INSTITUTE OF PHYSICS PUBLISHING NONLINEARITY

Nonlinearity **17** (2004) R1–R40 PII: S0951-7715(04)16394-9

INVITED ARTICLE

Nonlinear dynamics and statistical physics of DNA

Michel Peyrard

Laboratoire de Physique, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France

Received 19 September 2003
Published 27 January 2004
Online at stacks.iop.org/Non/17/R1 (DOI: 10.1088/0951-7715/17/2/R01)

Recommended by W Zakrzewski

Abstract

DNA is not only an essential object of study for biologists—it also raises very interesting questions for physicists. This paper discuss its nonlinear dynamics, its statistical mechanics, and one of the experiments that one can now perform at the level of a single molecule and which leads to a non-equilibrium transition at the molecular scale.

After a review of experimental facts about DNA, we introduce simple models of the molecule and show how they lead to nonlinear localization phenomena that could describe some of the experimental observations. In a second step we analyse the thermal denaturation of DNA, i.e. the separation of the two strands using standard statistical physics tools as well as an analysis based on the properties of a single nonlinear excitation of the model. The last part discusses the mechanical opening of the DNA double helix, performed in single molecule experiments. We show how transition state theory combined with the knowledge of the equilibrium statistical physics of the system can be used to analyse the results.

Mathematics Subject Classification: 70K75, 82B26, 92C05, 82C22

1. Introduction

The famous book of Schrödinger 'What is life?' [1] was one of the first attempts to use the laws of physics and chemistry to analyse the basic phenomena of life, and in particular the properties of DNA, the molecule that encodes the information that organisms need to live and reproduce themselves. Fifty years after the discovery of its double helix structure [2] DNA is still fascinating physicists, as well as the biologists, who try to unveil its remarkable properties.

It is now well established that the static structure of biological molecules is not sufficient to explain their function. This is particularly true for DNA which undergoes large conformational changes during transcription (i.e. the reading of the genetic code) or replication. This is why it is important to study the *dynamics* of the molecule. Owing to the large amplitude motions which are involved, its nonlinear aspect cannot be ignored, and this is what makes it particularly

interesting. However, pure dynamical studies are not sufficient because thermal fluctuations play a major role in DNA functioning. Therefore, statistical mechanics investigations must be included to yield useful results.

This paper, which emerged from a series of graduate lectures given in the university of Madrid, does not attempt to present a complete review of the numerous studies devoted to nonlinear dynamics of DNA, which is available in the book *Nonlinear Physics of DNA* [3]. Instead it focuses on a few nonlinear dynamical models for which not only nonlinear dynamics but also statistical mechanics has been investigated, and it tries to motivate the models and their parameters in order to allow the reader to build his/her own model and analysis.

2. DNA: structure, function and dynamics

2.1. DNA in biology

There are two major classes of biological molecules—proteins and nucleic acids. Proteins perform most of the functions in a living organism. They are active machines that catalyse some chemical reactions, transport other molecules such as oxygen (transported by haemoglobin), make controllable ion channels across membranes, self assemble to form some rigid structures in a cell such as the cytoskeleton, act as molecular motors responsible, for instance, for muscle contraction, and many other tasks.

The 'map' to build these proteins is given by the genetic code stored within the structure of DNA, deoxyribose nucleic acid. DNA makes up the genome of an organism and the human genome, for instance, is made up of 46 pieces, present in each cell, the chromosomes. Ingenious experiments have shown that each chromosome is made of a *single* DNA molecule that is 4–10 cm long [4, 5]! Therefore, the genome stored in each of our cells has a length of about 2 m. Also, humans do not have the privilege of the longest genome: the salamander has a genome of about 1 km. Of course, this very long genome has to be highly compacted to fit into a cell, into a very elaborate hierarchical structure, which is the subject of intense studies.

2.1.1. The structure of DNA. DNA is a polymer, or, more precisely, a set of two entangled polymers. Its structure is presented in detail in the books of Saenger [6] and Calladine [7] and shown in figure 1. Each of the monomers that make up these polymers is a *nucleotide* which is a compound of three elements: a phosphate group PO_4^-, a sugar ring, i.e. a five-atom cyclic group, and a base which is a complex organic group that can have one or two cycles. Figure 2 shows a schematic view of the chain of nucleotides. The backbone is formed by a sequence of phosphate groups and sugars, and it is oriented because on one side of the sugar the phosphate group is linked to a carbon atom that does not belong to the sugar ring while, on the other side, it is linked to a carbon atom which is part of the sugar ring.

The sugars and phosphates of all nucleotides are all the same but nucleotides differ by the bases which can be of four types, belonging to two categories: the purines, *adenine* (A) and *guanine* (G), have two organic cycles and the pyrimidines, *cytosine* (C) and *thymine* (T), which are smaller because they only have one organic cycle.

One important observation that led Watson and Crick to the famous discovery of the DNA double helix structure is that the bases tend to assemble in pairs through hydrogen bonds, and that pairs formed by a purine and a pyrimidine are of almost the same size, so that such pairs give a very regular structure to the two chains of nucleotides linked by the hydrogen bonds, as shown in figure 3.

To understand why these two linked nucleotide chains form a double helix, one must consider the interactions between the groups that make up the nucleotides. The base pairs are

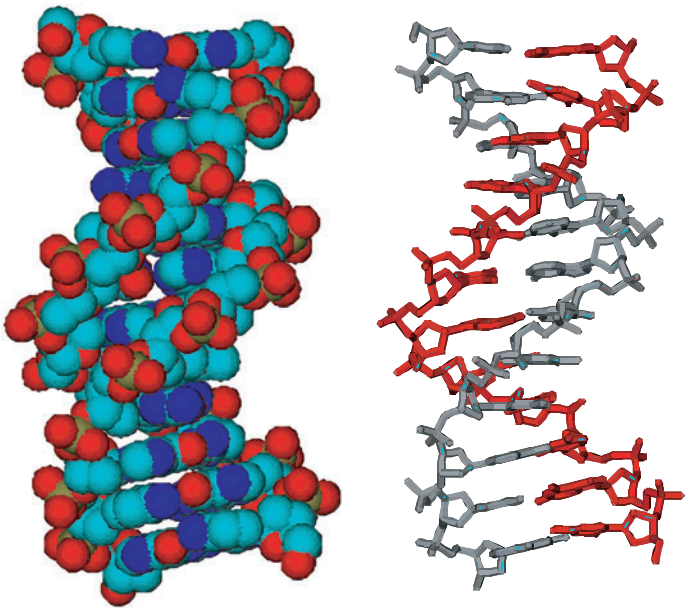

Figure 1. The double helix structure of DNA shown in a full atomic representation (left) and schematically (right).

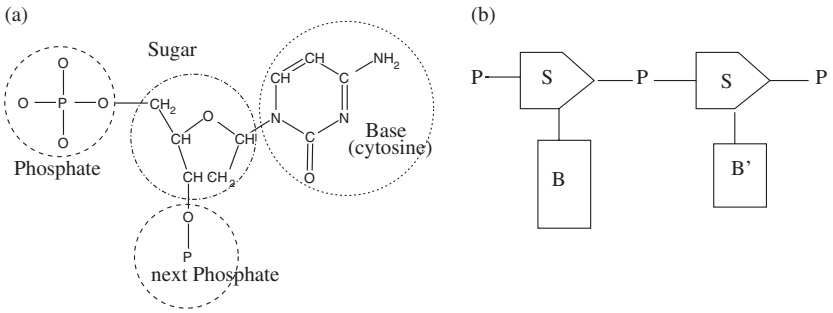

Figure 2. (*a*) An example of a nucleotide of DNA. (*b*) Schematic view of the chain of nucleotides along one DNA strand.

large flat groups having the shape of big plateaux, attached together by the sugar–phosphate strands which are fairly rigid. Many rotations around the bonds are possible within each strand but the distance between consecutive attachment points of the bases is, nevertheless, well defined. The interaction between the base plateaux is such that they tend to form a compact pile. This is due to the overlap of the π electrons of the cycles of the bases, and to hydrophobic forces: if water penetrates between the bases, the energetic cost is very high. The length of the strands is, however, too long to allow the plateaux to form a compact pile if they are simply piled up parallel to each other, as shown in figure 4. If one puts the plateaux on top of each other but rotated with respect to each other by about 30°, the strands become oblique and one can have the two plateaux in contact without compressing the length of the strands. Another possibility would be to create a skewed ladder, but a detailed examination of the structure of the bases shows that it leads to unacceptably close contacts between the atoms.

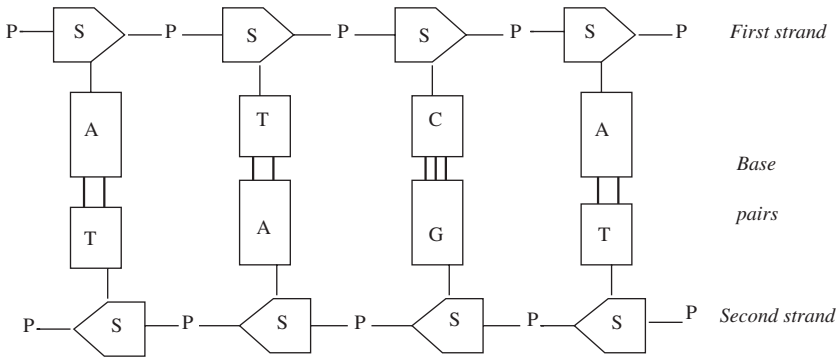

Figure 3. Schematic view of the two nucleotide chains, assembled by hydrogen bonds (thick lines) forming DNA.

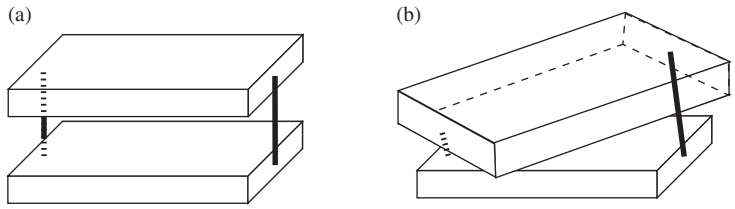

Figure 4. Schematic view of the stacking of the base plateaux DNA, taking into account the fixed length of the strands: (*a*) if the plateaux are parallel to each other, the strand is too long to allow a compact packing. (*b*) If the plateaux are rotated with respect to each other, the same strand length allows a compact packing.

Therefore, the basic topology of the groups, and their interactions, lead to the conclusion that the structure of DNA should be helicoidal [7]. However, these simple considerations are not sufficient to determine the details of the structure, and, in particular, they do not decide whether the helix should be left-handed or right-handed. In fact both situations can be found because DNA exists under different configurations, but the most common situation is a right-handed helix, which exists in two slightly different forms, known as *A* and *B* forms of DNA, depending on the degree of hydration of the molecule or the ionic strength of the solution. In cells, the typical form of DNA is the *B* form, where the base plateaux are approximately perpendicular to the axis of the helix. The basic dimensions of the molecule are illustrated in figure 5. One should note that the length of the molecule is always much greater than its diameter, which justifies models in which the molecule is described as a simple deformable string when one is not concerned with changes in its internal structure.

There are essentially two classes of forces that stabilize the double helix structure:

- Hydrogen bonding between complementary bases: only two types of base pairs are allowed by the Watson–Crick pairing: A–T which are linked by two hydrogen bonds and G–C which are linked by three hydrogen bonds and are, therefore, more robust. One should keep in mind that hydrogen bonds are weak with respect to the covalent bonds that form the strands or the cycles of the sugar and bases. A typical covalent bond, such as C–C, has a length of 1.54 Å, a bond energy of 3.6 eV, and can be lengthened by 0.1 Å by expending an energy of 0.14 eV. In contrast, a typical hydrogen bond, such as one linking two oxygens on an O–H $\cdots$ O bond, has a length of 2.75 Å, an energy of only 0.13–0.26 eV, and can be extended by 0.1 Å by expending only 0.004 eV.

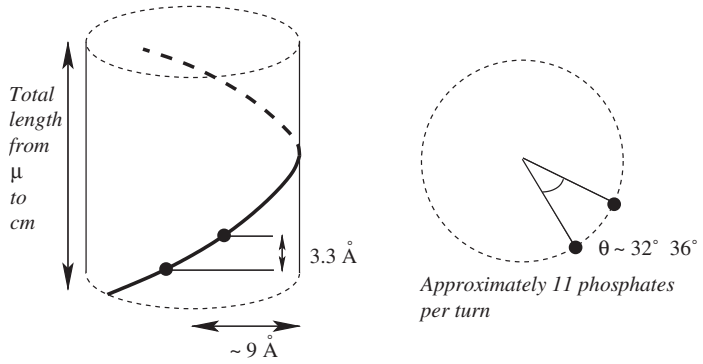

Figure 5. The basic dimensions of a DNA molecule.

- Stacking interactions of the base-pair plateaux: as mentioned earlier these interactions have a complex origin, coming partly from the overlap of the π electrons of the bases and partly from hydrophobic interactions. They impose a well-defined distance between the bases along the axis of the helix, giving rise to a high rigidity of the molecules along the axis. Bases can, however, be pulled out of the stack by sliding on each other, perpendicular to the axis. The stacking energies are highly dependent on the base sequence [6]. For instance, quantum chemistry calculations give the following energies for different stacked base pairs:

$$\begin{array}{cccc} \uparrow\!\!\begin{array}{l}\text{A–T}\\ \text{T–A}\end{array}\!\!\downarrow & \uparrow\!\!\begin{array}{l}\text{A–T}\\ \text{A–T}\end{array}\!\!\downarrow & \ldots & \uparrow\!\!\begin{array}{l}\text{C–G}\\ \text{G–C}\end{array}\!\!\downarrow \\ 0.14\,\text{eV} & 0.23\,\text{eV} & & 0.63\,\text{eV} \end{array}$$

2.1.2. Reading the genetic code. The sequence of base pairs in DNA codes the information for the synthesis of proteins, which are polymers made up of 20 different amino-acids. Because there are only four possible bases in DNA, one base cannot code for one amino-acid. The code uses three consecutive 'letters' on the DNA 'ribbon' to indicate a given amino-acid. For instance, the series CTT codes for the amino-acid leucine. With three letters per amino-acid, the code is redundant, and, for instance, CTC also codes for leucine. Moreover some specific sequences are not used to code for an amino-acid, such as the 'stop codon' TAA which marks the end of a gene coding for one particular protein.

In addition to the genes, DNA contains non-coding regions, the role of which is not yet fully understood. Some of them could be simply reminiscent of the evolution of the organisms.

Reading the genetic code is a complex operation because, as shown in figure 1, the bases, which contain the information, are buried within the double helix, and not readily available for chemical reaction. Therefore, the access to the code requires large distortions of the structure of DNA, which expose the bases outside the stack. During the *replication* of DNA, that occurs during cell division for instance, the full molecule is copied by opening it like a zipper. The *transcription* of DNA is the reading of a single gene to synthesize a protein, which occurs much more often than replication. It is performed under the control of an enzyme, the RNA polymerase, which scans the long DNA molecule to find the initiation site, i.e. the beginning of the gene. Then, it unwinds the double helix on a length of about 20 base pairs as schematized in figure 6. The bases which are exposed are copied into an RNA molecule that will be the template for the synthesis of the protein, and the 'transcription bubble' then moves along the

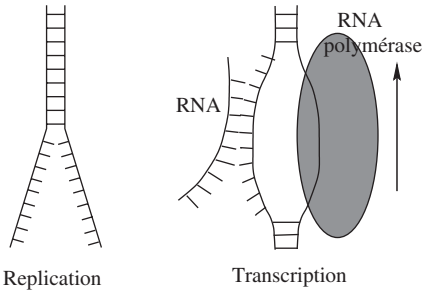

Figure 6. Schematic picture of the replication and transcription of DNA.

DNA, together with the enzyme, closing the base pairs which have been read and opening new ones, until the end of the gene is reached.

The process of transcription involves distortions of the molecule which are so large that they necessarily probe the nonlinear character of the forces linking the various groups. Due to the presence of the enzyme, which is itself a very complex molecule, it is currently beyond physical description, but, as discussed later, DNA thermal denaturation, which has some similarities with the transcription, can be described from a physical point of view. These studies can be viewed as a first step towards modelling transcription.

2.1.3. DNA is a dynamical entity. The famous discovery of the double helix structure of DNA put forward the notion that 'form is function'; but we have seen that the reading of genetic code requires large conformational changes. Such motions are observed even in the absence of enzymes. *DNA is a highly dynamical entity* and its structure is not frozen. The 'breathing' of DNA has been known to biologists for decades. It consists of the temporary opening of the base pairs. This is attested by proton–deuterium exchange experiments. DNA is put in solution in deuterated water, and one observes that the imino-protons, which are the protons forming hydrogen bonds between two bases in a base pair, are exchanged with deuterium coming from the solvent. As these protons are deeply buried in the DNA structure, the exchange indicates that bases can open, at least temporarily, to expose the imino-protons to the solvent [8]. The determination of the lifetime of a base pair, i.e. the time during which it stays closed, has been the subject of some controversy [9] because the rate limiting step in the exchange may be either the rate at which base pairs open, or the time necessary for the exchange. Accurate experiments, using NMR to detect the exchange, showed that the lifetime of a base pair is of the order of 10 ms. These experiments also show that the protons of one particular base pair can be exchanged while those of a base pair next to it are not exchanged. This indicates that the large conformational changes that lead to base pair opening in DNA are highly localized, which means that the coupling between successive bases along the DNA helix is weak enough to allow consecutive bases to move almost independently of each other.

2.2. Physical experiments on DNA

To model the nonlinear dynamics of DNA, one needs precise data to establish a meaningful model. They are provided by various physical experiments. Many data have been obtained by standard methods of condensed matter physics, but, in the last few years, powerful new methods have appeared, based on single molecule experiments.

2.2.1. Experiments based on standard methods of solid state physics. The static structure of DNA is obtained by standard x-ray or neutron diffraction. It is the earlier x-ray investigations of Rosalind Franklin, showing patterns characteristic of a helix, that put Watson and Crick on the track leading to their discovery of the double helix structure. Since then the accuracy of the determinations have been considerably improved and the structure of DNA is known to a high precision [6].

Dynamical data can be obtained by infrared or Raman vibrational spectroscopy [10, 11], or by quasi-elastic or inelastic neutron scattering, which allow a complete exploration of the dispersion curves of the small vibrational modes of DNA [12].

The main difficulty in all these experiments is to obtain oriented samples. DNA segments of a few tens of base pairs can form small crystals. Neutron diffraction needs much larger samples, which can be obtained by a rather remarkable method [12]: using natural DNA molecules a few microns long, the addition of alcohol leads to the precipitation of DNA fibres. A spinning method can be used to get a 'DNA wire' from these fibres. The 'wire' is rolled around a cylinder, leading to a sheet of parallel DNA fibres. Many of these sheets can then be mounted in a sample holder to give a sufficient volume of oriented DNA molecules. This technique is remarkable for its ability to manipulate molecules, but it also raises interesting theoretical questions. The precipitation of DNA into fibres is due to electrostatic forces, but it is not fully understood at present [13].

Another class of experiments which give useful data for modelling are the studies of DNA thermal denaturation, i.e. the separation of the two strands by heating, which is also called 'DNA melting'. It is easy to monitor the denaturation experimentally because the breaking of the base pairs is accompanied by a large increase in the absorbance of UV light at 260 nm. Experiments have been performed on artificial DNAs, which are homopolymers, i.e. have only one type of base pairs. In this case they detect a sharp transition between the double helix and the separated strands as shown in figure 7, showing that the 'melting appears as a genuine phase transition'. Natural DNA molecules show a denaturation that occurs in multiple steps and is highly sensitive to the details of the sequence [15]. The observation of DNA molecules under denaturation, using cryomicroscopy, shows that denaturation starts as local openings, called 'denaturation bubbles' that grow with temperature and invade the whole molecule at the denaturation temperature, causing the separation of the two strands, as schematically plotted in figure 8.

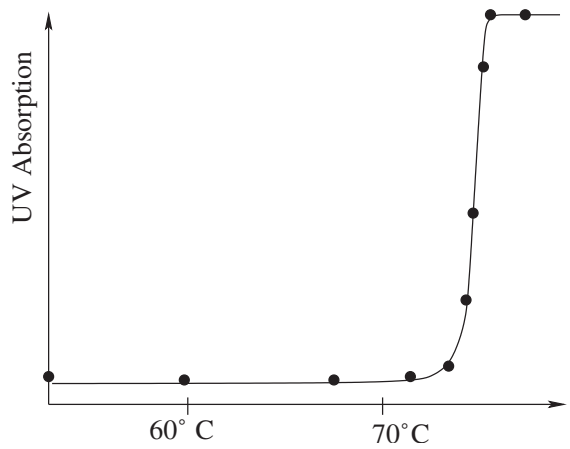

Figure 7. Denaturation curve of a homopolymer of DNA having only G–C base pairs (figure adapted from [14]).

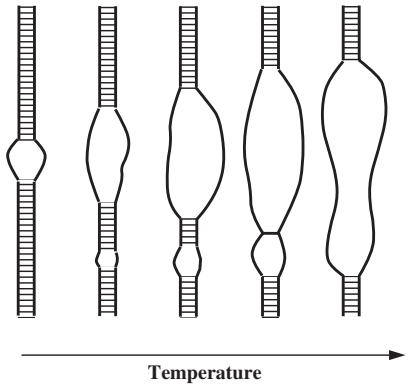

Figure 8. Schematic picture of the thermal denaturation of DNA, showing that it starts locally as 'denaturation bubbles'.

Besides thermal denaturation, it has recently been possible to study the fluctuational opening of DNA hairpins, which are single strands of DNA which have at their two ends groups of base pairs which are complementary to one another. As a result, these two groups tend to form a double helix, i.e. attach to each other, causing a folding of the strand which takes the shape of a hairpin. Such folded molecules show some temporary openings, which can be monitored very accurately. For this purpose one attaches a fluorescent molecule to one end of the strand and a quencher, i.e. a molecule that inhibits the fluorescence, at the other end. When the hairpin is formed the fluorophore and the quencher are close to each other, and the DNA molecule is not fluorescent. But when the hairpin opens, the fluorophore and the quencher are far apart and the DNA becomes fluorescent. Using confocal microscopy one can observe the fluctuations of the fluorescence of a single molecule, which give direct information about the large amplitude motions associated with the opening and closing of the DNA hairpin [16].

2.2.2. Single molecule experiments. Standard physical methods operate on a huge number of molecules and, therefore, they can only measure average properties. In the last few years a new class of investigations appeared. They are performed on a single molecule. Such experiments are possible for three reasons:

- As already mentioned, DNA is a very long molecule so that its manipulation is possible.
- Physicists can use methods developed by biologists. For instance, it is possible to use natural enzymes to perform reactions on a specific site of DNA, allowing, for instance, the chemical attachment of one particular point of the DNA molecule to a glass bead which can be manipulated.
- In line with the methods developed for scanning microscopy, the technical progress allows micromanipulations and observation of very small objects, although the experiments remain real 'tours de force'.

Figure 9 shows the result of such an experiment, the force–extension curve of a single DNA molecule, obtained for the first time in 1992 [17] and repeated with a higher accuracy by several groups a few years later [18, 19]. Such a curve is extremely rich in information about the molecule. Three different regions appear on the diagram.

- Weak forces ($f \lessapprox 5$ pN): in this domain the force essentially balances the fluctuations of the molecule and the elasticity of DNA is due to entropic effects. It can be analysed using

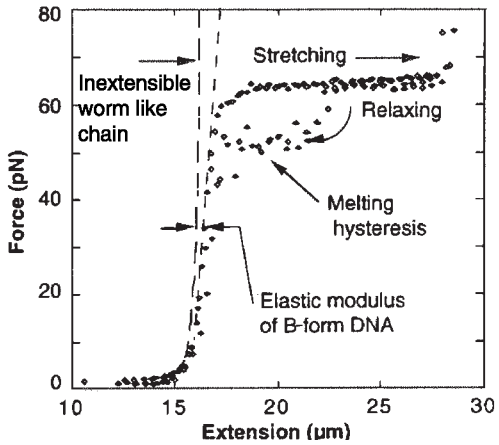

Figure 9. Force–extension curve of a single DNA molecule. Reproduced with permission from [17].

a simple model, known as the 'Worm Like Chain' (WLC) model [20, 21]. Comparison between theory and experiment gives the persistence length of the DNA molecule, or its bending rigidity.

- Intermediate forces ($5\,\mathrm{pN} < f < 60\,\mathrm{pN}$): when the force exceeds 5 pN one observes a systematic deviation from the WLC model which is due to the purely elastic contribution associated with the stretching of the double helix along its axis. This part of the curve allows the determination of the Young modulus of the molecule. Moreover, measurements in which DNA is attached to a magnetic bead which can be rotated can also provide the torsional rigidity of the molecule.
- Large forces ($f \gtrapprox 65\,\mathrm{pN}$): when the force becomes very large the molecule extends by a large amount (85–110%) almost at constant force. This is associated with a structural change induced by mechanical stress. Molecular modelling experiments have attempted to determine the structure of the extended phase [22] but this is still an open question.

Another mechanical experiment which raises interesting questions is the mechanical unzipping of DNA, which will be discussed in section 7.

3. A simple model for DNA

Our aim in this section is to examine to what extent one can understand theoretically some of the properties of DNA. Can we establish a mathematical model of the molecule which has solutions and properties in agreement with experimental observations?

The first question that one has to answer to establish such a model is to select the appropriate scale, which depends on the properties one is interested in. For instance, in order to analyse the force–extension curve of DNA, a model that ignores all the internal details of the molecule and simply describes it as a flexible string works very well.

In this paper we would like to focus our attention on properties of the molecule which are really *characteristic of DNA* and probe its *nonlinear dynamics*. The fluctuations of the molecule described by the WLC model do not meet these criteria because other polymers show similar features. Going to a much smaller scale, one could think of describing the dynamics of the molecule at the atomistic scale. However, besides the complexity of such a model which would confine the study to numerical simulations, this would not be a choice meeting our criteria

because the small amplitude motions of DNA are only weakly anharmonic, and moreover they are not fundamentally different from the motions of many other large molecules.

The specificity of DNA lies in its ability to store genetic information, and the large amplitude highly nonlinear motions that are necessary to read it. Thus, we want to establish a model which can describe these properties, and therefore, it must be *at the scale of the basic entity that encodes the information in DNA*, i.e. the base pair. One could think that a model at the atomic scale would be better since it would also describe the dynamics of groups of atoms (the bases for instance) and would be more accurate than a model at the scale of the bases. This would, however, not be practical, even if one uses only numerical simulations. We have seen that transcription involves an open region that extends over 20 base pairs. Therefore, to study it with a model, one has to include at least 100 base pairs. To study a phase transition such as thermal denaturation of DNA, the model has to extend over hundreds of base pairs because denaturation is a collective effect. Currently molecular dynamics can simulate 10 to 20 base pairs, only for timescales of a few nanoseconds, and such simulations are huge calculations. Increasing this by a few orders of magnitude is currently out of reach. Even if it were possible, such a calculation might be inappropriate because it would produce a huge amount of data, from which it would be very difficult to extract mesoscopic quantities such as the fraction of open base pairs at a given temperature, which can be compared with experiments.

This is why models at intermediate scales are useful. The fact that these models are less precise does not imply that their development is simpler. In fact, it can even be more difficult because one cannot rely on first principles to establish the models and determine their parameters. For instance, the interaction potentials between individual atoms are now rather well known. But this is not the case for the interaction potentials between complex groups such as the bases of DNA. Such potentials involve many atoms, but also solvent molecules or ions, so that they are hard to evaluate. That is why it is very important to check the validity of the models and to calibrate their parameters by comparison with well-controlled experiments.

3.1. The simplest model: the Ising model

At the scale of a base pair, the simplest model that one can think of is the Ising model. The state of a base pair is simply described by a two-state variable, equal to 0 if the base is closed and 1 if it is open.

Such models have been proposed and successfully used to study the thermal denaturation of DNA (see [23] and references therein). The drawbacks for our purpose are the following:

- They contain a large number of parameters which cannot be obtained (or even estimated) from basic physical laws. For instance, the energy to open a base pair can be evaluated, but, to describe collective effects, one also needs parameters such as the probability that a base pair opens if a neighbouring one is open, or if it is closed. They can only be obtained by fitting denaturation curves.
- As the model uses a two-state variable to describe the status of a base pair, it cannot represent the nonlinear dynamics of opening or closing, that requires the description of a continuum of intermediate states.

3.2. A simple model for nonlinear DNA dynamics

The next step is to keep only one degree of freedom per base pair, but use a real variable that describes the stretching of the bond linking the bases instead of the discrete $0/1$ variable of the Ising model. Let us denote by y_n the stretching of the nth base pair. The value $y_n = 0$ corresponds to a closed base pair as in the Ising model, but now y_n can increase continuously

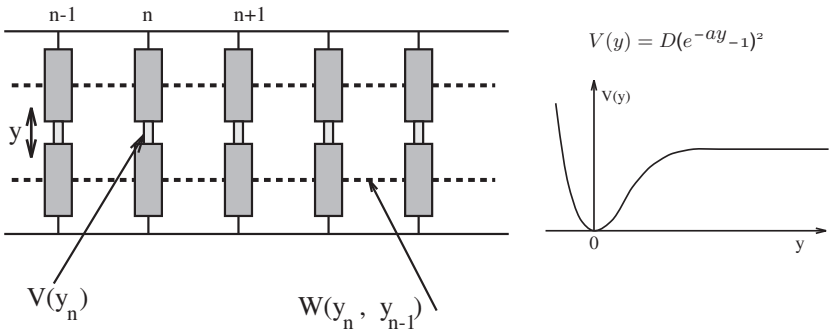

Figure 10. The simple dynamical model for DNA nonlinear dynamics, described by Hamiltonian (1).

to infinity if the two bases separate completely as in DNA denaturation. The variable y_n can even take negative values, corresponding to a compression of the bond linking the bases with respect to its equilibrium length. Large negative values will be forbidden by steric hindrance, which is introduced in the model by the potential linking the bases in a pair.

The model is shown in figure 10 and is defined by the Hamiltonian

$$H = \sum_n \frac{p_n^2}{2m} + W(y_n, y_{n-1}) + V(y_n) \qquad \text{with } p_n = m\frac{\mathrm{d}y_n}{\mathrm{d}t}, \tag{1}$$

where m is the reduced mass of the bases. At this stage we do not explicitly include the genetic code and all base pairs are considered to be the same.

The potential $V(y)$ describes the interaction between the two bases in a pair. We use a Morse potential

$$V(y) = D(\mathrm{e}^{-ay} - 1)^2, \tag{2}$$

where D is the dissociation energy of the pair and a a parameter, homogeneous to the inverse of a length, which sets the spatial scale of the potential. This expression has been chosen because it is a standard expression for chemical bonds and, moreover, it has the *appropriate qualitative shape*:

- it includes a strong repulsive part for $y < 0$, corresponding to the sterical hindrance mentioned earlier,
- it has a minimum at the equilibrium position $y = 0$,
- it becomes flat for large y, giving a force between the bases that tends to vanish, as expected when the bases are very far apart; this feature allows a complete dissociation of the base pair, which would be forbidden if we had chosen a simple harmonic potential.

The potential $W(y_n, y_{n-1})$ describes the interaction between adjacent bases along the DNA molecule. It has several physical origins:

- The presence of the sugar–phosphate strand, which is rather rigid and connects the bases. Pulling a base out of the stack in a translational motion tends to pull the neighbours due to this link. One should note, however, that we have not specified the three-dimensional motion of the bases in this simple model. An increase of the base pair stretching could also be obtained by rotating the bases out of the stack, around an axis parallel to the axis of the helix and passing through the attachment point between a base and the sugar–phosphate strand. Such a motion would not couple the bases through the strands. The potential

$W(y_n, y_{n-1})$ is an effective potential which can be viewed as averaging over the different possibilities to displace the bases.

- The direct interaction between the base pair plateaux, which is due to an overlap of the π-electron orbitals of the organic rings that make up the bases.

In the first stage, we shall use for $W(y_n, y_{n-1})$ the simplest expression, i.e. the expansion of the potential around its minimum which is reached when $y_n = y_{n-1}$

$$W(y_n, y_{n-1}) = \tfrac{1}{2}K(y_n - y_{n-1})^2. \tag{3}$$

Such an harmonic approximation would be good if the stacking interaction were strong enough to keep y_n close to y_{n-1} at all times. This is not true for DNA, but the harmonic approximation allows easier calculations, and it is sufficient to get some interesting results which agree with some experimental observations. However, we shall see that the expression for W has to be improved to provide a correct description of the thermal denaturation.

The choice of the *potential parameters* is a very difficult question because, as discussed earlier, the potentials entering the model are effective potentials, which combine many actual interactions. For instance, $V(y)$ includes the hydrogen bonds between the bases but also the repulsion between the charged phosphate groups, which is partly screened by the ions which are in solution.

The parameters that we use have been calibrated by comparison with experiments, in particular the thermal denaturation as discussed in section 6, but they are not accurately known. The parameters for $V(y)$ are $D = 0.03\,\text{eV}$, which is slightly above k_BT at room temperature (k_B being the Boltzmann constant) and $a = 4.5\,\text{Å}^{-1}$. For a stretching of the base pair distance of 0.1 Å, these parameters give a variation of energy of 0.006 eV, which is consistent with the values that we listed for hydrogen bonds in section 2. The value chosen for K is $K = 0.06\,\text{eV Å}^{-2}$, which corresponds to a weak coupling between the bases, as attested by the experimental results showing that proton–deuterium exchange can occur on one base pair without affecting the neighbours. The average mass of the nucleotides is 300 amu.

The values of the constants have been given with a system of units adapted to the scale of the problem: lengths in units of $\ell = 1\,\text{Å}$, energies in units of $e = 1\,\text{eV}$, mass in units of $m_0 = 1\,\text{amu}$. This defines a natural time unit t_0 through $e = m_0\ell^2 t_0^{-2}$, which is equal to $t_0 = 1.018 \times 10^{-14}\,\text{s}$, which is of the order of magnitude of the period of the vibrational motions of the base pairs.

Although actual units are important to compare the results with experiments, for theoretical calculations, it is very useful to express the problem in terms of dimensionless quantities. It is natural to introduce the dimensionless stretching of the base pairs as $Y = ay$. If we measure the energies in units of the depth D of the Morse potential, the dimensionless Hamiltonian is $H' = H/D$, and defining the dimensionless quantity $S = K/(Da^2)$, and a dimensionless time $\tau = \sqrt{Da^2/m}t$, we obtain the Hamiltonian H' only in terms of dimensionless quantities in the form

$$H' = \sum_n \frac{1}{2}P_n^2 + \frac{1}{2}S(Y_n - Y_{n-1})^2 + (\mathrm{e}^{-Y_n} - 1)^2 \qquad \text{with } P_n = \frac{\mathrm{d}Y_n}{\mathrm{d}\tau}, \tag{4}$$

from which one can derive dimensionless equations of motion, which depend on a single parameter S which is equal to $S = 0.0976$ with our potential parameters.

4. Observing the dynamics of the DNA model

A simple method to evaluate the ability of this model to describe DNA is to observe its dynamics and compare it with the experimental properties of DNA. Numerical simulations can be used,

but for meaningful comparisons with the actual properties of the molecule, they must take into account thermal fluctuations.

4.1. Methods for molecular dynamics at constant temperature

Simulating a thermal bath is not a simple task, but several methods have been designed to satisfactorily approximate the thermal fluctuations. A simple one is to add to the equations of motions a fluctuating force and a damping term, related by the fluctuation dissipation theorem, leading to a set of coupled Langevin equations. A more efficient way has been developed by Nosé [24] and improved by Martyna *et al* [25]. The idea is to simulate an extended system which includes not only the physical system of interest but a few additional dynamical variables corresponding to a chain of 'thermostats'. One of the thermostats is coupled to all the degrees of freedom of the physical system, according to the method proposed by Nosé that leads to exact canonical properties for the physical system, and the others contribute to properly randomize the first thermostat in order to ensure a proper exploration of the phase space of the system. This approach leads to a faster thermalization than Langevin simulations and one can control, not only that the average kinetic energy corresponds to the expected temperature, but also that the fluctuations of the kinetic energy have the value $N(k_BT)^2$ expected for a one-dimensional canonical system with N degrees of freedom.

4.2. Dynamics of the thermalized nonlinear DNA model

Figure 11 shows two typical results of the numerical simulation of the dynamics of a 256-base-pair segment of the DNA model in contact with a thermal bath.

Figure 11(a) shows a case where the model is kept at constant temperature $T = 340\,\mathrm{K}$, which is slightly below the denaturation temperature. It displays two types of characteristic patterns. The most apparent are large black spots, which correspond to regions where the base pair stretching is very large over a few tens of consecutive bases for a limited time. Such regions correspond to the 'denaturation bubbles' observed in the experiments. Another noticeable feature is the presence of vertical dotted lines, i.e. successions of black and white regions involving a few base pairs, located in one particular place on the DNA. Such patterns are created by very large amplitude localized vibrational modes of the molecule, and could correspond to the 'breathing of DNA' observed in experiments.

The simulation also points out an interesting feature of the model, which is an actual DNA property: the existence of localized large amplitude motions appears as a temporary breaking of energy equipartition in the system, although it is in equilibrium. Figure 11(a) clearly shows regions where the displacements are large separated by narrower 'cold' regions where the fluctuations of the base pairs are much smaller. Of course, energy equipartition is indeed obeyed, but to observe it one must follow the dynamics on very long timescales. On times which can be as large as thousands of periods of the small amplitude vibrations of the bases, one finds 'hot regions' coexisting with 'cold regions', but, on even larger timesales, the amplitude of the fluctuations in some hot regions will decrease while they grow in some other regions. A true uniformity in the fluctuations is never achieved but the inhomogeneities move from place to place. This phenomenon is reminiscent of turbulence, where pressure in a fluid shows large local fluctuations although an average pressure can be well defined. A quantitative analysis of the properties of a lattice of coupled nonlinear oscillators shows that the relationship with turbulence may go well beyond a simple qualitative analogy [26].

Figure 11(b) shows the results of a simulation that mimics an actual denaturation experiment: the temperature is raised in a linear manner from a low value to a value slightly

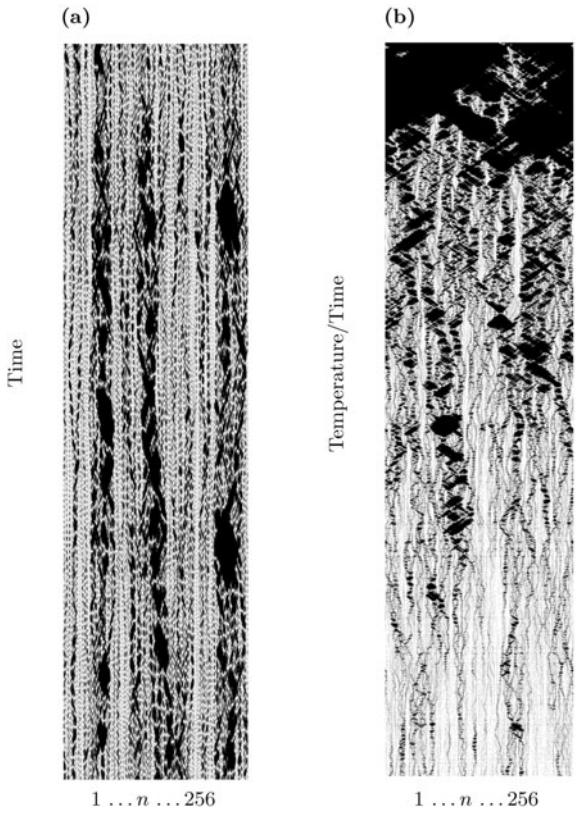

Figure 11. Numerical simulations of the dynamics of the DNA model in contact with a thermal bath simulated according to the method of Martyna *et al* [25]. The model has 256 base pairs. The amplitude of the base pair stretching is shown by a grey scale ranging from white (base pair fully closed) to black (base pair completely open). The horizontal axis extends along the molecule and the vertical axis corresponds to time (and temperature when we use a temperature ramp). (*a*) Simulation at constant temperature $T = 340\,\mathrm{K}$. (*b*) Simulation performed with a linear temperature ramp, corresponding to a denaturation experiment where the temperature is continuously raised from low T to a temperature beyond denaturation.

beyond the denaturation temperature. Such a numerical experiment can only give a crude picture of the denaturation process because, in order to keep the simulation within the times accessible to molecular dynamics, the heating has to be achieved in a few nanoseconds, which means that the system is strongly out of equilibrium. However, such a calculation gives an idea of the actual denaturation process and it shows that the model gives results which are in good agreement with the observations (see figure 8): the denaturation is preceded by the formation of large open regions that grow quickly when the denaturation temperature approaches, until they invade the whole sample, causing the separation of the strands. Moreover, the simulation shows that the localized modes described earlier appear as the precursors of the formation of the bubbles.

5. Nonlinear excitations in DNA: breathers and domain walls

Let us now try to gain some analytical understanding of the numerical observations discussed in the previous section.

5.1. Derivation of a nonlinear equation for the model

The starting point is the equations of motions of the model that derive from Hamiltonian (4)

$$\frac{\mathrm{d}^2 Y_n}{\mathrm{d}\tau^2} = S(Y_{n+1} + Y_{n-1} - 2Y_n) + 2\,\mathrm{e}^{-Y_n}(\mathrm{e}^{-Y_n} - 1) = 0. \tag{5}$$

They form a set of coupled nonlinear differential equations which cannot be solved exactly. The simulations show that, even at low temperatures, the stretching of the base pairs can become large in some regions and, therefore, a harmonic approximation has to be ruled out. We can, however, introduce a small amplitude expansion that keeps the first nonlinearities by defining

$$Y_n = \epsilon\phi_n \qquad \text{with } \epsilon \ll 1 \tag{6}$$

and keeping only the leading terms in the expansion,

$$\frac{\mathrm{d}^2\phi_n}{\mathrm{d}\tau^2} = S(\phi_{n+1} + \phi_{n-1} - 2\phi_n) - 2\left(\phi_n - \frac{3}{2}\epsilon\phi_n^2 + \frac{7}{6}\epsilon^2\phi_n^3\right) = 0. \tag{7}$$

One can look for a solution of this set of equations, which extends the simple plane wave solution that we would get in a linear approximation, under the form

$$\phi_n(\tau) = (F_n\,\mathrm{e}^{\mathrm{i}\theta_n} + F_n^\star \mathrm{e}^{-\mathrm{i}\theta_n}) + \epsilon(G_n + H_n\,\mathrm{e}^{2\mathrm{i}\theta_n} + H_n^\star\,\mathrm{e}^{-2\mathrm{i}\theta_n}), \tag{8}$$

with $\theta_n = qn - \omega t$. The choice of the additional terms will be justified more precisely later but it is easy to understand their origin: as the equations contain a term $\epsilon\phi_n^2$, the presence of the dominant term in solution (8) proportional to $\exp(\pm\mathrm{i}\theta_n)$ will naturally generate terms similar to the factor ϵ in the solution, i.e. without an exponential contribution or depending on $\exp(\pm 2\mathrm{i}\theta_n)$. Moreover, as the solution (8) appears as a modulated plane wave, which in the harmonic limit would keep a fixed amplitude, it is natural to assume that the coefficients F_n, G_n, H_n will only have a smooth spatial dependence when nonlinearity is included. These functions are assumed to depend only on 'slow' variables $X_1 = \epsilon x$, $X_2 = \epsilon^2 x$, $T_1 = \epsilon\tau$, $T_2 = \epsilon^2\tau$, and their spatial dependence is then rather well described by a continuum limit approximation, which amounts to replacing the functions at sites $n \pm 1$ by their Taylor expansion

$$F_{n\pm1} = F \pm \epsilon\frac{\partial F}{\partial X_1} \pm \epsilon^2\frac{\partial F}{\partial X_2} + \frac{\epsilon^2}{2}\frac{\partial^2 F}{\partial X_1^2}. \tag{9}$$

The time derivatives of these functions are of the form

$$\frac{\partial F_n}{\partial\tau} = \epsilon\frac{\partial F}{\partial T_1} + \epsilon^2\frac{\partial F}{\partial T_2} \tag{10}$$

with similar equations for G and H. Putting these expressions in the equations of motion, at order ϵ^0, the cancellation of terms in $\exp(\pm\mathrm{i}\theta_n)$ shows that the equations are satisfied if ω and q are linked by the dispersion relation that one gets for plane waves in the harmonic limit

$$\omega^2 = 2 + 4S\sin^2\frac{q}{2}. \tag{11}$$

Note that this dispersion relation corresponds to the *discrete* lattice, i.e. in spite of the continuum limit approximation performed for the envelope functions F, G, H, the calculation preserves some discreteness, which is important for DNA. That is why this approximation is called the *semi-discrete approximation* [27].

At order ϵ^1, the cancellation of terms in $\exp(\pm\mathrm{i}\theta_n)$ gives

$$\frac{\partial F}{\partial T_1} + v_g\frac{\partial F}{\partial X_1} = 0 \qquad \text{where } v_g = \frac{S\sin q}{\omega} \tag{12}$$

v_g is the group velocity of the waves having the dispersion relation (11). At the same order, terms without an exponential dependence and terms in $\exp(2i\theta_n)$ give

$$G = 3FF^\star \qquad H = -\frac{1}{2}\frac{F^2}{1+(8S/3)\sin^4(q/2)}. \tag{13}$$

One can note that the G and H terms in solution (8) are necessary to allow the cancellation of the terms of order ϵ^1 in the equations of motion.

The most interesting equation is obtained from the cancellation of the $\exp(i\theta_n)$ terms at order ϵ^2. In the frame mobile at speed v_g it reduces to the familiar NLS equation,

$$i\frac{\partial F}{\partial \tau_2} + P\frac{\partial^2 F}{\partial \xi_1^2} + Q|F|^2 = 0, \tag{14}$$

where the coefficients P and Q depend on the wavevector q of the carrier wave according to

$$P = \frac{S\omega^2\cos q - S^2\sin^2 q}{2\omega^3} \qquad Q = \frac{1}{2\omega}\left(11 - \frac{9}{3+8S\sin^4(q/2)}\right). \tag{15}$$

The PQ product is positive for all $q < \pi/2$ so that one can expect localized solutions when the carrier wave has a small wavevector.

5.2. Breathers and nonlinear energy localization

Using the soliton solution of NLS and computing F and G from (13), one obtains for $Y_n = \epsilon\phi_n$ the solution

$$Y_n(t) = 2a_0\,\mathrm{sech}\left[a_0\sqrt{\frac{Q}{2P}}(n - v_g t - U_e t)\right]\cos(q'n - \omega' t)$$
$$+ a_0^2\,\mathrm{sech}^2\left[a_0\sqrt{\frac{Q}{2P}}(n - v_g t - U_e t)\right]\{3 - \cos[2(q'n - \omega' t)]\}, \tag{16}$$

where a_0 and U_e are two parameters that characterize the NLS soliton and ω' and q' are corrected expressions for the frequency and wavevector of the carrier wave

$$\omega' = \omega + \frac{v_g U_e}{2P} + \frac{U_e^2}{4P} - \frac{Qa_0^2}{2}, \qquad q' = q + \frac{U_e}{2P}. \tag{17}$$

This solution is only approximate, but numerical simulations show that it is a good, stable, solution *provided ω' does not belong to the band of the allowed frequencies of the linearized lattice model*, which is the case when ω is taken near the bottom of the allowed frequencies of the lattice, i.e. $\omega = \sqrt{2}$. It corresponds to a large amplitude, localized oscillatory mode, very similar to the localized oscillations observed in figure 11(*a*). The 'breather' of NLS appears to be a good explanation for the 'breathing' of DNA and it is interesting that these names were given independently by applied mathematicians and biologists. However, to make sure that the nonlinear solution (16) of the equations of motion is meaningful for actual DNA, one must make sure that it can be spontaneously formed, and does not appear only when the appropriate initial condition is introduced in the system. This is the case for two reasons. First, extended solutions are unstable in such a system. They tend to spontaneously self modulate, leading to localized soliton-like modes [28]. These solutions may still have a moderate amplitude but, when they interact in collisions, the largest ones tend to grow at the expense of the smallest [29]. These combined effects lead to the formation of large amplitude modes, that can even emerge from the thermal fluctuations, as in the simulation of figure 11. The remarkable stability and the spontaneous formation of these breathers is due to the discreteness of the lattice [30]. But

discreteness has another role: when the amplitude of the excitations become very large, they tend to get narrower and narrower because the localizing effect of nonlinearity gets stronger. Their width can fall down to a few lattice spacings, and they strongly feel the discreteness of the lattice. They are no longer invariant by translation and they tend to be pinned to the lattice. This explains why the dotted lines on figure 11, corresponding to breathers, are vertical, i.e. correspond to an oscillatory mode which is always sitting at the same place. This extreme discreteness of DNA breathing is consistent with the experimental observations that show that imino-proton exchange can occur at one site and not at the neighbouring sites as discussed earlier [8].

5.3. Domain wall between open and closed regions

In the continuum limit, the dynamical equation resulting from Hamiltonian (4),

$$\frac{\partial^2 Y}{\partial t^2} - S\frac{\partial^2 Y}{\partial x^2} + \frac{\partial V(Y)}{\partial Y} = 0 \qquad \text{with } V(Y) = (\mathrm{e}^{-Y} - 1)^2, \tag{18}$$

has an exact static solution, which is completely different from the breather, but also has a meaning for the physics of DNA. Looking for a solution that does not depend on time, one obtains

$$-S\frac{\mathrm{d}^2 Y}{\mathrm{d}x^2} + \frac{\partial V(Y)}{\partial Y} = 0 \qquad \text{with } V(Y) = (\mathrm{e}^{-Y} - 1)^2, \tag{19}$$

which is easily solved by quadrature. After multiplication by $\mathrm{d}Y/\mathrm{d}x$ an integration gives

$$-\frac{1}{2}S\left(\frac{\mathrm{d}Y}{\mathrm{d}x}\right)^2 + V(Y) = C, \tag{20}$$

where C is a constant determined by the boundary conditions. Imposing the condition that the base pairs are in their equilibrium position $Y = 0$ for $x \to -\infty$, one gets $C = 0$. Therefore, equation (20) gives

$$\frac{\mathrm{d}Y\mathrm{e}^Y}{\mathrm{e}^Y - 1} = \sqrt{\frac{S}{2}}\,\mathrm{d}x \qquad \text{leading to } \ln(\mathrm{e}^Y - 1) = \sqrt{\frac{S}{2}}(x - x_0), \tag{21}$$

i.e.

$$Y(x) = \ln[1 + \mathrm{e}^{\sqrt{S/2}(x - x_0)}], \tag{22}$$

where x_0 is an integration constant that determines the position of the solution.

This solution describes a configuration where one part of the molecule ($x < x_0$) is closed, while for $x \gg x_0$ the base pair separation grows linearly with space and the molecule is fully denatured. It corresponds to a domain wall between two states of the DNA molecule.

Let us evaluate the energy of this solution for a finite chain of N base pairs. Sites with an index smaller than x_0 are such that $Y \simeq 0$. The Morse potential and the coupling energy between adjacent sites vanish. For sites with an index larger than x_0, $Y \gg 1$ and the Morse potential takes the value +1 while $\mathrm{d}Y/\mathrm{d}x \simeq \sqrt{2/S}$ corresponds to a coupling energy $\frac{1}{2}S(\sqrt{2/S})^2 = 1$. Therefore, each site with an index larger than x_0 contributes to the energy by $e = 2$. As a result, the energy of the domain wall is

$$E_P^+ = 2(N - x_0) + \mathcal{O}(N^0), \tag{23}$$

where the term $\mathcal{O}(N^0)$ corresponds to the core of the wall ($x \simeq x_0$) where Y evolves smoothly from the bottom of the Morse potential towards the plateau. In the limit $N \to \infty$ the energy of the domain wall becomes infinite.

For finite N, one can note that the energy of the wall gets smaller if x_0 increases, i.e. if the closed region of the molecule extends. The solution (22) is thus unstable. It tends to move in the direction that closes the base pairs. However, this result ignores the thermal fluctuations so that it is not surprising to find that the stable state of the DNA molecule is the closed state, i.e. the double helix. We shall see below that, when thermal effects are included, the analysis of the stability of the domain wall can be used to investigate the thermal denaturation of DNA.

6. Statistical physics of DNA thermal denaturation

Up to now, analytical studies have been devoted to the mechanics of DNA. Getting solutions for the nonlinear excitations of the model may be a nice result in nonlinear dynamics, but it is of little physical significance because these exact solutions never exist as such in the molecule due to the thermal fluctuations. That is why it is important to introduce temperature effects in the theoretical studies as we did in the simulations. Instead of working with the energy (or the Hamiltonian) one has to work with the free energy F which, for a system in contact with a thermal bath at temperature T, can be derived from the partition function Z. For the simple DNA models that we have introduced, it can be obtained analytically [31] using the standard methods of statistical physics.

6.1. The partition function of the simple DNA model

In the canonical ensemble the partition function of the model is related to the Hamiltonian by

$$Z = \int \prod_n \mathrm{d}p_n \, \mathrm{d}y_n \, \mathrm{e}^{-\beta H(p_n, y_n)} \tag{24}$$

with $\beta = 1/(k_B T)$. To compute the integral one can consider a finite chain of N base pairs and choose periodic boundary conditions, which amounts to adding a fictitious base pair with index 0 which has the same dynamics as base pair N. Taking into account the expression for the Hamiltonian, one has to compute

$$Z = \int\int \prod_{n=1}^{N} \mathrm{d}p_n \prod_{n=0}^{N} \mathrm{d}y_n \exp\left\{-\beta\left[\sum_{n=1}^{N} \frac{p_n^2}{2m} + \frac{1}{2}K(y_n - y_{n-1})^2 + V(y_n)\right]\right\} \delta(y_N - y_0), \tag{25}$$

where the δ function enforces the periodic boundary conditions. The integrals over p_n are decoupled Gaussian integrals, which can be easily calculated to give

$$Z = (\sqrt{2\pi m k_B T})^N Z_y \tag{26}$$

and due to the one-dimensional character of the model, the configurational part Z_y is a product of terms that couple two consecutive y_n

$$Z_y = \int \prod_{n=0}^{N} \mathrm{d}y_n \left[\prod_{n=1}^{N} \exp(-\beta f(y_n, y_{n-1}))\right] \delta(y_N - y_0) \tag{27}$$

with $f(y_n, y_{n-1}) = \frac{1}{2}K(y_n - y_{n-1})^2 + V(y_n)$. Therefore, calculation of Z_y can be performed by the transfer integral (TI) method. Let us define the TI operator $y_{n-1} \to y_n$ and its eigenfunctions ϕ_i by

$$\int \mathrm{d}y_{n-1} \exp[-\beta\{f(y_n, y_{n-1})\}] \times \phi_i(y_{n-1}) = \mathrm{e}^{-\beta\varepsilon_i}\phi_i(y_n). \tag{28}$$

The δ function of Z_y can be expanded in the basis of the eigenfunctions of the TI operator as $\delta(y_N - y_0) = \sum_i \phi_i^*(y_N)\phi_i(y_0)$. Performing successively the integrals over $y_0, y_1, y_2, \ldots, y_{N-1}$ one gets

$$\mathcal{Z}_y = \sum_i \mathrm{e}^{-\beta N \varepsilon_i} \int \mathrm{d}y_N |\phi_i(y_N)|^2 = \sum_i \mathrm{e}^{-\beta N \varepsilon_i}, \tag{29}$$

because ϕ_i is normalized. In the thermodynamic limit ($N \to \infty$) the result is dominated by the term having the smallest value of ε_i that we denote by ε_0. The free energy per site is, therefore,

$$f = -\frac{k_B T}{N} \ln \mathcal{Z} = \varepsilon_0 - \frac{k_B T}{2} \ln(2\pi m k_B T) \tag{30}$$

and similar calculations give the mean value of the base pair stretching $\langle y \rangle$ in terms of the eigenfunction ϕ_0 associated with ε_0 as

$$\sigma = \langle y \rangle = \int_{-\infty}^{+\infty} \mathrm{d}y \; y |\phi_0(y)|^2, \tag{31}$$

or the correlation function

$$C(n) = \langle (y_n - \sigma)(y_0 - \sigma) \rangle = \sum_{i=1}^{+\infty} |\langle \phi_i | y | \phi_0 \rangle|^2 \exp(-\beta(\varepsilon_i - \varepsilon_0)|n|). \tag{32}$$

The problem is, therefore, to find the eigenfunctions and eigenvalues of the transfer operator. The calculation can be made analytically in the limit of strong coupling between sites. This is a rough approximation for DNA but it is useful for a qualitative understanding of the melting transition. We want to solve

$$\int_{-\infty}^{+\infty} \mathrm{d}x \exp\{-\beta[\tfrac{1}{2}K(y-x)^2 + V(y)]\}\phi(x) = \exp(-\beta\varepsilon)\phi(y), \tag{33}$$

where we denote by x the base-pair stretching at site $n-1$ and by y the stretching at site n. For large K, when y differs significantly from x, the term $K(y-x)^2$ grows rapidly, so that the integral is dominated by the values of x which are close to y. Let us define z by $x = y + z$ and expand $\phi(y+z)$ in powers of z

$$\int_{-\infty}^{+\infty} \mathrm{d}z \; \exp\left(-\beta\left[\left(\frac{1}{2}\right) K z^2\right]\right)\left[\phi(y) + z\phi'(y) + \frac{1}{2}z^2\phi''(y)\right] = \exp(-\beta[\varepsilon - V(y)])\phi(y). \tag{34}$$

Then we perform the Gaussian integrals in z. The odd one vanishes and we get

$$\sqrt{\frac{2\pi}{\beta K}}\left[\phi(y) + \frac{1}{2\beta K}\phi''(y)\right] = \exp(-\beta[\varepsilon - V(y)])\phi(y). \tag{35}$$

The prefactor can be integrated in the exponent of the rhs to give

$$\left[\phi(y) + \frac{1}{2\beta K}\phi''(y)\right] = \exp\left\{-\beta\left[\varepsilon + \frac{1}{2\beta}\ln\left(\frac{2\pi}{\beta K}\right) - V(y)\right]\right\}\phi(y). \tag{36}$$

Let us define $\tilde{\varepsilon} = \varepsilon + (1/2\beta)\ln(2\pi/\beta K)$. For positive y (corresponding to the separation of the strands that we are investigating) the Morse potential is bounded by D. If $\beta D < 1$ one can expand the exponential to get

$$\left[\phi(y) + \frac{1}{2\beta K}\phi''(y)\right] \approx [1 + \beta V(y) - \beta\tilde{\varepsilon}]\phi(y). \tag{37}$$

Therefore, we arrive at an equation which is formally identical to the Schrödinger equation for a particle in the Morse potential,

$$-\frac{1}{2K}\frac{\mathrm{d}^2\phi(y)}{\mathrm{d}y^2} + \beta^2 V(y)\phi(y) = \beta^2\tilde{\varepsilon}\phi(y), \tag{38}$$

which has been solved by Morse himself.

One can get a qualitative understanding of the DNA denaturation transition from our basic knowledge of quantum mechanics. Such a potential well generally has a spectrum containing one or more localized states for $\beta^2\tilde{\varepsilon} < D$ and a continuous spectrum for $\beta^2\tilde{\varepsilon} > D$, but, contrary to the case of potentials bounded on both sides, the localized ground state only exists when the potential is deep enough. This leads to the following scenario for the transition:

- At low temperatures, the depth of the potential, equal to $\beta^2 D$, is large and equation (38) has at least one localized state ϕ_0. As a result $\langle y\rangle = \int y|\phi_0|^2\,\mathrm{d}y$ is finite. The DNA molecule is not denatured.
- When the temperature increases, the depth of the potential decreases and, for a critical temperature T_c the localized state merges into the continuum. The ground state is a non-localized eigenfunction of the continuum and $\langle y\rangle$ diverges. DNA is denatured.

This scenario is independent of the exact expression for the potential, provided it is qualitatively similar to the Morse potential. This is an important feature because, as we have discussed earlier, the potential is not exactly known.

For the Morse potential, the calculation shows that

$$T_c = \frac{2\sqrt{2KD}}{ak_B}. \tag{39}$$

6.2. Comparison with experiments: improving the model

As the theoretical calculation gives $\langle y\rangle$ versus T, one can also evaluate the fraction of base pairs that stay bound at a given temperature. This is a quantity that is not explicitly part of the model but one can define a threshold for the stretching above which a base pair is considered to be broken. Choosing $y_0 = 2\,\text{Å}$ is a reasonable value because hydrogen bonds are very sensitive to bond length and, for such a stretching, they are always broken. Moreover, the value of the fraction of broken base pairs is only very weakly sensitive to the value of y_0.

Figure 12 shows the temperature evolution of this fraction. It evolves very smoothly from almost 1 at a temperature equal to half the denaturation temperature to 0 at T_c. Qualitatively this is satisfactory because the simple model that we have introduced does show a denaturation transition at a given temperature T_c, but *quantitatively the simple model is not satisfactory* because for DNA the denaturation is very sharp as shown by figure 7. Choosing appropriate model parameters allows us to adjust the value of T_c to a value in quantitative agreement with the experiments, but this is not enough to get a good agreement for the full denaturation curve. The model has to be improved.

This can be done by improving the description of the stacking interaction. The harmonic expression $W(y_n, y_{n-1}) = \frac{1}{2}K(y_n - y_{n-1})^2$ resulted from an expansion of the interaction potential around its minimum. But, as previously mentioned, in DNA one base pair can be open while its neighbours are not, which implies large values of the difference $(y_n - y_{n-1})$. A small amplitude expansion is too rough.

To derive a more appropriate potential one has to take into account its physical origin. A large contribution to the stacking interaction comes from the overlap of the π electrons of the

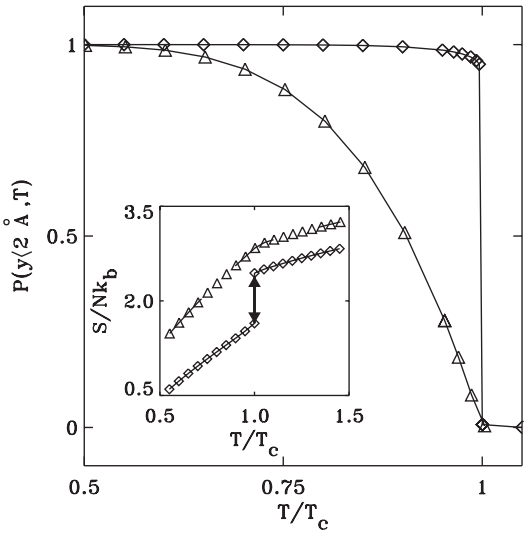

Figure 12. Theoretical result for the fraction of closed base pairs versus temperature (probability $P(y < 2\,\text{Å})$) for a model with harmonic coupling (△) or nonlinear coupling (◇). Inset: variation of the entropy versus temperature for the two models.

base plateaux. When a base moves out of the stack, the overlap decreases and its interaction with the neighbouring bases weakens. The redistribution of the π electron which occurs when hydrogen bonds are broken also contributes to this effect. This specificity of the stacking interaction can be described by an improved potential

$$W(y_n, y_{n-1}) = \tfrac{1}{2}K(1 + \rho\, \exp(-\alpha(y_n + y_{n-1})))(y_n - y_{n-1})^2. \tag{40}$$

It is important to note the plus sign in the exponential term, which is very important. As soon as either one of the two interacting base pairs is open (and not necessarily both simultaneously) the effective coupling constant drops from $K' \approx K(1 + \rho)$ down to $K' = K$.

The calculations performed with the harmonic stacking interaction in order to reduce the determination of the eigenfunctions of the transfer operator to a pseudo-Schrödinger equation are no longer valid because the coupling term is no longer a function of the difference $(y - x)$. Equation (28) has to be solved numerically. Discretizing space, the problem is turned into the diagonalization of a matrix. The accuracy of the calculation can be improved by using integration schemes such as Gauss–Hermite quadratures [32], and a finite scaling analysis [33] can be used to take care of the finiteness of the integration domain.

Figure 12 shows that the nonlinear stacking interaction drastically modifies the character of the denaturation transition of the model. It becomes very sharp, first-order like, in good agreement with experimental observations [34]. A careful study shows that the transition is not infinitely sharp (first order). It occurs smoothly but within a very narrow temperature domain, the width of which is controlled by the value of the parameter α. The entropy almost shows a jump at the transition, as shown in figure 12, which corresponds to an effective latent heat. The role of the nonlinear stacking to modify the character of the transition can be understood from simple qualitative arguments. Above denaturation, the stacking gets weaker, reducing the rigidity of the DNA strands, i.e. increasing their entropy. Although there is an energetic cost to open the base pairs, if the entropy gain is large enough, the opening may thus bring a gain in free energy $F = U - TS$, where U is the internal energy of the system and S its

entropy. The nonlinear stacking leads to a kind of 'self-amplification process' in the transition: when the transition proceeds, the extra flexibility of the strands makes it easier, which sustains the denaturation that becomes very sharp.

6.3. Another view of the transition: stability of the thermalized domain wall

We have seen that the equation of motion of the DNA model has a domain wall solution (22) which separates an open and a closed region. In the absence of thermal fluctuations the domain wall is unstable and it tends to move in the direction that closes the open region to reduce its energy. Let us now investigate the properties of the domain wall in *the presence of thermal fluctuations* and calculate its free energy.

In order to analyse the fluctuations around the domain wall, let us look for a solution of equation (18) of the form $Y(x,t) = Y_{\mathrm{DW}}(x) + f(x,t)$, where $Y_{\mathrm{DW}}(x)$ is the domain wall solution (22) centred at position x_0, and assume that $f(x,t)$ is small enough, which allows us to linearize the equation in f. Putting $Y(x,t)$ in equation (18) and linearizing in f, one gets

$$\frac{\partial^2 f}{\partial t^2} - S\frac{\partial^2 f}{\partial x^2} + \mathcal{V}(x) f(x,t) = 0, \tag{41}$$

where

$$\mathcal{V}(x) = \left(\frac{\partial^2 V[Y = Y_{\mathrm{DW}}(x)]}{\partial Y^2}\right) = -2\frac{\mathrm{e}^z - 1}{(1+\mathrm{e}^z)^2} \qquad \text{with } z = \sqrt{\frac{2}{S}}(x - x_0). \tag{42}$$

If we look for a solution of the form $f(x,t) = \mathrm{e}^{-\mathrm{i}\omega t} g(x)$ we get for $g(x)$ a Schrödinger-like equation

$$-S\frac{\partial^2 g}{\partial x^2} + \mathcal{V}(x) g(x) = \omega^2 g(x). \tag{43}$$

The eigenvalues ω^2 associated with the effective potential $\mathcal{V}(x)$ determine the spectrum of the small amplitude oscillations around the domain wall. Far from the centre of the wall, the effective potential $\mathcal{V}(x)$ tends to 0 on the open side and to 2 on the closed side of the wall. It is such that it has no bound states, but we can predict the existence of two kinds of extended states:

- For $\omega^2 < 2$ the states will only extend in the region $x > x_0$.
- For $\omega^2 > 2$ the states are extended on both sides of the wall, but the dispersion relation of the waves is different for $x < x_0$ and $x > x_0$.

The potential $\mathcal{V}(x)$ is such that equation (43) can be solved exactly [31], but to compute the free energy of the wall, it is sufficient to assume that the domain wall is sharp because the contribution of the heart of the wall becomes negligible in the thermodynamic limit. Let us consider the domain wall shown in figure 14, which is a simplified form of that in figure 13 showing the actual wall and the associated effective potential that it creates for small amplitude fluctuations.

The base pairs of index n with $n \neq x_0$ are assumed to be closed. In this domain, the dispersion relation of the small amplitude fluctuations is obtained by assuming $\mathcal{V}(x) = 2$. The fluctuations are solutions of $-S\,\mathrm{d}^2 g/\mathrm{d}x^2 + 2g = \omega^2 g$. The solutions $g(x) = \exp(\mathrm{i}qx)$, therefore, have the dispersion relation $\omega^2 = 2 + Sq^2$.

For $n \neq x_0$ the base pairs are open, i.e. Y has a value corresponding to the plateau of the Morse potential. In this range the effective potential is $\mathcal{V}(x) = 0$ and the dispersion relation of the fluctuations around the wall is $\omega = \sqrt{S}q$.

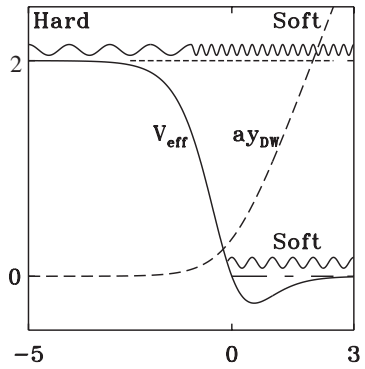

Figure 13. The domain wall of the DNA model for $x_0 = 0$ (- - - -), the effective potential $V_{\text{eff}} = \mathcal{V}(x)$ that it creates for small amplitude fluctuations, and a schematic view of the vibrational modes around the wall on both sides of the wall.

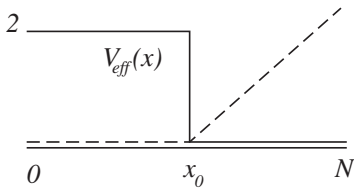

Figure 14. Simplified shape of the domain wall (- - - -) and the effective potential that it creates for small amplitude fluctuations (——), in a DNA segment of N base pairs.

The free energy of the DNA chain with a domain wall at position x_0 consists of two contributions,

$$F = F_{\text{Wall}} + F_{\text{Fluctuations}}. \tag{44}$$

The first term F_{Wall} is the free energy of the wall at a given position x_0. It is simply equal to its energy because, as the position is imposed, there is no entropy term. Each open base pair contributes for an energy $e = 2$ so that

$$F_{\text{Wall}} = 2(N - x_0). \tag{45}$$

The second term $F_{\text{Fluctuations}}$ is the free energy associated with the vibrational modes around the wall. These modes are obtained in the harmonic limit and, therefore, their energies are energies of independent harmonic oscillators of frequencies $\omega(q)$, i.e. $E(q) = n_q \hbar\omega(q)$, where n_q is the number of quanta excited for an oscillator (note that we do not introduce the zero point energy because we will be interested in the classical limit of our result). For a single mode, i.e. one of the frequencies $\omega(q)$, the partition function is the sum over all the possible states of excitation, which is easy to calculate as we have to sum a geometric series

$$Z_q = \sum_{n_q=0}^{\infty} \exp(-\beta\hbar\omega(q)n_q) = \frac{1}{1 - \exp(-\beta\hbar\omega(q))}, \tag{46}$$

which gives the usual Bose–Einstein expression. In the classical limit $\hbar\omega(q) \ll k_B T$ one can replace the exponential by its lowest order expansion $\exp(-\beta\hbar\omega(q)) \simeq 1 - \beta\hbar\omega(q)$ so that one gets

$$Z_q = \frac{1}{\beta\hbar\omega(q)}. \tag{47}$$

One can note that, for the dynamics of DNA base pairs, the classical limit is perfectly justified. Typical oscillation frequencies for the bases, which are big chemical groups, are 5×10^{11} Hz, giving $\hbar\omega \approx 2 \times 10^{-3}$ eV while $k_B T \approx 0.025$ eV at room temperature.

The partition function of the set of independent modes, which gives the number of states of the set of modes, is the product of the partition functions for each mode,

$$Z = \prod_q \frac{1}{\beta\hbar\omega(q)}, \tag{48}$$

giving the free energy of the fluctuations

$$F_{\text{Fluctuations}} = -k_B T \ln Z = k_B T \sum_q \ln(\beta\hbar\omega(q)). \tag{49}$$

This discrete sum is not easy to calculate but we can replace it with an integral as the number of modes is very large. For a lattice of p particles, the possible wavevectors are $q_i = i\pi/p$ ($i = 0, \ldots, p-1$), so that the variation of q for one mode to the next is $\Delta q = \pi/p$, which gives a density of states $1/\Delta q = p/\pi$. Using this density of states, one obtains

$$\sum_q f[\omega(q)] \simeq \frac{1}{\Delta q} \int \mathrm{d}q f[\omega(q)], \tag{50}$$

where $f[\omega(q)]$ is an arbitrary function of $\omega(q)$.

For a DNA molecule of N base pairs, having a domain wall at the base pair of index x_0, we have x_0 base pairs with the dispersion relation $\omega(q) = \sqrt{2 + Sq^2}$ and $N - x_0$ pairs with the dispersion relation $\omega(q) = \sqrt{S}q$. Therefore, one gets

$$F_{\text{Fluctuations}} = k_B T \frac{x_0}{\pi} \int_0^\infty \mathrm{d}q \ln \beta\hbar\sqrt{2 + Sq^2} + k_B T \frac{N - x_0}{\pi} \int_0^\infty \mathrm{d}q \ln \beta\hbar\sqrt{S}q, \tag{51}$$

where the integrals over q have been extended to infinity, which is consistent with the continuum limit. Setting apart the terms that depend on the position x_0 of the wall, one gets

$$F_{\text{Fluctuations}} = \frac{k_B T}{\pi} x_0 \int_0^\infty \mathrm{d}q \ln\left(\frac{\sqrt{2 + Sq^2}}{\sqrt{S}q}\right) + F_1, \tag{52}$$

where F_1 is a quantity that does not depend on x_0. The integral of equation (52), which we denote by I, can be calculated using

$$I = \int_0^\infty \mathrm{d}q \ln\sqrt{\frac{2}{Sq^2} + 1} = \frac{1}{2} \int_0^\infty \mathrm{d}q \ln 1 + \frac{2}{Sq^2} \tag{53}$$

with the help of

$$\int_0^\infty \mathrm{d}u \frac{\ln(1 + u^2)}{u^2} = \pi \tag{54}$$

and setting $u = \sqrt{2/S}q$. One gets

$$F_{\text{Fluctuations}} = k_B T \frac{x_0}{2} \sqrt{\frac{2}{S}} + F_1. \tag{55}$$

If one collects the results (45) and (55), one gets the following expression for the free energy of the DNA model with a domain wall at site x_0

$$F = x_0 \left[\frac{k_B T}{2} \sqrt{\frac{2}{S}} - 2\right] + F_1', \tag{56}$$

where F_1' *does not depend* on x_0.

- For small T (or for $T = 0$), the expression within the brackets is negative. The DNA segment can reduce its free energy by increasing x_0, which tends to close the open base pairs, as we already noted.

- In contrast, for large T the term within the brackets becomes positive and the DNA segment can reduce its free energy by decreasing x_0, i.e. by opening the base pairs that were closed.
- For $(k_B T/2)\sqrt{2/S} = 2$, i.e.

$$T = T_c = \frac{2}{k_B}\sqrt{2S} \tag{57}$$

the free energy of the DNA segment is *independent of the position of the domain wall*. Therefore, this temperature appears as the *transition temperature* at which the molecule changes from the double helix to the denaturated state.

It is important to note that, if one comes back to variables with dimensions by replacing S by its value $S = K/(Da^2)$, and taking into account that energies such as $k_B T$ must be multiplied by D to recover the true energy, one obtains $T_c = 2\sqrt{2KD}/(ak_B)$, which is *exactly the denaturation temperature given by the TI method* (equation (39)).

This calculation shows that, in addition to the standard approach to the denaturation transition, the calculation of the partition function using the TI operator, one can also study the transition by investigating the free energy of a nonlinear excitation, the domain wall. This method illustrates the interest of nonlinear science, and moreover it turns out to be *better than the conventional method* because it provides an easy method to take into account the effect of the discreteness of the lattice, which, for DNA, is not negligible since $S = 0.099$ [31].

6.4. Discussion: is a one-dimensional phase transition possible?

In statistical physics, it is often stated that there are no phase transitions in one-dimensional systems with short range interactions. We have described earlier *a simple system that does have such a phase transition*. There is no contradiction because the 'theorem' that forbids phase transitions in one-dimensional systems relies on some hypotheses:

- It is valid if the interactions are pair interactions, i.e. only depend on the difference of consecutive variables, such as $(y_n - y_{n-1})$. This is *not true* for the DNA model due to the term $V(y_n)$ coming from the interaction between the two strands of DNA. This term plays the same role as an external field on a magnetic system, for instance.
- It also requires that the domain walls between two regions have a finite energy. When this is the case, a simple argument due to Landau can show that, for instance, a spin lattice is never in an ordered state. The reason is that the cost of making a domain wall between a spin up region and a spin down one is finite and, contrary to the situation in higher dimensions, in one dimension it does not grow with the system size so that it becomes negligible in the thermodynamic limit with respect to the total interaction energy in the system. Therefore, the spin lattice can create many domains, increasing its entropy and breaking the order. *This is not true for the DNA model* because, in the thermodynamic limit, i.e. for an infinite chain, the energy of a domain wall separating an open from a closed region is infinite.

7. Modelling the mechanical denaturation of DNA

As mentioned in section 2.2.2 it is now possible to manipulate single molecules of DNA. It is, therefore, tempting to try to use this possibility to mechanically determine the sequence of DNA because a GC base pair, with three hydrogen bonds, is expected to be harder to break than an A–T base pair with only two bonds. Some experiments on the micromechanical denaturation of DNA have been performed. Figure 15 shows a schematic view of the actual experiment [35].

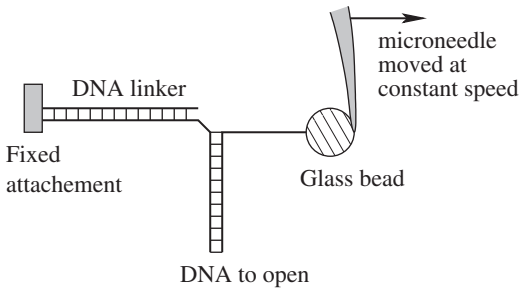

Figure 15. Schematic of the micromechanical denaturation experiment of DNA.

The DNA molecule to open is attached on one side to a DNA linker, which is attached to a glass plate providing a fixed reference point. On the other hand, one of the strands is attached to a glass bead, which is pulled by a glass micro-needle. The necessary attachments are provided by the 'biological glue' biotine–streptavidine, two molecules that bind strongly to each other. It is possible to buy glass beads coated with spectravidine, and biological techniques are used to chemically link biotine to the desired position of DNA. This experiment illustrates the power of the combination of physical techniques with biological methods that can take advantage of the specificity of biological reactions which are now well known and controlled, even if their mechanism is not yet fully understood at the molecular level. The micro-needle, moved at constant speed, pulls the glass bead, leading to the mechanical denaturation of the DNA segment under study, but, in addition, its elastic deformation which is recorded provides a measure of the force which is necessary to break the base pairs.

In fact a simple analysis shows that such an experiment is not able to distinguish a single base pair. The piece of DNA strand which is already open and the needle are equivalent to an elastic spring of rigidity k_1. When a base pair breaks, a length $\Delta\ell = 7.5\,\text{Å}$ is freed on each strand. As a result this spring shortens by $2\Delta\ell = 15\,\text{Å}$. The force that it exerts on DNA decreases, but the decrease of the elastic energy stored in the spring is very small, well below $k_B T$, and the denaturation does not stop immediately after the first mechanical denaturation has occurred. Assisted by thermal fluctuations, it goes on, along at least a few tens of base pairs, releasing the elastic stress sufficiently for the elastic force to fall well below the critical force that denaturates a base pair. Therefore, the experiment observes the breaking of groups of base pairs and cannot be used for mechanical sequencing of DNA. However, on a scale of 100 bases the experiment shows a good correlation between the G–C content and the force necessary to open the molecule [35], showing that mechanical denaturation can give some information on the sequence at low resolution. In order to analyse this information it is, however, necessary to examine the results of the experiment in detail because the relation between the measured force and the sequence is not trivial.

7.1. Numerical observation for a homopolymer

The simplest case that one can consider is the case of an artificial DNA molecule having a single type of base pair, such as the homopolymer that was used in some thermal denaturation experiments [14]. Actual single-molecule experiments on such systems have not yet been performed because it is hard to make long enough homopolymers. Let us examine the results of numerical simulations performed with the DNA model that we have introduced in section 3.

For such a study the model has to be slightly extended as shown in figure 16. As we have only one variable per base pair, the stretching of the bases, the overall translation of the

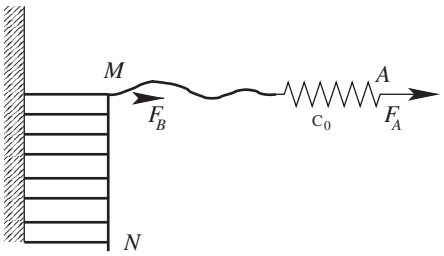

Figure 16. Extension of the one-dimensional DNA model to describe micromechanical denaturation experiments. F_A is the force pulling the glass bead, i.e. the force measured by the experiment. F_B is the force exerted by the strand which is already denaturated at the level of the base pair of index M, which is the next to break.

base pairs is not considered and, instead of the linker DNA, one can consider that one side of each base pair is attached to a fixed point. The elastic string, composed of the linker DNA (which is very rigid and plays a small role, and the micro-needle, is represented by a spring of constant c_0 which is attached to the base pair of index 1. One end of this spring (point A on the figure, whose position is denoted by y_A) is pulled at constant velocity v. The extension of the spring provides a measure of the force F_A that pulls on the DNA strand according to $F_A = c_0(y_A - y_1)$, where $y_A = vt$.

The Hamiltonian of this extended system is

$$H_M = \frac{1}{2}c_0(y_A - y_1)^2 + \frac{p_1^2}{2m} + V(y_1) + \sum_{n=2}^{N} \frac{1}{2}m\left(\frac{\mathrm{d}y_n}{\mathrm{d}t}\right)^2 + W(y_n, y_{n-1}) + V(y_n),$$

where, as before, V is the on-site Morse potential and W the stacking interaction between the bases.

Numerical simulations are performed at a controlled temperature T using a chain of Nose thermostats coupled to the variables $y_1 \cdots y_N$ but not to y_A which is constrained (and in practice attached to a macroscopic object, the glass bead and the micro-needle, itself connected to a micromanipulator). Figure 17 shows the typical results of a simulation. In the experiments the driving velocity is 20–40 nm s^{-1}, i.e. 20–40 base pairs break per second. Even with a simple model, simulations up to one second are not possible as they would require 10^{14} time steps. Simulations have been performed on timescales of the order of a ns, up to 0.5 μs for the longest. The driving speeds are, therefore, much higher than in the experiments as they are typically of the order of 2×10^{11} nm s^{-1}, down to 10^8 nm s^{-1} for the longest simulation. This is considerably larger than in experiments, but the main point is that one must pull slowly enough to be in a regime where the speed does not affect the results. This is not true for velocities of 2×10^{11} nm s^{-1} because typical forces for breaking are then approximately 64 pN, while they decrease to 16 pN for $v = 10^8$ nm s^{-1}. In this last case the forces no longer depend on v, so that one can consider that $v = 10^8$ nm s^{-1} is low enough to get meaningful results.

The average breaking force obtained in the simulation is 16 pN, while experiments give about 13 pN. *This provides an additional test for the model parameters* beyond the thermal denaturation, and confirms that they are correct because, with such a simple model, one cannot expect to reproduce experimental results with less than a 20% error.

The main result is that, although we are working with a homopolymer, *the breaking force shows very large oscillations*, associated with the variation of the speed at which the breaking propagates. We have to understand this point, and the theoretical analysis of the model will provide the clue. Moreover, simulations show that the breaking does not occur one base pair

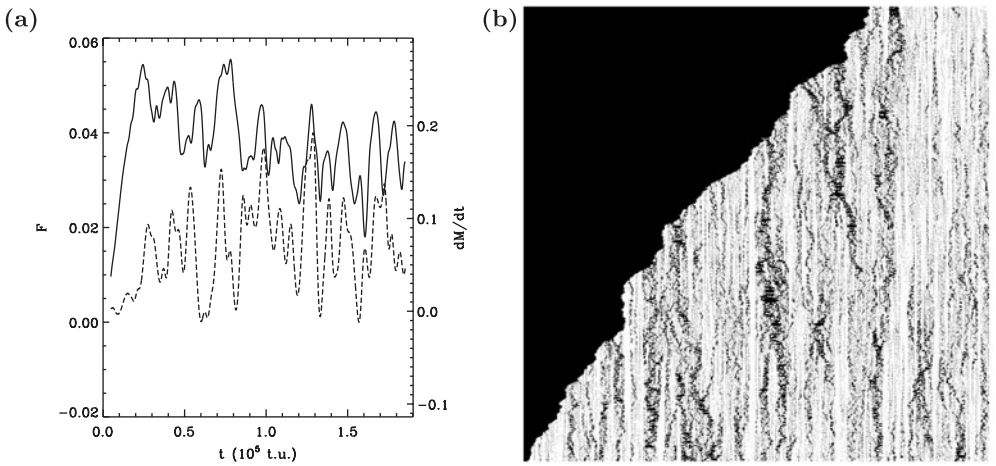

Figure 17. Results of the numerical simulations of the mechanical denaturation of a DNA segment of 512 base pairs at $T = 250\,\mathrm{K}$ and a stretching velocity $v = 0.0025\,\text{Å}\,\mathrm{tu}^{-1}$. (*a*) Variation of the force F_A (——) and of the time derivative $\mathrm{d}M/\mathrm{d}t$ of the number of broken base pairs (- - - -). The unit of force in these simulations is $1\,\mathrm{eV\,Å^{-1}} = 1.6 \times 10^{-9}\,\mathrm{N} = 1600\,\mathrm{pN}$. (*b*) Stretching of the base pairs versus time on a scale ranging from white (fully closed) to back (fully open). The horizontal axis extends along the molecule and the vertical axis is the time axis.

at a time. The propagation events involve 30–50 bases at once. This is in agreement with the experiments and provides a further test of the validity of the model.

7.2. Analysis: kinetics of the denaturation

7.2.1. Why does the molecule open? Of course it is because we are pulling on it, but is it the only reason? The maximum of the slope of the Morse potential, its slope at the inflection point, gives the value of the maximum force that the potential can oppose to the breaking force, $F_{\max} = aD/2$, which, with our model parameters, is equal to 216 pN. Therefore, the breaking is observed for forces which are *well below* the strength of the Morse potential.

Let us now consider the lifetime of a closed base pair. Simulations shows that 300 bases are broken in 2×10^5 tu, i.e. approximately 660 tu are required to break a single base pair (and much more in actual experiments). This has to be compared with the periods of oscillation at the bottom of the Morse potential $t_{\mathrm{Morse}} = 2\pi/\sqrt{2Da^2/m} = 98.7\,\mathrm{tu}$, which shows that the base pairs can perform several oscillations (very many in the actual experiments) in the bottom of the Morse potential before breaking.

In the presence of a pulling force F_B, the effective potential linking the bases becomes $V'(y) = V(y) - F_B y$. For a force that exceeds the maximum strength that the Morse potential can sustain, $V'(y)$ would lose its minimum, leading to a pure mechanical breaking. For the forces used in the experiments and in the simulations, $V'(y)$ still has a minimum separated from the dissociative state by a maximum for $y = Y_1$ (figure 18). Therefore, the stable state is the unbroken base pair, and, if breaking occurs, it is *because it is thermally activated*, i.e. it is assisted by the thermal fluctuations. Understanding the breaking amounts to understanding the escape from a potential well due to thermal fluctuations. This is a standard problem of *chemical kinetics*, similar to the dissociation of a molecule in a metastable state. It can be analysed using *transition state theory* (TST).

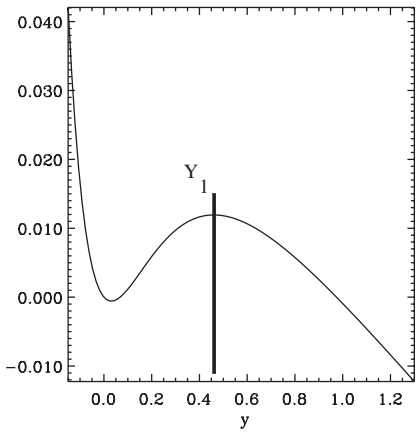

Figure 18. The effective potential acting on a base pair under the action of a pulling force.

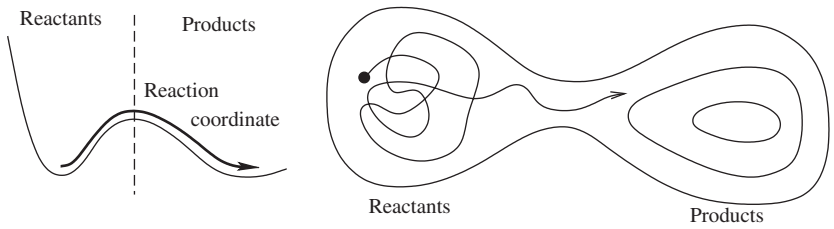

Figure 19. Schematic picture of the configuration space of the TST. Left: the reaction coordinates link reactants and products. Right: a typical trajectory between reactant and products, through a bottleneck corresponding to the transition state.

7.2.2. Introduction to TST. The theories of chemical reaction rates have looked at this problem from many viewpoints. The simplest approaches say that molecular collisions are required to induce a reaction. TST looks at the reaction pathway more completely [36]. The basic idea is that reactions may be represented by a particle motion in a potential hypersurface in configuration space (figure 19). The dimension of this configuration space is the number of degrees of freedom in the system. Among all the degrees of freedom, one variable can be considered as representative of the chemical reaction (for instance, the stretching of a bond that will dissociate). It is called the *reaction coordinate*. In this multidimensional space, the domain of *reactants* is the region where the parameters have values corresponding to the initial stage of the reaction. The domain of *products* is associated with the final state. These two regions are generally separated by a bottleneck in phase space, associated with a local maximum in energy. It is the *transition state*. The reaction is the evolution from the reactant domain to the product domain, through the transition state.

The basic assumptions of transition state theory are the following:

(i) Thermal equilibrium: all states of the system, whether they are stable or metastable states, are treated as if the system was in equilibrium. Strictly speaking, this cannot be true because the system is undergoing a chemical reaction. Particles escape from the reactant region to go into the product one. However, if the pathway between these two domains has to pass through a bottleneck, this approximation is satisfactory. This is illustrated in figure 19: the system will wander a very long time in the reactant region before

finding the way to the products. From a technical point of view, for the calculation, this assumption is equivalent to putting an infinite barrier at the transition state. One lets the system equilibrate in the reactant region, and then the barrier is removed. The system flows through the bottleneck at a rate determined by its (equilibrium) state just before the removing of the barrier.

This assumption of equilibrium seems to be wrong at a first glance since we are precisely interested in a non-equilibrium situation. However, a closer examination shows that is quite good because in this highly multidimensional space, only very few trajectories correspond to the reaction, i.e. the bottleneck appears as a tiny hole in the multidimensional boundary surface of the reactant zone.

(ii) No return assumption: the calculation assumes that, once a system has passed the barrier at the bottleneck, it does not come back to the reactant region and always evolves towards the final state of the reaction. This assumption is also related to the multidimensional aspect of the phase space and to the smallness of the bottleneck. As the products are the stable final stage of the reaction, the probability of return is even significantly smaller than the probability of passing the bottleneck from reactants to products. This assumption, which is better when the energy barrier at the bottleneck is sharp, leading the particle to quickly escape from the transition region, implies that TST *gives an upper bound for the reaction rate*.

It simplifies the analysis very much because one only needs to study the phase space trajectories in the reactant domain to determine if a particle escapes.

With these assumptions, the rate at which particles evolve from reactants to products, equal to the inverse of the lifetime τ of the reactants, is simply given by the *flow of the trajectories* in the phase space that pass through the transition state

$$k_{\mathrm{TST}} = \frac{1}{\tau} = \frac{\int \mathrm{d}p\,\mathrm{d}q\;\delta(q - q_{\mathrm{TST}})\dot{q}\theta(\dot{q})\exp(-\beta H(p,q))}{\int_{\mathrm{Reactants}} \mathrm{d}p\,\mathrm{d}q\exp(-\beta H(p,q))}, \tag{58}$$

where q denotes the coordinates in the phase space, p the corresponding momenta, q_{TST} the coordinates at the transition state, $H(p,q)$ the Hamiltonian of the system and θ the Heaviside step function, equal to 1 when its argument is positive and 0 otherwise. In this expression the denominator is simply the partition function of a system restricted to be in the reactant region only. It comes from assumption (i) above. In the numerator, the δ function means that we consider only the point of the trajectory which is exactly at the transition state. And, as $\dot{q}$ is the velocity in the configuration space, the integral weighted by the partition function of the denominator is the expression of the average *flow* across the bottleneck, the θ function ensuring that only the flow from reactants to products is considered (assumption (ii)).

7.2.3. Application of TST to the mechanical denaturation of DNA. Let us consider the system at the edge of the breaking of the Mth base pair (figure 16). We denote by F_B the force exerted on this base pair by the denaturated strand. Note that F_B is not equal to the external force F_A applied to cause the mechanical denaturation due to the fluctuations of the part of the strand which is already denaturated. The Hamiltonian of the system composed of base pairs M to N (end of the molecule) is

$$H_1 = \sum_{j=M+1}^{N} \frac{1}{2}m\left(\frac{\mathrm{d}y_j}{\mathrm{d}t}\right)^2 + W(y_j, y_{j-1}) + V(y_j) + \left[\frac{1}{2}m\left(\frac{\mathrm{d}y_M}{\mathrm{d}t}\right)^2 + V'(y_M)\right], \tag{59}$$

where $V'(y_M) = D[\exp(-\alpha y_M) - 1]^2 - F_B y_M$ is an effective potential which has a metastable minimum for $y_M \approx 0$, corresponding to the closed state of the Mth base pair (figure 18) and a barrier at $y_M = Y_1$.

In the language of chemical kinetics, the reaction coordinate is y_M and the transition state is obtained for $y_M = Y_1$. According to TST, the reaction rate, i.e. the inverse of the lifetime of the closed base pair is

$$k_{\mathrm{TST}} = \frac{1}{\tau} = \frac{\langle \delta(y_M - Y_1)\dot{y}_M \theta(\dot{y}_M)\rangle}{\langle 1 - \theta(y_M - Y_1)\rangle}. \tag{60}$$

The expression in the denominator ensures that the average is taken in the reactant domain, i.e. for $y_M \leqslant Y_1$. The brackets $\langle \ldots \rangle$ denote statistical averages, which have to be calculated with the Hamiltonian H_1.

Let us first evaluate the numerator of equation (60)

$$\mathrm{Num} = \frac{1}{Z}\int_{p_M=0}^{p_M=\infty} \mathrm{d}p_M \frac{p_M}{m} \exp(-\beta p_M^2/(2m)) \int_{-\infty}^{+\infty} \prod_{n=M+1}^{N} \exp(-\beta p_n^2/(2m)) \times N_y, \tag{61}$$

where we denote by N_y the contribution of the integrals over the coordinate positions y_n ($n = M \cdots N$), and Z is the partition function of the system. The integral over $p_M = \dot{y}_M/m$ is restricted to positive p_M due to the factor $\theta(\dot{y}_M)$ which ensures that we are only counting the flow leaving the reactant domain in the phase space.

The denominator in equation (60) is of the form

$$\mathrm{Den} = \frac{1}{Z}\int_{p_M=-\infty}^{p_M=\infty} \mathrm{d}p_M \exp(-\beta p_M^2/(2m)) \int_{-\infty}^{+\infty} \prod_{n=M+1}^{N} \exp(-\beta p_n^2/(2m)) \times D_y, \tag{62}$$

where D_y is the part depending on the integrations over the coordinates. There is no restriction on the range of p_M in this case because the denominator describes the trajectories in the reactant domain and all fluctuations are allowed.

In equations (61) and (62) the integrals over p_n for $n = M + 1 \cdots N$ cancel in the numerator and denominator and the integrals over p_M are Gaussian integrals which can be easily calculated to give

$$k_{\mathrm{TST}} = k_B T \frac{N_y}{D_y}. \tag{63}$$

The integrals over the y can be evaluated by the TI method, similar to the calculation of the partition function in section 6.1.

Let us define the eigenstates of the transfer operator by

$$\int \mathrm{d}y_n \exp[-\beta(W(y_n, y_{n-1}) + V(y_n))]\varphi_\ell(y_n) = \exp[-\beta\epsilon_\ell]\varphi_\ell(y_{n-1}). \tag{64}$$

Note that it is convenient to define them so that they decrease the index n by one unit. With definition (64) the kernel is not symmetric. For numerical calculations it might be convenient to symmetrize it. The eigenvalues are not changed but the eigenfunctions are modified by a factor $\exp[-\beta V(y)/2]$.

The expression for N_y is

$$\begin{aligned} N_y = \int_{y_M=-\infty}^{+\infty} \mathrm{d}y_M \delta(y_M - Y_1) \exp[-\beta V'(y_M)] \int \mathrm{d}y_{M-1} \exp[-\beta(W(y_M, y_{M+1}) + V(y_{M+1}))] \\ \times \int \mathrm{d}y_{M+2} \ldots \times \int \mathrm{d}y_{N-1} \exp[-\beta(W(y_{N-1}, y_{N-2}) + V(y_{N-1}))] \\ \times \int \mathrm{d}y_N \exp[-\beta(W(y_N, y_{N-1}) + V(y_N))]\delta(y_N). \end{aligned}$$

The last factor, $\delta(y_N)$, assumes that the last base pair of the molecule is fixed. This is convenient for the calculation but irrelevant for the final result on the lifetime of the base pair M because we shall take the thermodynamic limit $N \to \infty$.

The completeness relation for the eigenfunctions of the transfer operator gives

$$\delta(y_N - 0) = \sum_\ell \varphi_\ell^*(0)\varphi_\ell(y_N). \tag{65}$$

Putting this relation into N_y and performing successively the integrals over $y_N, y_{n-1}, \ldots, y_{M+1}$, we get

$$N_y = \sum_\ell \varphi_\ell^*(0) \int \mathrm{d}y_M \delta(y_M - Y_1) \exp[-\beta V'(y_M)]\varphi_\ell(y_M) \exp(-\beta\epsilon_\ell(N-M)). \tag{66}$$

Performing the integral over y_M, which includes a δ function, simply gives

$$N_y = \sum_\ell \varphi_\ell^*(0)\varphi_\ell(Y_1) \exp[-\beta V'(Y_1)] \exp(-\beta\epsilon_\ell(N-M)). \tag{67}$$

In the thermodynamic limit $N \to \infty$, the sum is dominated by the term having the lowest ϵ_ℓ, which we denote by ϵ_0

$$N_y = \varphi_0^*(0)\varphi_0(Y_1) \exp[-\beta V'(Y_1)] \exp(-\beta\epsilon_0(N-M)). \tag{68}$$

The calculation of D_y proceeds along the same lines. Most of the integrals are the same as the ones that appear in N_y. The only difference concerns the integral over y_M that no longer includes a δ function and must be performed for $y_M \leqslant Y_1$ (in the reactant region). One gets

$$D_y = \varphi_0^*(0) \int_{-\infty}^{Y_1} \mathrm{d}y_M \exp[-\beta V'(y_M)]\varphi_0(y_M) \exp(-\beta\epsilon_0(N-M)). \tag{69}$$

Combining equations (63), (67) and (69) one gets

$$k_{\mathrm{TST}} = \frac{1}{\tau} = k_B T \frac{\varphi_0(Y_1) \exp[-\beta V'(Y_1)]}{\int_{-\infty}^{Y_1} \mathrm{d}y_M \exp[-\beta V'(y_M)]\varphi_0(y_M)}. \tag{70}$$

The result shows that the reaction rate k_{TST} contains the Arrhenius term $\exp[-\beta V'(Y_1)]$ generally found for chemical reactions because $V'(Y_1)$ is the amplitude of the barrier that has to be overcome to break base pair M (figure 18). But *there is another important factor*, $\varphi_0(Y_1)$. As discussed in the analysis of the thermal denaturation, the function $\varphi_0(y)$ gives the weighting factor for the calculation of $\langle y \rangle = \int y|\varphi_0(y)|^2 \,\mathrm{d}y$. At low temperatures it is strongly peaked around the minimum of the Morse potential, and decays exponentially away from it (figure 20). Therefore, it is very small for $y = Y_1$ (unless the force is very close to the maximum force that the Morse potential can sustain so that Y_1 gets closer to 0; we have seen that this is not the case in the experiments.) In contrast, when T increases, and particularly when it approaches the denaturation temperature T_c, $\varphi_0(y)$ gets much broader, so that $\varphi_0(Y_1)$ increases drastically. Therefore, $\varphi_0(Y_1)$ appears as an *amplification factor* which increases the effect of temperature on the probability of opening. As a result temperature inhomogeneities, due to nonlinear localization, play a very large role in determining the lifetime of the base pairs under stretch. This explains why the propagation of the mechanical breaking is so irregular, even for a homopolymer.

Another indication of the significant effect of temperature inhomogeneities in the DNA model is the strong variation of $\bar{y}(x)$ with space, where

$$\bar{y}(x) = \int_0^{t_0} y(x,t)\,\mathrm{d}t, \tag{71}$$

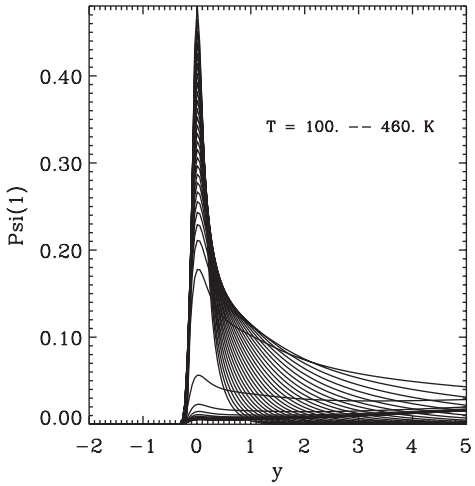

Figure 20. Evolution of the eigenfunction φ_0 of the transfer operator versus temperature in the range $T = 100$–460 K. The function has a peaked shape at low temperatures and broadens at high temperatures.

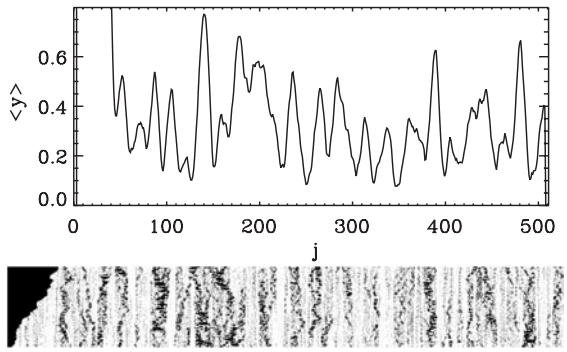

Figure 21. Variation of $\bar{y}(x)$ in the DNA model at $T = 250$ K (for $t_0 = 4 \times 10^4$ tu) and corresponding grey scale picture of the displacements during the same time interval (using the same coding as in figure 17).

t_0 being a time which is much larger than the period of oscillation of the bases in the Morse potential, but nevertheless finite (figure 21 shows the case of $t_0 = 4 \times 10^4$ tu, i.e. approximately 400 periods of oscillation of the bases). For each value of $\bar{y}(x)$, one can determine the temperature $T_{\rm eff}(x)$ that would give the same value for $\langle y \rangle$. For the example shown in figure 21, the values of $T_{\rm eff}(x)$ vary in the range 190–308 K, which should be compared with the mean temperature in the simulation, equal to 250 K. These data can be used to evaluate theoretically the expected range of the force fluctuations as the breaking propagates. In the experiment, the lifetime of a base pair is imposed by the speed v at which the DNA strand is pulled. It is equal to $\bar{\tau} = 1/v$. For each effective local temperature $T_{\rm eff}$, equation (70) can be used to determine which force F_B one should apply to get this lifetime, keeping in mind that F_B enters into the expression for the potential $V'(y)$. For the example in figure 21, one gets $0.024 < F_B < 0.050$ (corresponding to $38 < F_B < 80$ pN when one restores standard force units), in good agreement with the numerical simulations of figure 17.

The order of magnitude of the forces corresponds to values that are observable in actual experiments, but one must keep in mind that the calculation only involves the force F_B at the level of the base pair that will break (figure 16). This is not the force F_A, which is the measured force, because, between the next base to break and the place where the external force is applied, there is a segment of denaturated strand, which fluctuates due to Brownian motion. The tension of this denaturated segment has fluctuations ΔF, which are superimposed on the fluctuations of F_B that one would like to measure. To calculate ΔF, one could use the WLC discussed in section 2.2.2. A rough estimate can be obtained with a model that considers the denaturated strand as a lattice of M springs, with a coupling constant K, attached to the spring of constant c_0 used to measure the force. One gets $\Delta F = c_0\sqrt{Mk_BT/K}$. The important point is that these fluctuations grow as $\sqrt{M}$ because, as they appear as an added noise on the force of interest, this imposes a fundamental limitation on the experiments which cannot investigate the denaturation of a very long double helix.

Another fundamental limitation is the lifetime of the 'hot' and 'cold' regions in DNA. Figure 21 exhibits inhomogeneities in the thermal fluctuations of the model, which in turn determine the regions which open easily and the regions which are hard to denaturate. But, indeed, equipartition of energy applies to DNA as it does to any physical system. The lifetime of the thermal inhomogeneities is finite [38], and if one averages over a very long time, such inhomogeneities are no longer detectable. Therefore, the speed at which the experiment is carried out should affect the possible observation of nonlinear localization in DNA. The experiments performed by pulling a glass bead with a microneedle are very slow on the molecular timescale, and it is unlikely that they can detect the temporary thermal inhomogeneities. This is why they can probe the sequence, or at least the average sequence over a few tens of base pairs. One can, however, note that when the experiments are carried ten times faster than the original experiments they show additional fast fluctuations which are not understood, and not related to the sequence [37].

In order to try to bridge the gap between the numerical simulations and the experiment, we carried simulations up to $t = 0.4\,\mu$s, with a pulling speed $v = 10^{-5}\,\text{Å}\,\text{tu}^{-1}$. Figure 22 shows that the fluctuations in the speed at which the denaturation processes are still present, and even larger than at larger speed. Sometimes the speed becomes negative, i.e. a denaturated region of the molecule can close again, before reopening later. This gives rise to large force oscillations. It is tempting to speculate on the fact that these fluctuations could explain the experimental fluctuations observed in the fast experiments, but this is still an open question.

In order to reach a quantitative analysis of the experiments, there is another characteristic of DNA that should not be neglected. This is its helicoidal character. By pulling on the two strands of an helix, it is easy to convince oneself that the opening is accompanied by a torsional stress of the helix close to the open part. If the helix is short, the part of the molecule which is still in the helicoidal form, will rotate as a whole to release the stress. But for a long DNA molecule, which moreover may be partly in a globular form instead of being straight, the rotation due to the opening will lead to an accumulation of torsion. The build up of the torsional energy is likely to modify the speed of the opening, and the force necessary to open.

8. The role of the helicoidal geometry, another step towards realistic models

The helicity not only plays a role during a mechanical denaturation experiment; the geometry of the double helix naturally leads to a coupling between the opening and the torsion, as shown in figure 23. The role of the helicity is particularly important when dynamical effects are involved. The simple model introduced previously does not take into account the helicoidal geometry, which can only be described by a much more complex model, shown in figure 24 [39, 40].

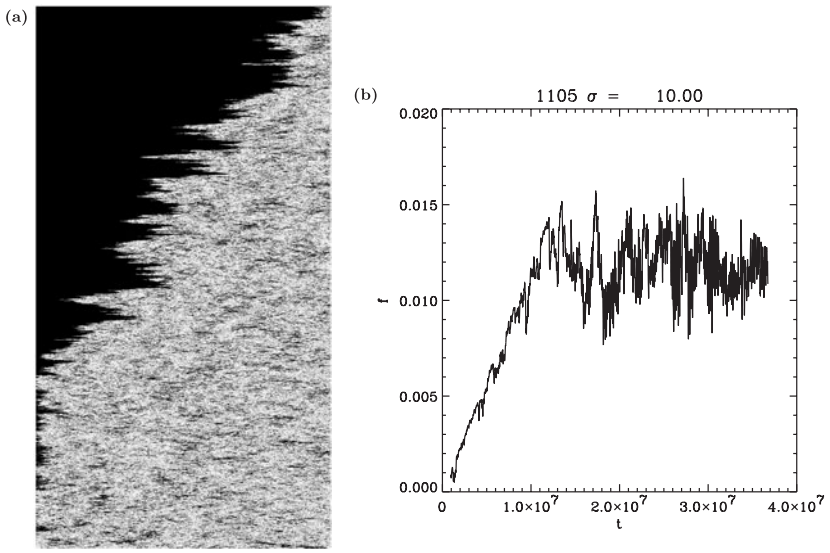

Figure 22. Results of the numerical simulations of the mechanical denaturation of a DNA segment of 1024 base pairs at $T = 320\,\mathrm{K}$ and a stretching velocity $v = 10^{-5}\,\text{Å}\,\text{tu}^{-1}$. The total simulation time is $0.4\,\mu\text{s}$. (*a*) Stretching of the base pairs versus time on a scale ranging from white (fully closed) to black (fully open). The horizontal axis extends along the molecule and the vertical axis is the time axis. (*b*) Variation of the force leading to the denaturation. The unit of force in these simulations is $1\,\text{eV}\,\text{Å}^{-1} = 1.6 \times 10^{-9}\,\text{N} = 1600\,\text{pN}$. The rise corresponds to the first stage of the simulation, during which the spring extends until its tension becomes large enough to begin to break the base pairs. Then, the average force stays constant but one observes very large fluctuations in the instantaneous value.

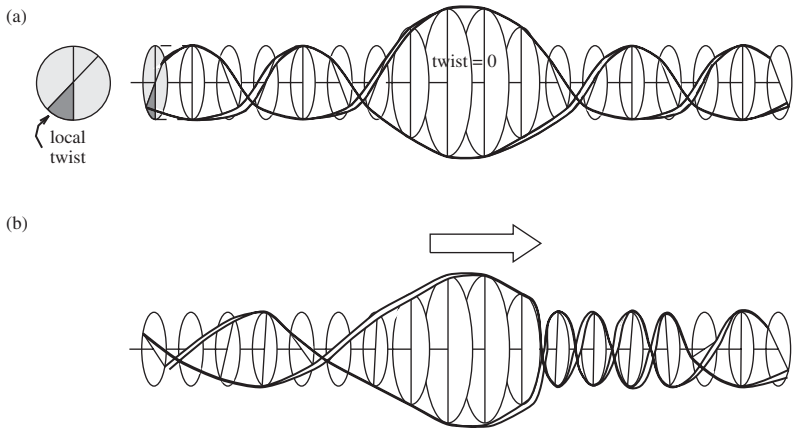

Figure 23. The coupling between opening and twist in DNA dynamics: (*a*) a local opening leads to a local untwisting of the helix; (*b*) when the opening propagates, it produces some overtwisting upstream and some undertwist downstream.

It must include at least two degrees of freedom for base pair n, the distance r_n between the two bases, which plays the same role as the stretching y_n of the simple model, and the angle ϕ_n of the bond linking the bases with respect to a reference axis. The helical structure of DNA is introduced essentially by the competition between a stacking interaction that tends to keep

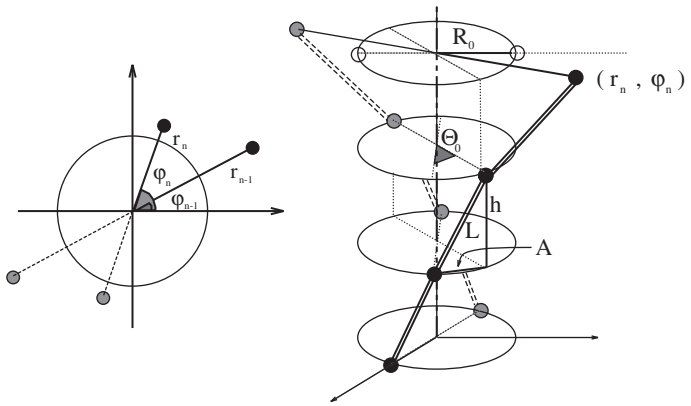

Figure 24. Helicoidal model for DNA.

the base-pairs close to each other (given by the fixed distance h between the base planes) and the length $L_0 > h$ of the backbone segment (described as an elastic rod of rest length L_0) that connects the attachment points of the bases along each strand. The ratio L_0/h fixes the strand slant and, therefore, the resulting helicity of the structure. This helicity is accounted for by the angle of rotation of a base-pair with respect to the previous one, namely the *equilibrium twist angle* θ, equal to $2\pi/10.4$ in B-DNA at room temperature. In the model, bases are described as point-like particles of equal masses m joined by elastic rods along each strand. Bases lying on the same plane are coupled by hydrogen bonds leading to an attractive force that tends to maintain their equilibrium distance equal to the DNA diameter $2R_0$. We assume that the two bases in each pair move symmetrically. The Lagrangian of the model is

$$\mathcal{L} = m\sum_n(\dot{r}_n^2 + r_n^2\dot{\phi}_n^2) - D\sum_n(\exp[-a(r_n - R_0)] - 1)^2 - K\sum_n(l_{n,n-1} - L_0)^2$$
$$- S\sum_n(r_n - r_{n-1})^2\exp[-b(r_n + r_{n-1} - 2R_0)], \tag{72}$$

where $L_0 = \sqrt{h^2 + 4R_0^2\sin^2(\theta/2)}$ and $l_{n,n-1}$ are, respectively, the equilibrium and the actual distance between the two bases n and $n-1$ along a strand,

$$l_{n,n-1} = \sqrt{h^2 + r_{n-1}^2 + r_n^2 - 2r_{n-1}r_n\cos(\phi_n - \phi_{n-1})}. \tag{73}$$

The first term is the kinetic energy. The second describes the interaction between the two bases of a pair with the same Morse potential as in the simple model. The third term describes the elastic energy of the strands, and the last term corresponds to the stacking interaction between the bases. It is similar to the nonlinear stacking interaction introduced in the simple one-component model (equation (40)), without the purely harmonic part which is here replaced by the elasticity of the strands.

The dynamics and statistical physics of this model are of course much more complex than for the one-component model that we discussed earlier, but, although there are some differences introduced by geometrical effects, the major results hold.

- The model sustains localized vibrational modes equivalent to the breathers in the one-component model [40, 41]. The opening is accompanied by a local decrease of the twist as expected from the geometrical constraint.

- The model shows a denaturation transition, which, in the presence of the nonlinear stacking term is very sharp and first-order like [42], although a more complete analysis shows that, in a very narrow range around T_c, the nature of the transition is in fact more complex, suggesting the presence of an underlying Kosterlitz–Thouless type transition [43].

The coupling between opening and twist is particularly important for dynamical effects and, in addition to a central peak also found in the one-component model, which can be attributed to the slow dynamics of the denaturation 'bubbles', the dynamical structure factor of the helicoidal model shows additional features that are not yet fully understood [43].

9. Discussion and conclusion

Physicists are more and more interested in 'complex systems' where the interplay of many individual components leads to new phenomena. Biological molecules provide fascinating examples of such complex systems and this is why they are attracting the attention of a growing number of physicists. But, contrary to proteins, DNA can be viewed as a 'simple complex system' because it has a fairly ordered structure that Schrodinger called a 'quasi-periodic crystal' [1]. A peculiarity of biological molecules is the very large conformational changes which they can undergo, and which are often essential for their function. They are highly deformable, hence highly nonlinear objects, and DNA is not an exception to this rule. Therefore, besides its fundamental interest for biology, it is a very rich system for nonlinear science, perhaps even too rich because this led physicists to speculate on the role of nonlinear excitations in DNA without paying enough attention to experimental realities, triggering a strong rejection reaction by some biologists [44]. As noted by Frank-Kamenetskii, 'there has hardly been a novelty or a vogue among solid-state physicists that has not been applied to DNA', and one cannot exclude that the idea of nonlinear localization, presented in this review, may belong to this category.

Hopefully, nonlinear localization may, however, have some reality in DNA although the 'breathers' that we described are unlikely to exist as perfectly periodic localized modes. What may be more relevant for DNA is the observation that nonlinearity facilitates the formation of localized motions, that can exist in many types of lattices and emerge from thermal fluctuations. For natural DNA, which includes the genetic code, the inhomogeneity of the sequence is certainly a first source for inhomogeneous fluctuations, similar to what was observed in proteins [45], but nonlinearity can greatly enhance this effect. The relevance of these effects to biology is not yet clearly established, but there are, however, some recent studies which strongly suggest that the thermally induced openings, which are predicted by the one-component DNA model that we have described, agree with the experimental observations on DNA opening detected by potassium permanganate footprinting, and, more importantly, coincide with the location of functionally relevant sites for transcription [46]. These observations were made on several sequences, and, moreover, when a mutation is made to remove an initiation site, the model also detects that a region undergoing large thermal fluctuations has been eliminated. Although these results are very preliminary and have to be confirmed by further studies they may open the way for a dynamical analysis of the genome, based on a model suggested by nonlinear science.

Another direction of research, which could well emerge as a major field in the next decade, is the use of DNA for purely physical applications, in particular thanks to its remarkable ability to perform highly selective self-assembly. We have shown that experiments can manipulate single DNA molecules. It is also remarkable that physical methods can be used to control the conformation of a single DNA molecule. This has been achieved with DNA single strands

which have at both ends sequences of complementary base pairs. As a result, the two ends tend to get together to form a double helix, folding the single strand into a hairpin shaped molecule. The closing of this hairpin can be precisely monitored because, using selective biological reactions, one can attach to one end of the DNA strand a chromophore and to the other end a quencher, i.e. a molecular group that suppresses the fluorescence of the chromophore when both are close to each other [47, 48]. The experimental possibilities, however, go beyond *observing* the conformation of the hairpin. It is now possible to *control* it electronically. Using selective reactions one can attach a short organic chain to the loop of the hairpin, and then bind to it a gold nanocrystal which can be heated by a microwave field. When the field is on, enough heat is transmitted to the DNA hairpin to cause the thermal denaturation of its double helix part. The hairpin opens and the fluorescence of the chromophore shows up because the two ends of the molecule are far apart and the quencher is, therefore, away from the chromophore. When the microwave field is switched off, the self-assembly of DNA closes the hairpin again, and the whole process can be repeated [49]. This remarkable experiment is only one example of what can now be achieved using denaturation and self-assembly of DNA. One can use this transition as the basis of a working nanomachine [50, 51], but also for the fabrication of objects with nanometre precision [52, 53]

Acknowledgments

MP wants to thank T Dauxois, N Theodorakopoulos, M Barbi, S Cocco, who participated in different aspects of the work reviewed in this paper, for their contribution to these studies.

References

[1] Schrödinger E 1944 *What is Life?* (Cambridge: Cambridge University Press)
[2] Watson J D and Crick F H C 1953 Molecular structure of nucleic acids *Nature* **171** 737–8
[3] Yakusheich L V 1998 *Nonlinear Physics of DNA* (Chichester: Wiley)
[4] Kavenoff R, Klotz L C and Zimm B H 1973 On the nature of chromosome-sized DNA molecules *Cold Spring Harbor Symp. on Quantitative Biology* vol 38, pp 1–8
[5] Kavenoff R and Zimm B H 1973 Chromisome-sized DNA molecules *Chromosoma* **41** 1–27
[6] Saenger W 1984 *Principles of Nucleic Acid Structure* (Berlin: Springer)
[7] Calladine C R and Drew H R 1992 *Understanding DNA* (London: Academic)
[8] Gueron M, Kochoyan M and Leroy J L 1987 A single mode of DNA base-pair opening drives imino proton exchange *Nature* **328** 89
[9] Frank-Kamenetskii M 1987 How the double helix breathes *Nature* **328** 17
[10] Urabe H and Tominaga Y 1982 Low-lying collective modes of DNA double helix by Raman spectroscopy *Biopolymers* **21** 2477
[11] Movileanu L, Benevides J M and Thomas G J Jr 2002 Temperature dependence of the Raman spectrum of DNA: II. Raman signatures of premelting and melting transitions of poly(dA)-poly(dT) and comparison with poly(dA-dT)-poly(dA-dT) *Biopolymers* **63** 181
[12] Grimm H and Rupprecht A 1995 Inelastic neutron scattering of oriented DNA *Nonlinear Excitations in Biomolecules* ed M Peyrard (Paris: Les Editions de Physique)
[13] Gelbart W M, Bruinsma R F, Pincus P A and Parsegian V A 2000 Dna-inspired electrostatics *Phys. Today* **53** 38–44
[14] Inman R B and Baldwin R L 1964 Helix–random coil transitions in DNA homopolymer pairs *J. Mol. Biol.* **8** 452–69
[15] Wartell R M and Benight A S 1985 Thermal denaturation of DNAmolecules: a comparison of theory with experiments *Phys. Rep.* **126** 67
[16] Bonnet G, Krichevsky O and Libchaber A 1998 Kinetics of conformational fluctuations in DNA hairpin-loops *Proc. Natl Acad. Sci. USA* **95** 8602–6
[17] Smith S B, Finzi L and Bustamante C 1992 Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads *Science* **258** 1122–6

[18] Smith S B, Cui Y and Bustamante C 1996 Overstretching B-DNA: the elastic response of individual double-stranded and single stranded DNA molecules *Science* **271** 795–9
[19] Cluzel P, Lebrun A, Heller C, Lavery R, Viovy J-L, Chatenay D and Caron F 1996 DNA: an extensible molecule *Science* **271** 792–4
[20] Bustamante C, Marko J, Siggia E and Smith B 1994 Entropic elasticity of λ-phage DNA *Science* **265** 1599–600
[21] Bouchiat C, Wang M D, Allemand J F, Strick T, Block S M and Croquette V 1999 Estimation of the persistence length of a worm-like chain molecule from force–extension measurements *Biophys. J.* **76** 409–13
[22] Lavery R, Lebrun A, Allemand J F, Bensimon D and Croquette V 2002 Structure and mechanics of single biomolecules: experiments and simulation *J. Phys.: Condens. Matter* **14** R383
[23] Theodorakopoulos N 2002 *Proc. Int. Conf. on Localization and Energy Transfer in Nonlinear Systems (Escorial, Spain, 2002)* (Singapore: World Scientific) at press
Theodorakopoulos N 2002 Phase transitions in homogeneous biopolymers: basic concepts and methods *Preprint* org/cond-mat/0210188
[24] Nose S 1984 A unified formulation of the constant temperature molecular dynamics methods *J. Chem. Phys.* **81** 511–9
[25] Martyna G J, Klein M L and Tuckerman M 1992 Nosé–Hoover chains: the canonical ensemble via continuous dynamics *J. Chem. Phys.* **97** 2635–43
[26] Daumont I and Peyrard M 2003 One dimensional 'turbulence' in a discrete lattice *Chaos* **13** 624–36
[27] Remoissenet M 1986 Low-amplitude breather and envelope solitons in quasi-one-dimensional physical models *Phys. Rev.* B **33** 2386–92
[28] Daumont I, Dauxois T and Peyrard M 1997 Modulational instability: first step toward energy localization in nonlinear lattices *Nonlinearity* **10** 617–30
[29] Bang O and Peyrard M 1996 Generation of high energy localized vibrational modes in nonlinear Klein–Gordon lattices *Phys. Rev.* E **53** 4143–52
[30] Flach S and Willis C 1998 Discrete breathers *Phys. Rep.* **295** 181
[31] Dauxois T, Theodorakopoulos N and Peyrard M 2002 Thermodynamic instabilities in one dimension: correlations, scaling and solitons *J. Stat. Phys.* **107** 869–91
[32] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes* (Cambridge: Cambridge University Press)
[33] Theodorakopoulos N 2003 Thermodynamic instabilities in one-dimensional particle lattices: a finite-size scaling approach *Phys. Rev.* E **68** 026109
[34] Theodorakopoulos N, Dauxois T and Peyrard M 2000 Order of the phase transitions in models of DNA thermal denaturation *Phys. Rev. Lett.* **85** 6
[35] Essevaz-Roulet B, Bockelmann U and Heslot F 1997 Mechanical separation of the complementary strands of DNA *Proc. Natl Acad. Sci. USA* **94** 11935–40
[36] Hänggi P, Talkner P and Borkovec M 1990 50 years after Kramers *Rev. Mod. Phys.* **62** 251
[37] Heslot F 1999 Private communication
[38] Peyrard M and Farago J 2000 Nonlinear localization in thermalized lattices: application to DNA *Physica* A **288** 199–217
[39] Barbi M, Cocco S, Peyrard M and Ruffo S 1999 A twist-opening model of DNA *J. Biol. Phys.* **24** 97
[40] Barbi M, Cocco S and Peyrard M 1999 Helicoidal model for DNA opening *Phys. Lett.* A **253** 358
[41] Cocco S, Barbi M and Peyrard M 1999 Vector nonlinear Klein–Gordon lattices: general derivation of small amplitude envelope soliton solution *Phys. Lett.* A **253** 161
[42] Cocco S and Monasson R 1999 Statistical mechanics of torque induced denaturation of DNA *Phys. Rev. Lett.* **83** 5178
[43] Barbi M, Lepri S, Peyrard M and Theodorakopoulos N 2003 Thermal denaturation of an helicoidal DNA model *Phys. Rev.* E **68** 061909
[44] Frank-Kamenetskii M 1987 Physicists retreat again *Nature* **328** 108
[45] Frauenfelder H 1989 The Debye–Waller factor: from villain to hero in protein crystallography *Int. J. Quantum Chem.* **35** 711
[46] Kalosakas G, Rasmussen K O, Bishop A R, Choi C H and Usheva A 2003 Structurally specific thermal fluctuations identify functional sites for DNA transcription *Preprint* cond-mat/0309157
[47] Bonnet G, Krichevsky O and Libchaber A 1998 Kinetics of conformational fluctuations in DNA hairpin-loops *Proc. Natl Acad. Sci. USA* **95** 8602–6
[48] Altan-Bonnet G, Libchaber A and Krichevsky O 2003 Bubble dynamics in double-stranded DNA *Phys. Rev. Lett.* **90** 138101
[49] Hamad-Schifferli K, Schwartz J J, Santos A T, Zhang S and Jacobson J M 2002 Remote electronic control of DNA hybridization through inductive coupling to an attached metal nanocrystal antenna *Nature* **415** 152

[50] Alberti P and Mergny J-L 2003 DNA duplex–quadruplex exchange as the basis for a nanomolecular machine *Proc. Natl Acad. Sci. USA* **100** 1569
[51] Yurke B, Tuberfield A J, Lils A P and Neuman J L 2000 A DNA-fuelled molecular machine made of DNA *Nature* **406** 605
[52] Winfree E, Liu F, Wenzler L A and Seeman N C 1998 Design and self-assembly of two-dimensional DNA crystals *Nature* **394** 539
[53] Zanchet D, Michell C M, Parak W J, Gerion D, Williams S C and Alivisatos A P 2002 Electropheric and structural studies of DNA-directed Au nanoparticle groupings *J. Phys. Chem.* B **106** 11758