Topology matters: Some aspects of DNA physics

Ralf Metzler
NORDITA, Blegdamsvej 17, 2100 København Ø, Denmark

ABSTRACT

Biological cells, in some sense, are all about topology: Biomembranes separating different volumes from one another, ions or even macromolecules having to cross these membranes in controlled fashion through membrane pores; or certain proteins moving along the DNA to find their target sequence instead of searching for this site in the full 3-dimensional cell volume. Even on the single biopolymer level, topology is an essential ingredient: Intriguingly, in bacteria DNA occurs knotted, i.e., in a state topologically different from a simply connected ring. It is a key question to understand the statistical behaviour of such knotted DNA to understand a number of physiological processes having to overcome this knottedness, or to quantify results from DNA separation techniques such as electrophoresis, in which the knottedness influences the mobility. At the same time, double-stranded DNA continuously opens up floppy single-stranded bubbles, which fluctuate in size, exposing the single Watson-Crick bases to binding proteins. Again, statistical mechanical tools can be employed to examine the bubble dynamics. Here, we introduce some recent results on DNA knots and bubble fluctuations.

INTRODUCTION

Biological macromolecules split up into fairly unspecific molecules such as polysaccharides (cellulose, chitin, starch, etc.), actin filaments, or microtubules; and the highly specific biopolymers with their unique sequence of building blocks, namely nucleic acids (DNA and RNA) and polypeptides (proteins). These latter are characterised by their specific biological function, which is due to some uncommon hierarchical structure, being fully determined by the primary structure, i.e., the sequence of nucleotides or peptides formed during synthesis. This primary structure uniquely defines the local interactions within the biopolymer (α-helices and β-sheets in proteins, or stacking interactions, local twist and curvatures in nucleic acids), the secondary structure. Finally, volume interactions (tertiary structure) of chemically remote monomers or segments of the biopolymer such as pseudo-knots in RNA or looping in DNA, ensure its 3-dimensional shape, making
the biopolymer ready to function.

The relation between DNA, RNA, and proteins is governed by the central dogma of molecular biology: Starting from its complementary double-strand formed by Watson-Crick bonds and stacking of the base pairs G-C and A-T, DNA is replicated (copied) through helicase and DNA polymerase by continuous separation into the two single strands, each of which acts as a template for two new double strands, which are completed by adding new nucleotides. Similarly, from DNA by transcription through RNA polymerase, messenger RNA is created, a single strand consisting of the nucleotides G, C, A, and U; and finally, via translation by ribosomes (and others) the polypeptidic proteins with their 22 different amino acids are synthesised. This opens up the possibility to produce perfect copies in the DNA → DNA, and DNA → mRNA → protein reaction chains, thereby avoiding typical shortcomings of regular polymer chemistry such as polydispersivity, comparatively short achievable chain lengths and parasitic reactions (branching etc.).

In the following, we mainly focus on double-stranded DNA (dsDNA). DNA can be extremely long and can reach the length of several $10^9$ nucleotides, or 1 m in animal cells. Bacterial DNA is shorter, often several μm long, and occurs mostly in a more or less free ring form within the comparatively unstructured bacteria cell. This kind of DNA is therefore well suited for physical studies. DNA has a geometric diameter of $d_g \approx 2\text{nm}$, but a persistence length of $l_b \approx 53\text{nm}$. Single DNA molecules can be manipulated in controlled fashion using optical tweezers, atomic force microscopes or micro-needles, and their dynamical behaviour made visible in single molecule spectroscopy or indirectly by labelling genes with green fluorescent protein (GFP). Thus, the individual behaviour of DNA including the random fluctuations caused by their thermal embedding can be probed.

**DNA KNOTS**

Bacterial DNA incessantly undergoes topological changes through the action of specific enzymes. These can, for instance, create knots in a formerly simple ring DNA, or cut a ring DNA into little pieces, each of which forms an independent DNA circle. We are mainly interested in the working principles of antagonistic enzymes, topoisomerase (topo) II. These can detect knots or permanent entanglements, and actively, under ATP → ADP hydrolisation, remove the knottedness [1]. Similarly, topo II removes links from concatenated (interlocked) kinetoplast DNA (kDNA) such that the now single rings
can enter an electrophoresis gel (there are some $10^4$ rings of size $\sim 1$ kbp [kilo base pairs] in such a high molecular kDNA) [2].

In Fig. 1, an example for a DNA trefoil knot is shown, after some preparative steps such that the image does not necessarily give information about the real conformation, compare reference [3]. Such knots are actively removed under energy consumption from ATP, by topo II by cutting a DNA segment, passing another segment through the gap, and resealing the cut of the former, until eventually the topology of the chain is that of an unknot. It can be shown that the accuracy of topo II is significantly better than thermal cutting and pasting [1]. This is a priori surprising as topo II can only have local information whereas a knot is defined globally. A possible physical model, which could explain the high accuracy, is the kinetic proofreading mechanism, in which it is assumed that within a DNA knot the collision frequency between segments is higher [4]. How might this work for really long DNA? One might expect that if a DNA-knot is completely delocalised and forms a random coil, the proofreading would no longer work, but topo II still works on long chains in vivo. Thus, one might ask whether possibly fluctuations due to the thermal embedding might assist the DNA-knot to become tight, creating a knot region with enriched topological entanglements, which would then, in turn, again allow a local detection mechanism to work. In contrast, the mobility of DNA-knots depends on the complexity of the knot in electrophoresis experiments [3], which would point towards a decloalised knot, either a generic property of the DNA-knot, or a finite size effect caused by bending rigidity. These phenomena prompt us to investigate the statistical interplay between (bio)polymer chains and a fixed topology.
Let us start with the simplest case, a polymer ring in which a pair contact is enforce by a slip-link, a little belt buckle, which separates two subloops in the original ring, such that the two subloops can exchange length with each other but none of them can fully escape (figure-eight structure [F8]). In the Gaussian chain model, in which the biopolymer is represented by connecting the points visited by a random walker (phantom chain), it is clear that the number of configurations (NOC) is given by the scaling relation

\[ \omega \sim \mu^L (L - l)^{-d/2} l^{-d/2}, \]

where \( L \) is the entire chain length, \( l \) the length of one subloop, \( d \) is the embedding dimension, and \( \mu \) is the connectivity factor (\( \mu = 2^d \) for Gaussian chains on a hypercubic lattice). The NOC, that is, includes the entropy loss due to the returning probabilities \( l^{-d/2} \) of a random walk of length \( l = na \), \( a \) being the monomer length. It is intuitively clear that it becomes an increasingly severe constraint for such a random walk to create a loop, i.e., return to its starting point after a certain number of steps, the higher the spatial dimension is. This is mirrored in the two poles of \( \omega \) at \( l \sim 0 \) and \( l \sim L \). In fact, it can be shown that in \( d = 3 \), on average the smaller loop is only of order \( \langle l \rangle \sim a^{1/2} L^{1/2} \), and that it is completely tight \( \langle l \rangle \sim a \) in four dimensions and above [5].

For a chain with volume interactions (non-phantom chain), the random walk model is replaced by the self-avoiding walk (SAW)\(^1\). For such an SAW, it is even with a higher entropy loss that a loop can form, and the result for the probability density \( p(l) \) to find the smaller loop with length \( l \),

\[ p(l) \sim a^{-1} l^{-c} : \begin{cases} \text{2.6875,} & d = 2 \\ \text{2.24,} & d = 3 \end{cases} \]

indicates that in the presence of self-avoiding interactions the small loop will always be tight, i.e., \( \langle l \rangle \sim a \) [5]. This result was obtained within the Duplantier scaling approach [8], which allows to calculate the scaling behaviour for any polymer network with fixed vertices. The explicit, self-consistent derivation is detailed in reference [5]. We note that the scaling prediction is beautifully verified by Monte Carlo simulations of an F8 structure [5]. With a similar approach, one can investigate more complicated structures, in which multiple loops are created through slip-links, so called paraknots.

\(^1\)We note that although a linear DNA behaves like a Gaussian walk up to rather high lengths [6], at fixed topology even a Gaussian chain behaves like an SAW-chain [7].
Figure 2: Standard minimal projection of the trefoil knot $3_1$, followed by its different possible contractions, arranged according to higher scaling orders. The uncontracted trefoil geometry is found at position III of the hierarchy. At I, the F8 structure is drawn. The diagrams II-VI show the multiply connected knot region of total length $\ell = \sum_{i=1}^{N-1} s_i$ where the protruding legs indicate the outgoing large loop of length $s_N = L - \ell$. Below the individual contractions, we include the network exponents $\gamma_G$, the number $m$ of free segment lengths, and the exponents $c$ from the probability density function $p(\ell) \sim \ell^{-c}$. See reference [5] for details.

at hands of which details of multiply connected single chains can be studied. Paraknots exhibit a rich behaviour in the localisation of individual elements, depending delicately on its global structure [9]. It should be noted that with present chemistry tools, such entropy-dominated molecules could in essence be produced [10].

To obtain information about real knots, let us start with flat knots, i.e., 3-dimensional knots, that are collapsed on a surface by, e.g., gravitation or adsorption, such that they can still equilibrate in this pseudo 2-dimensional geometry. Such flat knots can indeed be realised experimentally on both molecular scale by adsorption on membranes or mica [11, 12] and macroscopically with a knotted chain on a vibrating table [13], and they offer the possibility to investigate their localisation behaviour analytically. A careful study of the scaling function reveals a hierarchy of shapes, as shown in the example of the flat trefoil knot in figure 2. The different contractions in this hierarchy all emanate from the original trefoil configuration $3_1$, which corresponds to position III in Fig.2. However, while fluctuating, there exists shapes in which one or more of the six segments of the original trefoil configuration become so small that they have to be considered non-existent on the resolution of the scaling approach, and the vertices (the crossings of...
the segments of the SAW-knot) they connect become so close that effectively they behave like molten, i.e., like one vertex of higher order. These contractions correspond to corners in the configuration space of the original trefoil knot, and they turn out to cause poles in the involved scaling function, such that the associated critical exponents change. The smallest exponent of the trefoil hierarchy in fact is that of the F8 structure considered above. Thus, looking at snap-shots of the fluctuating knot, with the highest probability one finds the F8 structure, and with the smallest probability the contraction VI. This is intuitively reasonable, as it is extremely unlikely that an SAW with an 8-vertex is created. In other words, the entire probability density $p(l)$ to find the knot region with length $l$ will scale like

$$p(l) \simeq \alpha_1 l^{-c_1} \left\{ 1 + \alpha_{II} l^{-\Delta c_{II}} + \ldots + \alpha_{VI} l^{-\Delta c_{VI}} + \ldots \right\},$$

with $\Delta c_{II} = c_{II} - c_1$ etc., such that the leading order is due to the F8 contraction, and so forth. To measure this localisation behaviour from statistical quantities is possible. For instance, the gyration radius $R_g$ (mean squared extension) of the “flat knot will, to first order, behave like $R_g$ of a regular self-avoiding ring, as the knot region is of the size of the cutoff. However, to second order, the existence of the knot region will reduce the apparent size $L$ of this ring:

$$R_g \simeq A L^{-d v} \left\{ 1 - B L^{-\Delta} + \ldots \right\}. \quad (4)$$

High-quality simulations can in fact be employed to determine this correction [14]. Moreover, it is contained to leading order in the NOC

$$\omega \simeq \mu L^{1-d v}, \quad (5)$$

in which due to the migration of the knot region all around the large central ring, the additional factor $L$ in the entropy loss factor occurs, which has in fact been measured [14].

It can be shown that the hierarchy of each flat prime knot is universally given by the F8, and that composite knots split up into their individual prime components [5]. In 3 dimensions, no such exact analytical treatment is possible, and one has to resort to phenomenological models or computer simulations, see the discussions in references [9,14–19]. It seems to be well supported that simple 3-dimensional knots in long chains localise weakly in the sense that the ratio of the mean size of the knot region to the entire chain length tends to zero. However, a recent model approach predicts that
more complex knots in fact delocalise [18], a phenomenon related to the delicate balance between the entropy gain of the big loop circled by its satellite knot region, and the internal degrees of freedom of the knot region [16, 18]. Whereas simulations for such more complex knots are outstanding, the model is consistent with the scaling exponent measured for a 3-dimensional trefoil knot [19], and corresponds to an extension of the model proposed in reference [15].

These results pertain to long dilute dsDNA, in which energetic effects (persistence length, charges) can be incorporated into an effective persistence length, i.e., a renormalised size of the Kuhn segments. However, there exist situations of geometric confinement or multivalent counterion collapse, in which the chain becomes dense. It turns out that in such a dense phase both in 2 and 3 dimensions, all knots delocalise [20]. However, simulations indicate that fluctuations under such conditions might still assist in the local knot detection, as the entanglements appear to move to the surface of the densely coiled knot [21]. These issues are part of ongoing investigations, revealing a rich behaviour of fluctuating systems with fixed topology. Similar effects have been studied in the force-extension of chains with knots or entanglements [19, 22], or in the looping structure formed in bacterial DNA [23].

DNA BUBBLE FLUCTUATIONS

The above treatment of knotted DNA pertains to long chains, typically some 10 μm long (several 100 Kuhn segments). Let us now come to a dynamical phenomenon, which occurs on the scale of a few tens of nm, the local opening of floppy ssDNA bubbles. Such bubbles, i.e., regions of the dsDNA, which open up due to thermal fluctuations already at room temperatures, lead to the exposure of unsatisfied Watson-Crick bonds, and facilitate the docking of certain ss-binding proteins. The exact dynamics of these fluctuations turn out to be physiologically important, as their time scale competes with the time scale of the binding time of the proteins [24], which was recently verified in single molecule studies [25]. At higher temperatures, eventually the two strands of the dsDNA are fully separated (denatured), a phenomenon used extensively in modern labs in the Polymerase Chain Reaction technique, which vastly multiplies an original DNA. These phenomena prompt a more detailed analysis of the bubble fluctuations. Experimentally, these have been

2In shorter chains, the behaviour will be partially masked by finite size effects.
Figure 3: Schematic melting curve of dsDNA, forming bubbles of increasing size on heating, until eventually both strands are fully denatured [27].

recently measured on the single molecular level [26].

At physiological temperatures and under standard salt conditions, bubbles are of the mean size of 20 to 30 base pairs, rising to some 200 base pairs shortly below melting. The behaviour of a dsDNA sample under temperature increase is sketched in Fig. 3, in which the portion of remaining ds segments is plotted versus $T$. While the order of this melting transition is under debate [27–31], we here address the bubble dynamics.

Corresponding to the recent measurements on short designed DNA strands [26], in which an AT sequence is clamped by GC regions\(^3\), we restrict the discussion to the simple case of denaturation of a homopolymer sequence. In the classical Poland-Scheraga model of DNA melting, we view the DNA as a sequence of double-helical regions, which due to their much higher stiffness are regarded entropy-free; and floppy ss-bubbles, which are characterised by the entropy loss discussed above. We therefore obtain the free energy

$$
\beta \mathcal{F}(n) = -\log \{ W(n) \Omega(n) \} = n \gamma(T) + \gamma_1 + c \log(n + 1),
$$

for a dsDNA with one bubble of size $n$ base pairs.\(^4\) Here, $\beta \equiv 1/(k_B T)$,

---

\(^3\)**GC bonds open under considerably higher activation than AT bonds.**

\(^4\)**The discussion of a single bubble corresponds to the experimental situation in reference [26], but it is also a reasonable approximation for long DNA below melting [34],**
$W(n) = \exp\{-n\gamma\}$ is the energy loss from the formerly satisfied Watson-Crick and stacking bonds with stacking enthalpy $\gamma = \gamma(T)$ per base pair [32, 33], and $\Omega(n) = \sigma_0 f(n)$ with $\sigma_0 = \exp\{-\gamma_1\}$ and the entropy loss $f(n) = (n + 1)^{-c} \gamma_5$ [32, 33]. Finally, $\gamma_1$ is the so-called loop initiation energy, corresponding to breaking the first stacking interaction.$^6$ As $\gamma_1$ is larger than a Watson-Crick bond energy, once the loop is initiated by a large fluctuation, a whole domain of base pairs will open up, causing the co-operative behaviour of DNA denaturation [27, 29, 30]. In Fig. 4, we plot the Poland-Scheraga free energy for a homopolymer bubble at different temperatures.

The fluctuations of the bubble size as a function of time can be described in terms of the Fokker-Planck equation for the probability density $P(n, t)$, as the fluctuations (Rouse-Zimm modes) of the bubble itself occur on a faster time scale as the zipping/unzipping. In such a reduced, one-dimensional description in the space of bubble size $n$, the free energy (6) will play the role of an effective potential, driving the bubble to closing ($T < T_{\text{melt}}$) or further unzipping ($T > T_{\text{melt}}$). To establish this Fokker-Planck equation, we combine the continuity equation

$$\frac{\partial P(n, t)}{\partial t} + \frac{\partial j(n, t)}{\partial n} = 0$$

$^5$In this expression, there appears a correction compared to the usual scaling of an asymptotically long loop.

$^6$Stacking interactions correspond to the alignment of base pairs along the dsDNA chain, and they in fact cause the high stability of DNA.
with the expression for the corresponding probability flux,

$$j(n, t) = -D \left( \frac{\partial P(n, t)}{\partial n} + \frac{P(n, t)}{k_B T} \frac{\partial \mathcal{F}}{\partial n} \right)$$

(8)

in which the gradient of the free energy enters as a drift. Combining both equations, the Fokker-Planck equation

$$\frac{\partial P(n, t)}{\partial t} = D \left( \frac{\partial}{\partial n} \left( \gamma + \frac{c}{n+1} \right) + \frac{\partial^2}{\partial n^2} \right) P(n, t)$$

(9)

is obtained [33].

To gain insight into the nature of the bubble fluctuations, we determine the characteristic time scales for bubble closing at physiological temperatures, and slightly below the melting temperature (66°C for AT stacks). This corresponds to solving the first passage time problem, starting with an initial bubble size \(n_0\), which we assume to have been opened by a co-operative fluctuation event. The results are plotted in Fig. 4. Qualitatively, the behaviour can be understood as constant drift diffusion towards an absorbing barrier \((T < T_{melt})\), or as drift-free diffusion in a box of size \(n_0\) with one absorbing boundary \((T \leq T_{melt})\), causing a linear or quadratic dependence of the characteristic closing time as a function of \(n_0\), respectively. Under the constant drift approximation, our model reduces to the fit model used in reference [26]. The more detailed analysis shows, however, that the explicit inclusion of the entropy loss shifts the values of the fitted zipping/unzipping rates. We note that the dynamics of the bubble is not significantly influenced by the different loop closure exponents \(c\), see also reference [34].

We note that in a heteropolymer situation, bubbles will preferably be localised in AT-rich zones, and therefore the dynamics of bubble diffusion along the DNA chain becomes subdiffusive [34]. We also note that at higher temperatures, the single bubble approximation may become more severe, and the coalescence of bubbles comes into play. The Fokker-Planck approach can, however, be modified to account for such phenomena. The heteropolymer character, for instance, can be included by an additional variable, the position of the zipping base pair; the subdiffusive character of bubble diffusion along the chain may be accounted for by introducing a fractional time operator [35], similar to a recent comment on biopolymer translocation [36].
SUMMARY
We have presented two case studies within current DNA biophysics, namely, DNA knots and bubble fluctuations. These phenomena on the μm and nm scale demonstrate how statistical physics in unison with single molecule experiments can contribute to the further understanding of biological systems, and their technical applications. DNA knots and their manipulation through topological enzymes in vivo, or their investigation in electrophoresis motility assays pose sophisticated new questions to the interactions between fixed topology and thermal fluctuations in macromolecules. We found a delicate dependence of knot localisation on the environmental conditions. Similarly, for the bubble fluctuations, such conditions determine the lifetime of bubbles, and therefore the competition with single-strand binding proteins. Thus, the bubbles should have a long enough lifetime, such that certain proteins can bind; at the same time, they should be fast enough to avoid binding of other proteins, e.g., repair proteins. The biophysics of biopolymers is an extraordinary example of the mutual fertilisation of physics and biology.

ACKNOWLEDGEMENT
I would like to thank Andreas Hanke for very helpful discussions.

REFERENCES
12. F. Valle and G. Dietler, personal communication.