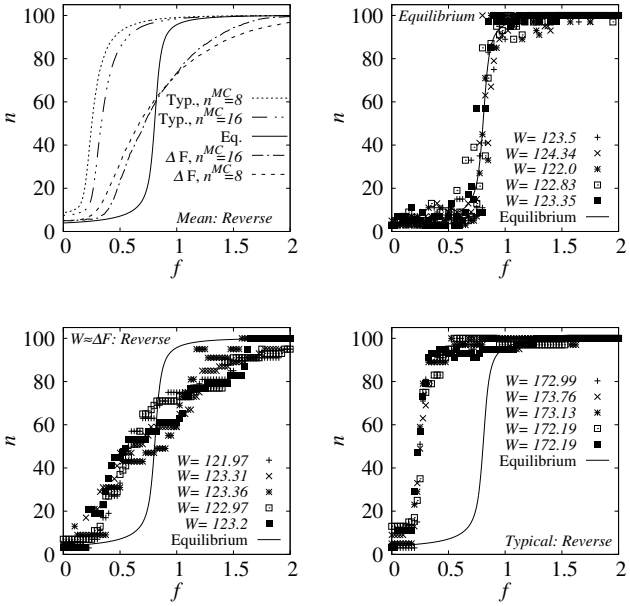

FIG. 9. Top left: Mean FECs, in equilibrium as well as in non-equilibrium for typical reverse processes and for work values near ΔF, for two different numbers of $n_{\rm MC}$ of sweeps at $T = 1$. Top right: Samples of such single FECs in equilibrium. Bottom left: samples of non-equilibrium reverse FECs with $n_{\rm MC} = 8$ for W near ΔF. Bottom right: samples of typical non-equilibrium reverse FECs, i.e., where $W \gg \Delta F$, with $n_{\rm MC} = 8$. The solid line represents always the mean equilibrium FEC.

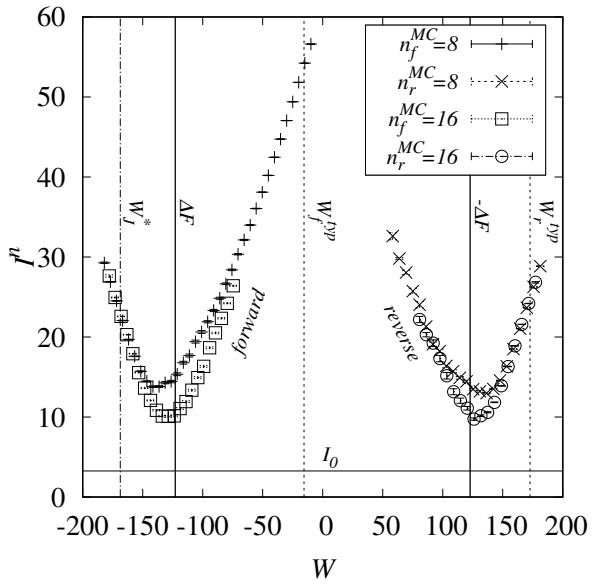

FIG. 10. Integrated extension difference I^n between averaged equilibrium and non-equilibrium at $T = 1$. For 8 sweeps the entire work range is plotted, where for 16 sweeps only a range around the minimum is shown, for better visibility. I_0, represented by a horizontal line, is the averaged value of I^n when comparing equilibrium FECs to the averaged equilibrium FEC. The left curves represent the forward, the right ones the reverse process. Vertical lines indicate work values at (from left to right) the maximum W_J^* of the Jarzynski integrand , the free energy difference ΔF, the maximum W_f^{typ} of the forward process work distribution , the negative free energy difference $-\Delta F$ and the maximum point W_r^{typ} of the reverse process work distribution.

learning not only about equilibrium characteristic scalar numbers from non-equilibrium measurements, but even investigating near equilibrium dynamics by performing very fast but biased non-equilibrium simulations. We anticipate that similar studies are feasible and useful for many different types of systems.

Clearly, we have studied so far one particular RNA sequence. Although we expect that in general the closeness of relevant rare-event trajectories to equilibrium carries over, details on the result will certainly depend on the actual sequence. One can imagine that there are biological sequences which fold easily and thus evolve closer to equilibrium anyway. On the other hand, there might be more complex structures, e.g., arising from random sequences, which have an even more complex behavior than the hairpin. Here, it would be very interesting to test whether our approach still works and, if yes, how close the most relevant trajecories to equilibrium are in this case.

For further studies, also beyond considering RNA secondary structures, one could also extend the approach, by storing the configurations of the close-to-equilibrium $W \approx \Delta F$ generated rare trajectories. Starting with these configurations, one could perform additional equilibrium simulations at fixed force values, i.e., without performing work, in the hope to get quickly close or even up to equilibrium. We have run some test simulations which show that one can indeed get even much closer to the equilibrium behavior by applying this add-on equilibration, apparently perfectly with respect to the force-extension curves, but this also depends on the temperature. Here more studies are needed, in particular a comparison of how good one can equilibrate by just using secondary-structure MC simulations when starting with empty configurations, i.e., the extended structure without any base pairs. Also it would be very interesting to see how these results depend of the actual RNA sequence and the corresponding energy landscape.

ACKNOWLEDGMENTS

The simulations were performed at the the HPC cluster CARL, located at the University of Oldenburg (Germany) and funded by the DFG through its Major Research Instrumentation Program (INST 184/157-1 FUGG) and the Ministry of Science and Culture (MWK) of the Lower Saxony State.

[1] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[2] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).
[3] J. Kurchan, JSTAT **2007**, P07005 (2007).
[4] U. Seifert, Eur. Phys. J. B **64**, 423 (2008).
[5] E. M. Sevick, R. Prabhakar, S. R. Williams, and D. J. Searles, Ann. Rev. Phys. Chem. **59**, 603 (2008).
[6] C. Jarzynski, Eur. Phys. J. B **64**, 331 (2008).
[7] M. Esposito, U. Harbola, and S. Mukamel, Rev. Mod. Phys. **81**, 1665 (2009).
[8] C. Jarzynski, Annual Review of Condensed Matter Physics **2**, 329 (2011).
[9] U. Seifert, Rep. Progr. Phys. **75** (2012), 10.1088/0034-4885/75/12/126001.
[10] R. Marsland III and J. England, Rep. Progr. Phys. **81** (2018), 10.1088/1361-6633/aa9101.
[11] A. K. Hartmann, *Big Practical Guide to Computer Simulations* (World Scientific, Singapore, 2015).
[12] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. **107**, 21441–21446 (2010).
[13] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993).
[14] D. J. Evans and D. J. Searles, Phys. Rev. E **50**, 1645 (1994).
[15] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**,

2694 (1995).
[16] G. Gallavotti and E. G. D. Cohen, J. Stat. Phys. **80**, 931 (1995).
[17] J. Kurchan, J. Phys. A: Math. Gen. **31**, 3719 (1998).
[18] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).
[19] C. Maes, J. Stat. Phys. **95**, 367 (1999).
[20] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).
[21] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[22] A. B. Adib, Phys. Rev. E **71**, 056128 (2005).
[23] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).
[24] G. Hummer and A. Szabo, Proc. Nat. Acad. Sci. **98**, 3658 (2001).
[25] M. Müller, F. Krzakala, and M. Mézard, Eur. Phys. J. E **9**, 67–77 (2002).
[26] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Science **296**, 1832 (2002).
[27] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, Nature **437**, 231– (2005).
[28] F. Liu, H. Tong, and Z. Ou-Yang, Biophys. J. **90**, 1895.
[29] A. K. Hartmann, Phys. Rev. E **65**, 056102 (2002).
[30] J. A. Bucklew, *Introduction to rare event simulation* (Springer-Verlag, New York, 2004).
[31] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
[32] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Ann. Rev. Phys. Chem. **53**, 291 (2002), pMID: 11972010.
[33] C. Giardinà, J. Kurchan, and L. Peliti, Phys. Rev. Lett. **96**, 120603 (2006).
[34] V. Lecomte and J. Tailleur, JSTAT **2007**, P03004 (2007).
[35] W. Staffeldt and A. K. Hartmann, Phys. Rev. E **100**, 062301 (2019).
[36] A. K. Hartmann, P. L. Doussal, S. N. Majumdar, A. Rosso, and G. Schehr, Europhys. Lett. **121**, 67004 (2018).
[37] A. K. Hartmann, Phys. Rev. E **89**, 052103 (2014).
[38] P. G. Higgs, Phys. Rev. Lett. **76**, 704 (1996).
[39] B. Burghardt and A. K. Hartmann, Phys. Rev. E **71**, 021913 (2005).
[40] R. Lorenz, S. H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker, Algor. Mol. Biol. **6**, 26 (2011).
[41] I. Tinoco and C. Bustamante, J. Mol. Biol. **293**, 271 (1999).
[42] M. Zuker, Nucl. Acid Res. **31**, 3406 (2003).
[43] Y. Ding, C. Y. Chan, and C. E. Lawrence, Nucleic Acids Research **32**, W135 (2004).
[44] J. Reuter and D. H. Mathews, BMC Bioinformatics **11**, 129 (2010).
[45] U. Gerland, R. Bundschuh, and H. T., Biophys J. **81**, 1324 (2001).
[46] R. Nussinov, G. Pieczenik, J. R. Griggs, and D. J. Kleitman, SIAM J. Appl. Math. **35**, 68 (1978).
[47] G. E. Crooks and D. Chandler, Phys. Rev. E **64**, 026109 (2001).
[48] C. Flamm, W. Fontana, I. L. Hofacker, and P. Schuster, RNA **6**, 325–338 (2000).
[49] E. E. Dykeman, Nucleic Acids Research **43**, 5708–5715 (2015).
[50] K. Darty, A. Denise, and Y. Ponty, Bioinformatics **25**, 1974 (2009).
[51] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, and C. Bustamante, Science **292**, 733 (2001).
[52] F. Liu and Z. Ou-Yang, Biophys. J. **88**, 76 (2005).
[53] M. Faber and S. Klumpp, Phys. Rev. E **88**, 052701 (2013).
[54] R. Bundschuh and T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999).
[55] M. Mézard, G. Parisi, and M. Virasoro, *Spin glass theory and beyond* (World Scientific, Singapore, 1987).