Force-induced melting of DNA hairpin: Unfolding pathways and phase diagrams

Sumitra Rudra,[1] Keerti Chauhan,[1] Amit Raj Singh,[2] and Sanjay Kumar[1]
[1]*Department of Physics, Banaras Hindu University, Varanasi 221005, India*
[2]*Department of Physics, Graphic Era Hill University, Dehradun 248002, India*

(Received 13 December 2022; accepted 27 April 2023; published 23 May 2023)

Using the exact enumeration technique, we have studied the force-induced melting of a DNA hairpin on the face centered cubic lattice for two different sequences which differ in terms of loop closing base pairs. The melting profiles obtained from the exact enumeration technique is consistent with the Gaussian network model and Langevin dynamics simulations. Probability distribution analysis based on the exact density of states revealed the microscopic details of the opening of the hairpin. We showed the existence of intermediate states near the melting temperature. We further showed that different ensembles used to model single-molecule force spectroscopy setups may give different force-temperature diagrams. We delineate the possible reasons for the observed discrepancies.

DOI: 10.1103/PhysRevE.107.054501

I. INTRODUCTION

Hairpin loops are frequently observed secondary structures of nucleic acids [1,2]. Besides their canonical forms, hairpins can also adopt noncanonical conformations that occur at a high frequency in transfer RNAs (tRNAs), when stabilized by the Wobble base pairs, which do not obey the Watson-crick pairing rules [3–5]. Hairpins participate in several cellular processes, such as DNA recombination [6,7], regulation of gene expression, facilitation of mutagenic events, and most importantly acts as nucleation sites for RNA folding into the final conformations [8–13]. A single-stranded DNA (ssDNA) (or RNA) hairpin structure has structurally and dynamically two distinct domains: a base paired stem part formed by the complementary bases, and a single-stranded loop formed by one type of nucleotides, which connect two halves of the stem [14,15]. The stem part shows the same response to change in solution conditions as a dsDNA oligomer, but the loop region shows a wide range of folding patterns that depend on the number and type of nucleotide in the loop. This stem-loop structure fluctuates thermodynamically between different conformations, which are broadly divided into two states: the open state or unzipped state, where all the complementary bases are separated, and the fully closed (folded) state, where all the complementary bases are paired and form the hairpin structure [16–19]. From polymer theory, it is known that a polymer chain will be in either a closed (folded) state or a swollen state depending on the temperature [20]. In the closed state which occurs at low temperature, the average end-to-end distance $< R >$ will be nearly equal to zero. At high temperatures, $< R >$ scales as N^{ν} with $\nu = 3/(d+2)$, where N is the length of the chain and d is the dimension [20–22].

Significant efforts have been made to understand the underlying mechanism of folding and unfolding of nucleic acids, which is a prerequisite for explaining the biological functions of such molecules. Thermal melting of DNA hairpins are investigated using various experimental techniques such as calorimetry [23–25], ultraviolet (UV) [26,27], circular dichroism (CD) [28], and some spectroscopic methods, including fluorescence resonance energy transfer (FRET) [28] and fluorescence correlation spectroscopy (FCS) [29,30], etc. Experiments performed on a single-stranded hairpin molecule often yield sigmoidal melting curves, which corresponds to a two-state process of DNA hairpin melting. Efforts were also made to study the effect of loop entropy and sequence sensitivity on the melting mechanism of the short oligonucleotides. However, the information showing fingerprints of intermediate states in FCS experiments require further understanding of closing-opening dynamics of DNA hairpin [31–33]. Efforts have also been made to explore T-jump in the energy landscape using ultra fast temporal resolution [34,35], which show the existence of intermediate states near the melting temperature.

Besides thermal studies, the behavior of nucleic acids (NAs) under the applied force also drew considerable attention in the recent past [36–39]. These studies were mostly motivated by the fact that certain enzymes and helices participate during cellular processes such as replication, transcription, translation, and chromatin remodeling [8,10,40] and apply force of the order of pN [36,41,42] on the biomolecules of interest. Thus, from the stability point of view, the force appears to be a critical thermodynamic parameter in numerous biological systems. Advent of single-molecule force spectroscopy (SMFS) techniques such as magnetic tweezers, optical tweezers, and atomic force microscopy permit us to manipulate biomolecules of interest by applying the force of pN [22,39,43,44]. Unprecedented details were obtained from the force-extension curves in the last couple of decades about the stability and functions of biomolecules. The existence of multistep plateaus in the force-extension curve shows the signature of intermediate states, which have been seen in recent unfolding experiments performed on homo-sequence DNAs such as poly(dA), poly(dC), etc. [36]. Danilowicz *et al.* [39] studied the elastic properties of ssDNA and showed that temperature has a significant impact on the force-extension curve. In the observed experiment by varying the force, the

polymer chain acquires the conformation of a stretched state with $\nu = 1$, which is otherwise not accessible.

Theoretical works based on the Ising-like models [45,46] (Poland and Scheraga (PS) model [47], the standard PBD model [48], self-avoiding walks model of polymer [49–51], etc.) which followed these experiments shed important information about the cellular processes. Some of these models successfully incorporate loops or bubbles in their description to model short oligonucleotides, hence could show the presence of intermediate states during the melting. The Gaussian binding energy model based on the exact calculations of the Helmholtz free energy of the all partially denatured states have been recently studied to describe thermal melting of DNA hairpin [33,52,53]. The model is capable of yielding an effective one-dimensional free-energy landscape, to provide the most probable pathway of the opening of DNA hairpin. Moreover, using the free-energy landscapes, one can probe the dynamical features of the intermediate states (bubble breathing dynamics for DNAs) in terms of calculating some direct experimental observables such as correlation functions and first passage times [33]. In this context, molecular dynamics simulations (Langevin and the atomistic simulations) based on coarse-grained [54] or atomistic models [55–60] have emerged as a powerful tool to uncover structural and dynamical details of the cellular processes.

In the past, the similar free-energy landscape was obtained using the classic exact enumeration technique, where DNA is considered as a mutual attracting self-avoiding walks (MASAWs) on regular 2D and 3D lattices, where all the possible configurations have been enumerated exactly [16,50]. The results based on these models are considered as "exact" for a small length of DNA, and results can be extrapolated using suitable techniques in the limit length tends to infinity. The total number of conformations for a single strand, $C_N \propto \mu^N N^{\gamma-1}$, where γ is the critical exponent and μ is the connectivity constant of the underlying lattice [21,22]. Among these lattices, the two-dimensional square ($\mu = 2.638$) and the three-dimensional simple cubic (SC) lattice ($\mu = 4.684$) have drawn the most effort, from calculating universal constants to predicting thermodynamics of bio-polymers. Efforts were also made on lattices to obtain universal constants with higher connectivity order, e.g., body-centered cubic (BCC) and face-centered cubic (FCC) structures [61,62]. The reported value of connectivity coefficients for the BCC lattice is $\mu = 6.530520$, whereas for FCC lattice it is $\mu = 10.037075$. However, thermodynamic properties and associated transitions in DNA remain elusive. Since the configurational entropy of FCC lattice is larger compared to square or cubic lattices, therefore, one can get a better insight into the role of bubbles in DNA melting.

In this paper, we study thermodynamics of a hairpin construct as a self-avoiding walks on the FCC lattices. We chose two short DNA sequences 5′-GGATAX-(T(4))-X'TATCC-3′, which were explored in experimental as well as theoretical studies [57,63]. The sequence has a loop of length 4 and a stem of 6 base pairs with a weak AT region at middle, which is connected through strong GC bonds at the free end of the stem and a variable base-pair (XX') closing the loop, called the closing base-pair (CBP) that can decisively affect melting of the sequence. Despite its short length, the chosen sequence shows several interesting structural regions, which need to be explored. The paper is organized as follows: in Sec. II, we briefly describe the model and the method to obtain the partition function. In Sec. III, we study the response of temperature on the intact base pairs to obtain the melting profile of DNA hairpin. We compare the melting profile obtained from the exact enumeration technique with Gaussian binding energy model and Langevin dynamics simulations also in this section. The free-energy landscapes of DNA hairpins obtained from the partition function has been discussed in Sec. IV. We also propose a method to calculate the opening pathways of the sequences in this section. In Sec. V, we focused our studies on the force-induced melting of hairpins with different CBPs, where we have shown that there is no qualitative change in force-extension curves; however, temperature-extension curves at different forces reveal the microscopic details of opening of hairpins. In Sec. VI, we have discussed two different ensembles to obtain the phase diagram: (i) by varying temperature, we monitor average number of intact base pairs at constant force, and (ii) by varying the applied force, we calculate average extension at constant temperature. We showed that the phase diagrams obtained in these two ensembles differ at high temperature. We delineate the possible reasons for the observed discrepancies. We summarized the results in Sec. VII.

II. MODEL AND METHODS

We have considered N-step self-attracting, self-avoiding walk with $(N + 1)$ vertices. Every vertex of the walk represents a nucleotide. Although the model is general enough to be defined in any dimension, to take account of proper entropy, we considered a three-dimensional FCC lattice in the present study. Here, we have considered the lattice constant 2 with step length $\sqrt{2}$. We consider here an effective native interaction between nucleotides (or bases) of the stem if they are the nearest neighbor on the lattice. The nearest-neighbor interaction mimics the short range nature of the hydrogen bonds thus the interacting pair may also be called a base pair. By native interaction in the stem, we mean that first nucleotide of a chain interacts with a complementary $(N + 1)$th nucleotide, second nucleotide interacts with complementary Nth nucleotide and so on (Fig. 1). This restriction allows that a base can at most pair with another complementary base only and is based on the assumption of the Go model that the energy of each conformation is proportional to the number of native contacts of the stem, and nonnative contacts incur no energetic cost. By construction, the native state is the lowest energy conformation of the closed state of a DNA hairpin in comparison to the open state. We have taken two sequences (GGATAX − (T(4)) − X′TATCC) of DNA hairpin [Fig. 1(a)] studied by Vellone *et al.* [63], which differ in closing base pairs. We have indexed the base pair from the free-end of stem to the loop-end. We have enumerated all possible conformations of ssDNA of step $N = 15$. The partition function of the system is defined as

$$Z_N(T) = \sum_{N_{GC},N_{AT}} C_N(N_{GC}, N_{AT}) u^{N_{GC}} v^{N_{AT}}, \qquad (1)$$

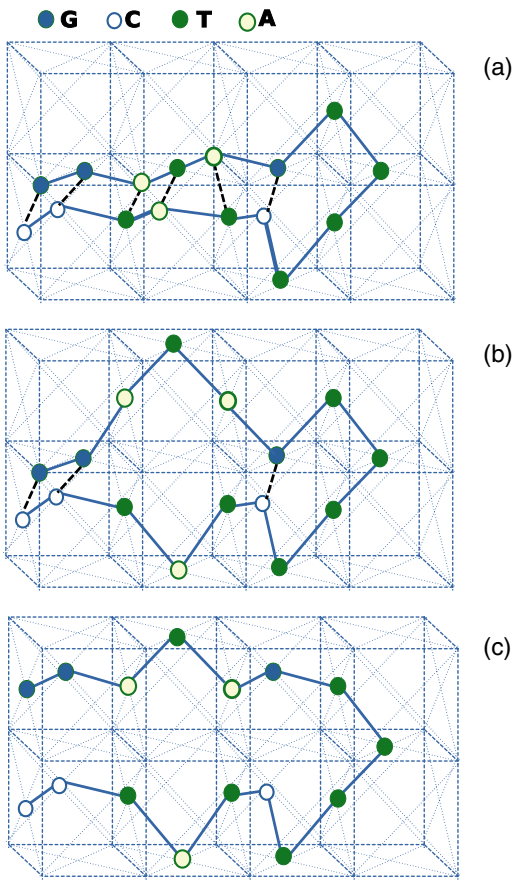

FIG. 1. Schematic representations of a ssDNA hairpin on a face-centered cubic lattice system. Panel (a) represents the native structure of a stem-loop hairpin structure; panel (b) shows the formation of bubble during the melting of hairpin. Panel (c) corresponds to the open state of the hairpin.

where $C_N(N_{GC}, N_{AT})$ is the total number of conformations corresponding to the walks of N steps. Here, N_{GC} and N_{AT} correspond to the numbers of G-C and A-T base-pairs in the stem, respectively. $u = \exp(-\epsilon_{GC}/k_BT)$, and $v = \exp(-\epsilon_{AT}/k_BT)$ are the Boltzmann weights of the G-C and A-T base pairing interactions, respectively. In the following, we set $k_B = 1$, $\epsilon_{GC} = -1.5$ and $\epsilon_{AT} = -1.0$ and do the calculations in the reduced units. The free energy of the system can be calculated exactly from the following relations:

$$F = -T \log Z_N(T). \tag{2}$$

III. MELTING PROFILE

Table I given in Appendix, provides the information about the exact density of states for the given value of N. From this table, it is straightforward to calculate the thermodynamic observable of interest. For example, the average number of closed base pair $\langle N_P(T)\rangle$ as a function of temperature describes the melting behavior of DNA, which can be obtained from the following relation:

$$\langle N_P(T)\rangle = \frac{1}{Z_N(T)} \sum_{N_{\rm GC},N_{\rm AT}} N_P C_N(N_{\rm GC}, N_{\rm AT}) u^{N_{\rm GC}} v^{N_{\rm AT}}, \tag{3}$$

where N_P is the number of closed base pairs of the stem. For the sake of the comparison, we also consider homo-sequences of the stem (only A-T or G-C) by changing the ϵ value. This will permit us to study the difference in melting temperature arising due to the change in sequence, and one expects that the melting profile of hetero-sequence of DNA hairpin will lie in between the profiles of homo-sequences of with stem consisting of all ATs or all GCs. In Fig. 2, we show the variation of $< N_P >$ as a function of temperature. The melting profile of different DNA hairpin sequences are qualitatively similar to the one seen in experiments [63]. The melting temperature T_m is defined as the temperature at which half of the base pairs are open. It is evident from the Fig. 2 that the hairpin structure with CBP as AT is less stable in comparison to the structure with CBP as GC. This is also in accordance with experimental observations [63].

We now focus our studies on the GBE model [33,52,53], which was studied in the context of unfolding of protein [64]. Singh and Granek developed the GBE model to study the effect of base pair and stacking on the melting of DNA [52,53]. Following the method developed in Refs. [52,53], we study the melting of DNA hairpin and show the influence of stacking on the melting profile. The qualitative behavior of

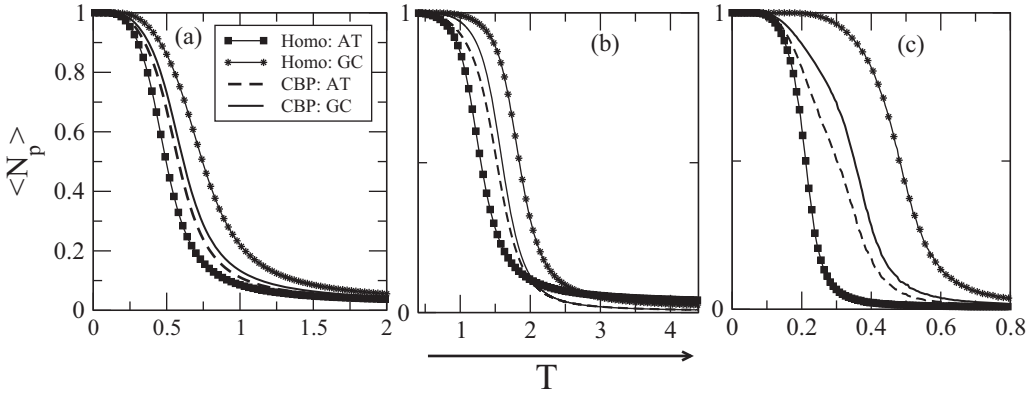

FIG. 2. Melting profiles of Homo(AT), Homo(GC), and DNA hairpins with CBP-AT and CBP-GC studied in the text obtained through (a) the exact enumeration technique, (b) the GBE model, and (c) the Langevin dynamic simulations. Difference in melting temperature is due to use of different energy in the description of the model.

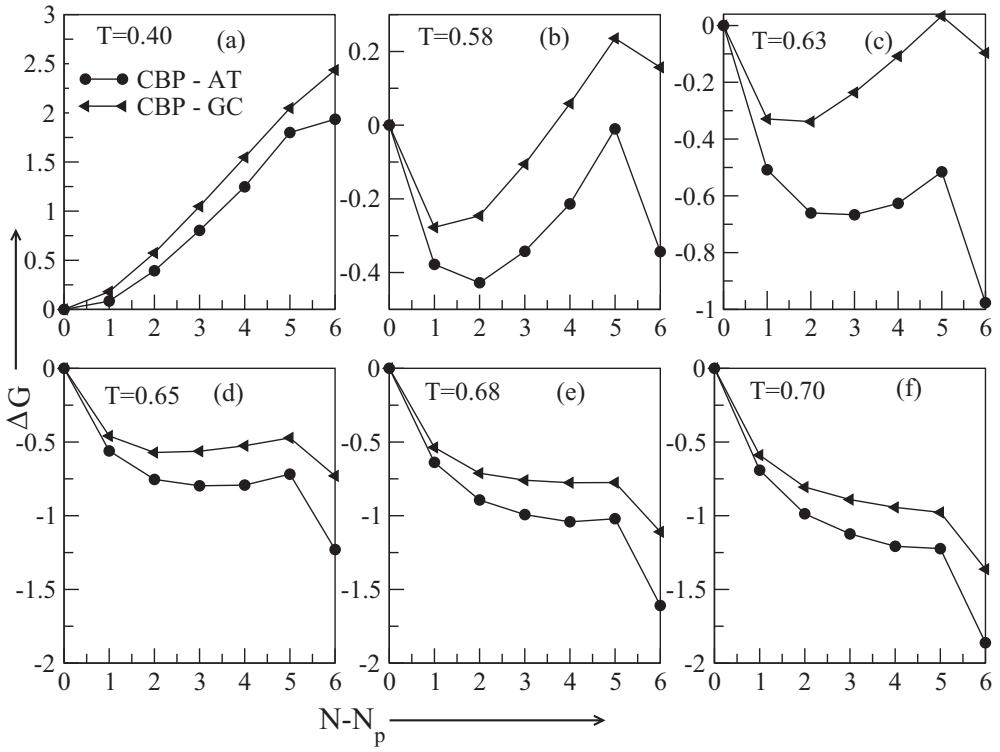

FIG. 3. Free-energy landscapes of DNA hairpin of different CBPs (AT/GC) at different temperatures as a function of open base pairs.

melting profile of hairpin remains the same as of the lattice model. One can see that the melting temperature obtained from the GBE model is higher than the one obtained from the exact enumeration technique. This is because the GBE model also contains the stacking energy in its description which is absent in the lattice model. We further substantiate our findings by Langevin dynamics simulations with the parameters used in Ref. [33]. We find a qualitatively similar melting profile shown in Fig. 2(c). Since, energy involved in these model systems differ from each other, therefore, quantitative comparison is not possible.

IV. FREE-ENERGY LANDSCAPE AND UNFOLDING PATHWAY

The effects of CBP (GC and AT) on the stability and opening pathways of unfolding remains a challenging issue. Here, we show that information of exact density of state may be used to predict the pathways of unfolding. For the present sequence, we have $2^6 = 64$ combinations for six base pairs of the stem, which may be either closed (1) or open (0). In Table I given in Appendix, we list the total number of conformations for the all possible combinations of closed and open base pairs. The equilibrium structure of the DNA hairpin at a given temperature can be obtained from the Eq. (2). The method developed in Ref. [34] has been used here to get the free-energy landscape, which can be described as follows: at low temperature, all the base pairs will be intact ($N_P = 6$) and the free energy will be the minimum. If the temperature of the system increases, here we assume that any one of the base pairs can open. We calculate the difference of free energy (ΔG) between the bound state ($N_P = 6$) and a single base pair open. In the next step, we choose any two of the base pairs open, and calculate the difference in free energy with respect to the bound state. We repeat this process to get the free-energy difference till all the base pairs get open.

In Figs. 3(a)–3(f), we show the variation of ΔG with open base pairs ($N - N_P$) at different temperatures for both CBPs. One can see from the plot [Fig. 3(a)] that at temperatures much below the melting temperature, there is an increase in ΔG as a function of ($N - N_p$) indicating that the barrier height increases which prohibits melting. The free-energy barrier decreases with rise in temperature [Figs. 3(b)–3(f)]. It is interesting to note that instead of leading directly to monotonic unfolding, at some intermediate temperatures [Figs. 3(b)–3(d)], there is emergence of a partial unfolded state which has a lower free energy. For example, at $T = 0.58$ in Fig. 3(b), the sequence having CBP-AT has minimum free energy when two bonds are open, whereas sequence having CBP-GC has minimum free energy when only one bond is open. As a result, a local minima occurs in between the completely unfolded state and folded state. Above the melting temperature, the free energy descends along the pathway, and unfolding occurs without any intermediate states [Figs. 3(e) and 3(f)], indicating that there is no barrier and the system approaches the global minima. This result is consistent with Lin *et al.* [65].

To get the most probable pathways of the unfolding, we calculate the probability of opening of individual base pair with temperature from the following relation:

$$P_i(T) = \frac{A(N_O, i)}{Z_N(T)}, \tag{4}$$

which are shown in Fig. 4. Here, the index "i" is chosen to be the position ($i = 1, 2, 3, 4, 5, 6$) of base pairs from the

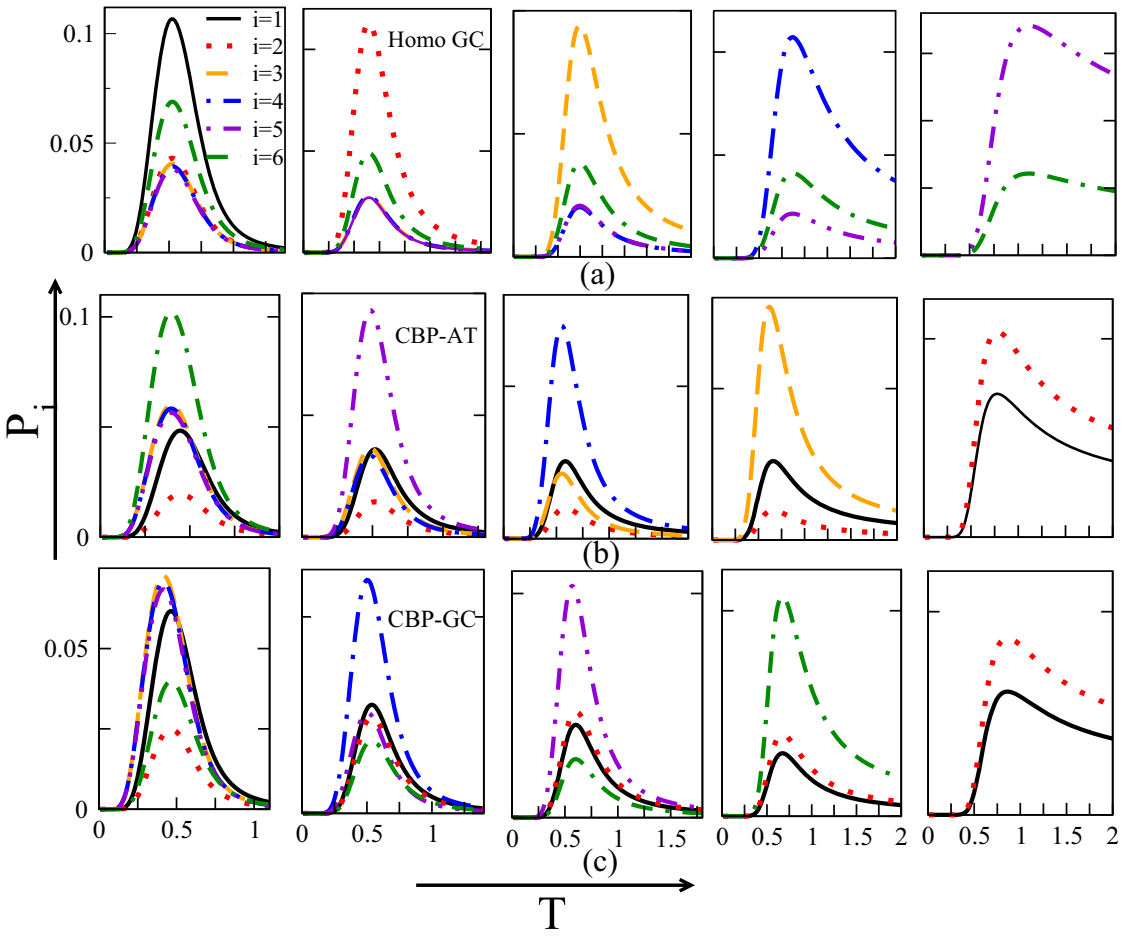

FIG. 4. Probability distribution of opening of base pairs (a) of stem having homo-sequence GC, (b) hairpin having AT as CBP, and (c) hairpin having GC as CBP as a function of temperature. The highest peak value corresponds to the opening of a particular base pair among the closed base pairs.

5′ end (Fig. 5) to the loop end of the hairpin. $A(N_O, i)$ is the partition function of ith bond opening where N_O corresponds to the sum of base pairs opened in the sequence having the information about position of opened base pair and differ from $N - N_P$, which only gives the number of open base pairs but does not have the information about the position of opened base pairs.

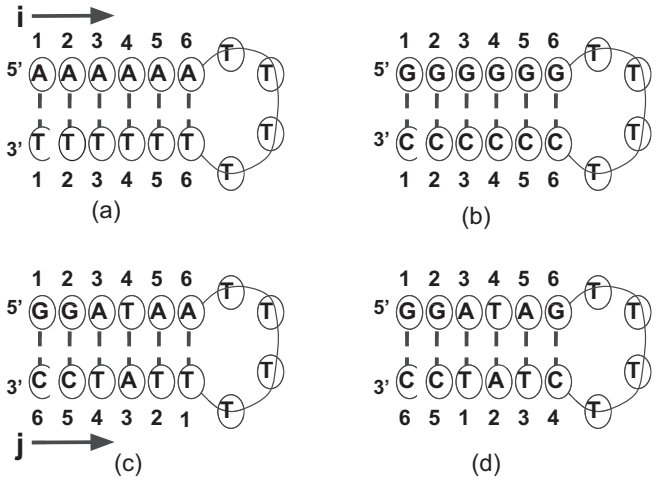

FIG. 5. Most probable opening sequence of hairpins. Top indexing i are the position of base pair and down indexing j are opening position of base pair.

The index "j" corresponds to the opening position of base pairs, i.e., which base pair opens first. By using the Eq. (4), it can be seen in Fig. 4 that the base pair indexed "i" has the maximum probability which corresponds to the weakest base pair and therefore will open first position, which we label by $j = 1$. Once the location of the weakest bond is identified, we next find out the highest probability of opening of the remaining base pairs which we identified as $j = 2$ and so on, keeping the preceding base pair(s) open. From these plots, one can get the information about the opening of the base pair sequence, which are shown in Fig. 5. It can be seen from Figs. 5(a) and 5(b) that for homo-sequences, the DNA hairpin opens from the 5′ end of the stem irrespective of type of CBPs. This is because the entropy associated with the 5′ end nucleotide is maximum compared to the CBP of the loop-end side, as the stem end is free.

Interestingly, Figs. 5(c) and 5(d) show that the opening of base pairs for heterosequences do depend on the type of CBPs. If the CBP is AT [Fig. 5(c)], the hairpin opens from the loop end of the stem. This is because of the fact that the four consecutive base pairs are ATs, which is weaker than the

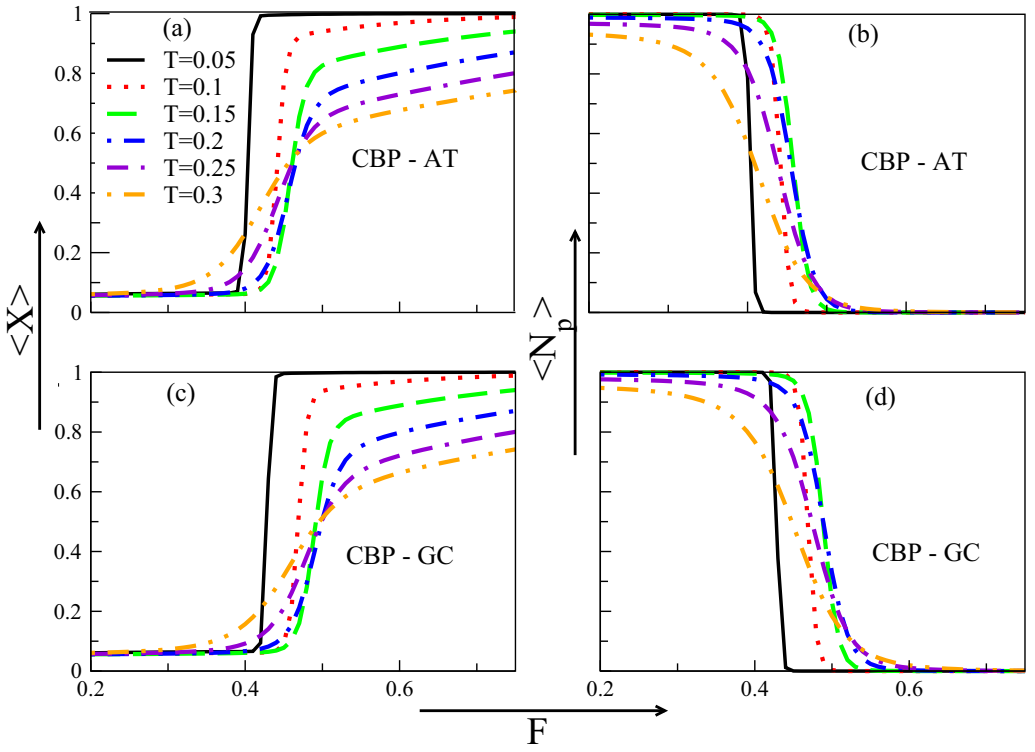

FIG. 6. Figures show the variation of the average extension and average number of base pairs as a function of the applied force at different temperatures. Panels (a, b) are for the DNA hairpin having CBP-AT, and (c, d) are for CBP-GC. The qualitative features remain the same except a shift in the unzipping force towards the right side for the hairpin sequence having CBP-GC.

GC at $5'-$ end of the hairpin. However, when the CBP is GC, one can see that the DNA hairpin opens first at $i = 3$, and there is a possibility of bubble formation as both the stem and CBP sides have GC base pairs, which is quite stronger than the AT base pairs. Although the sixth base pair (CBP) is GC [Fig. 5(d)], it opens at $j = 4$ followed by the base pair at $i = 2(j = 5)$ then $i = 1(j = 6)$ as it is sandwiched between two bubbles formed due to opening of AT base pairs and the loop made up of four thymine bases.

V. FORCE-INDUCED MELTING

Motivated by SMFS experiments [39] and recent simulations [57], now we explore effects of external force on the melting of DNA hairpin. Since, we are considering SASAWs on the FCC lattice in three dimensions, the force is applied at the one end of the chain along the diagonal (X-Y plane), keeping the other complimentary end fixed. This leads the system to undergo a force-induced transition from the zipped (closed) state to the unzipped (open) state. Interestingly, further rise in the force drives the system from the open state to the stretched state, which otherwise would have not been possible. The partition function defined in Eq. (1) can easily be extended to study the system of interest,

$$Z_N(F,T) = \sum_{N_{\rm GC},N_{\rm AT},x} C_N(N_{\rm GC},N_{\rm AT},x) u^{N_{\rm GC}} v^{N_{\rm AT}} \omega^x, \qquad (5)$$

where $C_N(N_{\rm GC},N_{\rm AT},x)$ is the total number of conformations corresponding to the walk of N steps, whose one end is at distance x apart from the fixed end. The Boltzmann weight for the applied force ω is defined as $\exp(F/k_BT)$. The average extension can be obtained from the following relation:

$$\langle x\rangle = \frac{1}{Z}\sum_{N_{\rm GC},N_{\rm AT},x}\sum_{x} x C_N(N_{\rm GC},N_{\rm AT},x) u^{N_{\rm GC}} v^{N_{\rm AT}} \omega^x. \qquad (6)$$

In Figs. 6(a) and 6(c), we have plotted the end-to-end distance x as a function of force F at different temperatures for the DNA hairpin, whose CBPs are either AT or GC. It can be seen that at very low temperature, the transition appears to be of first order, and as we increase the temperature the variation in the extension becomes continuous. We do not observe any qualitative change in case of force-induced melting of DNA hairpin having different CBPs except a shift in the unzipping force to the right for the sequence having CBP as GC. This may be attributed to the base pairing interaction of GC, which is higher compared to AT. In Figs. 6(b) and 6(d), we have shown the variation of $\langle N_P\rangle$ as a function of the applied force at different temperatures. The qualitative nature of the transition remains the same, whether one uses $\langle x\rangle$ or $\langle N_P\rangle$ as an order parameter. We noted that for the homo-sequence of stem (all AT or all GC), the variation of $\langle N_P\rangle$ or $\langle x\rangle$ are qualitatively similar except a right shift in F towards GC side. Thus, the CBP having GC base pair has higher stability and offers more resistance against the applied force. For homo-sequence, we observed that the hairpin opens from the free end of the stem side, and sequence has no such effect as seen in the case of thermal melting.

Danilowicz *et al.* [66] studied the elastic properties of single-stranded DNA (ssDNA) and showed that temperature has a significant impact on the force-extension curve. The

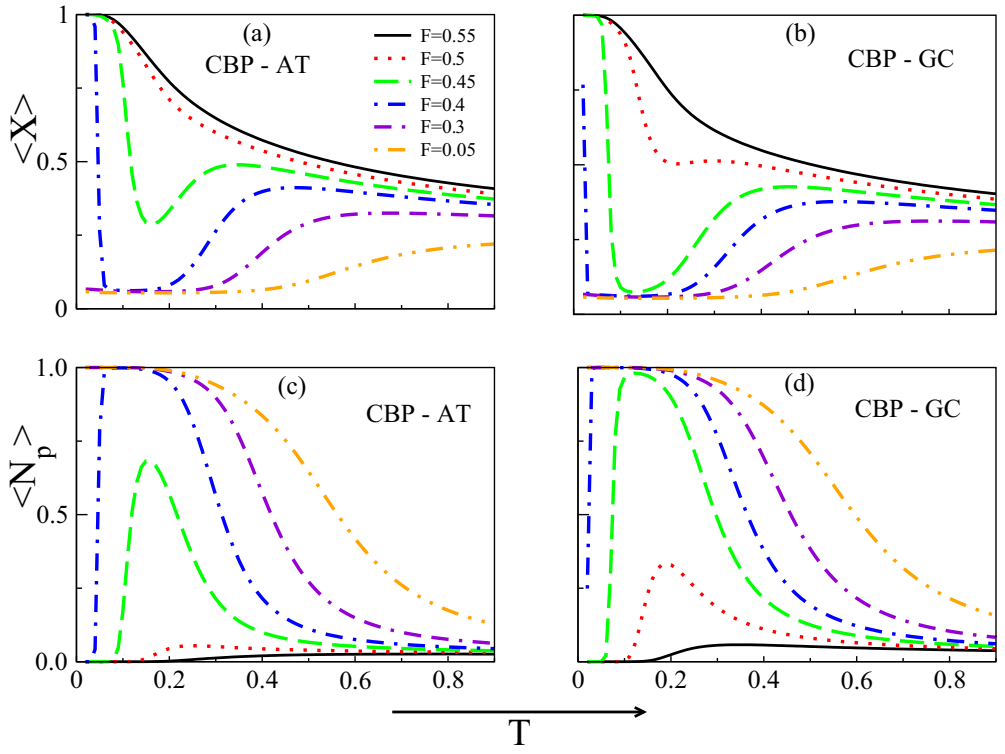

FIG. 7. Same as Fig. 6 but variation is with temperature at different forces. There is a significant difference between the response of AT and GC as CBP. It is apparent that the DNA hairpin remains in the unzipped state at low temperature and high forces. With a rise in temperature, it acquired the conformation of the closed state at the same force, and a further rise in temperature drives the system to the coil state.

extension increases with the temperature in the low force regime, however, for the higher forces, the extension decreases abruptly with the rise of the temperature. Kumar and Mishra [67] using the SAW model of polymer on the square lattice showed that the decrease in extension with temperature is due to the entropic contribution to the free energy. The model developed here is in 3D with higher coordination number and hence has a larger entropy. It would be interesting to study the effect of CBPs on the melting. Figures 7(a) and 7(b) shows the temperature versus extension curves for different CBPs at different forces. At low force, the extension increases with temperature and the hairpin melts. At high force and low temperature, the chain acquires the conformation of the stretched state up to a certain temperature, but as the temperature increases, the hairpin refolds to the closed state because of the increased contribution of entropy to the free energy. With further rise in temperature, the hairpin again melts and acquires the conformations of the swollen state. In Figs. 7(c) and 7(d), we have shown the variation of $\langle N_P \rangle$ as a function of the temperature at different applied force. Here also the qualitative nature of the transition of both CBPs remains the same. Thus, the simple model developed here captured the qualitative behavior as seen in experiments [39].

However, here one can notice the influence of closing base pairs (CBPs) from the plots in Fig. 7. At high force, say 0.55 and low temperature, the DNA hairpin is found to be in the stretched state irrespective of CBPs. As force decreases (say at 0.45), the temperature-extension curve of the hairpin having CBP as AT differs significantly with GC. One can note that the fall in extension decreases for both CBPs, but the magnitude of the fall is higher for the hairpin having CBP as GC compared to AT. In Fig. 8, we have plotted the probability of formation of $1st$ (free stem-end) and $6th$ (loop-end) base pairs as a function of temperature of the sequences having CBPs as AT and GC at the applied force 0.48. It is evident that the probability of closing of both ends (free-stem end and loop end) of the sequence having CBP as GC is much higher than the sequence having CBP as AT. Moreover, one

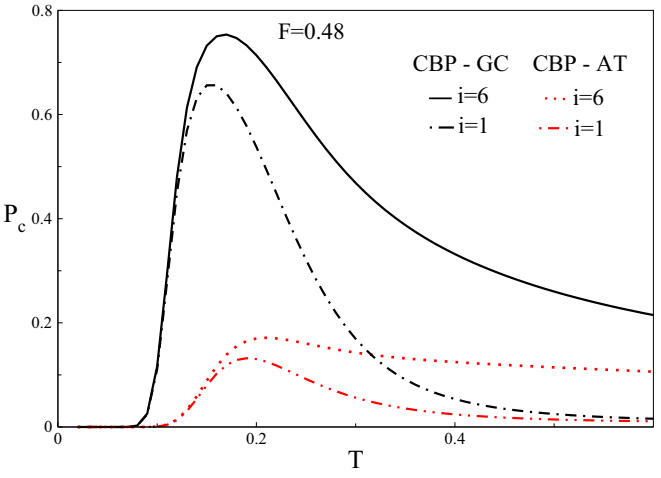

FIG. 8. Figures show comparative plots of probability of closing the free stem end ($i = 1$) and loop end ($i = 6$) base-pairs of the stem part of the sequence having CBP-AT and CBP-GC. In both cases, the closing of hairpin occurs from the loop side.

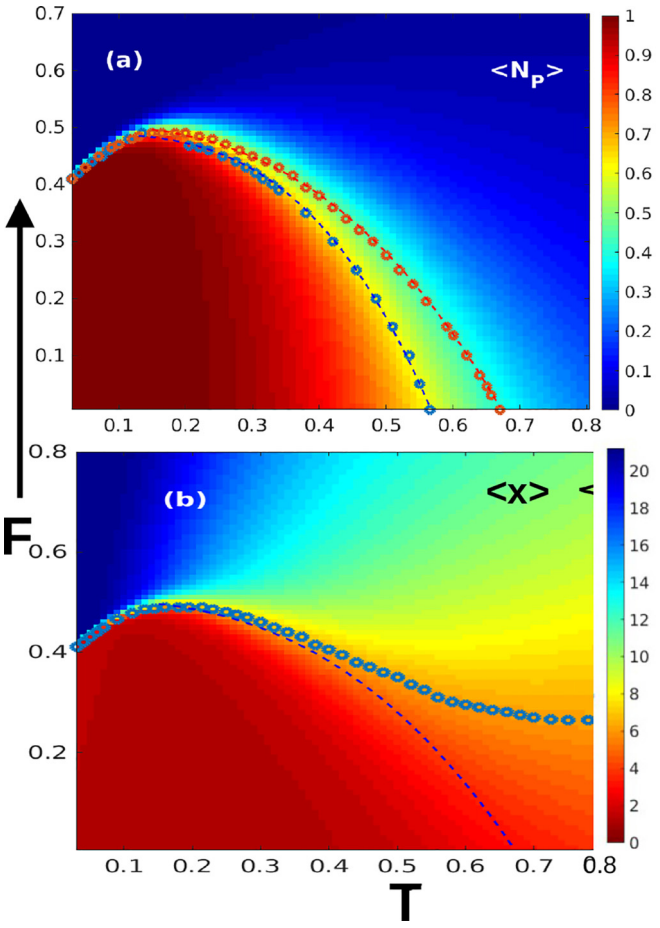

FIG. 9. Force-temperature diagrams show (a) variation of average number of base pairs (color coded) as a function of temperature at constant force or *vice versa*, and (b) variation in average extension (color coded) as a function of the applied force at constant temperature or *vice versa*. In panel (a) the blue circles are obtained from the peak of the contact fluctuation curves with temperature at fixed force and the red circles are obtained by fixing temperature and varying the applied force. In panel (b), the blue circles show the fluctuation in the extension with force at fixed temperature. The dashed line in panels (a) and (b) is a fit by using Eq. (8) is in excellent agreement with the phase boundary obtained from Eqs. (3) and (6) or fluctuation curve obtained from Eq. (7).

can notice that the probability of closing of loop-end (sixth) base pair is higher than the free stem-end (first) base pair for both sequences, which also indicates that the closing of hairpin will be initiated by the loop end side as its probability is higher.

VI. PHASE DIAGRAM

SMFS experiments are mostly performed over finite-size biopolymers, and therefore, the outcome depends crucially on whether the control parameter is the temperature or force or the extension. In constant temperature ensemble, energy used to be the control parameter, therefore, we have calculated $\langle N_P \rangle$ as a function of temperature to study the melting profile shown in Fig. 9(a). However, optical tweezers and AFM essentially control the position of the end monomer where a force is applied. In a constant force ensemble (CFE) the control parameter is the average extension $\langle x \rangle$, whereas for the constant distance ensemble, the control parameter is the average force $\langle F \rangle$. In Fig. 9(b), we plot the $\langle x \rangle$ order parameter as a function of force at fixed temperature. The values of $\langle N_P \rangle$ and $\langle x \rangle$ are color coded shown in the right of the Figs. 9(a) and 9(b). Here, we report the phase diagram of the hairpin having CBP as GC because the qualitative nature of the phase diagram remains the same irrespective of the CBPs.

Since the single-molecule experiments study systems of finite size, therefore, results may depend on the choice of ensembles. It is clear from the plots in Fig. 6 that at low temperature transition is first order (zipped to stretched state), and as the temperature increases it becomes continuous. In Fig 9(a), at high temperature, we observe the signature of intermediate states demarcated by varying colors. Sadhukhan and Bhattacharjee [68] showed that at constant temperature, the applied force does not affect the bound state below a certain force but penetrates up to a range of force followed by the stretched state. In Fig. 9(a), at constant temperature (say 0.5), the hairpin remains in the bound state (represented by the red region). One can see the variation of colours (yellow, green, cyan, . . .), where the bound state gets perturbed due to the applied force. After a certain force, the hairpin acquires the stretched state (the blue region of Fig. 9). As force increases, the melting temperature decreases. However, below the certain temperature, one can notice "reentrance" where decrease in temperature also leads to decrease in the unzipping force. The qualitative features of the phase diagram remains largely the same as those observed in previous studies [69]

The partition function defined in Eq. (5) is sufficient enough to obtain the phase boundaries between different states by calculating the fluctuations $\xi = < O^2 > - < O >^2$ (or specific heat) in the observables, with the kth moment given by

$$\langle O^k \rangle = \frac{1}{Z} \sum_{N_{GC},N_{AT},x} \sum_{x} O^k C_N(N_{GC}, N_{AT}, x) u^{N_{GC}} v^{N_{AT}} \omega^x. \quad (7)$$

For $k = 1$, Eq. (7) reduces to Eq. (3) for $\omega = 1$ and Eq. (6) for $\omega > 1$, where O corresponds to N_p and x, respectively. The transition points shown in the force-temperature diagram [Fig. 9(a)] is obtained from the peak position in the fluctuation in N_P either by keeping force constant and varying the temperature [shown by blue circles in Fig. 9(a)] or keeping temperature constant and varying the force [shown by red circles in Fig. 9(a)]. In both cases, we observed the occurrence of two peaks in the fluctuation that give rise to the reentrance at low temperature, whose physical origin is now well understood [44,50,70]. The different phases shown in the phase diagram [Fig. 9(a)] is indicated by different colours, which are obtained from the variation of $\langle N_P \rangle$ as a function of temperature at constant force or *vice versa*. In Fig. 9(b), we show the variation of $\langle x \rangle$ as a function of applied force at constant temperature, which is also in good agreement with the one obtained from the fluctuations. We fit the locus of bound and open states using the following equation:

$$F_c = f_b \left[1 - \left(\frac{T}{T_m} \right) \right]^{\alpha}, \quad (8)$$

where f_b is the maximum force to open the hairpin at low temperature and α is a fitting exponent [71]. The dashed line

shown in Figs. 9(a) and 9(b) is in excellent agreement with the one obtained from the fluctuation of N_P. The value of α determines whether the transition is weak or strong.

In Fig. 9(b), we have also shown the transitions obtained by fixing the temperature (blue open circles) and monitoring the fluctuation in the extension with force. It is interesting to note that at low temperature, transition is first order. However, at high temperatures, the F-T diagram differs significantly. One of the reasons for the observed difference might be due to the definition of the transition point. In case of thermal melting, the temperature at which half of the base pairs are open is termed melting temperature. In this case it is difficult to know whether the DNA hairpin is opening from the loop side or stem end side. Hence, it is possible that the hairpin may be closed at the stem side, but due to bubble formation, half of the base pairs get open and one gets the signature of melting. However, when we monitor extension, the hairpin opens from the free end of the stem and unless the CBP gets open, we do not get the signature of the unzipping.

More importantly, the fluctuation in the extension does not show transition near $F < 0.25$. This may be understood by recalling that the discontinuity in the free energy or its derivative as a function of the thermodynamic parameter gives the signature of transition. Since above the temperature $T = 0.6$ most of the bonds are open, and the applied force may be below the stochastic force, therefore, there is no discontinuity in the free energy. Hence, extension may be an appropriate parameter to study first order transitions at low temperature only, and may not give correct description of transition at higher temperatures.

VII. CONCLUSION

Despite the simplicity, the model studied here is quite successful in predicting local fluctuations, large-scale conformational motion, and unfolding behavior of DNA hairpin. Our study revealed that during the thermal melting DNA hairpin acquires low-entropy state (bound state) to a high-entropy (melted) state. Whereas force-induced melting leads the system from one low-entropy state (bound state) to the another low-entropy state (stretched state), where high-entropy (melted) state appears as an intermediate state. Using the probability distribution analysis of individual base-pairs, we first time obtained the information about the most probable pathways of opening of individual base pairs. For thermal melting, the hairpin sequence having CBP-AT, opens from the loop end side, whereas for CBP-GC, there is a possibility of formation of a bubble and chain opens from the middle of the stem in the form of a bubble. For the homo sequence, the hairpin always opens from the stem end side irrespective of the type of CBPs.

The method developed here along with Table I in Appendix can provide thermodynamic properties of any base sequence of length 15. The exact density of states given in Table I allowed us to calculate the free-energy landscape at different temperatures. At low temperature, the system shows two-state behavior of melting between bound and open states with a well defined energy barrier. At high temperature, it shows the monotonic two states unfolding of the hairpin. We showed the existence of the intermediate state around the melting temperature. Our results are consistent with the earlier molecular dynamic simulations and kinetic intermediates structure (KIS) model [65].

The force extension (F-x) curves show mechanical properties (stability) of the sequence, which differ from the thermal melting. The force temperature diagram obtained in different ensembles revealed that SMFS experiment results will depend on the choice of setup. There is no qualitative change in force extension curves for the hairpin sequences with different CBPs (AT/GC) at different temperatures except shift in the scale at the right side for the sequence having CBP as GC. The most remarkable observation is the response of the temperature on the extension at constant force which reveals the microscopic details of opening of the hairpin that depends on the type of CBPs.

It is pertinent to mention here that Woodside *et al.* [42] have studied kinetics and thermodynamics of the unfolding transition of hairpin in detail by varying the stem length, loop length and changing the GC contents in the constant force ensemble. At this stage, our results warrant similar experiments by varying temperature at constant force to explore the melting of hairpins of different stem and loop sizes having GC as a CBP. The proposed experiment may work as a potential candidate to observe the reentrance, which remains elusive in the experimental studies. This inference is based on the probability distribution analysis, which revealed that using GC as CBP in the hairpin may be an ideal candidate to observe reentrance.

ACKNOWLEDGMENTS

Financial assistance from the SERB, New Delhi, SPARC, IoE (BHU), MoE, New Delhi, and UGC, New Delhi is gratefully acknowledged.

APPENDIX

TABLE I. Table shows all possible (64) combinations of open (0) and closed (1) base pairs and its corresponding number of conformations (C_N). Here "i" is the position of base pairs: $i = 1$ corresponds to the position of stem-end base pair, whereas $i = 6$ corresponds to the loop-end base pair (Fig. 5). Sum of the numbers (0 and 1) of a row gives the information about the number of closed base pairs ($N_P = N_{GC} + N_{AT}$) and its corresponding $C_N(N_{GC}, N_{AT})$.

$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	C_N
0	0	0	0	0	0	1705109852909220
0	0	0	0	0	1	68838594367848
0	0	0	0	1	0	24519791398944
0	0	0	0	1	1	9094511944680
0	0	0	1	0	0	14016464422224
0	0	0	1	0	1	1831422657120
0	0	0	1	1	0	3505561646376
0	0	0	1	1	1	1225181071656
0	0	1	0	0	0	10149248529120
0	0	1	0	0	1	860730178848
0	0	1	0	1	0	711707532000
0	0	1	0	1	1	262570788528

TABLE I. (*Continued.*)

$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	C_N
0	0	1	1	0	0	2109000442656
0	0	1	1	0	1	270149269344
0	0	1	1	1	0	494331638664
0	0	1	1	1	1	169233992424
0	1	0	0	0	0	9327519141456
0	1	0	0	0	1	634913549328
0	1	0	0	1	0	377660878656
0	1	0	0	1	1	142146861552
0	1	0	1	0	0	476059143264
0	1	0	1	0	1	62670981408
0	1	0	1	1	0	118398156000
0	1	0	1	1	1	40865993184
0	1	1	0	0	0	1645015720944
0	1	1	0	0	1	139556249064
0	1	1	0	1	0	113391806640
0	1	1	0	1	1	41494717920
0	1	1	1	0	0	318158087184
0	1	1	1	0	1	40515450912
0	1	1	1	1	0	73401632328
0	1	1	1	1	1	24844491576
1	0	0	0	0	0	13381479775200
1	0	0	0	0	1	827239780872
1	0	0	0	1	0	430302439776
1	0	0	0	1	1	162740378616
1	0	0	1	0	0	409075822848
1	0	0	1	0	1	54162414720

TABLE I. (*Continued.*)

$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	C_N
1	0	0	1	1	0	103830486216
1	0	0	1	1	1	36140079192
1	0	1	0	0	0	616360010688
1	0	1	0	0	1	52740089424
1	0	1	0	1	0	43525224144
1	0	1	0	1	1	15915630624
1	0	1	1	0	0	126364953360
1	0	1	1	0	1	16216039248
1	0	1	1	1	0	29533203888
1	0	1	1	1	1	10023802704
1	1	0	0	0	0	2216073064968
1	1	0	0	0	1	150936295776
1	1	0	0	1	0	89479745616
1	1	0	0	1	1	33613194504
1	1	0	1	0	0	110706362544
1	1	0	1	0	1	14486612256
1	1	0	1	1	0	27397037664
1	1	0	1	1	1	9425123952
1	1	1	0	0	0	361816065960
1	1	1	0	0	1	30880874184
1	1	1	0	1	0	25123802832
1	1	1	0	1	1	9166607712
1	1	1	1	0	0	69988619760
1	1	1	1	0	1	8866199088
1	1	1	1	1	0	16031999256
1	1	1	1	1	1	5411291208

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell* (W.W. Norton & Company, New York, 2007).

[2] O. C. Uhlenbeck, Nature **346**, 613 (1990).

[3] J. Doherty and M. Guo, in *Encyclopedia of Cell Biology* (Elsevier, Amsterdam, 2016), pp. 309–340.

[4] W. K. Olson, S. Li, T. Kaukonen, A. V. Colasanti, Y. Xin, and X.-J. Lu, Biochemistry **58**, 2474 (2019).

[5] M. R. Giese, K. Betschart, T. Dale, C. K. Riley, C. Rowan, K. J. Sprouse, and M. J. Serra, Biochemistry **37**, 1094 (1998).

[6] D. B. Roth, J. P. Menetski, P. B. Nakajima, M. J. Bosma, and M. Gellert, Cell **70**, 983 (1992).

[7] A. K. Kennedy, A. Guhathakurta, N. Kleckner, and D. B. Haniford, Cell **95**, 125 (1998).

[8] E. Zazopoulos, E. Lalli, D. M. Stocco, and P. Sassone-Corsi, Nature **390**, 311 (1997).

[9] S. Froelich-Ammon, K. Gale, and N. Osheroff, J. Biol. Chem. **269**, 7719 (1994).

[10] T. Q. Trinh and R. R. Sinden, Genetics **134**, 409 (1993).

[11] J. yan Tang, J. Temsamani, and S. Agrawal, Nucleic Acids Res. **21**, 2729 (1993).

[12] G. Varani, Annu. Rev. Biophys. Biomol. Struct. **24**, 379 (1995).

[13] H. A. Heus and A. Pardi, Science **253**, 191 (1991).

[14] P. Svoboda and A. D. Cara, Cell. Mol. Life Sci. **63**, 901 (2006).

[15] Y. Zhang, in *Encyclopedia of Systems Biology* (Springer, New York, 2013), pp. 875–876.

[16] S. Kumar, D. Giri, and Y. Singh, Europhys. Lett. **70**, 15 (2005).

[17] M. Mosayebi, F. Romano, T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Phys. Chem. B **118**, 14326 (2014).

[18] S. V. Kuznetsov, C.-C. Ren, S. A. Woodson, and A. Ansari, Nucleic Acids Res. **36**, 1098 (2007).

[19] S. Kumar and G. Mishra, Phys. Rev. Lett. **110**, 258102 (2013).

[20] P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University, Ithaca, NY, 1979).

[21] C. Vanderzande, *Lattice Models of Polymers* (Cambridge University Press, Cambridge, UK, 1998).

[22] S. Kumar and M. S. Li, Phys. Rep. **486**, 1 (2010).

[23] M. M. Senior, R. A. Jones, and K. J. Breslauer, Proc. Natl. Acad. Sci. USA **85**, 6242 (1988).

[24] G. Vesnaver and K. J. Breslauer, Proc. Natl. Acad. Sci. USA **88**, 3569 (1991).

[25] S. S. Chan, K. J. Breslauer, R. H. Austin, and M. E. Hogan, Biochemistry **32**, 11776 (1993).

[26] P. Doty, J. Cell. Comp. Physiol. **49**, 27 (1957).

[27] C. J. Wienken, P. Baaske, S. Duhr, and D. Braun, Nucleic Acids Res. **39**, e52 (2011).

[28] C. A. Gelfand, G. E. Plum, S. Mielewczyk, D. P. Remeta, and K. J. Breslauer, Proc. Natl. Acad. Sci. USA **96**, 6113 (1999).

[29] G. Bonnet, O. Krichevsky, and A. Libchaber, Proc. Natl. Acad. Sci. USA **95**, 8602 (1998).

[30] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, Phys. Rev. Lett. **90**, 138101 (2003).

[31] M. Manghi and N. Destainville, Phys. Rep. **631**, 1 (2016).

[32] S. Srivastava and Y. Singh, Europhys. Lett. **85**, 38001 (2009).

[33] K. Chauhan, A. R. Singh, S. Kumar, and R. Granek, J. Chem. Phys. **156**, 164907 (2022).
[34] H. Ma, D. J. Proctor, E. Kierzek, R. Kierzek, P. C. Bevilacqua, and M. Gruebele, J. Am. Chem. Soc. **128**, 1523 (2006).
[35] R. Narayanan, L. Zhu, Y. Velmurugu, J. Roca, S. V. Kuznetsov, G. Prehna, L. J. Lapidus, and A. Ansari, J. Am. Chem. Soc. **134**, 18952 (2012).
[36] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997).
[37] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. E **58**, 2386 (1998).
[38] G. U. Lee, L. A. Chrisey, and R. J. Colton, Science **266**, 771 (1994).
[39] C. Danilowicz, Y. Kafri, R. S. Conroy, V. W. Coljee, J. Weeks, and M. Prentiss, Phys. Rev. Lett. **93**, 078101 (2004).
[40] D. Lohr, R. Bash, H. Wang, J. Yodh, and S. Lindsay, Methods **41**, 333 (2007).
[41] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).
[42] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, Proc. Natl. Acad. Sci. **103**, 6190 (2006).
[43] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, Science **276**, 1109 (1997).
[44] D. Marenduzzo, S. M. Bhattacharjee, A. Maritan, E. Orlandini, and F. Seno, Phys. Rev. Lett. **88**, 028102 (2001).
[45] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959).
[46] T. L. Hill, J. Chem. Phys. **30**, 383 (1959).
[47] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1456 (1966).
[48] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
[49] Y. Singh, S. Kumar, and D. Giri, J. Phys. A: Math. Gen. **32**, L407 (1999).
[50] S. Kumar, I. Jensen, J. L. Jacobsen, and A. J. Guttmann, Phys. Rev. Lett. **98**, 128101 (2007).
[51] S. Kumar and G. Mishra, Soft Matter **7**, 4595 (2011).
[52] A. R. Singh and R. Granek, J. Chem. Phys. **145**, 144101 (2016).
[53] A. Raj Singh and R. Granek, Phys. Rev. E **96**, 032417 (2017).
[54] T. A. Knotts, N. Rathore, D. C. Schwartz, and J. J. de Pablo, J. Chem. Phys. **126**, 084901 (2007).
[55] J. A. Bueren-Calabuig, C. Giraudon, C. M. Galmarini, J. M. Egly, and F. Gago, Nucleic Acids Res. **39**, 8248 (2011).
[56] S. Piana, J. Phys. Chem. A **111**, 12349 (2007).
[57] A. Upadhyaya and S. Kumar, Phys. Rev. E **103**, 062411 (2021).
[58] S. Kannan and M. Zacharias, Biophys. J. **93**, 3218 (2007).
[59] S. Kannan and M. Zacharias, Nucleic Acids Res. **39**, 8271 (2011).
[60] G. Portella and M. Orozco, Angew. Chem., Int. Ed. **49**, 7673 (2010).
[61] R. D. Schram, G. T. Barkema, R. H. Bisseling, and N. Clisby, J. Stat. Mech.: Theory Exp. (2017) 083208.
[62] S. McKenzie, J. Phys. A: Math. Gen. **12**, L267 (1979).
[63] P. M. Vallone, T. M. Paner, J. Hilario, M. J. Lane, B. D. Faldasz, and A. S. Benight, Biopolymers **50**, 425 (1999).
[64] A. Srivastava and R. Granek, Phys. Rev. Lett. **110**, 138101 (2013).
[65] M. M. Lin, L. Meinhold, D. Shorokhov, and A. H. Zewail, Phys. Chem. Chem. Phys. **10**, 4227 (2008).
[66] C. Danilowicz, V. W. Coljee, C. Bouzigues, D. K. Lubensky, D. R. Nelson, and M. Prentiss, Proc. Natl. Acad. Sci. USA **100**, 1694 (2003).
[67] G. Mishra, D. Giri, and S. Kumar, Phys. Rev. E **79**, 031930 (2009).
[68] P. Sadhukhan and S. M. Bhattacharjee, Indian J. Phys. **88**, 895 (2014).
[69] C. Danilowicz, C. H. Lee, V. W. Coljee, and M. Prentiss, Phys. Rev. E **75**, 030902 (2007).
[70] D. Marenduzzo, A. Maritan, A. Rosa, and F. Seno, Phys. Rev. Lett. **90**, 088301 (2003).
[71] C. Hyeon and D. Thirumalai, Proc. Natl. Acad. Sci. USA **102**, 6789 (2005).