Modélisation d'un simple brin d'ADN : Configuration en "épingle à cheveux"

Jalal Errami

▶ To cite this version:

Jalal Errami. Modélisation d'un simple brin d'ADN : Configuration en "épingle à cheveux". Data Analysis, Statistics and Probability [physics.data-an]. Ecole normale supérieure de lyon - ENS LYON, 2007. English. NNT : . tel-00156784

HAL Id: tel-00156784
https://theses.hal.science/tel-00156784
Submitted on 22 Jun 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Numéro d'ordre : 403
Numéro attribué par la bibliothèque : 07ENSL0 403

Laboratoire de Physique de l'École Normale Supérieure de Lyon

THÈSE
en vue d'obtenir le grade de
Docteur de l'École Normale Supérieure de Lyon

spécialité : physique
École doctorale de physique et astrophysique de Lyon

présentée et soutenue publiquement le 11 Mai 2007
par Monsieur Jalal ERRAMI

Modelling DNA Hairpins

Directeurs de thèse :	Michel PEYRARD et Nikos THEODORAKOPOULOS
Après avis de :	Monsieur Ralf BLOSSEY, membre et rapporteur Monsieur Wolfgang DIETERICH, membre et rapporteur

Devant la Commission d'examen formée de :

Monsieur Ralf BLOSSEY, membre et rapporteur
Monsieur Wolfgang DIETERICH, membre et rapporteur
Monsieur Ralf EVERAERS, membre
Monsieur Michel PEYRARD, membre
Monsieur Nikos THEODORAKOPOULOS, membre

Modelling DNA Hairpins

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
im Rahmen des grenzüberschreitenden Promotionsverfahrens
zwischen der
Universität Konstanz

vorgelegt von

Jalal Errami (DEA)
aus Vienne (Frankreich)

Tag der mündlichen Prüfung : 11 Mai 2007

Referent:	Dr. Ralf Blossey
Referent:	Prof. Dr. Wolfgang Dieterich

Remerciements

Je tiens tout d'abord à remercier Michel Peyrard pour m'avoir proposé ce sujet de thèse et pour son encadrement tout au long de ce périple malgré son emploi du temps déjà chargé. Je tiens aussi à le remercier au travers de ces quelques mots pour son soutien moral qu'il m'a apporté durant ces trois dernières années, alors un grand merci, du fond du coeur. Je remercie également Nikos Theodorakopoulos pour avoir accepté de co-encadré ce travail de thèse et pour son investissement durant mes séjours à l'Université de Constance. Je voudrais remercier également Johannes pour toute son aide concernant le coté germanique de cette thèse. Mes remerciements sont également adressés aux rapporteurs W. Dieterich et R. Blossey pour avoir accepté de juger mon travail, ainsi qu'à R. Everaers pour avoir accepté de faire partie du jury.

J'en arrive maintenant aux différentes personnes que j'ai pu cotoyé dans les couloirs de l'Ecole et avec qui nous avons passé de très bon moment, elles seraient trop nombreuses à citer mais je pense qu'elles se reconnaitront, donc merci à vous.

Je ne serais oublier mes amis d'enfance et de ping qui m'ont supporté toute ces années: Boubeche, Ahmed, Hatouf, Mehmet, Rachon, Chanouse, Moha, Ichamouse, Nasson, mes amis du quartier de l'Isle et j'en oubli pleins d'autres, vraiment merci à vous.

J'en viens à présent à ma famille. Merci à mes frères Mouns, Hims et Baguet et à ma soeur Noums, pour tout ce que l'on a vécu ensemble et pour leur soutien tout au long de mes études. Je ne serais comment remercier mes parents Fati et Ben, si ce n'est en leur dédiant ce manuscrit, qui n'aurait jamais vu le jour sans leur dévouement. Merci Maman. Merci Papa.

Contents

Introduction **xi**

I DNA molecule and Single-Stranded DNA **1**

1 The DNA molecule and Single Stranded DNA, Hairpins **3**
- 1.1 The DNA molecule 3
  - 1.1.1 DNA structure and conformation 4
  - 1.1.2 DNA properties 7
    - 1.1.2.1 Replication and Transcription 7
    - 1.1.2.2 Melting of DNA 7
  - 1.1.3 DNA melting models 10
    - 1.1.3.1 Microscopic model 10
    - 1.1.3.2 Poland and Scheraga model 10
    - 1.1.3.3 PBD model 12
    - 1.1.3.4 Helicoidal Model 13
- 1.2 Single stranded DNA 14
  - 1.2.1 How to get it? 14
  - 1.2.2 Why is it interesting to study ssDNA and their hairpin form? 15

2 Review of experimental properties of DNA hairpins. **21**
- 2.1 Bulk fluorescence 21
  - 2.1.1 Fluorescence Resonance Energy Transfer 21
  - 2.1.2 Fluorescence Bulk measurements 23
    - 2.1.2.1 Measurement principle 23
    - 2.1.2.2 Results 24
- 2.2 Fluorescence Correlation Spectroscopy(FCS): Kinetics 25
  - 2.2.1 Experimental protocol 26
  - 2.2.2 Results 26
- 2.3 Static Absorbance measurements 29
  - 2.3.1 Experiment 29
  - 2.3.2 Analysis 30

3 Review of some polymer and protein models 33
3.1 Polymer theory 33
3.1.1 Introduction 33
3.1.2 Freely jointed chain 34
3.1.2.1 End-to-end vector 35
3.1.2.2 End-to-end vector distribution 36
3.1.3 Freely rotating chain 37
3.1.3.1 End-to-end vector 38
3.1.3.2 End-to-end vector distribution 40
3.1.4 Kratky-Porod chain 41
3.1.4.1 An exact calculation of $P_N(\mathbf{r})$ 41
3.1.4.2 Effective Gaussian approach 44
3.1.5 Growth of a polymer chain 45
3.2 Protein models 49
3.2.1 Protein folding 49
3.2.2 Lattice models 50

II Modelling DNA hairpins 53

4 A two dimensional lattice model 55
4.1 Self assembly of DNA hairpins 55
4.1.1 Model 55
4.1.2 Metropolis-Monte Carlo scheme 57
4.2 Equilibrium properties of the opening-closing transition 59
4.2.1 The transition in the absence of mismatch 59
4.2.2 Role of the mismatches 62
4.3 Kinetics of the opening and closing 63

5 PBD-Polymer model for DNA Hairpins 69
5.1 Presentation of the model 69
5.2 Study of the stem 71
5.2.1 Partition function 73
5.2.2 Transfer integral in the continuum medium approximation . . 75
5.2.3 Results 77
5.3 The complete system 81
5.3.1 Partition function 82
5.3.2 Free Energy and Entropy 84
5.3.3 Kinetics: theoretical predictions 85
5.4 Case of S≡1 90
5.4.1 Thermodynamics 91
5.4.1.1 Role of the loop 92
5.4.1.2 Role of the stem 95

5.4.2 Kinetics . . . 96
5.4.2.1 FRC model . . . 97
5.4.2.2 Discrete Kratky-Porod chain . . . 99
5.5 Complete calculation: $S \neq 1$ . . . 100
5.5.1 Thermodynamics . . . 101
5.5.1.1 FRC model . . . 101
5.5.1.2 Discrete Kratky-Porod model . . . 107
5.5.2 Kinetics . . . 111
5.5.2.1 FRC model . . . 111
5.5.2.2 Discrete Kratky-Porod model . . . 114
5.5.3 Discussions . . . 116
5.5.4 Beyond the PBD-model for the stem . . . 117

Conclusion **123**

Summary **127**

Zusammenfassung **129**

Résumé **133**

III Appendices **135**

A Calculation of $P_N(R)$ for the Kratky-Porod chain **137**

B The Gaussian chain **141**
B.1 Theoretical predictions . . . 141
B.2 Monte Carlo simulation . . . 145

References **151**

Introduction

Introduction

DNA hairpins are molecules made of a single strand of DNA which has two complementary sequences of bases at its two ends. As a result the ends tend to bind to each other to form a short piece of double stranded DNA, called the stem of the hairpin. The remaining part of the strand makes a loop as shown on Fig. (1).

DNA hairpins have a dual interest. First they play important roles in biology such as the regulation of gene expression during transcription [2]. Second, hairpins provide a model system to study the self-assembly process that leads to the formation of the famous DNA double helix. This self-assembly can occur in solutions that contain a sufficient concentration of two complementary DNA species. But the process is complex because the complementary strands must first find each other in solution and then assemble. In a hairpin, the two parts that have to assemble are already attached to each other. Therefore the process leading to their assembly is simpler. Moreover, as explained later in the manuscript hairpins can be studied very precisely in experiments using some fluorescent dyes [3]. As a result accurate experimental results on the assembly-dis-assembly of the stem can be collected [4, 5]. The goal of our study is to propose a suitable model for the equilibrium statistical physics and kinetics of the closing and opening of DNA hairpins. As DNA hair-

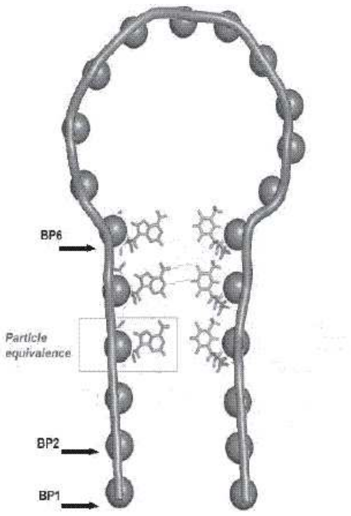

Figure 1: *schematic representation of a DNA hairpin configuration [1].*

pins are fairly simple biological molecules, their self-assembly in solution is a more tractable problem than either protein folding or DNA double helix formation and one can isolate more easily a plausible reaction coordinate, which is the end-to-end distance. In particular when one compares their assembly to protein folding, one could think that this task has already been completed. This is not the case. Of course some studies have been performed [6, 7], and we shall review them in Chap. 2, but they are phenomenological and rely on many empirical parameters which are difficult to evaluate quantitatively and have to be fitted on experimental results. The difficulties are not restricted to the theoretical level. Even the experiments raise puzzling questions because the studies of Libchaber and coworkers [4] disagree on some fundamental points with the measurements of Wallace *et al.* [8] and Ansari [6]. All experiments agree qualitatively on the equilibrium thermodynamics properties. The melting temperature T_m decreases with the length of the loop and T_m is lower for a poly(A) than for a poly(T) loop. Discrepancies appear in the kinetic studies. While all agree that the activation energy for the opening is positive and does not depend on the loop, different experiments disagree on the properties of the closing. Libchaber and coworkers measure a small **positive** activation energy of closing but Wallace and Ansari find instead a negative activation for closing. A careful analysis shows that the contradiction may be only apparent. First the experiments of Ansari *et al* [7] are made with very short loop (only 4 thymine bases T_4) and a stem of 6 base-pairs while Libchaber and coworkers [4] consider much longer loops (T_{12} to T_{30}) and a shorter stem (5 base-pairs). The experiments of Wallace *et al* consider hairpins which are similar to those studied by Libchaber and collaborators (A_{30} loop, and 5 base-pairs in the stem) but they have varied the solvent. In pure water their activation energy for closing is mostly negative (in the highest range of the temperature domain that has been investigated) but it becomes slightly positive at the lowest temperatures (275K). With a solvent containing $MGCl_2$ (20.10^{-3} mol/l) the activation energy is weakly positive in the whole temperature range which has been studied. In their analysis of the discrepancies between their measurements and those of the group of Libchaber, Ansari *et al.* invoke the possible role of misfolded loops. They could play a dominant role in the low temperature range (where positive activation energies are found by Wallace; similarly all experiments of the Libchaber group are performed significantly below T_m where traps by misfolded loops could play a role). Wallace *et al.* assign the non-Arrhenius behavior that they observe to intrachain interactions within the loop (the breaking of AA stacking interactions in the loop).

All these studies show that although rather complete set of data on DNA hairpins is available, those data are far from being properly understood. The studies by Ansari *et al.* [7, 6], are able to reach a reasonable fit of the experiments but at the expense of a complex loop model which includes a phenomenological cooperativity parameter [7].

Our aim in this work is to examine to what extend statistical physics can describe the properties of DNA hairpins in terms of a basic model with the minimal amount of ad-hoc assumptions and parameters that can be related to the interaction energies between the elements that make the structure of the hairpin. We will of course have to make some limitations, as discussed in this manuscript, but this kind of approach can be fruitful for understanding some properties of DNA hairpins. For instance we shall see in Chap. 5 that a positive activation energy for closing can be found even for a simple loop model.

The first model that we have developed is a two dimensional lattice model with two parameters only [9]. We model the favourable interaction between complementary bases by a parameter d, and introduce a parameter of flexibility ϵ to take into account the rigidity of the strands. We show that we can reproduce qualitatively some experimental results and we report on the role of the mismatches on the thermodynamics and the kinetics of this system by comparing two models one with mismatches, the other without. This first model reveals its limits when quantitative results are sought in particular because the entropy of the system is not properly described. So we have developed an another model, based on the same idea as the first one but some what more sophisticated. We divide the system into two parts, the loop and the stem. We apply for the loop the theory of polymers and for the stem we introduce the base pairing and stacking interactions following the work of Peyrard, Bishop, Dauxois and Theodorakopoulos [10, 11], which has been successful in describing many aspects of DNA denaturation. Our approach involves only fundamental entities relating either to the single-strand structure (polymer rigidity) or to H-bond and stacking interactions. The thermodynamics can be determined using the standard results of the statistical mechanics of systems in equilibrium between two limit states and the kinetics can also be addressed within the framework of the reaction rate theory for systems where it is possible to isolate a reaction coordinate. We will show in this work that the model of the single strand that forms the loop is crucial to reproduce properly the experimental properties of hairpins. In other words hairpins are very sensitive systems to test simple models of single stranded DNA. The interest of the development of such models is not only academic because single stranded DNA is closely related to RNA, which plays a very important role in biology, in particular because it can adopt complex configurations which often include hairpins.

The first chapter of this thesis gives some general backgrounds around the DNA molecule and DNA hairpins. It also presents briefly the previous works around the thermal denaturation of DNA. The second chapter presents a review of some experimental studies dealing with the problem of the self-assembly of single strands of DNA. It also gives a brief review of the problem of protein folding. The third chapter deals with the different polymer models commonly used to model single chains and that we have used for the modelling of the loop part of DNA hairpins. Finally, the fourth and the fifth chapters introduce and discuss the two models that we have developed in order to study the thermodynamics and the kinetics of DNA hairpins.

Part I

DNA molecule and Single-Stranded DNA

Chapter 1

The DNA molecule and Single Stranded DNA, Hairpins

Contents

1.1 The DNA molecule 3
  1.1.1 DNA structure and conformation 4
  1.1.2 DNA properties 7
  1.1.3 DNA melting models 10
1.2 Single stranded DNA 14
  1.2.1 How to get it? 14
  1.2.2 Why is it interesting to study ssDNA and their hairpin form? 15

1.1 The DNA molecule

Desoxyribonucleic acid (DNA) is the molecule which contains all the genetic information inside nucleotide sequences called genes. This molecule was found at the beginning of 20^{th} century [12], but its structure has only been precised in the middle of the century by Watson and Crick [13]. DNA is inside the core of each cell in several forms. For example during the mitose which is the cell division, DNA adopts the chromosomal form whereas for the rest of the time, the molecule is in the interphasic form. The genetic code stored in DNA is expressed during complex processes such as transcription and replication. It is important to notice that more than one meter of DNA is compacted in the nucleus of each cell which has a diameter of 10^{-7} m. Therefore DNA in the cell is not a linear molecule.

1.1.1 DNA structure and conformation

DNA is a very long helicoidal polymer composed of two chains which are twisted around each other. Each chain consists of nucleotides linked by covalent bonds. In the name desoxyribonucleic acid we find nucleic acid and desoxyribose. DNA is a nucleic acid because this molecule is in the core of each cell and is an acid according to Bronstëd. More precisely, in the DNA molecule, monomers of each chain are desoxyribonucleotides. Two of them are purines: Adenosine and Guanosine formed by a five-atom cycle plus a six-atom cycle. The other two are pyrimidines: Cytosine and Thymine formed by a single cycle of six atoms. A desoxyribonucleotide is composed of three molecular parts:

- a cyclic sugar of five carbon atoms (desoxyribose)
- a purine base: Adenine or Guanine or a pyrimidine: Cytosine or Thymine
- and a phosphate linked to the sugar by a phosphoester bond.

The sequences of single bonds between successive nucleotides give a flexibility to the backbone because the rotation around a single bond is quite easy. However the helicoidal configuration of the DNA restricts these rotations.

Each base is linked to the sugar-phosphate backbone, by a covalent bond (N-glycosidic bond) and the two nucleotidic chains are linked together by hydrogen bonds. These hydrogen bonds only exist between complementary bases called base-pairs: Guanine-Cytosine(G-C) and Adenine-Thymine(A-T). Therefore the double helix which has a complementary structure contains the same information in the two strands twisted around each other. Finally the sites where the bases are attached to the backbones are not exactly opposite on a diameter of the section, so that the helicoidal structure of the DNA presents a minor and a major groove.

Using the abbreviation of the bases one can easily describe any nucleotide sequence, which is also called the primary structure. The genetic information is stored in the primary sequence. The sequence is written in the direction from 5'-end to the 3'-end of the sugar phosphate backbone where 5' and 3' label two particular carbon atoms of the sugar 5'-ACCGGTTA-3'OH as shown in Fig. (1.1), or simply, ACCGGTTA (which is different from the opposite sequence, ATTGGCCA) [14]. In the native form, each strand is coupled into a duplex or double helix with its complementary strands.

Figure (1.2) gives some dimensions of the DNA components, Fig.(1.3) shows the double helix according to Crick and Watson and Fig. (1.4) presents the pairing between complementary bases.

There are several conformations of the DNA double helix. The more characteristic structures are called A,B and Z. A and B forms are right-handed helices which

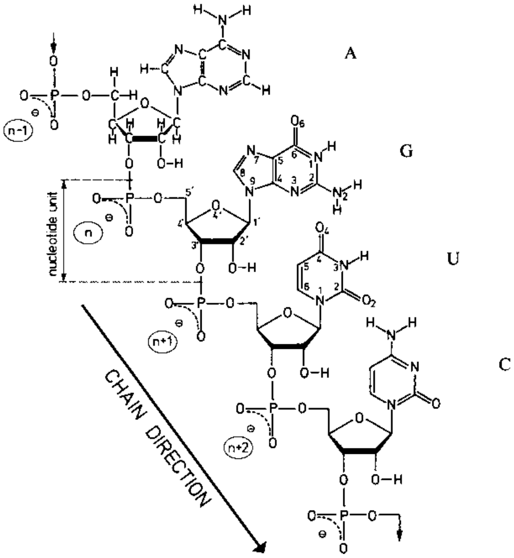
A
G
U
C
nucleotide unit
n-1
n
n+1
n+2
CHAIN DIRECTION

Figure 1.1: *Numeration of the carbon-atom in the sugar [14].*

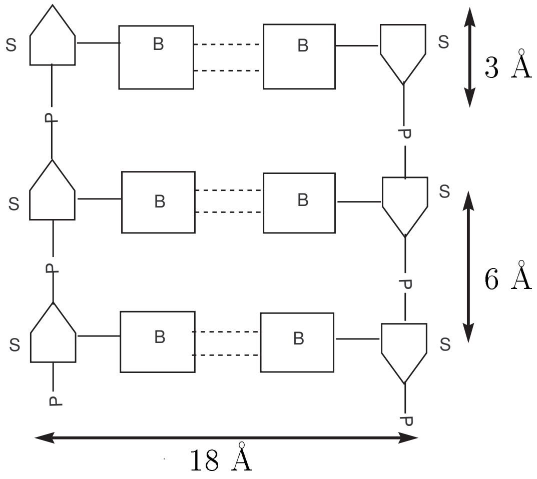
S
B
P
3 Å
6 Å
18 Å

Figure 1.2: *Schematic form of the double chain.*

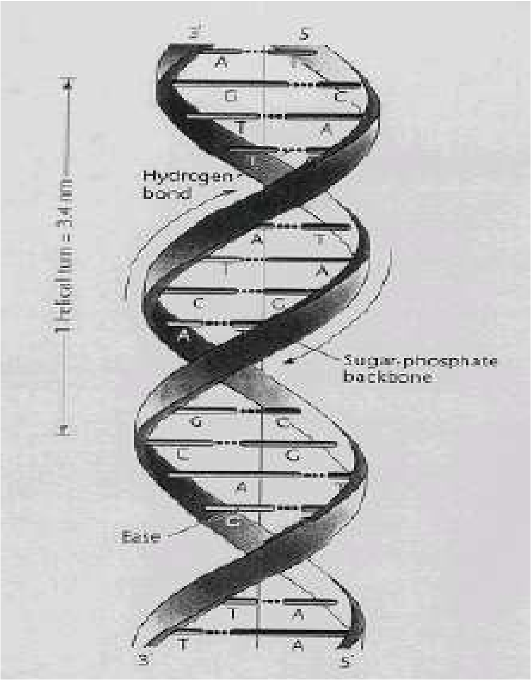
3'
5'
1 Helical turn = 3.4 nm
Hydrogen bond
Sugar-phosphate backbone
Base

Figure 1.3: *The double helix of Crick and Watson [12].*

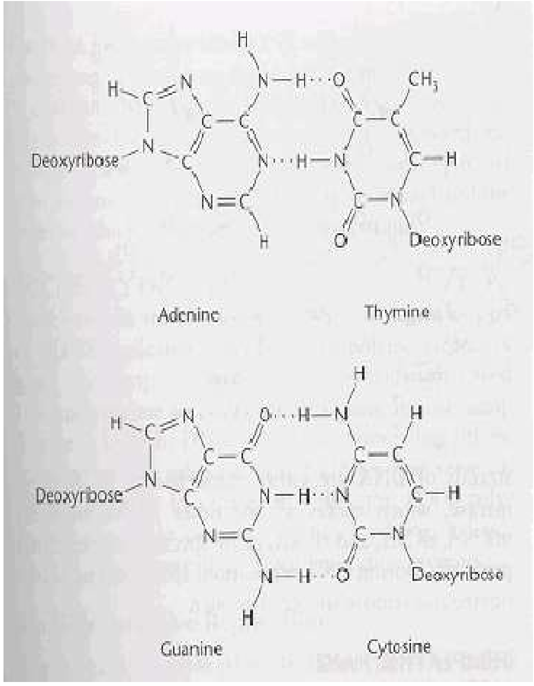
Deoxyribose
Adenine
Thymine
Guanine
Cytosine

Figure 1.4: *Pairing of complementary bases [12].*

turn around their axis counter-clockwise. The difference between these structures is the position of the bases around the axis of the helix and the inclination of the plateau formed by the bases with this axis. In the B helix, the plateaus of the bases is tilted by approximately fifteen degrees with respect to the helix axis. Moreover each base-pair turns about thirty six degrees around the helix axis compared to the previous base-pair. Thus, ten base-pairs are needed to get one full rotation. The B configuration is stable for approximately 92 % of relative humidity. While the A form is stable for approximately 75 % of relative humidity and needs the presence of counter ions such as sodium or potassium. A-T sequences are prone to the B configuration. The distance between base-pairs along the helix axis is 0.34 nm for B configuration and it is not very different for the A form. Another important form is the Z configuration which is a left-handed helix. In this configuration the monomer of the helicoidal chain is the dinucleotide and not the nucleotide. Moreover there are no large grooves and the backbone sugar-phosphate "zigzags" on the periphery of the helix. This conformation only exists in particular conditions: high salt concentrations, methylation of cytosines. Alternate sequences of purines and pyrimidines have a higher tendency to adopt the Z configuration. Figure (1.5) gives an idealized representation of the A,B and Z configurations.

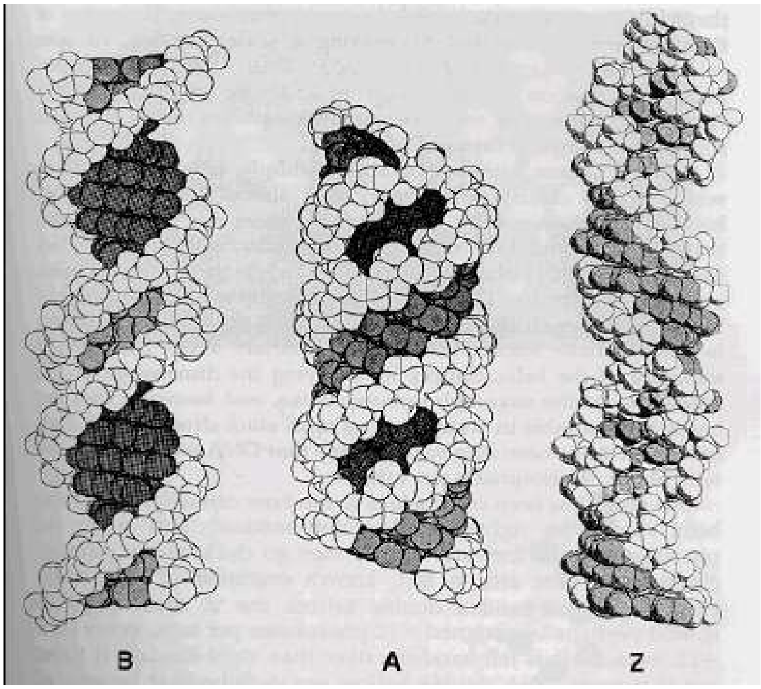

Figure 1.5: *A,B and Z form of the DNA double helix [12].*

1.1.2 DNA properties

The stability of DNA results from various interactions between atoms or groups of atoms of the molecule and interactions with the solvent, as for instance electrostatic interactions between cations such as magnesium and phosphates. Studies of the DNA [26, 15] reveal that its stability is essentially due to two types of interaction between the bases:

- Interaction between complementary bases: hydrogen bonds link the cycles of the two bases forming a pair

- Stacking interaction between base-pairs which are due to hydrophobic interactions and overlap of the π-electrons of the base plateaus

Finally it is important to note that the stacking interaction also exits between consecutive bases of the same chain and is very important in the case of single stranded DNA as we will show in the next sections.

1.1.2.1 Replication and Transcription

DNA is involved in two major events in biology: transcription and replication [14]. For these to occur the DNA double helix has to be untwisted or curved. The transcription is the copy of DNA into a messenger RNA that tells to the cell how to make a protein. DNA only unwinds over a short region, say 15-20 base-pairs, when making RNA. The bubble of unpaired bases can travel along the DNA very rapidly, at about 100 base-pairs per second. When DNA is copied into RNA, a copying enzyme called RNA polymerase attaches itself to one of the two DNA strands and carries out the process of copying DNA into RNA according to the rules of Watson-Crick pairing. There is one difference between RNA and DNA: the Thymine of DNA is replaced by the Uracil in RNA. Using the process called translation, the nucleotidic sequence of the RNA is read by group of three nucleotides, named triplets. Each triplet corresponds to a particular amino acid and sequences of amino acids determine the proteins synthesized by the cell.

The replication is the process by which DNA is copied into another DNA molecule just before a single cell divides into two cells. During this process the DNA double helix has to open completely and an enzyme called DNA polymerase carries out the process of copying DNA into DNA. Figures (1.6) and (1.7) give a schematic representation of replication and transcription of DNA.

1.1.2.2 Melting of DNA

The two strands of a DNA molecule can be dissociated into single polydeoxyribonucleotide strands (the process is also called denaturation or melting) by heat.

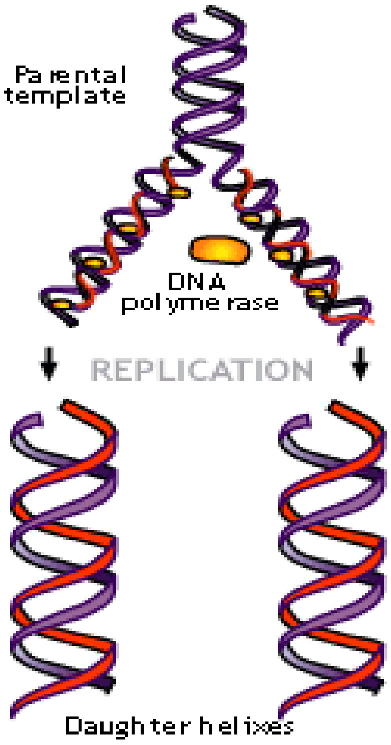

Figure 1.6: *Schematic representation of replication of DNA [16].*

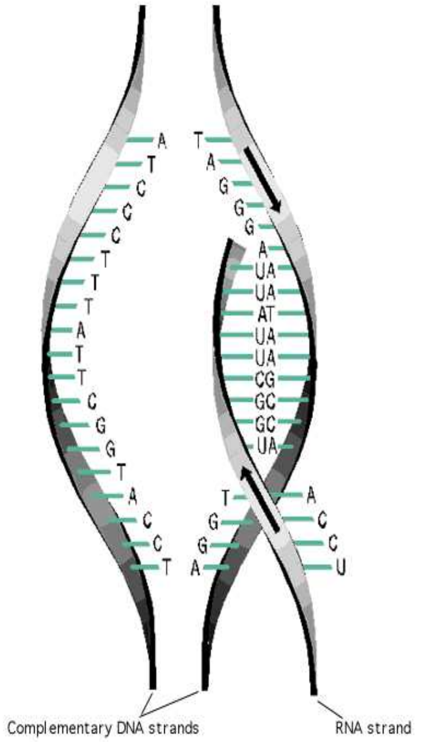

Figure 1.7: *Schematic representation of transcription of DNA [17].*

It occurs because of the breaking of the hydrogen bonds between complementary bases and the disruption of the base stacking. Knowing how denaturation proceeds is important for understanding DNA replication and manipulations of DNA in laboratory. Besides the denaturation due to a temperature increase, the separation of the strands can also be caused by a number of physical factors such as change in salt concentration, pH or other factors. Melting of DNA by heat is a standard method for preparing "single-stranded DNA" (ssDNA).

The denaturation of DNA occurs over a narrow temperature range and causes a number of physical changes. For instance, the ultraviolet absorption at 260 nm increases. The simplest characterization of DNA denaturation is via the melting temperature, T_m, the temperature at which half the melting has taken place. T_m depends on DNA length, sequence, ionic environment, pH, etc. Because GC-pairs are linked by three hydrogen bonds, while AT-pairs only have two, the temperature at which a particular DNA molecule "melts" usually will increase with higher percentage of GC-pairs. The relationship between melting temperature (T_m) and GC content for long DNA can be approximately described:

$$T_m = 69^\circ + 0.41 \times \%(G + C). \tag{1.1}$$

This equation emphasizes that GC-pairs are more stable than AT-pairs but it oversimplifies the phenomenon. As the ordered regions of stacked base-pairs in the DNA duplex are disrupted, the UV absorbance increases. This difference in absorbance between the duplex and single strand states is due to an effect called hypochromicity. Hypochromicity (meaning "less color") is the result of nearest neighbor base-pair interactions. When the DNA is in the duplex state (dsDNA), interactions between base-pairs decrease the UV absorbance relative to that of single strands. When the DNA is in the single strand state the interactions are much weaker, due to the decreased proximity, and the UV absorbance is higher than that in the duplex state. The profile of UV absorbance versus temperature is called a melting curve; the midpoint of the transition determines the melting temperature, T_m. The dependence of the melting temperature, T_m, on the salt concentration can be analyzed to yield quantitative thermodynamic data including ΔH, ΔG and ΔS for the transition from duplex to single stranded DNA. Alternatively, one can get this information by analyzing the whole melting curve.

Thermodynamic analyses of this type are done extensively in biochemistry research labs as well as in physics labs [18, 19, 20] particularly those involved in nucleic acid structure determination. In addition to providing important information about the conformational properties of either DNA or RNA sequences (mismatched base-pairs and loops have distinct effects on melting properties), thermodynamic data for DNA are also important for several basic biochemical applications. For example, information about T_m can be used to determine the minimum length of a oligonucleotide probe needed to form a stable double helix with a target gene at a particular temperature. Figure (1.8) gives a example of a melting curve.

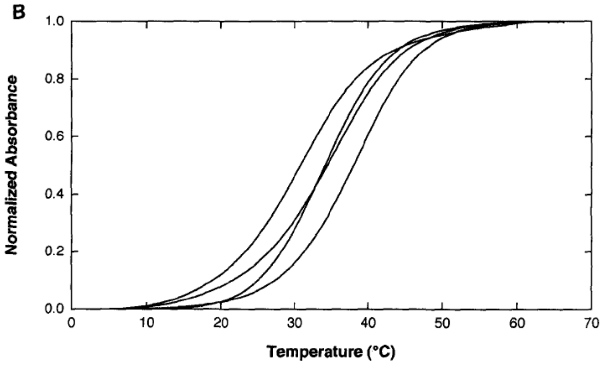

Figure 1.8: *Melting curves example. The solution conditions were 10 nM sodium phosphate, pH 7.0, 1.0 M sodium chloride and a strand concentration of 2μM. The duplex sequences are GCAAAGAC/GTCTTTGC, GCATAGAC/GTCTATGC, GCAGAGAC/GTCTCTGC, and GCACAGAC/GTCTGTGC, with melting temperature of 33.7, 30.6, 35.7, and 38.5 °C, respectively [18].*

1.1.3 DNA melting models

DNA melting can be viewed as a phase transition in a one-dimensional system and it has attracted the attention of theoreticians for the last fifty years. Various models have been developed to study the opening of the double helix and its fluctuational opening. We introduce some of them in this section because they provide a basis for a model for the stem of the hairpin.

1.1.3.1 Microscopic model

This model may appear the most natural at a first sight because it describes the molecule at the atomic scale. It includes all the interactions between the atoms of the macromolecule and must take into account the geometric constraints in the three dimensional space. In this model different types of interactions have to be considered: electrostatic, Van der Waals, angular and dihedral energies. Biophysicists use this type of models in particular to study the dynamics of proteins [21]. The common expressions for the interactions are the following:

- potential describing the stretching of covalent bonds $k_{\text{bond}}(r-r_0)^2$ where k_{bond} is a constant, r the bond length and r_0 the equilibrium length;

- potential of angular rigidity: $k_f(\theta - \theta_0)^2$, where k_f is constant and θ is the polar angle between two consecutive bonds and θ_0 the equilibrium value;

- potential of torsion(rotation around simple bonds): $k_g(1 + \cos\phi)$, where k_g is a fixed parameter and ϕ is the rotational angle around a bond;

- Lennard-Jones potential: $4\epsilon\big[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^6\big]$ for non-bonding interactions

One can easily imagine that this type of calculation needs a very long cpu-time in numerical simulations. And such a detailed study may not be relevant to study large DNA conformational changes. Indeed, the fast microscopic displacements of atoms are not responsible of physical properties of the molecule at mesoscale. We will come back to this point in the second part of this thesis. While microscopic models can be useful to observe the dynamics of the molecule for a short time scale, they cannot be applied to study the melting transition itself, which is a collective effect involving long segments of DNA on time scales which are beyond the possibilities of the present computers. This is even more obvious if one thinks that useful results for the melting can only be provided by the statistics of many individual events and not from a single molecular dynamics trajectory.

1.1.3.2 Poland and Scheraga model

The Poland-Scheraga model takes a completely opposite approach because it tries to use the simplest possible description of the molecule. It was introduced in 1966 by Poland and Scheraga [23, 24]. The model is built upon an original idea by Zimm [25]. The model consists of an alternating sequence (chain) of ordered and unordered states (loops), which represent denaturing DNA in terms of a sequence of double-stranded and single-stranded regions. In the original model [25], the base is assumed to exist in any of three states, bounded in the helix, unbound in free chains or in unbound sequences between two helicoidal portions. The helicoidal (ordered) sequences are energetically favoured over the unbound states and the contribution of the other two states is included in some phenomenological parameters. The nucleation of an ordered (helicoidal) region (a low-probability event controlled by a cooperativity factor [25]), is followed by helix growth, a high probability event controlled by the statistical weight w of the ordered (helicoidal) state. Figure (1.9) illustrates the Poland-Scheraga model schematically. The question which is addressed is the possible first order phase transition in one dimensional system. Indeed, experiments around melting of DNA suggest that the transition is first order [26].

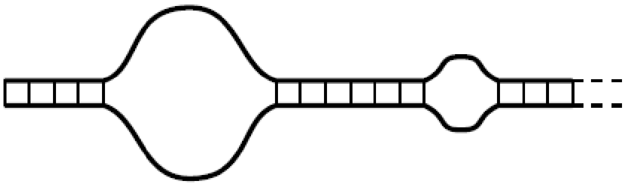

Figure 1.9: *Schematic representation of the Poland-Scheraga model.*

For such a simple model one can compute the partition function Z and the fraction of ordered states in a chain of N base-pairs given by

$$\theta = \frac{1}{N}\frac{\partial \ln Z}{\partial \ln w}, \tag{1.2}$$

where w is the statistical weight of an ordered state, which is not at the end of the ordered sequence. A phase transition occurs if θ has a discontinuity with temperature. But this one-dimensional model would not have a phase transition unless additional ingredients are included. In fact the most delicate aspect of these Ising-like model lies in the evaluation of the entropy of a loop. It must be explicitly included because the model is not rich enough to describe all the configurations of an open region since it uses a simple two-state variable. Poland and Scheraga asserted that the statistical weight of a denaturated sequence of length l is given by the change in entropy due to the added configurations arising from a loop of length $2l$. This has the general form $\frac{As^l}{l^c}$ for large l, where s is the entropy gain for the opening of a single base-pair. As shown by Poland and Scheraga, the value of the exponent c is crucial. No phase transition should occur for $c \leq 1$ and a first order transition arises if $c > 2$. If $1 < c \leq 2$ a phase transition of higher order should occur, although θ is continuous at the transition. They find that $c = d/2$ for ideal random walks, where d is the dimension, there is thus no transition at $d \leq 2$ ($c \leq 1$) and a continuous

transition for $2 < d \leq 4$ ($1 < c \leq 2$).
Fisher [27] has derived the entropy of the denaturated loops modelled as self-avoiding walks. Within this approach, the denaturation transition of DNA is continuous both in two and three dimensions. Indeed, He finds $c = 1.46$ for $d = 2$ and $c \approx 1.75$ for $d = 3$. The transition is thus sharper, but still continuous, in three dimensions.
The proper calculation of c turns out to be a very difficult problem which has only been solved recently. Kafri *et al* [28] and have shown that the DNA denaturation transition could be first order if the effects of excluded volume interaction inside the loop and with the rest of the chain is taken into account. Assuming that the entropy is still given by the expression showed below, they evaluate the exponent c by considering the entropy of a loop of length $2l$ embedded in a chain of length $2L$. Figure (1.10) gives a representation of a such configuration.
They find a lower entropy yielding a larger value of the exponent $c \approx 2.115$ which gives a first order phase transition in dimension 3.

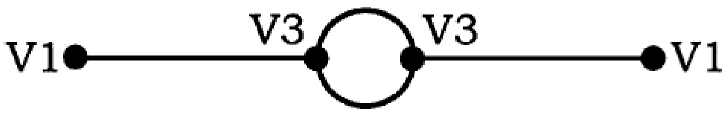

Figure 1.10: *Topology of the loop embedded in a chain. The vertices V_i correspond to the separation between bound and unbound states.*

Finally Blossey and Carlon [29] propose a reparametrizing of the helix nucleation parameters, reanalysing the data including the works of Kafri *et al.*
Besides the need of many parameters, these models are not adapted to short DNA segments and moreover they cannot describe intermediate states between closed and fully open. For instance one aspect which is missing is the actual distance between the strands. For hairpins this is also the distance between the two ends of the loop. This distance is very important to determine the properties of the loop. This is why we have chosen a model which includes this distance.

1.1.3.3 PBD model

This model was introduced by Peyrard and Bishop in 1989 [10] and was improved with Dauxois in 1993 [11, 32]. In this approach the molecule is supposed to be linear in one dimension, and its helicity is not taken into account. Each base-pair is represented by its stretching y and has a mass m. The idea in this approach is to use a potential at the scale of the base. Hydrogen bonds between complementary bases are modelled by a Morse potential and the coupling between consecutive base-pairs is either harmonic or nonlinear. In this last case the coupling constant depends on the state of the two base-pairs which interact. The displacements along the molecule are not considered because they are much weaker than transverse ones.

We will come back to this model in much more details in the second part of this thesis. The Hamiltonian of the system is given by (1.3)

$$H = \sum_n \left[\frac{p_n^2}{2m} + W(y_n, y_{n-1}) + V(y_n) \right], \tag{1.3}$$

where:

$$p_n = m\frac{dy_n}{dt}$$

$$W(y_n, y_{n-1}) = \frac{K}{2}\left[1 + \rho e^{-\alpha(y_n + y_{n-1})}\right](y_n - y_{n-1})^2$$

$$V(y_n) = D\left(e^{-ay_n} - 1\right)^2,$$

with, y_n which is the stretching of the base-pair and K, ρ, α, D and a which are positive constants. Figure (1.11) shows the different interaction potentials in the chain.

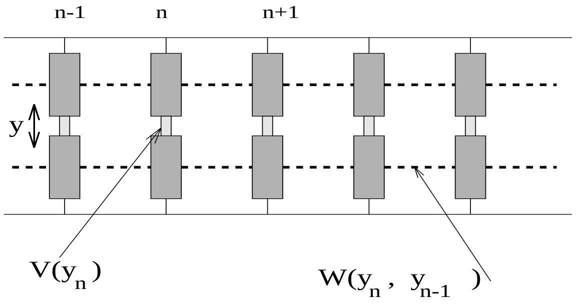

Figure 1.11: *Peyrard-Bishop model for DNA.*

1.1.3.4 Helicoidal Model

In order to be more realistic, Simona Cocco during her PhD [33] with Michel Peyrard, and Maria Barbi developed a DNA helicoidal model [34, 35].This model incorporates the helicity of the molecule [25, 36]. Figure (1.12) shows a schematic representation of the model. This approach, like the previous model uses a Morse potential (V_m) for hydrogen bonds as well as a stacking interaction (V_s). Moreover there is a potential (V_b) which represents the longitudinal vibration of the molecule which is coupled to the stretching of the base-pairs because the backbone is assumed to be rigid. Indeed, to take into account the helicity there is one more degree of freedom compared to the Peyrard-Bishop and Dauxois model. With the notations of Fig. (1.12), the expressions of the potentials are:

$$\begin{aligned} V_m(r_n, r_{n-1}) &= D\left(e^{-a(r_n-R)}-1\right)^2 \\ V_s(r_n, r_{n-1}) &= Ee^{-b(r_n+r_{n-1}-2R)}\left(r_n-r_{n-1}\right)^2 \\ V_b(r_n, r_{n-1}, h_n) &= K\left(h_n-H\right)^2, \end{aligned} \tag{1.4}$$

with E, b, R, K and H which are positive parameters. This model is more complete than the PBD model and it is not necessary to introduce such a complexity for the case of DNA hairpins because we are considering only very short stems. Taking into account the helicity is important for long DNA molecules where torsional energy can build up. For a short stem it can be easily released at the free end and therefore it is not essential for the physics of the system.

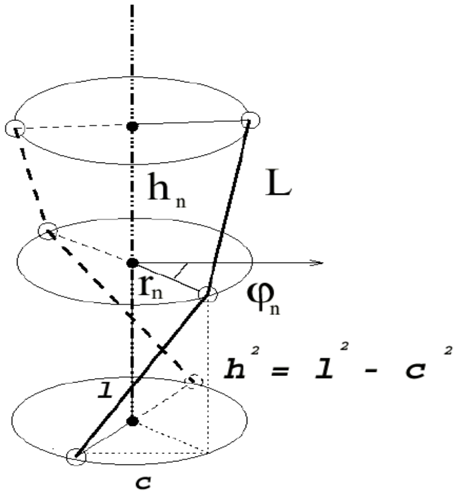

Figure 1.12: *DNA Helicoidal Model [33].*

1.2 Single stranded DNA

1.2.1 How to get it?

A single stranded DNA is one of the two nucleotidic chains of the double helix. In principle it is not difficult to get a ssDNA. Single stranded DNA can be produced experimentally by rapidly cooling heat-denatured DNA. Heating causes the strands to separate and rapid cooling prevents renaturation. Bases in ssDNA also seem to stack to give helicity to the chain. There is a lot of research [37, 38] to characterize the stacking of bases in ssDNA. In DNA the stacking interaction between base-pairs is “a priori” different from the case of ssDNA at least for the intensity of the interaction. Figure (1.13) gives a schematic representation of a ssDNA. The interest of ssDNA also lies on its strong analogy with RNA which plays a large role in biology.

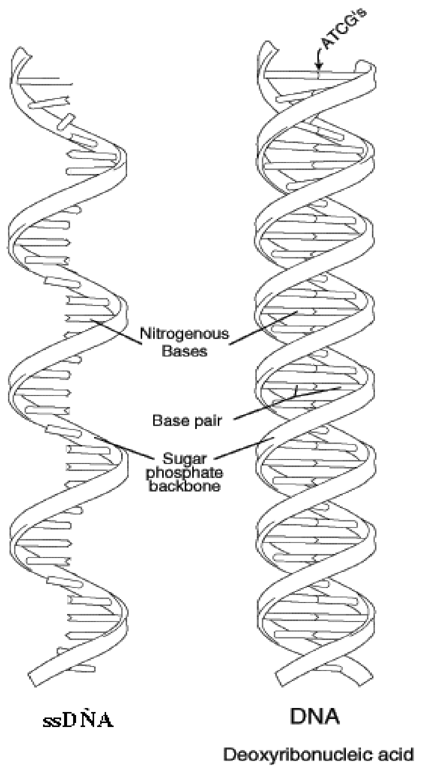

Figure 1.13: *Schematic representation of ssDNA.*

1.2.2 Why is it interesting to study ssDNA and their hairpin form?

ssDNA can form hairpin-loop configurations which are very interesting structures for physicists and biologists [41, 39, 40]. As explained in the introduction, DNA hairpins are short nucleotide strands which have, in their two terminating regions, complementary bases which can therefore self assemble to form a short double helix called the stem of the hairpin. They can exist in two states, the open and the closed state, and fluctuate between the two, being mostly closed at low temperature and mostly open at high temperature. For biologists, regions of DNA molecule where hairpin formation is possible, are believed to play a key role in DNA transposition and in global regulation of gene expression [2]. Moreover loop formation is a first step in the folding of the RNA molecule [14] and also serve as interaction sites for proteins [42]. DNA hairpins may provide very sensitive probes for short DNA sequences [43]: a loop which is complementary to a sequence to recognise can self assemble with it. It is proposed as an alternative to the DNA-chips [44]. This

prevents the hairpin from closing and it is detected by fluorescence. The hairpin configuration can be adopted by the molecular beacons which are single stranded oligonucleotide comprising a probe sequence embedded within complementary sequences that form the stem part of the hairpin. A fluorophore is covalently attached to one end of the oligonucleotide, and a quencher is covalently attached to the other end. In the absence of target, the stem of the hairpin holds the fluorophore so close to the quencher that fluorescence does not occur. When this probe binds to its target, the rigidity of the probe-target duplex forces the stem to unwind, causing the separation of the fluorophore and the quencher and the restoration of the fluorescence. This allows the detection of probe-target.
For the physicists hairpins provide a very simple system to study the self assembly of DNA with two pieces of strand which are maintained in the vicinity of each other for the assembly. Physical applications of DNA hairpins are beginning to be considered. One remarkable example is the use of DNA hairpins to make memory chips for computers [45]. These systems use the fluorophore/quencher method that we present in the next chapter to detect the opening of the hairpins and use a local laser heating to cause their opening. To construct a memory, transitions between bistable states are generally required. The bistable states correspond to a written state and an unwritten state, respectively. The transition between bistable states is realized by molecular reactions bases on hairpin DNA. DNA molecular memory is composed of two types of DNA: a hairpin DNA and a linear DNA. The hairpin acts as a memory molecule with a memory address, the linear DNA as a data molecule with an address tag of the memory. Figure (1.14) gives a schematic representation of such molecules. The loop region of memory DNA has a memory adress, which is recognized by the data DNA. The address tag part of the data DNA is composed of a complementary base sequence of the loop and the 3'-stem of the memory DNA. This memory exploits a hybridization reaction between the hairpin DNA and the linear DNA in memory addressing. Writing data on the memory is to make the linear

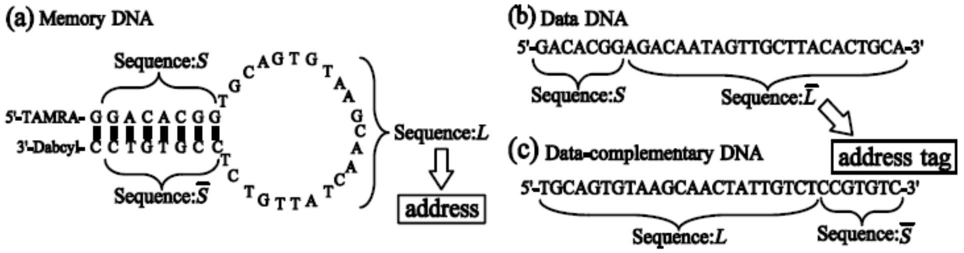

Figure 1.14: *Schematic representation of the memory DNA and the data DNA [45]. (a) Memory DNA: a fluorescent dye TAMRA is attached to the 5'-end and its quencher Dabcyl is attached to the 3'-end. (b) Data DNA: a data DNA has a complementary base sequence of the loop and the 3'-stem of the memory DNA. (c) Data-complementary DNA: a data-complementary base sequences of S and L, respectively.*

DNA hybridize with the hairpin DNA. The hairpin DNA changes from a closed to an open structure when the data is written on the memory. In practice the writing operation follows a serie of operations: heating up a solution of memory DNA and data DNA from room temperature T_R (=25°C) to the writing temperature T_W then cooling it down from T_W to T_R. At T_W the data DNA hybridizes with the memory DNA because the memory DNA opens and the memory-data DNA duplex is stable. Erasing data from the memory is to separate the linear DNA from the hairpin DNA. The hairpin DNA returns to the closed configuration when the data is erased from the memory through a series of operations: heating up the solution from T_R to the erasing temperature T_E and cooling it down quickly from T_E to T_R. The duplex of memory DNA and data DNA is completely dissociated at T_E. The quick cooling allows the memory DNA to close so that the data DNA can no longer access to the memory DNA. Figures (1.15) and (1.16) gives a schematic view of the written and the erasing process. The molecular reactions for addressing of a large amount of DNA molecular memories based on hybridization between the address part of hairpin DNA and the address tag of linear DNA proceed in parallel so that massively parallel addressing of a huge memory space will be possible in principle. There are some problems and the most important one is that the data are not completely erased during the erasing procedure which is due to the fact that the cooling rate of erasing is not fast enough to separate the memory DNA and the data DNA.

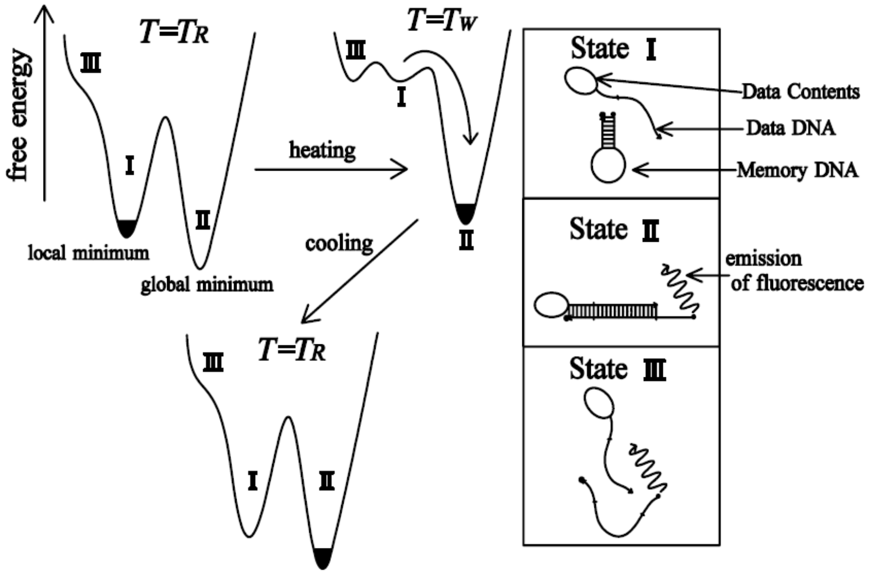

Figure 1.15: *Schematic representation of the writing process [45]. It is composed of the heating from T_R (room temperature) to T_W(writing temperature) then cooling from T_W to T_R.*

Figure (1.17) gives a schematic representation of hairpin-loop configuration for a

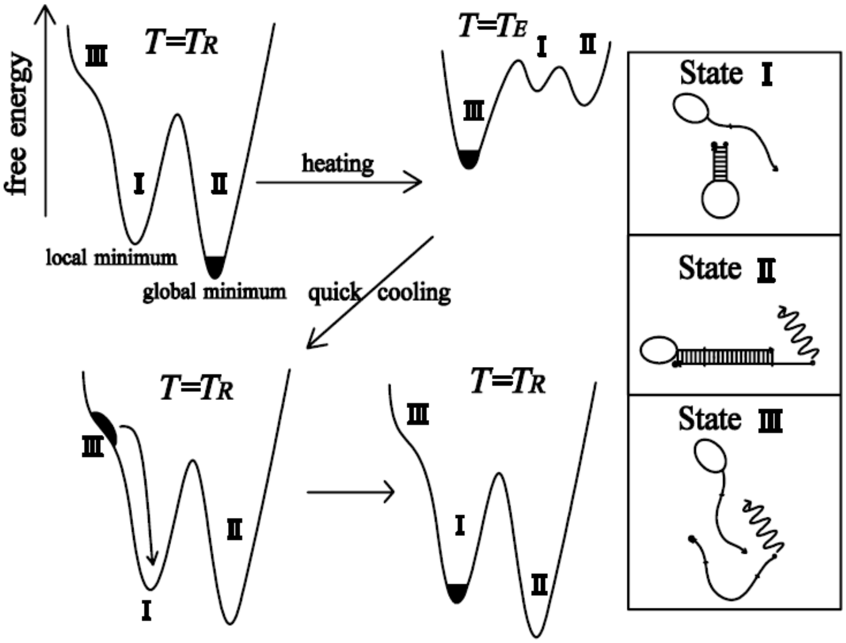
free energy
T=TR
III
I
II
local minimum
global minimum
heating
T=TE
I
II
III
quick cooling
T=TR
III
II
I
T=TR
III
I
II
State I
State II
State III

Figure 1.16: *Schematic representation of the erasing process [45]. It is composed of the heating from* T_R *(room temperature) to* T_E*(erasing temperature) then cooling quickly from* T_E *to* T_R.

RNA (for ssDNA Uracile is replaced by Thymine). Modelling the fluctuations of

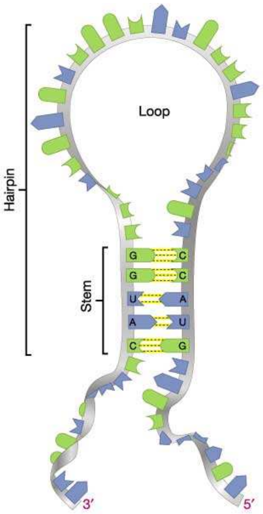
Loop
Hairpin
Stem
G
G
U
A
C
C
C
A
U
G
3′
5′

Figure 1.17: *Schematic representation of RNA loop.*

a hairpin is more challenging than modelling the thermal denaturation of DNA for two reasons:

- the self assembly of a structure is not simply the reverse process of its opening because the elements must find each other in space and then orient properly with respect to each other, before actually assembling in a final stage which is the only stage of the process which can be viewed as the reverse of the breaking;

- the time scales for the assembly can be very long (hundred of μs for instance), i.e. many orders of magnitude longer than the typical time scale of the microscopic dynamics of a macromolecule [46].

Chapter 2

Review of experimental properties of DNA hairpins.

Contents

2.1 Bulk fluorescence 21

2.1.1 Fluorescence Resonance Energy Transfer 21

2.1.2 Fluorescence Bulk measurements 23

2.2 Fluorescence Correlation Spectroscopy(FCS): Kinetics . 25

2.2.1 Experimental protocol 26

2.2.2 Results 26

2.3 Static Absorbance measurements 29

2.3.1 Experiment 29

2.3.2 Analysis 30

In this section we review some of the known experimental results [49, 50] of DNA hairpins and their analysis by the authors of the experiments. This will give us hints on the ingredients required to design a model and experimental facts against which this model can be tested.

2.1 Bulk fluorescence

2.1.1 Fluorescence Resonance Energy Transfer

Fluorescence Resonance Energy Transfer (FRET) is a powerful technique for characterizing distance-dependent interactions at a molecular scale [3]. It is one of the few tools available that is able to measure intermolecular and intramolecular distance interactions both in-vivo and in-vitro.

FRET involves the excitation of a donor fluorophore by incident light within its absorption spectrum. This radiative absorption elevates the donor fluorophore to a higher-energy excited state that would normally decay (return to the ground state) radiatively with a characteristic emission spectrum. If, however, another fluorophore molecule (the acceptor) exists in proximity to the donor with its energy state characterized by an absorption spectrum that overlaps the emission spectrum of the donor, then the possibility of non-radiative energy transfer between donor and acceptor exists. The radiationless energy transfer described above is mediated by dipole-dipole interactions (Van der Waals forces) between the donor and acceptor fluorophore molecules that vary as the inverse 6th power of distance between the two molecules. The rate of energy transfer from donor to acceptor, k_F, is approximately [47]:

$$k_F \approx K_D \left(\frac{r_0}{r}\right)^6 , \tag{2.1}$$

where k_D is the radiative decay rate of the donor fluorophore, or inverse of the fluorescence emission lifetime in the absence of the acceptor fluorophore (typically 1-50 ns), r is the distance between the two molecules, and r_0 is the "Förster distance" that characterizes the 50 % efficiency point of the energy transfer. The FRET efficiency depends on the sixth power [47] of the distance between the two dye molecules:

$$E = \frac{1}{1 + \left(\frac{r}{r_0}\right)^6} . \tag{2.2}$$

FRET is suited to measuring changes in distance on the order of the Förster distance, which is typically 20 to 90 Å. This length scale is far below the Rayleigh-criterion resolution limit of an optical microscope (typically 2500 Å for visible light at high numerical aperture), thus illustrating the power of FRET for measuring extremely small distance interactions.

As an example, Fig. (2.1) shows the overlap of the cyan fluorescent protein (CFP) emission spectrum and the yellow fluorescent protein (YFP) absorption spectrum; this pair supports a strong FRET interaction. After energy transfer occurs from donor to acceptor, the acceptor fluorophore is excited to its fluorescence emission state. Because the observed rate of fluorescence emission from the acceptor is rate-limited by energy transfer from donor to acceptor, the quantitative measurement of FRET emission can therefore provide an inferred measurement of distance using the equation above. Accurate FRET determination generally involves comparison of the donor and donor-acceptor fluorescence emission intensities in samples with and without the acceptor present. A ratio measurement is necessary because, as Fig. (2.1) demonstrates, there is typically overlap between the donor and acceptor emission spectra, thus making it difficult to determine with a single measurement exactly what fraction of the fluorescence measured with an acceptor emission filter derives from only the acceptor. Fluorescence lifetime measurements provide more direct results for the energy transfer rate, are not susceptible to concentration variations, and can be made using time domain or phase modulation lifetime

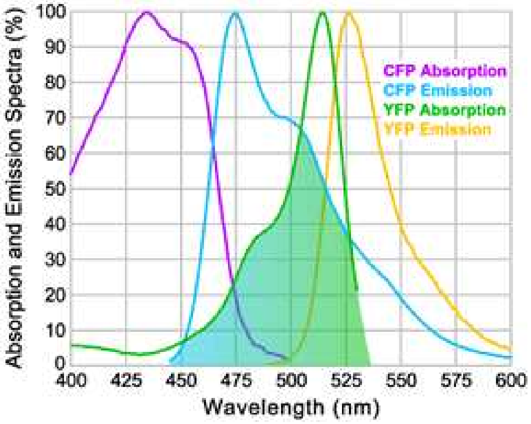

Figure 2.1: *Donor and acceptor absorption and emission spectra [3].*

measurement techniques. These types of measurement can also provide information regarding conformational changes due to molecular interactions.
This technique was used by the group of Libchaber [4] and others [48] to study DNA hairpin-loops and their conformational fluctuations. We present the thermodynamic results obtained by the group of Libchaber in the next section.

2.1.2 Fluorescence Bulk measurements

2.1.2.1 Measurement principle

DNA hairpin-loops are supposed to be in equilibrium between two states: the open state and the closed state. This equilibrium is characterized by an equilibrium constant and rates of opening and closing. In a more complex view one can imagine a transition state between the closed and the open configuration. Figure (2.2) gives a schematic representation of the equilibrium. In the experiments carried by the group of Libchaber, they used molecular beacons which are oligonucleotides capable

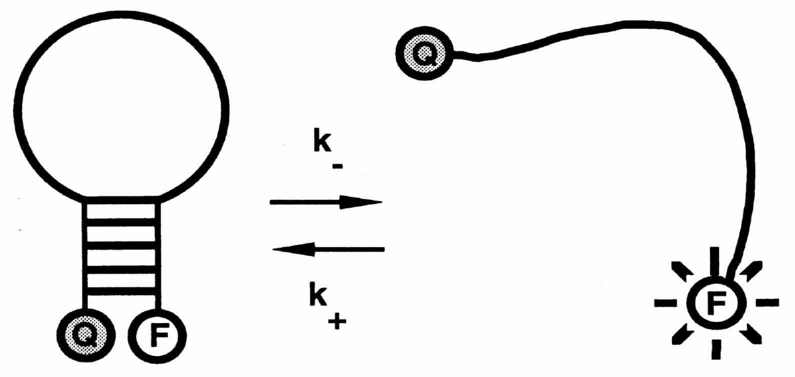

Figure 2.2: *Schematic representation of the two states [4].*

of forming a hairpin loop with a fluorophore and a quencher attached to the two ends of the stem. The conformational state is directly reported by its fluorescence according to the FRET principle: in the closed state the fluorophore is quenched by the quencher and the molecule is not fluorescent; in the open state the fluorophore and the quencher are far apart and the beacon is fluorescent. The sequences of the DNA hairpin-loop under study were 5'-CCCAA-$(N)_n$-TTGGG-3' with varying loop being alternatively $(T)_{12}$, $(T)_{16}$, $(T)_{30}$, or $(A)_{21}$. By monitoring the fluorescence I as a function of the temperature T they can deduce the normalized fluorescence:

$$f(T) = \frac{I(T) - I_c}{I_0 - I_c}, \tag{2.3}$$

where I_0 is the fluorescence of the open beacons and I_c is the fluorescence of the closed beacons. This quantity measures the percentage of open hairpins at a given temperature. Then the equilibrium constant is given by

$$K(T) = \frac{f(T)}{1 - f(T)}. \tag{2.4}$$

It is linked to chemical rates of opening and closing which are essential to deal with the conformational fluctuations of the structure (kinetics).

$$K(T) = \frac{k_-(T)}{k_+(T)}. \tag{2.5}$$

The derivation of Eq. (2.5) is presented in Chap. 4

2.1.2.2 Results

The first interesting result is the shape of the melting curves and the dependence of the melting temperature with the length and the nature of the sequence of the loop. The melting temperature T_m of the structure is defined as the temperature where closing and opening rates are equal, i.e. $K(T_m) = 1$ or $f = 0.5$. Figure (2.3) compares melting curves for a series of poly(A) and poly(T) hairpins. We can notice two important points. First, for poly(A) and poly(T), the melting temperature decreases with the length of the loop and the decay is most significant for Poly(A). One possibility is that the entropic effect produces constraints or forces at the beginning of the stem and induces the opening of the molecule. We will discuss more precisely the relation between the loop length and T_m in Chap. 5 where we analyse the results of our model. Second for a same length of the loop the melting temperature is higher for poly(T) than poly(A). The authors argue that the base stacking is at the origin of the difference from poly(A) to poly(T). Therefore the modelling of stacking interaction in the loop or at least the rigidity of the loop is therefore very important because it seems to explain how the sequence of ssDNA can affects the properties of hairpins. In order to be more precise these authors performed experiments to find the kinetic properties of DNA hairpins using Fluorescence Correlation Spectroscopy.

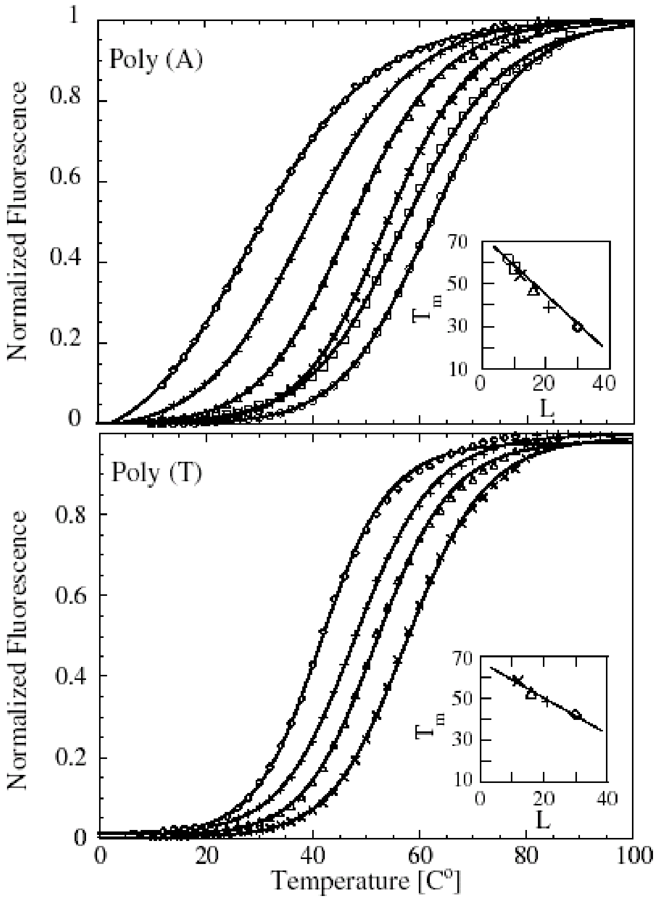

Figure 2.3: *Normalized melting curves. Loop lengths(number of bases) are described by the symbols, ○=8, □=12, ×=12, △=16, +=21, and ◇=30. Data are fit with a single equilibrium mass action law [4]*

2.2 Fluorescence Correlation Spectroscopy(FCS): Kinetics

The idea is to measure the auto-correlation function which reflects the fluctuations of the emitted fluorescence. The problem is that the sources of fluctuations in fluorescence are the diffusion of molecules in and out of the sampling volume and the opening and closing of the secondary structure. Therefore two independent measurements were performed:

1. measurements of the auto-correlation function of the molecular beacons G_{beacon} which contains both diffusion and kinetics contributions.

2. Measurements of the auto-correlation function G_{control} from a sample for which the correlation function consists of the diffusion contribution only. The ratio

of the two function gives the kinetics part and is linked to the sum of the kinetic rates k_- and k_+.

The theoretical form of the auto-correlation function G_{beacon} is a product of a diffusion term and kinetic term [4]:

$$\begin{aligned} G_{\text{beacon}} &= \frac{\langle I(0)I(t)\rangle - \langle I(0)\rangle^2}{\langle I(0)\rangle^2} \\ &= G_{\text{control}}\left(1 + \frac{1-f}{f} e^{-(k_+ + k_-)t}\right). \end{aligned} \tag{2.6}$$

Therefore fitting the ratio $G_{\text{beacon}}/G_{\text{control}}$ gives access to the sum of the rates. Then using the fluorescence bulk measurements k_- and k_+ can be deduced.

2.2.1 Experimental protocol

A laser beam is focused onto the sample with an objective lens and the emitted light is collected through the same objective. It is then focused onto 25 μm diameter pinhole. Then the beam is divided in two by a beam-splitter cube and focused onto two Avalanche photo-counting modules. Finally the signals from these two detectors are fed onto a correlator and the cross-correlation of the excited light is collected. Figure (2.4) gives a schematic drawing of the experimental setup.

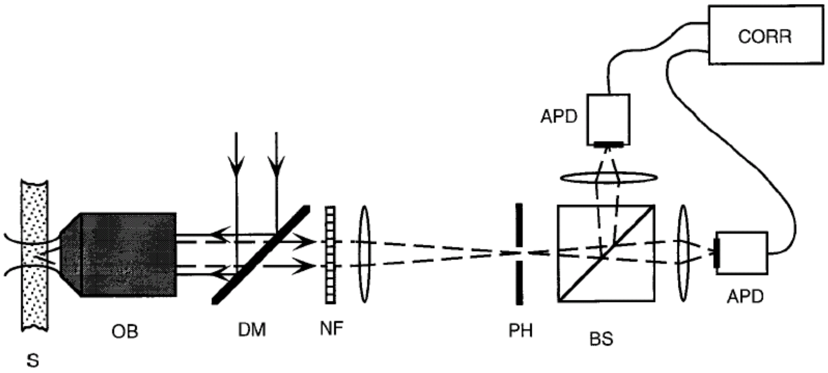

Figure 2.4: *Schematic drawing of the experimental setup. S, sample; OB, objective lens; DM, dichroic mirror; NF, notch filter; PH, pinhole; BS, beam-splitter; APD, Avalanche photo-counting detector; CORR, correlator. [4]*

2.2.2 Results

Figure (2.5) gives the evolution of the rates of opening and closing versus temperature for different loop lengths.

Figure (2.6) gives the evolution of the rates with temperature for the same loop

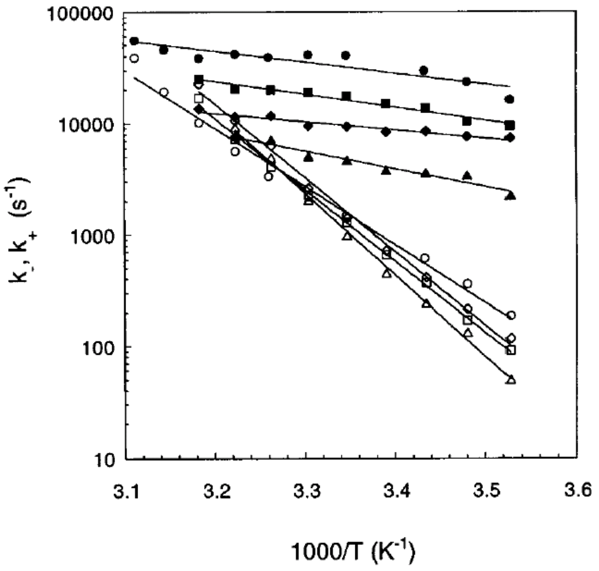

Figure 2.5: *Arrhenius plots of the opening rates (open symbols) and the closing rates (filled symbols) of beacons with different loop lengths: (T)$_{12}$ (circles), (T)$_{16}$ (squares), (T)$_{21}$ (diamonds), and (T)$_{30}$ (triangles). The lines are exponential fits to the data [4].*

length but with a different loop sequence, (A)$_{21}$ and (T)$_{21}$. First of all, rates of opening and closing seem to follow an Arrhenius law. Indeed, the fitting of the experimental points with an exponential $k(T) = k_{\infty}exp(-E_a/RT)$ is consistent with such a law. Therefore the activation energies of opening and closing could be deduced. In a first approximation the opening rate is not affected by the length and the nature of the loop. Consequently, the opening seems to be governed by the stem only: strength of the base-pairs and stacking interactions in the double helix part. This first evidence is very important for the modelling and we will come back to this point for quantitative comparison of the experimental and theoretical results. Second, the activation energy of closing for poly(T) is not affected by the length of the loop. Nevertheless the rate of closing is lower for bigger loops according to the increase of the loop entropy. Indeed bigger loops generates a bigger phase space and the meeting of the two ends of the ssDNA take more time. This indicates that the free energy of a poly(T) loop is mostly entropic and the base stacking does not seem to be very important in this case. Nevertheless, Fig. (2.6) shows that the activation energies of closing for poly(A) and poly(T) are very different and the activation energy of poly(A) is bigger than for poly(T). So, in poly(A) there is an additional enthalpic term due to the base stacking (perhaps also due to a bigger excluded volume in poly(A)).

Figure (2.7) shows the evolution of the activation energy of closing with the loop lengths for poly(A) and poly(T). In a first approximation the author of the study consider that the enthalpy of poly(T) does not depend on the loop length (−0.1

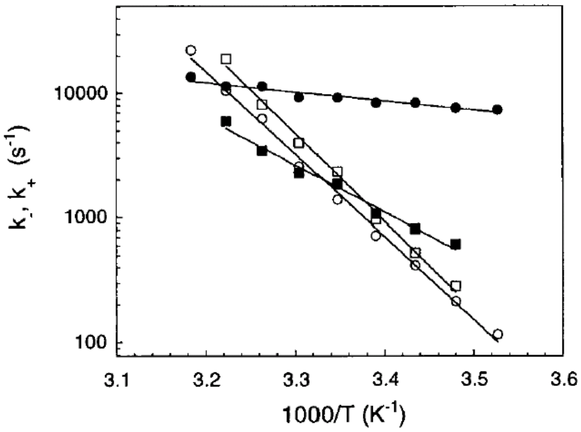

Figure 2.6: *Comparison of the opening rates (opening symbols) and the closing rates (filled symbols) for the beacons with loops of equal length but with different sequence:* $(T)_{21}$ *(circles) and* $(A)_{21}$ *(squares). The lines are exponential fits to the data [4].*

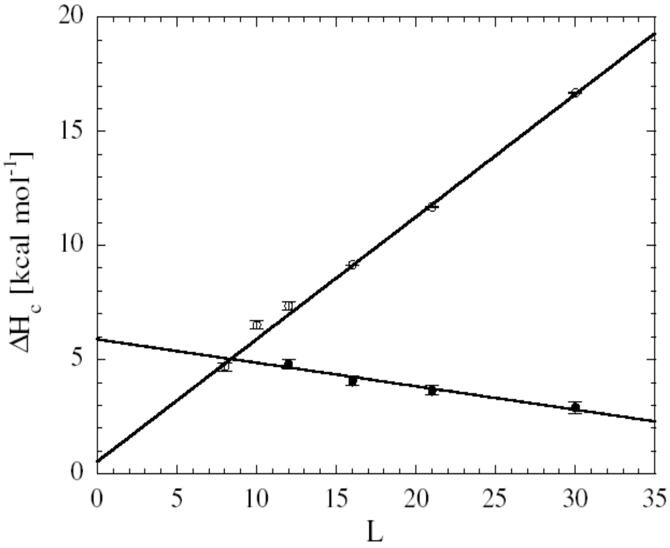

Figure 2.7: *Closing enthalpy vs loop lengths (number of bases) of (○) poly(A) and (●) poly(T)* [37].

kcal.mol^{-1}. base^{-1}) but for poly(A) ΔH_c increases with increasing loop length(+0.5 kcal.mol^{-1}.base^{-1}). This confirms two key points:

1. the loop sequence dependence of the closing properties;
2. a free energy mostly entropic for poly(T) but with an additional enthalpic term for poly(A).

According to the Libchaber's group the energetic barrier of closing comes from a distortion of the loop and a nucleation of the first base-pair in the stem while the linearity of ΔH_c with loop length in poly(A) reflects the base stacking energy in ssDNA.

All these results will help us in the design of a model for ssDNA. They give us ideas of the physical ingredients necessary to the modelling: hydrogen bonds + stacking interaction for the stem and rigidity + base stacking in the loop.

2.3 Static Absorbance measurements

Another type of measurement that can be used for hairpins is the common absorbance technique. We present briefly this technique as well as some results that can be found in the literature [49] in particular the results of Kuznetsov *et al* [6]. We also present in this section an interesting model developed by Kuznetsov *et al* which is in good agreement with absorbance results.

2.3.1 Experiment

As explained in Chap. 1, a DNA molecule is composed of nucleic acids which absorb UV light around 265 nm. This absorption depends on the composition and the structure of nucleic acids. The absorbance measurement is based on the Beer-Lambert law:

$$A = \epsilon.l.c \tag{2.7}$$

Where ϵ is the molecular absorption coefficient, l the distance of sample traversed by the UV-light and c the concentration of the system in the sample. The change of absorbance is directly proportional to the amount of substance which absorbs UV-light. Figure (2.8) gives a schematic representation of a possible experimental method to measure absorbance. For DNA the closed and open forms have very different absorption coefficients. Natural DNA, i.e. closed DNA, has a small value of ϵ while single strands, or more precisely unstacked bases, have a much higher ϵ. Therefore the opening of the stem of hairpins leads to a strong increase in absorbance. In

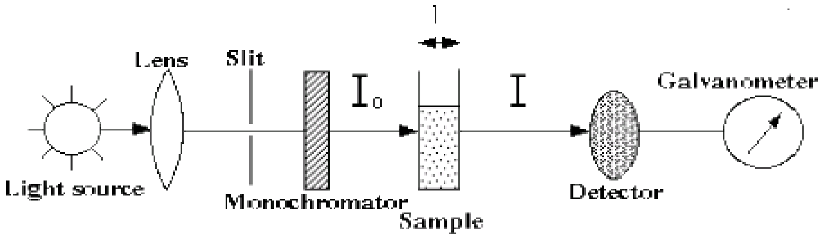

Figure 2.8: *Schematic representation of a spectrophotometer [6].*

their experiments, in order to increase the sensitivity of the detection, Kuznetsov *et al*, use a modified form of DNA. They change the base A in the base-pair A-T by 2-aminopurine (2AP), a fluorescent analog of the Adenine which absorbs at 266 nm and 330 nm. When the base-pair is formed there is no absorbance, so in the closed state a hairpin does not absorb.

2.3.2 Analysis

In order to analyse their experiments, Kuznetsov *et al* introduce a very simple model for the hairpin which has some similarities with the models that we discuss in details in the next chapter.

The model [6] is based on the simple one dimensional Ising model that we presented in Chap. 1 [23] (called also Poland and Scheraga model) but with the improvement brought by Benight and coworkers [26]: the introduction of nearest-neighbor sequence dependence in the stacking interaction. Of course this model is only valid for the stem. For the loop they used the wormlike chain model [51, 52] which we will present in more detail in the next chapter. To describe the partition function of the system they need three parameters: s_i, the statistical weight for each base-pair; σ, the cooperativity parameter and $w_{loop}(n)$, the end-loop weighting function for a loop consisting of n bases. The statistical weight corresponding to each base-pair formation, s_i, depends on the type of base-pair A-T or G-C and interactions with its neighbors, and includes the stability from hydrogen bonding as well as stacking interactions:

$$s_i = e^{-\frac{\Delta G_i}{RT}}, \tag{2.8}$$

where

$$\Delta G_i = \Delta H_i - T\Delta S_i + \frac{\delta G_{i-1,i} + \delta G_{i,i+1}}{2}. \tag{2.9}$$

ΔH and ΔS are the enthalpy and the entropy change, respectively, associated with base-pair formation. $\delta G_{i,i\pm1}$ are enthalpies associated to stacking interactions. The stacking interaction as well as base-pair formation are directly included in enthalpies and they do not deal with potential of interactions which could explain the physical origin of such phenomena. The cooperativity is associated with the junction between an intact and broken base-pair, and it depends on the specific type of base-pairs at the junction. The form of the cooperativity parameter is the following:

$$\sigma_{i,i+1} = \langle\sigma\rangle^{\frac{1}{2}}\, e^{\frac{\delta G_{i,i+1}}{2RT}}, \tag{2.10}$$

where $\langle\sigma\rangle$ is the average of the ten different stacking interactions and the value is taken according to Wartell and Benight's works [26]. The base-pair at the junction between the stem and the loop is always intact in their modelling (of course not in the coil state) therefore the end-loop weighting function $w_{loop}(n)$ is proportional to the probability of forming a loop with n bases (the end-to-end distance is therefore fixed):

$$w_{loop}(n) = \left(\frac{3}{2\pi b^2}\right)^{\frac{3}{2}} V_r g(n)\sigma_{loop}(n), \tag{2.11}$$

where n is the number of bases in the loop, $b = 2P$ is the statistical segment length (Kuhn's length), V_r is a characteristic reaction volume within which the bases at the two ends of the loop can form hydrogen bonds, $\sigma_{loop}(n)$ models the stabilizing interactions of the bases within the loop and between the loop and the stem, and finally $g(n)$ is the probability of forming a loop with n bases. Figure (2.9) gives a schematic representation of some microstates of the model and the corresponding statistical weights are given in Eq. (2.12)

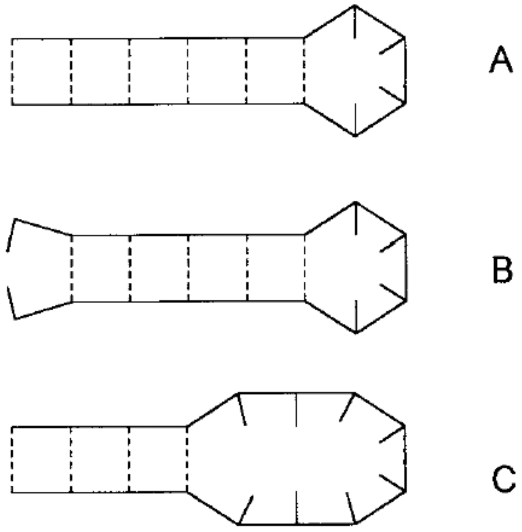

Figure 2.9: *Schematic representation of some microstates of the Kuznetsov et al model. [6]*

$$\begin{aligned} z_a &= \langle\sigma\rangle^{\frac{1}{2}} \left(\prod_{i=1}^{N_s} s_i\right) w_{loop}(N) \\ z_b &= \sigma_{1,2} \left(\prod_{i=2}^{N_s} s_i\right) w_{loop}(N) \\ z_c &= \langle\sigma\rangle^{\frac{1}{2}} \left(\prod_{i=1}^{N_s-2} s_i\right) w_{loop}(N+4). \end{aligned} \tag{2.12}$$

To fit the abosorbance measurements they derive the fraction of intact base-pairs summed over all the microstates, $\theta_I(T)$:

$$\theta_I(T) = \sum_j \frac{n_j}{N_s} \frac{z_j}{Q(T)}, \tag{2.13}$$

where $Q(T)$ is obtained by summing the statistical weights of all microstates $\{j\}$ and n_j is the number of intact base-pairs in the j^{th} microstate.
The absorbance melting profiles at 266 nm can be expressed as :

$$A(T) = \theta(T)[A_U(T) - A_L(T)] + A_L(T), \tag{2.14}$$

where $A_U(T)$ and $A_L(T)$ are the limiting baselines at high and low temperature, respectively and $\theta(T)$ is the net fraction of broken base-pairs which is calculated from Eq. (2.13) as

$$\theta(T) = 1 - \theta_I(T).$$

We only give one result that shows that, with appropriate parameters, the model is in good agreement with the experimental results. Figure (2.10) shows the melting profiles of 5'-CGGATAA(T_N)TTATCCG-3' with different value of N and the fits using the model presented below. The most important weaknesses of this model are the following:

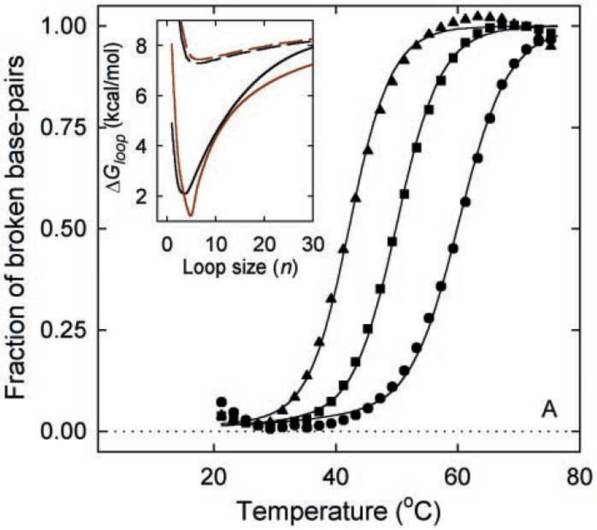

Figure 2.10: *Fits to the equilibrium melting profiles. The symbols are normalized absorbance: ●, N=4; ■, N=8; ▲, N=12; the lines are the fraction of broken base-pairs.* ΔG_{loop} *is the free energy of forming a loop closed by an A-T base-pair and is obtained by the model: red and black curve is the test of different* σ_{loop} *[6].*

1. the stem does not contain enough degrees of freedom and the end-to-end distance of the loop is fixed.

2. This model is too phenomenological. Its parameters are hard to connect with properties of DNA hairpins. The stacking is directly included in an enthalpic term and in the parameter σ.

Chapter 3

Review of some polymer and protein models

Contents

3.1 Polymer theory 33
  3.1.1 Introduction 33
  3.1.2 Freely jointed chain 34
  3.1.3 Freely rotating chain 37
  3.1.4 Kratky-Porod chain 41
  3.1.5 Growth of a polymer chain 45
3.2 Protein models 49
  3.2.1 Protein folding 49
  3.2.2 Lattice models 50

For hairpins the properties of the loop are important. In this chapter we review some polymer models [53] that could be used to describe the loop. Another aspect of our study is the formation of the hairpin, i.e. the folding of the single strand of DNA to form the stem. This process is qualitatively similar to the folding of proteins in their biologically active configuration. This is why, in this chapter, we also give a brief review of protein folding theory.

3.1 Polymer theory

3.1.1 Introduction

Since the birth of the interdisciplinary studies approximately fifty years ago, polymer theory has known a high development for its application in chemical technology as well as, of course, in biology. Indeed macromolecules play a key role in molecular biology with DNA, RNA and proteins. As one can imagine, polymers have complex properties due to their interaction both inside the molecule and with the environment, i.e. with the solvent and other identical molecules. In this chapter we will concentrate our attention on the equilibrium properties of polymers presenting three different models: the freely jointed chain, the freely rotating chain and finally the Kratky-Porod chain (or worm like chain) [54]. Dynamical properties of polymer in solution will not be considered in this thesis [53, 55] because they are not necessary for our purpose.

3.1.2 Freely jointed chain

The freely jointed chain (FJC) is the simplest model for a single polymer in solution. Each monomer occupies a point in three or two dimensional space. The conformation of the FJC is represented by the set of N+1 position vectors $\{\mathbf{R}_n\} \equiv (\mathbf{R}_0 \ldots \mathbf{R}_N)$ defining the position of the nodes in space. We can also define the bond vectors that connect together these monomers $\{\mathbf{r}_n\} \equiv (\mathbf{r}_1 \ldots \mathbf{r}_N)$, with

$$\mathbf{r}_n = \mathbf{R}_n - \mathbf{R}_{n-1}, \tag{3.1}$$

for n=1...N.

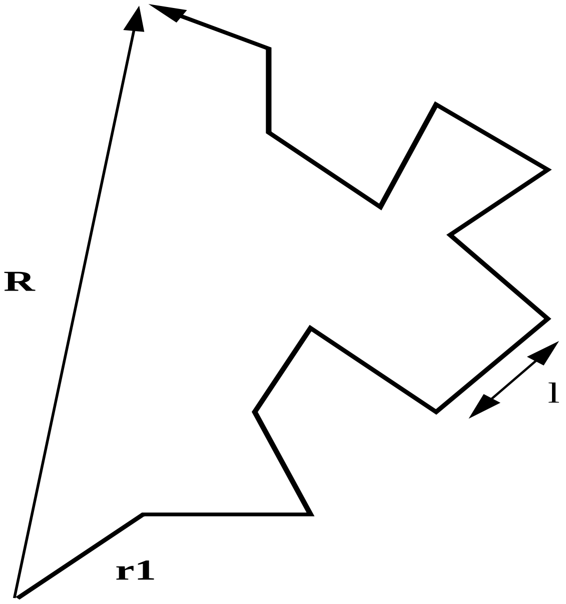

Figure 3.1: *Freely jointed chain.*

To construct a probabilistic model for the polymer, we say that the node n must be at a distance b from the node $n-1$, and each direction in space has the same probability. Therefore the distribution for the bond vector with, a constant length b, is the following:

$$\Phi(\mathbf{r}) = \frac{1}{4\pi b^2}\delta\left(|\mathbf{r}| - b\right). \tag{3.2}$$

This distribution is normalized to unity

$$\int d\mathbf{r}\Phi(\mathbf{r}) = 1. \tag{3.3}$$

Since the bond vectors $\mathbf{r}_n$ are independent of each other,

$$\Phi(\mathbf{r}_i, \mathbf{r}_j) = \Phi(\mathbf{r}_i)\Phi(\mathbf{r}_j). \tag{3.4}$$

so that the joint probability distribution can be factored into single bond vector probability distribution. For a chain of N bond vectors, the distribution function is written as

$$\Psi(\{\mathbf{r}_n\}) = \prod_{n=1}^{N} \Phi(\mathbf{r}_n). \tag{3.5}$$

Note that this is an unphysical model for a polymer since it allows two monomers to be arbitrarily close to each other: there is no "excluded volume" interaction between any two monomers. Note also that constructing the polymer chain with N bonds is equivalent to a random walk of N steps, which is the other name of this model.

3.1.2.1 End-to-end vector

We are interested in certain properties of this model. First, we want to know the properties of the end-to-end distance of the polymer.

$$\mathbf{R} = \mathbf{R}_N - \mathbf{R}_0 = \sum_{n=1}^{N} \mathbf{r}_n. \tag{3.6}$$

To define its statistical properties, we would like to know the moments of this quantity, in particular $\langle \mathbf{R} \rangle$ and $\langle \mathbf{R}^2 \rangle$. First, $\langle \mathbf{R} \rangle = \sum_{n=1}^{N} \langle \mathbf{r}_n \rangle = 0$ because

$$\langle \mathbf{r}_n \rangle = \int \mathbf{r}_n \Phi(\mathbf{r}_n) d\mathbf{r}_n = 0. \tag{3.7}$$

There is no preferred direction for any bond, so that the average is zero. Second, $\langle \mathbf{R}^2 \rangle$,

$$\begin{aligned}
\langle \mathbf{R}^2 \rangle &= \left\langle \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{r}_i \cdot \mathbf{r}_i \right\rangle \\
\langle \mathbf{R}^2 \rangle &= \sum_{i,j=1}^{N} \langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle \\
\langle \mathbf{R}^2 \rangle &= \sum_{i=1}^{N} \langle |\mathbf{r}_i|^2 \rangle + \sum_{i \neq j=1}^{N} \langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle \\
\langle \mathbf{R}^2 \rangle &= N b^2.
\end{aligned} \tag{3.8}$$

All of the cross terms vanish because the distribution of the individual bonds are statistically independent. There are N remaining terms, each of them giving a factor b^2. Also, note that this implies that $\sqrt{\langle \mathbf{R} \cdot \mathbf{R} \rangle} = \overline{R} = \sqrt{N}b$, i.e. that the root mean square end-to-end distance of a polymer grows as $\sqrt{N}$.

3.1.2.2 End-to-end vector distribution

We now consider the statistical distribution of the end-to-end vector of the FJC model. The probability distribution function $G(\mathbf{R})$ of the end-to-end vector is calculated using the distribution of the bonds:

$$G(\mathbf{R}) = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \cdots \int d\mathbf{r}_N \delta \left(\mathbf{R} - \sum_{n=1}^{N} \mathbf{r}_n \right) \Psi(\{\mathbf{r}_n\}), \tag{3.9}$$

which is rewritten using the integral representation of the delta function as

$$G(\mathbf{R}) = \frac{1}{(2\pi)3} \int \Psi(\{\mathbf{r}_n) \int \exp \left(-i\mathbf{k} \left(\mathbf{R} - \sum_{n=1}^{N} \mathbf{r}_n \right) \right) d\mathbf{k} \prod_{i=1}^{N} d\mathbf{r}_j$$

$$G(\mathbf{R}) = \frac{1}{(2\pi)3} \int e^{-i\mathbf{k}\cdot\mathbf{R}} \prod_{n=1}^{N} \left(\int \frac{1}{4\pi b^2} \delta \left(|\mathbf{r}_n| - b \right) \ e^{i\mathbf{k}\cdot\mathbf{r}_n} d\mathbf{r}_n \right) d\mathbf{k}. \tag{3.10}$$

It is possible to evaluate the integral within the parentheses for each n using polar coordinates with $\mathbf{k}$ pointing along the z direction. We get

$$\int_0^\infty \frac{1}{4\pi b^2} \delta \left(|\mathbf{r}_n| - b \right) \ e^{i\mathbf{k}\cdot\mathbf{r}_n} d\mathbf{r}_n = \frac{\sin kb}{kb}. \tag{3.11}$$

Using Eq. (3.11), the expression (3.10) becomes

$$G(\mathbf{R}) = \frac{1}{(2\pi)^3} \int e^{-i\mathbf{k}\cdot\mathbf{R}} \left(\frac{\sin kb}{kb} \right)^N d\mathbf{k}. \tag{3.12}$$

So far the calculation is exact for all N. To proceed, we need to make an approximation to evaluate the integral. We are interested in large N, since we are interested in long polymer chains. One can check that $\lim_{N\to\infty} (\sin kb/kb)^N = 0$ for all $kb > 0$. So the dominant part of the integral comes from the small values of kb. Therefore we can use the fact that

$$\frac{\sin kb}{kb} \approx 1 - \frac{(kb)^2}{3!} \approx \exp \left(-\frac{(kb)^2}{6} \right). \tag{3.13}$$

The distribution now becomes

$$G(\mathbf{R}) = \frac{1}{(2\pi)^3} \int e^{-i\mathbf{k}\cdot\mathbf{R}} e^{-\frac{k^2 b^2 N}{6}} d\mathbf{k}. \tag{3.14}$$

The integral over $\mathbf{k}$ is a standard Gaussian integral [57] which gives us

$$G(\mathbf{R}) = \left(\frac{3}{2\pi b^2 N}\right)^{\frac{3}{2}} e^{-\frac{3R^2}{2b^2 N}}. \tag{3.15}$$

We can notice that the probability distribution for the vector $\mathbf{R}$ only depends on its length R and is Gaussian. Moreover the distribution (3.15) has the unrealistic feature that $||\mathbf{R}||$ can be larger than the maximum extended length Nb of the chain which is due to the approximation made in the calculations. Finally we can express the probability distribution of the end-to-end distance R using

$$G(\mathbf{R})d\mathbf{R} = P(R)dR. \tag{3.16}$$

Therefore, replacing b by l,

$$P(R) = R^2\sqrt{\frac{2}{\pi}}\left(\frac{3}{2l^2 N}\right)^{\frac{3}{2}} e^{-\frac{3R^2}{2l^2 N}}. \tag{3.17}$$

Figure (3.2) gives a representation of $P(R)$ for different value of N and a fixed value of l=6 Å which approximately is the interbase distance in ssDNA.

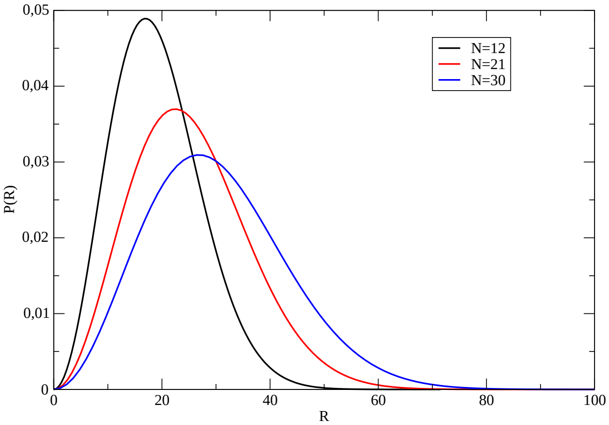

Figure 3.2: *Probability distribution of the end to end distance of a freely jointed chain.*

3.1.3 Freely rotating chain

A more realistic model to describe chains without long-range-interactions is the freely rotating chain (FRC) [56]. A drawing of a freely rotating chain is shown in Fig. (3.3). The angle θ is fixed for each segment; but each segment can freely rotate along the ϕ degree of freedom. The distribution function for the end-to-end vector

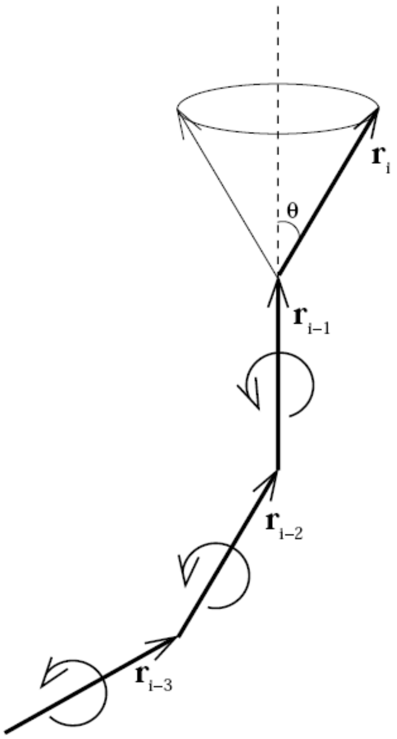

Figure 3.3: *Freely rotating chain.*

R, is not known for the discrete case but for very long chain this distribution tends to a Gaussian function. Nevertheless with numerical simulation it is quite easy to get this distribution. It is interesting to derive $\langle \mathbf{R}^2 \rangle$ of such a chain in order to introduce the notion of persistence length [54].

3.1.3.1 End-to-end vector

We can write back the expression of $\langle \mathbf{R}^2 \rangle$ as

$$\langle \mathbf{R}^2 \rangle = \sum_{i=1}^{N} \langle \mathbf{r}_i^2 \rangle + 2 \sum_{i=1}^{N} \sum_{j=1}^{N-i} \langle \mathbf{r}_i \cdot \mathbf{r}_{i+j} \rangle . \tag{3.18}$$

Thus a recursion relation is needed to calculate $\langle \mathbf{r}_i \cdot \mathbf{r}_{i+j} \rangle$. The relationship is derived by successively projecting each vector $\mathbf{r}_i$ onto the unit vector along the direction of the previous two vectors of the chain $\mathbf{r}_{i-1}$ and $\mathbf{r}_{i-2}$. Therefore

$$\mathbf{r}_i = -\cos\phi_i \mathbf{r}_{i-2} + \cos\theta \left(1 + \cos\phi_i\right) \mathbf{r}_{i-1} + \frac{\sin\phi_i}{l} \mathbf{r}_{i-2} \times \mathbf{r}_{i-1}, \tag{3.19}$$

where ϕ is the azimuthal rotation angle of the ith bond vector relative to the previous one. It follows that

$$\mathbf{r}_i \cdot \mathbf{r}_{i-2} = l^2 \left(\cos^2\theta - \sin^2\theta \cos\phi_i\right) . \tag{3.20}$$

The second term in Eq. (3.20) averages to zero (integration over the azimuthal angle). Therefore

$$\langle \mathbf{r}_i \cdot \mathbf{r}_{i-2} \rangle = l^2 \cos^2 \theta, \tag{3.21}$$

which can be generalized as

$$\begin{aligned} \langle \mathbf{r}_i \cdot \mathbf{r}_{i+j} \rangle &= (\cos\theta)^{j-1} \langle \mathbf{r}_{i+j-1} \cdot \mathbf{r}_{i+j} \rangle = l^2 (\cos\theta)^j \\ &\equiv l^2 e^{-\frac{jl}{\lambda}}, \end{aligned} \tag{3.22}$$

where $\lambda = -l/\ln\cos\theta$ is defined as the **correlation length**. Putting Eq. (3.22) into Eq. (3.18) and after some standard algebraic manipulations, we obtain

$$\left\langle \mathbf{R}^2 \right\rangle = Nl^2 \left(\frac{1+\cos\theta}{1-\cos\theta} - \frac{2\cos\theta}{N} \frac{1-(\cos\theta)^N}{(1-\cos\theta)^2} \right). \tag{3.23}$$

We clearly see that when N becomes large Eq. (3.23) simplifies into

$$\left\langle \mathbf{R}^2 \right\rangle = Nl^2 \frac{1+\cos\theta}{1-\cos\theta}, \tag{3.24}$$

which shows that, as in the case of the FJC, the end-to-end distance scales as $\sqrt{N}$. As Eq. (3.21) shows, the bonds are correlated and the chain is said to have “stiffness”. To characterize how stiff the chain is, we have to find the “memory” of the chain. Let us suppose that the first segment of the chain points in the direction $\mathbf{u}_0$. We ask, how does the end-to-end vector of the chain $\mathbf{R}$, correlate with the original orientation, $\mathbf{u}_0$? If $\mathbf{R}$ is on average along the same direction as the original, the chain is very stiff. If not, it is more flexible. Thus, it is natural to evaluate

$$\begin{aligned} \langle \mathbf{R} \cdot \mathbf{u}_0 \rangle &= \left\langle \mathbf{R} \cdot \frac{\mathbf{r}_1}{\|\mathbf{r}_1\|} \right\rangle \\ \langle \mathbf{R} \cdot \mathbf{u}_0 \rangle &= \frac{1}{l} \sum_{i=1}^{N} \langle \mathbf{r}_1 \cdot \mathbf{r}_i \rangle \\ \langle \mathbf{R} \cdot \mathbf{u}_0 \rangle &= l \sum_{i=1}^{N} (\cos\theta)^{i-1} \\ \langle \mathbf{R} \cdot \mathbf{u}_0 \rangle &= l \frac{1-(\cos\theta)^N}{1-\cos\theta}. \end{aligned} \tag{3.25}$$

In the limit of a long chain (only large N),

$$\lim_{N\to\infty} \langle \mathbf{R} \cdot \mathbf{u}_0 \rangle \equiv l_p = \frac{l}{1-\cos\theta}, \tag{3.26}$$

where l_p is called the **persistence length** of the chain. This describes the stiffness in the chain because it describes how long the orientation of the chain *persists* through its length. Clearly, the smaller θ is, the stiffer the chain will be. A θ-value of zero corresponds to a completely rigid rod [55]. It is interesting to look at the continuum limit defined by $l \to 0$, $N \to \infty$, $Nl \to L$ which is constant and $\theta \to 0$. We can write Eq (3.22) as

$$\begin{aligned}
\langle \mathbf{r}_0 \cdot \mathbf{r}_N \rangle &= l^2 (\cos\theta)^N \\
\langle \mathbf{r}_0 \cdot \mathbf{r}_N \rangle &= l^2 \exp\left(N \ln\left(\cos\theta\right)\right) \\
\langle \mathbf{r}_0 \cdot \mathbf{r}_N \rangle &= l^2 \exp\left(N\left(\cos\theta - 1 - \frac{(\cos\theta - 1)^2}{2} + \cdots\right)\right) \\
\langle \mathbf{r}_0 \cdot \mathbf{r}_N \rangle &= l^2 \exp\left(-Nl\left(\frac{(1-\cos\theta)}{l} + \frac{(1-\cos\theta)^2}{2l} + \cdots\right)\right) \\
\langle \mathbf{r}_0 \cdot \mathbf{r}_N \rangle &\approx \exp -\frac{Nl}{l_p}, \qquad (3.27)
\end{aligned}$$

which shows that the persistence length corresponds to the *correlation length* of the chain in the continuum limit approximation only.

3.1.3.2 End-to-end vector distribution

It is not possible to derive an exact expression for the end-to-end vector distribution for all $\mathbf{R}$ and all N. Nevertheless as Eq. (3.24) shows, the end-to-end distance scales with $\sqrt{N}$ for large N. Therefore we can expect, according to the central limit theorem that the probability distribution of the end-to-end vector to be Gaussian. In Ref. [54] it is shown that, in such a limit, the characteristic function, which is the Fourier transform of the probability distribution, is Gaussian:

$$K(\mathbf{k}) = \exp -\left(\frac{\mathbf{k}^2}{6} N l^2 \frac{1+\cos\theta}{1-\cos\theta}\right). \qquad (3.28)$$

Therefore the probability distribution $G(\mathbf{R})$ also is Gaussian for large N:

$$\begin{aligned}
G(\mathbf{R}) &= \frac{1}{(2\pi)^3} \int K(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{R}} d\mathbf{k} \\
G(\mathbf{R}) &= \frac{1}{8(\pi\sigma_N^2)^{\frac{3}{2}}} \exp -\frac{\mathbf{R}^2}{4\sigma_N^2}, \qquad (3.29)
\end{aligned}$$

where $\sigma_N^2 = \frac{Nl^2}{6}\frac{1+cos\theta}{1-\cos\theta}$ is the gyration radius of the polymer in such a limit. Therefore the end-to-end probability distribution is

$$P(R) = 4\pi R^2 G(\mathbf{R}) = \frac{1}{2\sqrt{\pi}} \frac{1}{\sigma_N} \left(\frac{R}{\sigma_N}\right)^2 e^{-\frac{R^2}{4\sigma_N^2}}. \qquad (3.30)$$

In practice we have to know when the approximation of large N is valid. For that we have compared the real probability of the FRC simulated numerically and the

Gaussian approximation. Figure (3.4) gives the comparison for two different values of the polar angle and for different values of the number of monomers. The length of one monomer is fixed to 6 Å, which is the appropriate value for a DNA strand.

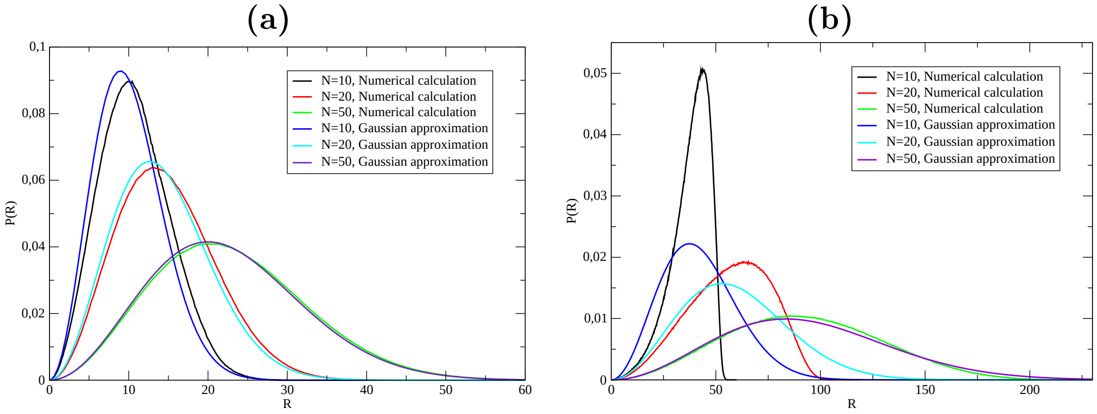

Figure 3.4: *Probability distribution of the Freely Rotating Chain for two values of* θ, *(a):* θ=*120°; (b):* θ=*45° and comparison with the Gaussian approximation. The length of one monomer is fixed to 6 Å.*

First of all, $P(R)$ is not Gaussian for all N and for all θ. Indeed for a small value of θ and N=10-20, the Gaussian approximation is not correct because the Gaussian approximation allows R to be larger than Nl and it is physically not possible. Nevertheless for bigger values of N like 50 the Gaussian approximation is better and in these conditions we can use such an approximation.

Second, for a large value of θ, the limit of large N is rapidly reached. Indeed for N=10 the probability distribution is approximately Gaussian and the greater N, the best is the Gaussian approximation. Therefore the validity of the large N limit depends on θ. If θ is large, the limit is reached rapidly but if θ is small, bigger values of N are needed.

We now understand why it is very difficult to derive an exact expression of the end-to-end distance probability distribution for all N.

3.1.4 Kratky-Porod chain

3.1.4.1 An exact calculation of $P_N(\mathbf{r})$

We consider the chain described by the Hamiltonian

$$H = -\epsilon \sum_{j=1}^{N-1} \left(\mathbf{r}_j \cdot \mathbf{r}_{j+1} - l^2 \right) , \tag{3.31}$$

where l is the length of the segment. If we define $\mathbf{X}_j = \mathbf{r}_j/l$, which is a unit vector

$$H = \epsilon l^2 \sum_{j=1}^{N-1} \left(\mathbf{X}_j \cdot \mathbf{X}_{j+1} - 1\right). \tag{3.32}$$

The partition function of the chain is given by

$$Z_N = \int d\Omega_1 ... d\Omega_N \prod_{j=1}^{N-1} e^{b(\mathbf{X}_j \cdot \mathbf{X}_{j+1} - 1)}, \tag{3.33}$$

with $b = \epsilon l^2/k_B T$ and Ω_j is the solid angle variation associated with a change of orientation of vector $\mathbf{X}_j$. This system is formally analogous to a one-dimensional Heisenberg chain in zero field studied in [58]. Using polar coordinates, θ_{j+1}, ϕ_{j+1} referred to $\mathbf{X}_j$ as the polar axis, the integrals separate yielding

$$\begin{aligned}
Z_N &= \int d\Omega_1 \left[\prod_{j=1}^{N-1} \int_{\theta_{j+1}=0}^{\pi} \int_{\phi_{j+1}=0}^{2\pi} e^{b\cos\theta_{j+1}} \sin\theta_{j+1} d\theta_{j+1} d\phi_{j+1}\right] e^{-b(N-1)} \\
Z_N &= 4\pi \left[2\pi \frac{e^b - e^{-b}}{b}\right]^{N-1} \\
Z_N &= (4\pi)^N \left[\frac{e^{-b}\sinh b}{b}\right]^{N-1}.
\end{aligned} \tag{3.34}$$

Or if we introduce the modified Bessel function of zeroth order $i_0(b) = \sinh b/b$, $Z_N = (4\pi)^N \left[e^{-b} i_0(b)\right]^{N-1}$.

A similar approach can be used to compute the correlation functions which give us the persistence length.

$$C_k = \langle \mathbf{X}_j \cdot \mathbf{X}_{j+k} \rangle = \langle \mathbf{X}_1 \cdot \mathbf{X}_{k+1} \rangle, \tag{3.35}$$

by setting $j = 1$ without loss of generality

$$\begin{aligned}
C_k = \frac{1}{Z_N} \int d\Omega_1 \mathbf{X}_1 \int d\Omega_2 e^{-b\mathbf{X}_1 \cdot \mathbf{X}_2} ... \int d\Omega_k e^{-b\mathbf{X}_{k-1} \cdot \mathbf{X}_k} \times \\
\int d\Omega_{k+1} \mathbf{X}_{k+1} e^{-b\mathbf{X}_k \cdot \mathbf{X}_{k+1}} \int d\Omega_{k+2} e^{-b\mathbf{X}_{k+1} \cdot \mathbf{X}_{k+2}} \times ... \times \\
\int d\Omega_{N-1} e^{-b\mathbf{X}_{N-1} \cdot \mathbf{X}_N} \times e^{-(N-1)b}.
\end{aligned} \tag{3.36}$$

The integrals over $\Omega_{k+2} ... \Omega_{N-1}$ simplify with the corresponding integrals in Z_N. Moreover we can use the relation for unit vectors

$$\int d\Omega_{j+1} \mathbf{X}_{j+1} e^{-b\mathbf{X}_j \cdot \mathbf{X}_{j+1}} = 4\pi i_1(b) \mathbf{X}_j, \tag{3.37}$$

where

$$i_1(b) = \frac{b\cosh b - \sinh b}{b^2}, \tag{3.38}$$

which can again be obtained by direct integration in polar angles [57].
This allows us to get an expression of C_k by integrations which involve successively $\mathbf{X}_{k+1}$, $\mathbf{X}_k$, ...$\mathbf{X}_1$. Each one gives a factor $i_1(b)$.
The result is

$$C_k = \langle \mathbf{X}_1 \cdot \mathbf{X}_{k+1} \rangle = \left(\frac{i_1(b)}{i_0(b)} \right)^k . \tag{3.39}$$

Using the definition of the persistence length

$$C_k = \langle \mathbf{X}_1 \cdot \mathbf{X}_{k+1} \rangle = e^{-kl/l_p}, \tag{3.40}$$

we obtain the persistence length as

$$\frac{l}{l_p} = -\ln \left[\frac{i_1(b)}{i_0(b)} \right] = -\ln \left(\coth b - \frac{1}{b} \right) . \tag{3.41}$$

It is interesting to notice that, in the limit of large b (ϵ large or low temperature T) we get

$$l_p = \frac{l}{\ln\left[\coth b - \frac{1}{b}\right]} \approx lb = l \times \frac{\epsilon l^2}{k_B T}, \tag{3.42}$$

which is the result obtained with the worm like chain model [51], i.e. the continuum limit of the Kratky-Porod chain.
As explained in Chap. 5 to model the statistical physics of DNA hairpins, we need the probability distribution function of the polymer $P_N(\mathbf{R})$, which makes up the hairpin. For the Kratky-Porod chain its calculation is much more complex than for a Gaussian chain. Even in the continuum limit (WLC model) the exact expression is not known. An approximate expression has been obtained by Wilhem and Frey [59].
It reads

$$P_N(R) = 4\pi R^2 \frac{1}{4\pi R^2} \frac{\kappa}{2\sqrt{\pi}} \sum_{n=1}^{\infty} \frac{1}{\kappa\,(1 - R/L)^{3/2}} \exp\left[-\frac{(n - 1/2)^2}{\kappa\,(1 - R/L)} \right] \times H_2\left(\frac{n - 1/2}{\sqrt{\kappa\,(1 - R/L)}} \right), \tag{3.43}$$

where $L = Nl$ is the total length of the polymer, $\kappa = \epsilon l^3 / k_B T L$ is the rigidity coefficient of the WLC.
In the case of the discrete Kratky-Porod chain the calculation is even harder and the probability distribution $P_N(R)$ is not known analytically. However a computationally efficient method for its accurate numerical calculation has recently been

proposed by N. Theodorakopoulos [60]. As we use this method in our numerical calculations, we give the calculation in Appendix A. The Fourier transform of $P_N(\mathbf{R})$ is expressed as a matrix element of the N^{th} product of a matrix F as

$$P_N(\mathbf{q}) = \left(F^N\right)_{00}, \tag{3.44}$$

where the elements F_{ll} of the semi-infinite matrix F are expressed as a finite sum of Bessel functions. (See Appendix A for their expression).

In practice the size of the matrix F has to be truncated to a finite l_{max}. For a semi-flexible chain $L >> l_p$ (for instance $N = 11$ segments and a persistence length of 2 segments) l_{max}=2 or 3 produces results which can hardly be distinguished from the exact results produced by Monte Carlo simulations. For rigid chains $L/l_p = O(1)$, for instance for $N = 10$ and a persistence length of 5 segments, $l_{max} = 4$ is necessary to get a good agreement with Monte Carlo simulations. These small values of l_{max} provide a rather efficient numerical method to compute $P_N(\mathbf{R})$ for the Kratky-Porod chain.

3.1.4.2 Effective Gaussian approach

In spite of its efficiency and the moderate values of l_{max} which are required, the calculation of $P_N(\mathbf{R})$ for a Kratky-Porod chain may become quite long when we want to scan a large number of temperatures to obtain a curve for the opening as a function of temperature. This is why it is useful to have a faster approximate calculation.

One possibility is to use an effective Gaussian approximation which has a double interest

1. it is faster than the complete Kratky-Porod calculation;

2. for Gaussian chain we know an exact expression for the conditional probability function $S(r|R)$ which enters into our hairpin calculation (the S function is presented in the next section).

The idea is to approximate $P_N(R)$ by the expression for a Gaussian chain that would lead to the persistence length that we calculated for the Kratky-Porod chain Eq. (3.42). This is can be done with

$$P_N^G(R) = \frac{1}{2\sqrt{\pi}}\frac{1}{\sigma_N}\left(\frac{R}{\sigma_N}\right)^2 e^{-R^2/4\sigma_N^2}, \tag{3.45}$$

with $\sigma_N = \frac{N}{6}\chi l^2$ and $\chi = \frac{1+\coth b-1/b}{1-\coth b+1/b}$. The corresponding conditional probability is given by Eq. (3.58) which exactly verifies Eq. (3.50).

Figure (3.5) compares the effective Gaussian approximation to the Kratky-Porod expression. In the case L/l_p=5.9 the effective Gaussian approximation is rough (but

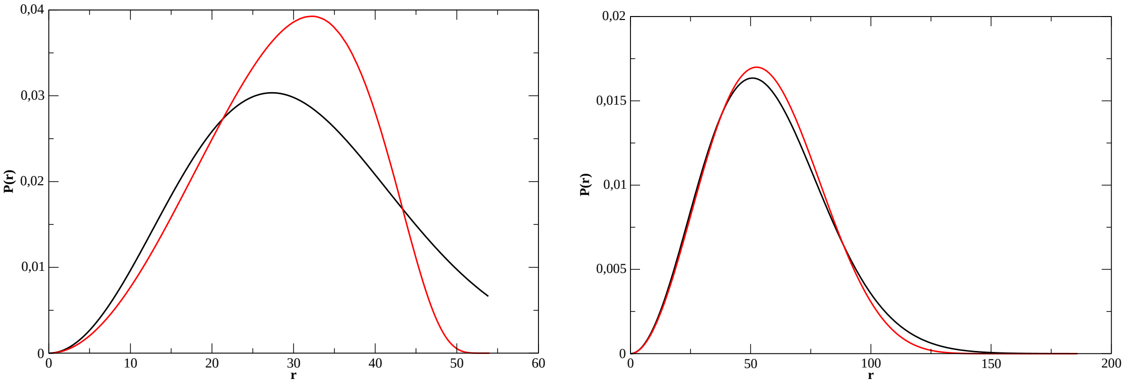

Figure 3.5: *Comparison of the effective Gaussian probability distribution function and the exact expression for N=10 and N=32. The parameters are T=300 K and ϵ=0.0015 eV.Å^{-2}.The black curve corresponds to the effective Gaussian function. Left:N= 10 and right:N= 32*

nevertheless better than the WLC expression of Wilhem and Frey), but for L/l_p=19 one can notice that the effective Gaussian approximation becomes very good. Therefore, in our hairpin calculation for small values of N we use the full discrete KP distribution and for higher values of N we use the effective Gaussian approximation. Moreover in the case of the Kratky-Porod chain, in any case for our hairpin calculation we have to use for $S(r|R)$ the Gaussian form.
In order to determine to what extend this approximation modifies the denaturation curves for hairpins (the calculation of such curves is given in Chap. 5) we have compared such curves for the two expressions $P_N^G(R)$ and $P_N^{KP}(R)$ as shown in Fig. (3.6). The difference between the two models for the loop are only perceptible for the shortest and fairly rigid loops ($N = 12$, ϵ=0.0022 eV.Å^{-2} giving l_p=15.4 Å or L/l_p=4.66). For larger loops ($N = 24$, i.e. L/l_p=9.32) the denaturation curves computed with $P_N^G(R)$ or $P_N^{KP}(R)$ can hardly be distinguished.

3.1.5 Growth of a polymer chain

Let us consider an effective Gaussian chain with a given number of monomers N, and an end-to-end distance vector $\mathbf{R}$. Its end-to-end distance probability distribution is given by Eq. (3.30). We introduce at this stage a new variable defined as

$$\sigma_N = \frac{Nl^2}{6}\chi. \tag{3.46}$$

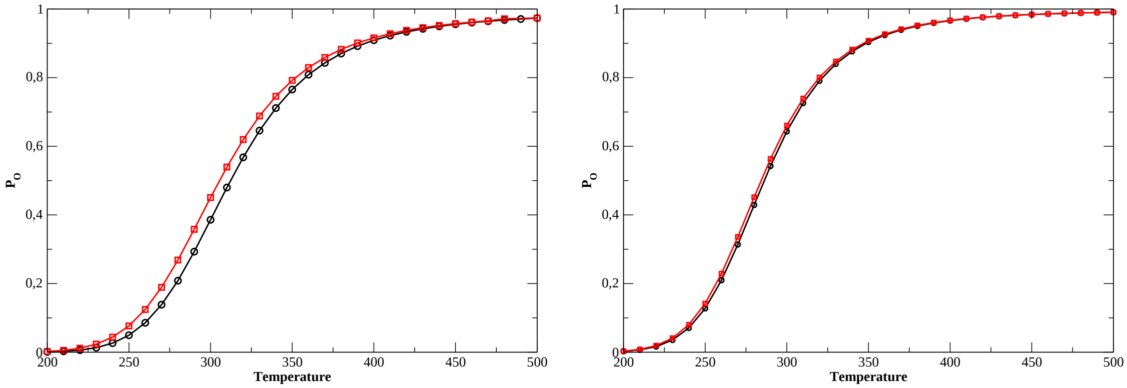

Figure 3.6: *Comparison between the melting curves obtained with the effective Gaussian and the exact expression of the probability distribution function for N=12 and N=24. The parameters are (see chapter 5 for their signification) D=0.090 eV, k=0.025 eV.Å^{-2}, α=6.9 Å^{-1}, δ=0.35, ρ=2.0 and ϵ=0.0022 eV.Å^{-2}. The black curve corresponds to the calculation with the Effective Gaussian. Left: N=12 and right: N=24.*

We immediately see that

$$\begin{aligned} \chi &= 1 && \text{(FJC)} \\ \chi &= \tfrac{1+\cos\theta}{1-\cos\theta} && \text{(FRC)} \\ \chi &= \tfrac{1+\coth b-1/b}{1-\coth b+1/b} && \text{(KP)}, \end{aligned} \tag{3.47}$$

if we use an approximate description for the FRC and the KP model. Suppose that the chain grows by the addition of one monomer at each end. Let the additional segments at the two ends be represented by the vectors $\mathbf{\Delta}_1$, $\mathbf{\Delta}_2$, respectively. The new end-to-end distance vector would then be $\mathbf{r} = \mathbf{R} + \mathbf{\Delta}_1 - \mathbf{\Delta}_2$. The unnormalized probability for the growth at each end by a vector $\mathbf{\Delta}_i$ will be proportional to

$$e^{-\frac{3|\mathbf{\Delta}_i|^2}{2\chi l^2}}. \tag{3.48}$$

We would like to derive the function $S(r|R)$ such as $S(r|R)dr$ is the conditional probability that, if the end-to-end distance of the polymer chain of N monomers is equal to R, the end-to-end distance of a chain of $N+2$ monomers, i.e. where one monomer have been added at each end, will be in the range $(r, r+dr)$. It is normalized to unity

$$\int_0^\infty dr S(r|R) = 1 \quad \forall R. \tag{3.49}$$

Furthermore, it satisfies

$$\int_0^\infty dR P_N(R) S(r|R) = P_{N+2}(r) \quad \forall r, N, \tag{3.50}$$

by definition. We shall see in Chap. 5 that this conditional probability is useful to calculate the partition function of a DNA hairpin.
The function $S(r|R)$ is defined by

$$S(r|R) = Ar^2 \int d\Omega_r \int d\boldsymbol{\Delta}_1 \int d\boldsymbol{\Delta}_2 e^{-\frac{\Delta_1^2+\Delta_2^2}{\tau^2}} \delta\left(\mathbf{r}-\mathbf{R}-\boldsymbol{\Delta}_1+\boldsymbol{\Delta}_2\right), \tag{3.51}$$

where $\tau = \frac{2\chi l^2}{3}$ and A is a normalization factor. The first integral is over all orientations of the vector $\mathbf{r}$, and the other two are meant over all space. The normalization constant will be specified at the end of the calculation. The r^2 factor appears because we only want the norm of $\mathbf{r}$ to fall in the specific range. The integral over $\boldsymbol{\Delta}_2$ can be done trivially. Abbreviating $\mathbf{r}-\mathbf{R}=\rho$, we obtain

$$S(r|R) = Ar^2 \int d\Omega_r \int_0^\infty d\Delta_1 \Delta_1^2 \int_{-1}^{1} d\mu e^{-\frac{\Delta_1^2}{\tau^2}} e^{-\frac{\Delta_1^2+\rho^2-2\rho\Delta_1\mu}{\tau^2}}, \tag{3.52}$$

where

$$\mu = \frac{\rho.\boldsymbol{\Delta}_1}{\rho\Delta}. \tag{3.53}$$

We are omitting a 2π factor from integration over the azimuthal angle of $\boldsymbol{\Delta}_1$ because this only changes the normalization. Performing the $d\mu$ integration, we get

$$S(r|R) = Ar^2 \int d\Omega_r \frac{1}{\rho} e^{-\frac{\rho^2}{\tau^2}} \int_0^\infty d\Delta_1 \Delta_1 e^{-\frac{2\Delta_1^2}{\tau^2}} \sinh\left(\frac{2\rho\Delta}{\tau^2}\right), \tag{3.54}$$

where we have again omitted constant factors to be fixed by normalization.
Using the definite integral

$$J(a,b) = \int_0^\infty dx\ x e^{-ax^2} \sinh bx = \frac{b}{2a}\left(\frac{\pi^{\frac{1}{2}}}{a}\right) e^{\frac{b^2}{4a}}, \tag{3.55}$$

we can do the integration over Δ_2. Reintroducing $\rho = \mathbf{r}-\mathbf{R}$

$$S(r|R) = Ar^2 \int d\xi e^{-\frac{r^2+R^2-2rR\xi}{2\tau^2}}, \tag{3.56}$$

where now

$$\xi = \frac{\mathbf{r}.\mathbf{R}}{rR}. \tag{3.57}$$

Finally, performing the integration over $d\xi$, and using Eq. (3.49) that fixes the constant A, we get

$$S(r|R) = \left(\frac{3}{\pi\chi l^2}\right)^{\frac{1}{2}} \frac{r}{R} \sinh\left(\frac{3rR}{2\chi l^2}\right) e^{-\frac{3}{4}\frac{r^2+R^2}{\chi l^2}}. \tag{3.58}$$

One can show, that the function $S(r|R)$ satisfies Eq. (3.50) with $P_N(R)$ given by Eq. (3.30) but it is slightly tedious. This equation assumes that $P_N(R)$ is Gaussian.

As we discussed above it is not always the case. Since we intend to use the conditional probability $S(r|R)$ in our hairpin calculations, it is useful to examine the error that it introduces when it is applied to a polymer which is not Gaussian such as the FRC or the KP chain. Let us compare $P_{N+2}(r)$ given by the exact polymer model and its value obtained with (3.50) where $P_N(R)$ is also described by the exact polymer model (FRC of KP). As we can see, for small values of N, the calculation of

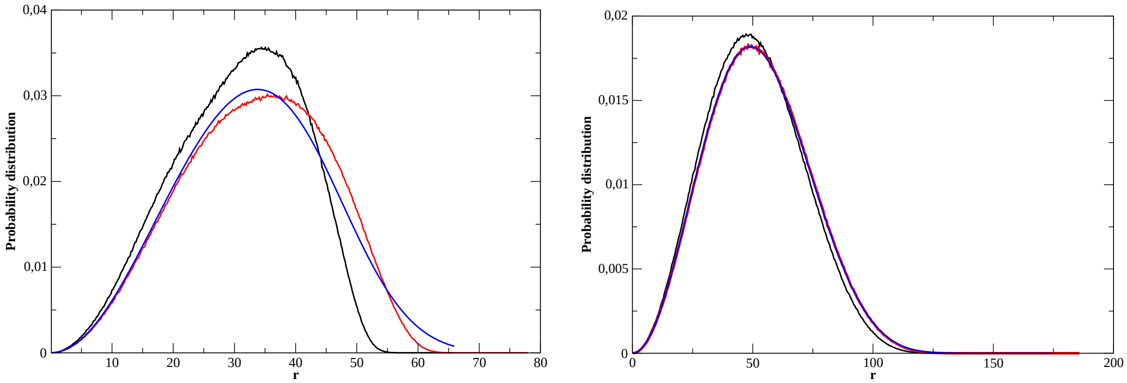

Figure 3.7: *Comparison of $P_{N+2}(r)$ obtained using Eq. (3.50) and the real form with the FRC. The length of one monomer is fixed to 6 Å, and θ=45°. The black curve represents $P_N(r)$, the red curve is for the exact $P_{N+2}(r)$ and the blue one is obtained using Eq. (3.50). Left: N=12 and right: N=30.*

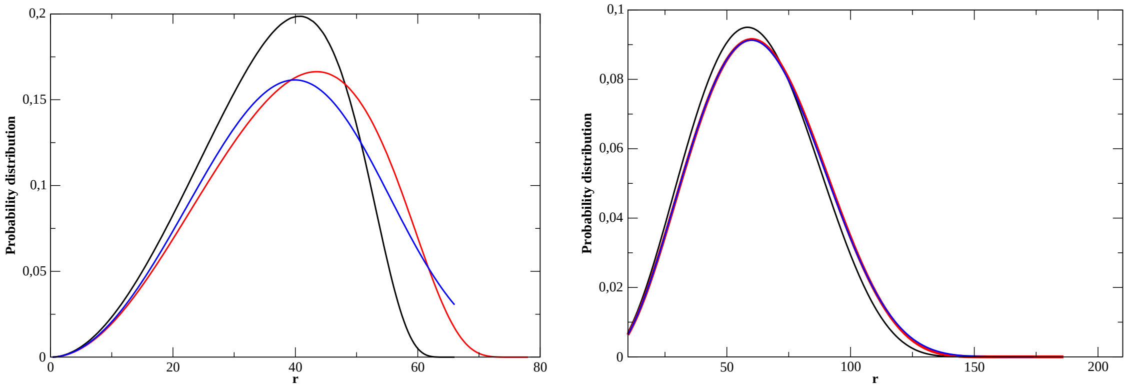

Figure 3.8: *Comparison of $P_{N+2}(r)$ obtained using Eq. (3.50) and the real form with the KP chain. The length of one monomer is fixed to 6 Åand ϵ=0.0020 eV.Å^{-2}. The black curve represents $P_N(r)$, the red curve is for the exact $P_{N+2}(r)$ and the blue one is obtained using Eq. (3.50). Left: $N = 12$ and Right: $N = 30$.*

$P_{N+2}(r)$ using Eq. (3.50) is not correct because $P_N(r)$ is not Gaussian. Nevertheless for $N = 30$ the “growth” of the polymer is correctly reproduced by the S function. In a more general way, we can say that better the Gaussian approximation for $P_N(r)$, the better the result obtained by Eq. (3.50), which is of course natural since (3.50) is exact in the Gaussian case.

3.2 Protein models

3.2.1 Protein folding

The formation of a DNA hairpin from a single strand of DNA is qualitatively similar to the folding of the amino-acid chain of a protein. The particular amino-acid sequence (or "primary structure") of a protein predisposes it to fold into its native conformation or conformations [61]. Many proteins do so spontaneously during or after their synthesis inside cells. While these macromolecules may be seen as "folding themselves," in fact their folding depends a great deal on the characteristics of their surrounding solution, including the identity of the primary solvent (either water or lipid inside cells), the concentration of salts, the temperature, and molecular chaperones. For the most part, scientists have been able to study many identical molecules folding together. It appears that in transitioning to the native state, a given amino acid sequence always takes roughly the same route and proceeds through roughly the same number of fundamental intermediates.

The essential fact of folding, however, remains that the amino acid sequence of each protein contains the information that specifies both the native structure and the pathway to attain that state: folding is a spontaneous process. The passage of the folded state is mainly guided by Van der Waals forces and entropic contributions to the Gibbs free energy: an increase in entropy is achieved by moving the hydrophobic parts of the protein inwards, and the hydrophilic ones outwards [62]. During the folding process, the number of hydrogen bonds does not change appreciably, because for every internal hydrogen bond in the protein, a hydrogen bond of the unfolded protein with the aqueous medium has to be broken.

The entire duration of the folding process varies dramatically depending on the protein of interest. The slowest folding proteins require many minutes or hours to fold, primarily due to steric hindrances. However, small proteins, with lengths of a hundred or so amino acids, typically fold on time scales of milliseconds. The very fastest known protein folding reactions are complete within a few microseconds. The Levinthal paradox, proposed by Levinthal in 1969 [21], states that, if a protein were to fold by sequentially sampling all possible conformations, it would take an astronomical amount of time to do so, even if the conformations were sampled at a rapid rate (on the nanosecond or picosecond scale). Based upon the observation that proteins fold much faster than this, Levinthal then proposed that a random conformational search does not occur in folding, and the protein must, therefore, fold by following a pre-determined path.

Folding and unfolding rates also depend on environment conditions like temperature, solvent viscosity, pH and more. The folding process can also be slowed down (and the unfolding sped up) by applying mechanical forces, as revealed by single-molecule experiments.

The study of protein folding has been greatly advanced, in recent years by the development of fast, time-resolved techniques [63]. These are experimental methods for rapidly triggering the folding of a sample of unfolded protein, and then observing the resulting dynamics. Fast techniques in widespread use include ultrafast mixing of solutions, photochemical methods, and laser temperature jump spectroscopy. For DNA hairpins the formation of the hairpin is similar to the folding, but, thanks to the use of FRET we have seen that the kinetics can be measured.

The protein folding phenomenon was largely an experimental endeavor until the groundbreaking formulation of the Energy Landscape theory by Bryngelson and Wolynes in the late 1980's [64]. The theory introduced the principle of minimal frustration, which asserts that evolutionary selection has designed the amino acid sequences of natural proteins so that interactions between side chains largely favor the molecule's acquisition of the folded state. Interactions that do not favor folding are selected against, although some residual frustration is expected to exist. A consequence of these evolutionarily designed sequences is that proteins are generally thought to have globally "funneled energy landscapes" (coined by Onuchic) that are largely directed towards the native state. This "folding funnel" landscape allows the protein to fold to the native state through any of a large number of pathways and intermediates, rather than being restricted to a single mechanism. The theory is supported by computational simulations [67], [68] of model proteins and has been used to improve methods for protein structure prediction and design. Ab initio techniques for computational protein structure prediction employ simulations of protein folding to determine the protein's final folded shape.

3.2.2 Lattice models

Lattice proteins are highly simplified computer models of proteins [66], [69] which are used to investigate protein folding. Because proteins are such large molecules, containing hundreds or thousands of atoms, it is not possible with current technology to simulate more than a few microseconds of their behaviour in all-atom detail. Hence real proteins cannot be folded on a computer. Lattice proteins [65], however, are simplified in two ways: the amino acids are modelled as single "beads" rather than modelling every atom, and the beads are restricted to a rigid (usually cubic) lattice. This simplification means they can fold to their energy minima in a time quick enough to be simulated. Lattice proteins are made to resemble real proteins by introducing an energy function, a set of conditions which specify the energy of interaction between neighbouring beads, usually taken to be those occupying adjacent lattice sites. The energy function mimics the interactions between amino acids in real proteins, which include steric, hydrophobic and hydrogen bonding effects. The beads are divided into types, and the energy function specifies the interactions depending on the bead type, just as different types of amino acid interact differently. Lattice protein models were studied in the last seventies to gain a deeper understanding of the Levinthal paradox. The main advantage of lattice models over more detailed ones is that in many cases their whole conformational space can be examined. However, even for such simple models the number of possible conformations is growing very quickly as the size of the polymer increases. For example, on the square lattice, a 18-mer has 5808335 different conformations unrelated by symmetries. Simply enumerating them is tricky in the above case, while in the 49-mer case it is out of reach (there are $\approx 10^{20}$ of them). However as shown by Go [70] and his collaborators, starting form a random conformation, the 49-mer can reach its ground state, that, is its lowest energy configuration, within a few thousands steps of a Monte Carlo simulation, as long as the energy surface is defined as follows. First, the lowest energy, compact 7x7 conformation, is chosen *a priori*. Figure (3.9) gives a schematic representation of the compact conformation of the 49-mer on the square lattice. Then, for all pairs of monomers which are close neighbours in this configurations, the contact energy is assumed to be attractive, while for all others it is not. So, when the ground-state is at the bottom of a deep funnel on the energy surface, then it is quite easy for a flexible polymer to find its way and reach it trough a random search biased by the average energy gradient. However, even if the funnel picture is nowadays the preferred view for understanding the folding process, there is no indication that protein energy surfaces are as funneled and as deep as in the Go model.

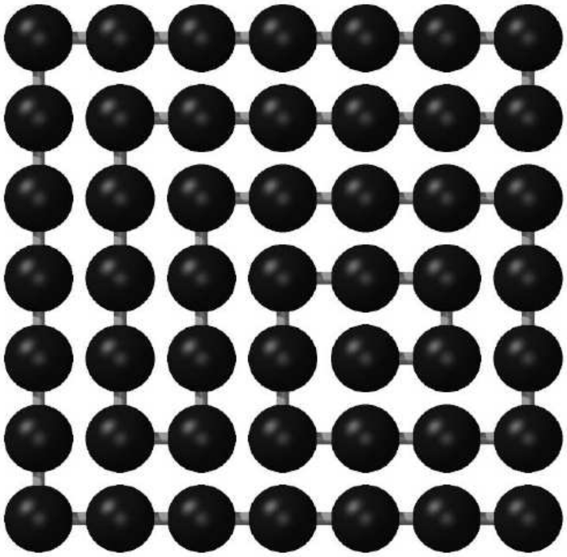

Figure 3.9: *A compact conformation of the 49-mer on the square lattice [21].*

Another popular lattice models, the HP model, features just two bead types - hydrophobic (H) and polar (P) - and mimics the hydrophobic effect by specifying a negative (favourable) interaction between H beads [21]. For any sequence in any particular structure, an energy can be rapidly calculated from the energy function. For the simple HP model, this is simply an enumeration of all the contacts between H residues that are adjacent in the structure, but not in the chain.

Most researchers consider a lattice protein sequence protein-like only if it possesses a single structure with an energetic state lower than in any other structure. This is the energetic ground state, or native state. The relative positions of the beads in the native state constitute the lattice protein's tertiary structure. By varying the energy function and the bead sequence of the chain (the primary structure), effects on the native state structure and the kinetics (rate) of folding can be explored, and this may provide insights into the folding of real proteins. In particular, lattice models have been used to investigate the energy landscapes of proteins, i.e. the variation of their internal free energy as a function of conformation.

Part II

Modelling DNA hairpins

Chapter 4

A two dimensional lattice model

Contents

4.1 Self assembly of DNA hairpins 55
  4.1.1 Model 55
  4.1.2 Metropolis-Monte Carlo scheme 57
4.2 Equilibrium properties of the opening-closing transition 59
  4.2.1 The transition in the absence of mismatch 59
  4.2.2 Role of the mismatches 62
4.3 Kinetics of the opening and closing 63

4.1 Self assembly of DNA hairpins

4.1.1 Model

As we explained in Chap. 2, a fluorophore and a quencher can be used to monitor the two limiting conformations of ssDNA. We propose here a very simple model which allows us to describe such an equilibrium. Our hairpin model is inspired by the lattice models which have been used to study protein folding [65]. It is a lattice model so that only discrete motions are allowed, thus it cannot describe the true dynamics of the hairpin. Instead we use a Monte-Carlo dynamics where the moves are discrete and determined by their probability at the temperature of the simulation, depending on their energy cost or gain. To carry such a calculation we only have to specify the energy of the model in each configuration. As a first approach to this problem we decided to choose the simplest underlying lattice, a planar square lattice. This choice of model restricts the number of accessible states with respect to a more complex three-dimensional lattice, but, as discussed below, it introduces some limitations on the ability of the model to describe actual hairpins. The energy of the DNA strand is assumed to depend on two terms only, a bending

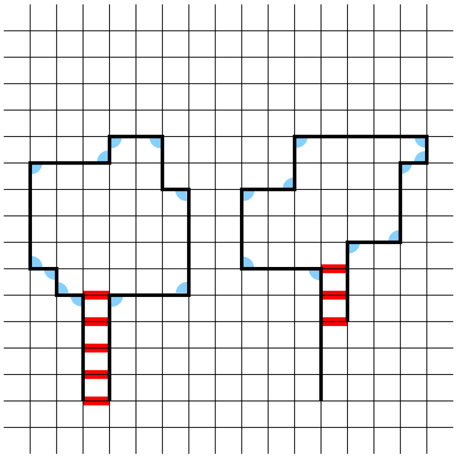

Figure 4.1: *Two configurations of the hairpin model in a lattice. The DNA strand is indicated by the thick line on the lattice. The hydrogen bonds are marked by the thick bonds connecting two points of the stand, and the shaded corners represent the bending energy contributions. The left case corresponds to the perfect closing, while the right figure shows an example of a mismatched partial closing.*

energy which appears when two consecutive segments are at some angle, and the energy of the base-pairs which can form in the stem. The total number of nucleotides in the DNA strand is denoted by N. The number of nucleotides which can form the stem is denoted by n_s. In order to specify the kind of pairing allowed in the stem, each nucleotide of the stem, denoted by index j is affected of a “type” t_j. Only two nucleotides having the same “type” are allowed to form a base-pair by hydrogen bonding. Thus, rather that actually specifying the type of a base (A, T, G, C) we specify the type of pairing that it can form. The energy of the model is written as

$$E = n_A E_A + \frac{1}{2} \sum_{j=1}^{n_s} \sum_{j'=1}^{n_s} e(j, j') \tag{4.1}$$

$$e(j, j') = \delta(t_j - t_{j'}) \delta(d_{jj'} - 1) a(j) a(j') E_{\mathrm{HB}}(t_j), \tag{4.2}$$

where

- n_A is the number of angles in the DNA strand on the lattice, and E_A is a positive model parameter giving the energy costs of a bent. In some calculations, E_A may be different for a bent in the stem or in the loop.
- $e(j, j')$ is the pairing energy between nucleotides j and j' of the stem. The factor $\delta(t_j - t_{j'})$ enforces the condition that the two nucleotides should be of the same “type”, $\delta(d_{jj'} - 1)$ indicates that the pairing is only possible if the two nucleotides are adjacent on the lattice. The factors $a(j)$ and $a(j')$ are equal to 1 only if the nucleotide is available for pairing, i.e. if it is not already involved in another pair. Otherwise the pairing is not formed and they are

set to 0. They are necessary because some geometries of the chain could put a nucleotide in a position adjacent to two sites occupied by nucleotides of the same type. Finally $E_{\mathrm{HB}}(t_j)$ is the pairing energy for nucleotides of type t_j. It is a negative quantity, which means that the pairing is favourable because it lowers the energy of the hairpin.

We studied this model using Monte Carlo simulations in the same spirit as the studies performed on lattice models of proteins, i.e. we generate a random walk of the DNA chain on the lattice with the condition that the system should be in thermal equilibrium at temperature T. A configuration of energy E must therefore have a probability proportional to $\exp(-E/T)$, where T is measured in units of energy. If the moves are selected in order to stay as close as possible to the actual motion of a polymer in a fluid, the method can even be used to study dynamical effects with a fictitious time scale which is simply given by the number of Monte Carlo steps [72]. For this reason we selected only local motions of the chain. On the two-dimensional square lattice, there are only three such motions: the change of the angle between the two segments at one end of the chain, the flipping of a corner of a lattice cell with respect to the diagonal of the cell and a crank mechanism. Figure (4.2) gives a representation of these displacements. If it does not lead to a clash with another part of the chain, an attempted motion is accepted with probability $P = \min[\exp(-\Delta E/T), 1]$, where $\Delta E = E_2 - E_1$ is the difference between the energy after and before the move, using a standard Metropolis algorithm.

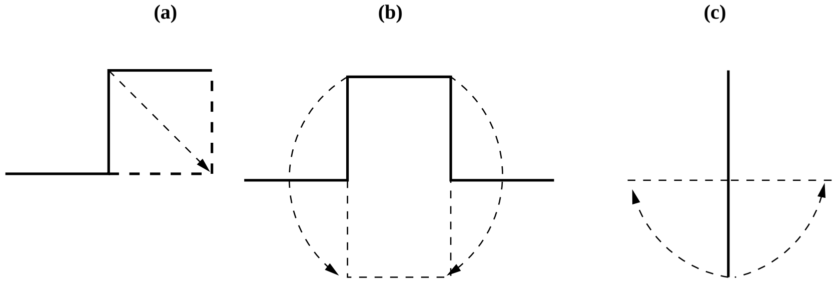

Figure 4.2: *three possible motions: (a), flipping of a corner of a lattice cell with respect to the diagonal of the cell; (b) crank mechanism; (c), change of the angle between the two segments at one end of the chain.*

4.1.2 Metropolis-Monte Carlo scheme

We are interested in the thermodynamics and the kinetics of the system, and we studied them with the Monte Carlo-Metropolis scheme [72]. This technique is frequently used for equilibrium properties nevertheless we also use it for kinetics assuming that local displacements give a dynamic with time scales proportional to reality. When we are interested in the statistical properties, we have to determine the partition function of the system, which is in the discrete case:

$$Z = \sum_i \exp(-\beta U(i)), \tag{4.3}$$

where the sum is over all the configuration of the system. In practice, the number of configuration in too large and it not possible to determine this sum numerically. We have the same problem for the calculation of integrals in the continuous case. Therefore we need specific methods to estimate these integrals. Monte Carlo algorithm consists in replacing the calculation of an integral by a discrete sum over points which are judiciously distributed. Indeed, one does not have to calculate the value of the integral where the integrand is negligible. Thus, we can determine in a reasonable number of step the value of the integral. Let us come back to the problem of statistical mechanics. We assume that we fix the temperature to T. We are often interested in the determination of averages quantities such as:

$$\langle A \rangle = \frac{\sum_i A_i \exp(-\beta U_i)}{Z}. \tag{4.4}$$

In Eq. (4.4) we can see:

$$P_i = \frac{\exp(-\beta U_i)}{Z}. \tag{4.5}$$

This quantity defines the probability of the configuration of energy U_i at equilibrium. If we can generate configurations with this weight, then the average of A will be estimated by

$$\langle A \rangle \simeq \frac{1}{N_r} \sum_i^{N_r} A_i. \tag{4.6}$$

So with the Monte Carlo method we can estimate the average of A if we can generate configurations with the equilibrium probability. Therefore, the problem consists in determining a method that generates a stochastic dynamic in order to get the equilibrium distribution. Then, the averages will simply be done by the relation (4.6). In 1953, to generate such a stochastic dynamics, Metropolis, Rosenbluth and Teller, proposed a method based on the detailed balance relation (in the canonical ensemble and at equilibrium):

$$W(j \to i) P_j^e = W(i \to j) P_i^e, \tag{4.7}$$

where $W(i \to j)$ is a transition probability of the state i to the state j and P_i^e is the equilibrium probability of the state i which is given by Eq. (4.5). We can rewrite relation (4.7) as:

$$\frac{P_i^e}{P_j^e} = \frac{W(j \to i)}{W(i \to j)} = e^{-\beta(U(i)-U(j))}. \tag{4.8}$$

Therefore the system will converge to the equilibrium state if at each transition of a state i to a state j the transition probabilities obey the relation (4.8). We only have to find a simple expression for the transition probability W. The choice of Metropolis *et al* which gives the Monte Carlo-Metropolis algorithm is the following:

$$W(i \to j) = \begin{cases} 1, & U(j) - U(i) \leq 0 \\ e^{-\beta(U(j)-U(i))}, & U(j) - U(i) > 0. \end{cases} \tag{4.9}$$

A possible algorithm to implement it is:

1. We generate a state j from state i using a deterministic rule or a random process
2. We calculate $\Delta U = U(j) - U(i)$.
3. - If $\Delta U \leq 0$, then $W(i \to j) = 1$ and we keep the new state j.
   - If $\Delta U > 0$, then $W(i \to j) = e^{-\beta \Delta U}$ and we pick a number r randomly in the interval [0,1]. We keep the state j if $r \leq e^{-\beta \Delta U}$, or we reject it if not.
4. We come back to the beginning of the procedure in 1.

Using this scheme, the system reaches its equilibrium state after a number of step that is difficult to estimate “a priori”. In practice the number of steps is chosen large enough to observe steady state values of the observed quantities averaged over a large number of individual steps. After that, we repeat the procedure with a different initial condition and another set of random numbers to get averages or equilibrium probability distributions from different realizations. Finally new algorithms based on Monte Carlo scheme [73] have been introduced to allow the study of bigger systems.

4.2 Equilibrium properties of the opening-closing transition

4.2.1 The transition in the absence of mismatch

Let us consider first the *equilibrium* properties of DNA hairpins in the simple case when they can only close with a correct matching of the bases in the stem. This would be the case if the base sequence in the stem forbids any mismatch. In order to compare with experimental results [4] we considered the case of a stem having 5 base-pairs ($n_s = 5$). Since there are only 4 types of bases, at least one has to appear twice in the stem. Thus the Watson-Crick pairing rules allow at least one mismatched pairing, but it may be very unfavourable because, if it occurred, the other bases of the stem would not be paired and may even experience some steric hindrance. In the model it is easy to strictly forbid any mismatched closing by using a sequence $t_i = \{1, 2, 3, 4, 5\}$ where all base-pairs have different types. Besides this condition, in our calculations we gave same energy $E_{HB} = -1$ to all types of base-pairs. This value sets the energy scale, and thus the temperature scale. With these parameters, the model does not attempt to mimic any real DNA hairpin, but it is designed to stay as simple as possible in order to exhibit the basic mechanisms that govern the hairpin properties.

Figure (4.3) shows the variation of the number of hydrogen-bonded base-pairs versus temperature for chains having different numbers N of nucleotides. The number of nucleotides in the loop is $N - 10$ since the stem is always made of two segments of 5 nucleotides. In these calculations, the bending energy E_A has been set to $E_A = 0.02$, and it has the same value along the whole DNA strand. The results have been obtained with different initial conditions: we start either from a closed hairpin or a random coil. Each point in the figure is an average of 100 calculations with different sets of random numbers to generate the initial conditions and the stochastic motions of the chains on the lattice, each calculation involving between $4\,10^8$ and $8\,10^8$ Monte Carlo steps (depending on temperature and chain length). The first $2\,10^7$ steps are discarded in the analysis to allow the model to equilibrating to the selected temperature. For $T \geq 0.15$ a good equilibration is achieved, while results at lower temperatures show some dependence on the initial conditions because an equilibrium state has not been reached. This is why they are not shown in Fig. (4.3).

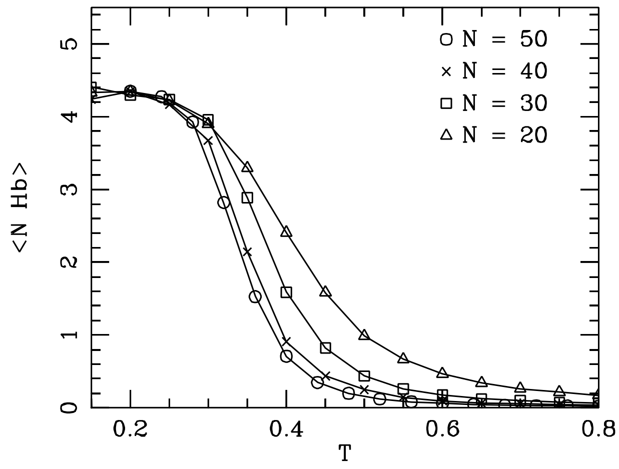

Figure 4.3: *Variation versus temperature of the number of hydrogen-bonded pairs in the stem for hairpins of different lengths N, in the absence of mismatches.*

As expected, when temperature increases we observe a fairly sharp decrease of the number of hydrogen-bonded base-pairs. It corresponds to the opening of the hairpin, which occurs over a temperature range of about 0.2 energy units, around the so-called "melting temperature" $T_m \approx 0.35$, which is well below to the temperature

$T = 1$ corresponding to the binding energy of a base-pair. This indicates that the entropy gain provided by the opening of the hairpin contributes to lower the free energy barrier for opening. Increasing the length of the loop lowers T_m, in agreement with the experiments [4]. It also makes the transition sharper, which is not observed in the experiments.

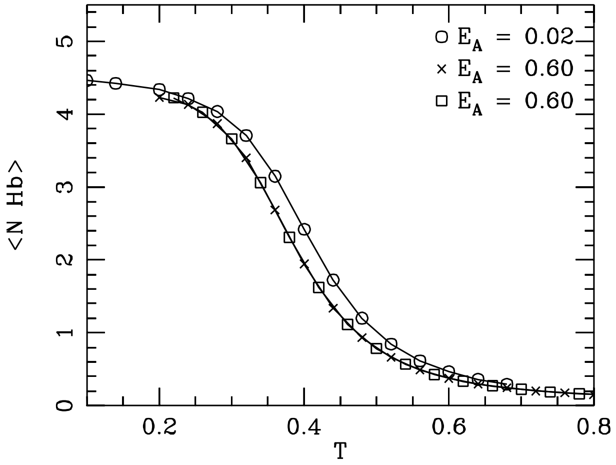

Figure 4.4: *Effect of the rigidity of the loop on the opening of the hairpin: variation versus temperature of the number of hydrogen-bonded pairs in the stem for loops with different bending energies $E_A = 0.02$ and 0.60, in the absence of mismatches. In the stem the bending energy has been set to $E_A = 0.02$ for both calculations. The two sets of points for $E_A = 0.6$ (crosses and squares) have been obtained in two independent calculations, with different sets of temperatures and different initial conditions. The crosses show results obtained with a closed hairpin initial condition, while the squares have been obtained with random initial conditions. Each point on this figure is an averaging over 100 sets of initial conditions and random numbers.*

The role of the rigidity of the loop can be tested by changing the value of the bending energy E_A for all the bends in the loop, without changing its value in the stem. Figure (4.4) shows that a more rigid loop leads to an opening at lower temperature, in agreement with the experimental observations [4]. However the variation of T_m given by the model appears to very small, and moreover, as discussed below, the effect of the rigidity of the loop on the thermodynamics of the hairpin is not correctly described in our model. This points out some limitations of the simplified model, although a quantitative comparison with the experiments is difficult because, in the experiments, the rigidity was varied by changing the bases from T to A. The larger purine bases A are assumed to give a higher rigidity to the strand but this could only be related to the variation of E_A by extensive all-atom numerical simulations [1]. Moreover, the role of base stacking in the loop is certainly more complex than the simple change of the rigidity of the chain that our simplified model can describe.

4.2.2 Role of the mismatches

One feature of DNA hairpins is that, unless they have a specifically designed sequence, they may close with a wrong pairing in the stem (see figure (4.1)). These imperfect, mismatched, closings have a higher energy that the perfectly closed hairpin, but they can be very long-lived.

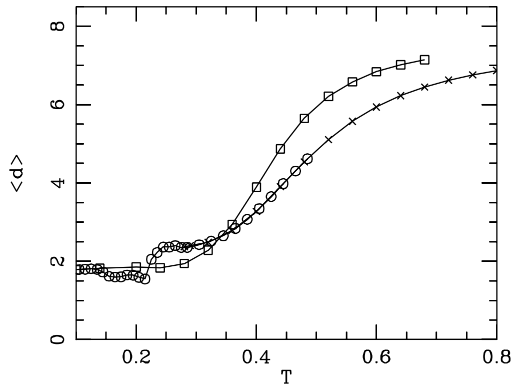

Figure 4.5: *Comparison of melting curves with and without mismatches. The mean value* $\langle d \rangle$ *of the distance between the first and last nucleotide is plotted versus temperature. The chain has* $N = 20$ *nucleotides, with* $E_{HB} = -1$ *for all base-pairs of the stem,* $E_a = 0.02$. *The squares show data without mismatch* ($t_i = \{1, 2, 3, 4, 5\}$), *while the circles and crosses show data with mismatches* ($t_i = \{1, 1, 1, 1, 1\}$). *In this case two sets of calculations have been performed. The circles have been obtained with* $8\,10^8$ *Monte Carlo steps, while the crosses involve only* $4\,10^8$ *Monte Carlo steps. For* $T > 0.25$ *the two sets give identical results, but, at low* T, *a smaller number of Monte-Caro steps slightly affects the results.*

They affect the opening-closing transition as shown in Fig. (4.5) which compares the melting curves in the presence and in the absence of mismatches. In order to allow mismatches, the sequence of bases of the stem has been set to $t_i = \{1, 1, 1, 1, 1\}$, i.e. all base-pairs are of the same type so that many mismatched pairings are possible, with 1,2,3,4 hydrogen-bonded base-pairs. In this case we show the mean value $\langle d \rangle$ of the distance between the first and last nucleotide of the chain rather than the number of hydrogen-bonded stem base pairs because $\langle d \rangle$ provides a more complete picture of the configuration of the hairpin.

On Fig. (4.5), the case without mismatch shows a smooth melting curve, similar to the results of Fig. (4.3). In the low temperature domain where the hairpin is closed, $\langle d \rangle$ is larger than the value $\langle d \rangle = 1$ that could be expected from a static image of the closed hairpin because there are fluctuations. They are particularly important at the free end of the stem, as schematised on Fig. (4.6).

When mismatches are allowed, the curve $\langle d(T) \rangle$ shows a fairly sharp kink around $T = 0.215$, and then an increase, qualitatively similar to cases without mismatch, but occurring however more smoothly and at higher temperature. The kink, which corresponds to a jump of $\langle d \rangle$ of about one unit, is due to the formation of a mismatched closing where only 4 base-pairs of the stem are formed (Fig. (4.6), right part). As temperature is raised further, the number of paired bases in the stem keeps decreasing, but, as there are many more possibilities for binding than in the no-mismatch case, the opening of the hairpin is more gradual.

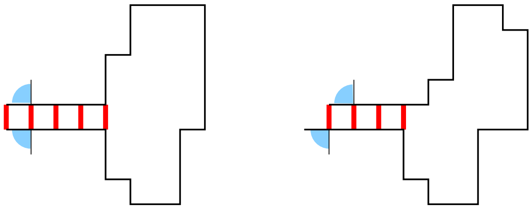

Figure 4.6: *Schematic plot of the fluctuations of the free end of the chain in a perfectly closed state (left) and in a mismatched state (right).*

4.3 Kinetics of the opening and closing

Up to now we spoke of the opening transition of the hairpin as if the hairpin should be closed at low T and open at high T. It is actually more complex because, in a small system like the hairpin, a phase transition between two states does not exist. Actually we always have a equilibrium between the open form O and the closed form C

$$C \underset{k_{cl}}{\overset{k_o}{\rightleftarrows}} O \; , \tag{4.10}$$

which can be studied like a chemical equilibrium rather than a phase transition. At low T the equilibrium is displaced towards closing and at high T it is displaced towards opening.

This suggests that the methods of chemical kinetics can be used to analyse the dynamics of the fluctuations of the hairpin. Let us consider that the hairpin is a two-state system. This is obviously an approximation which becomes very crude when mismatches are allowed since, in this case, the hairpin can also exist in some intermediate states where it is incompletely closed. In the absence of mismatch, the two-state picture is a satisfactory approximation, as shown in Fig. (4.7). This figure shows the histogram of the distance d between the two ends of the chains, and the histogram of the number of hydrogen-bonded base-pairs at temperature $T = 0.36$ for a model without mismatch with $N = 50$. This temperature is close to the melting temperature T_m for this model, and the histograms clearly show the coexistence of two populations of states: (i) an open state, where there are no hydrogen-bonded pairs in the stem, which corresponds to the hump for $d > 5$ on Fig. (4.7-a), (ii) a closed state corresponding to the sharp maximum for $d < 4$ in Fig. (4.7-a) and to the existence of 2 to 5 hydrogen-bonded base-pairs in Fig (4.7-b) (with a maximum at 4, due to the opening fluctuations at the end of the stem as discussed above and schematised in Fig. (4.6), left).

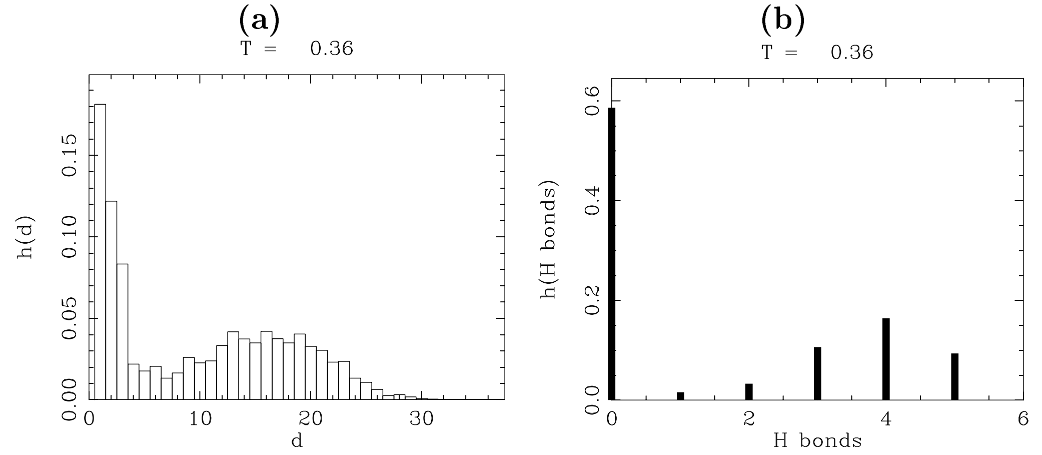

Figure 4.7: *Normalised histograms of the distance d between the two ends of the chain (a), and number of hydrogen bonds (b) for a hairpin with $N = 50$ and no mismatches, at temperature $T = 0.36$. This temperature is close to the opening temperature T_m of this hairpin. Model parameters $E_{HB} = -1$, $E_a = 0.02$. The histograms show the coexistence of two populations: one population of completely open hairpins (large values of d and 0 hydrogen bonds) and a population of closed hairpins in which some of the hydrogen bonds are formed, the highest probability being with 4 hydrogen bonds formed.*

The two-state picture allows us to write standard kinetic equations for the populations $[C]$ and $[O]$ of the closed and open states as

$$\frac{d[C]}{dt} = -k_o[C] + k_{cl}[O] \tag{4.11}$$

$$\frac{d[O]}{dt} = +k_o[C] - k_{cl}[O] \ , \tag{4.12}$$

where k_o and k_{cl} are the kinetic constants for the opening and closing events respectively. This system has the solution

$$[C](t) = \frac{C_0 k_o}{k_o + k_{cl}} e^{-(k_o+k_{cl})t} + \frac{C_0 k_{cl}}{k_o + k_{cl}} \ , \tag{4.13}$$

where C_0 is the value of $[C]$ at time $t = 0$. This shows that, if we start from a population of closed hairpins, we expect it to decay exponentially with a characteristic time $\tau = 1/(k_o + k_{cl})$ until an equilibrium is reached with

$$\frac{[O]}{[C]} = \frac{k_o}{k_{cl}} = K_e \ , \tag{4.14}$$

where K_e is the equilibrium constant.

Therefore, if we follow the evolution of the population of closed hairpins in a Monte Carlo simulation which starts from C_0 closed configurations, we can determine separately τ (from the decay of the closed population) and K_e from the final equilibrium state, so that we can determine the kinetic constants for opening and closing, given by

$$k_o = \frac{1}{\tau}\,\frac{1}{1+K_e} \qquad k_{cl} = \frac{1}{\tau}\,\frac{K_e}{1+K_e}. \tag{4.15}$$

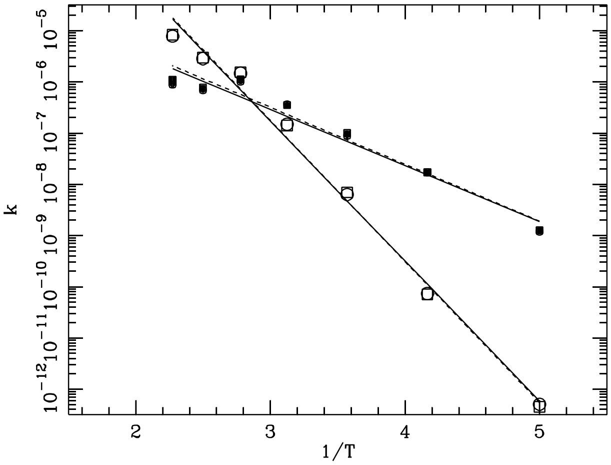

Figure 4.8: *Arrhenius plot of the kinetic constants k_{op} (open symbols) and k_{cl} (closed symbols) versus $1/T$ for a model without mismatch, $N = 50$, $E_{HB} = -1$, $E_a = 0.02$. The time unit is a Monte Carlo step. The lines are least square fits of the points (full lines for opening state defined by $d > 4$, and dashed lines for opening defined by the absence of hydrogen bonded base pairs).*

Figure (4.8) shows the results of such an analysis for a case without mismatches. The open/closed state of the chain was measured with two different criteria: from the distance d between the two ends (a value $d > 4$ is considered as an open state) or from the number of hydrogen-bonded base-pairs (an open state must not have any bound base-pair). Both give very similar results, in agreement with the above discussion of Fig. (4.7) which shows that both criteria can be used to separate between the open and closed states. When they are plotted in logarithmic scale versus $1/T$, the kinetic constants are well fitted by straight lines, which allows us to define activation energies E_o and E_{cl} for the opening and closing events by

$$k_o = K_o e^{-E_o/T} \qquad k_{cl} = K_{cl} e^{-E_{cl}/T}. \tag{4.16}$$

The fits of Fig. (4.8) give $E_o = 6.3$ and $E_{cl} = 2.5$. Figure (4.8) is very similar to the figures showing k_o and k_{cl} which can be obtained experimentally [4] (see figure (2.5)). The experiments also find an opening activation energy much larger than the closing energy. The experimental ratio E_o/E_{cl} is even larger than the ratio that we derive from our model. Owing to the simplicity of the model, it would be meaningless to try to adjust parameters to get the experimental ratio. What is more interesting is the meaning of this result $E_o \gg E_{cl}$, which can be related to the need to break the hydrogen bonds linking the base-pairs to open the hairpin, while the kinetic of the closing is dominated by entropic effects because it occurs when the two sides of the stem managed to reach the correct spatial position after a random walk in the configuration space.

Experiments show that the opening kinetics is almost insensitive to the length of the loop, while the closing slows down significantly when the length of the loop increases (k_{cl} decreases) while its activation energy does not depend on the length of the loop. The model confirms that the activation energies do not vary when we change N, but it only finds a very small variation of k_{cl} as a function of N, contrary to the experiments. This points out one of its severe limitations: the entropy of the loop is not sufficiently well described when its motions are constrained on a two-dimensional square lattice. This limitation also appears when we study the effect of the rigidity of the loop. As noticed above, the effect is very small and to obtain some noticeable influence of the rigidity, we have to increase the bending energy very significantly, for instance up to $E_A = 0.6$ (figure 4.4). In this case the activations energies become $E_o = 5.5$ and $E_{cl} = 2.5$, i.e. the opening activation energy is reduced by about 12 % and the closing energy is only weakly affected, while the experiments found a large increase of the closing activation energy and almost no change for E_o . This shows that, for this study, our model does not correctly describe the experiment. Besides an incorrect description of entropic effects in the model, that we already mentioned above, other phenomena could enter, and particularly a possible role of the mismatches in the experimental sequence. While the model strictly forbids mismatches, in the experiments, changing the bases in the loop from A to T modifies the possible mismatches.

As one could expect, the kinetics of the hairpin fluctuations is strongly affected by the presence of mismatches. The two-state approach is no longer valid. Mismatched states are open if we define them in terms of the distance between the ends but still show many hydrogen-bonded base-pairs. Although the time evolution of the closed states is no longer a simple exponential decay, an approximate fit by an exponential gives the order of magnitude of the characteristic time τ. Figure (4.9) shows the values of τ determined with two definitions of an open state: (i) a state where the distance of the two ends of the chain is $d > 2$, (ii) a state where all the hydrogen bonds linking the bases in the stem have been broken. Figure (4.9) shows that the lifetime of closed hairpins defined according to these criteria vary by several orders of magnitude. This is not surprising because a hairpin which is closed in a mismatched state may be counted for open for the first criterion ($d > 2$) but closed with respect to the second one since some of its base-pairs are hydrogen bonded. In this case the above analysis to calculate k_o and k_{cl} loses its meaning.

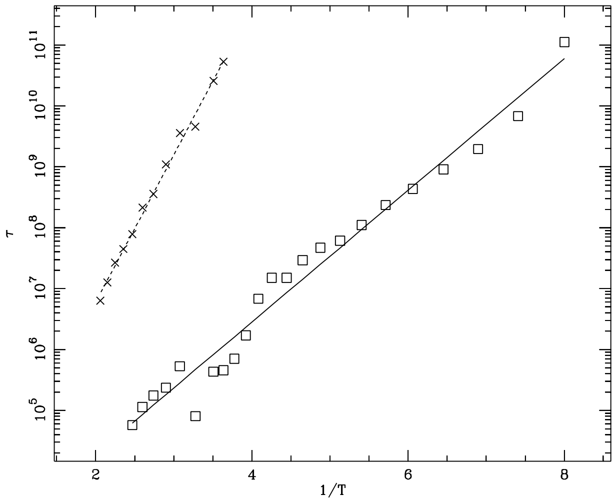

Figure 4.9: *Logarithmic plot of the characteristic time for opening τ versus $1/T$ for a case with mismatches. The squares (fitted by the full line) correspond to a definition of the opening from the distance of the two ends ($d > 2$) and the crosses (fitted by the dashed line) define opening by the absence of any hydrogen-bonded base-pair. The time unit is a Monte Carlo step.*

The role of the mismatches in the experimental studies of molecular beacons [4] has not been investigated so that we cannot compare the results of the model with experimental data. Although the sequence used in [37] could in principle allow wrong closing, there were certainly much less likely than in our study where all base-pairs of the stem are the same. Moreover, studies using a fluorophore and a quencher are only probing the distance d between the ends of the chain, so that they are not sensitive to wrong closings. For such a study the hairpin is still a two-state system.

Chapter 5

PBD-Polymer model for DNA Hairpins

Contents

5.1 Presentation of the model 69
5.2 Study of the stem 71
  5.2.1 Partition function 73
  5.2.2 Transfer integral in the continuum medium approximation 75
  5.2.3 Results 77
5.3 The complete system 81
  5.3.1 Partition function 82
  5.3.2 Free Energy and Entropy 84
  5.3.3 Kinetics: theoretical predictions 85
5.4 Case of S≡1 90
  5.4.1 Thermodynamics 91
  5.4.2 Kinetics 96
5.5 Complete calculation: $S \neq 1$ 100
  5.5.1 Thermodynamics 101
  5.5.2 Kinetics 111
  5.5.3 Discussions 116
  5.5.4 Beyond the PBD-model for the stem 117

5.1 Presentation of the model

The previous model shows some weaknesses especially on the modelling of the entropy of the system. So we have developed an off lattice model that still is a highly simplified model but is nevertheless much richer, in particular regarding the modelling of the loop, which plays a large role in the properties of DNA hairpins. A simple view of DNA hairpins can consider them as a single short polymer with hydrogen bonds as well as base-pair stacking between the two ends of the chain. So the idea is to combine models of polymers with the PBD-model for the double helix. Our model is based in this point of view. We have chosen to divide the model of the hairpin in two parts:

- the loop formed by a sequence of identical bases which is treated as a simple polymer, in practice made of a single type of base, A or T.

- The stem which is an extension of the two ends of the loop (with a polymer behaviour) but with additional interactions according to the pairing of complementary monomers or bases (given by the PBD-model).

In practice we construct our model beginning from the simplest loop which is a sequence of A or T-bases, i.e. an homogeneous polymer. The loop is modelled by a polymer chain in three dimensions. One major question of our study is what is the appropriate model for the loop? We will examine it in detail in this chapter but at this level, we can already make some comments that set the framework of our study. We have tested the three different polymer models that we have presented in the Chap. 3. The FJC is the simplest but we can expect it to be oversimplified because the experiments show that the stacking interaction of the bases inside the loop is important regarding the physical properties of the hairpin. Fixing the value of θ in the FRC could perhaps model in some sense the stacking interaction and the rigidity even if the rotation around the bond is free because, as we have shown in Chap. 3, the value of θ determines the persistence length of the chain, i.e. its rigidity. Thus this model deserves an investigation. The Kratky-Porod model which seems to be a good model for the modelling of long DNA chains could be a good candidate for the loop because it includes a parameter which represents the rigidity of the chain. The question is to know whether this model remains correct for single chain where the persistence length is very different from that of double stranded DNA for which it was experimentally tested, and for short chains less than ten times the persistence length.

As we are interested in a very short stem, it is not necessary to take into account the helicity of the DNA molecule [33], [34]. As for the previous model, the goal is to find thermodynamics and kinetics properties of this system [37], [4]. Before doing that, we will study separately a short stem in order to see the difference with the infinite case and it will also give us the qualitative properties of this part on the complete system. Figure (5.1) gives a schematic representation of the model.

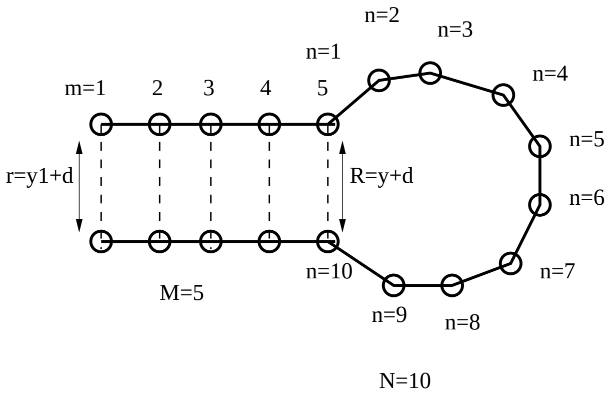

Figure 5.1: *Plot of the model to define some notations. Index m=1$\cdots$M will be used for the stem base-pairs. Index n=1...N+1 will be used for the bases in the loop. Note we have 2M+N-1 bases in total. The variables y_m are the stretching of the base pairs $y_m = 0$ means that the distance between the bases is d=10Å, which is the value that we use for the equilibrium distance of bases in a pair. The variable r will be used for the variation of the distance between the two bases at the end of the hairpin, i.e. r=y_1+d. The variable R is the distance between the two ends of the loop. Therefore R=y_M+d.*

5.2 Study of the stem

In this part we study the stem with the condition that the two strands are confined because we must keep in mind that we have the loop which limits their separation. In practice we will impose this condition through the potential $V(y)$. In order to illustrate the transfer integral method we have chosen a very simple version of the PBD-model which allows analytical calculations. Figure (5.2) gives a schematic representation of the model of the stem.

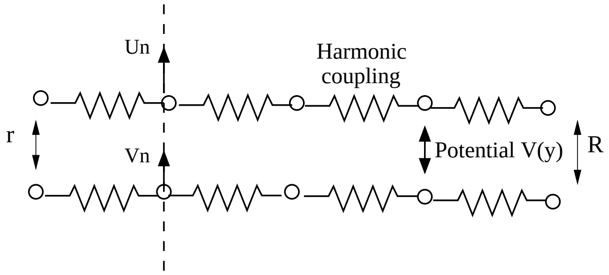

Figure 5.2: *Schematic representation of the stem.*

The characteristics of the stem are the following:

- The displacements along the chain are not considered because their amplitude is much smaller than the perpendicular ones [32]. The transverse displacements are represented by u_n and v_n for the two bases.

- The coupling between two consecutive bases is harmonic.
- To model the combined effect of the hydrogen bond, the repulsive part of the phosphate as well as the effect of the solvent, we put an effective potential. The PBD-model uses a Morse potential. In this section we use a simpler square potential shown on Fig. (5.3). It has qualitatively the shape that we can expect for the interaction within a base-pair of the stem. The well describes the binding of the bases. The plateau corresponds to the open state. But the bases are confined to a finite distance by the loop. This effect is described by the infinite barrier at distance L.

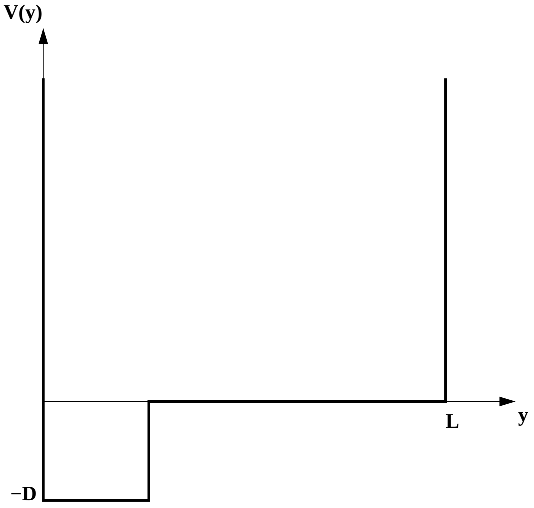

Figure 5.3: *Schematic representation of the potential V(y) where y is the stretching of the hydrogen bonds between the bases. The infinite wall at y=0 means that the bases cannot overlap, while the infinite wall at y=L comes from the maximum separation of the strands, limited by the length of the loop.*

Therefore, the Hamiltonian of the model is:

$$H = \sum_n \left[\frac{1}{2} m \left(\dot{u_n}^2 + \dot{v_n}^2 \right) + \frac{1}{2} K \left[\left(u_n - u_{n-1} \right)^2 + \left(v_n - v_{n-1} \right)^2 \right] + V \left(u_n - v_n \right) \right], \tag{5.1}$$

where the three terms represent the kinetic energy of the transverse vibrations, the potential energy of the chain and the bonds connecting bases in pairs, respectively. m is the mass of a base and K, the spring constant. This Hamiltonian can be used for various calculations [10], [11], [19] but here we are interested in the statistical mechanics only. It is convenient to introduce new variables x_n and y_n linked to u_n

and v_n by:

$$\begin{aligned} x_n &= \frac{1}{\sqrt{2}}(u_n + v_n) \\ y_n &= \frac{1}{\sqrt{2}}(u_n - v_n). \end{aligned}$$

The Hamiltonian takes the following form:

$$\begin{aligned} H &= \sum_n \left[\frac{1}{2}m\dot{x_n}^2 + \frac{1}{2}K\left(x_n - x_{n-1}\right)^2\right] + \sum_n \left[\frac{1}{2}m\dot{y_n}^2 + \frac{1}{2}K\left(y_n - y_{n-1}\right)^2\right] + V(y_n) \\ H &= H_x + H_y. \end{aligned} \tag{5.2}$$

We immediately see that the Hamiltonian is divided in two parts: H_x describes the harmonic center of mass motion and H_y contains all the anharmonicities expressed in $V(y_n)$. In the next section, we will focus our attention on H_y only because it is this part of the Hamiltonian that contains the physics of the hairpin opening because it is the variable y_n that describes the opening or the closing of a base-pair.

5.2.1 Partition function

In statistical physics, if we are able to derive the partition function of a system, then we get all the thermodynamic quantities. The problem is that we must sum over all the configurations and it is generally impossible. That's why numerical approximations like Monte Carlo Metropolis scheme or other more sophisticated methods are sometimes used [72]. Here we present an exact analytical calculation of the partition function for a finite homogeneous stem. In the case of a non homogeneous stem numerical calculation are necessary [71].
The partition function that we have to calculate is the following:

$$Z_s = \int \prod_{i=1}^{N} dy_i dp_i e^{-\beta \sum_i \frac{p_i^2}{2m}} e^{-\beta\left[\sum_i V(y_i) + \sum_{i=2}^{N} \frac{K}{2}(y_i - y_{i-1})^2\right]}. \tag{5.3}$$

The momentum part in the partition function gives:

$$Z_{sp} = \left(\frac{2\pi m}{\beta}\right)^{\frac{N}{2}}.$$

To go further in the calculation, we introduce the eigenfunctions and eigenvalues of the non symmetric transfer integral operator:

$$\int dy_{i-1} e^{-\beta\left(\frac{K}{2}(y_i - y_{i-1})^2 + V(y_i)\right)} \phi_k^R(y_{i-1}) = e^{-\beta\epsilon_k} \phi_k^R(y_i) \tag{5.4}$$

$$\int dy_{i-1} e^{-\beta\left(\frac{K}{2}(y_i - y_{i-1})^2 + V(y_{i-1})\right)} \phi_k^L(y_{i-1}) = e^{-\beta\epsilon_k} \phi_k^L(y_i), \tag{5.5}$$

with:

$$\int dy \phi_k^R(y)\phi_k^L(y) = 1 \tag{5.6}$$

$$\sum_k \phi_k^R(y)\phi_k^L(x) = \delta(x-y) \tag{5.7}$$

$$\phi_k^L(y) = e^{\beta V(x)}\phi_k^R(y). \tag{5.8}$$

Now it is convenient to use the identity:

$$\int dr \delta(r - y_1) = 1.$$

Therefore we can introduce this integral in the partition function without changing anything:

$$\begin{aligned} Z_s = Z_{sp} \int \prod_{i=2}^{N} dy_i e^{-\beta\left[\sum_i^N V(y_i) + \sum_{i=3}^N \frac{K}{2}(y_i - y_{i-1})^2\right]} \\ \int dy_1 \int dr \ \delta(r - y_1) e^{-\beta\left(V(y_1) + \frac{K}{2}(y_2 - y_1)^2\right)}. \end{aligned}$$

Using Eq. (5.7), we get:

$$\begin{aligned} Z_s = Z_{sp} \int dr \sum_k \phi_k^R(r) \int \prod_{i=2}^{N} dy_i e^{-\beta\left[\sum_{i=2}^N V(y_i) + \sum_{i=3}^N \frac{K}{2}(y_i - y_{i-1})^2\right]} \\ \underbrace{\int dy_1 e^{-\beta\left(V(y_1) + \frac{K}{2}(y_2 - y_1)^2\right)} \phi_k^L(y_1)}_{e^{-\beta\epsilon_k}\phi_k^L(y_2)}. \end{aligned}$$

Then we can perform the same integration over the variables y_2 to y_{N-1}:

$$Z_s = Z_{sp} \sum_k e^{-\beta(N-1)\epsilon_k} \int dr \phi_k^R(r) \int dy_N e^{-\beta V(y_N)} \phi_k^L(y_N).$$

Finally using Eq. (5.8) we get the following expression for the partition function:

$$Z_s = \left(\frac{2\pi m}{\beta}\right)^{\frac{N}{2}} \sum_k e^{-\beta(N-1)\epsilon_k} \left[\int dy \phi_k^R(y)\right]^2. \tag{5.9}$$

Thus if we are able to find the eigenstates and the eigenvalues of the transfer integral operator, we can compute the thermodynamic quantities such as the free energy, the entropy and the heat capacity.

5.2.2 Transfer integral in the continuum medium approximation

If we use the continuum medium approximation it is possible to get the eigenfunctions and the eigenvalues that we need. Due to the Gaussian function in the transfer integral operator $exp\left(-\beta K(y_i - y_{i-1})^2/2\right)$, the kernel takes very small values except in the vicinity of y_i. Consequently we can perform a Taylor expansion of $\phi_k^R(y_{i-1})$ around y_i and then integrate over y_{i-1}:

$$\begin{aligned}
e^{-\beta\epsilon_k}\phi_k^R(y_i) &= \int dy_{i-1}e^{-\beta\left(\frac{K}{2}(y_i-y_{i-1})^2+V(y_i)\right)}\phi_k^R(y_{i-1}) \\
&= e^{-\beta V(y_i)}\int dy_{i-1}e^{-\beta\frac{K}{2}(y_i-y_{i-1})^2}\phi_k^R(y_{i-1}) \\
&= e^{-\beta V(y_i)}\int dy_{i-1}e^{-\beta\frac{K}{2}(y_i-y_{i-1})^2}\Bigg[\phi_k^R(y_i)+ \\
&\qquad \frac{d\phi_k^R}{dy}(y_i-y_{i-1})+\frac{1}{2}\frac{d^2\phi_k^R}{dy^2}(y_i-y_{i-1})^2+\cdots\Bigg] \\
&= e^{-\beta V(y_i)}\left[\phi_k^R(y_i)+0+\frac{1}{2}\frac{d^2\phi_k^R}{dy^2}\left(\frac{-2}{K}\right)\frac{\partial}{\partial\beta}\sqrt{\frac{2\pi}{\beta K}}+\cdots\right] \\
&= e^{-\beta V(y_i)}\sqrt{\frac{2\pi}{\beta K}}\left[1+\frac{1}{2\beta K}\frac{d^2}{dy^2}+\cdots\right]\phi_k^R(y_i) \\
e^{-\beta\epsilon_k}\phi_k^R(y_i) &= e^{-\beta V(y_i)}\sqrt{\frac{2\pi}{\beta K}}\left(e^{\frac{1}{2\beta K}\frac{d^2}{dy^2}}\right)\phi_k^R(y_i).
\end{aligned}$$

Indeed, we recognize the expansion of an exponential. Putting $e = \frac{1}{2\beta}ln\left(\frac{\beta K}{2\pi}\right)$, $\alpha = \frac{1}{2\beta^2 K}$ and $E_k = \epsilon_k - e$ we get the following Schrödinger equation:

$$-\alpha\frac{d^2\phi_k^R(y)}{dy^2}+V(y)\phi_k^R(y) = E_k\phi_k^R(y). \tag{5.10}$$

Consequently finding the eigenfunctions and eigenvalues is equivalent to solving a Schrödinger equation for a particle in the potential $V(y)$. The solution of this equation is quite easy to derive and we will only give the result here. We must consider two cases, one for $E_k < 0$ and the other for $E_k > 0$.

Bound states: -D < E < 0 In the solution of the Schrödinger equation in the book of Peyrard and Dauxois [74] with a similar potential, but without the restriction y<L, we see that a localized ground state exists only under a temperature $T_m^\infty = \frac{2a\sqrt{2KD}}{\pi k_b}$. In our case $L \gtrsim 100a$, which means that the constraint y<L does not change qualitatively the results, although the system now has a discrete spectrum for all E. When the particle is in the well, it lies in a localized ground state, which exists for $T < T_m$ with $T_m \approx T_m^\infty$.

One can show that the ground state has the following form:

$$\phi_0^R(y) = \begin{cases} A_0 \sin k_0 y & 0 \leq y \leq a, \\ \\ A_0 \frac{\sin k_0 a}{\sinh \rho_0 (L-a)} \sinh \rho_0 \left(L - a\right) & a < y \leq L. \end{cases} \tag{5.11}$$

With $k_0^2 = \frac{D+E_0}{\alpha}$ and $\rho_0^2 = -\frac{E_0}{\alpha}$. One must be careful for the normalisation. Indeed the correct normalisation is given by the Eq. (5.6). So that we have:

$$\begin{aligned} \frac{1}{A_0^2} = \frac{e^{-\beta D}}{k_0} & \Big[k_0 a - \sin k_0 a \cos k_0 a \Big] \\ & + \frac{\sin^2 k_0 a}{\rho_0} \left[\coth \rho_0 \left(l - a\right) - \frac{\rho_0 \left(L - a\right)}{\sinh^2 \rho_0 \left(L - a\right)} \right]. \end{aligned} \tag{5.12}$$

The eigenvalue E_0 is solution of the equation :

$$\tan k_0 a = -\frac{k_0}{\rho_0} \tanh \rho_0 \left(L - a\right). \tag{5.13}$$

In practice we solve this equation numerically.

Extended states: E>0 As the potential $V(y)$ goes to infinity for $y > L$, we get a infinite but discrete number of eigenfunctions. Indeed, the confining aspect of the potential leads to a quantization of the eigenvalues. In this case, the eigenfunctions are given by :

$$\phi_n^R(y) = \begin{cases} A_n \sin k_n y & 0 \leq y \leq a, \\ \\ A_n \frac{\sin k_n a}{\sin k_n^{'} (L-a)} \sin k_n^{'} \left(L - a\right) & a < y \leq L. \end{cases} \tag{5.14}$$

With $k_n^2 = \frac{D+E_n}{\alpha}$ and $k_n^{'2} = \frac{E_n}{\alpha}$. The condition of normalisation gives the A_n :

$$\begin{aligned} \frac{1}{A_n^2} = \frac{e^{-\beta D}}{k_n} & \Big[k_n a - \sin k_n a \cos k_n a \Big] \\ & + \frac{\sin^2 k_n a}{k_n^{'}} \left[\cot k_n^{'} \left(L - a\right) - \frac{k_n^{'} \left(L - a\right)}{\sin^2 k_n^{'} \left(L - a\right)} \right]. \end{aligned} \tag{5.15}$$

And the eigenvalues are given by :

$$\tan k_n a = -\frac{k_n}{k_n^{'}} \tan k_n^{'} \left(L - a\right). \tag{5.16}$$

In this case we also find the solutions numerically. Figures (5.4) and (5.5) give some eigenfunctions for $T < T_m$ and the evolution versus temperature of the eigenstates corresponding to the lowest eigenvalues versus temperature.

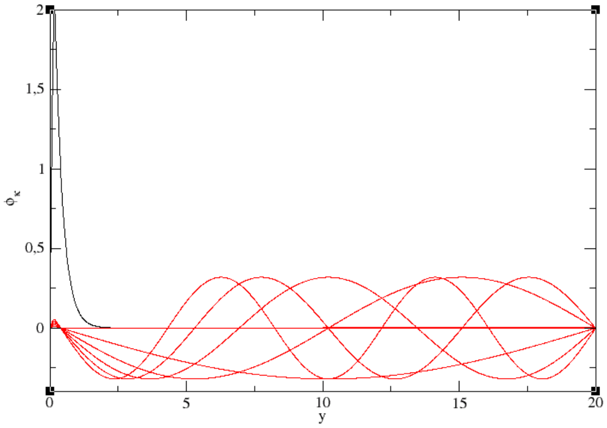

Figure 5.4: *Representation of eigenfunctions.*

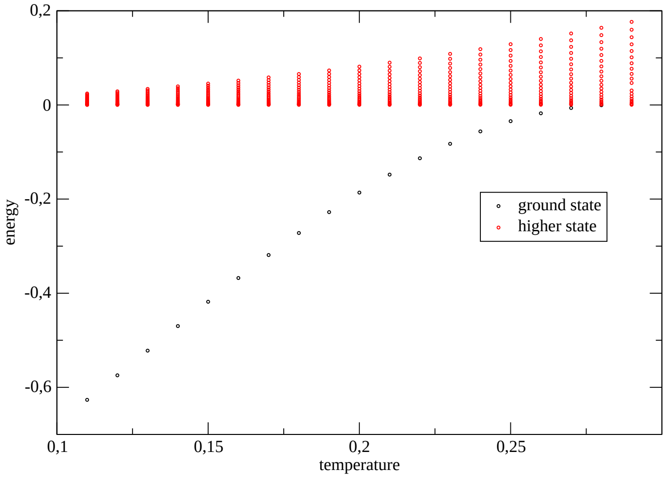

Figure 5.5: *Evolution of the eigenvalues as a function of temperature.*

Now we have the eigenfunctions and eigenvalues necessary to compute the partition function of the stem.

5.2.3 Results

Free energy and Entropy Using the expression of the partition function and the relation $F(T) = -k_b T \ln\ Z_s$ we can compute the total free energy of the stem.

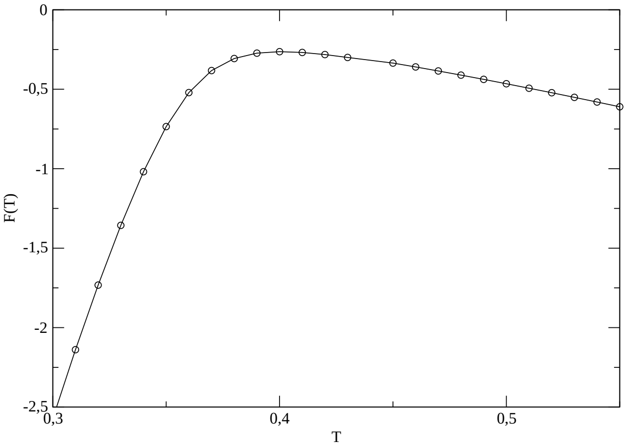

Figure 5.6: *Free energy of a finite stem.The parameters are the following: D=4; a=0.1, K=6 and N=5 in arbitrary units*

And the derivative of the free energy determines the evolution of the entropy of the system with temperature.

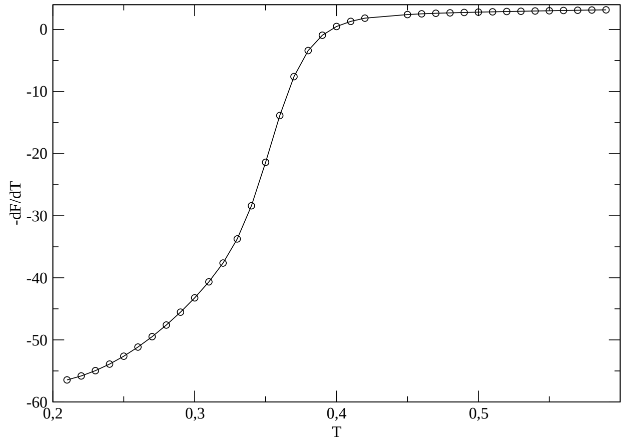

Figure 5.7: *Temperature variation of the entropy of the stem. The parameters are the following: D=4, a=0.1, K=6 and N=5 in arbitrary units*

The graphic of the entropy does not show a transition because there is no discontinuity or angular point in the free energy. The entropy grows continuously with the temperature but there is nevertheless a temperature range in which the entropy increases faster. It corresponds to the temperature domain in which the system changes form closed to open. Instead of a transition, for the finite system that we consider here, we can expect the coexistence of closed and open state with a gradual shift from a mostly closed to a mostly open situation. To verify this hypothesis we can select a "reaction coordinate" and compute the free energy versus this coordinate. For the hairpin the appropriate coordinate is r, the stretching of the base-pair that terminates the hairpin. This parameter is appropriate because it is related to the experiments that use FRET to detect the variation of distance between a fluorophore and a quencher.

Free energy as a function of r Let us calculate this new quantity which will be very important for the study of the hairpin. We must calculate the partition function for a given r. The derivation is quite similar to the previous calculation. So we have to integrate $e^{-\beta H_s}$ over all the variables of the stem excepted the first variable y_1. That is equivalent to integrating over the first variable y_1 but putting also a delta function $\delta(r - y_1)$. Therefore the partition function is given by:

$$Z_s(r) = Z_{sp} \int \prod_{i=2}^{N} dy_i e^{-\beta\left[\sum_i^N V(y_i) + \sum_{i=3}^N \frac{K}{2}(y_i - y_{i-1})^2\right]} \int dy_1 \delta(r - y_1) e^{-\beta\left(V(y_1) + \frac{K}{2}(y_2 - y_1)^2\right)}.$$

Then we perform the same calculation as for Z_s introducing the eigenstates of the transfer integral operator and finally we get:

$$Z_s(r) = \left(\frac{2\pi m}{\beta}\right)^{\frac{N}{2}} \sum_k e^{-\beta(N-1)\epsilon_k} \phi_k^R(r) \int dy \phi_k^R(y). \tag{5.17}$$

In practice the summation over k is truncated to the 100 lowest values of ϵ because the other contributions are negligible. Consequently we can easily compute the free energy landscape $F_{s,T}(r) = -k_b T ln Z_{sr}$. Figure (5.8) gives the evolution of the free energy landscape of the stem as a function of temperature.

We get a free energy with a well for a small value of r, which represents the closed configuration, and a large plateau for higher value of r which represents the open configurations. The fact that we have a plateau comes from the form of the potential $V(y)$. The shape of the free energy $F(r)$ indicates that only one state is really stable, the closed state. But due to the large plateau, states with large r will also be populated at any temperature. And when T increases their weight will increase

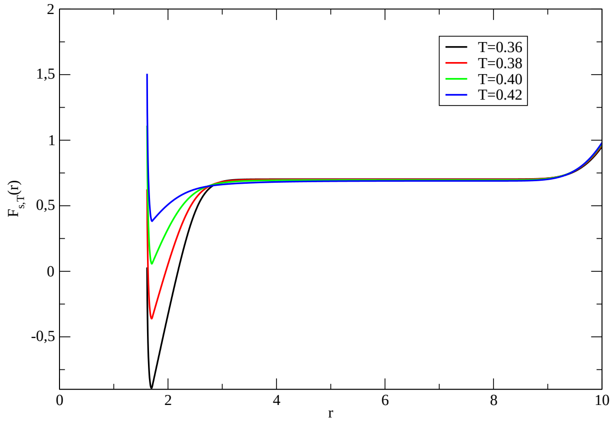

Figure 5.8: *Free energy landscape for different temperature.*

because the depth of the well corresponding to the closed state decreases. Therefore the free energy $F(r)$ shows that the stem opens gradually when temperature increases. However for the stem alone we cannot speak of a transition since only one stable state exists. The expression of $Z(r)$ allows us to compute the mean value of r versus T, which is a measure of the opening of the double stranded DNA. Notice that the value of $\langle r \rangle$ involves the summation over all eigenstates (in practice 100). On the contrary in the limit $N \rightarrow \infty$ the sum is dominated by the lowest eigenvalue ϵ_0. It is interesting to evaluate the influence of the excited states ϵ_k $(k > 0)$ on the mean distance of the first base-pair $\langle r \rangle$. The expression of $\langle r \rangle$ is given by

$$\langle r \rangle = \frac{\int dr \; r Z_s(r)}{\int dr Z_s(r)}. \tag{5.18}$$

Figure (5.9) shows $\langle r \rangle$ calculated with respectively 1 term (ϵ_0 only), 2, 5, 10 terms in the summation.

With one term we note sharp rise of $\langle r \rangle$ while the transition appears smoother when we include additional terms. This is because the summation restricted to the lowest term corresponds to the thermodynamic limit for which a true transition would exist (at least in the limit $L \rightarrow \infty$) while the introduction of the extra terms allow us to properly take into account the finite size of the stem. The simple square potential that we have chosen is convenient for this study because we can get the eigenfunctions of the transfer operator in an analytic form. For $L \rightarrow \infty$ and the Morse potential of the PBD-model an analytical expression exists (but is very tedious to manipulate and leads to numerical difficulties) but for a finite L, only the numerical approach would have been possible if we had not chosen the simple square potential.

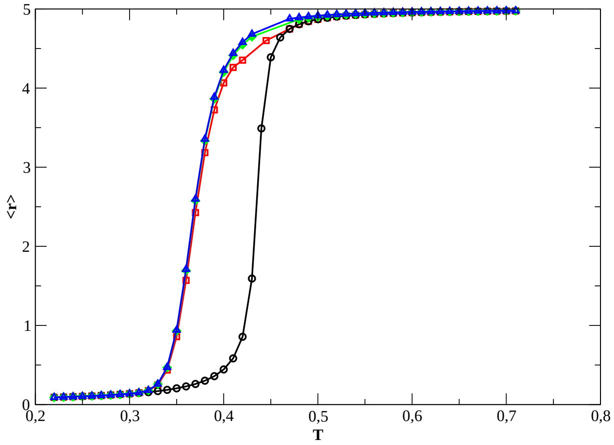

Figure 5.9: *Influence of the excited states on the mean distance of the first base-pair. The parameters are D=4, a=0.1, L=10 and K=6 in arbitrary units. ●: one term; □: two terms; ◇: five terms and △: ten terms in the summation.*

To conclude, we have seen that the study of a **finite** stem requires several eigenstates and with the simple version of the PBD-model it is quite easy to calculate them. Nevertheless, we know that to be more realistic we have to use the complete version of the PBD-model that we have presented in Chap. 1 with a non linear stacking and a Morse potential. Indeed, the work on the DNA molecule has shown that the stacking is more important when two consecutive base-pairs are closed than one intact and the other broken. To take this into account the PBD-model includes a non linear stacking given by W in Eq. (1.3). Moreover the potential which characterises hydrogen bonds is the Morse potential. The coupling in the Hamiltonian of the stem given by Eq. (5.2) (without the H_x) is now replaced by Eq. (1.3). In the case of the complete model we cannot use the transfer integral method because it is difficult to find all the eigenstates and eigenvalues of the transfer operator. A numerical calculation of the eigenstates could be possible but, even this approach is technically difficult due to overflows and numerical accuracy problems. Moreover, the approximation of continuous media is not correct for small chains as it is shown in Ref. [32]. For these reasons we have used a direct numerical integration of the partition function for the complete system. We present our calculation in the next section.

5.3 The complete system

Now we can come back to the problem of the hairpin. The goal is to find the partition function of the system in order to get the free energy landscape. With this quantity we will be able to find thermodynamics and kinetics properties and compare them to the experimental ones.

5.3.1 Partition function

As experiments probe the opening of hairpins by using a fluorophore/quencher system which is sensitive to the distance between the ends of the hairpin, it is useful to compute the partition function of the system for a given distance r between the two ends of the chain as we did for the stem in the previous section. Therefore we introduce a delta function in the calculation of the partition function as we have done for the stem only. In order to see how we construct our partition function let's begin by a system without stacking interaction and hydrogen bonds, i.e a polymer alone.

First of all the partition function for a given end-to-end distance $r_M = R$ is linked to the end-to-end probability distribution

$$P_N(r_M) = \frac{\int \prod_N d\alpha_N \delta\left(\| \sum_{i=1}^{N-1} \mathbf{r}_i \| - r_M\right) e^{-\beta\ H_N(\alpha_N)}}{\int \prod_N d\alpha_N e^{-\beta H_N(\alpha_N)}} = \frac{Z_N(R)}{Z_N^{tot}}. \tag{5.19}$$

Where N is the number of monomers, $\{\alpha_N\}$, the generic variables of the loop and H_N, the Hamiltonian of the loop. In order to build the partition function of the hairpin we shall start from the reduced partition function of the loop made of N monomers $Z_N(r_M)$, where r_M is the distance between the ends of the loop which is also the distance between the two bases making the last base-pair of the stem, which is at the end of the loop (see Fig. (5.1)). Then we shall extend the loop by adding the segments forming the stem. In a first step let us ignore the stacking and Morse potential interactions which are specific to the stem and only consider the polymer made by the DNA strand. When we add one base-pair to the stem we add two segments to the polymer. The extended loop with $N + 2$ monomers has now the distance r_{M-1} between its ends. So that its restricted partition function is

$$Z_{N+2}(r_{M-1}) = P_{N+2}(r_{M-1}) Z_{N+2}^{tot}. \tag{5.20}$$

But the probability $P_{N+2}(r_{M-1})$ can be expressed as a function of $P_N(r_M)$ if we introduce the conditional probability $S(\rho'|\rho)$ that if a polymer has the distance ρ between its ends, the polymer with two additional monomers has the distance ρ' between its ends as schematized on Fig. (5.10).

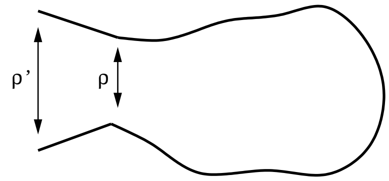

Figure 5.10: *Schematic representation of the growth of the polymer.*

This conditional probability function can in principle be calculated if we have a model for the polymer. We have shown in Chap. 3 how it can be obtained for an effective Gaussian model.
With this function we can express $P_{N+2}(r)$ in term of $P_N(R)$ as

$$P_{N+2}(\rho') = \int d\rho S(\rho'|\rho) P_N(\rho), \tag{5.21}$$

or, in the context of our calculation

$$P_{N+2}(r_{M-1}) = \int dr_M S(r_{M-1}|r_M) P_N(r_M), \tag{5.22}$$

which gives the reduced partition function for a stem with two base-pairs as

$$Z_{N+2}(r_{M-1}) = Z_{N+2}^{tot} \int dr_M S(r_{M-1}|r_M) P_N(r_M). \tag{5.23}$$

The same process can be repeated if we add the third base-pair in the stem. From

$$\begin{aligned} Z_{N+4}(r_{M-2}) &= P_{N+4}(r_{M-2}) Z_{N+4}^{tot} \\ &= Z_{N+4}^{tot} \int dr_{M-1} S(r_{M-2}|r_{M-1}) P_N(r_{M-1}), \end{aligned} \tag{5.24}$$

we get

$$Z_{N+4}(r_{M-2}) = Z_{N+4}^{tot} \int dr_{M-1}\ dr_M\ S(r_{M-2}|r_{M-1}) S(r_{M-1}|r_M) P_N(r_M). \tag{5.25}$$

We can continue the process until we have added $(M-1)$ base-pairs to the one that is next to the loop, in order to get the complete stem, with M base-pairs, which corresponds to the total of $(N+2(M-1))$ monomers in the polymer forming the hairpin.
We get the reduced partition function

$$Z_{N+2(M-1)} = Z_{N+2(M-1)}^{tot} \int_0^{+\infty} dr \prod_{i=2}^{M} S(r_{i-1}|r_i) P_N(r_M). \tag{5.26}$$

Up to now we have ignored the contribution of the Morse potential and stacking interaction. Let us now examine how it enters.
Consider again the loop alone with its terminal base-pair. Due to the Morse potential $V(r_M)$, the probability $P_N(r_M)$ must be multiplied by $e^{-\beta V(r_M)}$. Its reduced partition function is then

$$Z_N(r_M) = e^{-\beta V(r_M)}\ P_N(r_M) Z_N^{tot}. \tag{5.27}$$

When we add one base-pair, i.e. two monomers we add one stacking interaction $W(r_{M-1}, r_M)$ and one Morse potential $V(r_M)$. So that Eq. (5.20) becomes

$$\begin{aligned} Z_{N+2}(r_{M-1}) = Z_{N+2}^{tot} & \\ e^{-\beta V(r_{M-1})} & \int dr_M\ e^{-\beta(W(r_{M-1},r_M)+V(r_M))} S(r_{M-1}|r_M) P_N(r_M). \end{aligned} \tag{5.28}$$

This shows that, in our previous calculation we can formally replace $S(r_{i-1}|r_i)$ by

$$S(r_{i-1}|r_i) \rightarrow S(r_{i-1}|r_i)\exp\left(-\beta\left(V(r_i)+W(r_{i-1},r_i)\right)\right), \tag{5.29}$$

and multiply the final result by the $e^{-\beta V}$ term corresponding to the base-pair closing the system. Therefore the reduced partition function of the hairpin with the interactions in the stem is finally given by

$$\begin{aligned} Z(r) =& Z_{\mathrm{loop}(N+2(M-1))} e^{-\beta V(r_1)} \times \\ & \int_0^{+\infty} \prod_{i=2}^{M} dr_i \prod_{i=2}^{M} S(r_{i-1}|r_i) e^{-\beta[V(r_i)+W(r_{i-1},r_i)]} P_N(r_M), \end{aligned} \tag{5.30}$$

where $r_i = y_i + d$ according to the notations of Fig. (5.1). Note also that $r = r_1$ and $R = r_M$ in these notations. V and W have the following expressions

$$\begin{cases} V(r_i) & = D\left[\left(\exp\left(-\alpha\left(r_i-d\right)\right)-1\right)^2-1\right], \\ W(r_i,r_{i+1}) & = \frac{K}{2}\left[1+\rho\exp\left(-\delta\left(r_i+r_{i-1}-2d\right)\right)\right]\left(r_i-r_{i-1}\right)^2. \end{cases} \tag{5.31}$$

5.3.2 Free Energy and Entropy

It is interesting to see the form of the total free energy as well as the entropy of the system. The free energy is given by

$$F(T) = -k_B T \ln Z, \tag{5.32}$$

where Z is obtained by integrating $Z(r)$ over r

$$Z = \int dr Z(r). \tag{5.33}$$

And the entropy $S(T)$ is given by the first derivative of F

$$S(T) = -\frac{\partial F}{\partial T}. \tag{5.34}$$

Of course the expressions of F and S depend on the model of the loop we are using through $P_N(R)$. However the behavior of the temperature evolution of F and S stays qualitatively the same for different loop models. Figure (5.11) gives the evolution of $F(T)$ and $S(T)$ with temperature for the FRC model and without the growth of the polymer ($S \equiv 1$).

We can see a change of the slope in the free energy around 310 K which could be defined as the melting temperature. The entropy profile shows a sharp increase when the system goes from the closed state to the open one by increasing the temperature. To be more precise we have to derive melting curves as well as rates of opening and closing for different parameters of the model and different loop models. Before doing that we present the derivation of the rates of opening and closing in the case of an equilibrium between the open and the closed state with a transition state between the two.

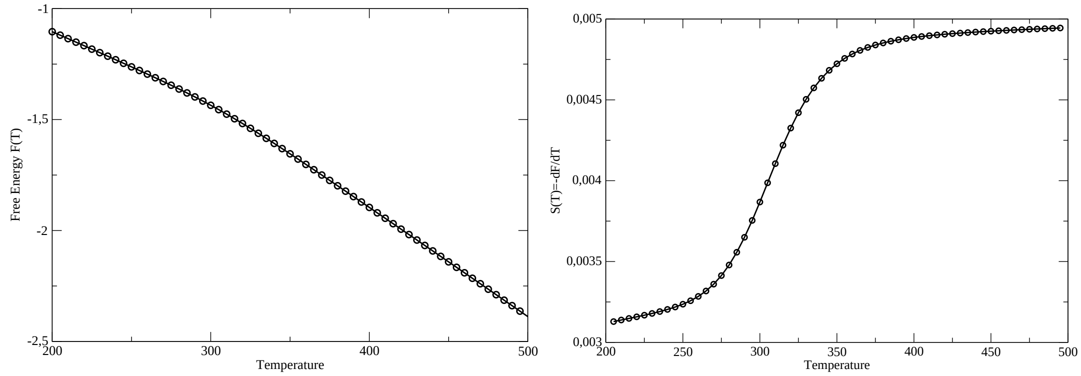

Figure 5.11: *Example of free energy profile and entropy with the FRC model for the loop.The parameters of the stem are: D=0.107 eV, k=0.025 eV.Å^{-2}, α=6.9 Å^{-1}, $\delta = 0.35$, $\rho = 5$, $\theta = 45°$ and N=21. Left: Free energy. Right: Entropy calculated by* $S(T) = \frac{\partial F}{\partial T}$

5.3.3 Kinetics: theoretical predictions

In order to study the kinetics of the opening-closing fluctuations, we view them from the point of view of a chemical equilibrium between two states (C closed, O open) separated by a transition state (T) as schematized on Fig. (5.12)

$$\mathbf{C} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathbf{T} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathbf{O}$$

Figure 5.12: *Chemical equilibrium.*

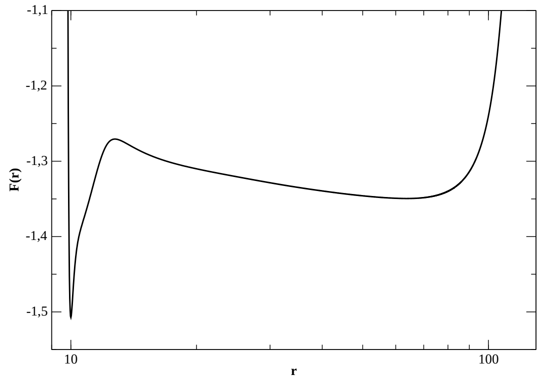

Figure 5.13: *Example of a free energy profile obtained with S ≡1 and a loop modeled by the FRC. The parameters are the following: D=0.107 eV, k=0.025 eV.Å^{-2}, α=6.9 Å^{-1}, δ=0.35, ρ=5, θ=45° and N=21.*

Here k_1, k_{-1}, k_2 and k_{-2} designate the kinetic constants. Let us denote by C with indices C, T, O the concentrations of the different species. Therefore we have

$$\begin{aligned}\dot{C}_C &= -k_1 C_C + k_{-1} C_T \\ \dot{C}_T &= -\left(k_{-1} + k_2\right) C_T + k_1 C_C + k_{-2} C_O \\ \dot{C}_O &= -k_{-2} C_O + k_2 C_T.\end{aligned} \tag{5.35}$$

We then assume that the concentration of the transition state stays constant. This is the quasi-stationary state approximation:

$$\dot{C}_T = 0. \tag{5.36}$$

Then we get

$$C_T = \frac{k_1 C_C + k_{-2} C_O}{k_{-1} + k_2}. \tag{5.37}$$

Now if we insert Eq. (5.37) in (5.35) we get

$$\begin{aligned}\dot{C}_C &= -k_1 C_C + k_{-1} \frac{k_1 C_C + k_{-2} C_O}{k_{-1} + k_2} \\ &= -\frac{k_1 k_2}{k_{-1} + k_2} C_C + \frac{k_{-1} k_{-2}}{k_{-1} + k_2} C_O \\ \dot{C}_C &= -k_f C_C + k_b C_O,\end{aligned} \tag{5.38}$$

where k_f and k_b are the rates of opening and closing, respectively, we would like to derive. The assumption (5.36) amounts to assuming $k_{-1} + k_2 >> k_b,\ k_f$, which means that the stationary state for T is reached because the time scales for going in and out of the transition state are shorter than the time scales to open or close. Moreover

$$\dot{C}_C + \dot{C}_O = 0, \tag{5.39}$$

and at the equilibrium $\dot{C}_C = \dot{C}_O = 0$, so that

$$\frac{\bar{C}_C}{\bar{C}_O} = \frac{k_b}{k_f} = \frac{k_1 k_{-2}}{k_{-1} k_2}. \tag{5.40}$$

Finally we obtain

$$k_f^{-1} = k_1^{-1} + \frac{\bar{C}_C}{\bar{C}_O} k_{-2}^{-1} \tag{5.41}$$

$$k_b^{-1} = k_{-2}^{-1} + \frac{\bar{C}_O}{\bar{C}_F} k_1^{-1}. \tag{5.42}$$

The ratio in Eq. (5.46) is given by thermodynamics

$$\frac{\bar{C}_C}{\bar{C}_O} = \frac{Z_C}{Z_O}. \tag{5.43}$$

The opening-closing of a hairpin is a complex process involving many degrees of freedom but in the spirit of our equilibrium thermodynamics calculation, it is natural to introduce a reaction coordinate r, which is the distance between the ends of the hairpin.

In this spirit, we can consider that the system is evolving on a one-dimensional free energy surface, which has the qualitative shape shown in Fig. (5.13). The closed and open states are minimum of this surface $F(r)$ and the transition state corresponds to the maximum. We can select the origin so that the transition state is at $r = 0$. In term of the free energy $F(r)$ the partition functions for the closed and the open states are

$$Z_C = \int_{-\infty}^{0} dr e^{-\beta F(r)} \tag{5.44}$$

$$Z_O = \int_{0}^{\infty} dr e^{-\beta F(r)}, \tag{5.45}$$

and the kinetics of the opening-closing fluctuations is an evolution on this free energy surface, which can be described by a Fokker-Planck formalism. Therefore we have to derive the expression of k_1 and k_{-2} to get the rates of opening and closing.

To do that we suppose that the system diffuses on the one dimensional effective potential and we would like to know the mean passage time [75] for the system which is in one of the two wells to go in the other one through the barrier. If we call $P(r)$ the probability distribution, i.e. $P(r)dr$ is the probability of the system to be in the range $[r, r + dr]$, it obeys to the usual Fokker-Planck equation:

$$\begin{cases} \frac{\partial P}{\partial t} & = -\frac{\partial j(r)}{\partial r} \\ j(r) & = -D(r)\left[\frac{\partial P}{\partial r} + \beta F' P\right] . \end{cases} \tag{5.46}$$

We assume some boundary conditions associated to our problem:

- Reflecting boundary also to the left: $r \to -\infty$: $\lim_{r\to-\infty} j(r,t) = 0 \ \forall \ t$. In practice we use a hard core at r=9.7 Å.

- Absorbing boundary in $r = r_{max}$: $j(r_{max}, t) = \Lambda P(r_{max}, t)$ with $\Lambda \to +\infty$ which means that once it has passes the maximum the system evolves to the second minimum.

The mean first passage time is given by [76]

$$\tau = \int_{0}^{+\infty} dt \int_{-\infty}^{r_{max}} dr P(r,t). \tag{5.47}$$

First of all let's integrate Eq. (5.46) over r:

$$\int_{-\infty}^{r'} \frac{\partial P(r,t)}{\partial t} dr = -j(r', t),$$

so that

$$j(r_{max},t) = \Lambda P(r_{max},t) = -\int_{-\infty}^{r_{max}} \frac{\partial P(r,t)}{\partial t} dr. \tag{5.48}$$

Using Eq. (5.46), we also get

$$\begin{aligned}\int_{-\infty}^{r'} \frac{\partial P(r,t)}{\partial t} dr &= D(r') \left[\frac{\partial P}{\partial r'} + \beta F' P \right] \\ &= D(r') e^{\beta F} \frac{\partial}{\partial r'} \left(e^{\beta F} P \right). \end{aligned} \tag{5.49}$$

Now we can integrate (5.49) over r'

$$\begin{aligned}\int_{R}^{r_{max}} dr' \frac{\partial}{\partial r'} \left(e^{\beta F} P \right) &= \int_{R}^{r_{max}} \frac{dr'}{D(r') e^{-\beta F}} \int_{-\infty}^{r'} dR' \frac{\partial P(R',t)}{\partial t} \\ e^{\beta F(r_{max})} P(r_{max},t) - e^{\beta F(R)} P(R,t) &= \int_{R}^{r_{max}} \frac{dr'}{D(r') e^{-\beta F}} \int_{-\infty}^{r'} dR' \frac{\partial P(R',t)}{\partial t}. \end{aligned} \tag{5.50}$$

Putting Eq. (5.48) in Eq. (5.50)

$$\begin{aligned} P(R,t) = &-\frac{e^{-\beta F(R)}}{e^{-\beta F(r_{max})}} \frac{1}{\Lambda} \int_{-\infty}^{r_{max}} dR' \frac{\partial P(R',t)}{\partial t} - \\ &e^{-\beta F(R)} \int_{R}^{r_{max}} \frac{dr'}{D(r') e^{-\beta F}} \int_{-\infty}^{r'} dR' \frac{\partial P(R',t)}{\partial t}, \end{aligned} \tag{5.51}$$

and putting

$$p_0(R) = \frac{e^{-\beta F(X)}}{\int_{-\infty}^{r_{max}} dR e^{-\beta F(R)}},$$

with $\int_{-\infty}^{r_{max}} p_0(R)\, dR = 1$, we get

$$\begin{aligned} P(R,t) = &-\frac{P_0(R)}{P_0(r_{max})} \frac{1}{\Lambda} \int_{-\infty}^{r_{max}} dR' \frac{\partial P(R',t)}{\partial t} - \\ &P_0(R) \int_{R}^{r_{max}} \frac{dr'}{D(r') P_0(r')} \int_{-\infty}^{r'} dR' \frac{\partial P(R',t)}{\partial t}. \end{aligned} \tag{5.52}$$

Now let us integrate Eq. (5.52) over R and t which is exactly the definition of τ that we are looking for

$$\begin{aligned} \tau &= \int_{0}^{\infty} dt \int_{-\infty}^{r_{max}} dR\, P(R,t) \\ \tau &= \frac{1}{\Lambda P_0(r_{max})} \int_{-\infty}^{r_{max}} dy P(y,0) + \\ &\quad \int_{-\infty}^{r_{max}} dx P_0(x) \int_{x}^{r_{max}} \frac{dr}{D(r) P_0(r)} \int_{-\infty}^{r} dy P(y,0) \end{aligned} \tag{5.53}$$

where we have assumed that $\lim_{t\to+\infty} P(y,t) = 0 \ \forall \ y$. At $t=0$ let us assume that the system is at the thermodynamic equilibrium, so that $P(y,0) = P_0(y)$, then

$$\begin{aligned}
\tau &= \frac{1}{\Lambda P_0(r_{max})} + \int_{-\infty}^{r_{max}} dx P_0(x) \int_x^{r_{max}} \frac{dr}{D(r)P_0(r)} \int_{-\infty}^{r} dy P_0(y) \\
&= \frac{1}{\Lambda P_0(r_{max})} + \int_{-\infty}^{r_{max}} dx P_0(x) \int_x^{r_{max}} dr H(r) \\
&= \frac{1}{\Lambda P_0(r_{max})} + \int_{-\infty}^{r_{max}} dx P_0(x) \int_{-\infty}^{r_{max}} dr H(r) \Theta(r-x) \\
&= \frac{1}{\Lambda P_0(r_{max})} + \int_{-\infty}^{r_{max}} dr H(r) \int_{-\infty}^{r} dx P_0(x) \\
&= \frac{1}{\Lambda P_0(r_{max})} + \int_{-\infty}^{r_{max}} dr \frac{1}{D(r)P_0(r)} \int_{-\infty}^{r} dy P_0(y) \int_{-\infty}^{r} dx P_0(x) \\
\tau &= \frac{1}{\Lambda P_0(r_{max})} + \int_{-\infty}^{r_{max}} dr \frac{1}{D(r)P_0(r)} \left(\int_{-\infty}^{r} dx P_0(x) \right), \qquad (5.54)
\end{aligned}$$

where $\Theta(x)$ is the Heaviside function. Finally, taking $\Lambda \to +\infty$, we get

$$\begin{cases} \tau & = \int_{-\infty}^{r_{max}} \frac{dr}{D(r)P_0(r)} I^2(r) \\ I(r) & = \int_{-\infty}^{r} dx P_0(x). \end{cases} \qquad (5.55)$$

Now we can apply the expression of τ to our special case

$$k_1^{-1} = \tau_{CT} = \int_{-\infty}^{r_T} dr \frac{I_C^2(r)}{D(r)P_0^{(C)}(r)}, \qquad (5.56)$$

with

$$I_C(r) = \int_{-\infty}^{r} dx P_0^{(F)}(x),$$

and

$$P_0^{(F)}(r) = \frac{e^{-\beta F(r)}}{Z_C} \ \forall \ r < r_T.$$

We also need the expression of k_{-2}^{-1}

$$k_{-2}^{-1} = \tau_{OT} = \int_{r_T}^{\infty} dr \frac{I_O^2(r)}{D(r)P_0^{(O)}(r)}, \qquad (5.57)$$

with

$$I_O(r) = \int_{r}^{\infty} dx P_0^{(O)}(x),$$

and

$$P_0^{(O)}(r) = \frac{e^{-\beta F(r)}}{Z_O} \ \forall \ r > r_T.$$

Therefore

$$k_f^{-1} = \int_{-\infty}^{r_T} dr \frac{I_C^2(r)}{D(r)P_0^{(C)}(r)} + \frac{Z_C}{Z_O} \int_{r_T}^{\infty} dr \frac{I_O^2(r)}{D(r)P_0^{(O)}(r)}$$
$$k_f^{-1} = Z_C \left(\int_{-\infty}^{r_T} \frac{dr}{Z_C} \frac{I_C^2(r)}{D(r)P_0^{(C)}(r)} + \int_{r_T}^{\infty} \frac{dr}{Z_O} \frac{I_O^2(r)}{D(r)P_0^{(O)}(r)} \right)$$
$$k_f^{-1} = Z_C \int_{-\infty}^{+\infty} dr \frac{e^{\beta F(r)} I^2(r)}{D(r)}, \tag{5.58}$$

with

$$I(r) = \begin{cases} \int_{-\infty}^{r} dx \frac{e^{-\beta F(x)}}{Z_C} & \forall\ r < r_T \\ \\ \int_{r}^{+\infty} dx \frac{e^{-\beta F(x)}}{Z_O} & \forall\ r > r_T. \end{cases} \tag{5.59}$$

Finally $k_b^{-1} = \frac{Z_O}{Z_C} k_f^{-1}$. In order to avoid numerical problems during integrations we transform Eq. (5.58) as

$$k_f^{-1} = Z_C \int_{-\infty}^{+\infty} dr \frac{e^{-\beta F(r)} J^2(r)}{D(r)}, \tag{5.60}$$

with

$$J(r) = \begin{cases} \int_{-\infty}^{r} dx \frac{e^{-\beta (F(x)-F(r))}}{Z_C} & \forall\ r < r_T \\ \\ \int_{r}^{+\infty} dx \frac{e^{-\beta (F(x)-F(r))}}{Z_O} & \forall\ r > r_T. \end{cases} \tag{5.61}$$

5.4 Case of S≡1

In order to get a first idea of the behavior of the hairpin, it is convenient to start from a zeroth-order approximation in which the stem and the loop are decoupled in the calculation. This can be obtained if we set $S \equiv 1$ in the general expression (5.29). This approximation simply replaces $e^{-\beta V(r_M)}$ by $e^{-\beta V(r_M)} P_N(r_M)$ in the expression for the stem alone. Strictly speaking this is not correct because the transformation gives an expression of $Z(r)$ which does not have the expected dimension for a reduced partition function. We nevertheless introduce this approximation as a preparation for the discussion of the complete calculation of Section 5, keeping in mind that it can only give the general behavior of $Z(r)$, up to a factor. In this case, the reduced partition function is given by

$$Z(r) = e^{-\beta U(r)} \int \prod_{i=2}^{M-1} dr_i \int dr_M P_N(r_M) T(r_M - d, r_{M-1}) \cdots T(r_2, r - d), \tag{5.62}$$

where $T(r_i, r_{i-1}) = \exp\left(-\beta \left[V(r_i) + W(r_i, r_{i-1})\right]\right)$ and $U(r) = V(r - d)$.

5.4.1 Thermodynamics

The free energy landscape $F(r) = -k_b T \ln\ Z(r)$, with $Z(r)$ defined by (5.54) has the shape plotted in Fig. (5.13).
It is interesting to compare this figure to Fig. (5.9) for the stem alone. In the presence of the loop besides, the deep minimum around r=10 Å, we have a second minimum for large values of r. One can understand its presence in term of the entropy of the loop. The idea is similar to rubber elasticity. When the loop is stretched it can only occupy a small number of conformations and thus has a lower entropy. When r increases the loop can access many configurations and its entropy increases, hence decreasing the free energy. But whatever the loop model, too low values of r also lead to a penalty in free energy. For the Kratky-Porod chain model the penalty is energetic, while for the FRC very low values of r again reduce the number of configurations or are even not accessible. This explains why, when r decreases below r_2 the free energy raises gain to a maximum for $r = r_c$ before the large drop at $r = r_1$ which is due to the large energy gain when the hydrogen bonds in the stem are formed.
This shape of the curve $F(r)$ justifies the image of the two-state system that we have used for the kinetics. Those states are the closed state for $r \approx r_1$ and the open state for $r \approx r_2$. In the view of a chemical equilibrium between the two states, one can define an equilibrium constant

$$K_{eq} = \frac{P_O}{P_C}. \tag{5.63}$$

Where, P_O, and P_C are the probabilities to be open or closed, respectively. We define the probabilities by

$$P_O = \frac{\int_{r_c}^{+\infty} dr Z(r)}{\int_0^{+\infty} dr Z(r)}, \tag{5.64}$$

and $P_C + P_0 = 1$. The parameter r_c is the value of the reaction coordinate at the maximum of the free energy (transition state) between the two wells which corresponds to the open and the closed state. Then the melting curves which are equivalent to the normalized fluorescence measured in the experiments are given by P_O. Indeed, we have

$$f = \frac{K_{eq}}{1 + K_{eq}} = \frac{\frac{P_O}{P_C}}{1 + \frac{P_O}{P_C}} = P_O. \tag{5.65}$$

Let us now give a first qualitative view of the properties of the hairpin as a function of the model parameters. A more quantitative picture will be given for $S \neq 1$ but this first approach is useful to get an idea of the separate influence of the loop and stem.

5.4.1.1 Role of the loop

FRC model First of all we propose to compare the melting curve obtained for a stem of five base-pairs and with and without loop to see its effect. Figure (5.14) gives such a comparison.

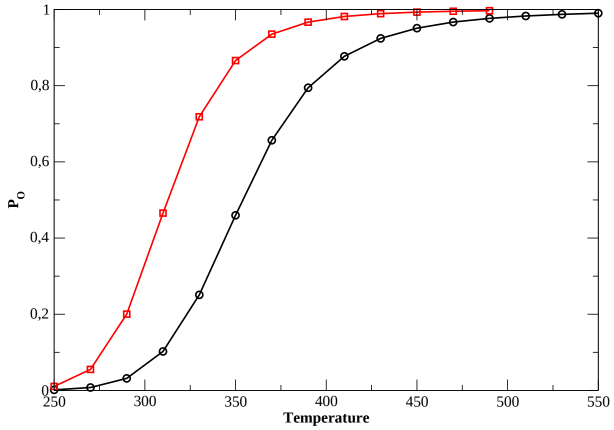

Figure 5.14: *Melting curve obtained for a stem of five base-pairs with and without a loop. The loop is described by the FRC model. The black curve corresponds to the stem alone.*

We see that the stem tends to open at lower temperatures in presence of the loop which is due to the additional entropy brought by the loop. Therefore T_m is smaller for the hairpin than for a stem alone. Moreover the transition is a bit sharper in the case of the hairpin but this is not a strong effect. The results are summarized in the next table

	T_m	$\frac{\Delta P}{\Delta T}T_m$
stem	350	3.9
stem+loop	325	3.1

where we indicate the melting temperature and the quantity $\frac{\Delta P}{\Delta T}T_m$ which is a dimensionless measure of the slope at T_m, multiplied by T_m to get a dimensionless quantity. It measures the width of the transition.

In order to study the effect of the loop in more details, we now present the results obtained by varying the properties of the loop. Figure (5.15) and (5.16) give the melting curves for different loop lengths as well as the evolution of T_m for two different fixed angles θ. First of all, for the two values of θ the melting temperature T_m decreases with the loop length. The decrease is most important for $\theta = 60°$. T_m varies from 350 K to 323 K for N going from 12 to 30 but for $\theta = 45°$, ΔT_m=15 K only. Secondly, for the same value of the loop length, T_m decreases with decreasing θ. Theses results are in qualitative agreement with some of the experimental results. Indeed T_m is smaller for Poly(A) than Poly(T) for the same loop length. The stacking interaction which is expected to be more important in the case of A-sequence is equivalent to smaller values of θ because it maintains the chain more rigid. Moreover, the larger the loop length, the larger the entropy, which tends to destabilize the hairpin configuration. However the model is not fully satisfactory because the observed variation ΔT_m of the melting temperature is larger for poly (A) than poly(T) which is not the results given by the model. We must also notice that the width of the transition given by the model is about 100 K which is much larger than in the experiments.

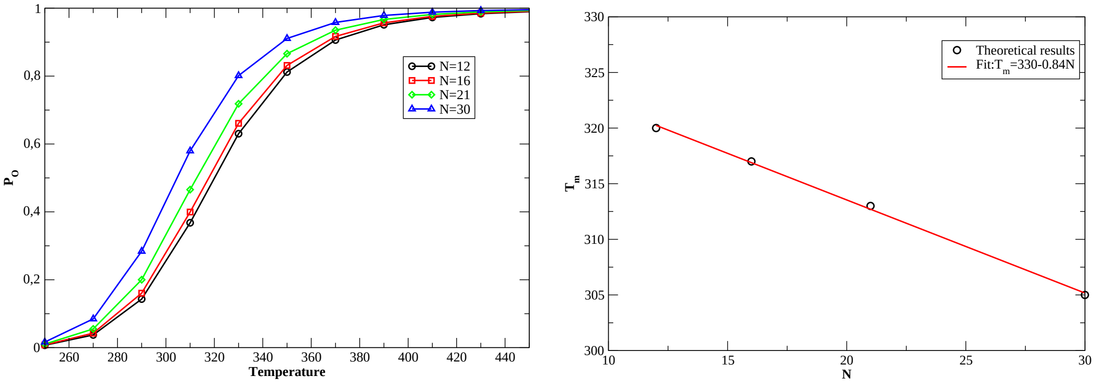

Figure 5.15: *Melting curves with the FRC model:* $\theta = 45°$. *The parameters of the stem are:* D=*0.107 eV,* k=*0.025 eV.*$Å^{-2}$, α=*6.9* $Å^{-1}$, $\delta = 0.35$, $\rho = 5$, $\theta = 45°$. *Left: Melting profiles,* ○: *N=12;* □: *N=16;* ◇: *N=21;* △: *N=30. Right: evolution of the melting temperature with N.* ○: *theoretical results, line: linear fitting.*

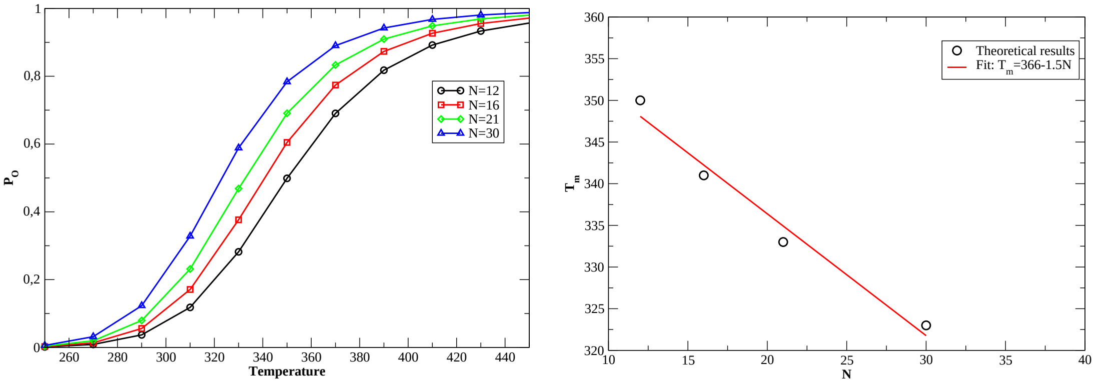

Figure 5.16: *Melting curves with the FRC model:* $\theta = 60°$. *The parameters of the stem are:* D=*0.107 eV,* k=*0.025 eV.*$Å^{-2}$, α=*6.9* $Å^{-1}$, $\delta = 0.35$, $\rho = 5$, $\theta = 60°$. *Left: Melting profiles,* ○: *N=12;* □: *N=16;* ◇: *N=21;* △: *N=30. Right: evolution of the melting temperature with N.* ○: *theoretical results, red line: linear fitting.*

Discrete Kratky-Porod chain If we change the model of the loop, it is interesting to see the change in the thermodynamics. Let us now consider the discrete version of the Kratky-Porod chain as we presented in Chap. 3 which includes an additional energetic contribution in the probability distribution of the end-to-end distance. Figures (5.17) and (5.18) give the melting profiles and the melting temperature T_m for different loop lengths and for two different values of the rigidity parameter ϵ.

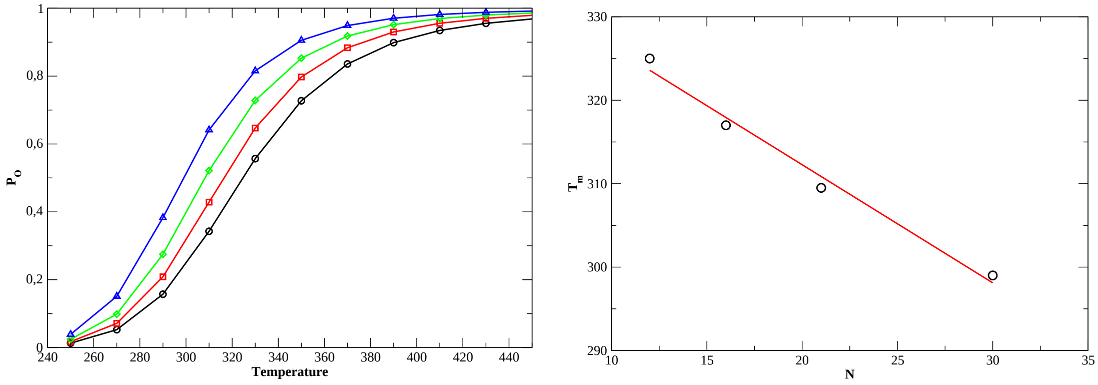

Figure 5.17: *Melting curves with the Kratky-Porod chain:* ϵ*=0.0019 eV.*Å^{-2}*. The parameters of the stem are: D=0.102 eV, k=0.025 eV.*Å^{-2}*,* α*=6.9* Å^{-1}*,* $\delta = 0.35$*,* $\rho = 5$*,* ϵ*=0.0019 eV.*Å^{-2}*. Left: Melting profiles, ●: N=12; □: N=16; ◇: N=21; △: N=30. Right: evolution of the melting temperature with N. ○: theoretical results, line: linear fitting.*

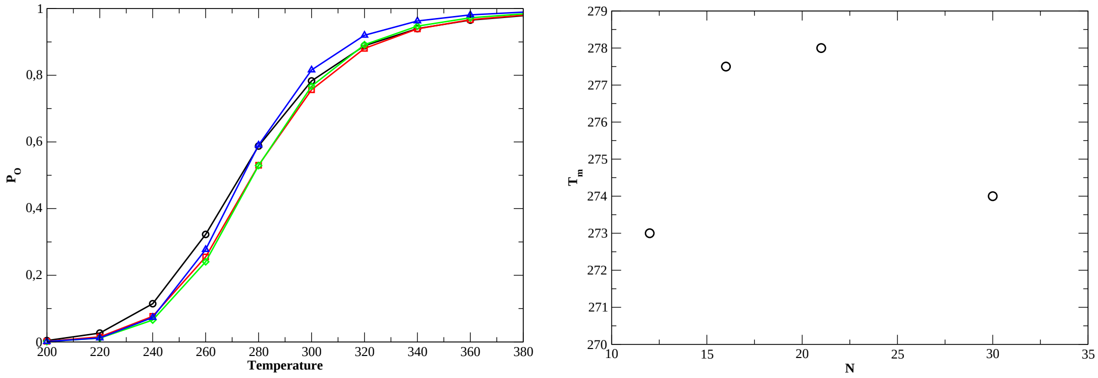

Figure 5.18: *Melting curves with the Kratky-Porod chain* ϵ*=0.0040 eV.*Å^{-2}*. The parameters of the stem are: D=0.107 eV, k=0.025 eV.*Å^{-2}*,* α*=6.9* Å^{-1}*,* $\delta = 0.35$*,* $\rho = 5$*,* ϵ*=0.0040 eV.*Å^{-2}*. Left: Melting profiles, ●: N=12; □: N=16; ◇: N=21; △: N=30. Right: evolution of the melting temperature with N. ○: theoretical results.*

For ϵ=0.0019.eV.Å^{-2} we find the correct tendency: T_m decreases with the length of the loop as in the case of the FRC and the experiments. T_m varies from 325 K to 299 K for N going from 12 to 30 which is comparable to the experimental results. However for ϵ=0.0040 eV.Å^{-2}, we obtain something quite surprising because the evolution of T_m as a function of N is not monotonous. Indeed, for N going from 12 to 21 T_m increases and for N higher than 21 it decreases. As ϵ is large, the probability to form small loops, which are necessary to form hydrogen bonds in the stem, is very small. Consequently the phase space corresponding to the closed configuration is smaller. But when we increase the number of monomers in the loop, even if ϵ is large, the tendency to get a closed loop is higher, which allows the formation of base-pairs in the stem. To see this effect, Fig. (5.19) gives the end-to-end probability distribution of the Kratky-Porod chain for different loop lengths and for two different values of ϵ.

For ϵ=0.0019 eV.Å^{-2}, near the equilibrium distance of the hydrogen bonds (10 Å approximately), for smaller N we get a larger end-to-end probability that tends to stabilize the hairpin configuration. On the contrary, for the case of ϵ=0.0040 eV.Å^{-2} there is an inversion of this phenomenon for $N < 21$. For $N < 21$, reducing N reduces the value of the end-to-end probability distribution for small R, whereas for $N > 21$, reducing N increases the end-to-end probability distribution at R small. That explains the evolution of T_m as a function of N.

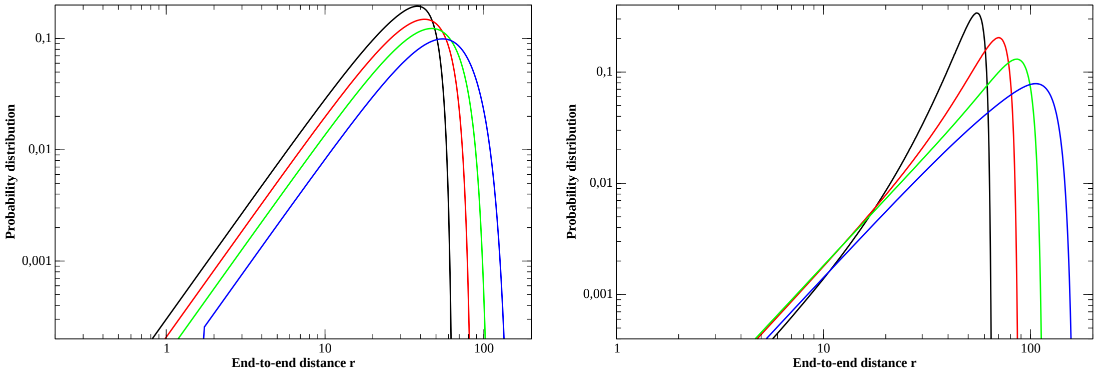

Figure 5.19: *Plot of the probability distribution of the Kratky-Porod chain. Left:T=330 K, ϵ=0.0019 eV.Å^{-2}; black: N=12; red: N=16; green: N=21; blue: N=30. Right: T=275 K, ϵ=0.0040 eV.Å^{-2}; black: N=12; red: N=16; green: N=21; blue: N=30*

5.4.1.2 Role of the stem

Let us now study the effect of the stem parameters on the properties of the hairpins. Figure (5.20) gives the evolution of the melting curves with the change of D and k, and Fig. (5.21) shows the same quantity but with the change of α and ρ. First when we increase the value of D, which is the depth of the Morse potential, the closed conformation is more stable and the transition to the open state takes place at higher temperatures as expected because the thermal fluctuations must be large enough to allow the system to overcome the free energy barrier represented in Fig. (5.12). Second, when we change the value of k, we affect the rigidity of the stem and the larger k, the larger the rigidity. Then, as for the stem alone, the closed configuration is more stable for larger values of k and the equilibrium is shifted to higher temperatures. Only the kinetic results will tell us if this evolution should be attributed to entropic or energetic effects. The value of ρ has a small effect on the melting profiles because we consider short stems such as the five base-pairs stem. This is different from the effect of ρ on the double stranded DNA. For long double helices large values of ρ lead to a large entropy increase when some regions are on the plateau of the Morse potential, and thereby lead to a sharper transition. Finally we see that the bigger the width of the Morse potential (small values of a), the larger the melting temperature T_m. When we increase the width of the Morse potential, we also increase the width of the first well of the free energy landscape which represent the closed configuration. Thus the closed conformation is more stable and the system again needs more thermal fluctuations to open. In fact we find qualitatively the same influence of the parameters on T_m as in the long dsDNA with a square potential and a linear stacking: $T_m \sim \frac{\sqrt{kD}}{\alpha}$. To finish with this part we also give the influence of ϵ as well as the influence of D on the melting profiles with the Kratky-Porod chain in Fig. (5.22). For the influence of D we get the same dependence as in the FRC case. Moreover, the effect of ϵ is comparable to the effect of θ in FRC, the bigger the value of ϵ, the smaller T_m. Therefore when we increase the rigidity, the hairpin is subjected to forces from the loop part which tend to destabilize it.

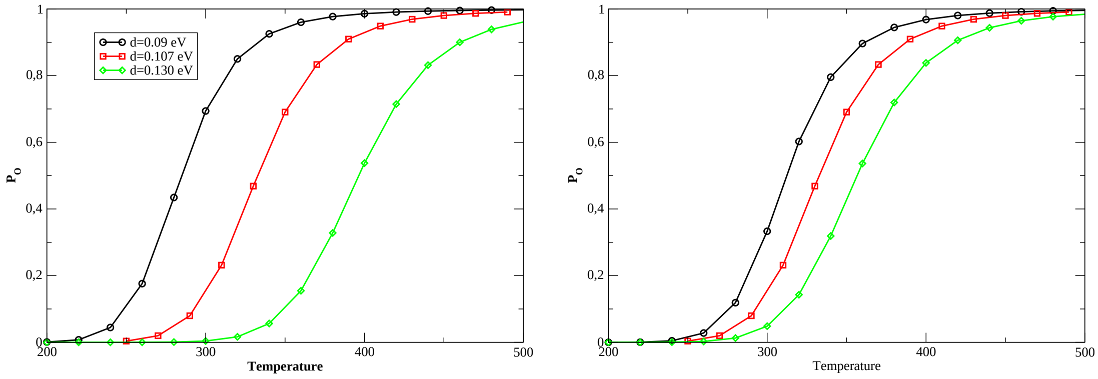

Figure 5.20: *Effect of D and k on the melting curve. The parameters are the following: α=6.9 Å^{-1}, δ = 0.35, ρ = 5, θ = 60°, N=21. Left: Effect of d, k=0.025 eV.Å^{-2}; •: D=0.09 eV; □: D=0.107 eV; ◇: D=0.13 eV. Right: Effect of k, D=0.107 eV; • k=0.013 eV.Å^{-2}; □ k=0.025 eV.Å^{-2}; ◇ k=0.050 eV.Å^{-2}.*

5.4.2 Kinetics

Let us discuss the kinetic results for the two models of the loop that we studied.

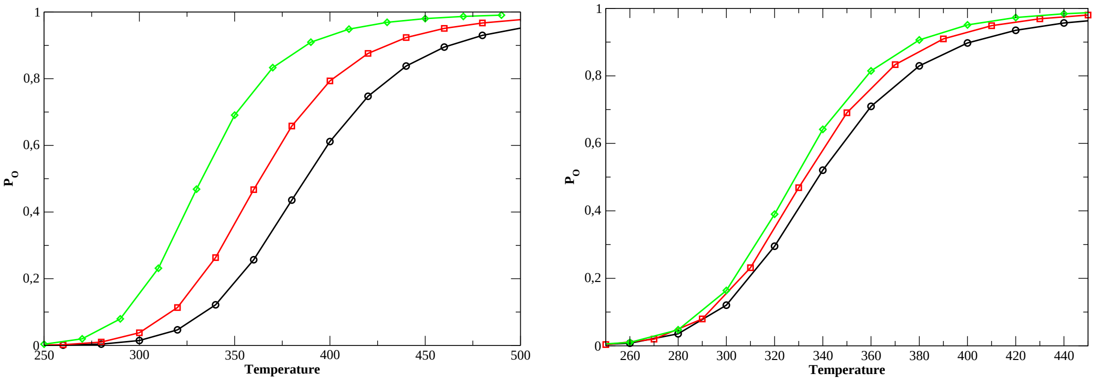

Figure 5.21: *Effect of α and ρ on the melting curve. The parameters are the following: D=0.107 eV, k=0.025 eV.Å^{-2}, α=6.9 Å^{-1},* $\delta = 0.35$, $\theta = 60°$. *Left: Effect of α, ●: α=4.0 Å^{-1}, ; □: α=5 Å^{-1}; ◇: α=6.9 Å^{-1}. Right: Effect of ρ, ●:* $\rho = 2$; *□: ρ=5; ◇: ρ=10.*

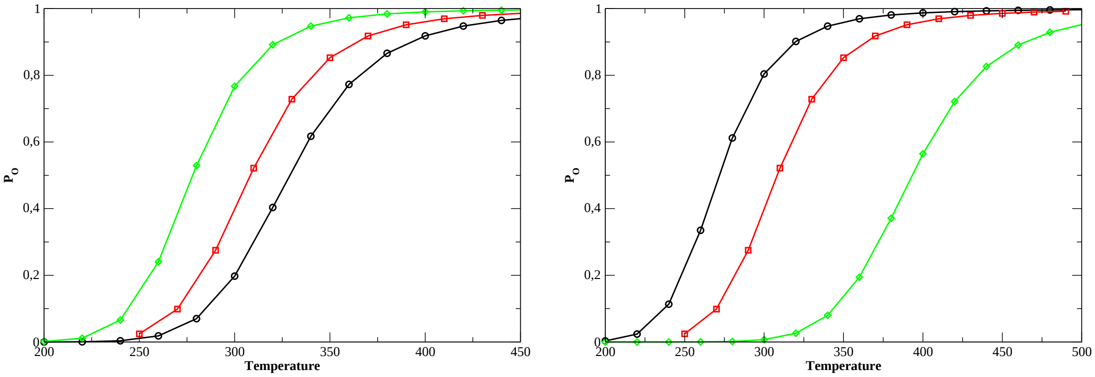

Figure 5.22: *Effect of ϵ and D on the melting profiles. The parameters are: k=0.025 eV.Å^{-2}, α=6.9 Å^{-1},* $\delta = 0.35$, $\rho = 5$, *N=21. Left: D=0.102 eV; ●: ϵ=0.0010 eV.Å^{-2} ; □: ϵ=0.0019 eV.Å^{-2} ; ◇: ϵ=0.0040 eV.Å^{-2} . Right: ϵ0.0019 eV.Å^{-2}; ●: D=0.09 eV; □: D=0.102 eV; ◇: D=0.13 eV*

5.4.2.1 FRC model

The effects of the length of the loop and of the θ angle of the FRC model are shown on Fig. (5.23) which displays the kinetic constants k_{op} and k_{cl} versus temperature in a semi-logarithmic plot.

The main points which appear on the curves are the followings

1. the variation of both constants is linear on this plot, showing that they obey Arrhenius laws
$$k_{op} \approx e^{-\frac{E_{op}}{k_B T}} \text{ and } k_{cl} \approx e^{-\frac{E_{cl}}{k_B T}}. \tag{5.66}$$

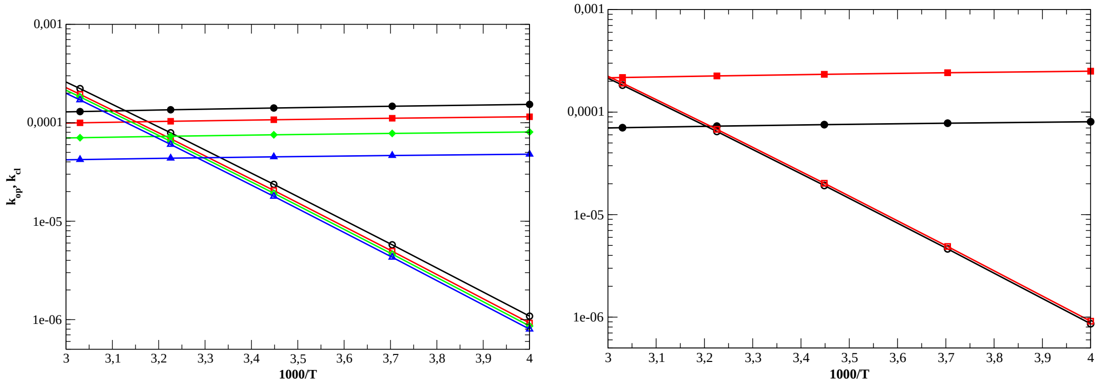

Figure 5.23: *Rates of opening and closing with the FRC model in an Arrhenius plot. Open and closed symbols represent the rates of opening and closing, respectively. The parameters are: D=0.107 eV, k=0.025 eV.Å*$^{-2}$*, α=6.9 Å*$^{-1}$*,* $\delta = 0.35$*,* $\rho = 5$*. Left:*$\theta = 45°$*;* $\bullet$*: N=12;* $\square$*: N=16;* $\diamond$*: N=21;* $\triangle$*: N=30. Right: N=21; black:* $\theta = 45°$*, red:*$\theta = 60°$

2. Changing the loop parameters (loop length N and θ angle of the FRC model) does not affect the kinetics of the opening. This means that the opening is only determined by the stem in this model.

3. The opening activation energy E_{op} is positive, i.e. the transition state has a higher energy than the closed one, in agreement with the experiments. This is consistent with point (2) because E_{op} can be viewed as the energy necessary to break the base-pairing in the stem.

4. The closing activation energy is negative. This implies that the energy of the transition state is lower than the energy of the open state. There is nevertheless a free energy barrier for closing, but it can only come from entropy effects. Going from the open to the transition state leads to an energy gain, which must be attributed to the stem because the freely rotating chain model of the loop has no energetic contribution. This is confirmed by the independence of the slope E_{cl} from the change of the loop parameters N or θ. But the entropy of the open state is much higher than the entropy of the transition state because the open loop can explore a much larger domain of the phase space.

Fig. (5.23) shows that longer loops lead to longer closing times (smaller k_{cl}). This is consistent with the entropic role of the loop. Longer loop lengths increase the phase space accessible to the system and the time that it needs to explore this phase space before reaching the transition state. The role of θ can also be understood in the same framework. When we increase θ the loop is less constrained when it forms the transition state. It can form this closed state in more manners than when θ is

lower, i.e. it has a higher entropy at the transition state. As a result the closing rate is higher for larger values of θ. The variation of k_{op} and k_{cl} with other parameters confirms the conclusions that we have drawn from the study of N and θ.
As shown in Fig. (5.24) a variation of D and k has little effect on the closing rate because closing is mostly controlled by the entropy of the loop. On the contrary the variation of D and k significantly influences the opening which is controlled by the stem. Raising D increases the depth of the free energy well associated to the closed state. Therefore it increases E_{op} and slows down the opening. Changing k we notice only a very small effect on E_{op}. This seems surprising because k enters into an energetic term in the stem and therefore we would expect it to play a role in the opening. We will come back to this point in the complete calculation ($S \neq 1$) but we can anticipate on this discussion by noticing that the coupling along the DNA strand is weak. Most of the energetic contribution lies in the Morse potential, i.e. in the contribution of D. But k has nevertheless an entropic role. Increasing k decreases the opening rate. This can be understood because the opening of the stem gives more freedom to its components to fluctuate. Therefore there is an entropy gain. This entropy gain is smaller when k increases because the relative motions of the elements of the stem are more constrained. This explains why opening is slower for larger k.

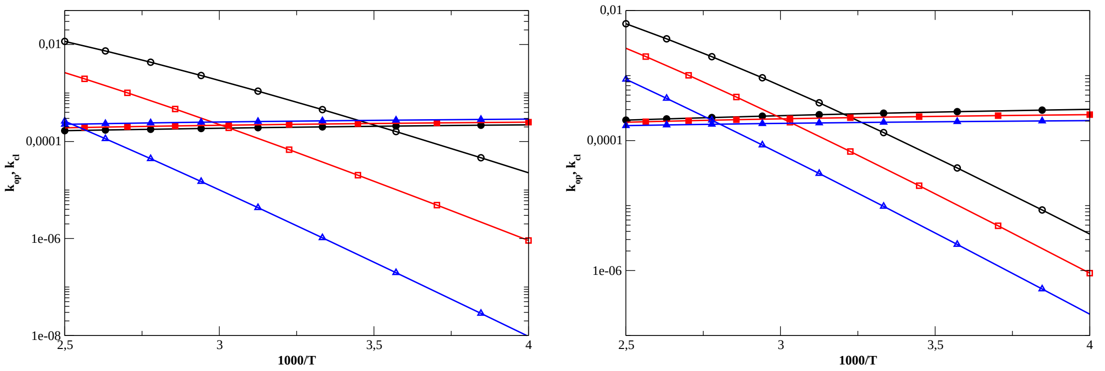

Figure 5.24: *Effect of D and k on the kinetics with the FRC model in an Arrhenius plot. Open and closed symbols represent the rates of opening and closing, respectively. The parameters are:* α*=6.9* $\AA^{-1}$, δ = 0.35, ρ = 5, *N=21. Left: k=0.025 eV.*$\AA^{-2}$*; ●: D=0.009 eV; □: D=0.107 eV; ◇: D=0.130 eV. Right: D=0.107 eV; ●: k=0.013 eV.*$\AA^{-2}$*; □: k=0.025 eV.*$\AA^{-2}$*; ◇: k=0.050 eV.*$\AA^{-2}$

5.4.2.2 Discrete Kratky-Porod chain

Figures (5.25) and (5.26) show the kinetic results for the Kratky-Porod model of the loop. They confirm and complete the analysis that we made from the FRC model. As for the FRC model we see that a change of the parameters of the loop mainly affects closing (Fig. 5.25). The main difference is that the closing activation

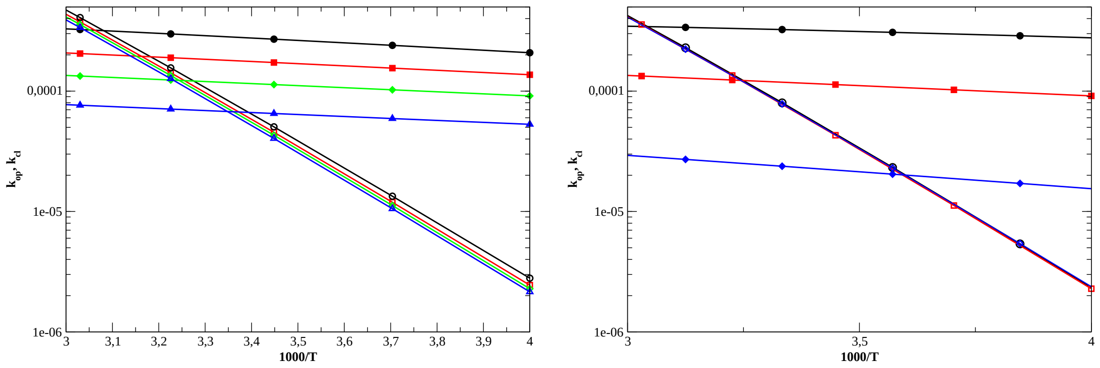

Figure 5.25: *Rates of opening and closing with the Kratky-Porod chain in an Arrhenius plot. Open and closed symbols represent the rates of opening and closing, respectively. The parameters are: D=0.102 eV, k=0.025 eV.*Å^{-2}, α*=6.9* Å^{-1}, $\delta = 0.35$, $\rho = 5$. *Left: variations as a function of the loop size N,* ϵ*=0.0019 eV.*Å^{-2} ; ●: *N=12;* □: *N=16;* ◇: *N=21;* △: *N=30 Right: for a fixed loop size , N=21 variations as a function of the loop rigidity;* ●: ϵ*=0.0010 eV.*Å^{-2}; □: ϵ*=0.0019 eV.*Å^{-2}; ◇: ϵ*=0.0040 eV.*Å^{-2}

energy is now **positive**, in agreement with some experimental results. This can be understood because, due to the ϵ-term in the Hamiltonian of the Kratky-Porod chain, there is now an energetic cost for closing. Increasing ϵ costs more energy for closing (k_{cl} decreases). The effect of ϵ is however more subtle because, as shown on Fig. (5.25) the Arrhenius plots for different values of ϵ show almost parallel curves. This indicates that E_{cl} is not simply proportional to ϵ. The closing rate is still strongly affected by entropic effects, which also depend on ϵ. Therefore the rigidity parameter plays a double role, i.e. an enthalpic and an entropic effect. The last point is very interesting because it shows that the Kratky-Porod chain could be a good candidate for the modelling of the loop, i.e. it could allow the differencing of poly(T) and poly(A) as the experiments point out.

Finally, Fig. (5.26) gives the variation of the kinetic rates as a function of D and k with the Kratky-Porod chain. The effects are exactly the same as in the FRC case and we arrive at the same conclusion that the stem only affects the physics of the opening.

This first part allows us to understand qualitatively the effects of the different parameters of the model.

5.5 Complete calculation: $S \neq 1$

We now use the complete calculation of the partition function. The calculation of the partition function involves therefore the conditional probability $S(r|R)$ that, if a polymer of N segments has the distance R between its ends, the polymer of $N+2$ segments has the end-to-end distance r. This function should depend on the

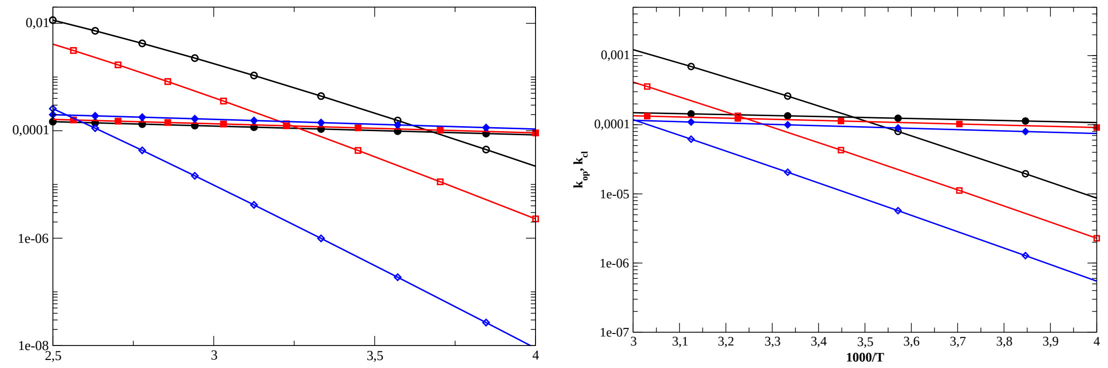

Figure 5.26: *Effect of D and k on the kinetics with the Kratky-Porod chain in an Arrhenius plot. Open and closed symbols represent the rates of opening and closing, respectively. The parameters are: α=6.9 Å^{-1}, $\delta = 0.35$, $\rho = 5$, $\epsilon = 0.0019$ eV.$^{-2}$, N=21. Left: k=0.025 eV.Å^{-2}; ●: D=0.09 eV; □ D=0.102 eV; ◇: D=0.130 eV. Right: k=0.025 eV.Å^{-2}; ●: k=0.013 eV.Å^{-2} ; □: k=0.025 eV.Å^{-2} ; ◇: k=0.050 eV.Å^{-2}.*

polymer model but we can only get its analytical expression in the case of a Gaussian polymer. We have discussed this point in Section (3.1.5) and we have shown that we can evaluate $S(r|R)$ with an effective Gaussian model which provides a good approximation for the FRC and the Kratky-Porod polymer models. In this section we use this effective Gaussian approximation of $S(r|R)$ and we examine in a more quantitative way the various points that we discussed in the previous section.

5.5.1 Thermodynamics

5.5.1.1 FRC model

First of all, it is interesting to look at the difference between the case $S \equiv$1 and the complete calculation which couples the loop and the stem in the polymer model. Figure (5.27) shows that there is not a big difference between the two calculations. Although the case of $S \neq$1 adds entropy in stem, the confinement of the part of the polymer making the stem by the Morse potential and stacking interaction does not allow large fluctuations within the stem as soon as at least one base-pair is made. This particularly true for a short stem. Taking into account the conditional probability $S(r|R)$ is important for the internal consistency of the calculation but it only brings small quantitative changes in the results. Including $S(r|R)$ properly, as we do in this section, would probably become more important for hairpins with a very long stem (20 base-pairs or more) because it would be able to form open bubbles with a large entropy. The next table gives the width of the melting curve, measured by $\frac{\Delta P}{\Delta T} T_m$ defined in Section (5.4.1.1), and compares it with the experimental value for poly(T).

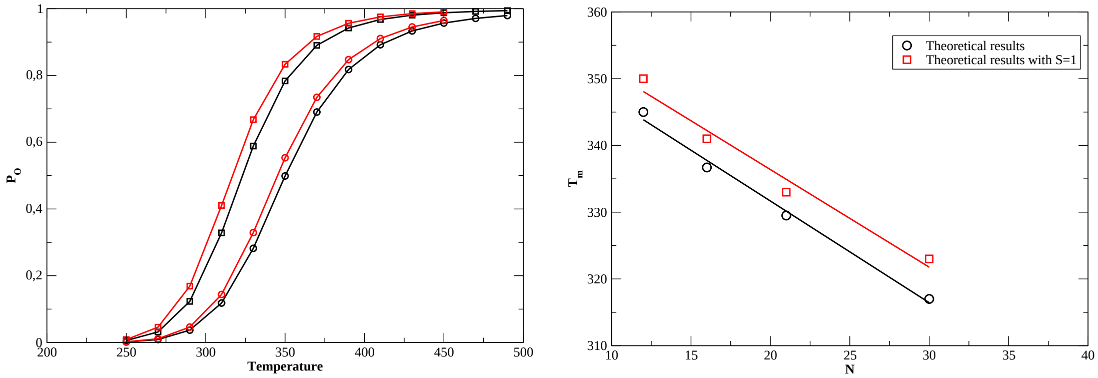

Figure 5.27: *Comparison of the melting curves with $S \equiv 1$ and $S \neq 1$ and with the FRC model: $\theta = 60°$.The parameters of the stem are: D=0.107 eV, k=0.025 eV.Å^{-2}, α=6.9 Å^{-1}, $\delta = 0.35$, $\rho = 5$, $\theta = 60°$. The black colour is for the case of $S \equiv 1$. Left: Melting profiles, ○: N=12; □: N=30. Right: evolution of the melting temperature as a function of N. ○: $S \neq 1$, square: $S \equiv 1$.*

N	$S = 1$, $\frac{\Delta P}{\Delta T} T_m$	$S \neq 1$, $\frac{\Delta P}{\Delta T} T_m$	Exp, Poly(T)
12	3.6	3.7	11
16	3.7	3.8	11
21	3.7	3.8	11
30	3.9	4.0	11

We can notice that the introduction of $S(r|R)$ in the calculation has a very small effect on the width. Whatever the theoretical approach, the calculation gives a width of the melting curves which is significantly higher than the experiments. It is one important weakness of our calculation and we will come back to this point in the discussion of our work. Using the FRC model we have adjusted our parameters in order to compare the results given by the model and the experimental ones in a quantitative way. We have used the following approach to chose the parameters and study the validity of the model. We use the experimental results for poly(T) as the reference. We look for the parameter set that give the best fit of these results as a function of the loop size N. Then we consider the case of poly(A). In this case, as all stem parameters have been fixed by the poly(T) study, we only have one free parameter (θ or ϵ, depending on the polymer model). Figure (5.28) shows melting curves obtained with two different sets of parameters. Both give the melting temperature found in experiments for a poly(T) loop of 12 bases. The difference lies in the variation of T_m as a function of the loop length N and this difference allows us to choose the optimal set of parameters as shown in Fig. (5.29). Indeed the best fit of the black curve which represents the experimental results for poly(T) is provided by the red curve obtained with D=0.112 eV, θ=50°, k=0.025 eV.Å^{-2}, α=6.9 Å^{-1}, δ=0.35 and ρ=5.

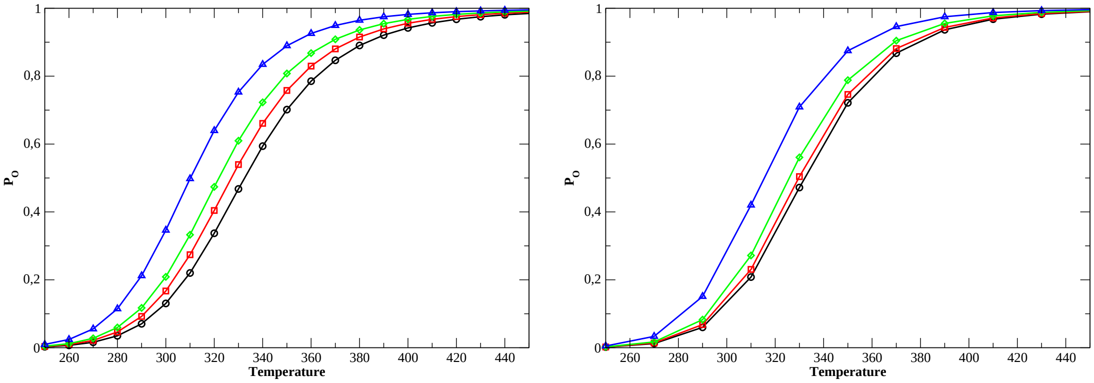

Figure 5.28: *Melting curves equivalent to poly(T) with the FRC model.The parameters of the stem are:k=0.025 eV.*$\AA^{-2}$, α*=6.9* $\AA^{-1}$, $\delta = 0.35$, $\rho = 5$. *Left: Melting profiles, D=0.112 eV,* $\theta = 50°$,$\circ$: *N=12:* $\square$: *N=16;* $\diamond$: *N=21;* $\triangle$: *N=30. Right: melting profiles, D=0.119 eV,* $\theta = 45°$; $\circ$: *N=12;* $\square$: *N=16;* $\diamond$: *N=21;* $\triangle$: *N=30.*

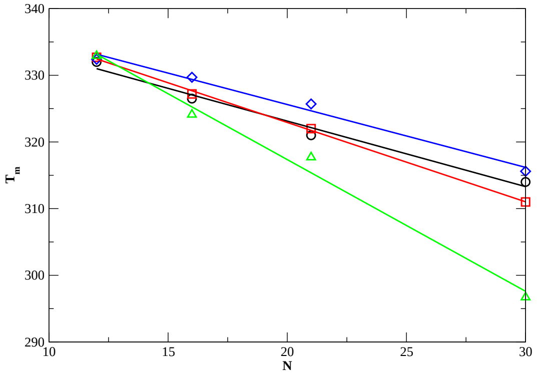

Figure 5.29: *Variation of* T_m *as a function of the loop length N for different sets of parameters.* $\circ$: *experimental results for poly(T);* $\square$: *D=0.112 eV, k=0.025 eV.*$\AA^{-2}$, α*=6.9* $\AA^{-1}$, $\delta = 0.35$, $\rho = 5$, $\theta = 50°$; $\diamond$: *D=0.119 eV, k=0.025 eV.*$\AA^{-2}$, α*=6.9* $\AA^{-1}$, $\delta = 0.35$, $\rho = 5$, $\theta = 45°$; $\triangle$: *D=0.100 eV, k=0.025 eV.*$\AA^{-2}$, α*=6.9* $\AA^{-1}$, $\delta = 0.35$, $\rho = 5$, $\theta = 64°$

Once these parameters have been fixed let us consider the poly(A) case. For the FRC model we can only select θ. As mentioned before the stacking interaction is larger in the case of a poly(A) loop, and we model that by a decrease of θ. As for the case $S \equiv 1$, this leads to a lowering of T_m in agreement with experiments. Figure (5.30) gives the results obtained with θ=48° and the same stem parameters as for the poly(T) case. We also show the comparison of the melting temperature variation as a function of N with the experimental results.

We can see that we are able to reproduce quantitatively the variation of T_m as a function of the loop length for poly(A) putting $\theta = 48°$. T_m varies from 326 K for N=12 to 304 K for N=30 in agreement with experimental results. Nevertheless the width of the transition stays to large as the next table shows. Between experiments

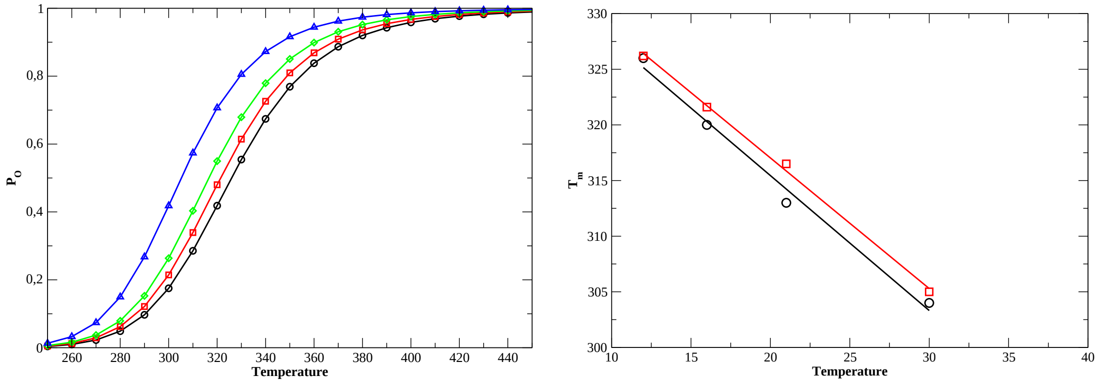

Figure 5.30: *Melting curves equivalent to poly(A) with the FRC model.The parameters of the stem are: D=0.112 eV, k=0.025 eV.*$Å^{-2}$, *α=6.9* $Å^{-1}$, $\delta = 0.35$, $\rho = 5$, $\theta = 48°$. *Left: Melting profiles,* $\circ$*: N=12;* $\Box$*: N=16;* $\diamond$*: N=21;* $\triangle$*: N=30. Right: evolution of the melting temperature with N. black: theoretical results, red: experimental data.*

and our calculation we a have a difference of a factor two for the poly(A) case and a factor three for the poly(T) case. The question is to understand why we get such a difference and if we can do something to improve this aspect. To help us in this discussion we present in the next section the same study with the Krakty-Porod chain model.

N	$\theta = 50°$, $\frac{\Delta P}{\Delta T} T_m$	$\theta = 48°$, $\frac{\Delta P}{\Delta T} T_m$	Poly(T)	Poly(A)
12	3.6	3.7	11	9
16	3.7	3.8	11	8.5
21	3.7	3.8	11	8.5
30	3.9	4.0	11	7.5

To complete the study with the FRC model for the loop, we give the evolution of T_m and of the width of the transition as a function of D, α and k, the depth of the Morse potential, the width of the Morse potential and the rigidity of the stem, respectively. Figure (5.31) shows the variation of T_m as a function of D. We can notice that T_m increases linearly with D. In the case of a single very long stem treated in the approximation of continuum media, one can find that T_m increases with the square root of D using the PBD-model. To properly describe the experimental properties of hairpins we must use a small value of the coupling constant k. This is consistent with the experimental observations on DNA which show that a single base-pair can break without breaking the neighbours. This means that the continuum limit approximation is not valid for DNA. Most of the energy when the stem opens comes from the pairing of the bases and this is why T_m depends linearly on D. The discreteness of the stem is very important and it is why we have not used the transfer integral method presented at the beginning of the chapter. Moreover, the kinetic results for $S \equiv 1$ confirm that the activation energy of opening

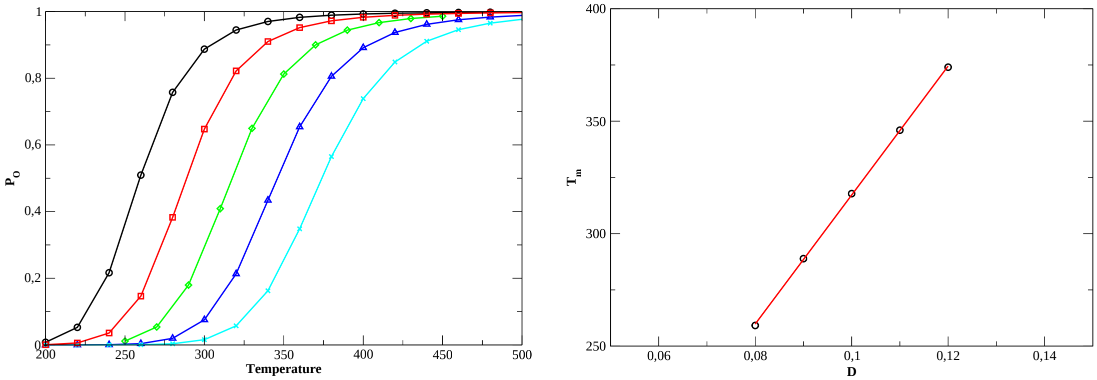

Figure 5.31: *Effect of the depth of the Morse potential on the melting profiles with the FRC modelling.The parameters of the stem are: k=0.025 eV.*Å^{-2}, *α=6.9* Å^{-1}, $\delta = 0.35$, $\rho = 5$, $\theta = 50°$ *and N=21. Left: Melting profiles,* $\bullet$*: D=0.08 eV;* $\square$*: D=0.09 eV;* $\diamond$*: D=0.10 eV;* $\triangle$*: D=0.11 eV,* $\times$*: D=0.12 eV. Right: evolution of the melting temperature with D. black* $\bullet$*: theoretical results, red line: linear fitting.*

only comes from D and not from k. Therefore it is not surprising to find such a linear dependence of T_m with D. Nevertheless, as the next table shows, the width of the transition is not significantly affected by the variation of D.

D (eV)	$S \neq 1$, $\frac{\Delta P}{\Delta T} T_m$
0.08	3.5
0.09	3.9
0.10	3.8
0.11	3.8
0.12	3.9

This shows us that the depth of the Morse potential serves as the fitting of the melting temperature by changing the depth of the first well of the reduced free energy only. Let us now examine the effect of the width of the Morse potential on the thermodynamics presented in Fig. (5.32). As in the case $S \equiv 1$, the larger α, the smaller the melting temperature T_m. The region that represents the closed configuration in the free energy profile is reduced when we increase α. Although it is more difficult to overcome the barrier between the closed and the open state (kinetic effects), the equilibrium is nevertheless displaced to the open state with the increase of α because the volume of the phase space corresponding to a closed state decreases. Moreover the width of the transition is slightly affected by the change of α and as one can expect the smaller the width of the Morse potential, the smaller the width of the transition.

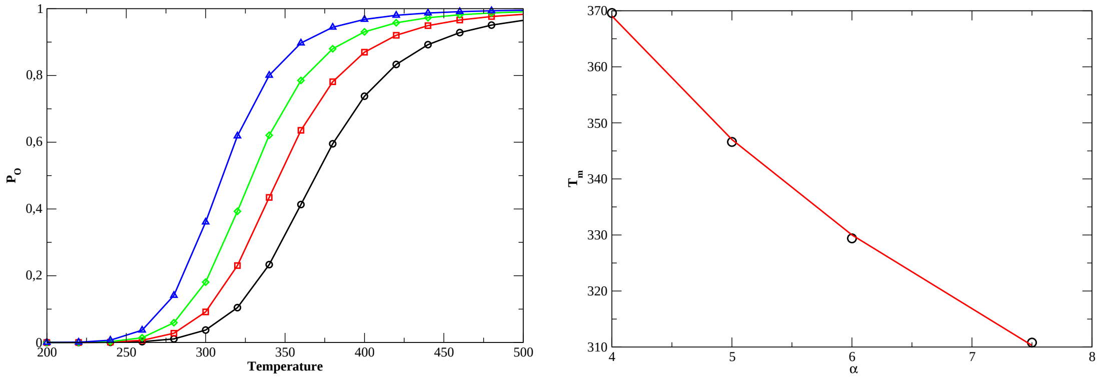

Figure 5.32: *Effect of the width of the Morse potential on the melting profiles with the FRC model. The parameters of the stem are: D=0.112 eV, k=0.025 eV.*$\mathring{A}^{-2}$*,* $\delta = 0.35$*,* $\rho = 5$*,* $\theta = 50°$ *and N=21. Left: Melting profiles,* $\bullet$*:* α*=4.0* $\mathring{A}^{-1}$*;* $\square$*:* α*=5.0* $\mathring{A}^{-1}$*;* $\diamond$*:* α*=6.0* $\mathring{A}^{-1}$*;* $\triangle$*:* α*=7.5* $\mathring{A}^{-1}$*. Right: evolution of the melting temperature with* α*.*

a ($\mathring{A}^{-1}$)	$S \neq 1$, $\frac{\Delta P}{\Delta T} T_m$
4	3.4
5	3.5
6	3.8
7.5	4.1

Finally, Fig. (5.33) gives the evolution of the melting profiles as a function of k. When we increase k we also increase the melting temperature T_m but we slightly decrease the width of the transition from the closed to the open state. The closed

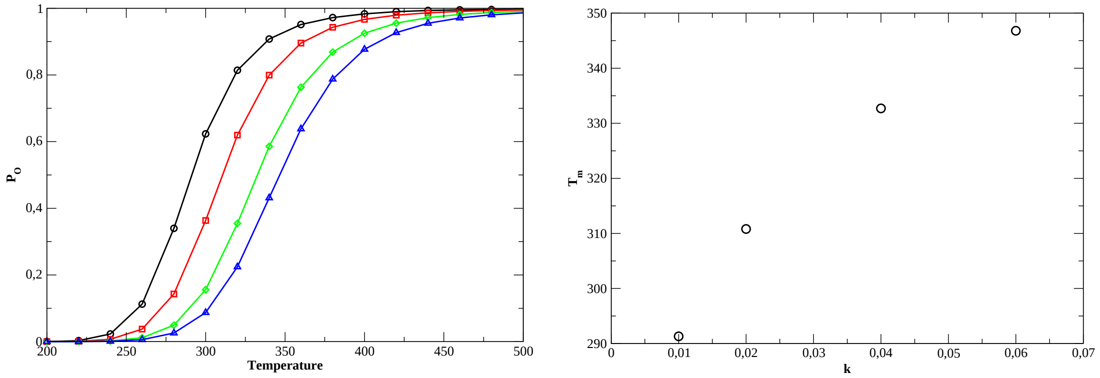

Figure 5.33: *Effect of the rigidity of the stem on the melting profiles with the FRC model.The parameters of the stem are: D=0.112 eV,* α*=6.9* $\mathring{A}^{-1}$*,* $\delta = 0.35$*,* $\rho = 5$*,* $\theta = 50°$ *and N=21. Left: Melting profiles,* $\bullet$*: k=0.010 eV.*$\mathring{A}^{-2}$*;* $\square$*: k=0.020 eV.*$\mathring{A}^{-2}$*;* $\diamond$*: k=0.040 eV.*$\mathring{A}^{-2}$*;* $\triangle$*: k=0.060 eV.*$\mathring{A}^{-2}$*. Right: evolution of the melting temperature with* k*.*

configuration is stabilized by the cooperative effects which are more important when k increases. As the stem is composed of five base-pairs only, the effect of k is less important than in the case of a very long stem. Indeed in the case of a very long stem, in the approximation of continuous medium, $T \propto \sqrt{k}$ but here the dependence is weaker.

k(eV.Å^{-2})	$S \neq 1$, $\frac{\Delta P}{\Delta T} T_m$
0.01	4.1
0.020	4
0.040	3.8
0.06	3.7

As for the case S ≡1, we now present the thermodynamics obtained with the Kratky-Porod chain. As mentioned before, this polymer model presents the advantage of having an explicit energetic term in the probability distribution.

5.5.1.2 Discrete Kratky-Porod model

It is interesting to see the effect of the S function in the case of the Kratky-Porod chain for the loop. Figure (5.34) gives the comparison of the two calculations. In the case of the KP model, the effect of the S function is more important than for the FRC polymer. Indeed, T_m changes from 325 K to 312 K for N=12 when we introduce the S function. We cannot say that it is only due to entropic effects because the KP chain contains energetic contributions, but we can say that the S function tends to destabilize the closed configuration. The next table gives the change of the width of the transition with and without the S function.

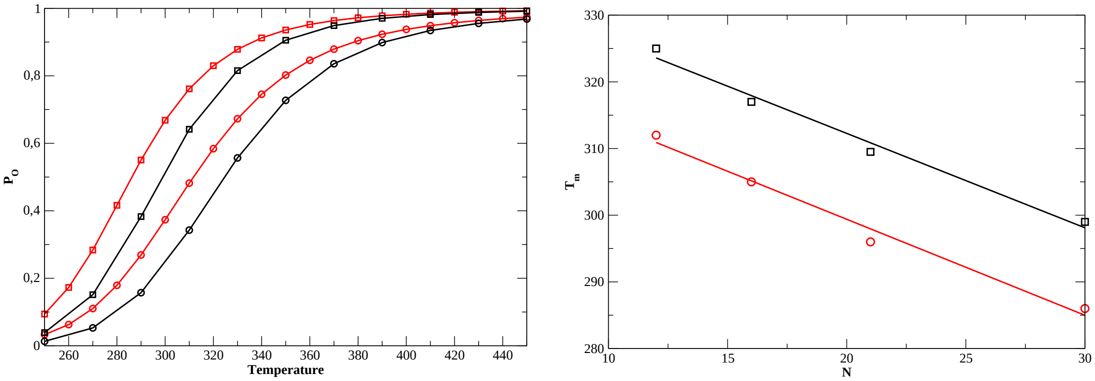

Figure 5.34: *Comparison of the melting curves with S ≡1 and S ≠1 and with the Kratky-Porod model:* ϵ = 0.0019 *eV.Å^{-2} .The parameters of the stem are: D=0.102 eV, k=0.025 eV.Å^{-2}, α=6.9 Å^{-1},* $\delta = 0.35$, $\rho = 5$. *The black color is for the case of S ≡1. Left: Melting profiles,* ○: *N=12;* □: *N=30. Right: evolution of the melting temperature as a function of N.* ○: *S ≠1,* □: *S ≡1. The curves correspond to a linear fitting.*

N	$S \equiv 1$, $\frac{\Delta P}{\Delta T} T_m$	$S \neq 1$, $\frac{\Delta P}{\Delta T} T_m$
12	3.3	3.2
30	4.1	3.7

As we can see, the width of the transition seems to be slightly larger in the presence of the S function but the change is not significant enough to allow a quantitative comparison with experiments. Moreover we have seen that the evolution of T_m as a function of N is not monotonous for ϵ=0.0040 eV.Å^{-2}. It is interesting now to see what happens when we put the S function. To give a quantitative comparison, Fig. (5.35) shows the evolution of $T_m(S \equiv 1, N) - T_m(S \equiv 1, N = 12)$ and $T_m(S \neq 1, N) - T_m(S \neq 1, N = 12)$ as a function of N.

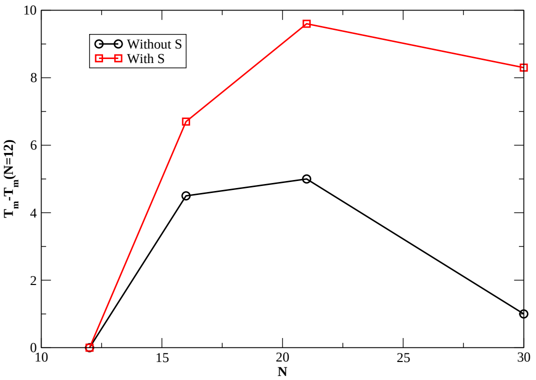

Figure 5.35: *Variation of T_m as a function of N with and without the S function. The black curve represents $T_m(S \equiv 1, N) - T_m(S \equiv 1, N = 12)$ and the red one is for $T_m(S \neq 1, N) - T_m(S \neq 1, N = 12)$.*

We can notice that we get the same tendency with and without the S function. The maximum of the curve stays around N=21 which shows us that the growth of the loop inside the stem represented by the S function has no effect on this maximum. Therefore this maximum is only governed by the evolution of the end-to-end probability distribution with N.

As we have done before we now give the comparison of the experimental results with our model in the case of the KP modelling for the loop in order to determine which is the best loop model. Figure (5.36) shows the melting curves obtained for ϵ=0.0018 eV.Å^{-2} which corresponds to a persistence length equal to 12.3 Å. The right graphic gives the comparison of the evolution of T_m as a function of N obtained experimentally for the poly(T) and obtained in our simulation. We can see that our results are in semi-quantitative agreement with the experiments since T_m varies from 333 K for N=12 to 305 K for N=30 which is comparable to the experimental case where T_m goes from 332 K to 314 K for the same variation of N. Our main problem stays in the width of the transition which is really too large compared to the experiments as shown in the next table.

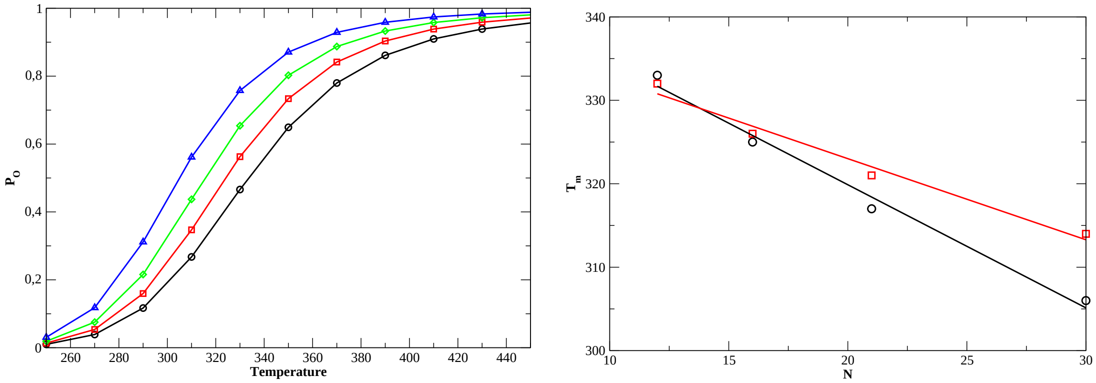

Figure 5.36: *Melting curves equivalent to poly(T) with the KP model.The parameters of the stem are: D=0.107 eV, k=0.025 eV.Å^{-2}, α=6.9 Å^{-1}, δ = 0.35, ρ = 5, ϵ = 0.0018 eV.Å^{-2}. Left: Melting profiles, ●: N=12; □: N=16; ◇: N=21; △: N=30. Right: evolution of the melting temperature as a function of N. black: theoretical results, red: experimental data.*

N	ϵ=0.0018 eV.Å^{-2}, $\frac{\Delta P}{\Delta T}T_m$	Poly(T), $\frac{\Delta P}{\Delta T}T_m$
12	3.2	11
16	3.4	11
21	3.45	11
30	3.8	11

The parameter ϵ represents the rigidity of the chain as mentioned before. The rigidity for the poly(A) loops is larger than the poly(T) because the stacking interaction is most important with A-bases. Therefore in order to model the difference between poly(T) and poly(A) we have increased the value of ϵ and we have adjusted our value to get T_m which agree with experiments. Figure (5.37) gives the melting curves obtained with ϵ=0.00195 eV.Å^{-2} which corresponds to a persistence length equal to 13.5 Å. We can see that T_m goes from 327 K for N=12 to 300 K for N=30 which is comparable to the experimental result where ΔT_m is equal to 22 K for the same variation of N and with T_m equal to 326 K for N=12. Nevertheless we still find larger transitions than the experimental case as shown in the next table.

N	ϵ=0.00195 eV.Å^{-2}, $\frac{\Delta P}{\Delta T}T_m$	Poly(A), $\frac{\Delta P}{\Delta T}T_m$
12	3.25	9
16	3.45	8.5
21	3.6	8.5
30	3.8	7.5

One can notice that to model the difference between poly(T) and poly(A) we do not need to significantly change the value of the persistence length. We will come back to this point in the discussion section after the presentation of the kinetic results. To complete this part we give the evolution of the melting profiles with the change

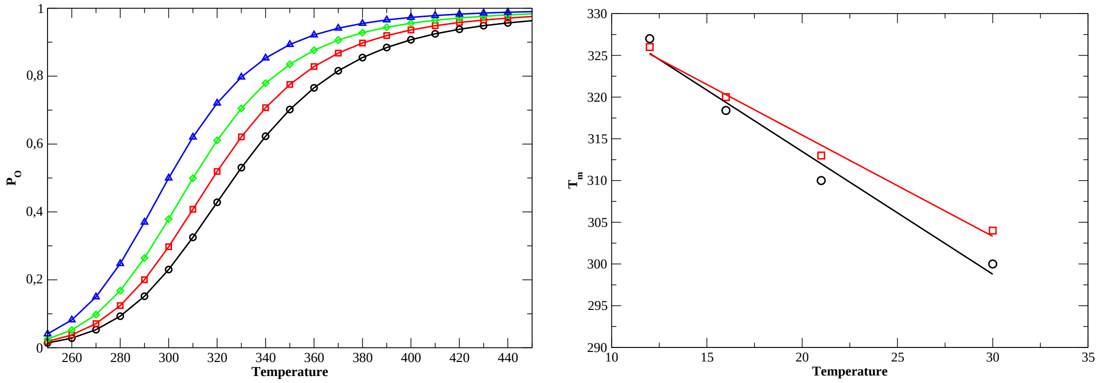

Figure 5.37: *Melting curves equivalent to poly(T) with the KP model.The parameters of the stem are: D=0.107 eV, k=0.025 eV.Å^{-2}, α=6.9 Å^{-1}, $\delta = 0.35$, $\rho = 5$, $\epsilon = 0.00195$ eV.Å^{-2}. Left: Melting profiles, $\bullet$: N=12; $\Box$: N=16; $\diamond$: N=21; $\triangle$: N=30. Right: evolution of the melting temperature as a function of N. black: theoretical results, red: experimental data.*

of D. Figure (5.38) shows such an evolution. We find a linear evolution, as for the

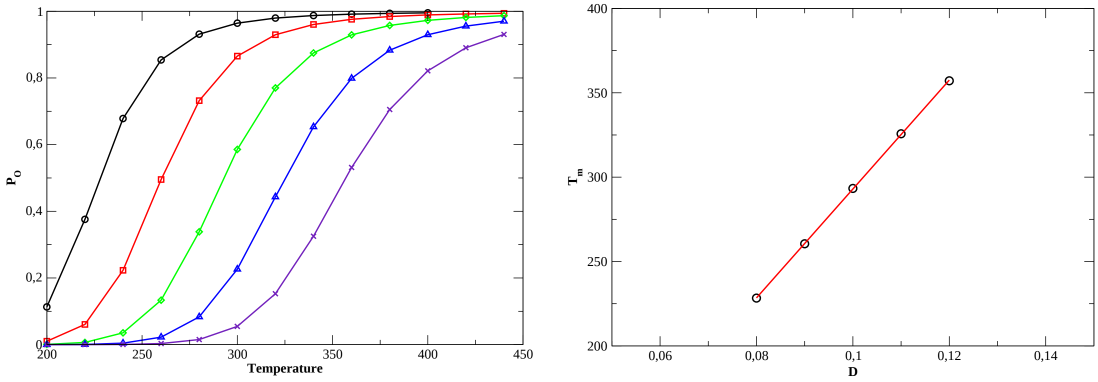

Figure 5.38: *Effect of the depth of the Morse potential on the melting profiles with the KP model.The parameters of the stem are: k=0.025 eV.Å^{-2}, α=6.9 Å^{-1}, $\delta = 0.35$, $\rho = 5$, ϵ0.0018 eV.Å^{-2} and N=21. Left: Melting profiles, $\circ$: D=0.08 eV; $\Box$: D=0.09 eV; $\diamond$: D=0.10 eV; $\triangle$: D=0.11 eV, $\times$: D=0.12 eV. Right: evolution of the melting temperature with D. black $\circ$: theoretical results, red line: linear fitting.*

FRC loop model which is not really surprising. Moreover, as the next table shows, the width of the transition is not significantly affected by the variation of D.

D (eV)	$S \neq 1$, $\frac{\Delta P}{\Delta T} T_m$
0.08	3.4
0.09	3.3
0.10	3.5
0.11	3.4
0.12	3.4

After dealing with the thermodynamics of the model we propose to study the kinetics and compare our results to the experimental ones.

5.5.2 Kinetics

5.5.2.1 FRC model

Let us first compare the kinetic result obtained with and without S in one particular case to see if there is a significant difference. Figure (5.39) gives such a comparison.

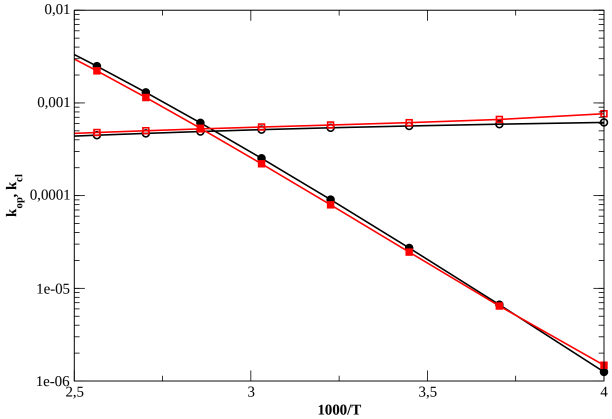

Figure 5.39: *Comparison of the kinetic rates with and without S with the FRC model in an Arrhenius plot. Open and closed symbols represent the rates of opening and closing, respectively. The parameters are the following: D=0.107 eV, k=0.025 eV.Å^{-2}, α=6.9 Å^{-1}, δ=0.35, ρ=5, $\theta = 60°$ and N=21. Black: S ≡1. Red: S ≠1.*

As we can show there is no perceptible difference between the two calculations. Even if the case $S \equiv 1$ is conceptually not satisfactory, it gives quite correct results. As discussed for the FRC case, this comes from the fact that the stem is confined by the Morse potential, so that the effect of the polymer part in the stem is small.

Let us now compare the kinetics obtained by the model and the experiments. The parameters have been selected by the thermodynamic studies so that we cannot do any fitting at this level.

Figure (5.40) gives the rates of opening and closing for different loop lengths and

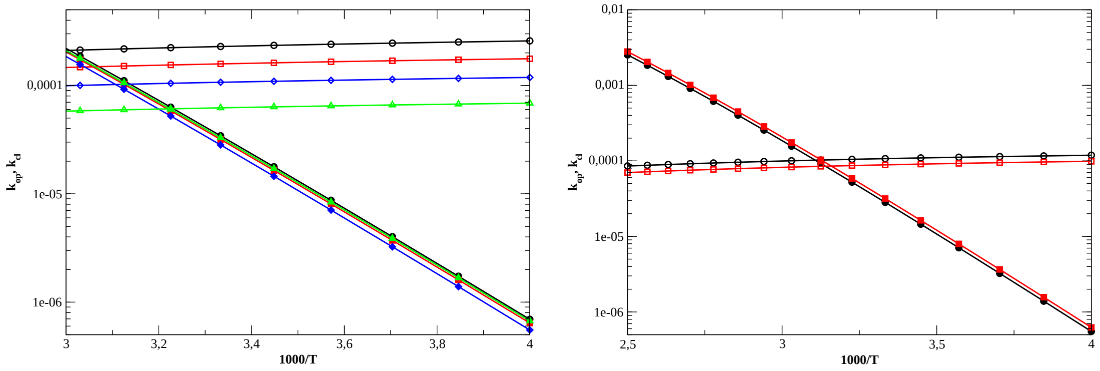

Figure 5.40: *Rates of opening and closing with the FRC model in an Arrhenius plot. Open and closed symbols represent the rates of closing and opening, respectively. The parameters are: D=0.112 eV, k=0.025 eV.*Å^{-2}, *α=6.9* Å^{-1}, $\delta = 0.35$, $\rho = 5$. *Left:* $\theta = 50°$*; •: N=12; □: N=16; ⋄: N=21; △: N=30. Right: N=21, black:* $\theta = 50°$, *red:*$\theta = 48°$

for $\theta = 50°$ and 48°. For the FRC model it is not possible to do a quantitative comparison of the theoretical results and the experimental ones, because, firstly we get negative activation energies for closing which is not the case of experiments and secondly we have a factor approximately three between the activation energy of opening obtained with our model and obtained in the experiments. Moreover the kinetics is only marginally modified when θ is varied in the range which correctly models the difference between poly(A) and poly(T) in the thermodynamics. However, as in the experiments, the increase of the loop length tends to decrease the rate of closing and it does not affect the rate of opening. As mentioned before when we increase the loop length, the available phase space is then bigger, therefore the hairpin takes more time to close.

The theoretical results as well as the experimental ones concerning the kinetics with the FRC model are summarized in the next table.

	E_{op}, model	E_{cl}, model	E_{op}, exp	E_{cl}, exp
Poly(T)	11.5	-0.33	32	3.4
Poly(A)	11.5	-0.33	32	17.4

As we can see in the table our model does not provide a quantitative agreement with experiments for the kinetics. This shows us that the single stranded DNA is not only a simple polymer. We will come back to this point after presenting the kinetics obtained with the Kratky-Porod chain which is a more realistic polymer model.

To complete this section, we present the evolution of the activation energies as a function of D, k and α. Figure (5.41) gives the rates of opening and closing with N=21 for different values of D.

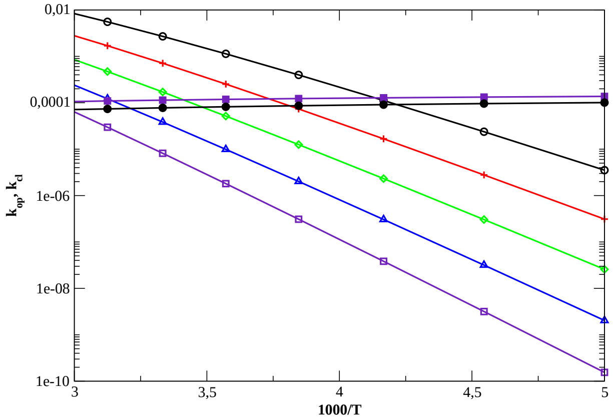

Figure 5.41: *Effect of D on the kinetics with the FRC model in an Arrhenius plot. Open and closed symbols represent the rates of opening and closing, respectively. The parameters are the following:* k=0.025 eV.$\mathring{A}^{-2}$, α=6.9 $\mathring{A}^{-1}$, δ=0.35, ρ=5, $\theta = 50°$ *and* N=21. *Rates of opening:*○: D=0.08 eV; +: D=0.09 eV; ◇: D=0.10 eV; △: D=0.11 eV; □: D=0.12 eV. *Rates of closing:* ●: D=0.08 eV; ■: D=0.12 eV.

First of all, we can notice that the rates of opening and closing are well described by an Arrhenius law even if we change the width of the Morse potential D. Moreover we can see that the closing is not really affected by the change of D as the case of S ≡1 which shows us that the closing is almost governed by the loop part of the hairpin. Moreover, when we increase D, we also increase the activation energy of opening E_{op}. Figure (5.42) gives the evolution of the activation energy of opening as a function of D. The red curve represents 5D in Kcal.mol^{-1} units.

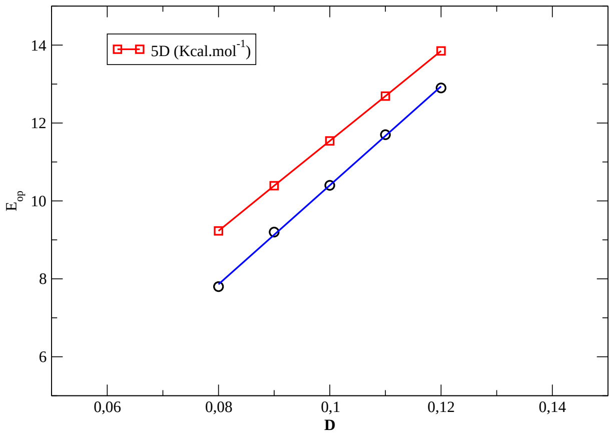

Figure 5.42: *Evolution of the activation energy of opening as a function of D. The parameters are the following:* k=0.025 eV.$\mathring{A}^{-2}$, α=6.9 $\mathring{A}^{-1}$, δ=0.35, ρ=5, $\theta = 50°$ *and* N=21. *The red curve represents* 5×D *in* Kcal.mol^{-1} *units.* ○: *theoretical results. The blue curve is a linear fitting.*

As we can see, the variation of the activation energy of opening as a function of D is linear. Moreover for a given value of D, E_{op} is close to $M \times D$ but it always stays lower than this value. As we also put stacking interaction in the stem we expect activation energies of opening of the order of $M \times D$ plus something coming form the stacking. On the contrary, we get the reverse, here. Moreover if we look at Fig. (5.43), the activation energy of opening and closing are not significantly affected by k which represents the force of the stacking interaction and by α. Stacking interactions only have an entropic effect (the curves are only translated). Before concluding on the kinetics let us examine the results obtained with the Kratky-Porod chain.

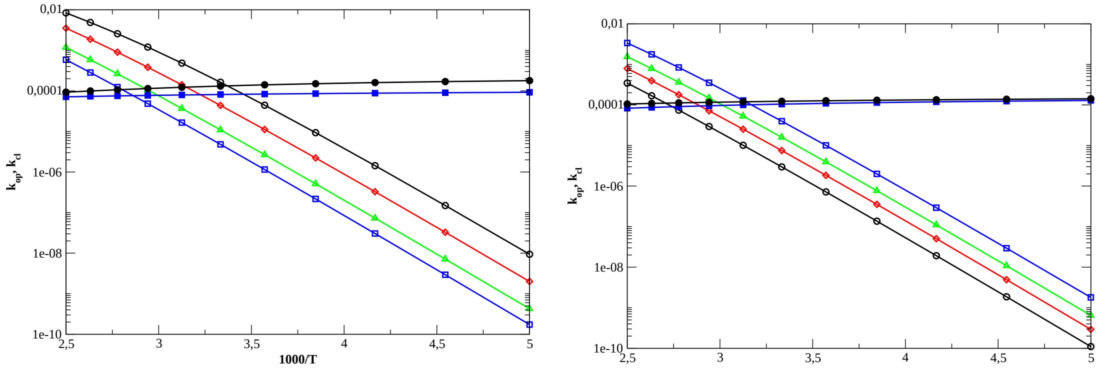

Figure 5.43: *Effect of k and α on the kinetics with the FRC in an Arrhenius plot. Open and closed symbols represent the rates of closing and opening, respectively. The parameters are: D=0.112 eV, $\delta = 0.35$, $\rho = 5$ and N=21. Left: α=6.9 Å^{-1}; ○: k=0.01 eV.Å^{-2}; ⋄: k=0.02 eV.Å^{-2}; ; △: k=0.04 eV.Å^{-2}; □: k=0.06 eV.Å^{-2};. Right: k=0.025 eV.Å^{-2}. ○: α=4.0 Å^{-1}; ⋄: α=5.0 Å^{-1}; △: α=6.0 Å^{-1}; □: α=7.5 Å^{-1}.*

5.5.2.2 Discrete Kratky-Porod model

First of all, as in the previous case, let us begin by the comparison of one kinetic result obtained with and without S to see the influence of S. Figure (5.44) gives such a comparison. We can notice that the closing rate is not significantly affected by the use of the complete calculation and it is not so surprising because, as we pointed out, the closing is mostly governed by the loop composed of the N monomers and not by the stem. Nevertheless, the opening is slightly affected by S which tends to slightly decrease the opening activation energy without changing the entropy because at high temperatures the two curves meet.

Figure (5.45) gives the rates of opening and closing for two different values of ϵ and for different values of the loop length N. We have used the parameters presented in the section thermodynamics, which provide the optimal comparison with the experimental results.

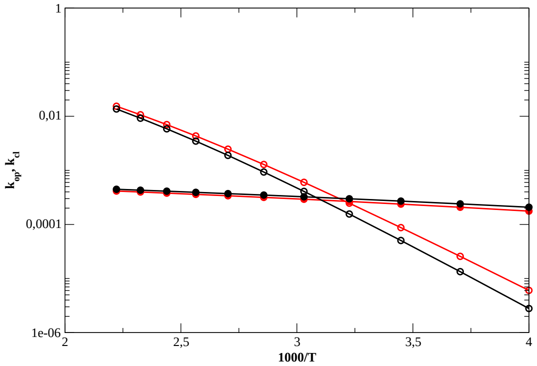

Figure 5.44: *Comparison of the kinetic rates with and without S with the KP model in an Arrhenius plot. Open and closed symbols represent the rates of opening and closing, respectively. The parameters are the following:* D*=0.102 eV,* k*=0.025 eV.*$\AA^{-2}$, α*=6.9* $\AA^{-1}$, δ*=0.35,* ρ*=5,* ϵ*=0.0019 eV.*$\AA^{-2}$*. Black:* $S \equiv 1$*. Red:* $S \neq 1$*.*

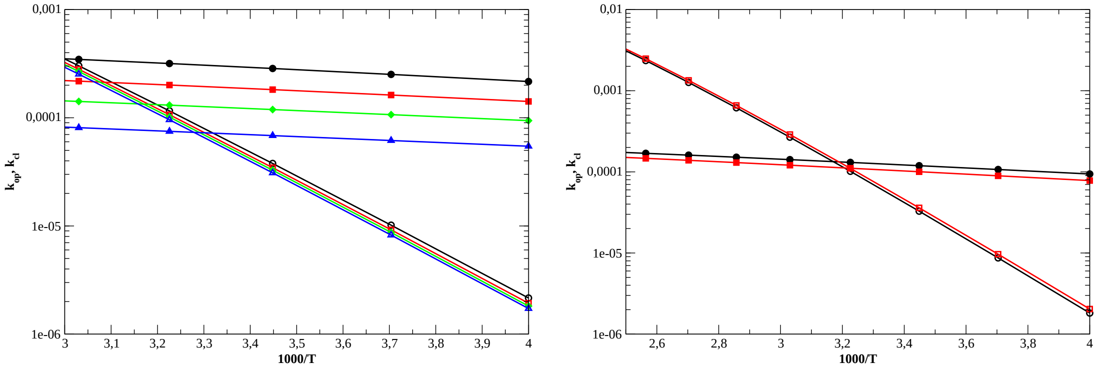

Figure 5.45: *Rates of opening and closing with the KP model in an Arrhenius plot. Open and closed symbols represent the rates of closing and opening, respectively. The parameters are:* D*=0.107 eV,* k*=0.025 eV.*$\AA^{-2}$, α*=6.9* $\AA^{-1}$, $\delta = 0.35$, $\rho = 5$*. Left:* ϵ*=0.0018 eV.*$\AA^{-2}$*;* $\circ$*: N=12;* $\square$*: N=16;* $\diamond$*: N=21;* $\triangle$*: N=30. Right: N=21, black:* ϵ*=0.0018 eV.*$\AA^{-2}$*, red:* ϵ*=0.00195 eV.*$\AA^{-2}$

As for the FRC model the kinetic of opening in not affected by the change of the number of monomers in the loop. The opening activation energy E_{op} is equal to 0.43 eV (10 kcal.mol^{-1}) for $D = 0.107$ eV. Concerning the kinetic of closing, we find that the larger the number of monomers, the smaller the rate of closing. Indeed if we increase the entropy of the loop by increasing N, then the loop takes more time to find the transition state in the phase space. Nevertheless, the closing activation energy is not significantly affected by the change of N. We find a closing activation energy E_{cl} equals to 0.04 eV (1 kcal.mol^{-1}). The next table gives the comparison with the experimental results.

	E_{op}, model	E_{cl}, model	E_{op}, exp	E_{cl}, exp
Poly(T)	10	+1	32	3.4
Poly(A)	10	+1	32	17.4

We see that we are not able to get quantitative agreement between our results and the experimental ones. Moreover if we increase the value of ϵ which gives us the difference between poly(A) and poly(T) in the thermodynamics, we get almost no difference in kinetics. This is in agreement with what we can see in literature where they claim that regarding the difference in the kinetics, the persistence length of poly(A) must be four times larger approximately than the poly(T) case to reproduce such a difference [7]. But if we impose such a change in the persistence length in order to get the correct kinetic results, it is then the thermodynamic results which are wrong. This shows us that the single stranded DNA is not a simple polymer. To model it one must elaborate more complex models. We will come back on this point in the conclusion because this an important lesson learned from the analysis of DNA hairpins.

5.5.3 Discussions

Our model allows us to derive thermodynamics and kinetics properties of DNA hairpins. We find that the thermodynamic results are in semi-quantitative agreement with the experimental ones. Indeed, we get correct values of the melting temperature T_m and a good dependence on the loop length. Moreover, the difference between poly(A) and poly(T) can be reproduced by increasing the rigidity of the loop. Nevertheless, we have shown that a slight change of the rigidity is sufficient to get the change of T_m. Therefore, the persistence length l_p would be comparable for poly(A) and poly(T) in our study. We must point out that the transition width that we get is approximately two times larger than expected in experiments. It could explain why we only need a small change of the rigidity parameter to get the correct variation of T_m.

For the kinetics, we have supposed that the system diffuses in a free energy surface that we derive from the thermodynamic study and we have derived the rates of opening and closing using the transition state theory and not only the Kramers'theory. At this stage we have fixed the diffusion coefficient to a constant. We find that the kinetics of opening does not depend on the loop properties as in the experiments. Moreover we get positive activation energies of opening but the values differ from a factor three from the results obtained by Libchaber. As we have shown, we can increase E_a by increasing D, which is the depth of the Morse potential but it would also change T_m to values that do not agree with experiments.

For the kinetics of closing the results are mixed. First of all, we are not able to get results in quantitative agreement with experiments. Nevertheless we can bring some contributions to the debate of the sign on the activation energy of closing that we raised in the introduction. First, we have shown that the Arrhenius law is only valid at low temperatures, i.e. below the melting temperature T_m. Moreover we have seen that it is possible to get negative or positive activation energies of closing putting or not energetic contributions in the loop. But we now that the stacking interaction is important within the loop as Libchaber and coworker show in their study and it is more important in poly(A) loops. Therefore the model of the loop must include energetic contributions. In this hypothesis, we find a positive activation energy of closing. As mentioned in the introduction, in their analysis of their discrepancy with the experiments of the Libchaber group, Ansari *et al.* attribute the positive activation energy for closing to mismatches. While we are not able to give a quantitative assessment of the effect of mismatches because we have not studied them, we can however show that mismatches are **not a necessary condition** to get a positive activation energy for closing. It can come from the rigidity of the loop only.

5.5.4 Beyond the PBD-model for the stem

Up to now we have described the stem by the PBD-model which has the interest of being fairly simple while describing the melting properties of DNA to a good accuracy as tested in some experiments [77]. We have obtained interesting results on the effect of the loop but we are still facing quantitative disagreement with experiments for the width of the melting transition. The model finds that the opening of the hairpin extends on a much broader range than in the experiments.

This problem of the broad melting was also met in the first studies of the double helix thermal denaturation. For a long double helix (or in the limit of an infinite double stranded DNA) the problem was solved by the introduction of the nonlinear stacking

$$W(y_i, y_{i-1}) = \frac{K}{2}\left[1 + \rho e^{-\delta(y_i + y_{i-1})}\right](y_i - y_{i-1})^2 . \tag{5.67}$$

Its effect is to increase the entropy of the melted part of the helix with respect to that of the closed part because the coupling decreases when either one of the two base-pairs is open.

However the coupling never vanishes, even when y_i, y_{i-1} are very large due to the constant 1 in the expression. This is necessary in the PBD-model because the DNA strands do not break, even when the double helix is denaturated.

In our hairpin model the stacking interaction does not have to describe the covalent bonds within the strands because this part of the physics of the hairpin is described by the polymer model. Since the stacking potential only describes the interaction by the plateaus made by the bases, in particular through the overlap of their π-electrons, it is now acceptable to let the stacking decay to 0 when the stem is fully open, as schematized in Fig. (5.46). To test the consequences of a complete vanishing of the stacking interaction, we have considered the case of the stacking potential

$$W_1(y_i, y_{i-1}) = \frac{1}{2} K_1 \rho e^{-\delta(y_i - y_{i-1})}(y_i - y_{i-1})^2 , \tag{5.68}$$

instead of the potential W. To allow a comparison with our previous results we have chosen

$$K_1\rho = K\,(1+\rho)\,, \tag{5.69}$$

which ensures that, for the closed stem, the stacking is not modified.

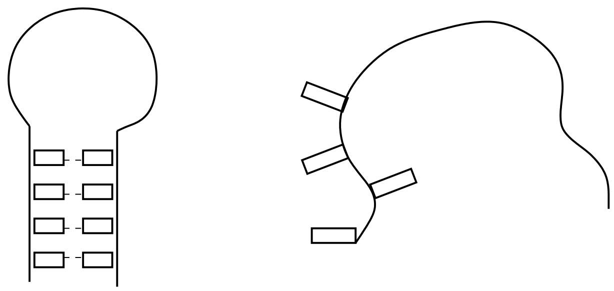

Figure 5.46: *Schematic representation of the stacking in the closed and the open configuration. Left: closed stem, the base-pairs interact. Right: open stem, the position of the bases is random and their stacking energy may vanish*

Figure (5.47) compares melting curves obtained with stacking described by W and W_1.

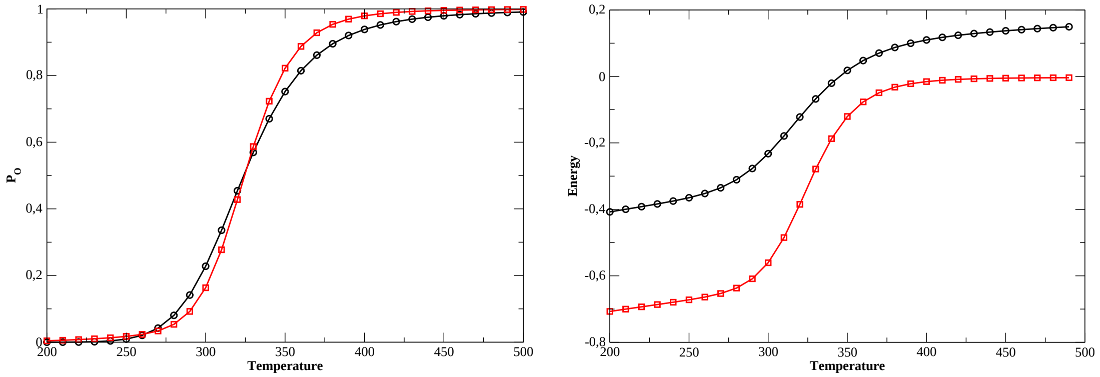

Figure 5.47: *Comparison of the melting curves and the energies obtained with two stacking potentials W and W_1. These calculations have been performed with a loop described by the Kratky-Porod chain (effective Gaussian approximation). Left: melting curves. Right: energy. The black color corresponds to D=0.112 eV, k=0.025 eV.$\mathring{A}^{-2}$, α=6.9 $\mathring{A}^{-1}$, δ=0.35, ρ=5, $\epsilon = 0.0019$ eV.$\mathring{A}^{-2}$, N=24 and stacking W. The red color corresponds to D=0.170 eV, k=0.030 eV.$\mathring{A}^{-2}$, stacking described by W_1 and identical others parameters.*

A stacking potential W_1 leads to a slightly sharper melting curve, which is therefore in better agreement with experiments, although the opening transition given by the model is still broader than the observed transition. It should be noticed that, in order to preserve the melting temperature, when we use the stacking potential W_1 we increase significantly the depth of the Morse potential. As shown by Fig. (5.47) showing the energy versus temperature for the two cases of stacking W and W_1, using stacking W_1 leads to an energy increase of 0.6 eV at the opening transition instead of 0.4 eV when we use the stacking W. This higher value is in better agreement with experimental measurements which give approximately 34 kcal/mol (1.47 eV) for hairpins with five base-pairs stem but still lower than the experimental values.

Conclusion

Conclusion

We have presented a simple model for DNA hairpins which contains the main physical ingredients, i.e. a polymer describing the DNA strands and the main features of the stem, base pairing and stacking. It allows us to understand the main features of hairpin properties, in particular the role played by the loop in the opening-closing hairpins:

- with respect to the stem alone, hairpins open at significantly lower temperatures. We have shown that it can be understood in terms of entropy gain when the loop opens.

- larger loops decrease the opening temperature even more, in agreement with experiments.

Kinetic studies have been very useful to complete our understanding because:

- they give results separably on opening and closing; allowing us to analyse the data more completely and in particular determine what has to be attributed to the stem and what comes from the loop

- they also help us determining what comes form energetic or entropic effects in the properties of hairpins.

The model is successful on some aspects:

- the effect of the size of the loop,

- the correct order of magnitude for E_{op}, E_{cl} (in particular positive activation energy for closing, while other models do not get this experimental feature), although our values are smaller than the experimental ones.

But the model is still not fully satisfactory:

- the melting transition that we calculate is too broad,

- the variation of T_m versus N is smaller for more rigid loops than for softer ones in our calculations while experiments show the contrary.

This indicates that some physical aspects are not properly described in our approach. Our results suggest that this problem cannot be solved by improving the polymer model because we have used two very different polymer models and they give the same qualitative behavior. The FRC model has no energetic term in the loop while the Kratky-Porod model (or its continuous counterpart the worm like chain) includes a bending energy. The Kratky-Porod model is an improvement because it can give a positive E_{cl} but it does not solve the quantitative disagreement that we noticed above.

The solution can neither come from a simple improvement of the model for the stem. We have used the PBD-model but we have shown for instance that changing drastically the model for the stacking by allowing the stacking energy to vanish completely in the open state narrowers slightly the melting transition but does not bring a major quantitative change. However this attempt to improve the model might give a clue to improving the theoretical description of DNA hairpins, because it suggests that an increase in the entropy change when the hairpin opens could bring the model closer to experiments. The simplifying assumptions that we have made to establish the model are indeed leading to an underestimation of the entropy. The main restriction is that bases are described as points. This allowed us to use a simple polymer model for the strand of the stem and loop but it ignores the entropy associated to the fluctuations of the orientation of the bases. When the stem is formed the bases have restricted motions, but when the pairing is broken the bases acquire a large orientational freedom which is not described in our model. Similarly, for the loop the polymer model completely ignores the orientational fluctuations of the bases. Moreover the properties of the loop could be strongly affected by the tendency of the bases, particularly the large purines such as A, to stack on each other.

Our results show that DNA hairpins are very good test to study the properties of DNA single strands. When this work started, our aim was to learn how to describe DNA self assembly and we had in mind that the effort would have to be focused mainly on a correct description of the stem. But as the study developed we got evidence that a good model of the loop was crucial. Hairpins provide precise experimental results so that their models are submitted to strict testing. Obviously we have not fully succeeded in describing DNA hairpins theoretically. We would however like to point out that the difficulties appear when one tries to describe **all** the experimental results (thermodynamics and kinetics, for various types of loops poly(A) or poly(T) and various loop lengths). To our knowledge all previous attempts to model DNA hairpins have only considered some aspects when a subset of the experimental results is considered. But, when they are considered on all their facets, DNA hairpins appear to be very complex.

The study shows that the description of the loop plays a large role for the validity of a model. This is why we had to investigate different possibilities.

Although they give interesting results none of the models is perfect and this study shows that a DNA strand is not a **simple polymer**! On a very long scale (hundreds of bases) a WLC model might be enough. On a very small scale (2 or 3 bases) any simplified model is bound to fail due to the complex geometry and interactions of the element making the strand (phosphates, sugars, bases). The intermediate range that hairpins allow to study (10 to 30 or 50 bases) could have been expected to be approximately described by the Kratky-Porod model which is a discrete version of the continuous WL chain. According to our study this is probably the best polymer model that one can use, but we have nevertheless shown that it is still not sufficient to describe all the properties of the DNA strand forming the loop of a hairpin.

Summary

DNA beacons are made of short single strands of DNA with terminal regions consisting of complementary base sequences. As a result, the two end-regions can self-assemble in a short DNA double helix, called the stem, while the remaining central part of the strand makes a loop. In this closed configuration, the single strand has the shape of a hairpin. Such hairpin conformations are important in determining the secondary structure of long single strands of DNA or RNA. A short single strand of DNA which can form a hairpin becomes a so-called « DNA beacon » when one of its ends is attached to a fluorophore while the second end is attached to a quencher. When the fluorophore and the quencher are within a few Angströms, the fluorescence is quenched due to direct energy transfer from the fluorophore to the quencher. As a result, in a closed hairpin configuration, the beacon is not fluorescent, while in the open configuration it becomes fluorescent. This property opens many interesting applications for molecular beacons in biology or physics. Biological applications use the possible assembly of the single strand which forms the loop with another DNA strand which is complementary to the sequence of the loop. The assembly of a double helix replacing the single strand of the loop forces the opening of the hairpin, leading to a fluorescent signal. This technique provides very sensitive probes for sequences which are complementary to the loop. In the same spirit it has been suggested that DNA beacons could be used in vivo to detect the single stranded RNA which is synthetized during the transcription of genes. This opens the possibility to recognise cancer cells by targeting some genes which are heavily transcribed in such cells.

For physics DNA beacons are very interesting too. They can for instance be used as the basis of some devices such as molecular memories read by the detection of fluorescence, or to perform molecular computation. The most important aspect for our purpose is that molecular beacons allow accurate observations of the opening and closing of DNA hairpins. The « melting profile » of the stem, induced by heating, can be recorded accurately versus temperature and the auto-correlation function of the fluorescence can be used to extract the kinetics of the opening/closing fluctuations. Measurements have been made for different loop lengths and different bases in the loop, providing a complete set of data which can be used to understand what governs the properties of DNA hairpins. This is the goal of this thesis. The analysis goes beyond the properties of hairpins themselves because, as shown below, the results are very sensitive to the properties of the loop. Therefore the comparison of

experimental data with the results of various models is a very sensitive test of our ability to model single strands of DNA. This is important in other related contexts such as the properties of RNA.
We have developed two different models in order to study the thermodynamics and the kinetics of such systems. The first one is a planar square lattice model inspired by the lattice models which have been used to study protein folding. The energy of the DNA strand depends on two terms only, a bending energy when two consecutive segments form a right angle and the energy of the base-pair which can form in the stem. Using Monte Carlo simulation, we compute the equilibrium properties and the kinetics of the system. The results obtained by this model are in qualitative agreement with the experiments showing that the main properties of DNA hairpin rely on very simple and general ideas. Nevertheless, the main weakness of the model is that it does not have enough degrees of freedom, so that a quantitative comparison with experiments is not possible. Therefore we have proposed another model which includes the physical ingredients of the lattice model but without the constraint of the lattice. It combines polymer theory and the Peyrard-Bishop and Dauxois (PBD) model of DNA melting. The model treats the hairpin as consisting of two subsystems:

- the loop which is modelled by a polymer

- the stem which is modelled by the PBD + additional terms that take into account the growth of the loop inside the stem.

With this approach we can compare our results quantitatively with the experimental ones. We find a good agreement for the dependence of the melting temperature with the characteristics of the loop, i.e. the length and the nature of the sequence. Moreover the kinetic results are in qualitative agreement with the experiments. We find that the kinetics of opening is governed by the stem only and that the rate of closing decreases with the length of the loop. However we are not able to get a quantitative agreement with experiments on all aspects. The temperature range in which the transition takes place in the experiments is much narrower than given by the model, irrespectively of the model that we choose for the loop. Although it sounds disappointing, this negative result is perhaps the most important in the thesis because we show clearly that a single strand of DNA cannot be modelled as a simple polymer on a length scale of the order of a few tens of base-pairs, in spite of the claims in the literature that such a picture is valid. Actually studies that claim the validity of such a description either consider much longer segments over which the subtleties of DNA structure are averaged out, or only take into account some aspects of the experimental results so that the discrepancies are hidden.

Zusammenfassung

DNA beacons bestehen aus kurzen DNA Einzelsträngen, die komplementäre Sequenzen in den Regionen der zwei Enden aufweisen. Die Endregionen eines Einzelstrangs können aufgrund dieser Eigenschaft eine kurze DNA Doppelhelix bilden, die mit Stamm bezeichnet wird. Der verbleibende zentrale Teil des Strangs formt eine Windung, den so genannten Loop. In dieser geschlossenen Anordnung bildet der Einzelstrang eine Hairpin-Struktur. Hairpins spielen eine besondere Rolle für die Bestimmung der Sekundärstruktur langer DNA- oder RNA-Einzelstränge. Ein kurzer DNA Einzelstrang, der eine Hairpin-Struktur bilden kann, formt einen so genannten DNA beacon, wenn ein Ende mit eine fluoreszierenden Marker und das andere Ende mit einem Quencher versehen wird. Sind diese Marker nur wenige Angström voneinander entfernt, so verschwindet die Fluoreszenz durch direkten Energietransfer vom fluoreszierenden Molekül zum Quencher. Folglich ist für einen geschlossenen Hairpin keine Fluoreszenz zu beobachten, sie tritt jedoch erneut auf, sobald das Molekül seine Struktur verändert. Diese Eigenschaft ermöglicht den Einsatz molekularer beacons für zahlreiche Anwendungen in der Physik und Biologie. Biologische Anwendungen nutzen die Bildung von Komplexen, bestehend aus dem Einzelstrang, der den Loop beinhaltet, und einem weiteren komplementären DNA Strang. Die Komplexbildung zu einer Doppelhelix erzwingt die Entfaltung des Hairpins, und ein Fluoreszenzsignal wird messbar. In diesem Zusammenhang wurde erwogen, dass DNA beacons in vivo dazu verwendet werden könnten, um einzelne RNA Stränge, die im Verlaufe der Transkription von Genen synthetisiert werden, nachzuweisen. Auf diese Weise wäre es möglich, Krebszellen zu erkennen, indem man gezielt einige Gene beobachtet, die besonders oft in den Krebszellen entschlüsselt werden.

Auch für die Physik sind DNA beacons von besonderem Interesse. Sie können beispielsweise für das Auslesen molekularer Speichereinheiten oder für molekulare Rechenvorgänge verwendet werden. Ihre herausragende Eigenschaft im Hinblick auf das Thema der vorliegenden Arbeit ist ihre Fähigkeit, den Vorgang des Öffnens und des Schließens von DNA Hairpins akkurat wiederzugeben. Eine "Schmelzkurve" des Stamms, hervorgerufen durch Erhitzen, kann auf diese Weise gegen die Temperatur aufgetragen werden; die Autokorrelationsfunktion der Fluoreszenz ermöglicht es, die Kinetik des Öffnens/Schließens zu bestimmen. Es existieren zahlreiche solcher Messungen für unterschiedliche Loop-Längen und Sequenzen, sie bilden einen vollständigen Datensatz und können dazu verwendet werden, das Verständnis der Eigen-

schaften von DNA Hairpins zu erweitern. Dies ist das Ziel der vorliegenden Arbeit. Die Untersuchungen in dieser Arbeit gehen über die Eigenschaften von Hairpins hinaus, da, wie im folgenden gezeigt wird, die Ergebnisse sehr wesentlich von den Eigenschaften des Loops abhängen. Der Vergleich zwischen experimentellen Daten und den Ergebnissen unterschiedlicher Modelle ist daher ein empfindlicher Test für das theoretische Verständnis der Physik einzelner DNA Stränge. Dies schließt Probleme in anderen Bereichen, so zum Beispiel die Modellierung der Eigenschaften von RNA, mitein.
In dieser Arbeit werden zwei Modelle vorgestellt, die die Thermodynamik und die Kinetik solcher Systeme untersuchen. Das erste Modell ist ein zweidimensionales Gittermodell, das auf den Gittermodellen für die Untersuchung der Proteinfaltung beruht. Die Energie des Einzelstrangs wird darin aus lediglich zwei Beiträgen berechnet, einem Beitrag der Krümmungsenergie, die für zueinander rechtwinklig angeordnete Segmente auftritt, und einem Beitrag aus der Bindung von Basenpaaren, die den Stamm bilden. Mithilfe von Monte Carlo Simulationen können die Eigenschaften im thermodynamischen Gleichgewicht und die Kinetik des Systems untersucht werden. Die Ergebnisse stimmen qualitativ mit experimentellen Beobachtungen überein und zeigen, dass die wesentlichen Eigenschaften von DNA Hairpins auf sehr einfache theoretische Überlegungen zurückgeführt werden können. Gleichwohl liegt die Hauptschwäche dieses Modells in der geringen Anzahl von Freiheitsgraden, so dass ein quantitativer Vergleich mit Experimenten nicht möglich ist. Aus diesem Grund wurde ein weiteres Modell entwickelt, das die physikalischen Eigenschaften des Gittermodells berücksichtigt, jedoch auf die räumliche Einschränkung des Gitters verzichtet. Das Modell verknüpft Ideen aus der Polymertheorie mit dem Peyrard-Bishop-Dauxois (PBD) Modell für DNA Schmelzen, und unterteilt ein Hairpin Molekül in zwei Untersysteme:

- den Loop, der als Polymer modelliert wird,

- den Stamm, wiedergegeben durch das PBD Modell unter Verwendung zusätzlicher Terme, die das Wachstum des Loops im Stamm mit in Betracht ziehen.

Dieser neue Zugang ermöglicht es, einen quantitativen Vergleich mit experimentell ermittelten Daten durchzuführen. Es zeigt sich, dass eine gute Übereinstimmung bezüglich der Abhängigkeit der Schmelztemperatur von den Eigenschaften des Loops (Länge und Sequenz) erzielt wird. Ein weiteres Ergebnis ist der Befund, dass die Kinetik des Öffnungsprozesses lediglich von den Eigenschaften des Stamms abhängt und die Rate des Schließungsprozesses mit steigender Loop-Länge abnimmt. Dessen ungeachtet ist es nicht möglich, eine quantitative Übereinstimmung mit allen experimentellen Beobachtungen zu erreichen. So ist das experimentell bestimmte Temperaturintervall, in dem der Übergang stattfindet, deutlich kleiner als durch das Modell vorhergesagt, unabhängig von der genauen Modellierung des

Loops. Obzwar diese Feststellung enttäuschen mag, ist dieses negative Ergebnis möglicherweise die zentrale Aussage der vorliegenden Arbeit: Auf der Längenskala von wenigen Dutzend Basenpaaren kann DNA nicht durch die klassische Polymertheorie erfasst werden, im Widerspruch zu gegenteiligen Behauptungen in der Literatur. Tatsächlich verwendet ein Teil der Studien, die zu solchen Behauptungen kommen, wesentlich längere Segmente, und die lokalen strukturellen Eigenschaften der DNA treten aufgrund von Mittelung nicht hervor. Der andere Teil der Studien schließt experimentelle Beobachtungen bereits in die Modellierung mitein, so dass die Abweichungen vom Polymerverhalten in den Ergebnissen nicht offensichtlich werden.

Résumé

Les “DNA beacons” sont des molécules composées de simple brins d’ADN dont les deux bouts contiennent des bases complémentaires et auxquels on attache un fluorophore et un quencher. Ainsi, ces deux extrémités peuvent s’assembler pour former un bout de double hélice d’ADN que nous appelons “stem”, la partie centrale du brin forme alors une sorte de boucle. On appelle cette structure la configuration en “épingle à cheveux”. Cette configuration joue un rôle important dans la détermination de la structure secondaire des long brins d’ARN ou d’ADN. Lorsque le fluorophore et le quencher sont à proximité l’un de l’autre, c’est-à-dire quelques Å, la fluorescence est bloquée du fait d’un transfert direct d’énergie du fluorophore vers le quencher. Donc, dans la configuration fermée, “l’épingle à cheveux” n’est pas fluorescente. Néanmoins, dans la configuration dite ouverte où les deux extrémités sont désappariées, la fluorescence réapparaît. Cette propriété permet un grand nombre d’applications des “molecular beacons” en Biologie et en Physique. En biologie, ces molécules ont été proposées comme une alternative aux puces à ADN. En effet, si la séquence d’un simple brin d’ADN est complémentaire de la séquence du brin formant la boucle d’une “épingle à cheveux”, il y a appariement entre cette séquence et la boucle. Cela implique une ouverture de “l’épingle à cheveux”, car la rigidité du double brin est bien plus grande que celle du simple brin d’ADN et la molécule devient alors fluorescente. Dans le même esprit, ces molécules ont été proposées pour la détection des cellules cancéreuses en ciblant l’ARN synthétisé par certains gènes de ces cellules. Pour les physiciens, ces molécules sont également très intéressantes. Elles sont à la base de mémoires moléculaires. En effet, la partie boucle d’une “épingle à cheveux” peut servir comme une mémoire où l’on stocke de l’information en utilisant la complémentarité des bases. Le processus d’écriture ou d’effacement est alors suivi par la mesure de fluorescence de ces molécules. Pour notre travail, l’aspect le plus important est qu’elles représentent des systèmes simples permettant une étude détaillée de l’assemblage/désassemblage de la double hélice d’ADN. Les courbes de dénaturation, qui représentent l’évolution de la fluorescence en fonction de la température ainsi que les fonctions d’auto-corrélation de fluorescence peuvent être mesurées très précisément, ce qui permet d’extraire les propriétés thermodynamiques et cinétiques de cette structure en “épingle à cheveux”. Des mesures ont été faites avec différents types de bases et différentes longueurs de boucle, donnant ainsi un grand nombre de données. Ce sont ces propriétés physiques qui nous intéressent dans cette thèse. La comparaison des résultats expérimentaux et des

résultats obtenus par différents modèles est un excellent moyen pour tester notre capacité à modéliser les propriétés de l'ADN.
Nous avons développé deux modèles différents pour étudier la thermodynamique et la cinétique de ces systèmes. Le premier est un modèle sur réseau inspiré des modèles sur réseau utilisés pour l'étude des repliements des protéines. Dans ce modèle, l'énergie du simple brin d'ADN, dépend seulement de deux termes, un terme pour le coût énergétique associé à un angle entre deux bases consécutives et un terme de gain énergétique pour la formation d'une paire de bases. A partir de simulations Monte Carlo, nous avons étudié les propriétés d'équilibre et la cinétique du système. Les résultats obtenus à l'aide de ce modèle sont en accord qualitatifs avec les résultats expérimentaux montrant ainsi que les principales propriétés des "épingles à cheveux" sont gouvernées par des phénomènes physiques simples. Néanmoins, la principale faiblesse de ce modèle réside dans le manque de degrés de liberté qui ne permet donc pas une comparaison quantitative avec les expériences. Nous avons donc élaboré un autre modèle qui inclut les ingrédients physiques du premier modèle mais sans la contrainte apportée par le réseau. Il combine la théorie des polymères et le modèle de Peyrard-Bishop et Dauxois (PBD) pour la double hélice. Le système est alors divisé en deux sous-système:

- la boucle qui est modélisée par un polymère,

- la partie double brin d'ADN qui est modélisée par le modèle PBD et complété par des termes pour tenir compte de l'agrandissement de la boucle le long du stem.

Avec cette nouvelle approche, nous sommes capable de comparer quantitativement nos résultats théoriques avec les résultats expérimentaux. Nous trouvons un bon accord pour la dépendance de la "température de transition" avec les caractéristiques de la boucle, à savoir, la longueur et la nature de la séquence. De plus, les résultats de cinétique sont en accord qualitatif avec les résultats expérimentaux. En effet, nous trouvons que la cinétique d'ouverture est déterminée par les propriétés du "stem" seulement et que la vitesse de fermeture décroît avec la longueur de la boucle. Cependant, nous ne sommes pas capable d'obtenir une comparaison quantitative complète. Nous obtenons une largeur de transition environ deux fois plus grande que celle obtenue dans les expériences, indépendamment du modèle de boucle. Aussi surprenant que cela puisse paraître, ce résultat négatif est peut-être l'un des résultats les plus important de ce travail de thèse parce qu'il montre clairement qu'un simple brin d'ADN ne peut pas être modélisé par un simple polymère à l'échelle de quelques dizaines de paires de bases, en dépit de ce que dit la littérature portant sur ce sujet.

Part III

Appendices

Appendix A

Calculation of $P_N(R)$ for the Kratky-Porod chain

This appendix explains the method proposed by N.Theodorakopoulos to compute the probability distribution function of the end-to-end distance of a Kratky-Porod chain.

Our calculation for the hairpin involves the probability distribution function for the extension of the chain $S(r|R)$. But for a chain like the Kratky-Porod chain which includes an energy contribution depending on the angle between segments, the probability distribution of an $(N+1)^{\text{th}}$ segment depends on the spatial orientation $\mathbf{X}_N$ of the n^{th} segment. This suggests that the appropriate distribution for the Kratky-Porod chain is not

$$P_N(\mathbf{R}) = \frac{1}{Z_N} \int d\Omega_1 ... d\Omega_N e^{-\beta H(\mathbf{X}_j)} \delta\left(\mathbf{R} - \sum_{j=1}^{N} \mathbf{X}_j\right), \tag{A.1}$$

but the end-to-end vector distribution function at fixed direction $\mathbf{X}_N$ of the N^{th} segment, i.e.

$$\widetilde{P}_N\left(\mathbf{R}; \mathbf{X}_N\right) = \frac{1}{Z_N} \int \prod_{j=1}^{N-1} d\Omega_j e^{-\beta H(\mathbf{X}_j)} \delta\left(\mathbf{R} - \sum_{j=1}^{N} \mathbf{X}_j\right). \tag{A.2}$$

The probability distribution A.1 for the end-to-end vector is related to $\widetilde{P}_N\left(\mathbf{R}; \mathbf{X}_N\right)$ by

$$P_N(\mathbf{R}) = \int d\Omega_N \widetilde{P}_N\left(\mathbf{R}; \mathbf{X}_N\right). \tag{A.3}$$

The method proposed by N.Theodorakopoulos uses an expansion of $\widetilde{P}_N\left(\mathbf{R}; \mathbf{X}_N\right)$ in terms of spherical harmonics

$$\widetilde{P}_N\left(\mathbf{R}; \mathbf{X}_N\right) = \sum_{lm} \widetilde{Q}_{lm}^{(N)}(\mathbf{R}) Y_{lm}(\Omega_N), \tag{A.4}$$

where the expansion coefficients are defined as

$$\widetilde{Q}_{lm}^{(N)}(\mathbf{R}) = \int d\Omega_N \widetilde{P}_N\left(\mathbf{R};\mathbf{X}_N\right) Y_{lm}^{*}(\Omega_N). \tag{A.5}$$

The end-to-end distribution function is obtained from the lowest coefficient by

$$P_N(\mathbf{R}) = \sqrt{4\pi} Q_{00}^{(N)}(\mathbf{R}). \tag{A.6}$$

The idea of the calculation is to build $P_N(\mathbf{R})$ by gradually adding segments to an initial segment. Therefore one needs to define a recurrence relation

$$\widetilde{P}_{N+1}\left(\mathbf{R};\mathbf{X}_{N+1}\right) = \frac{Z_N}{Z_{N+1}} \int d\Omega_N d\mathbf{r}' \delta\left(\mathbf{R}-\mathbf{r}'-\mathbf{X}_{N+1}\right) \times \\ e^{b(\mathbf{X}_N.\mathbf{X}_{N+1}-1)} \widetilde{P}_N\left(\mathbf{R};\mathbf{X}_N\right). \tag{A.7}$$

Using the expression of Z_N as a function of $i_0(b)$, one gets

$$\widetilde{P}_{N+1}\left(\mathbf{R};\mathbf{X}_{N+1}\right) = \int d\Omega_N d\mathbf{r}' \delta\left(\mathbf{R}-\mathbf{r}'-\mathbf{X}_{N+1}\right) \times \\ \phi\left(\mathbf{X}_N,\mathbf{X}_{N+1}\right) \widetilde{P}_N\left(\mathbf{R};\mathbf{X}_N\right), \tag{A.8}$$

with

$$\phi\left(\mathbf{X}_N,\mathbf{X}_{N+1}\right) = \frac{e^{b(\mathbf{X}_N.\mathbf{X}_{N+1}-1)}}{4\pi i_0(b)}, \tag{A.9}$$

which can be expanded in terms of spherical harmonics

$$\phi\left(\mathbf{X}_N,\mathbf{X}_{N+1}\right) = \sum_{l,m} \widehat{i}_l(b) Y_{lm}(\Omega_N) Y_{lm}^{*}(\Omega_{N+1}), \tag{A.10}$$

with

$$\widehat{i}_l(b) = \frac{i_l(b)}{i_0(b)}, \tag{A.11}$$

expressed in terms of modified Bessel functions. With the spherical harmonic expansion of ϕ, the angular integral of A.8 can be performed. The result is

$$\widetilde{P}_{N+1}\left(\mathbf{R};\mathbf{X}_{N+1}\right) = \int \frac{d\mathbf{q}}{(2\pi)^3} d\mathbf{r}' e^{i\mathbf{q}.(\mathbf{R}-\mathbf{r}')} e^{-i\mathbf{q}.\mathbf{X}_{N+1}} \times \\ \sum_{l,m} \widehat{i}_l(b) \widetilde{Q}_{lm}^{(N)}(\mathbf{r}') Y_{lm}(\Omega_{N+1}), \tag{A.12}$$

in which we have introduced the Fourier transform of the δ function.
Multiplying both sides by $Y_{l'm'}^{*}(\mathbf{X}_N)$ and integrating over Ω_{N+1} extracts the expression of $\widetilde{Q}_{l'm'}^{(N+1)}$

$$\widetilde{Q}_{l'm'}^{(N)}(\mathbf{R}) \int \frac{d\mathbf{q}}{(2\pi)^3} d\mathbf{r}' e^{i\mathbf{q}.(\mathbf{R}-\mathbf{r}')} \times \\ \sum_{l} \widehat{i}_l(b) f_{ll'}^{(m')}(q) \widetilde{Q}_{lm'}^{(N)}(\mathbf{r}'), \tag{A.13}$$

where

$$f_{ll'}^{(m')}(q) = \int d\Omega_j e^{-i\mathbf{q}.\mathbf{X}_j Y_{lm}}(\Omega_j) Y_{l'm}^{*}(\Omega_j). \tag{A.14}$$

As we are interested in the case $m' = 0$ because we need $\widetilde{Q}_{00}^{(N)}$, Eq. (A.13) reduces to

$$\widetilde{Q}_{l'0}^{(N)}(\mathbf{R}) = \int \frac{d\mathbf{q}}{(2\pi)^3} d\mathbf{r}' e^{i\mathbf{q}\cdot(\mathbf{R}-\mathbf{r}')} \times \sum_{l} \widehat{i}_l(b) f_{ll'}^{(0)}(q) \widetilde{Q}_{l0}^{(N)}(\mathbf{r}'), \tag{A.15}$$

where

$$f_{ll'}^{(0)}(q) = \frac{1}{2}\sqrt{(2l+1)(2l'+1)} \int_{-1}^{+1} d\mu e^{-iq\mu} P_l(\mu) P_{l'}(\mu), \tag{A.16}$$

where P_l is a Legendre polynomial. In Fourier space Eq. (A.15) becomes

$$\widetilde{Q}_{l'0}^{(N)}(\mathbf{q}) = \sum_{l} \widehat{i}_l(b) f_{ll'}^{(0)}(q) \widetilde{Q}_{l0}^{(N)}(\mathbf{q}), \tag{A.17}$$

which can be expressed in a matrix form by defining a vector $\mathbf{Q}^{(N)}$ and a symmetric matrix $\mathbf{F}$ by

$$Q_l^{(N)}(\mathbf{q}) = \sqrt{\widehat{i}_l(b)} \widetilde{Q}_{l0}^{(N)}(\mathbf{q}) \tag{A.18}$$

$$F_{ll'}(\mathbf{q}) = \sqrt{\widehat{i}_l(b)\widehat{i}_{l'}(b)} f_{ll'}^{(0)}(\mathbf{q}). \tag{A.19}$$

The recurrence relation is now

$$\mathbf{Q}^{(N+1)} = \mathbf{F}\mathbf{Q}^{(N)}, \tag{A.20}$$

and the end-to-end distribution function is given by

$$P_N(\mathbf{R}) = \sqrt{4\pi} Q_0^{(N)}(\mathbf{R}). \tag{A.21}$$

The recurrence relation (A.20) provides the basis for the calculation of $P_N(\mathbf{R})$. For this one needs to start from $N = 1$

$$\widetilde{P}_1(\mathbf{R}; \mathbf{X}_1) = \frac{1}{4\pi} \delta(\mathbf{R} - \mathbf{X}_1). \tag{A.22}$$

So that

$$P_1(\mathbf{R}) = \int d\Omega_1 \widetilde{P}_1(\mathbf{R}; \mathbf{X}_1) = \frac{1}{4\pi} \delta(R - 1). \tag{A.23}$$

From the expansion of $\widetilde{P}_1(\mathbf{R}; \mathbf{X}_1)$ we get

$$\widetilde{Q}_{lm}^{(1)}(q) = \frac{1}{\sqrt{4\pi}} f_{l0}^{(0)}(q) \delta_{m0}, \tag{A.24}$$

or

$$Q_l^{(1)} = \frac{1}{\sqrt{4\pi}} F_{l0}. \tag{A.25}$$

Now with the recurrence relation we get

$$Q_l^{(N)} = \frac{1}{\sqrt{4\pi}} \left[\mathbf{F}^N\right]_{l0}. \tag{A.26}$$

Therefore the Fourier transform of the end-to-end distribution is given by

$$P_N(\mathbf{q}) = \left[\mathbf{F}^N\right]_{00}. \tag{A.27}$$

If we know the matrix elements of $\mathbf{F}$, we can then get $P_N(\mathbf{q})$ and $P_N(\mathbf{R})$ by inverse Fourier transform. Their calculation is possible with the expansion

$$e^{-iq\mu} = \sum_{k=0}^{\infty}(2k+1)(-i)^k j_k(q) P_k(\mu), \tag{A.28}$$

where the j_k are the spherical Bessel functions (e.g. $j_0(q) = \sin q/q$).
Putting this expression into formula for $f_{ll'}^{(0)}(q)$, and using the integral formula for the product of three Legendre polynomials [60], it is possible to express the matrix elements of $\mathbf{F}$ as a finite sum of Bessel functions. (Eq.(31) in [60]).

Appendix B

The Gaussian chain

B.1 Theoretical predictions

We consider the case of a chain with monomer modelled by springs which are randomly oriented and totally independent from each other. Each monomer has a fixed equilibrium length l_0. We assume that the spring konstant K does not depend on T and we consider the case $l_0 \neq 0$, contrary to the case usually presented in the litterature. We will see even in such a simple polymer model that the calculations could be non trivial. Figure (B.1) gives a representation of the Gaussian chain.

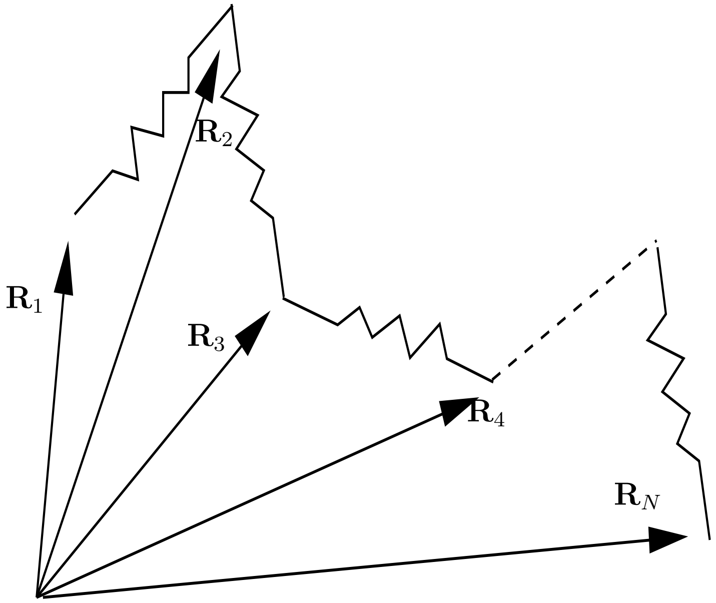

Figure B.1: *Modelling of the Gaussian chain.*

Using this model we can define the energy of such a chain, which is in this purely harmonic case:

$$
\begin{aligned}
U &= \frac{1}{2}K\sum_{i=1}^{N}(\|\mathbf{R}_i - \mathbf{R}_{i-1}\| - l_0)^2 \\
U &= \frac{1}{2}K\sum_{i=1}^{N}(r_i - l_0)^2
\end{aligned}
\tag{B.1}
$$

We would like to calculate: $\langle \mathbf{r}_i \rangle$, $\langle \|\mathbf{r}_i\| \rangle$, $\langle \mathbf{r}_i^2 \rangle$, $\langle (\mathbf{R}_N - \mathbf{R}_0)^2 \rangle$, the gyration radius R_g^2 and finally $\langle U \rangle$. The same method could be used to calculate other quantities.
First of all $\langle \mathbf{r}_i \rangle$:

$$
\begin{aligned}
\langle \mathbf{r}_i \rangle &= \frac{\displaystyle\int \prod_{j=1}^{N} d\mathbf{r}_j \ \mathbf{r}_i \ \exp(-\frac{\beta K}{2}\sum_{j=1}^{N}(\|\mathbf{r}_j\| - l_0)^2)}{\displaystyle\int \prod_{j=1}^{N} d\mathbf{r}_j \ \exp(-\frac{\beta K}{2}\sum_{j=1}^{N}(\|\mathbf{r}_j\| - l_0)^2)} \\
\langle \mathbf{r}_i \rangle &= \frac{\displaystyle\int d\mathbf{r}_i \ \mathbf{r}_i \ \exp(-\frac{\beta K}{2}(\|\mathbf{r}_i\| - l_0)^2)}{\displaystyle\int d\mathbf{r}_i \ \exp(-\frac{\beta K}{2}(\|\mathbf{r}_i\| - l_0)^2)} \\
\langle \mathbf{r}_i \rangle &= 0
\end{aligned}
\tag{B.2}
$$

This result is trivial because in this model each monomer is independent from the others and randomly oriented.
Let us now consider $\langle \|\mathbf{r}_i\| \rangle$:

$$
\begin{aligned}
\langle \|\mathbf{r}_i\| \rangle &= \frac{\displaystyle\int d\mathbf{r}_i \ \|\mathbf{r}_i\| \ \exp(-\frac{\beta K}{2}(\|\mathbf{r}_i\| - l_0)^2)}{\displaystyle\int d\mathbf{r}_i \ \exp(-\frac{\beta K}{2}(\|\mathbf{r}_i\| - l_0)^2)} \\
\langle r \rangle &= \frac{\displaystyle\int_0^\infty dr \ r^2 \ \exp(-\frac{\beta K}{2}(r - l_0)^2)}{\displaystyle\int_0^\infty dr \ r \ \exp(-\frac{\beta K}{2}(r - l_0)^2)}
\end{aligned}
\tag{B.3}
$$

Due to the presence of l_0, the calculation of the two previous integrals is not dimmediate. Nevertheless one can easily show that:

$$\begin{aligned} Z_1 &= \int_0^{\infty} dr\ r\ \ \exp(-\frac{\beta K}{2}(r-l_0)^2) \\ Z_1 &= \frac{1}{\beta K}\exp(-\frac{\beta K}{2}l_0^2) + \frac{l_0}{2}\sqrt{\frac{2\pi}{\beta K}}\Big(\mathrm{Erf}(\frac{1}{2}l_0\sqrt{2\beta K})+1\Big) \end{aligned} \tag{B.4}$$

Where Erf is the error function [57]. In the same way we have:

$$\begin{aligned} \int_0^{\infty} dr\ r^2\ \ \exp(-\frac{\beta K}{2}(r-l_0)^2) =& \frac{l_0}{\beta K}\exp(-\frac{\beta K}{2}l_0^2) + \\ & \frac{1}{2}\sqrt{\frac{2\pi}{\beta K}}\Big(\mathrm{erf}(\frac{1}{2}l_0\sqrt{2\beta K})+1\Big)\Big(\frac{\beta K l_0^2+1}{\beta K}\Big) \end{aligned} \tag{B.5}$$

Putting (B.4) and (B.5) in (B.3), we get:

$$\langle r\rangle = \frac{1}{Z_1}\left(\frac{l_0}{\beta K}\exp(-\frac{\beta K}{2}l_0^2) + \frac{1}{2}\sqrt{\frac{2\pi}{\beta K}}\Big(\mathrm{erf}(\frac{1}{2}l_0\sqrt{2\beta K})+1\Big)\Big(\frac{\beta K l_0^2+1}{\beta K}\Big)\right) \tag{B.6}$$

In the same spirit we can calculate $\langle {\mathbf{r}_i}^2\rangle$:

$$\begin{aligned} \langle {\mathbf{r}_i}^2\rangle &= \frac{\displaystyle\int d\mathbf{r}_i\ {\mathbf{r}_i}^2\ \ \exp(-\frac{\beta K}{2}(\|\mathbf{r}_i\|-l_0)^2)}{\displaystyle\int d\mathbf{r}_i\ \ \exp(-\frac{\beta K}{2}(\|\mathbf{r}_i\|-l_0)^2)} \\ \langle r^2\rangle &= \frac{\displaystyle\int_0^{\infty} dr\ r^3\ \ \exp(-\frac{\beta K}{2}(r-l_0)^2)}{\displaystyle\int_0^{\infty} dr\ r\ \ \exp(-\frac{\beta K}{2}(r-l_0)^2)} \end{aligned} \tag{B.7}$$

Using (B.4), (B.5) and usual integration methods we get:

$$\begin{aligned} \langle r^2\rangle = \frac{1}{Z_1}\Bigg(&\Big(\frac{2+\beta K l_0^2}{(\beta K)^2}\Big)\exp(-\frac{\beta K}{2}l_0^2) + \\ & \frac{l_0}{2}\sqrt{\frac{2\pi}{\beta K}}\Big(\mathrm{erf}(\frac{1}{2}l_0\sqrt{2\beta K})+1\Big)\Big(\frac{\beta K l_0^2+3}{\beta K}\Big)\Bigg) \end{aligned} \tag{B.8}$$

We can now easily derive the mean end to end distance of the chain using the fact that the monomers are independent from each other:

$$\begin{aligned}
\left\langle (\mathbf{R}_N - \mathbf{R}_0)^2 \right\rangle &= \left\langle ((\mathbf{R}_N - \mathbf{R}_{N-1}) + (\mathbf{R}_{N-1} - \mathbf{R}_{N-2}) + \cdots + (\mathbf{R}_1 - \mathbf{R}_0))^2 \right\rangle \\
\left\langle (\mathbf{R}_N - \mathbf{R}_0)^2 \right\rangle &= \sum_{i=1}^{N} \left\langle \mathbf{r}_i^2 \right\rangle \qquad \text{(B.9)} \\
\left\langle (\mathbf{R}_N - \mathbf{R}_0)^2 \right\rangle &= N \left\langle r^2 \right\rangle
\end{aligned}$$

Therefore, we immediately have the expression of $< U >$:

$$\begin{aligned}
\langle U \rangle &= \frac{1}{2} K \sum_{i=1}^{N} \left\langle (r_n - l_0)^2 \right\rangle \\
\langle U \rangle &= \frac{1}{2} N K \left(l_0^2 + \left\langle r^2 \right\rangle - l_0 \langle r \rangle \right)
\end{aligned} \qquad \text{(B.10)}$$

Before giving the expression of the gyration radius, let us notice that, if $l_0 \equiv 0$, then we find the usual results for a harmonic system with two degrees of freedom:

$$\begin{aligned}
\langle r \rangle &= \frac{1}{2} \sqrt{\frac{2\pi k_b T}{K}} \\
\left\langle r^2 \right\rangle &= \frac{2 k_b T}{K} \\
\langle U \rangle &= N k_b T
\end{aligned} \qquad \text{(B.11)}$$

Finally, we give the exact result of the radius gyration as well as its value in the limit of big N:

$$\begin{aligned}
R_g^2 &= \frac{1}{2N^2} \sum_{n,m} \left\langle (\mathbf{R}_n - \mathbf{R}_m)^2 \right\rangle \\
R_g^2 &= \frac{1}{2N^2} \sum_{n} \sum_{m} |n - m| \left\langle r^2 \right\rangle \\
R_g^2 &\approx \frac{1}{2N^2} \int_0^N \int_0^N |n - m| \left\langle r^2 \right\rangle \\
R_g^2 &\approx \frac{N}{6} \left\langle r^2 \right\rangle
\end{aligned} \qquad \text{(B.12)}$$

B.2 Monte Carlo simulation

We have developed a program which models this Gaussian chain. In our simulation we have chosen for simplicity $K \equiv 1$, $l_0 \equiv 1$ and $k_b \equiv 1$.
We have used the Monte Carlo algorithm presented in chapter 4. Here we present the mean values obtained numerically and compare it to the theoretical results. One can notice that the numerical results are in total agreement with the theoretical ones. This valid *a posteriori* the theoretical expressions derived for such quantities.

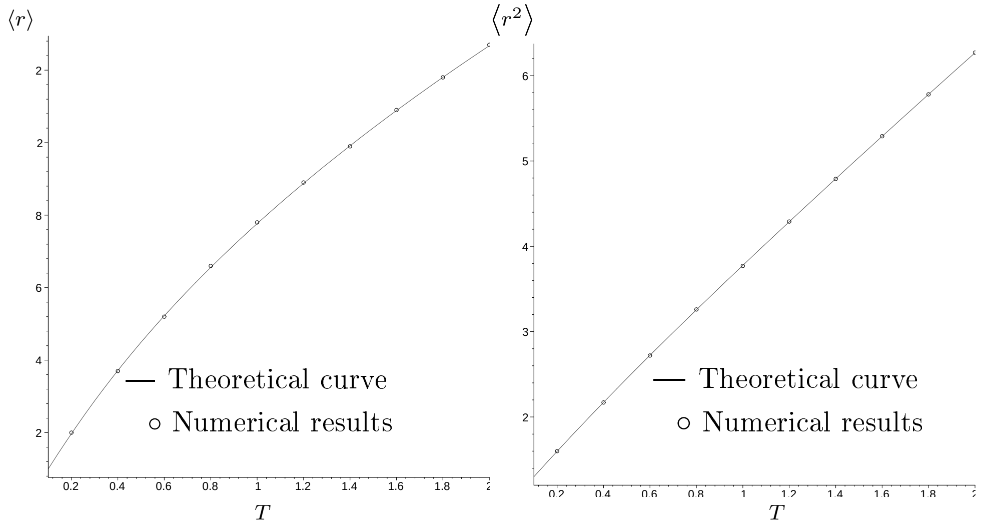

Figure B.2: *Square mean length and mean length of a monomer. Left: mean length. Right: square mean length*

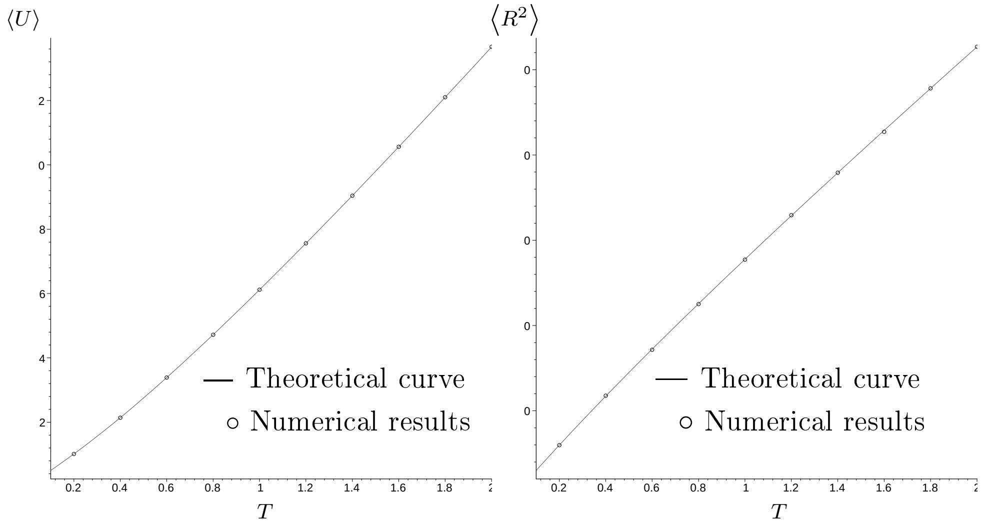

Figure B.3: *potential energy of the chain and square mean end-to-end distance. Left: potential of the chain. Right: square mean end-to-end distance.*

List of Figures

1 Schematic representation of a DNA hairpin configuration xi

1.1 Numeration of the carbon-atom in the sugar 5
1.2 Schematic form of the double chain 5
1.3 The double helix of Crick and Watson 5
1.4 Pairing of complementary bases 5
1.5 A,B and Z form of the DNA double helix 6
1.6 Schematic representation of replication of DNA 8
1.7 Schematic representation of transcription of DNA 8
1.8 Melting curves example 9
1.9 Schematic representation of the Poland-Scheraga model 11
1.10 Topology of the loop embedded in a chain 12
1.11 Peyrard-Bishop model for DNA 13
1.12 DNA Helicoidal Model 14
1.13 Schematic representation of ssDNA 15
1.14 Schematic represensation of the memory DNA and the data DNA . . 16
1.15 Schematic representation of the writing process 17
1.16 Schematic representation of the erasing process 18
1.17 Schematic representation of RNA loop 18

2.1 Donor and acceptor absorption and emission spectra 23
2.2 Schematic representation of the two states 23
2.3 Normalized melting curves for different loop lengths 25
2.4 Schematic drawing of the experimental setup 26
2.5 Arrhenius plots of the opening and the closing rates 27
2.6 Comparison of the opening and the closing rates 28
2.7 Closing enthalpy vs loop lengths 28
2.8 Schematic representation of a spectrophotometer 29
2.9 Schematic representation of some microstates 31
2.10 Fits to the equilibrium melting profiles 32

3.1 Freely jointed chain 34
3.2 Probability distribution of the end-to-end distance 37
3.3 Freely rotating chain 38

3.4 Probability distribution of the Freely Rotating Chain 41
3.5 Comparison of the effective Gaussian probability distribution function and the exact expression for N=10 and N=32. 45
3.6 Comparison between the melting curves obtained with the effective Gaussian and the exact expression of the probability distribution function for N=12 and N=24. 46
3.7 Comparison of $P_{N+2}(r)$ obtained using Eq. (3.50) and the real form with the FRC 48
3.8 Comparison of $P_{N+2}(r)$ obtained using Eq. (3.50) and the real form with the KP 48
3.9 A compact conformation of the 49-mer on the square lattice 51

4.1 Two configurations of the hairpin model in a lattice 56
4.2 The three possible motions 57
4.3 Variation versus temperature of the number of hydrogen-bonded pairs 60
4.4 Effect of the rigidity of the loop on the opening of the hairpin 61
4.5 Comparison of melting curves with and without mismatches 62
4.6 Schematic plot of the fluctuations of the free end of the chain 63
4.7 Normalized histograms of the distance d (a), and number of hydrogen bonds (b) 64
4.8 Arrhenius plot of the kinetic constants 65
4.9 Logarithmic plot of the characteristic time for opening τ versus $1/T$. 67

5.1 Drawing representation of the system 71
5.2 Schematic representation of the stem 71
5.3 Schematic representation of the potential $V(y)$ 72
5.4 Representation of eigenfunctions 77
5.5 Evolution of the eigenvalues as a function of temperature 77
5.6 Free energy of a finite stem 78
5.7 Temperature variation of the entropy of the stem 78
5.8 Free energy landscape for different temperature 80
5.9 Influence of the excited states on the mean distance of the first base-pair 81
5.10 Schematic representation of the growth of the polymer 82
5.11 Example of Free energy profile and Entropy 85
5.12 Chemical equilibrium 85
5.13 Example of a free energy profile. 85
5.14 Melting curve obtained for a stem of five base-pairs with and without a loop 92
5.15 Melting curves with the FRC model: $\theta = 45°$ 93
5.16 Melting curves with the FRC model: $\theta = 60°$ 93
5.17 Melting curves with the Kratky-Porod chain: ϵ=0.0019 eV.Å^{-2} 94
5.18 Melting curves with the Kratky-Porod chain: ϵ=0.0040 eV.Å^{-2} 94
5.19 Plot of the probability distribution for the Kratky-Porod chain 95

5.20 Effect of D and k on the melting curve 96
5.21 Effect of α and ρ on the melting curve 97
5.22 Effect of ϵ and D on the melting curves with the Kratky-Porod chain 97
5.23 Rates of opening and closing with the FRC model in an Arrhenius plot. 98
5.24 Effect of D and k on the kinetics with the FRC in an Arrhenius plot. 99
5.25 Rates of opening and closing with the Kratky-Porod chain in an Arrhenius plot 100
5.26 Effect of D and k on the kinetics with the Kratky-Porod chain in an Arrhenius plot 101
5.27 Comparison of the melting curves with $S \equiv 1$ and $S \neq 1$ with the FRC model: $\theta = 60°$ 102
5.28 Melting curves equivalent to poly(T) with the FRC model 103
5.29 Variation of T_m as a function of N 103
5.30 Melting curves equivalent to poly(A) with the FRC model 104
5.31 Effect of the depth of the Morse potential on the melting profiles with the FRC model 105
5.32 Effect of the width of the Morse potential on the melting profiles with the FRC model 106
5.33 Effect of the rigidity of the stem on the melting profiles with the FRC model 106
5.34 Comparison of the melting curves with $S \equiv 1$ and $S \neq 1$ with the KP model: $\epsilon = 0.0019\ eV.\text{Å}^{-2}$ 107
5.35 Variation of T_m as a function of N with and without the S function . 108
5.36 Melting curves equivalent to poly(T) with the KP model 109
5.37 Melting curves equivalent to poly(A) with the KP model 110
5.38 Effect of the depth of the Morse potential on the melting profiles with the KP model 110
5.39 Comparison of the kinetic rates with and without S with the FRC model in an Arrhenius plot 111
5.40 Rates of opening and closing with the FRC model in an Arrhenius plot. 112
5.41 Effect of D on the kinetics with the FRC model in an Arrhenius plot. 113
5.42 Evolution of the activation energy of opening as a function of D. . . . 113
5.43 Effect of k and α on the kinetics with the FRC in an Arrhenius plot. 114
5.44 Comparison of the kinetic rates with and without S with the KP model in an Arrhenius plot 115
5.45 Rates of opening and closing with the KP model in an Arrhenius plot. 115
5.46 Schematic representation of the stacking in the closed and the open configuration 118
5.47 Comparison of the melting curves and the energies obtained with two stacking potentials W and W_1. 118

B.1 Modelling of the Gaussian chain 141
B.2 Square mean length and mean length of a monomer. 145

B.3 Potential energy of the chain and square mean end-to-end distance. . 145

Bibliography

[1] S. Cuesta-López, M. Peyrard and D. J. Graham, *Model for DNA hairpin denaturation.* Eur. Phys. J. E **16**, 235-246 (2005).

[2] H. Lodish, A. Berk, L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, in *Molecular cell Biology*, W.H. Freeman, New York, 4th edition (2000)

[3] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Plenum Publishing Corporation, 2nd edition (1999).

[4] G. Bonnet, O. Krichevsky and A. Libchaber, *Kinetics of conformational fluctuations in DNA hairpin-loops*, Proc. Natl. Acad. Sci. USA **95**, 8602-8606 (1998).

[5] G. Bonnet, S. Tyagi, A. Libchaber and F. R. Kramer, *Thermodynamic basis of the enhanced specificity of structured DNA Probes*, Proc. Natl. Acad. Sci. USA **96**, 6171-6176 (1999).

[6] S. V. Kuznetsov, Y. Schen, A. S. Benight and A. Ansari, *A semiflexible Polymer Model applied to loop formation in DNA hairpin*, Biophys. J. **81**, 2864-2875 (2001).

[7] A. Ansari, Y. Schen, and S. V. Kuznetsov,*Configuration diffusion down a folding funnel describes the dynamics of DNA hairpins*, Proc. Natl. Acad. USA **98**, 7771-7776 (2001).

[8] M. I. Wallace, L. Ying, S. Balasubramanian, and D. Klenerman, *Non-Arrhenius kinetics for the loop closure of a DNA hairpin*, Proc. Natl. Acad. Sci. USA **98**, 5584-5589 (2001).

[9] S. Cuesta-López, J. Errami, F. Falo, and M. Peyrard, *Can We Model DNA at the Mesoscale?*, J. Biol. Phys. **31**, 273-301 (2005).

[10] M. Peyrard and A. Bishop, *Statistical mechanics of a nonlinear model for DNA denaturation*, Phys. Rev. Lett. **62**, 2755-2758 (1989).

[11] N. Theodorakopoulos, T. Dauxois and M. Peyrard, *Order of the phase transition in models of DNA thermal denaturation*, Phys. Rev. Lett. **85**, 6-9 (2000).

[12] C.R. Calladine and H. R. Drew, *Understanding DNA*, Academic Press (1992)

[13] J.D. Watson and F.H.C. Crick, *Molecular structure of nucleic acids*, Nature **171**, 737-738 (1953).

[14] W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag (1984)

[15] A. Wada, S. Yabuki and Y. Husimi, *Fine structure in the thermal denaturation of DNA: High temperature-resolution spectrophotometric studies*, Critical Reviews in Biochemistry **9**, 87-144 (1980).

[16] Internet site, *http://nobelprize.org*

[17] Internet site, *http://www.eas.slu.edu*

[18] M. J. Doktycz, M. D. Morris, S. J. Dormdy, K. L. Beattie, and K. B. Jacobson, *Optical Melting of 128 Octamer DNA Duplexes*, J. Biol. Chem. **270**, 8439-8445 (1995).

[19] T. Dauxois, M. Peyrard and A.R. Bishop, *Dynamics and thermodynamics of a nonlinear model for DNA denaturation*, Phys. Rev. E **47**, 684-695 (1993).

[20] T. Dauxois, N. Theodorakopoulos and M. Peyrard, *Thermodynamic instabilities in one dimension: correlations, scaling and solitons*, J. Stat. Phys. **107**, 869-883 (2002).

[21] T.Dauxois, A. Litvak-Hinenzon, R. Mackay, and A. Spanoudaki, *Energy Localisation and Transfer*, Advanced Series in Nonlinear Dynamics **22**, World Scientific publishing (2004).

[22] C. Kittel, *Phase Transition of a Molecular Zipper*, Am. J. Phys. **37**, 917-920 (1969).

[23] D. Poland and H. A. Scheraga, *Phase Transitions in One Dimension and the Helix-Coil Transition in Polyamino Acids*, J. Chem. Phys. **45**, 1456-1463 (1966).

[24] D. Poland and H. A. Scheraga, *Ocurrence of a Phase Transition in Nucleic Acid Models*, J. Chem. Phys. **45**, 1464-1469 (1966).

[25] B. H. Zimm, *Theory of Melting of the Helical Form in Double Chains of the DNA Type*, J. Chem. Phys. **33**, 1349-1346 (1960).

[26] R.M. Wartell and A.S. Benight, *Thermal denaturation of DNA molecules: a comparison of theory with experiments*, Physics Reports **126**, 67-107 (1985).

[27] M. E. Fisher, *Effect of Excluded Volume on Phase Transitions in Biopolymers*, J. Chem. Phys. **45**, 1469-1473 (1966).

[28] Y. Kafri, D. Mukamel and L. Peliti, *Why is the DNA Transition First Order*, Phys. Rev. Lett. **85**, 4988-4991 (2000).

[29] R. Blossey and E. Carlon, *Reparametrizing the loop entropy weights: Effect on DNA melting curves*, Phys. Rev. E **68**, 061911 (2003).

[30] M. Y. Azbel, *Random Two-Component-Dimensional Ising Model for Heteropolymer Melting*, Phys. Rev. Lett. **31**, 589-592 (1973).

[31] T. Garel and H. Orland, *Generalized Poland-Scheraga model for DNA hybridization*, Biopolymers **75**, 453-467 (2004).

[32] T. Dauxois, *Dynamique non linéaire et mécanique statistique d'un modèle d'ADN*, thèse, Université de Bourgogne (1993).

[33] S. Cocco, *Uno studio teorico delle vibrazioni della struttura elicoidale del DNA B e della transizione di denaturazione*, tesi, Università Degli Studi di Roma La Sapienza (2000).

[34] M. Barbi, S. Cocco and M. Peyrard, *Helicoidal model for DNA opening*, Phys. Lett. A. **253**, 358-369 (1999).

[35] S. Cocco, R. Monasson and J. F. Marko, *Force and kinetic barriers to unzipping of the DNA double helix*, Proc. Natl. Acad. Sci. USA **98**, 8608-8613 (2001).

[36] M. Barbi, S. Lepri, M. Peyrard and N. Theodorakopoulos, *Thermal denaturation of a helicoidal DNA model*, Phys. Rev. E. **68**, 061909 (2003).

[37] N. L. Goddart, G. Bonnet, O. Krichevsky and A. Libchaber, *Sequence dependence rigidity of single-stranded DNA*, Phys. Rev. Lett. **85**, 2400-2403 (2000).

[38] D. P. Aalberts, J. M. Parman, and N. L. Goddard, *Single-Strand Stacking Free Energy from DNA Beacon Kinetics*, Biophys. J. **84**, 3212-3217 (2003).

[39] K. Hamad-Schifferli, J. J. Schwartz, A. T. Santos, S. Zhang and J. M. Jacobson, *Remote electronic control of DNA hybridization through inductive coupling to an attached metal nanocrystal antenna*, Nature **415**, 152-155 (2002).

[40] M. Sales-Pardo, R. Guimerà, A. A. Moreira, J. Widom, and L. A. N. Amaral, *Numerical evidence for zipping in the opening and closing of DNA hairpins*, unpublished.

[41] A.V. Tkachenko, *Unfolding and unzipping of single stranded DNA by stretching*, Phys. Rev. E **70**, 051901 (2004).

[42] P.S. Shockett and D.G. Schatz, *DNA hairpins Opening Mediated by the RAG1 an RAG2 proteins*, Mol. Cell. Biol. **19**, 4159-4166 (1999).

[43] S. Tyagi, S. A. E. Marras, and F. R. Kramer, *Wavelength-shifting molecular beacons*, Nature Biotechnology **18**, 1191-1196 (2000).

[44] M. Schena, *DNA microarrays -a practical approach*, Oxford University Press (1999).

[45] M. Takinoue and A. Suyama, *Molecular reactions for a molecular based on hairpin DNA* Chem-Bio Informatics **4**, 93-100 (2004).

[46] A. Ansari, Y. Schen, and S. V. Kuznetsov,*Misfolded loops decrease the effective rate of DNA hairpin formation*, Phys. Rev. Lett. **88**, 069801 (2002).

[47] R. Guha, *Using Quantum Dots for a FRET Based Immunoassay Technique*, Pennsylvania State University (2003), Internet site: *http://cheminfo.informatics.indiana.edu/ rguha/writing/pub/qdfret-rep.pdf.*

[48] J. R. Grunwell, J. L. Glass, T. D. Lacoste, A. A. Denize, D. S. Chemla and P. G. Schultz, *Monitoring the conformational Fluctuations of DNA Hairpin Using Single-Pair Fluorescence Resonance Energy Transfer* J. Am. Chem. Soc. **123**, 4295-4303 (2001).

[49] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers and D. Herschlag, *Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins*, Proc. Natl. Acad. Sci. USA **103**, 6190-6195 (2006).

[50] V. Muñoz, E. R. Henry, J. Hofrichter, and W. A. Eaton, *A statistical mechanical model for β-hairpin kinetics*, Proc. Natl. Acad. Sci. USA, **95**, 5872-5879 (1998).

[51] W. Gobush, H. Yamakawa, H. Stockmayer, and W.S. Magee, *Statistical Mechanics of Wormlike Chains I. Asymptotic Behavior*, J. Chem. Phys. **57**, 2839-2854 (1972).

[52] W. Gobush, H. Yamakawa, H. Stockmayer, and W.S. Magee, *Statistical Mechanics of Wormlike Chains II. Excluded Volume Effects*, J. Chem. Phys. **57**, 2843-2854 (1972).

[53] P.-G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press (1979)

[54] H. Yamakawa, *Modern Theory of Polymer Solutions.* Harper's Chemistry series (2001).

[55] M. Doi and S. F. Edwards, *The theory of Polymer Dynamics*, Oxford Science Publication (1986).

[56] K. Bury, *Statistical distributions in engineering*, Cambridge university press (1999).

[57] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Wiley-Interscience Publication (1972).

[58] M. E. Fisher, *Magnetism in One-Dimensional Systems. The Heisenberg Model for Infinite Spin*, Am. J. Phys. **32**, 343-346 (1964).

[59] J. Wilhelm and E. Frey, *Radial Distribution Function of Semiflexible Polymers*, Phys. Rev. Lett. **77**, 2581-2584 (1996).

[60] N. Theodorakopoulos, *End-to-end distribution of discrete Kratky-Porod chains*, submitted to Phys. Rev. E.

[61] A. Sali, E. Shakhnovich, and M. Karplus, *How does a protein fold?*, Nature **369**, 248-251 (1994).

[62] D. K. Klimov and D. Thirumalai, *Criterion that Determines the Foldability of Proteins*, Phys. Rev. Lett. **76**, 4070-473 (1996).

[63] M. Terazima, *Protein dynamic detected by the time-resolved transient grating technique*, Pure. Appl. Chem. **73**, 513-517 (2001).

[64] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, *Diffusive Dynamics of the Reaction coordinate for Protein Folding Funnels*, J. Chem. Phys. **104**, 5860-5868 (1996).

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

[66] A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, *Chain Length of Protein Folding Time*, Phys. Rev. Lett. **77**, 5433-5436 (1996).

[67] A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, *Evolution-like selection of fast-folding model proteins*, Proc. Natl. Acad. Sci. USA **92**, 1282-1286 (1995).

[68] R. Mélin, H. Li, N. S. Wingreen and C. Tang, *Designability, thermodynamic stability, and dynamics in protein folding: A lattice model study*, J. Chem. Phys. **110**, 1252-1262 (1999).

[69] V. S. Pande and D. S. Rokhsar, *Folding pathway of a lattice model for proteins*, Proc. Natcl. Acad.Sci. USA **96**, 1273-1278 (1999).

[70] N. Go and H. Taketomi, *Respective Roles of Short-and Long-Range Interactions in Protein Folding*, Proc. Natl. Acad. Sci. USA **75**, 559-563 (1978).

[71] T. S. van Erp, S. Cuesta-López, J.-G. Hagmann and M.Peyrard, *Can One Predict DNA Transcription Start Sites by Studying Bubbles*, Phys. Rev. Lett. **95**, 218104 (2005).

[72] D. P. Landau and K. Binder, *Monte Carlo Simulations in Statistical Physics* Canbridge Univ. Press (2000)

[73] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, *New approach to Monte Carlo calculation of the Free energy: Method of expanded ensembles*, J. Chem. Phys. **96**, 1776-1783 (1992).

[74] T. Dauxois, M. Peyrard, *Physics of solitons*, Cambridge University Press (2005).

[75] P. Hanggi, P. Talkner, and M. Borkovec, *Reaction-rate theory: fifty years after Kramers*, Rev. Mod. Phys. **62**, 251-341 (1990).

[76] K. Schulten, Z. Schulten, and A. Szabo, *Dynamics of reactions involving diffusive barrier crossing*, J. Chem. Phys. **74**, 4426-4432 (1981).

[77] A. Campa and A. Giansanti, *Experimental tests of the Peyrard-Bishop model applied to the melting of very short DNA chains*, Phys. Rev. E **58**, 3585-3588 (1998).