AIP | The Journal of Chemical Physics

The effect of topology on the structure and free energy landscape of DNA kissing complexes

Flavio Romano, Alexander Hudson, Jonathan P. K. Doye, Thomas E. Ouldridge, and Ard A. Louis

Citation: The Journal of Chemical Physics **136**, 215102 (2012); doi: 10.1063/1.4722203
View online: http://dx.doi.org/10.1063/1.4722203
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/136/21?ver=pdfcov
Published by the AIP Publishing

The effect of topology on the structure and free energy landscape of DNA kissing complexes

Flavio Romano,[1] Alexander Hudson,[1,a)] Jonathan P. K. Doye,[1,b)] Thomas E. Ouldridge,[2] and Ard A. Louis[2]

[1]*Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford, OX1 3QZ, United Kingdom*

[2]*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford, OX1 3NP, United Kingdom*

(Received 16 March 2012; accepted 11 May 2012; published online 5 June 2012)

We use a recently developed coarse-grained model for DNA to study kissing complexes formed by hybridization of complementary hairpin loops. The binding of the loops is topologically constrained because their linking number must remain constant. By studying systems with linking numbers −1, 0, or 1 we show that the average number of interstrand base pairs is larger when the topology is more favourable for the right-handed wrapping of strands around each other. The thermodynamic stability of the kissing complex also decreases when the linking number changes from −1 to 0 to 1. The structures of the kissing complexes typically involve two intermolecular helices that coaxially stack with the hairpin stems at a parallel four-way junction. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4722203]

I. INTRODUCTION

Not only is DNA the genetic information carrier of life, but, given the degree of control achieved in the chemistry of DNA (Ref. 1) – molecule synthesis is fast, reliable, and relatively cheap – these information-rich building blocks can be exploited to reliably self-assemble two- and three-dimensional structures[2–4] and to build functional nanodevices.[5]

Hairpins probably represent the simplest structure that DNA can form besides the standard double helix. These are secondary structure motifs formed by single-stranded DNA molecules that have complementary regions that self-hybridize. The intramolecular double helix formed from the self-complementary sections is known as stem or neck, while the section that connects two of the stem ends is called a loop.

In contrast to RNA, which for the most part is single-stranded *in vivo*, so that hairpins are a common structural element,[6] DNA *in vivo* is mostly in its duplex form. Nevertheless, there are occasions when it is single-stranded, and examples have been identified where DNA hairpins play a biological role,[7] including in replication, transcription, and recombination.[8,9] However, hairpin formation can sometimes be an undesired process and has been implicated in certain diseases.[10,11]

DNA hairpins also play an important role in DNA nanotechnology[12–18] and can be used as the "fuel" to provide the energy to power autonomous DNA motors, although they can also be an unwanted secondary structural motif in DNA designed to be unstructured.[19]

In this paper, we study "kissing" complexes that can form when the loops of two DNA hairpins are complementary and partially hybridize. In particular, we focus on the interplay between topology and the structure and stability of these complexes. For example, when two hairpin loops hybridize, the right-handed wrapping of the DNA strands in the intermolecular double helix must be compensated by a region where the loops wrap around each other in the opposite sense. Thus, even when the two loops are fully complementary, topological effects will restrict the number of bonds that can be formed.

Kissing loop interactions are also an important RNA tertiary structure motif[20] and play key biological roles in processes such as the regulatory action of antisense RNAs and the dimerization of viral genomic RNA.[21] They have, therefore, been much better characterized for RNA, both in terms of their structure[22–24] and mechanical properties,[25,26] than for DNA. This structural knowledge has even been exploited in structural RNA nanotechnology, where kissing loop interactions have also been used as a means to join RNA components with a well-defined geometry.[27–29] Although these RNA systems provide an interesting comparison, the kissing loop interactions typically involve shorter sequences of complementary bases than for the DNA systems we consider here, and so topological effects are less significant. Interestingly, the NMR solution structure of a DNA kissing complex has been obtained for sequences analogous to that of a previously characterized RNA kissing complex.[30] Although there are differences in the details of the two structures, they are generally very similar.

The topological effects associated with kissing loop interactions have been exploited in DNA nanotechnology, particularly to allow the design of autonomous motors.[13,14,31,32] One way of driving a DNA nanodevice through a cycle is through the use of complementary single-stranded DNA strands as "fuel", where the first strand is designed to partially hybridize

a)Present address: Department of Chemistry, University of California, Berkeley, California 94720, USA

b)Author to whom correspondence should be addressed. Electronic mail: jonathan.doye@chem.ox.ac.uk.

with the device to induce a conformational change, and the second then reverses this change by displacing it to form a "waste" duplex. The first example of such a device was DNA nanotweezers, where the strands induced the device to open and close.[33] However, one problem with such a device is the two strands have to be added sequentially, since, if both are present at the same time, they will preferentially directly hybridize with each other rather than with the device.

One way to circumvent this problem is through the use of fuel strands that can form hairpins,[13,14] since the topological restriction on binding between the hairpin loops will effectively prevent the duplexes from being formed, even though the duplexes are more stable. Given that the two strands are complementary, these hairpins naturally form kissing complexes. While the hairpins are unable to open each other's stems by displacement, the motor can be designed to be a catalyst for the hybridization. By having a single-stranded DNA section that both is partially complementary to one of the hairpins and has a free end, the motor can open the hairpin by displacement, unconstrained by topological effects. A second, and similar, solution is to prepare the fuel strands complexed to partially complementary protective strands that bind to either end of the fuels but leave a loop region in the middle unhybridized.[31,32] A variety of autonomous motors have been designed based on these principles.[15–18]

Here, we investigate the system of fully complementary 40-base DNA hairpins studied by Bois *et al.*[13] using computer simulations of a nucleotide-level coarse-grained model of DNA.[19,34] This recently introduced model provides an excellent description of the structural, thermodynamic, and mechanical properties of both single-stranded and duplex DNA, and has now made it feasible to study the free energy landscapes of such DNA nanotechnology systems in detail, as previously illustrated for DNA nanotweezers.[19] In particular, we focus on the effects of topology on the free energy landscape for the binding of the hairpin loops, and how the structure of the resulting kissing complex reflects these topological constraints. To further illustrate the role played by the topology, we also consider kissing complex formation in systems of linked hairpins.

II. MODEL AND METHODS

A. Model

We use the coarse-grained DNA model developed by Ouldridge *et al.*[19,34] In this model, a DNA strand is described as a polymer of nucleotides that interact via excluded volume repulsions and anisotropic attractive potentials that mimic Watson-Crick base-pairing, stacking, cross stacking, and coaxial stacking. The model has been parameterized to reproduce the structural and thermodynamic properties of single-stranded and double-stranded DNA molecules at the high salt concentrations that are typically used in DNA nanotechnology applications. Since this model is described in detail in Ref. 34 we shall repeat here only the fundamental ingredients.

Each nucleotide is represented as a rigid body with three interaction sites, all on the same axis. Although the interaction sites are collinear, we stress that a nucleotide does not possess cylindrical symmetry, since the potential also depends on a vector perpendicular to the nucleotide axis to capture the effects of the orientation of a base on the interactions.

The potential energy V can be written as

$$V = \sum_{\text{nn}} (V_{\text{backbone}} + V_{\text{stack}} + V'_{\text{exc}}) + \sum_{\text{other pairs}} (V_{\text{HB}} + V_{\text{cross.stack}} + V_{\text{coaxstack}} + V_{\text{exc}}). \quad (1)$$

The first sum runs over all pairs of nucleotides that are adjacent along a strand (neighbours in our terminology) and the second sum runs over all other pairs. The interaction between neighbours consists of a backbone term that is designed to represent the connectivity of a DNA strand, a stacking term that is designed to mimic stacking interaction between nucleotides, and an excluded volume part that prevents nucleotides from overlapping. The interaction between non-neighbouring pairs consists of four different terms: (i) a hydrogen-bonding term that mimics directional Watson-Crick base-pairing; (ii) a cross-stacking term that accounts for stacking interactions between nucleotides that are second neighbours on different strands; (iii) a coaxial stacking term that is designed to capture stacking interactions between non-neighbouring bases; and (iv) an excluded volume term. We should note that the first three terms are not isotropic but depend on the relative orientations of the nucleotides. It is this angular modulation that ensures the double helix is right handed and the strands pair in an anti-parallel fashion. The full forms of each of these terms has been reported in Ref. 34, with the exception of the coaxial stacking term, which is described in Ref. 35 and whose parameterization will be discussed in detail in Ref. 36. A simulation code incorporating the potential can now be accessed from Ref. 37.

Features of the model that are particularly important for the current study are the relative flexibility of single-stranded DNA and its ability to describe the thermodynamics of hairpin formation accurately, as well as hybridization in general. We are also confident of the general robustness of the model, based on the wide range of DNA systems on which we[19,34,35] and others[38] have tested the model. These so far include DNA nanotweezers,[33] "burnt bridges",[39] and two-footed[40] DNA walkers, as well as processes such as DNA displacement,[41] overstretching,[42] cruciform formation,[43] and the formation of liquid crystalline phases.[44]

However, we should also note that the model does introduce a significant level of coarse graining and neglects several features of the DNA structure and interactions. First, all four nucleotides have the same structure and interaction properties, except for the hydrogen-bonding term, for which interactions are only allowed between Watson-Crick complementary bases. Although this approximation of course precludes the study of much of the sequence dependence of properties and behaviour, it is not a problem when, as here, we are interested in the generic behaviour of a system. Second, the double helix in our model is symmetrical rather than having different sizes for the minor and major grooves. Again this is

unlikely to be an issue, unless we are interested in the DNA structure at a quite fine level of detail. Finally, the interactions have been fitted for a single, fairly high, salt concentration (namely, 0.5 M), where the Debye screening length is short. This is the regime relevant to most DNA nanotechnology experiments.

B. Simulation methods

Throughout this work, we use Monte Carlo simulations employing the virtual move algorithm (VMMC) introduced by Whitelam and co-workers.[45,46] The latter is a modification to the standard (Metropolis) Monte Carlo algorithm specifically designed to promote the collective diffusion of strongly interacting clusters that would otherwise be suppressed. In our model, DNA strands are effectively clusters of interacting nucleotides and thus VMMC significantly speeds up sampling, particularly when strand diffusion is important, which is the case when studying hybridization processes. Our implementation of VMMC reflects that we are using it to enhance sampling rather than as a dynamical Monte Carlo method. Although VMMC can indeed be implemented in a way that mimics realistic dynamics by generating cluster moves with clusters distributed as N^{-1}, since we are interested only in the thermodynamic properties of the system, we do not implement this part of the algorithm, accepting or rejecting cluster moves solely on energetic grounds. In our implementation, each cluster move is either a rotation or a translation, chosen randomly. Rotations and translations are selected from a Gaussian distribution with standard deviation of 0.085 nm and 0.1 radians, respectively. These values have been chosen to lead to an acceptance of roughly 40% of the attempted moves. To avoid large moves that possibly break topology constraints, we automatically reject a move if it causes any particle to move more than 0.4 nm.

Because of the presence of large free-energy barriers, we have used umbrella sampling[47] to accurately sample transitions between different states. In practice, this is accomplished by adding an additional term to the system Hamiltonian designed to flatten the free energy profile along a particular reaction coordinate, and then subsequently unbiasing the results.[48] In the present case, the natural choice for the reaction coordinate is the number of intermolecular base pairs that are part of the target structure. This choice requires a definition of base pairing, and we define a pair of nucleotides as base paired if the hydrogen bonding interaction term between them is at least 0.093 times the well depth. Of course this choice is somewhat arbitrary, but changing the threshold does not significantly alter the results.

In our umbrella sampling implementation, we start preproduction simulations with a flat bias (equivalent to no bias) along the reaction coordinate, and iteratively adjust the bias to achieve a flat sampling. Once a bias that leads to roughly flat sampling is found, we start production runs with fixed bias to collect statistics. For each of the free energy profiles we discuss later, at least ten parallel runs of 10^{11} cluster moves were required. It should be pointed out that using the number of base pairs as a reaction coordinate is a good choice to sample the free energy profile apart from the complete attachment/detachment transition. This is because the bias in this case does not speed up the diffusion-limited attachment process. The first free energy jump in our profiles is thus the most difficult interval to sample along the reaction coordinate, and we have taken special care that this part of the free energy profile is properly equilibrated.

C. DNA sequences

We have studied the same 40-base nucleotide sequences as in Ref. 13. The two DNA strands are fully complementary, and so can form a duplex as well as hairpins with a stem of 10 base pairs and a loop of 20 bases. The sequences of the two strands are, in 5′ to 3′ direction

```
gcgttgctgc-attttactcttctcccctcg-
  gcagcaacgc
```

and

```
gcgttgctgc-cgaggggagaagagtaaaat-
  gcagcaacgc
```

where the hyphens separate stem and loop regions. All the results we present are at room temperature, taken as $T = 296.15$ K (23 °C). This compares to a hairpin melting temperature of around 350 K. Hence, at the temperature we consider, the probability of spontaneous hairpin opening is effectively zero in our simulations.

D. DNA topology

A commonly used property to characterize systems with respect to their topology is the linking number, Lk, a number that describes how two closed curves are linked in three-dimensional space. Given the projection of two closed curves onto any plane, a crossing is taken to be positive (negative) if the upper curve can be superimposed onto the lower by a counterclockwise (clockwise) rotation (see Fig. 1(a)). According to this definition, each crossing in the right-handed

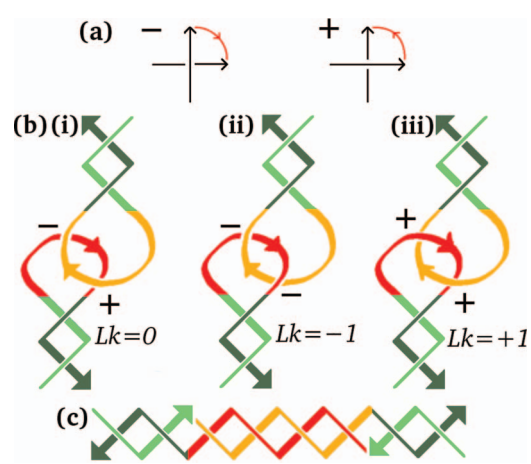

FIG. 1. (a) The definitions that we use for the signs associated with the crossing of two curves. Schematic representations of (b) the three topological configurations of DNA hairpins studied in this paper and (c) a control system with no topological or geometric constraints. The different topological configurations in (b) are: (i) Topologically unlinked, linking number $Lk = 0$. (ii) Topologically favoured, $Lk = -1$. (iii) Topologically frustrated $Lk = +1$.

helix formed by dsDNA is negative as the strands in the helix are antiparallel.[49] The linking number Lk is then defined as

$$Lk = \frac{1}{2}\sum_i c_i, \tag{2}$$

where the index i runs over the crossings with $c_i = +1$ for a positive crossing and $c_i = -1$ for a negative crossing. In more intuitive terms, Lk is the number of times that each curve wraps around the other. For a detailed discussion of the linking number in the context of nucleic acids we refer the reader to Ref. 50.

Although we are considering a system of two unclosed DNA strands, because the rate of hairpin opening is negligible in our simulations at the temperature we consider, the hairpin loops can be effectively considered as closed loops, and so can exhibit topologically different states (Fig. 1(b)). The most experimentally relevant state is that with $Lk = 0$, as the likelihood that two hairpin loops would interlink during their formation process is very low. Since linking number is conserved, in this state any negative crossings of the two hairpin loops due to hybridization of the complimentary loops must be compensated by positive crossings, i.e., sections where the loops wind around each other in the opposite sense. This topological effect will frustrate the hybridization process.

III. RESULTS

A. Topologically unlinked complex; $Lk = 0$

First, we consider two hairpins that are topologically unlinked (Fig. 1(b)(i)) and are free to bind through the complementary loop regions. One of the unbound hairpins is illustrated in Fig. 2(a). The room temperature free energy profile for bonding is reported in Fig. 3. The jump associated with the formation of the first bond is due to the loss of translational entropy associated with hybridization and is dependent on concentration as well as temperature. Our data were collected with two strands in a volume of 4944 nm^3, i.e., 0.336 mM.

For the hybridization of a duplex in our model, we have previously shown that, after the barrier for forming the first base pair, there is then a linear decrease in the free energy as the number of base pairs increases (aside from a possible small rise at the end due to the fraying of the ends of the duplex).[34] The behaviour seen for the binding of the two hairpin loops is significantly different from this scenario. After an initial roughly linear decrease (up to about 3–4 base pairs), the line begins to exhibit some positive curvature reaching a minimum at 14 base pairs (roughly one and a half helical turns) before rising steeply. This curvature is a result of the topological constraint that as the two hairpin loops wind around each other to form a duplex the linking number must remain constant, and this constraint is increasingly felt as the number of base pair increases.

To compare the thermodynamics to a system where topological constraints do not play a role we introduce a control system where the hairpin loops have been opened by breaking the backbones of both strands between bases 10 and 11, as shown in Fig. 1(c). The free energy profile of the control in the bound state has been computed and shows the expected

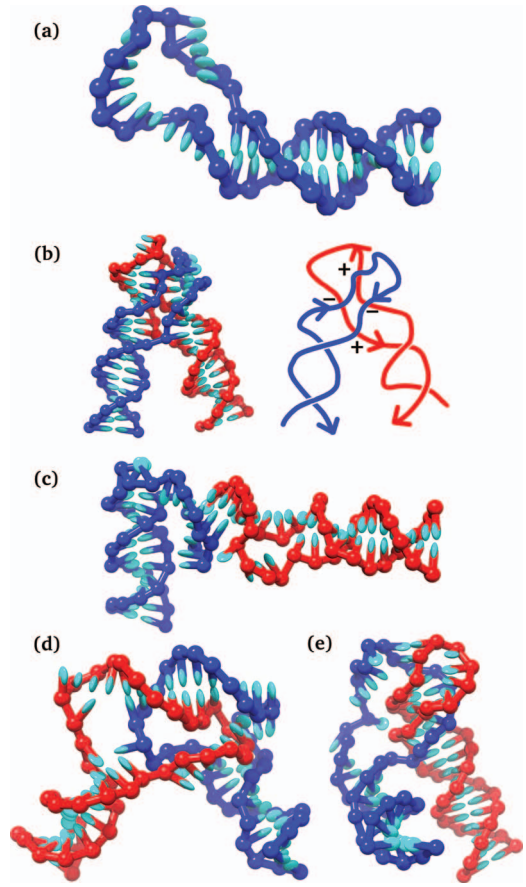

FIG. 2. Typical structures assumed by (a) a single hairpin and (b) the topologically unlinked ($Lk = 0$) kissing complex. Note that in (a) it almost looks as if the stem is longer than 10 base pairs, because the stacking tends to propagate beyond the end of the stem at this temperature. For (b) the chosen structure has 14 interstrand base pairs. To its right is a topological sketch of the configuration illustrating that the zero linking number is achieved by balancing positive and negative crossings. In panels (c)–(e) we show example structures for partially formed complexes with a total of 2, 6, and 10 base pairs formed between the loops, respectively. In our visualisation of the DNA structures, each backbone site is represented by a sphere and each base by an ellipsoid connected to the backbone site.

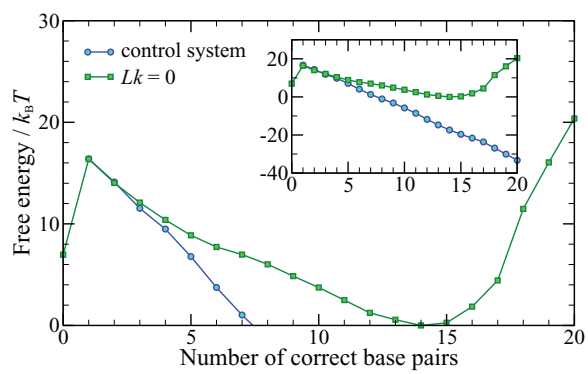

FIG. 3. Free energy profile of two complementary hairpins that are topologically unlinked (i.e., $Lk = 0$) at $T = 23\,^{\circ}$C at a single strand concentration of 0.336 mM (squares). The free energy profile for hybridization of the control system (Fig. 1(c)) is also plotted (circles). The control system can also form 20 intermolecular base pairs but without any geometric or topological constraints. In the inset, the full profiles are plotted, showing the large ($>30\,k_B T$) free energy difference between the most stable states of the kissing complex and the control system. For comparison, the free energy profiles have been set to have the same value when the number of base pairs is 1.

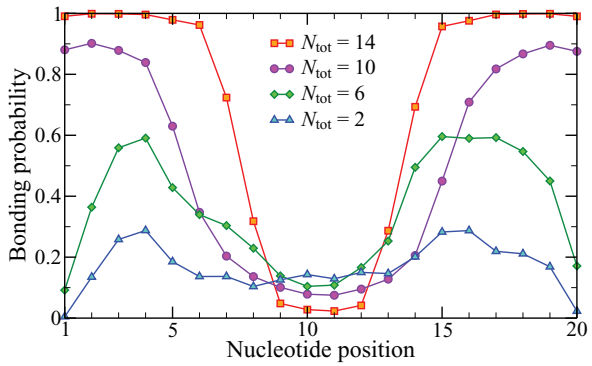

FIG. 4. Bonding probability as a function of nucleotide position in the loop for different N_{tot}, the total number of correct base pairs in the kissing complex. Note that the probability is not completely symmetric around the centre of the loop sequence because the propensity to form non-native base pairs depends on which native base pairs are formed.

linear decrease in free energy for hybridization in the absence of topological effects. It is also worth noting that this system gains 30 $k_B T$ more in free energy from hybridization than the hairpins do from forming the kissing complex. This value is roughly the amount of free energy stored in the metastable kissing complex (the most stable state for this system is the full duplex, since the two sequences are fully complementary). In DNA nanotechnology systems where such hairpins are used as fuels, this would be the amount of free energy that is potentially available to do work.

It is also interesting to look at the structure of the kissing complex as hybridization progresses. Figure 2 shows example structures with different numbers of base pairs, and in Fig. 4 the probability that a given base is bound as a function of its position along the loop is depicted for different numbers of total base pairs. For kissing complexes with a few base pairs there is not a strong thermodynamic preference for binding at a particular position in the loops, hence the distribution in Fig. 4 is roughly uniform. The exceptions are the first and last bases in the loops for which base-pairing is disfavoured, presumably because the more crowded environment that would result makes binding at these positions entropically less favourable.

By contrast, for kissing complexes with the most favourable number of base pairs ($N_{tot} = 14$), there is a very clear pattern for bonding. The six bases closest to each stem are invariably base-paired, while the four central bases have virtually no probability of binding. Therefore, the structure of the ensemble of configurations with 14 base pairs are all very similar to that in Fig. 4(b). We should also note that this structure is very different from the typical schematics of kissing complexes that tend to assume a single hybridized region, normally between the central regions of the loops.

The reason for this well-defined pattern of bonding becomes clear when we examine the structure in more detail. It is simply the best way to maximize the base pairing while satisfying the topological constraints. In particular, the base pairing adjacent to each stem continues the two helices formed by the stems (i.e., there is coaxial stacking) and there is a parallel four-way junction[51] at the coaxial stacking site associated with the exchange of strands between the two helices. Importantly, as the junction is parallel, the strand exchange leads to a crossing with positive sign that helps to counterbalance the negative crossings associated with each region of base pairing between the loops (Fig. 2(a)). This positive crossing comes with little free energy cost because it does not lead to any loss of base pairing. By contrast, the second positive crossing near to the centre of each hairpin loop is associated with a reversal of the direction of wrapping of the two chains around each other and is responsible for the inability of the two loops to hybridize further without significant free energy cost.

The base pairing probability distributions for kissing complexes with 6 and 10 base pairs in Fig. 4 illustrate how this tendency to base pair at the extremes of the loops becomes more pronounced as the number of base pairs is increased and the system becomes more topologically constrained. However, the ensembles of such structures are still much more diverse than for the fully formed kissing complex. Both the examples in Figs. 2(d) and 2(e) only have a single hybridized region, with only the latter being adjacent to one of the hairpin stems.

The positive crossings of DNA strands associated with parallel four-way junctions have previously been used to offset the negative crossings associated with hybridization in so-called "paranemic crossover" motifs.[52,53] In this motif, hybridization occurs between bubbles (a series of unpaired bases) in two duplexes leading to parallel four-way junctions at either end of the newly hybridized section (rather than at just one end as for the hairpin loops). These junctions can exactly offset half a turn in each of the helices that result, and so can lead to complete hybridization between topologically closed species. These paranemic crossover motifs have been proposed as an alternative to "sticky" single-stranded ends as a means for binding together different molecules in DNA nanotechnology[52,54] and have been used in making DNA triangles[55] and octahedra.[56] Such paranemic crossovers have also been shown to form between negatively supercoiled homologous duplexes because their zero linking number helps to alleviate the supercoiling.[57]

Interestingly, a very similar structure to that depicted in Fig. 2(b) has been identified for the inhibitory complex between the antisense RNA CopA and its target messenger RNA CopT (Refs. 58–60) and has also been suggested for other such complexes.[61] CopA and CopT have small hairpin loops that associate to form an initial kissing complex, which then progresses to form an "extended" kissing complex, where some of the base pairs in the hairpin stems are lost in favour of two intermolecular helices that coaxially stack with the rest of the hairpin stems at a parallel four-way junction like that in Fig. 2(a). This progression is dependent on the presence of bulges in the hairpin stems[60] that presumably aid the transformation by destabilizing the stems. Intriguingly, the number of base pairs in the two intermolecular helices is thought to be 15 with six bases in each of the loops connecting the ends of these helices,[58,59] which is extremely similar to the detailed structure of our kissing complexes.

Our results for the topologically unlinked system can be compared to the experimental results of Refs. 13 and 14 on the stability of these DNA kissing complexes. Reference 13 reported a high yield (nearly 100%) of kissing complexes at

a single strand concentration of 8 μM in a buffer of relatively high salt concentration, while Ref. 14 reported that kissing complexes were not stable for their 21-base loop hairpins but at a significantly lower strand concentration (0.1 μM) and at low salt. The relative probability Φ of the system being in a bound state (one or more interstrand base pairs) compared to an unbound state can be inferred from the data in Fig. 3. We find $\Phi \simeq 3100$. Assuming high dilution, this ratio can be extrapolated to a different simulation volume v' simply by dividing Φ by the ratio v'/v, where v is the original simulation volume. The relative probabilities of bound and unbound states can than be related to the bulk yields f_∞ as described in Ref. 62. Extrapolating our results to the conditions of Ref. 13 we get $\Phi = 73.7$ and a bulk yield $f_\infty = 0.89$, which is consistent with the experimental result that the kissing complexes were significantly more stable than the unbound state. By contrast, extrapolating our results to the concentration used in Ref. 14 we get $\Phi = 0.92$ and $f_\infty = 0.37$, indicating that the kissing complexes are less stable than the unbound states, which is again consistent with the experimental findings, especially when we take into account that our model is fitted to a much higher salt concentration than that used in Ref. 14 and thus is expected to give an overestimate for the bulk yield in this case since the electrostatic penalty for bringing two strands close together would be larger at the experimental salt concentration. It should also be pointed out that the sequences we study are quite asymmetrical in guanine-cytosine (GC) content, but that our model does not account for sequence-dependent effects. It is thus possible, depending on temperature and salt concentration, that the resulting bonding pattern is actually a single helix between the GC-rich regions of the loops.

B. Topologically favoured complex; $Lk = -1$

We next consider topologically linked hairpins. Although they are less experimentally relevant than the unlinked case, they nicely further illustrate the effect of topology on hybridization. First, we consider hairpins with a linking number of -1 (Fig. 1(b)(ii)). In this case the linkage has the same sense as the wrapping in duplex DNA and we thus expect hybridization of the two hairpin loops to be easier than for the unlinked case. The typical structure of the resulting kissing complex and the free energy profile for hybridization are shown in Fig. 5. There is a much lower entropic cost for initial binding as compared to the topologically unlinked system, because the two strands are already constrained to be close to each other due to the linkage. The free energy profile also exhibits a much closer to linear decrease with the number of base pairs formed than for the unlinked case, and has a minimum at 17 base pairs.

The structure of the resulting kissing complex is quite similar to that for the unlinked case in that it also has a parallel four-way junction at the point where the two stems end. The reason for this structure is again that the junction provides a positive crossing of the strands without any base pairs being lost. The typical structure assumed by the kissing complex is effectively two parallel helices and a small (2–4 base pairs) unbound region where the strands bend back on themselves. The topological sketch in Fig. 5(a) shows how the two positive crossings (at the four-way junction and at the end of the helices) and the linking number of -1 allow the system to have four negative crossings, which is topologically sufficient to form two double helical turns. Thus, that there is still a free energy cost associated with the formation of the last base pairs is due not so much to topological constraints but to geometric constraints arising because the backbones have to bend around to bridge the two helices.

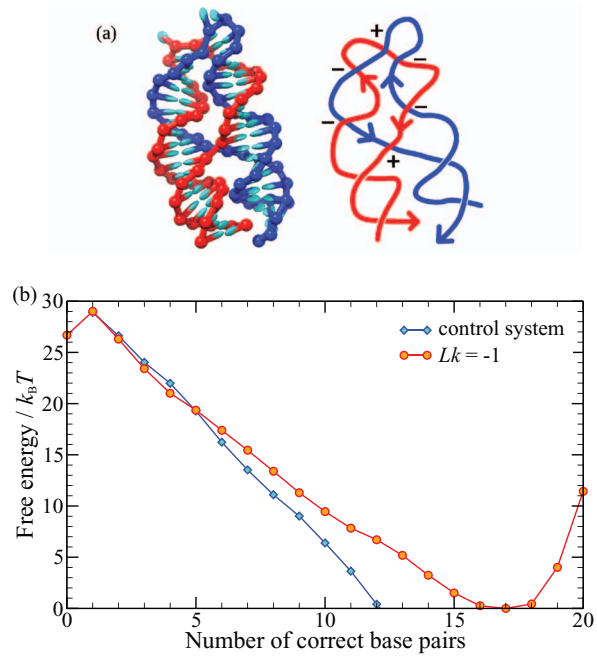

FIG. 5. (a) Typical structure and topological sketch of the kissing complex with $Lk = -1$ and (b) free energy profile associated with the formation with the kissing complex, compared to that for the control system in Fig. 1(c). To aid this comparison, the two free energy profiles were set to have the same value at 1 base pair.

One interesting feature of the structure shown in Fig. 5(a) is that the stems of both hairpins are only nine base pairs in length because the hairpin loops have displaced one base pair from each stem. This then raises the question of whether the four-way junction could migrate further and lead to the opening of both hairpins. We note that there are a number of features hindering the junction diffusion. First, junction migration is easiest when the junction adopts an "open" configuration where there is no stacking across the junction,[51,63] rather than the parallel stacked configuration typical of the kissing complex. Second, the junction migration is resisted by the topology. If the total number of base pairs is to remain constant during migration, then the number of base pairs in the duplex regions of the hybridized hairpin loops must increase. However, as the linking number must also remain constant, this also means that the unfavourable left-handed wrapping of the unhybridized sections of the loops must increase. Our simulations corroborate this picture. We observed that the position of the crossover between the end of the two stems is rather stable, although it occasionally did move one or two base pairs down.

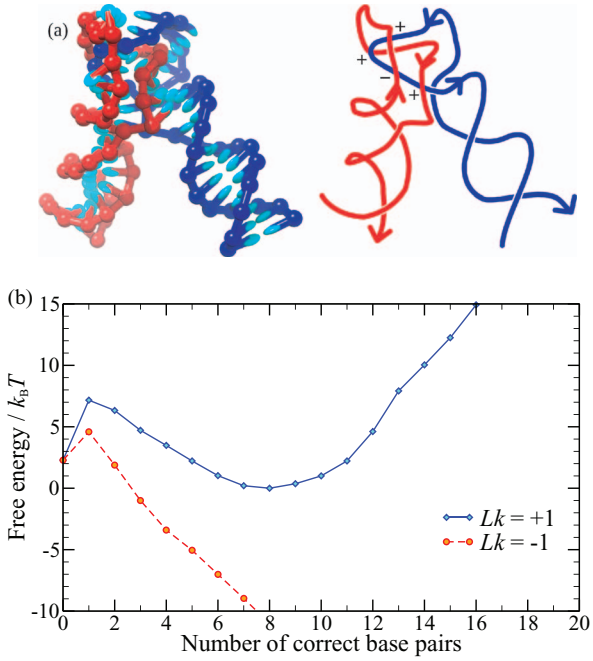

FIG. 6. (a) Typical structure and topological sketch of the kissing complex with $Lk = +1$ and (b) free energy profile associated with the formation of the kissing complex, compared to that for $Lk = -1$.

C. Topologically disfavoured complex; $Lk = +1$

Finally, we consider topologically linked hairpins with a linking number of +1 (Fig. 1(b)(iii)). In this case, the crossings associated with the linkage are of the opposite sign to that for a duplex, and so hinder base pairing between the loops. The effects of this topological frustration are clear from the free energy profile in Fig. 6. Now, the most stable kissing complex has only 8 base pairs between the hairpin loops and is only a few k_BT more stable than the unhybridized state. Indeed, it is likely that for a slightly shorter loops the topological frustration would be sufficient to totally inhibit binding.

The effects of topology are underlined by the comparison with the topologically favoured configuration with linking number $Lk = -1$, for which a further 25 k_BT drop in free energy is obtained on forming the kissing complex. Visual inspection of the structure in Fig 6(a) indicates a much more distorted structure compared to the previous cases. In this configuration, there is only a single negative crossing (roughly enough for one half turn of the double helix), but three positive crossings associated with the strands wrapping in the wrong sense around each other.

D. Role of the backbone excluded volume

Here, we consider how the results for kissing complexes depend on our parameterization of the excluded volume interaction between backbone sites. We do this first because this interaction term will play a key role in determining how easy it is for the DNA chains to wrap around each other in the wrong sense, and hence how many base pairs can be formed between hairpin loops. But, second, in our original parameterization of the model many of the properties to which we

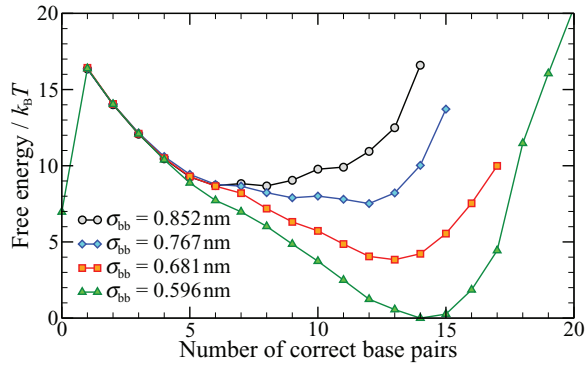

FIG. 7. Effect of the backbone excluded volume on the free energy profiles for a kissing complex with $Lk = 0$. The original value of σ_{bb} is 0.596 nm.

fitted are relatively insensitive to this interaction, and so it is possible that this parameter does not have its optimal value. For example, in the duplex state, in the high-salt concentration to which our DNA model is fitted, the backbone sites are too far away from each other for their mutual excluded volume to significantly affect duplex properties. Other properties such as the single-stranded persistence length played a greater role in its parameterization. The shape of the interaction between backbone sites, modelled as a soft repulsion, is also a significant approximation especially when two strands are close together.

We have, therefore, repeated the calculation of the free energy profile for the complex with $Lk = 0$, but where we have changed the amount of repulsion between backbones by increasing the effective radius σ_{bb} of the coarse-grained backbone site by up to 30%. As shown in Fig. 7, as the repulsion is increased, the average number of base pairs in the loop is diminished, and the free energy gain for association is significantly lowered. Of course, this change induces a large change in the yield of kissing complexes at this temperature. Since our model's predictions for yields using the original value of σ_{bb} are reasonably in line with the experimental studies reported in Refs. 13 and 14, the original parameterization appears to be physically reasonable. Moreover, that the detailed pattern of base pairing is consistent with known structures for RNA kissing complexes[22–24] further corroborates this conclusion.

We also note that at the largest value studied for the range of the repulsion that we studied the typical structure of the complex has a single intermolecular helix. Since one could regard the increase in the range of the repulsion as a very crude way to extrapolate the predictions of the model to a lower salt concentration than the one at which it was parameterized, it is possible that under those conditions binding between the hairpins' loops involves a single intermolecular helix between the GC-rich regions.

IV. CONCLUSIONS

Our simulations of the systems of kissing hairpins considered experimentally by Bois *et al.*[13] using a recently introduced coarse-grained DNA model clearly illustrates the effects of topology (due to the constraint that the linking

number must remain constant) on the free energy landscape for the formation of a kissing complex. For the unlinked case, this topological frustrations lead to 30% of the bases being unpaired, and the binding free energy of the kissing complex is significantly smaller than for a fully formed duplex with equal strand length (it is equivalent to the binding of a duplex with about 7 or 8 bases).

The free energy landscapes of the linked hairpins dramatically illustrate how manipulation of the linking number can increase or decrease this topological frustration. Compared to the unlinked system, the number of base pairs in the most stable kissing complex increases by 3 in the topologically more favourable case ($L_k = -1$), but decreases by 6 in the topologically less favourable case ($Lk = 1$). Even though the two sequences are fully complementary, the topological frustration also prevents the stems being opened by a displacement reaction involving the propagation of the intermolecular helices formed between the loops. It is this inhibition of duplex hybridization that underlies the use of hairpins as fuel for autonomous DNA nanodevices.

The structure of the kissing complex is also of particular interest, as this information is not straightforward to extract experimentally. We found that the kissing complex has a somewhat unusual structure. In particular, rather than having a single hybridized region between the loops, as might have been anticipated, the kissing complex involves two intermolecular helices that coaxially stack with the hairpin stems and involves a parallel four-way junction. This structure is favoured because there is a positive crossing of the strands at the junction that helps to offset the negative crossings associated with hybridization, but without any loss of base pairing. By contrast, the positive crossing nearer the centre of the loops is associated with unhybridized bases. For similar topological reasons, parallel four-way junctions have also been observed for paranemic motifs in which bulges in two separate duplexes cross hybridize. Furthermore, a structure very similar to that reported here has also been identified for an "extended" kissing complex between messenger and antisense RNA that also involves parallel helices and a four-way junction.[58–60]

As with any study with a coarse-grained model, one needs to consider how robust the results are and whether they might reflect any weaknesses of the model. We explicitly checked this for the repulsion between backbone sites in Section III D and found that although the results could change significantly when varying this parameter, our current value appears to be physically most reasonable. In particular, the thermodynamics in our model is consistent with the experimental stability of kissing complexes for 20-base hairpin loops in Refs. 13 and 14 (after taking into account differences in DNA and salt concentration). Furthermore, that similar structures are seen for systems with similar topological constraints suggests that our findings are physically robust.

One of the approximations in the model that we should particularly consider is the "average base" approximation, namely, that the bases in our model have identical interaction properties, except that hydrogen bonding can only occur between Watson-Crick base pairs. Although the G-C content of the hairpin loops is close to half, 7 of those G-C base pairs occur in one half of the loop. The consequences of this specific sequence might be to make the kissing complex asymmetrical with a longer intermolecular helix associated with the G-C rich half, or even to lead to the total loss of the second more weakly bound helical section. In this regard, it is interesting to note that the catalyst strand used by Bois *et al.* to open the kissing complex binds to that half of one of the hairpins that is more weakly bound in the kissing complex.[13]

Although different topological states are only strictly well defined for closed-loop molecules, as we have shown here, topological effects can be significant for linear DNA due to long-lived secondary structure that leads to the formation of internal loops. These topological constraints can inhibit hybridization and prevent the system reaching the lowest free energy state. DNA nanotechnology takes advantage of these effects when using DNA hairpins as fuel for autonomous motors, but they could also potentially be an obstacle to the successful self-assembly of DNA nanostructures. For example, if a strand hybridizes at its two ends to parts of a second long strand, an internal loop results that will be potentially restricted in its binding by topological effects, unless one of the already hybridized ends unbinds, either due to melting or displacement. Therefore, the longer the strands involved in a structure, the more likely that topological constraints will have a significant effect on the ability of the system to self assemble. This argument suggests that for DNA origamis,[64] the shortness of the staple strands (typically having two or three binding domains) probably has the effect of reducing the potential for topological effects to hinder self-assembly. Furthermore, the excess of staple strands means that a topologically constrained bound strand can be displaced by one of the equivalent staple strands from the reservoir in solution.

ACKNOWLEDGMENTS

We would like to thank the Engineering and Physical Sciences Research Council for financial support.

[1] R. Carlson, Nat. Biotechnol. **27**, 1091 (2009).
[2] N. C. Seeman, J. Theor. Biol. **99**, 237 (1982).
[3] N. C. Seeman, Nature (London) **421**, 427 (2003).
[4] N. C. Seeman, Annu. Rev. Biochem. **79**, 65 (2010).
[5] J. Bath and A. J. Turberfield, Nat. Nanotechnol. **2**, 275 (2007).
[6] G. Varani, Annu. Rev. Biophys. Biomol. Struct. **24**, 379 (1995).
[7] D. Bikard, C. Loot, Z. Baharoglu, and D. Mazel, Microbiol. Mol. Biol. Rev. **74**, 570 (2010).
[8] M. A. Glucksmann-Kuis, X. Dai, P. Markiewicz, and L. Rothman-Denes, Cell **84**, 147 (1996).
[9] M. A. Oettinger, Nature (London) **432**, 960 (2004).
[10] S. V. Santhana Mariappan, A. E. Garcia, and G. Gupta, Nucleic Acids Res. **24**, 775 (1996).
[11] S. L. Lam, F. Wu, H. Yang, and L. M. Chi, Nucleic Acids Res. **39**, 6260 (2011).
[12] R. M. Dirks and N. A. Pierce, Proc. Natl. Acad. Sci. U.S.A. **101**, 15275 (2004).
[13] J. S. Bois, S. Venkataraman, H. M. T. Choi, A. J. Spakowitz, Z.-G. Wang, and N. A. Pierce, Nucleic Acids Res. **33**, 4090 (2005).
[14] S. J. Green, D. Lubrich, and A. J. Turberfield, Biophys. J. **91**, 2966 (2006).
[15] S. Venkataraman, R. M. Dirks, P. W. K. Rothemund, E. Winfree, and N. A. Pierce, Nat. Nanotechnol. **2**, 490 (2007).
[16] S. J. Green, J. Bath, and A. J. Turberfield, Phys. Rev. Lett. **101**, 238101 (2008).
[17] P. Yin, H. M. Choi, C. R. Calvert, and N. A. Pierce, Nature (London) **451**, 318 (2008).

[18]R. A. Muscat, J. Bath, and A. J. Turberfield, Nano Lett. **11**, 982 (2011).
[19]T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, Phys. Rev. Lett. **104**, 178101 (2010).
[20]I. Tinoco, Jr., and C. Bustamante, J. Mol. Biol. **293**, 271 (1999).
[21]C. Brunel, R. Marquet, P. Romby, and C. Ehresmann, Biochimie **84**, 925 (2002).
[22]E. Bindewald, R. Hayes, Y. G. Yingling, W. Kasprzak, and B. A. Shapiro, Nucleic Acids. Res. **36**, D392 (2008).
[23]A. J. Lee and D. M. Crothers, Structure (London) **6**, 993 (1998).
[24]K. Réblová, N. Špačková, J. E. Šponer, J. Koča, and J. Šponer, Nucleic Acids Res. **31**, 6942 (2003).
[25]P. T. X. Li, C. Bustamente, and I. Tinoco, Jr., Proc. Natl. Acad. Sci. U.S.A. **103**, 15847 (2006).
[26]P. T. X. Li and I. Tinoco, Jr., J. Mol. Biol. **386**, 1343 (2009).
[27]S. Horiya, X. Li, G. Kawai, R. Saito, A. Katoh, K. Kobayashi, and K. Harada, Chem. Biol. **10**, 645 (2003).
[28]I. Severcan, C. Geary, A. Chworos, N. Voss, E. Jacovetty, and L. Jaeger, Nat. Chem. **2**, 772 (2010).
[29]W. W. Grabow, P. Zakrevsky, K. A. Afonin, A. Chworos, B. A. Shapiro, and L. Jaeger, Nano Lett. **11**, 878 (2011).
[30]F. Barbault, T. Huynh-Dinh, J. Paoletti, and G. Lancelot, J. Biomol. Struct. Dyn. **19**, 649 (2002).
[31]A. J. Turberfield, J. C. Mitchell, B. Yurke, A. P. Mills, M. I. Blakey, and F. C. Simmel, Phys. Rev. Lett. **90**, 118102 (2003).
[32]G. Seelig, B. Yurke, and E. Winfree, J. Am. Chem. Soc. **128**, 12211 (2006).
[33]B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel, and J. Neumann, Nature (London) **406**, 605 (2000).
[34]T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Chem. Phys. **134**, 085101 (2011).
[35]T. E. Ouldridge, "Coarse-grained modelling of DNA and DNA self-assembly," Ph.D. dissertation (University of Oxford, 2011), see http://tinyurl.com/7ycbx7c. The full potential is given in Chap. 2.
[36]T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "Coaxial stacking in a coarse-grained model of DNA" (unpublished).
[37]See http://dna.physics.ox.ac.uk for simulation code.
[38]C. De Michele, L. Rovigatti, T. Bellini, and F. Sciortino, "Self-assembly of short DNA duplexes: From a coarse-grained model to experiments through a theoretical link," Soft Matter (in press); arXiv:1204.0985.
[39]J. Bath, S. J. Green, and A. J. Turberfield, Angew. Chem., Int. Ed. **117**, 4432 (2005).
[40]J. Bath, S. J. Green, K. E. Allan, and A. J. Turberfield, Small **5**, 1513 (2009).
[41]D. Zhang and E. Winfree, J. Am. Chem. Soc. **131**, 17303 (2009).
[42]S. B. Smith, Y. Cui, and C. Bustamente, Nature (London) **271**, 795 (1996).
[43]T. Ramredday, R. Sachidanandam, and T. R. Strick, Nucleic Acids Res. **39**, 4275 (2011).
[44]M. Nakata, G. Zanchetta, B. D. Chapman, C. D. Jones, J. O. Cross, R. Pindak, T. Bellini, and N. A. Clark, Science **318**, 1276 (2007).
[45]S. Whitelam and P. L. Geissler, J. Chem. Phys. **127**, 154101 (2007).
[46]S. Whitelam, E. H. Feng, M. F. Hagan, and P. L. Geissler, Soft Matter **5**, 1521 (2009).
[47]G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[48]S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**, 1011 (1992).
[49]Sometimes an alternative convention is used where the strands in a DNA duplex are considered to be parallel, so that positive linking numbers result (Ref. 50).
[50]A. D. Bates and A. Maxwell, *DNA Topology* (Oxford University Press, 2005).
[51]D. M. J. Lilley, Q. Rev. Biophys. **33**, 109 (2000).
[52]X. Zhang, H. Yan, Z. Shen, and N. C. Seeman, J. Am. Chem. Soc. **124**, 12940 (2002).
[53]N. C. Seeman, Nano Lett. **1**, 22 (2001).
[54]Z. Shen, H. Yan, T. Wang, and N. C. Seeman, J. Am. Chem. Soc. **126**, 1666 (2004).
[55]W. Liu, X. Wang, T. Wang, R. Sha, and N. C. Seeman, Nano Lett. **8**, 317 (2009).
[56]W. M. Shih, J. D. Quispe, and G. F. Joyce, Nature (London) **427**, 618 (2004).
[57]X. Wang, X. Zhang, C. Mao, and N. C. Seeman, Proc. Natl. Acad. Sci. U.S.A. **107**, 12547 (2010).
[58]F. A. Kolb, C. Malmgren, E. Westhof, C. Ehresmann, B. Ehresmann, E. G. H. Wagner, and P. Romby, RNA **6**, 311 (2000).
[59]F. A. Kolb, H. M. Engdahl, J. G. Slagter-Jäger, B. Ehresmann, C. Ehresmann, E. Westhof, E. G. H. Wagner, and P. Romby, EMBO J. **19**, 5905 (2000).
[60]F. A. Kolb, E. Westhof, C. Ehresmann, B. Ehresmann, E. G. H. Wagner, and P. Romby, Nucleic Acids Res. **29**, 3145 (2001).
[61]F. A. Kolb, E. Westhof, B. Ehresmann, C. Ehresmann, E. G. H. Wagner, and P. Romby, J. Mol. Biol. **309**, 605 (2001).
[62]T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Phys.: Condens. Matter **22**, 104102 (2010).
[63]I. G. Panyutin and P. Hsieh, Proc. Natl. Acad. Sci. U.S.A. **91**, 2021 (1994).
[64]P. W. K. Rothemund, Nature (London) **440**, 297 (2006).