

# The Statistical Thermodynamics of Stretched Multi-stranded Biomolecules

Hemant Tailor

*Department of Physics and Astronomy  
University College London*

Thesis submitted to  
University College London

For the degree of  
**Doctor of Philosophy in**  
**Theoretical Physics**

2013

I, Hemant Tailor confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Hemant Tailor  
July 19, 2013

# Acknowledgements

First and foremost I offer my sincerest gratitude to my supervisor, Prof Ian Ford, who has supported and guided me throughout my thesis with his patience and knowledge. I attribute the level of my degree to his encouragement and continuous effort. One simply could not have wished for a better, or friendlier supervisor. Thanks also to my second supervisor Dr Laurent Bozec for his useful inputs during the course of my research.

To my friends; I would like to thank Dr Alex Rutherford, Dr Alastair Dunn, Dr Alex Grimwood, Dr Matthew Wolf, Matthew Sheirs, and Savvas Papasavva. Your company and friendship have made these years some of the most memorable, thank you for the good times!

To the people I love; a special thanks goes to Prakash and Hina Tailor, for there are fewer greater comforts in moments of adversity than having loving family at my side. Finally, I would like to thank my loving parents, Arvind and Nalini Tailor, for their unconditional support, both financially and emotionally throughout my degree. I dedicate this thesis to you.

## **Abstract**

Mechanical experiments on single polymers have provided force-displacement data for a number of species, including multi-stranded polymers such as DNA and collagen. The interpretation of this information at the mesoscale requires the use of statistical mechanics to calculate the free energies of various structures. In this thesis, we look initially at the behaviour of a single stranded polymer in extension and go on to consider two particular multi-stranded polymers, DNA and collagen, both of which have different but similar tertiary structures. In simplifying the multi-stranded structures we evaluate the partition function by a transfer matrix approach in one dimension. The numerical solution of the equations provide the restoring force as a function of the extension of a single strand, as well as a description of the molecular distortion that is induced. In DNA we investigate molecular breakage schemes for base-pair potentials that simulate the severing of bonds when a single strand is pulled, and the tearing of one strand from another in triple-stranded collagen.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Models</b>	<b>7</b>
2.1	Freely Jointed Chain (FJC) . . . . .	7
2.1.1	The Free Energy of Stretching a Freely Jointed Chain . . . . .	9
2.2	Worm-Like Chain (WLC) . . . . .	9
<b>3</b>	<b>Transfer Matrix Method</b>	<b>12</b>
3.1	One-Dimensional Ising Model . . . . .	13
3.2	Two-Dimensional Ising Model . . . . .	16
<b>4</b>	<b>Biopolymers DNA and Collagen</b>	<b>23</b>
4.1	Deoxyribonucleic Acid (DNA) . . . . .	23
4.1.1	DNA Structure . . . . .	24
4.1.2	DNA Mechanics . . . . .	26
4.1.3	DNA Models . . . . .	30
4.2	Collagen . . . . .	41
4.2.1	Collagen Structure . . . . .	41
4.2.2	Collagen Mechanics . . . . .	44
<b>5</b>	<b>The Extendable Freely Jointed Chain</b>	<b>50</b>
5.1	Probability Distribution of the Freely Jointed Chain . . . . .	50
5.2	Construction of the Partition Function Integral . . . . .	53
5.2.1	Quadratic Potential . . . . .	55
5.2.2	Double Potential . . . . .	56
5.2.3	Evaluating $Z_q(\mathbf{R})$ for the non-extendable Freely Jointed Chain	58
5.3	Extendable Freely Jointed Chain Model in One Dimension . . . . .	60
5.3.1	Potential for an Extendable Freely Jointed Chain . . . . .	60
5.3.2	Extendable Freely Jointed Chain . . . . .	62

5.3.3	Imposing the limit of $p$ by the small approximation of $\sin\left(\frac{\omega p}{2}\right)$	63
5.3.4	Using the Taylor expansion of $\sin\frac{\omega p}{2}$ to include higher order terms . . . . .	63
5.3.5	Expanding the Fourier Transform of the Potential as a Series . . . . .	65
5.4	Extendable Freely Jointed Chain in Three Dimensions . . . . .	71
5.4.1	Construction of the Partition Function . . . . .	71
5.4.2	Evaluating $Z_q$ for the non-extendable Freely Jointed Chain . . . . .	72
5.4.3	The Shell Potential and its Fourier Transform . . . . .	73
5.4.4	Evaluation of the Partition Function by the small $p$ approximation . . . . .	74
5.4.5	Evaluating the Partition Function for arbitrary $p$ . . . . .	76
5.5	Conclusion . . . . .	80
<b>6</b>	<b>DNA Model</b>	<b>81</b>
6.1	Construction of the Hamiltonian . . . . .	81
6.2	Partition Function Integral . . . . .	84
6.3	Transfer Matrices . . . . .	85
6.4	Boundary Conditions . . . . .	87
6.5	Eigenvalue Equations . . . . .	87
6.5.1	Analytical solutions for eigenvectors and eigenvalues of $T$ . . . . .	88
6.5.2	Analytical solutions for eigenvectors and eigenvalues of $\hat{T}$ . . . . .	90
6.6	Transfer Matrix Method . . . . .	92
6.6.1	Intact DNA . . . . .	93
6.6.2	Frayed DNA . . . . .	94
6.6.3	Double Frayed DNA . . . . .	96
6.6.4	Bubbling DNA . . . . .	97
6.7	Mean Axial Displacement . . . . .	99
6.7.1	Calculation of average $\eta_\alpha$ . . . . .	100
6.7.2	Calculation of average $\xi_\alpha$ . . . . .	100
6.8	Results for DNA eigensystem calculations . . . . .	102
6.8.1	Conclusion . . . . .	103
6.9	Results for DNA Model Calculations . . . . .	109
6.9.1	Conclusion . . . . .	116
6.10	Results for anharmonic DNA calculations . . . . .	117
<b>7</b>	<b>Collagen Model</b>	<b>122</b>
7.1	Collagen Model with a general residue potential . . . . .	122

7.1.1	Construction of the Hamiltonian . . . . .	122
7.1.2	Partition Function Integral . . . . .	124
7.1.3	Transfer Matrices . . . . .	125
7.1.4	Boundary Conditions . . . . .	127
7.1.5	Eigenvalue Equations . . . . .	127
7.1.6	Transfer Matrix Method . . . . .	129
7.1.7	Mean Axial Displacement . . . . .	130
7.1.8	Calculation of average $\rho_\alpha$ and $\lambda_\alpha$ . . . . .	131
7.1.9	Average residue positions and the calculation of average $R_\alpha$ .	133
7.2	Collagen Model with a Harmonic Potential . . . . .	136
7.2.1	Transfer Matrices . . . . .	137
7.2.2	Eigenvalue Equations . . . . .	138
7.2.3	Transfer Matrix Method . . . . .	139
7.2.4	Mean Axial Displacement . . . . .	142
7.2.5	Calculation of $\langle R \rangle$ and average residue positions for an applied extension . . . . .	146
7.3	Results for collagen eigensystem calculations . . . . .	149
7.4	Results for collagen model calculations . . . . .	155
7.5	Results for collagen calculations . . . . .	160

**8 Conclusion****163**

# Chapter 1

## Introduction

Modern experimental techniques have given us the tools to probe the microscopic scale of biomolecules. The most famous example was the use of X-ray diffraction to deduce the structure of DNA in 1953 [1, 2]. Since then many mechanical experiments performed on single biomolecules have provided insights into their static structure and function [3, 4, 5]. However, within the natural system biomolecules respond dynamically to their environment to operate the mechanisms that help support biological processes, and therefore to limit discussions to zero temperature structural states is an oversimplification in understanding of the biopolymer function. By studying the free energy we are able to move towards a view that encompasses biopolymers as a dynamical system with structural disorder [6, 7].

In thermodynamics the free energy is a property of a system that measures the amount of energy that can be converted into external work during an isothermal process or the system's ability to do isothermal work. The mechanism of protein folding from linear chains to unique structures that regulate cellular function can be successfully understood by the theoretical framework of the free energy landscape [8, 4]. By understanding the free energy behaviour we can begin to understand the entropic effects in protein stability [9, 10, 11, 12], protein-ligand binding [13, 14, 15] and protein folding [16, 17, 18]. Free energy behaviour also gives a better understanding of chemical processes that take place in drug partitioning across cell membranes. These processes are of paramount importance in the field of computer-aided drug design, and cannot be predicted reliably without knowledge of the associated free energy changes [19, 20]. Another example involves DNA. Supercoiled DNA contains a large amount of free energy that can be used to drive biological reactions [21]. The biological process of transcription and replication requires an input of work and hence a change of free energy to open and unwind the DNA double helix to expose

---

the chemical identity of the bases at the centre of the helix [22].

While molecular simulation provides an alternative approach to describe biomolecular mechanics from a bottom-up perspective this does not easily provide a thermal description of the behaviour. We can instead use semi-analytical statistical mechanics to provide a thermodynamic model to describe biopolymer mechanics. The canonical partition function

$$Z = \int e^{-\beta H} d\tau \quad (1.1)$$

where we integrate over the Boltzmann factors with  $H$  being the classical Hamiltonian, describes the statistical properties of a system in thermodynamic equilibrium which can be applied to biopolymers to calculate free energies for a given microstate. It is through this method we can investigate the structural properties at finite  $T$ , as well as potential phase transitions brought about during DNA shearing. In this thesis we simplify the structure of DNA allowing us explicitly to write a Hamiltonian describing the partition function of the system. We focus on characterising the free energy as a function of an axial displacement  $u$ , thus reproducing the effect of pulling one of the two polynucleotide strands in DNA.

Collagen is another multi-stranded biomolecule where in recent years much work has been focused on characterising the mechanical properties of collagen molecules. The structure is made up of three backbone strands twisted together into a right-handed coil known as a triple helix. Some of the earliest experiments by Sasaki and Odajima [23] estimated the Young's modulus of collagen molecules by X-ray diffraction techniques. Others include that of Bozec and Horton [24, 25] who used an AFM to investigate the topographical and mechanical properties of collagen molecules. However, the limitation of experimental analysis still doesn't allow the details at the nanoscale to be probed. We consider a simple structure for collagen, in which the additional strand adds an extra degree of freedom to the Hamiltonian in the partition function, increasing the complexity of the calculation. Despite this, we are still able to proceed with the calculation of the free energy, to study the axial pulling of one of the three strands in collagen.

We begin in chapter 2 by briefly introducing ourselves to models that describe simple linear polymers. Then in chapter 3 we briefly review the Transfer Matrix Method using the Ising Model before moving on to the structure of DNA and collagen in chapter 4. In this chapter we also consider some models that describe their conformations and force-extension behaviour. Since the early 1950's numerous models have been proposed to describe linear polymer conformation and statistics, the simplest being the Freely Jointed Chain (FJC). Because the method of evaluating a

partition function with a constrained mechanical extension is central to our analysis, we begin by describing our own version of the FJC using this method in chapter 5. We then provide an alternative model to the FJC, the Extendable Freely Jointed Chain (EFJC), which allows for extendable links in the chain. Chapter 6 introduces our analysis of DNA, the first of the two multi-stranded polymers investigated in this thesis. It covers the creation of a thermodynamic model for DNA extension, and describes the transfer integral method which we use to evaluate the partition function. The simplification of the DNA structure allows us to take a deeper look into the effects of base-pair breakages when a strand in the DNA molecule is pulled in an axial direction. Chapters 7 and 8 cover the analysis of collagen. We introduce a triangular prism geometric structure as a simplification of the collagen molecule and apply a similar method to calculate the partition function. We then determine the force-extension curves for a variety of inter-strand potentials. Conclusions are given in chapter 9.

# Chapter 2

## Models

### 2.1 Freely Jointed Chain (FJC)

The simplest mathematical model for a polymer is the freely jointed chain. In this hypothetical model the polymer is a set of points joined by  $n$  links of fixed length  $l$  in a linear succession. The angle between bond links in the  $xy$ -plane,  $\theta$ , and the  $zx$ -plane,  $\phi$ , can assume any value from 0 to  $2\pi$  with equal probability without any restriction. Thus, we can deduce that the directions for a given bond link are all equally probable, irrespective of the directions of its neighbouring bond links. These assumptions for the freely jointed chain model were first mathematically described by Kuhn, Guth and Marks by the analogy with an unrestricted random flight [26, 27]. Although this is covered extensively in most polymer physics textbooks we review some of the basic principles [28, 29, 30, 31].

We represent the conformation of a freely jointed chain by a set of  $N+1$  position vectors in space  $\mathbf{R}_i$  for  $i = 0, 1, \dots, N$ , or alternatively by the set of  $N$  bond vectors  $\mathbf{r}_i$  for  $i = 1, \dots, N$ , where  $\mathbf{r}_i = \mathbf{R}_i - \mathbf{R}_{i-1}$ . We can then characterise the size of a polymer by considering the end-to-end vector  $\mathbf{R}$  of the chain,

$$\mathbf{R} = \sum_{i=1}^N \mathbf{r}_i \quad (2.1)$$

Since the bond orientations are uncorrelated in the chain, the of symmetry of orientation implies  $\mathbf{R}$  will be zero [27]. Even if bond orientations were correlated the average would still be zero Eq. 2.2.

$$\langle \mathbf{R} \rangle = 0 \quad (2.2)$$

The mean square of the end-to-end distance which characterises the spatial configurations of the chain molecules can be expressed as,

$$\langle \mathbf{R}^2 \rangle = \sum_{i,j=1}^N \langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle \quad (2.3)$$

$$= \sum_{i=1}^N \langle |\mathbf{r}_i|^2 \rangle + 2 \sum_{i \neq j=1}^N \langle |\mathbf{r}_i|^2 \rangle \delta_{ij} \quad (2.4)$$

$$= Nl^2 \quad (2.5)$$

The root mean square length for the chain will then be

$$\mathbf{R}_{rms} \simeq N^{\frac{1}{2}} l \quad (2.6)$$

Eq. 2.6 shows that  $\mathbf{R}_{rms}$  is much less than the total length  $Nl$  measured along the path of the chain. This important conclusion implies that in the set of conformations that the freely jointed chain assumes in the process of thermal motion, stretched conformations only constitute a minor fraction. The majority of chain conformations are strongly coiled in space in thermodynamic equilibrium of an ideal freely jointed chain.

A freely jointed chain can also be treated as a simple random walk where  $P(\mathbf{R})$  obeys Gaussian statistics in the limit of  $N \rightarrow \infty$ . The general Gaussian probability distribution in 3 dimensions that is normalised to  $\int P(\mathbf{R}) d^3 R = 1$  is

$$P(\mathbf{R}) = (2\pi\beta)^{-\frac{3}{2}} \exp\left(-\frac{(\mathbf{R} - \alpha)^2}{2\beta}\right) \quad (2.7)$$

The condition imposed by Eq. 2.2 means that the Gaussian distribution is symmetrical about  $\mathbf{R} = 0$ , implying  $\alpha = 0$ . The other condition set by Eq. 2.5 means that the integral,  $\int |\mathbf{R}|^2 P(\mathbf{R}) d^3 R = Nl^2$  is satisfied when  $\beta = Nl^2/3$  in 3 dimensions. The probability distribution then becomes,

$$P(\mathbf{R}) = \left(\frac{3}{2\pi Nl^2}\right)^{\frac{3}{2}} \exp\left(-\frac{3\mathbf{R}^2}{2Nl^2}\right) \quad (2.8)$$

Here, Eq. 2.8 describes the probability distribution of end-to-end distance for the freely jointed chain. We can then obtain the entropy of the chain by using Eq. 2.8

$$S(\mathbf{R}) \propto k \ln(P(\mathbf{R})) \quad (2.9)$$

and

$$S = S_0 - k \frac{3\mathbf{R}^2}{2Nl^2} \quad (2.10)$$

The free energy of the chain with the end-to-end point distance  $\mathbf{R}$  only comes from the entropy.

$$F(\mathbf{R}) = U - TS = F_0 - kT \frac{3\mathbf{R}^2}{2Nl^2} \quad (2.11)$$

### 2.1.1 The Free Energy of Stretching a Freely Jointed Chain

We will now calculate the free energy cost of stretching the ends of the freely jointed chain in order to understand the entropic elasticity. The entropy has already been defined by Eq. 2.10, where  $P(\mathbf{R})$  is the number of obtainable micro-conformations. For a chain with an end-to-end vector  $\mathbf{R}$  the number  $P(\mathbf{R})$  is the probability of that micro-conformation being occupied. The entropy difference between a chain held with an end-to-end distance  $\mathbf{R}$  and one held with the end-to-end vector of zero is

$$\Delta S(\mathbf{R}) = k \log P(\mathbf{R}) - k \log P(0) = k \log \frac{P(\mathbf{R})}{P(0)} \quad (2.12)$$

Hence the change in free energy of the chain is

$$\Delta F = -T\Delta S = \frac{3kT\mathbf{R}^2}{2Nl^2} \quad (2.13)$$

This is an entropic spring with spring constant

$$k = \frac{3kT}{Nl^2} \quad (2.14)$$

## 2.2 Worm-Like Chain (WLC)

Real polymers have mechanical properties that allow them to flex and bend. Different polymers exhibit different flexible properties that are quantitatively described by the persistence length  $L_p$ , the length scale at which the specific polymer bends. Polyethylene is an example of polymer that is extremely flexible,  $L_p \approx 0.5\text{nm}$ , in comparison, DNA is a much stiffer molecule due to its double helix structure,  $L_p \approx 50\text{nm}$ .

The worm-like chain accounts for the semi-flexible polymer by assuming that the polymer is a uniform rod that is continuously flexible and each segment in the chain resists the bending force [32]. The WLC model was first described by Kratky and Porod when they did experiments which investigated X-ray scattering from

cellulose in a colloidal suspension [33]. The model has since been used to describe semi-flexible polymers.

The worm-like chain represents a polymer on a contour of fixed length  $L$ . We can define a position on the chain by a vector  $\vec{x}(s)$ . The tangent vector to the point  $s$  on the curve is then

$$\vec{t}(s) = \frac{\partial \vec{x}(s)}{\partial s} \quad (2.15)$$

and the end-to-end distance becomes

$$\vec{R} = \int_0^L \vec{t}(s) ds \quad (2.16)$$

the mean square end-to-end distance becomes

$$\langle \mathbf{R}^2 \rangle = \left\langle \vec{R} \cdot \vec{R} \right\rangle \quad (2.17)$$

$$= \left\langle \int_0^L \vec{t}(s) ds \cdot \int_0^L \vec{t}(s') ds' \right\rangle \quad (2.18)$$

$$= \int_0^L ds \int_0^L \langle \vec{t}(s) \cdot \vec{t}(s') \rangle ds' \quad (2.19)$$

It can be shown that the orientational correlation function for a worm-like chain follows an exponential decay [34]

$$\langle \vec{t}(s) \cdot \vec{t}(s') \rangle = \langle \cos \theta(s - s') \rangle = \exp \left( -\frac{s - s'}{L_p} \right) \quad (2.20)$$

Where the persistence length  $L_p$  is the length scale over which the orientation correlation is considered lost. Inserting Eq. 2.20 into Eq. 2.19 the root mean square end-to-end distance of a worm-like chain becomes

$$\langle \mathbf{R}^2 \rangle = 2L_p L - 2L_p^2 \left( 1 - \exp \left( -\frac{L}{L_p} \right) \right) \quad (2.21)$$

Eq. 2.21 is important in two limiting cases. For  $L \gg L_p$ , Eq. 2.21 simplifies to

$$\langle \mathbf{R}^2 \rangle = 2L_p L \quad (2.22)$$

such that the worm-like chain behaves as a random coil, and for  $L \ll L_p$

$$\langle \mathbf{R}^2 \rangle = L^2 \quad (2.23)$$

which shows the WLC behaving as a rigid rod. Models like the WLC fit biopolymers

quite well for weak pulling forces where the elasticity of DNA is due to entropic effects [35]. When the force exceeds 5 pN stretching of the double helix along its axis takes place and experimental results begin to deviate from the WLC.

# Chapter 3

## Transfer Matrix Method

The Transfer Matrix Method (TMM) is a mathematical method for calculating exact solutions to certain statistical mechanical models. The method works by expressing the partition function in terms of an object called the transfer matrix and the calculation of the exact solution is reduced to computing its eigenvalues and eigenvectors. The transfer matrix encapsulates the thermodynamic properties of the model through the eigenspectrum of the matrix [36].

In polymer physics we can use the transfer matrix formalism by treating a polymer as a one-dimensional array of units, much like a one dimensional lattice [37, 38]. Adjacent units make a contribution to the statistical weight of the conformation in a manner that is characterised by the transfer matrix. Since the polymer segments are connected, the state of a given segment depends on the states of the other segments. The partition function of the total system is then given by successive multiplication and summation of all transfer matrices corresponding to the monomer units [39, 40].

The transfer matrix method also has many points in common with a discrete formulation of quantum statistical mechanics, in particular the Feynman formulation in terms of a path integral. Here the transfer matrix method links classical systems described by statistical mechanics in  $d$  dimensions with quantum systems in  $(d - 1)$  dimensions [41].

The transfer matrix method was first introduced into statistical mechanics by Kramers and Wannier [42] to find an approximate solution to the statistical mechanics of a two-dimensional ferromagnet, but it was Onsager [43] in 1944 who demonstrated its power by calculating the exact partition function of the zero-field version of the Ising model. The exact solution is covered extensively in the literature [44, 29, 36, 45, 46] but for the purpose of understanding the transfer matrix method in its simplest case we begin by introducing the exact solution of the Ising model in

one dimension.

### 3.1 One-Dimensional Ising Model

The Ising model in one dimension is a classical model for ferromagnetism described by a system of  $N$  spins situated regularly on a straight line Fig. 3.1. We denote the value of spin  $i$  by  $\sigma_i$  which can only take discrete values -1 and 1. In the system we have an external magnetic field interaction of strength  $B$ , a coupling constant between spin pairs with strength  $J$ , and the atomic magnetic moment  $\mu$ . The Hamiltonian for a one-dimensional Ising model is [46, 36]

$$H = -\mu B \sum_{i=1}^N \sigma_i - J \sum_{i=1}^N \sigma_i \sigma_{i+1} \quad (3.1)$$

where we impose a periodic boundary condition such that

$$\sigma_{N+1} = \sigma_1 \quad (3.2)$$

This changes the geometry of the straight line lattice into a circle where  $\sigma_N$  and  $\sigma_1$  become adjacent neighbours Fig. 3.2. In the Hamiltonian this allows the last term in the sum of interactions to become  $\sigma_N \sigma_1$ . Inserting the Hamiltonian into the

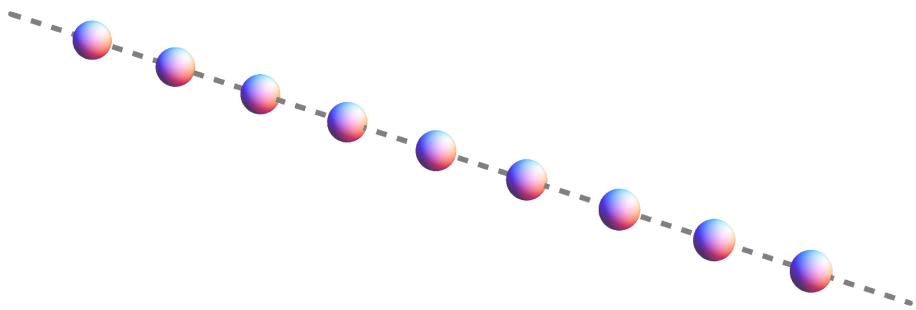


Figure 3.1: A lattice in one dimension is a set of regularly spaced units.

partition function we get

$$Z = \sum \exp(-\beta H) \quad (3.3)$$

$$= \sum_{\{\sigma_i=\pm 1\}} \exp \left( \beta \mu B \sum_{i=1}^N \sigma_i + \beta J \sum_{i=1}^N \sigma_i \sigma_{i+1} \right) \quad (3.4)$$

$$= \sum_{\{\sigma_i=\pm 1\}} \prod_{i=1}^N \exp(\beta \mu B \sigma_i) \exp(\beta J \sigma_i \sigma_{i+1}) \quad (3.5)$$

The product of contributions from each spin in the exponential allows us to group adjacent spins together. If we do this Eq. 3.5 becomes [36]

$$\begin{aligned} Z = & \sum_{\{\sigma_i=\pm 1\}} \exp \left( \beta J \sigma_1 \sigma_2 + \frac{\beta B (\sigma_1 + \sigma_2)}{2} \right) \exp \left( \beta J \sigma_2 \sigma_3 + \frac{\beta B (\sigma_2 + \sigma_3)}{2} \right) \dots \quad (3.6) \\ & \dots \exp \left( \beta J \sigma_{N-1} \sigma_N + \frac{\beta B (\sigma_{N-1} + \sigma_N)}{2} \right) \exp \left( \beta J \sigma_N \sigma_1 + \frac{\beta B (\sigma_N + \sigma_1)}{2} \right) \end{aligned}$$

Next we introduce  $T(\sigma_i, \sigma_{i+1})$ , which are the elements of a transfer matrix  $T$ . The rows are labelled by the values of  $\sigma_i$  and columns by the values  $\sigma_{i+1}$ .

$$T(\sigma_i, \sigma_{i+1}) = \exp \left( \beta J \sigma_i \sigma_{i+1} + \frac{\beta B (\sigma_i + \sigma_{i+1})}{2} \right) \quad (3.7)$$

If we write out the transfer matrix explicitly we get a  $2 \times 2$  matrix

$$\begin{array}{cc} \sigma_i = 1 & \sigma_{i+1} = -1 \\ \sigma_i = 1 & \left( \begin{array}{cc} \exp(\beta J + B) & \exp(-\beta J) \\ \exp(-\beta J) & \exp(\beta J - B) \end{array} \right) \\ \sigma_{i+1} = -1 & \end{array} \quad (3.8)$$

Substituting Eq. 3.7 into Eq. 3.6, the transfer matrix allows us to express the partition function as a product of matrices [36]

$$Z = \sum_{\{\sigma_i=\pm 1\}} T(\sigma_1, \sigma_2) T(\sigma_2, \sigma_3) \dots T(\sigma_{N-1}, \sigma_N) T(\sigma_N, \sigma_1) = \text{tr}(T^N) \quad (3.9)$$

This equation reduces the calculation of the partition function to the calculation of the trace of an unknown matrix  $T^N$ . We know from linear algebra that the trace is invariant under cyclic permutations, and therefore using this property in conjunction with an appropriate unitary matrix  $U$  and its inverse  $U^{-1}$  we can convert the transfer

matrix into a diagonal form

$$U^{-1}TU = \Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \quad (3.10)$$

Where  $\lambda_1$  and  $\lambda_2$  are the eigenvalues of the transfer matrix  $T$ . Inserting this into the partition function we get

$$Z = \text{tr}(T^N) = \text{tr}(\Lambda^N) = \lambda_1^N + \lambda_2^N \quad (3.11)$$

The eigenvalues of the transfer matrix can be calculated using the eigenvalue equation without explicitly knowing the form of the corresponding eigenvectors

$$\det |T - \lambda_i I| = \begin{vmatrix} \exp(\beta J + \beta B) - \lambda_1 & \exp(-\beta J) \\ \exp(-\beta J) & \exp(\beta J - \beta B) - \lambda_2 \end{vmatrix} = 0 \quad (3.12)$$

where the eigenvalues are [29, 36]

$$\lambda_{1,2} = \exp(\beta J) \cosh \beta B \pm \sqrt{\exp(2\beta J) \sinh^2 \beta B + \exp(-2\beta J)} \quad (3.13)$$

The analysis using the transfer matrix gives surprising results in the thermodynamic limit  $N \rightarrow \infty$  where the boundary effects become negligible only the largest eigenvalue is relevant. In the thermodynamic limit the free energy per spin is given by

$$F = -k_B T \lim_{N \rightarrow \infty} \frac{1}{N} \ln Z_N \quad (3.14)$$

$$= -k_B T \lim_{N \rightarrow \infty} \frac{1}{N} \ln \{ \lambda_1^N + \lambda_2^N \} \quad (3.15)$$

For the more general case where the transfer matrix is larger than a  $2 \times 2$  matrix the set of eigenvalues go beyond  $\lambda_1$  and  $\lambda_2$ , and Eq. 3.15 becomes

$$F = -k_B T \lim_{x \rightarrow \infty} \frac{1}{N} \ln \left\{ \lambda_1^N \left( 1 + \sum_i \frac{\lambda_i^N}{\lambda_1^N} \right) \right\} \quad (3.16)$$

With  $\lambda_1$  being the largest eigenvalue. Since the ratio  $\lambda_i/\lambda_1$  is always less than unity, taking  $N \rightarrow \infty$  leads to  $(\lambda_i/\lambda_1)^N \rightarrow 0$  and hence

$$F = -k_B T \ln \lambda_1 \quad (3.17)$$

This elegantly shows that thermodynamic quantities in the thermodynamic limit  $N \rightarrow \infty$  depend only on the largest eigenvalue of the transfer matrix, which is often easier to calculate than the entire spectrum of the matrix. Since this is also true for other interacting many-particle systems treated with the transfer matrix method the problem of finding the exact solution within the transfer-matrix approach is reduced to finding the largest eigenvalue.

## 3.2 Two-Dimensional Ising Model

For a two dimensional model we consider a spin- $\frac{1}{2}$  Ising model on a square lattice in a vanishing external field. The lattice is defined by  $R$  rows and  $C$  columns and the Hamiltonian can be written as [47]

$$H = -J \left( \sum_{i=1}^R \sum_{j=1}^C \sigma_{i,j} \sigma_{i+1,j} + \sum_{i=1}^R \sum_{j=1}^C \sigma_{i,j+1} \sigma_{i,j} \right) \quad (3.18)$$

The summations contain two spin coupling interactions between nearest-neighbours: one is for adjacent rows of the square lattice, while the other takes into account adjacent columns. Again, we impose boundary conditions such that the square lattice is continuous

$$\sigma_{R+1,j} = \sigma_{1,j} \quad \text{where } j = 1, 2, 3, \dots, C \quad (3.19)$$

$$\sigma_{i,C+1} = \sigma_{i,1} \quad \text{where } i = 1, 2, 3, \dots, R \quad (3.20)$$

The periodicity of the square lattice makes it topologically equivalent to a 2-torus where the  $C^{th}$  column is coupled to the first column, and the  $R^{th}$  row is coupled to the first row. We then set the overall spin configuration of the  $j^{th}$  column of spins by a new variable  $\xi_j$  which has  $2^R$  possible spin configurations.

$$\xi_j = (\sigma_{1,j}, \sigma_{2,j}, \sigma_{3,j}, \sigma_{4,j}, \dots, \sigma_{R,j}) \quad (3.21)$$

The Hamiltonian can then be written as a sum of the interaction energy of individual

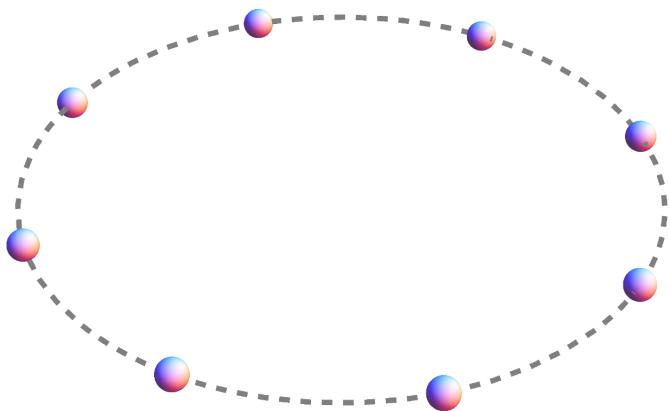


Figure 3.2: A lattice in one dimension with periodic boundary conditions forms a circle.

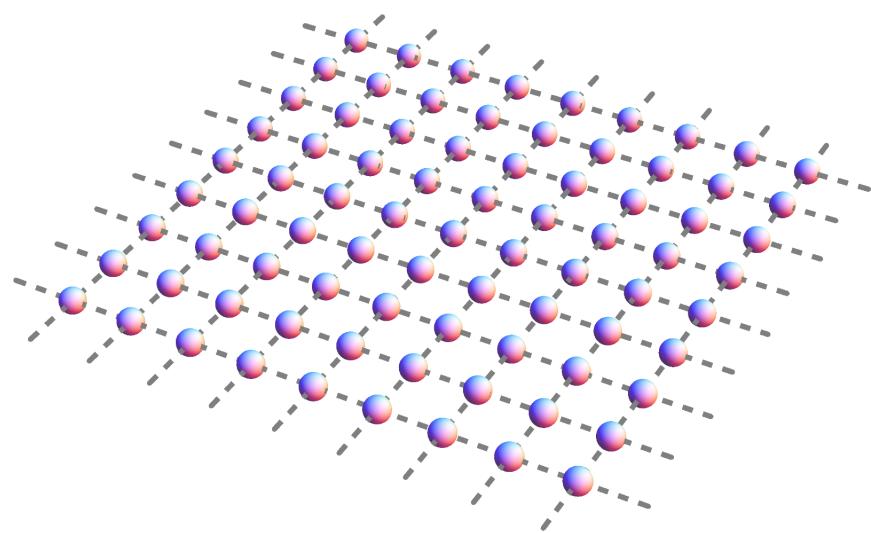


Figure 3.3: A lattice in two dimensions.

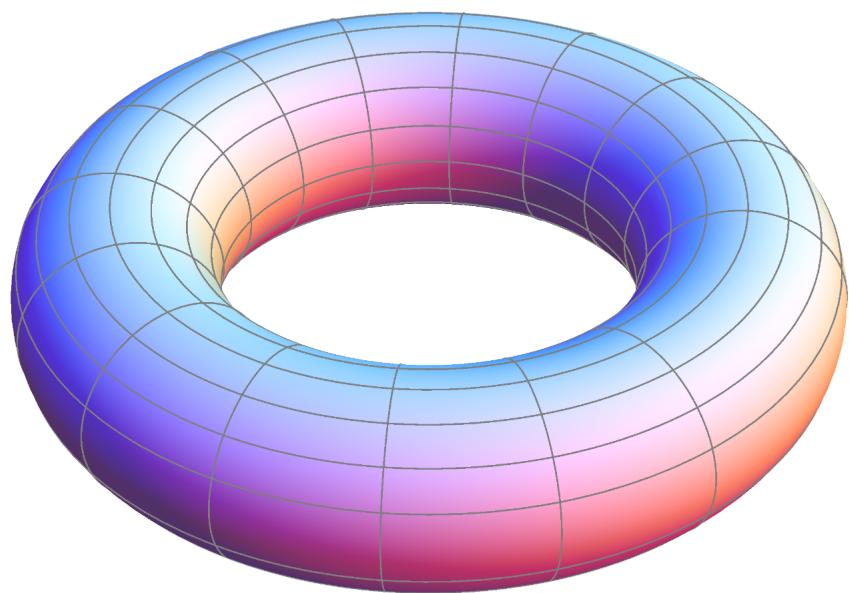


Figure 3.4: The boundary conditions in the vertical and horizontal directions wraps the square lattice into a 2-torus.

columns,  $H_1$ , and the interaction energy between adjacent columns,  $H_2$ .

$$H_1(\xi_j) = -J \sum_{i=1}^R \sigma_{i,j} \sigma_{i+1,j} \quad (3.22)$$

$$H_2(\xi_j, \xi_{j+1}) = -J \sum_{i=1}^R \sigma_{i,j} \sigma_{i,j+1} \quad (3.23)$$

The total Hamiltonian is simply

$$H(\xi) = \sum_{j=1}^C H_1(\xi_j) + H_2(\xi_j, \xi_{j+1}) \quad (3.24)$$

In the symmetric form the Hamiltonian is

$$H(\xi) = \sum_{j=1}^C \frac{1}{2} H_1(\xi_j) + \frac{1}{2} H_1(\xi_{j+1}) + H_2(\xi_j, \xi_{j+1}) \quad (3.25)$$

Inserting this into the partition function we get the relation

$$\begin{aligned} Z &= \sum_{\xi_1} \sum_{\xi_2} \sum_{\xi_3} \dots \sum_{\xi_C} \exp \left( \sum_{j=1}^C -\frac{\beta}{2} (H_1(\xi_j) + H_1(\xi_{j+1})) - \beta H_2(\xi_j, \xi_{j+1}) \right) \\ &= \sum_{\xi_1} \sum_{\xi_2} \sum_{\xi_3} \dots \sum_{\xi_C} \prod_{j=1}^C \exp \left( -\frac{\beta}{2} (H_1(\xi_j) + H_1(\xi_{j+1})) - \beta H_2(\xi_j, \xi_{j+1}) \right) \end{aligned} \quad (3.26)$$

which evidently demonstrates the transfer matrix for the 2D Ising model to be

$$\begin{aligned} T(\xi_j, \xi_{j+1}) &= \exp \left( -\frac{\beta}{2} (H_1(\xi_j) + H_1(\xi_{j+1})) - \beta H_2(\xi_j, \xi_{j+1}) \right) \\ &= \exp \left( \frac{\beta J}{2} \left( \sum_{i=1}^R \sigma_{i,j} \sigma_{i+1,j} + \sum_{i=1}^R \sigma_{i,j+1} \sigma_{i+1,j+1} \right) + \beta J \sum_{i=1}^R \sigma_{i,j} \sigma_{i,j+1} \right) \end{aligned} \quad (3.27)$$

Using these transfer matrices in the partition function we get

$$Z = \sum_{\xi_1} \sum_{\xi_2} \dots \sum_{\xi_C} T(\xi_1, \xi_2) T(\xi_2, \xi_3) \dots T(\xi_{C-1}, \xi_C) T(\xi_C, \xi_1) \quad (3.28)$$

At this point we introduce a Kronecker delta defined by the boundary conditions  $\delta_{\xi_{C+1}, \xi_1}$  and then apply the completeness relation to express the delta function as

set of eigenfunctions,  $\psi_\mu$ , of the transfer matrix  $T(\xi_i, \xi_{i+1})$ .

$$\delta_{\xi_{C+1}, \xi_1} = \sum_{\mu} \psi_{\mu}^*(\xi_1) \psi_{\mu}(\xi_{C+1}) \quad (3.29)$$

The eigenvalue equation is

$$\sum_{\xi_i} T(\xi_i, \xi_{i+1}) \psi_{\mu}(\xi_i) = \lambda_{\mu} \psi_{\mu}(\xi_{i+1}) \quad (3.30)$$

The partition function with the complete set of eigenfunctions becomes

$$Z = \sum_{\mu} \sum_{\xi_1} \sum_{\xi_2} \dots \sum_{\xi_C} \sum_{\xi_{C+1}} \psi_{\mu}^*(\xi_1) T(\xi_1, \xi_2) T(\xi_2, \xi_3) \dots T(\xi_{C-1}, \xi_C) T(\xi_C, \xi_{C+1}) \psi_{\mu}(\xi_{C+1}) \quad (3.31)$$

Now we use equation Eq. 3.30 to contract each transfer matrix in the sum with an appropriate eigenfunction, leaving behind an eigenvalue which is a scalar quantity. We get

$$Z = \text{tr}(T^C) = \sum_{j=1}^{2^R} \lambda_j^C \quad (3.32)$$

In the thermodynamic limit the free energy per spring for a 2D Ising model becomes

$$F = -k_B T \lim_{R \rightarrow \infty} \lim_{C \rightarrow \infty} \left\{ \frac{1}{RC} \ln (\lambda_1^C + \lambda_2^C + \dots + \lambda_{2^R}^C) \right\} \quad (3.33)$$

$$= -k_B T \lim_{R \rightarrow \infty} \frac{1}{R} \ln \lambda_1 - k_B T \lim_{R \rightarrow \infty} \lim_{C \rightarrow \infty} \frac{1}{RC} \ln \left[ 1 + \sum_i^{2^R} \left( \frac{\lambda_g}{\lambda_1} \right)^C \right] \quad (3.34)$$

$$= -k_B T \lim_{R \rightarrow \infty} \frac{1}{R} \ln \lambda_1 \quad (3.35)$$

The calculation of the largest eigenvalue of the  $T(\xi_i, \xi_{i+1})$  matrix goes beyond the scope of this thesis but here again we have clearly demonstrated the use of the transfer matrix method in finding the exact solution to the 2D Ising model. Expanding on the two dimensional model we can consider the lattice subject to just one boundary condition, Eq. 3.20, and find an expression for the partition function when the topology of the lattice is folded into a cylinder. We continue from Eq. 3.28, but

omit the final transfer matrix  $T(\xi_C, \xi_1)$ . The partition function becomes

$$Z = \sum_{\xi_1} \sum_{\xi_2} \dots \sum_{\xi_C} T(\xi_1, \xi_2) T(\xi_2, \xi_3) T(\xi_3, \xi_4) \dots T(\xi_{C-1}, \xi_C) \quad (3.36)$$

$$\times \exp\left(-\frac{\beta H(\xi_1)}{2}\right) \exp\left(-\frac{\beta H(\xi_C)}{2}\right) \quad (3.37)$$

Here we introduce a Kronecker delta function with a dummy variable  $\xi_\alpha$  such that

$$\delta_{\xi_1, \xi_\alpha} = \sum_{\mu} \psi_{\mu}^*(\xi_1) \psi_{\mu}(\xi_\alpha) \quad (3.38)$$

Inserting this into Eq. 3.36

$$\begin{aligned} Z(\xi_\alpha) &= \sum_{\mu} \sum_{\xi_1} \sum_{\xi_2} \dots \sum_{\xi_C} \psi_{\mu}^*(\xi_1) T(\xi_1, \xi_2) T(\xi_2, \xi_3) T(\xi_3, \xi_4) \dots T(\xi_{C-1}, \xi_C) \psi_{\mu}(\xi_\alpha) \\ &\quad \times \exp\left(-\frac{\beta H(\xi_1)}{2}\right) \exp\left(-\frac{\beta H(\xi_C)}{2}\right) \end{aligned} \quad (3.39)$$

We then proceed to contract each transfer matrix in the sum from  $\xi_1$  to get an expression for the partition function

$$Z(\xi_\alpha) = \sum_{\mu} \sum_{\xi_C} \lambda_{\mu}^{C-1} \psi_{\mu}^*(\xi_C) \psi_{\mu}(\xi_\alpha) \exp\left(-\frac{\beta H(\xi_\alpha)}{2}\right) \exp\left(-\frac{\beta H(\xi_C)}{2}\right) \quad (3.40)$$

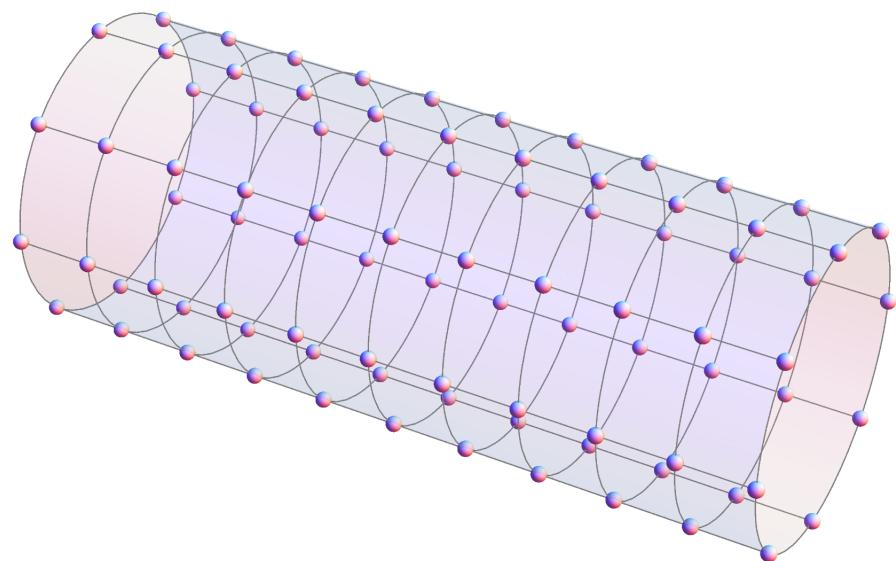


Figure 3.5: A single boundary condition in the vertical direction wraps the square lattice into a cylinder.

# Chapter 4

## Biopolymers DNA and Collagen

Macromolecules formed by biochemical reactions are called biopolymers, and like ordinary polymers, they are made up of repeating chemical units called monomers. Homologous biopolymers such as proteins are made up of one type of monomer unit whereas heterologous biopolymers are made up of more than one class of monomeric unit. In contrast to many synthetic polymers the fundamental characteristic of biopolymers is their hierarchical structure. At the atomic level the primary structure of biopolymers specifies the sequence of its monomeric subunits. The build up of monomeric units then forms a secondary structure which describes the macroconformation the biopolymer takes in three dimensional space. Another feature that characterises polymers is the architecture which can also be linear, branched or even cross-linked.

### 4.1 Deoxyribonucleic Acid (DNA)

DNA is a biomolecule within cells of living organisms that contain genetic information essential for biological functions. In the early 1950's X-ray diffraction patterns first gave an insight into the double helical structure of DNA [1, 2]. Further research in the later years would provide detailed views of DNA as well as refining ideas about the DNA structure and the interactions with other proteins [48]. It still continues to be an area of active research for both physical and life sciences.

Today advances in technology enable us to observe and manipulate these molecules in a controlled environment using force spectroscopy techniques; the most common methods include using optical tweezers [49, 50], magnetic tweezers and atomic force microscopy (AFM) [51, 52, 53]. The well publicised structure of the molecule is only a static picture and it is now understood that the dynamics of biological

molecules are also essential for their function. Molecular dynamics simulation has shown that chemical reactions that seem to be impossible according to the static molecular structure might indeed take place due to temporary large molecular distortions [54, 55]. An example of when this might occur in DNA is when it needs to replicate itself, or when DNA goes through the transcription process with the help of enzymes through biochemical reactions [51, 56]. Here DNA can either be bent, twisted, stretched, compressed, sheared or even locally destroyed requiring a significant amount of energy. Experiments have measured some of these bulk properties where the twist modulus of DNA was found to be 440 pN nm<sup>2</sup> [57], and the stretch modulus was measured to be 1000 pN [52, 58].

### 4.1.1 DNA Structure

Made out deoxyribonucleotide monomers, DNA is a biopolymer that belongs to a group of macromolecules called nucleic acids. The composition of each nucleotide consists of three components: the sugar deoxyribose, the heterocyclic (5-carbonic) base that binds to the 1' carbon of the sugar, and the phosphate group that binds to the 5' carbon of the sugar. The polymerisation of these nucleotides form the two backbone strands of DNA where the phosphate group bonds with the 3' carbon of another nucleotide by a 3',5' phosphate-diester covalent bond. Together they form a regular polynucleotide chain with an alternating sugar and phosphate group that is characterised by its polarity from its 5'-end to its 3'-end [8, 59].

In DNA the two strands are of equal length and are aligned anti-parallel to each other coiling round an axis to form a right-handed helix. In this arrangement the negatively charged sugar-phosphate backbones are positioned on the exterior of the double helix structure with the nucleotide bases on the inside being exposed through the major and minor grooves of the double helix structure Fig. 4.1b. The first type of bases which contain two heterocyclic rings are classified as purines, called adenine (A) and guanine (G) (Fig. 4.2a); the second type contain only a single heterocyclic ring and are classified as pyrimidines, called thymine (T) and cytosine (C) (Fig. 4.2b). These bases participate in hydrogen bonding where two bases are connected by hydrogen bonds to form complimentary AT, and GC base-pairs. Both base-pairs have the same dimensions giving them the same planar geometry allowing the base-pairs to stack on top of each other without disturbing the sugar-phosphate backbones. The AT base-pairs are linked by two hydrogen bonds whereas the GC base-pairs are linked by three hydrogen bonds. Although individual hydrogen bonds are relatively weak when compared to covalent bonds, the

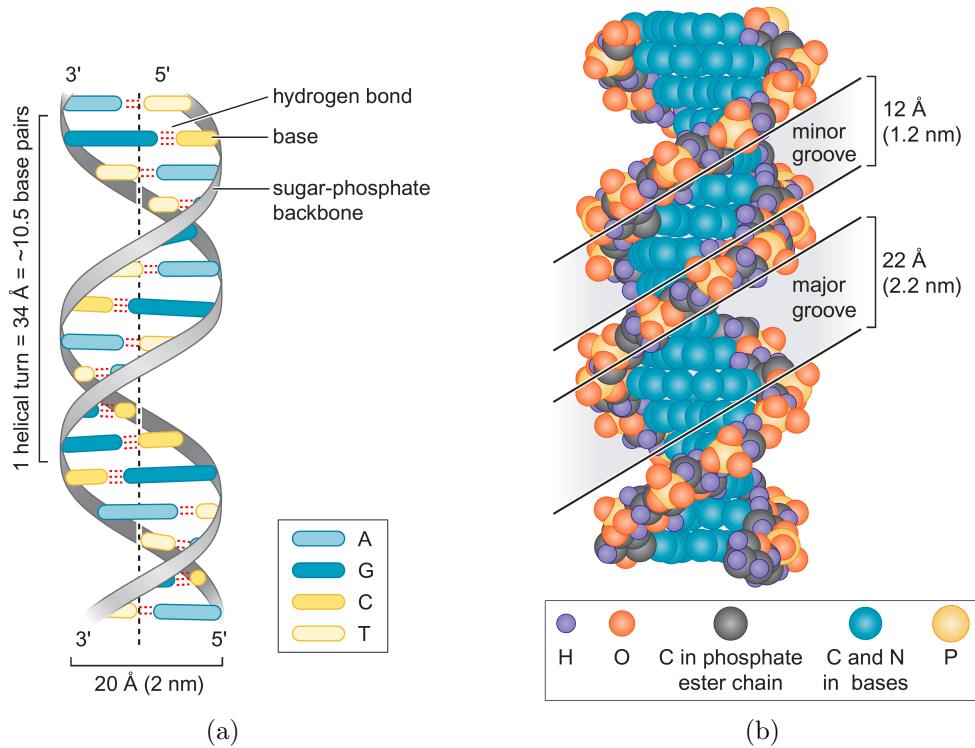


Figure 4.1: On the left is a schematic model of the double helix illustrating the base-pairs, and the sugar-phosphate backbone. The space-filled illustration on the right highlights the overall twist of the DNA structure with the bases located on the inside of the cylindrical structure. These illustrations are from figure 6-1 in reference [59].

bonding between the polynucleotides lies perpendicular to the fibre axis making the strength of the hydrogen bonds additive [8, 59]. Although these hydrogen bonds may contribute to some of the stability of the double helix structure the main contribution comes from the stacking interaction between the bases. A combination of dipole-dipole interactions,  $\pi-\pi$  stacking interaction, van der Waals forces and hydrophobic forces, produce an overall complex attractive interaction between the heterocyclic rings of the base-pairs. In its natural conditions, DNA is a right-handed double-helix which contains about 10.5 base-pairs per helical turn. It has two neighbouring base-pairs that are separated axially by 0.34 nm and the pitch length, the length of a helical turn of the double-helix along the central axis, is about 3.4 nm Fig. 4.1a [1].

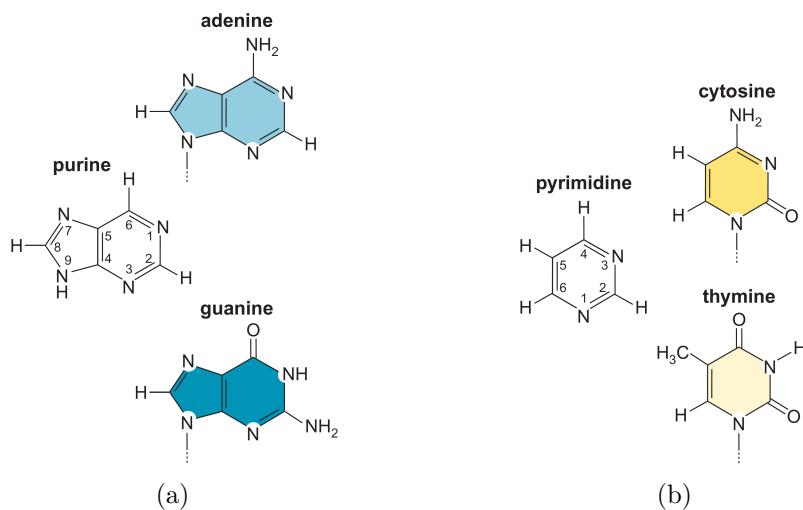


Figure 4.2: The types of bases that exist in DNA. On the left are the two purine bases that contain two heterocyclic rings, and on the right are the two pyrimidine bases that contain a single heterocyclic ring. These illustrations are from figure 6-1 in reference [59].

### 4.1.2 DNA Mechanics

Double stranded DNA behaves like any other polymer in aqueous solution, exchanging energy with its surrounding environment at thermal equilibrium. The thermal agitation causes the DNA to bend and curve locally by small forces that continuously change the macroscopic configuration of the molecule. Commonly described as Brownian motion this has the effect of reducing the end-to-end distance of the molecule, maximising its conformational entropy such that its most probable configuration is a random coil. Purely entropic in origin, the free energies of deformations of the molecule are of order  $kT \sim 4$  pN nm which sets the lower limit of the forces needed to stretch a DNA molecule to its contour length [60, 53].

Some of the first experiments measuring the entropic elasticity of DNA were conducted by Smith et al (1992). Using a method where they chemically attached one end of the DNA to a glass surface and tethered the other end to a magnetic bead they were able to measure the force extension of DNA using a combination of magnetic fields and hydrodynamic drag [61]. The results shown from their experiments in Fig. 4.5 describe DNA having a linear relationship in the limit of low forces with a Hooke's constant that is inversely proportional to the length of the molecule. Here we see the FJC model agreeing well with such force extension curves for DNA. As the applied force increases the experimental results clearly dismiss the FJC model as a suitable candidate for describing DNA.

A better description is provided by the Worm Like Chain (WLC) model which

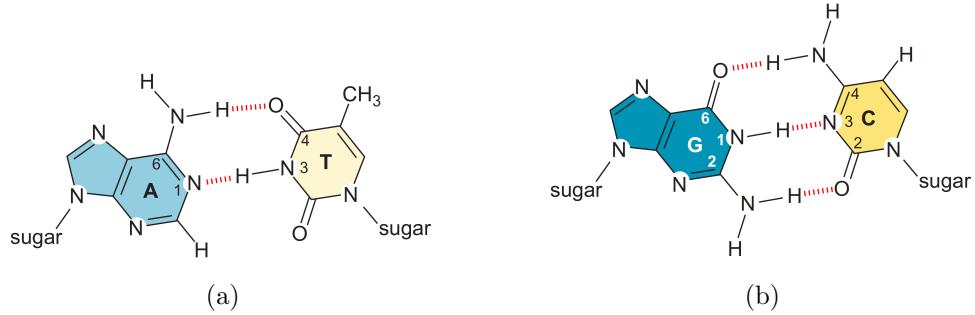


Figure 4.3: Chemical structure of the complementary base-pairs in DNA. On the left is the adenine-thymine(AT) base-pair with two hydrogen bonds and on the right is the guanine-cytosine base-pair with three hydrogen bonds. These illustrations are from figure 6-3 figure 6-6 in reference [59].

considers DNA to be a continuous flexible worm-like chain [62]. In this model the persistence length  $L_p$  is a quantitative measure of chain stiffness which defines a length scale over which segments of the counter length are correlated. Since the FJC model is only flexible between fixed segments and ignores the bending fluctuations of the segments themselves, the WLC provides a more realistic description of entropy change in stiffer polymers. The first complete treatment of the WLC model was achieved by Marko and Siggia using two different methods; for the exact solution they numerically evaluated a line integral of two terms in the WLC model where the first term described the resistance of the chain bending, and the second term described the stretching energy resulting from an applied force. Following the terminology used in [62] we describe the DNA conformations by a space curve  $\mathbf{r}(s)$  of a contour length,  $L$ , where  $s$  is the arc length, and the energy in the WLC model is expressed as:

$$E_{wlc} = k_B T \int_0^L ds \left( \frac{A}{2} \left| \frac{d\mathbf{r}(s)}{ds} \right|^2 - F \cos(\theta(s)) \right) \quad (4.1)$$

where  $\theta(s)$  is the angle between  $\mathbf{r}(s)$  and the axis along the direction of force  $F$ . In the second method they expanded the first term to the quadratic order that led to the Gaussian approximation of the line integral. They were able to derive an interpolation formula that was close to the exact numerical solution of the force-extensions [62, 63]:

$$F(x) = \left( \frac{k_B T}{L_p} \right) \left[ \frac{1}{4(1 - \frac{x}{L})^2} - \frac{1}{4} + \frac{x}{L} \right] \quad (4.2)$$

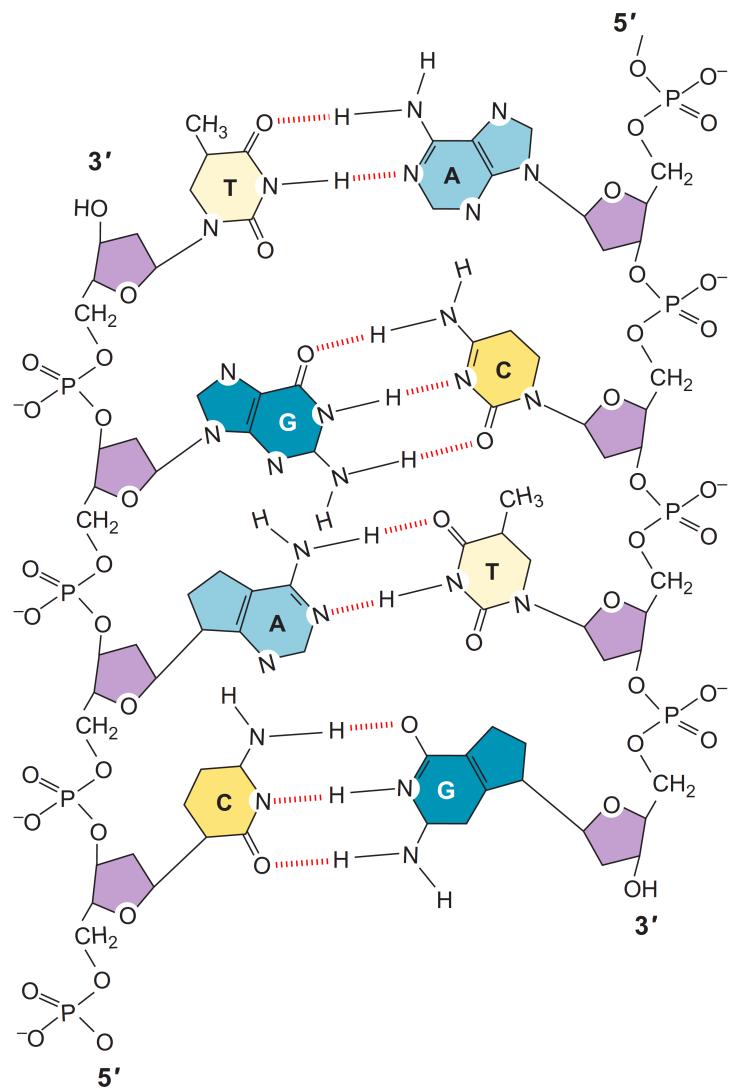


Figure 4.4: Another schematic diagram of the DNA molecule showing the different molecular groups within the backbone and base-pairs. This illustration is from figure 6-3 in reference [59].

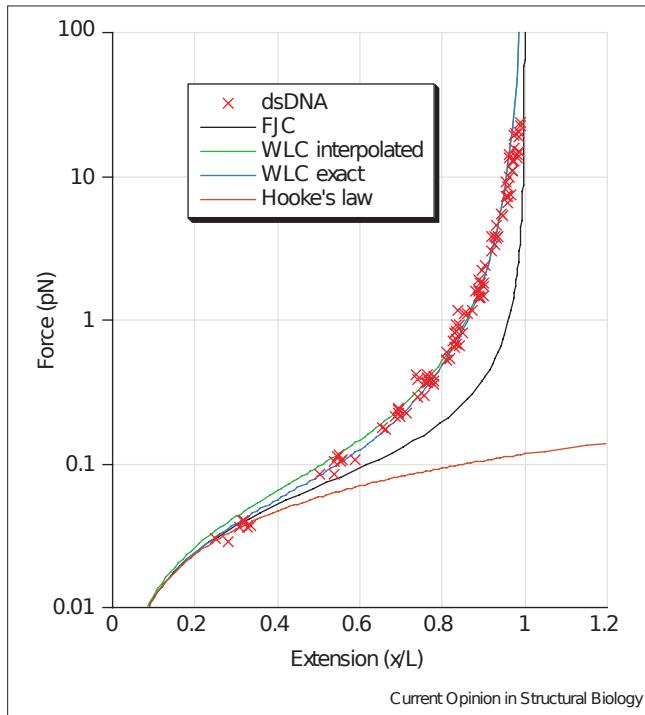


Figure 4.5: Force extension behaviour of double stranded DNA. The experimental data taken from Smith et al. (1992) is fitted to the WLC model using the modified WLC interpolated formula Eq. 4.2, the exact WLC model solved numerically, and the linear spring Eq. 4.3 are also shown. In these plots  $L_p = 53$  nm where the FJC assumes twice the persistence length. This original plot was reported in [58].

where  $L$  is the length of the molecule,  $L_p$  is the persistence length and  $x$  is the extension. For small values of  $\frac{x}{L}$ , the approximation simplifies to

$$F(x) = \frac{3k_B T}{2L_p} \left( \frac{x}{L} \right) \quad (4.3)$$

which demonstrates that the molecule behaves as a linear spring with a stiffness constant  $k_{DNA} = \frac{3k_B T}{2LL_p}$ . Further corrections, and improvements have been made to Eq. 4.2 to include effects of stretching DNA in physiological buffer [49, 64]. For the results shown in Fig. 4.6 the stretch modulus was found to be approximately 1000 pN in a buffer solution [58]. The failure to account for the enthalpic correction at small forces (5-10 pN) leads to underestimation of  $L_p$  resulting in a deviation from the model above 10 pN [64, 49]. The experimental results shown in Fig. 4.5 were taken from Smith et al. (1992) where the persistence length of DNA in the WLC model was  $L_p \sim 53$  nm in order to fit the experimental data. Other experiments including the more recent experiment by Van Maneren et al. (2009) have reproduced these force extension curves for DNA which have shown that the predictions of the

WLC model provide an excellent agreement with the data measured [65, 49, 66, 50]. This can also be observed in Fig. 4.5. At forces above 65 pN DNA structurally

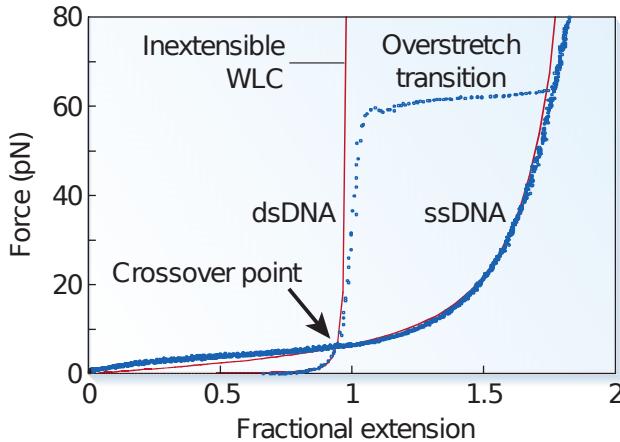


Figure 4.6: Force extension behaviour of DNA and ssDNA [67].

changes under mechanical stress causing the DNA molecule to extend 1.7 times the original length. We see this extra length being released over a narrow range of force shown by the plateau in Fig. 4.6. Experiments have highlighted that the force at which the over-stretching transition occurs is dependent on the details of terminal attachments,  $\sim 65$  pN if the molecule is torsionally relaxed at both ends, and  $\sim 110$  pN if it is not. Two competing models [50] have been proposed to describe the structural behaviour in this overstretched region: In the first model it is thought that DNA gradually unwinds under tension, with the base-pairs intact, into the S-form DNA resembling a parallel ladder. The second model interprets the overstretched phase as a denatured DNA resulting from strand separation caused by breaking of hydrogen bonds between complementary bases. Recent experiments show that there is a region where both these models can coexist with base-pair melting causing strand separation from DNA to single stranded DNA (ssDNA) [50, 68, 69].

### 4.1.3 DNA Models

In an attempt to study the DNA transcription process Peyrard et al. [70] found that developing a mathematical model that included the roles played by the RNA polymerase enzymes would be far too complex a task. Similarities were noted with DNA thermal denaturation, a mechanism where the backbone DNA strands separate by heating, also known as DNA melting. This function initiates locally by the formation of bubbles within the DNA structure and therefore was considered to be a preliminary step in understanding the transcription process [70, 71, 56, 35].

Simplifying the DNA structure to a 1D ladder system Peyrard et al. were able to write a Hamiltonian as a function of base pair separation that included temperature independent parameters, and then studied the statistical mechanics of the system by employing the transfer integral method. The model has two degrees of freedom where in each strand the displacements of the  $n^{th}$  base from their equilibrium position is labelled  $u_n$  and  $v_n$  along the direction of the hydrogen bond that connects base pair. The structure is considered homogeneous and periodic with each nucleotide base being represented by a point mass  $m$  [70].

The effective base pair potential  $V$  was modelled by a Morse potential to represent the 2 or 3 hydrogen bonds that connect the two bases in a pair, and along each backbone strand the neighbouring bases are connected by a stacking harmonic potential that has a stiffness constant  $\kappa$ . For system with  $N$  base pairs, the Hamiltonian is expressed as [70],

$$H = \sum_n \frac{m}{2} (\dot{u}_n^2 + \dot{v}_n^2) + \frac{\kappa}{2} (u_n - u_{n-1})^2 + \frac{\kappa}{2} (v_n - v_{n-1})^2 + V(u_n - v_n) \quad (4.4)$$

with

$$V(u_n - v_n) = D [\exp(-a(u_n - v_n)) - 1]^2 \quad (4.5)$$

where  $D$  is the dissociation energy of the base pair, and  $a$  is the parameter that sets the spatial range of the base pair potential. The first term in Eq. 4.4 represents the kinetic energy of the system, the second term corresponds to the effect of stretching along the backbone, and the third term represents the interaction of the base pair.

Further simplifications are made to the Hamiltonian by transforming the coordinate system into variables that represent the in-phase  $x_n$ , and out-of-phase  $y_n$  motions of the base positions. The transformations are [70]

$$x_n = \frac{u_n + v_n}{\sqrt{2}} \quad (4.6)$$

$$y_n = \frac{u_n - v_n}{\sqrt{2}} \quad (4.7)$$

which readily separates the Hamiltonian to give

$$H = H(x_n) + H(y_n) \quad (4.8)$$

with

$$H(x) = \sum_n \frac{p_n^2}{2m} + \frac{\kappa}{2} (x_n - x_{n-1})^2 \quad (4.9)$$

$$H(y) = \sum_n \frac{q_n^2}{2m} + \frac{\kappa}{2} (y_n - y_{n-1})^2 + V'(y_n) \quad (4.10)$$

where  $p_n = m\dot{x}_n$ ,  $q_n = m\dot{y}_n$  and  $V'(y_n) = V(\sqrt{2}y_n)$ . In the partition function integral we find that the momentum terms and the terms in  $x_n$  integrate analytically to provide a constant factor in the total partition function. For the remaining terms in  $y_n$  the partition function integral can be written as [70]

$$Z(\beta) = \int \prod_{n=1}^N dy_n \exp(-\beta H(y_n, y_{n-1})) \delta(y_N - y_1) \quad (4.11)$$

with the delta function fulfilling a periodic boundary condition. This partition function integral can also be expressed using a product of Boltzmann factors to give

$$Z(\beta) = \int \prod_{n=1}^N dy_n \left[ \prod_{i=1}^N \exp(-\beta f(y_n, y_{n-1})) \right] \delta(y_N - y_1) \quad (4.12)$$

where the term  $f(y_n, y_{n-1})$  can be generalised to a sum of the backbone interactions  $W(y_n, y_{n-1})$  and the base pair interactions  $V'(y_n)$ , that is,

$$f(y_n, y_{n-1}) = W(y_n, y_{n-1}) + V'(y_n) = \frac{\kappa}{2} (y_n - y_{n-1})^2 + V'(y_n) \quad (4.13)$$

In this one-dimensional problem the integral can be evaluated exactly in the thermodynamic limit of a large system using the eigenvalues and eigenfunctions of a transfer matrix. Introducing the the transfer integral operator  $T$  we have an eigenvalue equation [70]

$$\int dy' T(y, y') \psi_i(y') = \lambda_i \psi_i(y) \quad (4.14)$$

with

$$T(y, y') = \frac{\kappa}{2} (y - y')^2 + V'(y) \quad (4.15)$$

where  $\psi_i$  are the orthonormal eigenfunctions and  $\lambda_i$  are the corresponding eigenvalues. From Eq. 4.12 the delta functions are expanded as a set of orthonormal eigenfunctions  $\delta(y_N - y_1) = \sum_i \psi_i^*(y_N) \psi_i(y_1)$ , such that the partition function

now becomes

$$\begin{aligned} Z(\beta) = & \sum_i \int dy_1 \exp(-\beta f(y_1, y_2)) \psi_i(y_1) \int dy_2 \exp(-\beta f(y_2, y_3)) \\ & \dots \int dy_{N-1} \exp(-\beta f(y_{N-1}, y_N)) \psi_i^*(y_N) \end{aligned} \quad (4.16)$$

Using the eigenvalue equation we can contract the integrals to give a reduced form of the partition function equation that is a sum over all eigenvalues [70]

$$Z(\beta) = \sum_i \lambda_i^{N-1} \quad (4.17)$$

In the limit of a large DNA structure the largest eigenvalue  $\lambda_1$  dominates from the contractions. The free energy as a function of temperature can then be written as

$$f(\beta) = -\frac{(N-1)}{\beta} \log \lambda_i \quad (4.18)$$

In order to study the mean stretching of the base-pairs,  $\langle y \rangle$  can be expressed as

$$\langle y \rangle = \int dy \psi_1^2(y) y \quad (4.19)$$

In the original model [70] Peyrard et al. found that the results from the calculation of  $\langle y \rangle$  showed the mean stretching diverged at a temperature  $T_c$ , thus demonstrating a one-dimensional phase transition corresponding to DNA melting. However, it failed at producing the sharpness of the melting transition measured in experiments. In order to fit experimental data, the stacking interaction was later modified through an anharmonic potential to describe the coupling between the neighbouring base pairs [56]. This improved stacking potential can be described by

$$W(y_n, y_{n-1}) = \frac{\kappa}{2} (1 + \rho \exp(-\alpha(y_n - y_{n-1}))) (y_n - y_{n-1})^2 \quad (4.20)$$

where  $\rho$  and  $\alpha$  are constants. This model is referred to as the Peyrard-Bishop-Dauxois (PBD) Model and compared very well with experiments for short heterogeneous DNA chains in solution [72]. The results are shown in Fig. 4.7 where we have a comparison of theoretical calculations with experimental melting curves demonstrating a phase transition occurring at finite temperature.

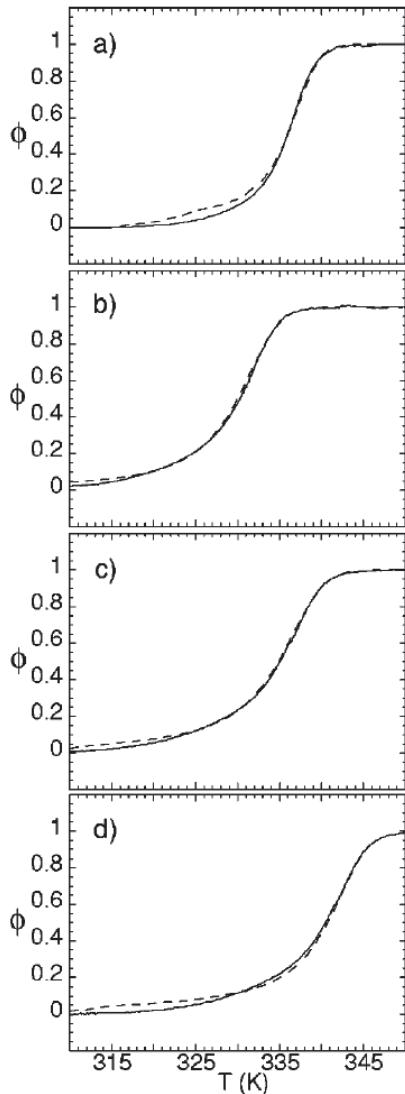


Figure 4.7: Experimental melting profiles (full lines) and theoretical results (dashed lines) for three different types of short heterogeneous DNA chains. The temperature is plotted against the average fraction of melted base pairs  $\phi$ . Subplots (a), (b) and (c) are the melting profiles for the three different DNA chains in a low concentration phosphate buffer solution. Subplot (d) uses the same DNA chain as in (c) but at a much higher concentrations [72].

Other studies which have investigated the thermodynamics of a heterogeneous PBD model have predicted a multi-step melting behaviour for some long DNA sequences agreeing with experimental observations [73]. This has been investigated more recently in [74].

The calculations of the eigenvectors and eigenvalues of the PB model go beyond the scope of this review, however. We have outlined the basic principles Peyrard et al. used to apply the transfer integral method. This method will be central to our

axial shearing models for both DNA and collagen.

The FJC and WLC models describe the behaviour at a scale where complex polymers are coarse-grained to resemble a singular flexible strand. For DNA these models do not take into account the mesoscopic structure of the two backbones, and complementary base-pairs. As we are interested in DNA shearing and its effects on the base-pairs we will focus our attention on models that include some form of base-pair description. One such model proposed by de Gennes (2001) took a relatively simply approach in studying this behaviour which later became a foundation to other models that we will briefly review [75, 76, 77, 78]. Collectively, these models showed excellent agreement with experiments when they were compared with results produced by Hatch et al (2008) [79].

To model the force dependence of axial shear pulling de Gennes simplified the structure of DNA to a ladder ignoring the double helical twist, and approximated the backbone, and base-pair potentials to be harmonic with different stiffness constants  $Q$  and  $R$ . By ignoring effects of finite temperature he described the Hamiltonian as a set of one-dimensional displacements for each strand distorted by a force  $F$  with the extremities being pulled at either the 3'3' or 5'5' terminals of DNA [75, 79], namely

$$H = \sum_{-l/2}^{\infty} \frac{Q}{2} (u_{n+1} - u_n)^2 + \sum_{-\infty}^{l/2} \frac{Q}{2} (v_{n+1} - v_n)^2 + \sum_{-l/2}^{l/2} \frac{R}{2} (u_n - v_n)^2 \quad (4.21)$$

where  $u_n$  and  $v_n$  are the axial displacements for each strand with indices  $n$  in the range of  $-l/2 \leq n \leq l/2$ . By defining a critical force for the breakage of a single base-pair to be  $R|v_n - u_n| > f_c$  the equilibrium conditions derived from the Hamiltonian allow the overall rupture force  $F_c$  to be expressed as [75]:

$$F_c = 2f_c \kappa^{-1} (1 - 2 \exp^{-\kappa l}) \quad (4.22)$$

where  $\kappa^2 = 2R/Q$ , and  $\kappa^{-1}$  is defined as the number of base-pairs for which the tension is mainly on a single strand; a dimensionless quantity also known as the adjustment length. In this model most of the shear distortion is concentrated near the ends of the ladder structure resulting in larger base-pair axial displacements near  $n = \pm l/2$ . As we approach the centre of the structure the differential force across the base-pairs tend to zero [75]. Experiments were able to test this model to a high degree of accuracy using magnetic tweezers to probe the shearing force of DNA [79]. There is an excellent agreement between experiment and the results predicted by

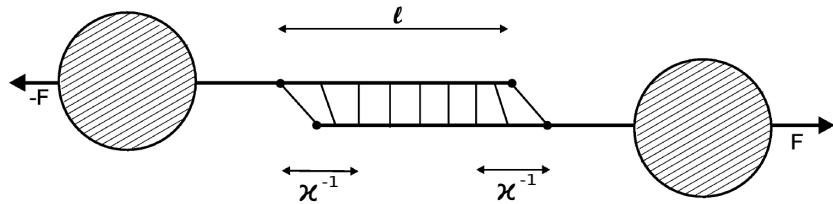


Figure 4.8: A schematic diagram of the ladder model used by de Gennes. The equal and opposite forces distort the base-pairs over an adjustment length  $\kappa^{-1}$  at both ends of the ladder structure of length  $l$ . This figure was edited from [75].

Eq. 4.21 if we include the effects of a finite number of frayed ends in DNA. Results shown in Fig. 4.9 show that the best fit of de Gennes theory to experiment actually occurs when we assume 7 frayed base-pairs at each end of the ladder structure with  $f_c = 3.9 \text{ pN}$ ,  $\kappa^{-1} = 6.8 \text{ bp}$  and stiffness constant ratio of  $Q/R = 92.5$  [79]. Matched

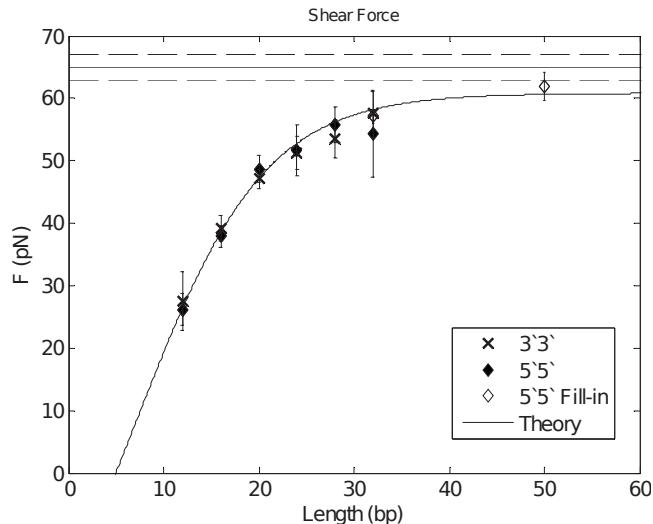


Figure 4.9: A match between the experimental results of pulling either 3'3' and 5'5' ends produced by Hatch et al with the best fit of Eq. 4.22. The horizontal line at  $65 \pm 3 \text{ pN}$  represents the forces characteristic of the over-stretching transitions as measured for the molecule. The upper and lower dashed lines correspond to 110% and 90% of the overstretched transition. This figure was taken from in [79].

by theory, the shear force as a function of number of base-pairs first increases linearly for short DNA lengths, and for larger lengths the critical force saturates below the asymptotic limit of entropic elasticity before the overstretched transition near 65 pN. Both experiment and theory imply that the mechanical shear stress is localised at the ends of the molecule where the shear force is applied, and that the pulling techniques from 3'3' and 5'5' ends make no difference on the shearing force [79].

A generalisation of de Gennes model was later studied by Chakrbarti et al (2009)

using a semi-microscopic vector model. Based on the same structure as shown in Fig. 4.8 the model adds further detail to the Hamiltonian by including a stacking interaction term that is expressed as the interactions between nearest neighbours on the same strand, and the interactions between the nearest, and next nearest-neighbour nucleotides of the opposite strand. Keeping the backbone strands harmonic, and using a Lennard-Jones potential for the interaction along the complementary base-pairs the non-linear Hamiltonian can be expressed as [76]:

$$H = \frac{Q}{2} \sum_{n=1}^L (|\vec{r}_{n+1} - \vec{r}_n| - a)^2 + (|\vec{s}_{n+1} - \vec{s}_n| - a)^2 + \sum_{n=1}^L V_{LJ}(|\vec{r}_n - \vec{s}_n|) + V'_{LJ}(|\vec{r}_{n+1} - \vec{s}_n|) + V'_{LJ}(|\vec{s}_{n+1} - \vec{r}_n|) \quad (4.23)$$

where  $\vec{r}_n$  and  $\vec{s}_n$  are the position vectors of the  $n$ th nucleotide along the two backbone strands,  $a$  is the equilibrium spacing between the bases along the backbone, and  $L$  is the size of the overlap region shown in Fig. 4.8. The two additional potential terms labelled  $V'_{LJ}$  represent the inter-strand stacking interactions that exist between the base-pairs of DNA. Though this vector model calculation takes a more complex approach to de Gennes, when coarse-grained, by solving and eliminating all displacements except those that couple directly to the force, Eq. 4.23 reduces to a general non-linear scalar model in terms of displacements projected along the chain. With the base-pair interaction still modelled by a Lennard-Jones potential the Hamiltonian becomes [76]

$$H = \frac{Q}{2} \sum_{n=1}^L (u_{n+1} - u_n)^2 + \frac{Q}{2} \sum_{n=1}^L (v_{n+1} - v_n)^2 + \sum_{n=1}^L V_{LJ}(|u_n - v_n|) - Fu_1 + Fv_L \quad (4.24)$$

where  $u_n$  and  $v_n$  are now the displacements of the top and bottom parts of the ladder structure. Expanding the base-pair potential around the minimum and retaining terms up to the quadratic order we recover a harmonic approximation first studied by de Gennes where the force exceeds a critical value  $F_c \sim f_c N$  for short chains, and  $F_c \sim 2f_c \kappa^{-1}$  for long chains [76].

Extending this work further, the statistical mechanics of DNA shearing was later studied at a finite temperature using a Hamiltonian that was similar to the non-linear scalar displacement model of Eq. 4.24. Calculating the shear force  $F_c$  of DNA as a function of the number of base-pairs, this model developed by Prakash et al.

(2011) found that an important factor in getting accurate results was dependent on the choice of values used for the adjustment length and single base-pair breakage force [77]. Keeping with the non-linear displacement model where the backbone strands are pulled along the molecular axis by an applied force  $F$ , the displacements of  $i$ th nucleotide from its equilibrium position are labelled by  $u_i$ , and  $v_i$  leading to a Hamiltonian similar to Eq. 4.23. Transforming the co-ordinate system the displacements of the Hamiltonian can be expressed as two independent components  $H_x$ , and  $H_y$  in the constant force ensemble [77].

$$H = H_x + H_y \quad (4.25)$$

where

$$H_x = \sum_{i=-N/2}^{N/2-1} \frac{Q}{2} (x_{i+1} - x_i)^2 - \frac{F}{\sqrt{2}} (x_{N/2} - x_{-N/2}) \quad (4.26)$$

and

$$H_y = \sum_{i=-N/2}^{N/2-1} \frac{Q}{2} (y_{i+1} - y_i)^2 + V(y_i) - \frac{F}{\sqrt{2}} (y_{N/2} - y_{-N/2}) \quad (4.27)$$

The base-pair interactions in this model are represented by a potential that encompasses a hardcore repulsion to prevent the compression of base-pairs with respect to its equilibrium, as well as long range interactions. The potential is expressed as

$$V(y_i) = -\frac{\epsilon}{\left(1 + \frac{2y_i^2}{\sigma^2}\right)^3} \quad (4.28)$$

where  $\epsilon$  is the depth of the potential, and  $\sigma$  is the diameter of DNA. In calculating the average positions,  $H_x$  simplifies  $\langle x_n \rangle$  to an integral of a Gaussian that can be solved analytically to give

$$\langle x_n \rangle = \frac{F}{\sqrt{2}\kappa} n \quad (4.29)$$

which agrees with the result found by de Gennes [77, 75]. The potential term  $V(y_i)$  gives the partition function integral of  $H_y$  a non-Gaussian form and therefore needs to be calculated numerically. Calculating the average displacement quantities Prakash et al. was able to study how the DNA shearing was distributed along the molecule as well as the energies associated with backbone and base-pair stretching. In this constant force model, the shear force needed to separate two strands of DNA was calculated by defining a critical distance  $\bar{y}$  needed to rupture a single base-pair. By differentiating the potential Eq. 4.28 with respect to  $y$ , the critical force for a

single base-pair was fitted to be close to the values used by Hatch et al. to fit their experimental data of  $f_c = 4.1$  pN if  $y = 2.38$  Å. The expression for the critical force as derived by Prakash et al. was found to be [77]:

$$F_c = \sqrt{2}\kappa (\langle y_{N/2} \rangle - \langle y_{(N/2-1)} \rangle) + 2f_c \quad (4.30)$$

When the base-pairs at the ends of that ladder structure  $\langle y_{N/2} \rangle$  reach an axial displacement of  $\bar{y}$ , the point of shearing, the values of  $\langle y \rangle$  can be determined. Prakash's results for molecules with 10-60 base-pairs are shown as the dotted curve plotted in Fig. 4.10 and compare well with the experimental results of Hatch et al.

Alternatively, for a constant extension method the average force is calculated from the amount of work needed to keep the extension of the ladder a distance  $y$ . From the Hamiltonian the work done by the base-pairs was calculated using the partition function with the constrained displacements imposed using two delta functions for each end,  $\delta(y_{-N/2} - y) \delta(y_{N/2} - y)$ , Prakash et al. described the work done as a function of  $y$  by the relations [77]:

$$W(y) = \frac{1}{\beta} [\ln Z_{N+1}(y) - \ln Z_{N+1}] \quad (4.31)$$

where

$$Z_{N+1}(y) = \int \prod_{i=-N/2}^{N/2} dy_i \delta(y_{-N/2} - y) \delta(y_{N/2} - y) \exp(-\beta H_y) \quad (4.32)$$

and

$$Z_{N+1}(y) = \int \prod_{i=-N/2}^{N/2} dy_i \exp(-\beta H_y) \quad (4.33)$$

The critical force,  $F_c$ , is now simply the first derivative of Eq. 4.31 with respect to  $y$  taking a value that represents the separation length needed to break a single base-pair  $y_c$ . The dashed curve shown in Fig. 4.10 is plot using the constant extension ensemble with  $y$  having a value of 2.0 Å.

Even though both these methods produce slightly different results for molecules of small lengths which later converge for molecules of larger  $N$ , both these methods agree well with results found by Hatch [77].

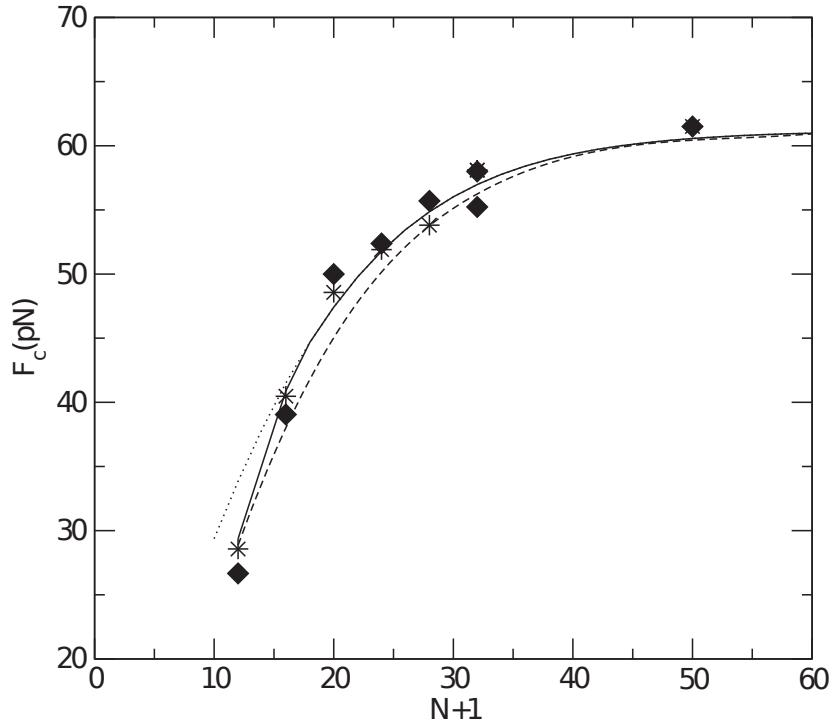


Figure 4.10: Results from the constant force method are shown by the dotted curve, and the constant force method is shown by the dashed curve of the critical force  $F_c$  as a function of the base-pairs. The experimental results of Hatch et al are shown by diamonds for 5'5' pulling, and stars for 3'3' pulling. This figure was taken from [77].

Compared with experimental data [77, 79] these models agree well with axial shearing of DNA Fig. 4.10. For long molecular lengths the models reach an important asymptotic limit giving reasonable fitting parameters to the shearing force of a single base-pair, and stiffness constant ratio between the backbone and base-pair. These models do not take into account breakage of base-pairs.

## 4.2 Collagen

Collagen is a naturally occurring protein that is a central structural component of multicellular organisms most abundantly found in animals. Forming the major component of the extracellular matrix and connective tissue collagen exist in several distinct types discussed extensively in [80, 81, 8, 82]. The collagen subtypes can be found in tissues such as skin, tendon, bone, and cartilage with 80 to 90 percent of the collagen in the human body consisting of types I, II, and III. These collagen molecules pack together to form elongated fibrils of similar structure.

Reinforcing biological tissues, collagen molecules also pack together to create fibrous polymers that are the major building blocks of all types of load-bearing tissues making the mechanical properties of collagen extremely important [83]. In bone and dentin, collagen is combined with mineral to yield very stiff tissues that transmit the force from muscles to bones enabling mammals to physiologically move. In tendon or the cornea, collagen is combined with other organic molecules, such as proteoglycans for different biological functions [81, 24, 84, 85].

Due to the limitations in performing mechanical testing on the nanometre and micrometre scale only very recent studies have been initiated to measure the mechanical properties of substructures like collagen fibrils, and tropocollagen [24, 84, 86].

### 4.2.1 Collagen Structure

Similar to DNA as well as other proteins, the collagen molecule has a hierarchical structure arising from the interactions of amino acid molecular groups at different levels. Three parallel polypeptide chains, also known as  $\alpha$ -chains, are made up of repeating amino acid triplets that intertwine in an overall right-handed coil to form the secondary structure of the collagen molecule which is approximately 300 nm long and 1.5 nm wide. This structure is known as a triple helix [80, 81, 8, 82, 87, 85, 24, 84, 86].

The amino acids are bound together by peptide bonds which are strong covalent bonds between the carboxyl group and the amino group of each amino acid. Each of the three polypeptides in collagen contains 1000 amino acid residues forming a primary structure that is coiled into a left-handed non- $\alpha$  extended helix. The helix is caused by a steric repulsion of the five-membered heterocycle Pro rings in the hydroxyproline and the proline residues [80, 89].

The tight molecular packing of the three polypeptide chains results in a glycine residue being close to the central core of the triple helix [80, 89, 8]. The amino acids in the Xaa and Yaa positions of collagen are often proline and hydroxyproline

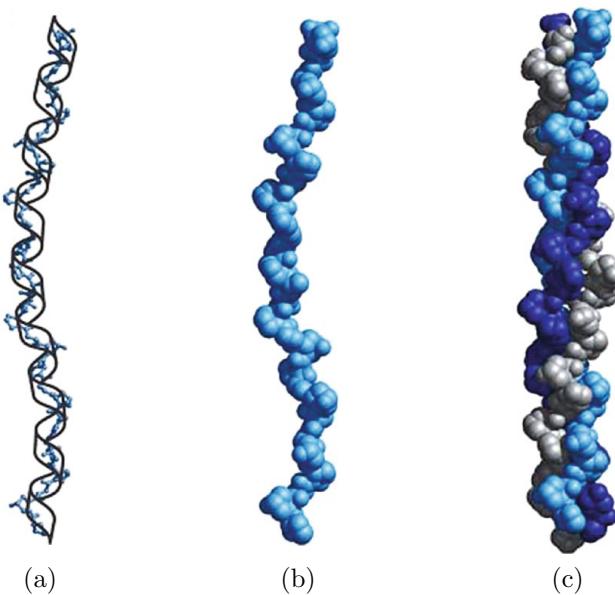


Figure 4.11: Formation of the collagen triple helix from three polypeptide chains. Each polypeptide has a left-handed helical structure as shown in Fig. 4.11a. A space-filling diagram of the same polypeptide is shown in Fig. 4.11b. Fig. 4.11c illustrates three polypeptide chains wrapping around one another with a right-handed twist. (Figure 4.12a, b and c taken from [88])

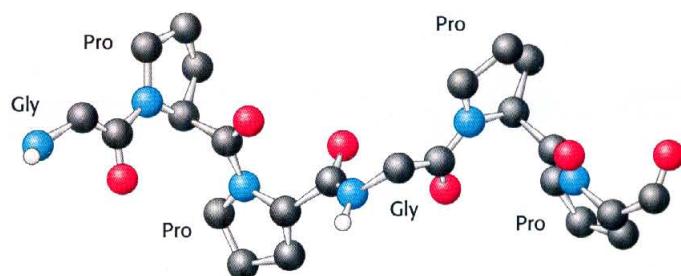


Figure 4.12: A section of a polypeptide showing two amino acid triplets of Gly-Pro-Hyp covalently bonded by a peptide bond. It also shows the steric repulsion of the heterocyclic rings. (From Fig. 2-39 in [8])

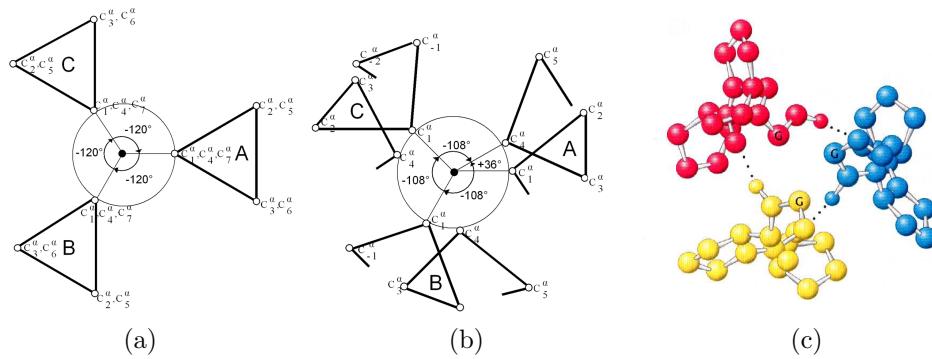


Figure 4.13: A top-down view of the collagen triple helix. The far left Fig. 4.13a was the original structure proposed by Ramachandra [87] (Figure 1a from [89]). This was later refined by Ramachandra, proposing a coiled structure [90] in Fig. 4.13b (Figure 1b from [89]). Fig. 4.13c is a molecular top-down illustration of coiled coil triple helix.

residues respectively. The compact space near the centre of the triple helix is unable to accommodate any of the larger groups of any amino acid and therefore to maintain stability the large Pro rings are positioned on the outermost part of the triple helix.

Ramachandra and Kartha first proposed the first prototype of the right-handed triple helix structure for the collagen molecule in 1954 using X-ray diffraction pictures from different sources [87]. Their initial structure comprised of three staggered left-handed polypeptides related by a three fold screw symmetry about a common axis, with approximately three residues per turn (Fig. 4.13a) [87, 89]. All peptide bonds were positioned in a trans position and had two hydrogen bonds within each triplet stabilising the triple helix [87]. This model was later updated to fit observations from X-ray diffraction of stretched collagen fibres which proposed a coiled coil triple helical structure indicating approximately 3.33 residues per turn of the left handed minor helix of each chain Fig. 4.13b [90]. The major right handed helix had 30 residues per turn. The neighbouring helices in the triple helical assembly are thus related by a twist of  $-108^\circ$  and rise of  $\sim 2.86\text{\AA}$  [90, 89]. However, this model still caused problems until a refinement by Rich & Crick showed only one inter-strand hydrogen bond existed per amino acid triplet Fig. 4.14a [91].

The collagen triple helix is stabilised by inter-chain hydrogen bond networks. Within each amino acid triplet, one hydrogen bond forms between the amide hydrogen (NH) atom of glycine, a hydrogen bond donor, in one chain and the carbonyl oxygen (C=O) atom of residue X in an adjacent strand, a hydrogen bond acceptor [87, 91] Fig. 4.14a. Additional stability comes from other weak hydrogen bond systems where the hydrogen of a residue is weakly bonded to the carbonyl oxygen atom

of another residue in an adjacent strand, and the two hydrogen atoms of glycine are weakly bonded to the carbonyl oxygen atom of another glycine located in one of the adjacent strands [92].

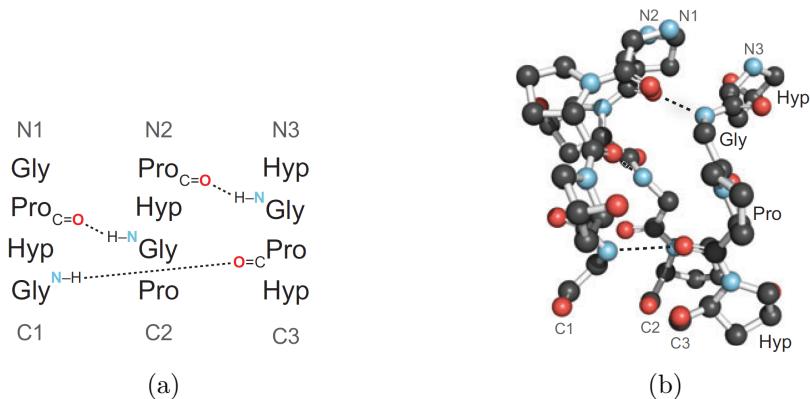


Figure 4.14: An illustration showing a segment of the collagen triple helix. A schematic diagram of the segment shows the hydrogen bonds between the amino acid triplet in each polypeptide Fig. 4.14a and the ball and stick image shown in Fig. 4.14b shows a 3D representation of the strong hydrogen bonds, and three polypeptides [80]. (Figures 1c and 1d from [80].)

## 4.2.2 Collagen Mechanics

Extensive studies of the mechanical properties of collagen molecules have been made. In single molecule pulling experiments different set-ups have shown that the force-extension behaviour in the entropic elasticity region can be modelled accurately by using the WLC model. At larger forces the increase in tensile strain causes the mechanical behaviour of the molecule to move into the energetic elasticity region where the WLC model fails. We expect to observe the stretching and breaking of hydrogen bonds followed by the deformation of covalent bonds [93]. Results from some of the early experiments probing the entropic response of collagen molecules are shown as force-extension curves in Fig. 4.15. Investigations by Sun et al. used optical tweezers to stretch human procollagen molecules of type I and II, the precursor form of collagen monomers. The chemical attachment to the beads was achieved by using the abundant amount of disulfide bonds present at the terminal ends of the procollagen molecules. Biotinylated type II procollagen was purified through a desalting column and coated on streptavidin-coated polystyrene beads; one terminal was fixed to a stationary bead on a moving plane while the other terminal was attached to a free bead trapped by optical tweezers [94]. The molecule was stretched

by moving the plane of the fixed bead and measuring the elastic response of the trapped bead using an optical laser [95]. Fitting their data to the WLC model they found that the collagen monomer had a persistence length of  $L_p \sim 15$  nm and a contour length of  $L \sim 300$  nm [95, 94].

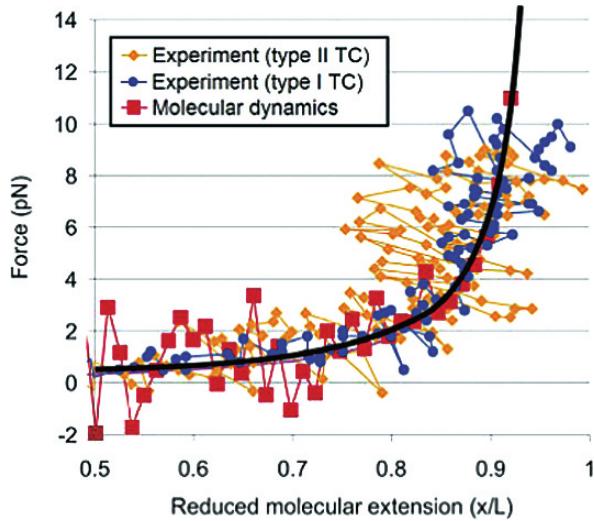


Figure 4.15: A plot of molecular extension of a single collagen molecule in the entropic regime ( $F < 14$  pN). The experimental results on type I and type II collagen molecules were reported in [96, 94] and the MD results were obtained from a study by Buehler et al. using a mesoscale model to predict the force-extension response of a collagen molecule [93]. Image was taken from [81].

In AFM experiments by Bozec et al. insoluble tropocollagen molecules with cleaved terminals were pulled from a surface without a covalent coupling between the tip and the molecule. This method had no control over the binding processes resulting in cases where multiple molecules were being pulled off the surface, and variations in position where the probe becomes bound to the molecule along its length. Repeat experiments were used to differentiate between single and multiple stretching events. The mechanical properties obtained from this method of stretching collagen molecules was determined to be only representative of the actual length of the monomer stretched rather than the entire molecule [24]. Shown in Fig. 4.16 are the results from these measurements fitted to the WLC model for small extensions. Similar experiments on rat tail tendon by Gutsmann et al. found that the steep increase in force was a result of stretching individual collagen molecules. Their force-extension curves were similar to the ones found by Bozec et al. [24, 97]. They determined the effective contour length of the pulled collagen monomer by fitting their data to the WLC model using a persistence length of  $L_p \sim 0.4$  nm.

Observations made by Bozec et al. did suggest that a proportion of the force-

extension curves obtained did feature a discontinuity, thought to be the result of structural changes within the sample, exhibiting a transition to different mechanical behaviour. In data sets where peaks did not exhibit a discontinuity it was found that the WLC model failed at larger extensions [24].

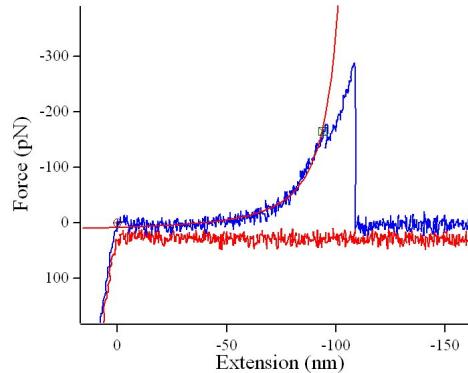


Figure 4.16: A force extension curve from pulling collagen molecules from a surface using atomic force microscopy. A persistence length of  $L_p \sim 0.4$  nm was used to fit the WLC model to the data. The discontinuity found in this plot at an extension of approximately 100 nm occurred in 20% of cases across an entire data set.

Recently, molecular dynamic simulations have been used to get a better understanding of the nanomechanical properties of tropocollagen molecules [93, 98, 86, 99]. In the full atomistic model calculations by Buehler et al. simulations were run using two types of force field parameters to include the atomistic interactions in proteins: a non-reactive CHARMM force field which provides harmonic and anharmonic terms to describe covalent, van der Waals, and electrostatic interactions, and a reactive ReaxFF force field which includes the dissociation of chemical bonds under deformation [86, 93]. Fixing one end of the tropocollagen molecule by applying constraints to each of the three carbon atoms, a mechanical load is applied axially using steered molecular dynamics to obtain a force-extension relationship. In these types of MD simulations an external force is applied through a virtual spring with a known stiffness constant to steer the tropocollagen molecule along a path to simulate mechanical stretching also known as steered molecular dynamics (SMD). In both cases Buehler et al. found the results to agree well with experiment for small strains up to 10% strain but deviated strongly at large strains. The maximum tensile force in the collagen molecule reached ,  $F \sim 2.35 \times 10^4$  pN Fig. 4.17a. At  $T = 300K$  the persistence length was found to be  $L_p \sim 23.4$  nm. Running a simulation for pulling a single polypeptide out of the tropocollagen molecule he found the strength to be  $\sim 0.713 \times 10^4$  pN reaching 3.7% tensile strain [86].

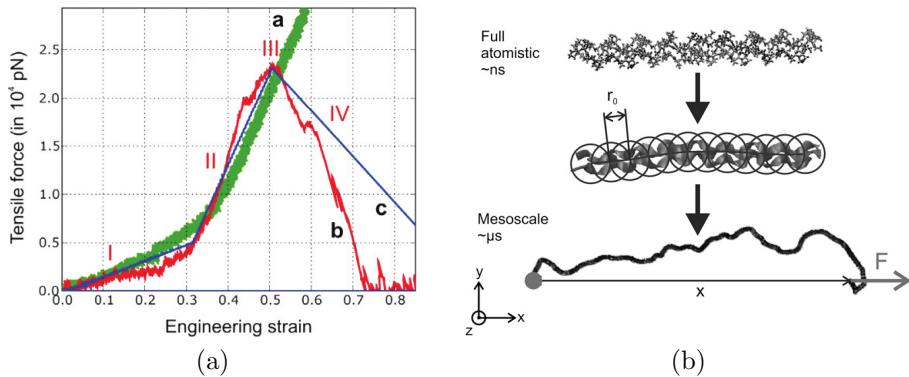


Figure 4.17: The force extension behaviour from the atomistic simulations of a single collagen molecule with  $L \sim 8.4$  nm are plotted in Fig. 4.17a. The methods used in these MD simulations had a non-reactive CHARMM force field (a), a reactive ReaxFF force field (b), and a reactive mesoscale model (c) shown in Fig. 4.17b. The tensile deformation of the molecule starts with the uncoiling of the triple helical structure (I), at larger strains the modulus gradient increases to reflect the stretching of covalent bonds (II). In phase (III) the molecule breaks with a rapid decay of force in phase IV [86, 93].

Taking a coarse-grained approach in the mesoscopic model, Buehler et al. simplified the tropocollagen molecule to a collection of beads. At this level information about the biochemical features within the molecule is lost Fig. 4.17b. Each bead represents approximately 10 protein atoms having a radius of  $7 \text{ \AA}$ . Again, one end of the molecule in the simulation is fixed while a force is applied to the other, thus extending the molecule. The MD results shown in Fig. 4.15 demonstrate a very good agreement with experimental results in the entropic regime. A fit to the WLC model found the persistence length to be  $L_p \sim 16$  nm where the contour length of the molecule was  $L = 301$  nm [93].

In a separate MD study, Gautieri et al. used a coarse-grained MARTINI force-field to model the collagen molecule after making modifications to include parameters that described a hydroxyproline amino acid residue and incorporated the triple helical conformational structure of collagen. All the amino acids are modelled by mapping four non-hydrogen atoms into one bead, and the number of beads used to model a specific residue is decided from the dimensions of the side chain of the amino acid. In determining the force field parameters that describe the triple helix structure of collagen, bond lengths between backbone beads, bonding angles, and dihedral angles were analysed using five collagen-like molecules. These parameters were then used in SMD simulations to obtain a reference force-extension behaviour to determine the elastic stiffness for the potential energy terms in bond stretching, angle deformation, and dihedral deformations [99].

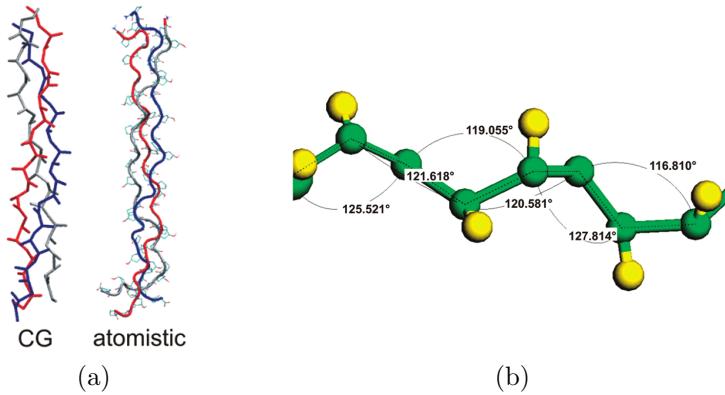


Figure 4.18: A comparison between the full atomistic and coarse-grained structures of collagen. Image Fig. 4.18b is a detailed view of the coarse grained polypeptides in collagen showing the angles between the backbone beads. The bond lengths between backbone beads, bonding angles, and dihedral angles were computed on the basis of five different collagen molecules. Gautieri et al. obtained a bond reference length  $d_b = 0.365 \pm 0.07$  nm, a bonding reference angle  $\varphi_a = 119.2 \pm 8.72^\circ$ , and a dihedral reference angle  $\Psi_d = -89.3 \pm 9.76^\circ$  [99].

Focusing our attention on the backbone, the spring constant  $K_b$  was calculated by simulating a single glycine-proline structure and then matching a best fit of the force-extension curve to the full atomistic simulations computed using GROMACS with GROMOS96 43a1 force field parameters. Shown in Fig. 4.19, a best fit was found for  $K_b = 1250$  kJ mol<sup>-1</sup> nm<sup>-2</sup>. Further details of the Extended MARTINI force fields are described comprehensively in [99, 100, 101]. The simulation of a 8 nm collagen molecule using the coarse-grained model had a stiffness of  $1052 \pm 51.23$  pN nm<sup>-1</sup> up to 15% strain as shown in Fig. 4.20. In simulating a 300 nm human type I collagen molecule Gautieri et al. obtained a persistence length of  $L_p = 51.5 \pm 6.7$  nm.

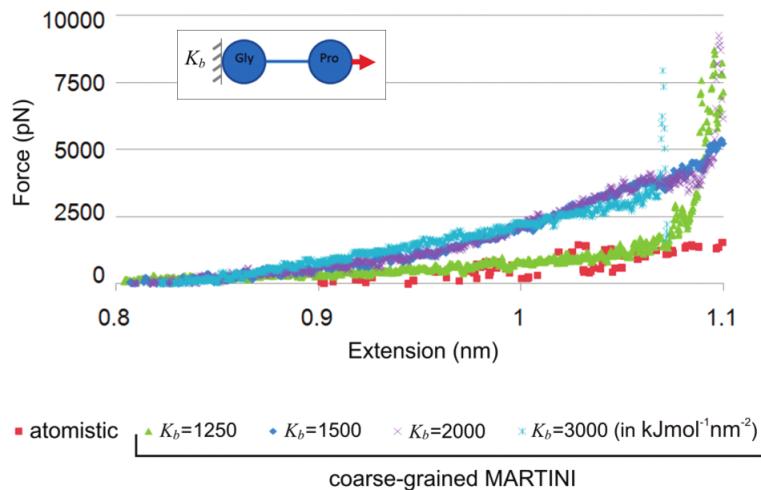


Figure 4.19: A plot determining the elastic stiffness of the backbone spring constant. The inset diagram is an illustration of the MD simulation studied with the results plotted for various values of  $K_b$ . Data comparison with the atomistic calculations suggest  $K_b = 1250 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  [99].

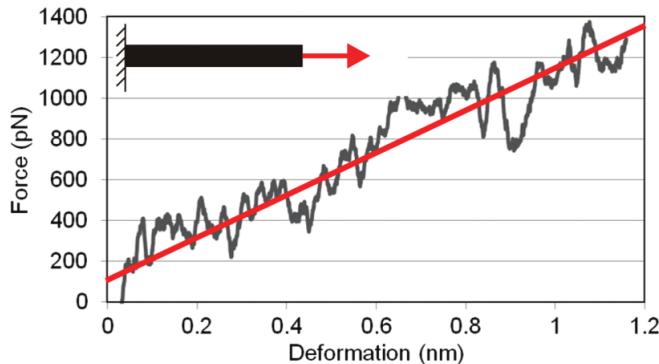


Figure 4.20: Force-extension plot of a 8 nm collagen molecule using the Extended MARTINI force field up to 15% strain (1.2 nm).

Apart from a few pulling experiments of collagen molecules, most of the work described has been focused on the work done by MD simulations giving valuable information on the collagen backbone stiffness. Unfortunately no experimental investigations on the axial shearing of collagen molecules have been found.

# Chapter 5

## The Extendable Freely Jointed Chain

Simple polymer models, excluding the WLC model, are variations of the ideal chain where each link in the chain is described as a uniform rod. To introduce a more realistic behaviour in the FJC we need to add further detail to the description of the polymer on the microscopic scale. We can do this by making each link in the chain extendable, providing extra statistical configurations in space and subsequently altering the entropy and free energy of the system. By considering each segment in the chain to stretch, a modification of the FJC model is obtained, an Extendable Freely Jointed Chain (EFJC) model. In the new model it is assumed that each link within the chain may extend by a small amount contributing to the overall extension of the FJC. By taking each link independently, we can employ a potential energy function to allow for the extensibility. We begin by modelling the FJC as analogous to a one dimensional random walk, and then find a relationship between the end position of the random walk and the partition function of an FJC with the same end-to-end separation of the path taken to reach the end position. We then discuss a suitable potential energy function to use for an EFJC, and evaluate the partition function for a FJC and EFJC in one dimension before obtaining them in three dimensions.

### 5.1 Probability Distribution of the Freely Jointed Chain

The FJC model can be used to describe the conformations of polymer molecules since the links have the ability to rotate freely around individual nodes and assume a limitless number of orientations [34]. Fixing one end of the chain, the other end has

a certain probability to lie at any other position in space depending on the position of the previous links. If we set our chain to have  $N$  such links each with a length  $a$ , then one of many possible configurations, and the simplest, is when the chain is fully extended linearly. Here the end-to-end distance of the chain is  $l = Na$ , which has all the links in the same orientation. The fact that the choice of the orientation of each link is random and that the probabilities of a FJC occupying some configuration in space is finite means that we can treat the FJC model as a random walk through space where the trajectory is the path taken by successive random steps [34]. Much like a chain, the length of the link is similar to the step size of the random walk, the number of steps taken in the random walk is equal to the number of links in the FJC, and the probability of a step taken in a random walk is the same as the probability that a link should take a particular configuration in space. The latter does not depend on the previous step or link configuration.

We can determine the probability distribution of end to end distance in the FJC by solving the master equation for a random walk using the multiplicative and additive laws of probability. Starting with,

$$P_{N+1}(x_k) = \sum_{k'=-\infty}^{\infty} P_N(x_{k'}) T(x_k - x_{k'} | x_{k'}) \quad (5.1)$$

where  $P_{N+1}(x_k)$  is the probability that the walk should end at position  $x_k$  after step  $N+1$  and  $T(\Delta x|x)$  is the transition probability for making a step of size  $\Delta x$  given a starting position of  $x$ .  $P_{N+1}(x_k)$  is a sum of probabilities of all the possible previous histories up to this point. For a symmetric random walk in 1-D, where  $x_k = ka$ ,

$$T(x_k - x_{k'} | x_{k'}) = \frac{1}{2} (\delta_{k,k'+1} + \delta_{k,k'-1}) \quad (5.2)$$

The terms in the brackets represent steps to the right  $k = k' + 1$  and left  $k = k' - 1$ . Hence, the master equation becomes

$$P_{N+1}(x_k) = \frac{1}{2} P_N(x_{k-1}) + \frac{1}{2} P_N(x_{k+1}) \quad (5.3)$$

which when solved gives the result, for  $|k| \leq N$ , and even  $(N - k)$  [102]:

$$P_N(x_k) = \frac{1}{2^N} \frac{N!}{\left(\frac{N-k}{2}\right)! \left(\frac{N+k}{2}\right)!} \quad (5.4)$$

A plot of this distribution in Fig. 5.1 shows that the probability is greatest for  $x_k = 0$ . This is true for all  $N$ . For larger values of  $N$  the random walk is able

to follow more trajectories reaching higher values of  $x_k$ , hence we see a broader probability distribution for  $N = 10$  when compared to  $N = 2$ .

Taking the sum over all possible configurations we can express the probability density function of walk displacement  $R$  as

$$P_N(R) = \sum_k P_N(x_k) \delta(x_k - R) \quad (5.5)$$

where the delta function specifies integer values for the continuous walk displacement  $R$ . With the expression for  $P_N(x_k)$  being a set of binomial coefficients we can represent Eq. 5.5 as

$$P_N(R) = \frac{1}{2^N} \sum_{k=-N}^N \binom{N}{\frac{N+k}{2}} \delta(x_k - R) \quad (5.6)$$

$$= \frac{1}{2^N} \sum_{K=0}^N \binom{N}{K} \delta(x_{2K-N} - R) \quad (5.7)$$

where  $K = \frac{N+k}{2}$  [102]. The probability that the walk should end with a displacement in  $R \rightarrow R + dR$  in the continuous limit would then be  $\int P_N(R) dR$ , but since  $R$  is an integer the probability of obtaining the result in  $R_1 \leq R \leq R_2$  is  $\int_{R_1}^{R_2} P_N(R) dR$ .

When  $x_k = Na$  the FJC is linear; This being in only one possible configuration we find it has the least probability of being in this state. This is demonstrated in Fig. 5.1 and Fig. 5.2. Treating the FJC now in statistical mechanics we can expect the partition function of the system with end-to-end separation  $R$  to be the greatest when  $R = 0$ . The end-to-end polymer length  $R$  is analogous to the displacement  $x_k$  in the random walk. In relation to higher values of  $N$  we can expect the distribution to broaden.

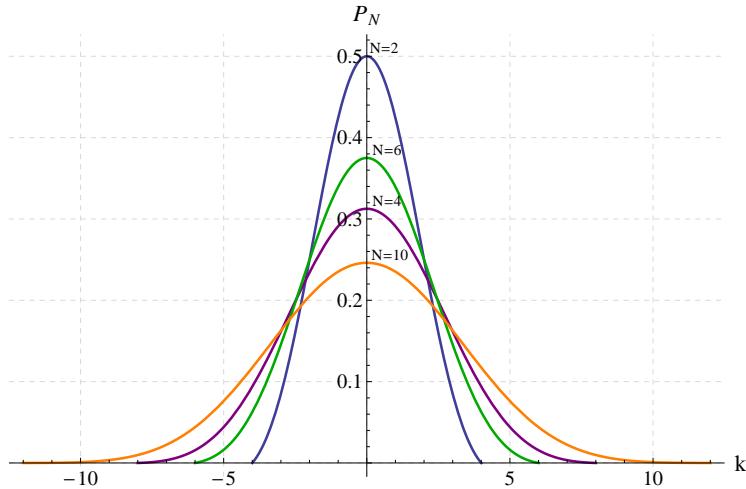


Figure 5.1: A plot showing an envelope of the expected probability distribution over position  $x_k$  for  $N = 2, 4, 6, 10$ . The probability distributions are the sum of contributions over all paths.

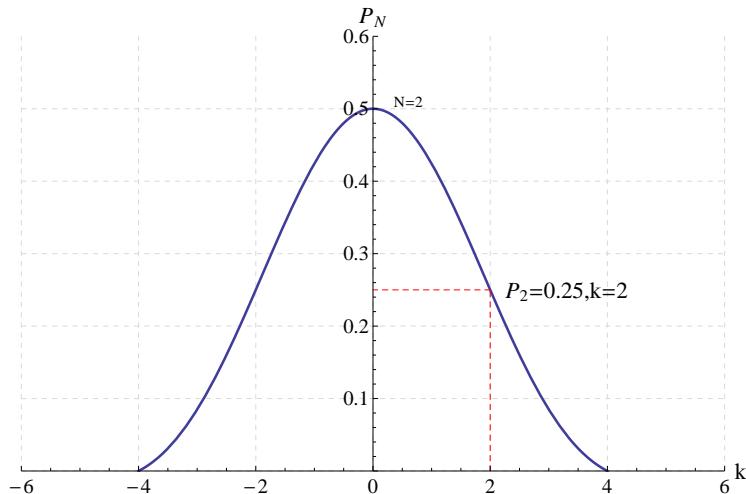


Figure 5.2: A plot showing an envelope of the expected probability distribution of displacement  $x_k$  for  $N = 2$ . The dashed lines indicate that the end position is twice as likely to be at the origin,  $x_k = 0$  than to be at either  $x_k = 2a$  or  $x_k = -2a$ .

## 5.2 Construction of the Partition Function Integral

In constructing our partition function for an EFJC we will consider a system consisting of  $N + 1$  particles each of which interacts with its neighbouring particles. The partition function is an integral over phase-space of the exponential of the system's Hamiltonian. The integral represents the sum over all configurations for a  $N$ -particle

system. The Hamiltonian will later allow us to include our desired potential. We work initially in one dimension. The partition function in a system with  $N$  links is

$$\mathcal{Z} = \frac{1}{h^N} \int \prod_{i=1}^N dq_i dp_i \exp(-H(p_i, q_i)/kT) \quad (5.8)$$

where  $h$  is Planck's constant,  $q_i$  is the phase space co-ordinate of the  $i^{th}$  particle,  $p_i$  is its momentum,  $H(p_i, q_i)$  is the Hamiltonian,  $k$  is the Boltzmann constant, and  $T$  is the absolute temperature. The zeroth particle is held static at the origin.

The Hamiltonian, which is a function of both position and momentum, represents the total energy of the system. It is the sum of the kinetic energy and the potential energy of all the constituents,

$$\hat{H}(p_i, q_i) = \sum_{i=1}^N \left( \frac{p_i^2}{2m} + \Phi(q_i, q_{i-1}) \right) \quad (5.9)$$

where  $m$  is the mass of each particle and  $\Phi(q_i, q_{i-1})$  is the potential controlling the length of each link. Because the position and momentum parameters are independent in the Hamiltonian it allows us to fully separate the spatial and momentum integrations of the partition function Eq. 5.8.

$$\begin{aligned} \mathcal{Z} &= \frac{1}{h^N} \int \prod_{i=1}^N dp_i \exp\left(-\frac{p_i^2}{2mkT}\right) \int \prod_{i=1}^N dq_i \exp\left(-\frac{\sum_{i=1}^N \Phi(q_i, q_{i-1})}{kT}\right) \\ &= \frac{1}{h^N} Z_p Z_q \end{aligned} \quad (5.10)$$

The momentum partition function in Eq. 5.10 is simply a Gaussian integral raised to the power of  $N$ . This allows us to integrate over the momentum for all the individual links in the chain. It can be simplified to a constant leaving the position partition function integral to be evaluated.

$$Z_p = (2\pi mkT)^{\frac{N}{2}} \quad (5.11)$$

The breakdown of the dimensionless partition function gives the position integral  $Z_q$  a dimension of length to the power of  $N$ . To simplify the position integral in Eq. 5.10 the potential  $\Phi(q_i, q_{i-1})$  needs to be written in terms of a set of variables,  $q_{i,i-1} = (q_i - q_{i-1})$ . The Jacobian for this change in variables is unity. A Fourier transform representation of the Dirac delta function is then inserted into the partition function to impose the following constraint. We consider the case where the end-to-end

polymer length is equal to  $R$ , where  $R = q_N - q_0$ . We write

$$Z_q = \int_{-\infty}^{\infty} dR Z_q(R) \quad (5.12)$$

with

$$Z_q(R) = \int \prod_{i=1}^N dq_i \delta(q_N - q_0 - R) \exp\left(-\frac{\sum_{i=1}^N \Phi(q_i - q_{i-1})}{kT}\right) \quad (5.13)$$

and employ

$$\begin{aligned} \delta(q_N - q_0 - R) &= \delta\left(\sum_{i=1}^N q_{i,i-1} - R\right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \exp\left[i\omega\left(\sum_{i=1}^N q_{i,i-1} - R\right)\right] \end{aligned} \quad (5.14)$$

Inserting Eq. 5.14 into  $Z_q(R)$ , we get

$$\begin{aligned} Z_q(R) &= \int_{-\infty}^{\infty} \prod_{i=1}^N \left[ dq_{i,i-1} \exp\left(-\frac{\Phi(q_{i,i-1})}{kT}\right) \right] \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \exp\left[i\omega\left(\sum_{i=1}^N q_{i,i-1} - R\right)\right] \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \exp(-i\omega R) \left[ \int_{-\infty}^{\infty} dq_{i,i-1} \exp\left(-\frac{\Phi(q_{i,i-1})}{kT}\right) \exp(i\omega q_{i,i-1}) \right]^N \end{aligned} \quad (5.15)$$

Using a suitable potential we can solve Eq. 5.15 to find the partition function. In the next section we will discuss potential energy functions to describe a FJC and an EFJC.

### 5.2.1 Quadratic Potential

For an illustration of how one might solve the partition function using the Fourier transform partition function, Eq. 5.15, we will begin by fixing the configuration of the EFJC to be linear, and then allow each link to extend harmonically. The focus on a linear configuration takes the model away from a theory of polymers to a theory of an elastic rod. Considering only the 1-D case along the  $x$ -axis the harmonic potential is

$$\Phi(x) = \frac{1}{2}\alpha(x - a)^2 \quad (5.16)$$

where  $a$  corresponds to the length of each link in the absence of a load, and  $\alpha$  is the spring constant. Combining the elastic potential with Eq. 5.15 gives

$$Z_q^{rod}(R) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \exp(-i\omega R) \left[ \int_{-\infty}^{\infty} dx \exp\left(-\frac{\alpha(x-a)^2}{2kT}\right) \exp(i\omega x) \right]^N \quad (5.17)$$

which when evaluated gives an expression

$$Z_q(R) = \left(\frac{2\alpha\pi}{NkT}\right)^{\frac{1}{2}} \left(\frac{2kT\pi}{\alpha}\right)^{\frac{N}{2}} \exp\left(-\frac{\alpha(R-aN)^2}{2kTN}\right) \quad (5.18)$$

The case described by Eq. 5.18 corresponds to one configuration of the EFJC with  $R = Na$ . The results from Fig. 5.3 show that Eq. 5.18 for  $N=4$  describes a Gaussian distribution peaking at  $R = 4a$ . This is the unstretched length of the rod. Irrespective of the many random configurations each link can take within the EFJC, all the links described by this partition function follow the same direction to make the chain comparable to a rod of length  $Na$ .

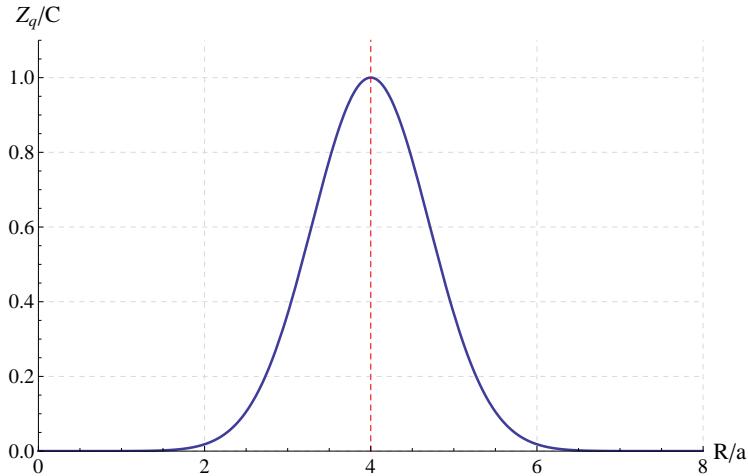


Figure 5.3: A plot showing the value of  $Z_q/C$  as a function of  $R$  using an harmonic potential for  $N = 4$ . The constant  $C$  is a scaling factor and  $\alpha/2kTN$  is set to unity.

### 5.2.2 Double Potential

We will begin by looking at how we can create a suitable potential to describe the behaviour of the links that make up the FJC. In modelling each link we will imagine that one end of the link is fixed while the other end is free to lie in any direction characterised by a potential. By subsequently using this method for  $N$  links we create an FJC. For example, if the length of a link is  $a$ , and consider one end of

the link to be fixed at a point  $\beta$  on the  $x$ -axis, the other end would be positioned at either  $\beta + a$  or  $\beta - a$ . We can allow each link to have this spatial freedom by setting up a potential,  $\Phi(x)$ , that is infinite everywhere except at  $a$  and  $-a$ . The exponential of the potential as seen in Eq. 5.15, should correspond to two Dirac delta weighting functions with zero weight everywhere except at  $a$  and  $-a$ . The link cannot be extended further than its unperturbed length,  $a$ . This potential would describe a non-extendable FJC in one dimension. The potential for the FJC is inserted such that

$$\exp\left(-\frac{\Phi(q_{i,i-1})}{kT}\right) = \frac{A}{2} (\delta(q_{i,i-1} - a) + \delta(q_{i,i-1} + a)) \quad (5.19)$$

where  $A$  is a constant with dimensions of length.

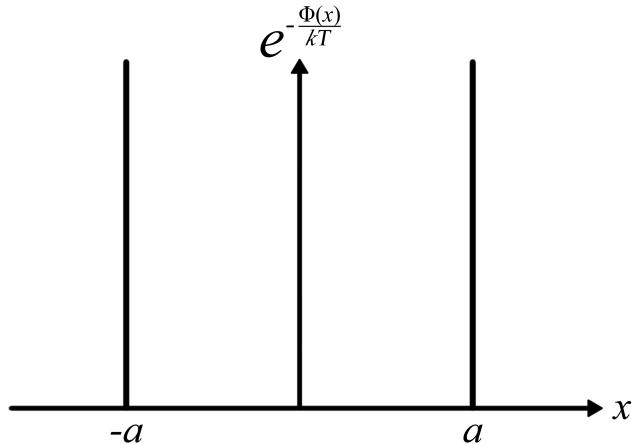


Figure 5.4: A plot of the double delta function weighting function where the value of  $\exp\left(-\frac{\Phi(x)}{kT}\right)$  is infinite at  $\pm a$ .

The partition function is then

$$Z_q(R) = \int \prod_{i=1}^N dq_{i,i-1} A^N \delta\left(\sum_{i=1}^N q_{i,i-1} - R\right) \prod_{i=1}^N \left(\frac{1}{2} (\delta(q_{i,i-1} - a) + \delta(q_{i,i-1} + a))\right) \quad (5.20)$$

This is intuitively proportional to a sum over paths, constrained such that the total displacement is  $R$  through the first delta function in Eq. 5.20. Each step is  $\pm a$  imposed by each of the second set of delta functions. It is therefore the sum of all possible outcomes of the symmetric random walk with given  $R$ . This establishes the analogy between  $Z_q(R)$  and  $P_N(R)$  derived in 5.1. The normalisation of the

probability distribution for  $R$  requires that

$$P_N(R) = \frac{Z_q(R)}{\int dR Z_q(R)} \quad (5.21)$$

The transition probability of the analogous random walk is

$$T(q_{i,i-1}) = \frac{1}{2} (\delta(q_{i,i-1} - a) + \delta(q_{i,i-1} + a)) \quad (5.22)$$

This is normalised such that

$$\int_{-\infty}^{\infty} T(q_{i,i-1}) dq_{i,i-1} = 1 \quad (5.23)$$

In order to pursue the analogy to a random walk, we need to employ a weighting function such that the transition probability in the analogous random walk is

$$T(q_{i,i-1}) = \frac{\exp\left(-\frac{\Phi(q_{i,i-1})}{kT}\right)}{\int_{-\infty}^{\infty} \exp\left(-\frac{\Phi(q_{i,i-1})}{kT}\right) dq_{i,i-1}} \quad (5.24)$$

which is correctly normalised.

### 5.2.3 Evaluating $Z_q(R)$ for the non-extendable Freely Jointed Chain

Following on from Eq. 5.20, where the exponential term is replaced by two delta weighted functions, the partition function becomes

$$Z_q(R) = \frac{A^N}{2\pi} \int_{-\infty}^{\infty} d\omega e^{-i\omega R} \left[ \frac{1}{2} \int_{-\infty}^{\infty} dx \{ \delta(x-a) + \delta(x+a) \} e^{i\omega x} \right]^N \quad (5.25)$$

Taking the terms that correspond to the Fourier transform of the weighting function as  $\kappa(\omega)$ , and using the sifting property of the Dirac delta function we get the following [103],

$$Z_q(R) = \frac{A^N}{2\pi} \int_{-\infty}^{\infty} d\omega e^{-i\omega R} [\kappa(\omega)]^N \quad (5.26)$$

where

$$\kappa(\omega) = \frac{1}{2} \int_{-\infty}^{\infty} \delta(x-a) e^{i\omega x} dx + \frac{1}{2} \int_{-\infty}^{\infty} \delta(x+a) e^{i\omega x} dx \quad (5.27)$$

$$= \frac{e^{i\omega a} + e^{-i\omega a}}{2} \quad (5.28)$$

$$= \cos \omega a \quad (5.29)$$

The partition function then becomes

$$Z_q(R) = \frac{A^N}{2\pi} \int_{-\infty}^{\infty} d\omega e^{-i\omega R} \cos^N \omega a \quad (5.30)$$

The cos term which is raised to the  $N^{th}$  power can be expressed as a sum of exponentials using the binomial theorem,

$$\cos^N x = \frac{1}{2^N} \sum_{k=0}^N \binom{N}{k} e^{ix(N-2k)} \quad (5.31)$$

for all integer  $k$ . Combining this and the general integral representation of the Dirac delta function [103]

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} dt \quad (5.32)$$

we get

$$Z_q(R) = \frac{A^N}{2^{N+1}\pi} \sum_{k=0}^N \binom{N}{k} \int_{-\infty}^{\infty} e^{-i\omega(R+2k-N)} d\omega \quad (5.33)$$

$$= \frac{A^N}{2^{N+1}\pi} \sum_{k=0}^N \binom{N}{k} \delta(R+2k-N) \quad (5.34)$$

Here we obtain the partition function for a FJC in one dimension. Evaluating the integral of Eq. 5.34 with respect to  $R$  we can show that

$$\int Z_q(R) dR = \frac{A^N}{2^{N+1}\pi} \sum_{k=0}^N \binom{N}{k} = \frac{A^N}{2\pi} \quad (5.35)$$

and hence

$$P_N(R) = \frac{Z_q(R)}{\int Z_q(R) dR} = \frac{1}{2^N} \sum_{k=0}^N \binom{N}{k} \delta(R+2k-N) \quad (5.36)$$

as derived earlier in Eq. 5.7. Taking this further, the partition function for the EFJC will next be calculated using an appropriate weighting function.

## 5.3 Extendable Freely Jointed Chain Model in One Dimension

### 5.3.1 Potential for an Extendable Freely Jointed Chain

To make the FJC extendable we replace the double delta function weighting factors with two top-hat functions of finite width  $p$ . This corresponds to a pair of infinite square wells and since they have a finite width  $p$ , it allows each link is able to extend and contract by a length  $\frac{p}{2}$  without a cost in energy. The normalised transition probability for the random walk analogous to the EFJC would be  $T(q_i, q_{i-1})$  such that

$$\int_{-\infty}^{\infty} T(q_{i,i-1}) dq_{i,i-1} = \frac{1}{2p} \left[ \int_{-a-\frac{p}{2}}^{-a+\frac{p}{2}} dq_{i,i-1} + \int_{a-\frac{p}{2}}^{a+\frac{p}{2}} dq_{i,i-1} \right] = 1 \quad (5.37)$$

The effect of taking the limit of  $p \rightarrow 0$  whilst preserving the normalisation makes the height of each weighting function infinitely large such that they become delta functions. Analytically, we can show that our top-hat weighting functions become proportional to delta functions in the limit of  $p \rightarrow 0$  by using a combination of Heaviside functions to express an integral over finite limits as an integral over all space.

$$H(x) = \int_{-\infty}^x \delta(t) dt \quad (5.38)$$

Starting with the top-hat weighting functions in one dimension we can express the weight of the potential as two separate integrals where the width of weighting function  $\varsigma(x)$  is contained in the integral limits.

$$\int_{a-\frac{p}{2}}^{a+\frac{p}{2}} \varsigma(x) dx + \int_{-a-\frac{p}{2}}^{-a+\frac{p}{2}} \varsigma(x) dx \quad (5.39)$$

Representing a top-hat weighting function with two Heaviside step functions we have [104],

$$\square(x; p) \equiv H\left(x + \frac{p}{2}\right) - H\left(x - \frac{p}{2}\right) \quad (5.40)$$

Where  $x$  would be the centre position of the top-hat function. The weighting factor in the integrand, and implicitly the potential  $\Phi$ , will be written

$$\exp\left(-\frac{\Phi(q_{i,i-1})}{kT}\right) = \frac{A}{2p} (\square(x-a;p) + \square(x+a;p)) \quad (5.41)$$

Where  $A$  again has dimensions of length, and will turn out to be analogous to the  $A$  in Eq. 5.19. Using Eq. 5.40 each integral in Eq. 5.39 can now be expressed as an integral over all space,

$$\int_{a-\frac{p}{2}}^{a+\frac{p}{2}} \varsigma(x) dx = \int_{-\infty}^{\infty} \square(x-a;p) \varsigma(x) dx \quad (5.42)$$

$$\int_{-a-\frac{p}{2}}^{-a+\frac{p}{2}} \varsigma(x) dx = \int_{-\infty}^{\infty} \square(x+a;p) \varsigma(x) dx \quad (5.43)$$

The positions  $a$  and  $-a$  correspond to the positions of the top-hat potentials as shown in Fig. 5.5. Since the value of  $p$  is small and finite we can use the mean value theorem for integration to express each integral as,

$$\int_{-\infty}^{\infty} \frac{1}{p} \square(x-a;p) \varsigma(x) dx = \varsigma(\tau) \quad (5.44)$$

$$\int_{-\infty}^{\infty} \frac{1}{p} \square(x+a;p) \varsigma(x) dx = \varsigma(\tau') \quad (5.45)$$

for  $\tau$  such that  $x-a-\frac{p}{2} \leq \tau \leq x-a+\frac{p}{2}$  and  $\tau'$  such that  $x+a-\frac{p}{2} \leq \tau' \leq x+a+\frac{p}{2}$ .

Then taking the limit of  $p \rightarrow 0$  we get,

$$\lim_{p \rightarrow 0} \left[ \int_{-\infty}^{\infty} \frac{1}{p} \square(x-a;p) \varsigma(x) dx \right] = \lim_{p \rightarrow 0} \left[ \varsigma\left(x-a \pm \frac{p}{2}\right) \right] = \varsigma(x-a) \quad (5.46)$$

$$\lim_{p \rightarrow 0} \left[ \int_{-\infty}^{\infty} \frac{1}{p} \square(x+a;p) \varsigma(x) dx \right] = \lim_{p \rightarrow 0} \left[ \varsigma\left(x+a \pm \frac{p}{2}\right) \right] = \varsigma(x-a) \quad (5.47)$$

Thus in the  $p \rightarrow 0$  limit this procedure gives rise to two delta functions at the positions  $a$  and  $-a$ ,

$$\lim_{p \rightarrow 0} \left[ \frac{1}{p} \square(x-a;p) \right] = \delta(x-a) \quad (5.48)$$

$$\lim_{p \rightarrow 0} \left[ \frac{1}{p} \square(x+a;p) \right] = \delta(x+a) \quad (5.49)$$

The analysis demonstrates that the top-hat potentials reduce to delta functions in the limit of  $p \rightarrow 0$  which we would expect. Moreover, by applying the appropriate

$p \rightarrow 0$  limits in the final expression for  $Z_q(R)$  for an EFJC, we can check that it behaves like a non-extendable FJC. In other words, Eq. 5.41 should tend towards Eq. 5.19.

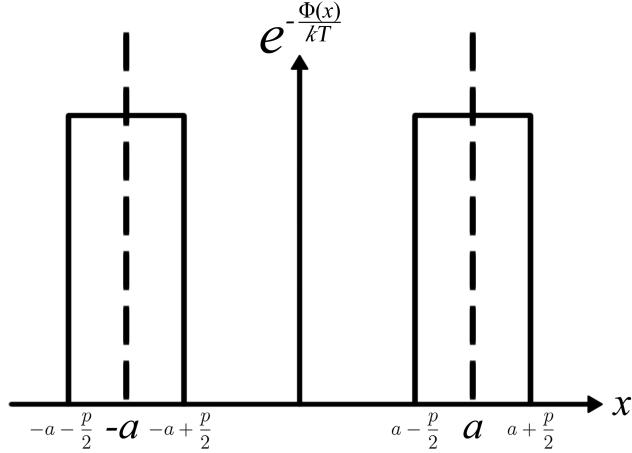


Figure 5.5: A plot showing a pair of top-hat functions as the weighting function with a finite thickness  $p$  and finite height. The top-hats are centred about  $a$  and  $-a$ .

### 5.3.2 Extendable Freely Jointed Chain

Working from Eq. 5.15 taking our one dimensional model along the  $x$ -axis in a Cartesian co-ordinate system we have,

$$Z_q(R) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{-i\omega R} \left[ \int_{-\infty}^{\infty} d\Delta x e^{-\frac{\Phi(\Delta x)}{kT}} e^{i\omega \Delta x} \right]^N \quad (5.50)$$

We have already obtained the partition function for the FJC by replacing the weighting factor in the integrand with a sum of two delta functions. For a EFJC we will now use two top-hat functions as the weighting factor for the potential, the Fourier transform weighted potential,  $G(\omega)$ , is then written as two integrals over finite limits. The Fourier transform of the weighting factor for the EFJC becomes,

$$G(\omega) = \int_{-\infty}^{\infty} e^{-\frac{\Phi(\Delta x)}{kT}} e^{i\omega \Delta x} d\Delta x \quad (5.51)$$

$$= \frac{A}{2p} \int_{-a-\frac{p}{2}}^{-a+\frac{p}{2}} e^{i\omega \Delta x} d\Delta x + \frac{A}{2p} \int_{a-\frac{p}{2}}^{a+\frac{p}{2}} e^{i\omega \Delta x} d\Delta x \quad (5.52)$$

$$= \left( \frac{2A}{\omega p} \right) \cos \omega a \sin \frac{\omega p}{2} \quad (5.53)$$

Combining Eq. 5.50 and Eq. 5.53 gives a simpler expression for the partition function,

$$Z_q(R) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{-i\omega R} \left[ \left( \frac{2A}{\omega p} \right) \cos \omega a \sin \frac{\omega p}{2} \right]^N \quad (5.54)$$

At this point we can begin to consider how  $p$  affects the solution and enquire whether an approximation of small  $p$  helps to simplify Eq. 5.54 to produce a neat solution. It is perfectly acceptable to take the small approximation of  $p$  since we have stated that the width of the top-hat weighted potential is small and finite. For the EFJC different methods can be used to impose the limit where  $p$  becomes finite producing the desired result for  $Z_q(R)$ . Each method starts from Eq. 5.54.

### 5.3.3 Imposing the limit of $p$ by the small approximation of $\sin(\frac{\omega p}{2})$

Applying the small approximation of  $p$  we linearise the trigonometric function such that  $\sin(\frac{\omega p}{2}) \simeq \frac{\omega p}{2}$  before the integral is evaluated. Eq. 5.54 becomes,

$$Z_q(R) = \frac{A^N}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega R} \cos^N \omega a d\omega \quad (5.55)$$

Then using Eq. 5.31 and the integral representation of the Dirac delta function we recover the partition function for the FJC,

$$Z_q(R) = \frac{A^N}{2^{N+1}\pi} \sum_{k=0}^N \binom{N}{k} \delta(R + 2k - N) \quad (5.56)$$

With the small approximation of  $p$ , we have effectively approximated  $G(\omega)$  to first order which gives the model for a FJC. Next we will include the higher order terms in the expansion of  $\sin(\frac{\omega p}{2})$ .

### 5.3.4 Using the Taylor expansion of $\sin \frac{\omega p}{2}$ to include higher order terms

By employing the expansion of  $\sin(\frac{\omega p}{2})$  we are including higher order terms in the calculation of  $Z_q$  before imposing the limit of  $p \rightarrow 0$ . Working from Eq. 5.54 we can write  $\cos^N \omega a$  as a series of exponentials using Eq. 5.31. The Taylor expansion of

the sine term can be written as

$$\frac{\sin^N\left(\frac{\omega p}{2}\right)}{\left(\frac{\omega p}{2}\right)^N} = \sum_{m=0}^{\infty} A_m (\omega p)^m \quad (5.57)$$

The partition function for an EFJC then becomes

$$Z_q(R) = \frac{A^N}{2^{N+1}\pi} \sum_{k=0}^N \binom{N}{k} \int_{-\infty}^{\infty} e^{-i\omega(R+2k-N)} \sum_{m=0}^{\infty} A_m (\omega p)^m d\omega \quad (5.58)$$

The complex integral maybe be written [104]

$$\int_{-\infty}^{\infty} \omega^b e^{-i\alpha\omega} d\omega = 2\pi i^b \delta^{(b)}(\alpha) \quad (5.59)$$

where  $\delta^{(b)}(\alpha)$  is the  $b^{th}$  derivative of the delta function. So inserting this solution into the power series we get

$$Z_q(R) = \frac{A^N}{2^{N+1}\pi} \sum_{k=0}^N \binom{N}{k} \sum_{m=0}^{\infty} A_m 2\pi (ip)^m \delta^m(\alpha) \quad (5.60)$$

where  $\alpha = R + 2k - N$ . We can then express the differential of the delta functions as,  $\delta^n(\gamma) = (-1)^n n! \delta(\gamma) / \gamma^n$ , so

$$Z_q(R) = \frac{A^N}{2^{N+1}\pi} \sum_{k=0}^N \binom{N}{k} \sum_{m=0}^{\infty} G_m p^m \delta(\alpha) \quad (5.61)$$

where  $G_m = A_m 2\pi (-i)^m m!$ . Having the Dirac delta function in the result for  $Z_q$  is important because it shows that only when  $\alpha = 0$  do we have a non-zero result for the partition function. At this stage we can take the limit of  $p \rightarrow 0$ ,

$$Z_q(R) = \frac{A^N}{2^{N+1}\pi} \sum_{k=0}^N \binom{N}{k} \left( G_0 \delta(\alpha) + \lim_{p \rightarrow 0} \sum_{m=1}^{\infty} G_m p^m \delta(\alpha) \right) \quad (5.62)$$

which becomes,

$$Z_q(R) = \frac{A^N}{2^{N+1}\pi} \sum_{k=0}^N \binom{N}{k} G_0 \delta(\alpha) \quad (5.63)$$

as required, as in Eq. 5.34.

Both methods described above yield the same result for the small approximation for  $p$ , with the solution being the same as the FJC. However, the analysis is somewhat

misleading since the general partition function appears to be a modified set of delta functions, rather than a continuous function of  $R$ . This outcome is a consequence of assuming that the series in  $p$  produces convergent integrals which is unclear. In order to obtain the correct partition function for the EFJC we will need to consider the exact solution of Eq. 5.54.

### 5.3.5 Expanding the Fourier Transform of the Potential as a Series

With the Fourier transform of the weighted potential raised to the  $N^{th}$  power we can express the trigonometric functions as a series of exponentials with sum over  $k'$  and  $k$  using the binomial theorem. The expression for  $\cos^N(x)$  is given by Eq. 5.31 and the term for  $\sin^N(x)$  is given by,

$$\sin^N x = \frac{1}{(2i)^N} \sum_{k'=0}^N \binom{N}{k'} (-1)^{k'} e^{ix(N-2k')} \quad (5.64)$$

Using Eq. 5.31 and Eq. 5.64 for the trigonometric terms, Eq. 5.54 becomes

$$Z_q(R) = \frac{A^N}{2^{N+1}\pi i^N p^N} \sum_{k=0}^N \sum_{k'=0}^N \binom{N}{k} \binom{N}{k'} (-1)^{k'} \int_{-\infty}^{\infty} \frac{e^{i\omega(a(N-2k)+\frac{p}{2}(N-2k')-R)}}{\omega^N} d\omega \quad (5.65)$$

The complex multipole integral takes the form,

$$\int_{-\infty}^{\infty} \frac{e^{i\omega\zeta}}{\omega^N} d\omega = \frac{\pi i^N \zeta^{N-1}}{(N-1)!} \text{sgn}(\zeta) \quad (5.66)$$

where

$$\text{sgn}(\zeta) = \begin{cases} -1 & \text{for } \zeta < 0 \\ 0 & \text{for } \zeta = 0 \\ 1 & \text{for } \zeta > 0 \end{cases} \quad (5.67)$$

Using Eq. 5.66 to evaluate the integral, Eq. 5.65 becomes

$$Z_q(R) = \frac{A^N}{2^{N+1}p^N} \sum_{k=0}^N \sum_{k'=0}^N \binom{N}{k} \binom{N}{k'} (-1)^{k'} \frac{\eta^{N-1}}{(N-1)!} \text{sgn}(\eta) \quad (5.68)$$

where  $\eta = a(N-2k) + \frac{p}{2}(N-2k') - R$ .

Referring to the case where  $N = 2$ , we see in Fig. 5.6b that the partition function given by Eq. 5.68 is twice as large at  $R = 0$  than at  $R = 2$ . This is what we expect

to find for the FJC chain in 1D. For the case when  $N = 1$  only one link exists in the FJC which means that the end point can only lie at one of the two positions denoted by  $R = 1, -1$ . A plot of  $Z_q(R)$  where  $N = 1$  and  $p = 0.1a$  is shown in Fig. 5.6a. For the cases where  $N$  is an odd number the partition function at an even position,  $k = 0, 2, 4, 6, 8, \dots$ , will always be zero since the end of the EFJC will never reach those positions unless  $p$  is larger.

As the number of links increases in the EFJC we see from Fig. 5.6e and Fig. 5.6f that Eq. 5.68 begins to break down giving irregular results. The nature of this problem is due to the fact that as  $N$  gets larger the cancellations due to  $\text{sgn}(\eta)$  in the polynomials become extremely delicate. The sums in  $Z_q(R)$  for low  $N$  involves relevant cancellations to give the required result, as shown in Table. 5.1.

To overcome this problem we can impose constraints from the analysis of Eq. 5.68 to assist with the cancellations that take place. In the polynomial  $\eta^{N-1}\text{sgn}(\eta)$  we will rewrite  $\eta$  such that,

$$\eta = F(k, k') - R \quad (5.69)$$

Where

$$F(k, k') = a(N - 2k) + \frac{p}{2}(N - 2k') \quad (5.70)$$

To impose the correct constraints let us examine in detail the nature of Eq. 5.68. The partition function is a sum of terms labelled by  $k$  and  $k'$ . We shall refer to  $k$  as an index denoting the coarse scale structure of  $Z_q(R)$  and  $k'$  as a label of the fine scale structure for reasons that will become clear later. For a given value of  $k$ , the core of the expression is

$$\sum_{k'=0}^N (-1)^{k'} \eta^{N-1} \text{sgn}(\eta) \quad (5.71)$$

From Eq. 5.69 and Eq. 5.71 we can deduce that for  $R < \min_{k'}(F(k, k')) = F(k, N) = a(N - 2k) - \frac{Np}{2}$ ,  $\eta > 0$  for all  $k'$ . Therefore  $\text{sgn}(\eta) = 1$ . Hence for  $R$  below such a threshold, the core expression reduces to

$$\sum_{k'=0}^N (-1)^{k'} \eta^{N-1} \quad (5.72)$$

By writing  $\eta$  as  $A + Bk'$ , the evaluation of Eq. 5.72 requires us to consider

$$\sum_{k'=0}^N \binom{N}{k'} (-1)^{k'} k'^m \quad (5.73)$$

with  $m$  taking integer values between zero and  $N - 1$ . However, it may be shown

that all such summations are zero, for  $m < N$ . Similarly, for  $R > \max_{k'}(F(k, k')) = F(k, 0) = a(N - 2k) + \frac{Np}{2}$  we find that  $\eta < 0$  and  $\text{sgn}(\eta) = -1$  for all  $k'$ . Once again the core expression reduces to

$$\sum_{k'=0}^N \binom{N}{k'} (-1)^{k'} \eta^{N-1} \quad (5.74)$$

and by similar reasoning, this vanishes.

We find, therefore, that contributions to  $Z_q(R)$  for a given  $k$  in Eq. 5.68 only arise for the region  $a(N - 2k) - \frac{Np}{2} \leq R \leq a(N - 2k) + \frac{Np}{2}$ . This corresponds to a region of width  $Np$  about a central position  $a(N - 2k)$ . Within this range, a non-zero contribution to  $Z_q(R)$  is made from the sum over  $k'$ , with modulation due to the change in sign of  $\eta$  at some point in the sum over  $k'$ .

Thus each value of  $k$  in the expression for  $Z_q(R)$  defines a coarse region of  $R$  within which non-zero contributions are made due to the summation over  $k'$ . For this reason we characterise  $k$  as the coarse structure label and  $k'$  as a fine structure label.

Returning to Eq. 5.71, we see that it is a sum of polynomials in  $R$ , of order  $N - 1$ , each with a different zero corresponding to  $R = F(k, k')$ . For values of  $R$  further and further away from each zero, the polynomial  $(F(k, k') - R)^{N-1}$  increases rapidly, especially for large  $N$  and yet the sum of all the terms vanishes. This is due to a very delicate cancellations of terms. In numerical implementations, such a cancellation will fail in detail, as is shown by the breakdown in the calculation of  $Z_q(R)$  for larger  $N$  in Fig. 5.6e and Fig. 5.6f. This difficulty, however, may be removed if we help the core expression to vanish for  $R$  outside the coarse scale range of values. We multiply it by a top hat function which is zero outside the range  $a(N - 2k) - \frac{Np}{2} < R < a(N - 2k) + \frac{Np}{2}$  and unity within. This function is

$$\Theta(R, k) = H(R - F(k, N))H(F(k, 0) - R) \quad (5.75)$$

and hence the expression for  $Z_q(R)$  may be revised to,

$$Z_q(R) = \frac{A^N}{2^{N+1}p^N} \sum_{k=0}^N \sum_{k'=0}^N \binom{N}{k} \binom{N}{k'} (-1)^{k'} \frac{\eta^{N-1}}{(N-1)!} \text{sgn}(\eta) \Theta(R, k) \quad (5.76)$$

A plot of the partition function using Eq. 5.76 can be seen in Fig. 5.7. For the specific cases where  $N = 8$  and  $N = 10$  we can see that Fig. 5.7e and Fig. 5.7f show stable results. A detailed breakdown of the modified partition function for  $N = 2$

is shown in Table. 5.2.

$R$	$\binom{N}{k} \binom{N}{k'} (-1)^{k'} \frac{\eta^{N-1}}{(N-1)!} \text{sgn}(\eta)$	$\sum_{k,k'}^N \binom{N}{k} \binom{N}{k'} (-1)^{k'} \frac{\eta^{N-1}}{(N-1)!} \text{sgn}(\eta)$
-2	$\{4.2, -8, 3.8, 4.4, -8, 3.6\}, 0.2, 0, 0.2$	0.4
-1	$\{3.2, -6, 2.8, 2.4, -4, 1.6, 0.8, -2, 1.2\}$	0
0	$\{2.2, -4, 1.8\}, 0.4, 0, 0.4, \{1.8, -4, 2.2\}$	0.8
1	$\{1.2, -2, 0.8, 1.6, -4, 2.4, 2.8, -6, 3.2\}$	0
2	$0.2, 0, 0.2, \{3.6, -8, 4.4, 3.8, -8, 4.2\}$	0.4

Table 5.1: A breakdown of the partition function Eq. 5.68 for  $N=2$ . The relevant summations in the curly brackets are the cancellations that occur in the partition to give a desired result. For small  $N$  these cancellations are simple which become more difficult for larger  $N$ . This is seen in Fig. 5.6f.

$R$	$\binom{N}{k} \binom{N}{k'} (-1)^{k'} \frac{\eta^{N-1}}{(N-1)!} \text{sgn}(\eta) \Theta(R, k)$	$\sum_{k,k'}^N \binom{N}{k} \binom{N}{k'} (-1)^{k'} \frac{\eta^{N-1}}{(N-1)!} \text{sgn}(\eta) \Theta(R, k)$
-2	$\{0, 0, 0, 0, 0, 0\}, 0.2, 0, 0.2$	0.4
-1	$\{0, 0, 0, 0, 0, 0, 0, 0, 0\}$	0
0	$\{0, 0, 0\}, 0.4, 0, 0.4, \{0, 0, 0\}$	0.8
1	$\{0, 0, 0, 0, 0, 0, 0, 0, 0\}$	0
2	$0.2, 0, 0.2, \{0, 0, 0, 0, 0, 0\}$	0.4

Table 5.2: A breakdown of the modified partition function Eq. 5.76 for  $N=2$ . The values that contributed to the cancellations in the partition function Eq. 5.68 have been set to zero by the function  $\Theta$ . Only values that contribute to the overall partition function are included in the summations. This works for all  $N$ .

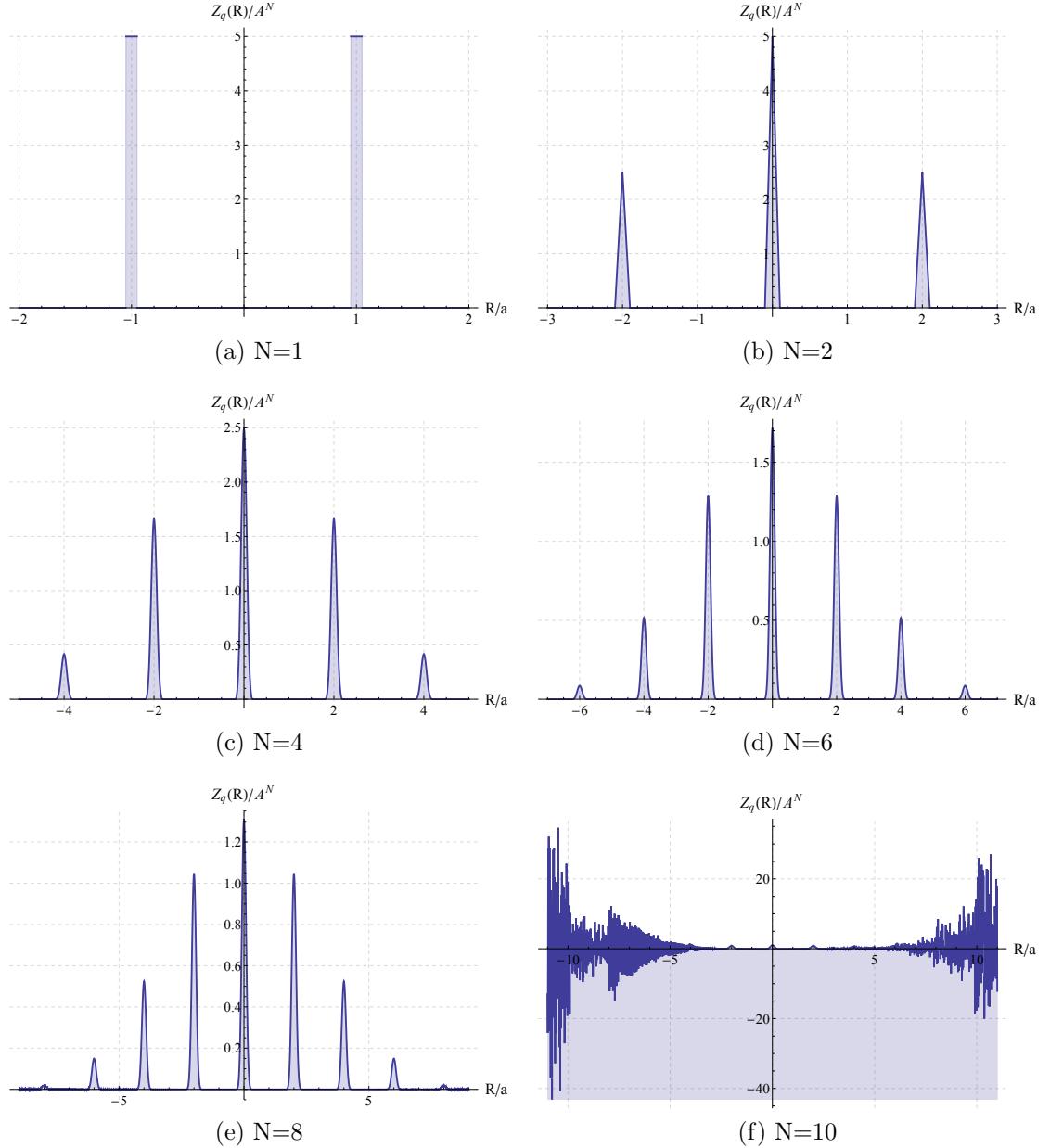


Figure 5.6: Plots showing the partition function for an extendable FJC with  $p = 0.1a$ . The original partition function used to plot these graphs Eq. 5.68. Notice the breakdown in the calculations for  $N = 10$ .

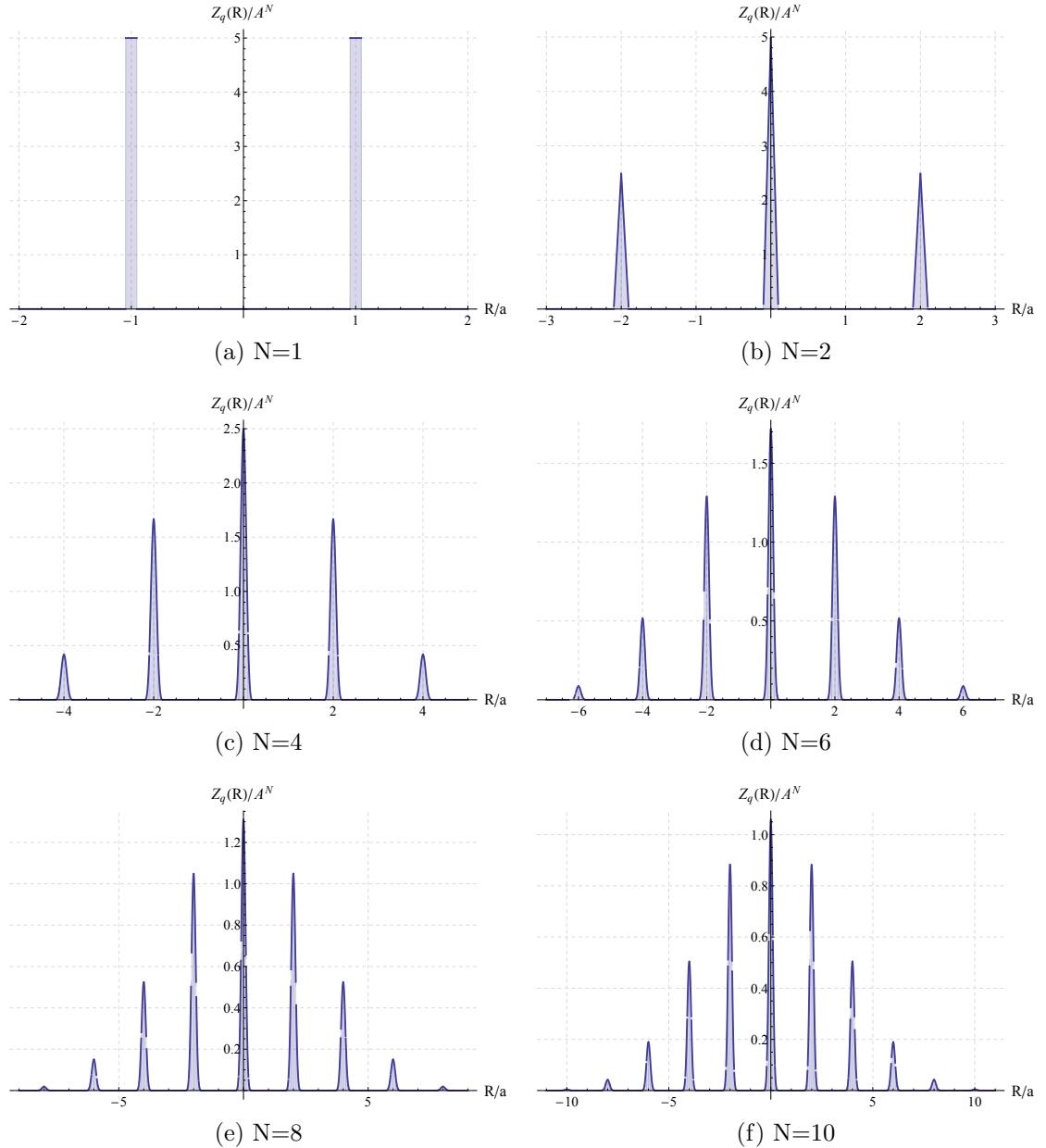


Figure 5.7: Plots showing the partition function for an extendable FJC with  $p = 0.1a$ . The partition function used to plot these graphs has been modified with the insertion of the top-hat function  $\Theta(R, k)$ . Notice that the failure to cancel terms in the expression has been avoided and the calculation for  $N=10$  converges.

## 5.4 Extendable Freely Jointed Chain in Three Dimensions

Evaluating the partition function for a one dimensional EFJC gave an expression for  $Z_q(R)$  which with slight adjustment worked well to give a probability density distribution for the end-to-end separation of the EFJC. In the small approximation of  $p$ , we saw that the behaviour of  $Z_q$  tended towards that of a non-extendable FJC, consistent with the treatment of the FJC as a one dimensional random walk. Modelling the FJC as a stochastic process we were able to determine an expectation of  $R$ . The results were consistent with each other. The next stage is to take a 3-D realisation of the EFJC model, allowing each link to move independently in three spatial dimensions, employing a finite potential to characterise the links.

### 5.4.1 Construction of the Partition Function

In three dimensional space the current partition function integral is inadequate as it has been constructed in the Cartesian co-ordinate system. A better method would be to employ spherical co-ordinates where one end of the EFJC is fixed at the origin. We have

$$Z_q = \int \prod_{i=1}^N d^3 \mathbf{r}_i \exp \left( -\frac{\Phi(\{\mathbf{r}_i\})}{kT} \right) \quad (5.77)$$

Each link in the 3D EFJC will be characterised by a potential which is function of  $\mathbf{r}_{i,i-1} = \mathbf{r}_i - \mathbf{r}_{i-1}$ , with  $\mathbf{r}_0 = 0$ , which allows it to extend or contract by a finite length  $\frac{p}{2}$ . This allows the potential,  $\Phi$ , to separate such that

$$\Phi(\{\mathbf{r}_i\}) = \sum_{i=1}^N \phi(\mathbf{r}_{i,i-1}) \quad (5.78)$$

Inserting the constraint  $\mathbf{r}' = \sum_{i=1}^N \mathbf{r}_{i,i-1}$  through a delta function, we fix the total vector sum of all the particle separations. We then define

$$Z_q(\mathbf{r}') = \int \prod_{i=1}^N d^3 \mathbf{r}_{i,i-1} \delta^3 \left( \mathbf{r}' - \sum_{i=1}^N \mathbf{r}_{i,i-1} \right) \prod_{i=1}^N \exp \left( -\frac{\phi(\mathbf{r}_{i,i-1})}{kT} \right) \quad (5.79)$$

such that  $Z_q = \int Z_q(\mathbf{r}') d^3 \mathbf{r}'$ .

Intuitively, we expect  $Z_q(\mathbf{r}')$  to depend on  $R = |\mathbf{r}'|$  only. It counts the (weighted) numbers of configurations with polymer vector length  $\mathbf{r}'$ , and the orientation of  $\mathbf{r}'$

does not alter this count. The probability distribution of the length  $R$  is actually

$$P_N(R) = \frac{R^2 Z_q(R)}{\int R^2 Z_q(R) dR} \quad (5.80)$$

which is correctly normalised such that  $\int_0^\infty P_N(R) dR = 1$ . Using the integral representation of the delta function constraint as

$$\delta\left(\mathbf{r}' - \sum_{i=1}^N \mathbf{r}_{i,i-1}\right) = \frac{1}{(2\pi)^3} \int d^3\mathbf{k} \exp\left(-i\mathbf{k} \cdot \left(\mathbf{r}' - \sum_{i=1}^N \mathbf{r}_{i,i-1}\right)\right) \quad (5.81)$$

and inserting it into Eq. 5.79 we obtain

$$Z_q(\mathbf{r}') = \frac{1}{(2\pi)^3} \int d^3\mathbf{k} \exp(-i\mathbf{k} \cdot \mathbf{r}') \left( \int_{-\infty}^{\infty} d^3\mathbf{r}_{i,i-1} \exp\left(-\frac{\phi(\mathbf{r}_{i,i-1})}{kT}\right) \exp(i\mathbf{k} \cdot \mathbf{r}_{i,i-1}) \right)^N \quad (5.82)$$

Representing the Fourier transform of the weighting function as  $F(\mathbf{k})$ , we can simplify Eq. 5.82 so that it becomes a function of  $R$ ,

$$Z_q(R) = \frac{1}{(2\pi)^3} \int_0^{\infty} k^2 \left( 2\pi \int_{-1}^1 e^{-ikR \cos \theta} d(\cos \theta) \right) dk F^N(\mathbf{k}) \quad (5.83)$$

$$= \frac{1}{(2\pi)^2} \int_0^{\infty} dk \frac{k}{R} \sin kR F^N(\mathbf{k}) \quad (5.84)$$

where the Fourier transform integral  $F(\mathbf{k})$  is

$$F(\mathbf{k}) = \int_{-\infty}^{\infty} d^3\mathbf{r}_{i,i-1} \exp\left(-\frac{\phi(\mathbf{r}_{i,i-1})}{kT}\right) \exp(i\mathbf{k} \cdot \mathbf{r}_{i,i-1}) \quad (5.85)$$

Using this general expression for  $Z_q(R)$  any potential may be studied.

### 5.4.2 Evaluating $Z_q$ for the non-extendable Freely Jointed Chain

We previously calculated the partition function for a FJC in one dimension using two delta functions as the weighting factor. Performing the calculation in three dimensions the analogous potential as a function of  $r = |\mathbf{r}_{i,i-1}|$  is

$$\phi(\mathbf{r}_{i,i-1}) = -kT \ln(D\delta(r - a)) \quad (5.86)$$

Where  $D$  is a constant with dimensions of length. Inserting Eq. 5.86 into Eq. 5.85 we get

$$\begin{aligned} F(\mathbf{k}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d^3\mathbf{r} D\delta(r-a) \exp(-i\mathbf{k}\cdot\mathbf{r}) \\ &= D \int_{-1}^1 2\pi d(\cos\theta) \int_0^{\infty} \delta(r-a) r^2 \exp(-ikr \cos\theta) dr \\ &= D \int_0^{\infty} \delta(r-a) \frac{4\pi r}{k} \sin kr dr \end{aligned} \quad (5.87)$$

Such that the Fourier transform of the exponentiated potential for a FJC becomes

$$F(k) = \left( \frac{4\pi a D}{k} \right) \sin ka \quad (5.88)$$

Inserting Eq. 5.88 into Eq. 5.84 the partition function integral becomes

$$Z_q(R) = \frac{(4\pi a D)^N}{4\pi^2 R} \int_0^{\infty} \frac{\sin kR \sin^N ka}{k^{N-1}} dk \quad (5.89)$$

and using Eq. 5.64 to express  $\sin^N ka$  as a sum of exponentials and  $\sin kR$  as the real part of  $ie^{-ikR}$  the partition function becomes

$$Z_q(R) = \frac{(4\pi a D)^N i}{4\pi^2 R (2i)^N} \sum_{l=0}^N \binom{N}{l} (-1)^l \int_0^{\infty} \frac{e^{ik(a(N-2l)-R)}}{k^{N-1}} dk \quad (5.90)$$

Since the integral has already been calculated in Eq. 5.66, the solution for a FJC in 3-D becomes

$$Z_q(R) = \frac{(2\pi a D)^N}{4\pi R (N-2)!} \sum_{l=0}^N \binom{N}{l} (-1)^l [a(N-2l)-R]^{N-2} \operatorname{sgn}(a(N-2l)-R) \quad (5.91)$$

### 5.4.3 The Shell Potential and its Fourier Transform

In the 1-D model we saw that a double top-hat weighting function was able to generate the partition function for an EFJC. In three dimensions, due to the rotational symmetry the potential is only dependent on the  $r$  co-ordinate. A potential that is non-zero only within the range of  $r$  at  $a - \frac{p}{2} \leq r \leq a + \frac{p}{2}$  gives a weighting function that can be viewed as a shell with a thickness of length  $p$ .

Working with the partition function, let us insert  $\exp(-\phi(r)/kT) = C$  for  $a - \frac{p}{2} \leq r \leq a + \frac{p}{2}$ , and zero elsewhere,  $\exp(-\phi(r)/kT) = 0$ , where  $C$  is a dimensionless

constant. Contributions to the partition function arise only where the potential is finite so the exponentiated potential Fourier transform integral simplifies to

$$F(k) = C \int_{\cos \theta = -1}^1 2\pi d(\cos \theta) \int_{a-\frac{p}{2}}^{a+\frac{p}{2}} r^2 dr e^{ikr \cos \theta} \quad (5.92)$$

This result was obtained by using  $d^3\mathbf{r} = d\phi \sin \theta d\theta r^2 dr$  and rewriting  $\sin \theta d\theta$  as  $-d(\cos \theta)$ . Performing the integral with respect to  $r$  in the potential range we have the final result

$$F(k) = \frac{4\pi C}{k^2} \left( \frac{2 \cos ka \sin \frac{kp}{2}}{k} + 2a \sin ka \sin \frac{kp}{2} - p \cos ka \cos \frac{kp}{2} \right) \quad (5.93)$$

With this expression for  $F(k)$  the final partition function for a EFJC can be analytical evaluated in 3-D.

#### 5.4.4 Evaluation of the Partition Function by the small $p$ approximation

In the small  $p$  approximation we once again insert Taylor expansions of all the trigonometric functions in the Fourier transform potential with respect to  $p$ . For small  $p$  in Eq. 5.93 we have a much simpler equation for  $F(k)$

$$F(k) = C \left( \frac{4\pi ap}{k} \right) \sin ka \quad (5.94)$$

Using Eq. 5.94 with Eq. 5.84 we get

$$Z_q(R) = \frac{(4\pi apC)^N}{2\pi^2 R} \int_{-\infty}^{\infty} \frac{\sin kR}{k^{N-1}} \sin^N ka dk \quad (5.95)$$

Which evaluates to give

$$Z_q(R) = \frac{(2\pi apC)^N}{2\pi R(N-2)!} \sum_{l=0}^N \binom{N}{l} (-1)^l [a(N-2l) - R]^{N-2} \operatorname{sgn}(a(N-2l) - R) \quad (5.96)$$

From the plot in Fig. 5.8 we see that in the large  $N$  limit the partition function with constrained end-to-end separation  $R$  gives a distribution that peaks about the origin. This is expected for the behaviour of a FJC. Fig. 5.9 shows the probability density function of the end-to-end separation.

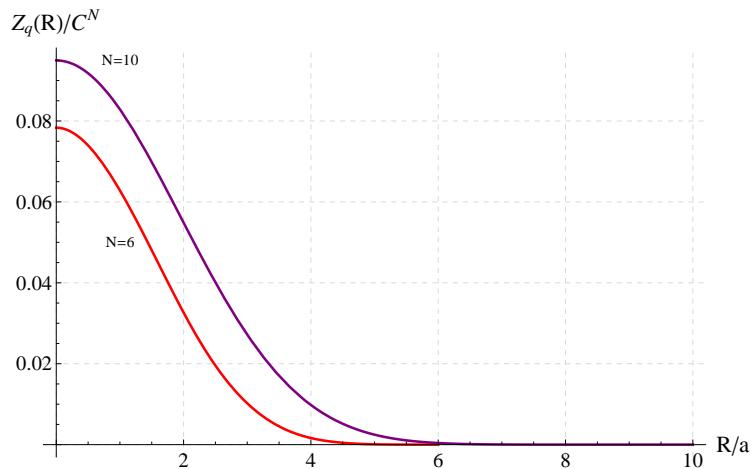


Figure 5.8: A plot of Eq. 5.91 for  $N = 6, 10$ .

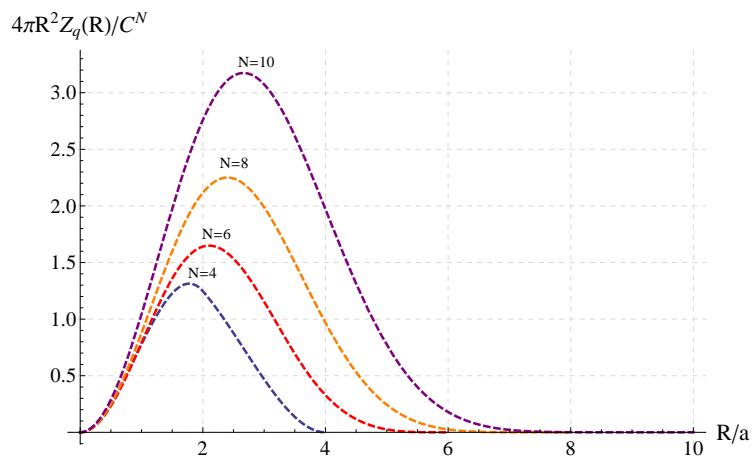


Figure 5.9: A plot of Eq. 5.91 multiplied by  $4\pi R^2$  for  $N = 4, 6, 8, 10$ .

### 5.4.5 Evaluating the Partition Function for arbitrary $p$

In the previous section we saw that when the small  $p$  approximation was applied to the EFJC in one dimension the resulting partition function became a description of the FJC, a similar case was seen when the small approximation was made in three dimensions. Let us now evaluate the partition function for the EFJC without approximation. We have the Fourier transform as an exact expression, and to combine the various terms into a single complex integral we can first use the binomial theorem to collect the terms which have different factors of  $\frac{1}{k}$

$$F^N(k) = (2\pi C)^N \sum_{\alpha=0}^N \binom{N}{\alpha} \frac{1}{i^\alpha} \frac{1}{k^{2N+\alpha}} X^\alpha Y^{N-\alpha} \quad (5.97)$$

where

$$X = e^{ik(a+\frac{p}{2})} - e^{ik(a-\frac{p}{2})} + e^{ik(-a+\frac{p}{2})} - e^{ik(-a-\frac{p}{2})} \quad (5.98)$$

and

$$Y = \left(-a - \frac{p}{2}\right) e^{ik(a+\frac{p}{2})} - \left(-a + \frac{p}{2}\right) e^{ik(a-\frac{p}{2})} + \left(a - \frac{p}{2}\right) e^{ik(-a+\frac{p}{2})} - \left(a + \frac{p}{2}\right) e^{ik(-a-\frac{p}{2})} \quad (5.99)$$

The multinomial theorem is then used for the  $X^\alpha$  and  $Y^{N-\alpha}$  terms. This approach reduces the terms to a single exponent as a function of  $k$  such that the integral over  $dk$  is simplified. Combining this result with Eq. 5.84 we get a final expression for  $Z_q$  as

$$Z_q(R) = \frac{\pi (-1)^N (2\pi)^{N-2} C^N}{R} \sum_{\alpha=0}^N \binom{N}{\alpha} \frac{1}{(2N+\alpha-2)!} \sum_{j_1, j_2, j_3, j_4}^{\alpha} \sum_{l_1, l_2, l_3, l_4}^{N-\alpha} I(N, \alpha, a, p, R, \{j\}, \{l\}) \quad (5.100)$$

with

$$I(N, \alpha, a, p, R, \{j\}, \{l\}) = \binom{N}{j_1, j_2, j_3, j_4} \binom{N-\alpha}{l_1, l_2, l_3, l_4} (-1)^{j_2+j_4} \left(-a - \frac{p}{2}\right)^{l_1+l_4} \times \\ \left(a - \frac{p}{2}\right)^{l_2+l_3} (J(a, p, \{j\}) + L(a, p, \{l\}) - R)^{2N+\alpha-2} \operatorname{sgn}(J(a, p, \{j\}) + L(a, p, \{l\}) - R) \quad (5.101)$$

and

$$\begin{aligned}\binom{N}{j_1, j_2, j_3, j_4} &= \frac{N!}{j_1! j_2! j_3! j_4!} \\ \binom{N - \alpha}{l_1, l_2, l_3, l_4} &= \frac{(N - \alpha)!}{l_1! l_2! l_3! l_4!}\end{aligned}$$

$$\begin{aligned}L(a, p, \{l\}) &= \left(a + \frac{p}{2}\right)(l_1 - l_4) + \left(a - \frac{p}{2}\right)(l_2 - l_3) \\ J(a, p, \{j\}) &= \left(a + \frac{p}{2}\right)(j_1 - j_4) + \left(a - \frac{p}{2}\right)(j_2 - j_3)\end{aligned}$$

$$\begin{aligned}\sum_{k=1}^4 j_k &= N \\ \sum_{k=1}^4 l_k &= N - \alpha\end{aligned}$$

A plot of Eq. 5.100 in Fig. 5.10 shows the probability distributions for  $N = 3, 4, 5, 6$ . As  $N$  increases in Fig. 5.11 we see that the distributions increasingly centre around  $r = 0$ . Fig. 5.11e show results for large  $N$ , where we begin to see that the results become unstable just like in Fig. 5.6e.

$$4\pi R^2 Z_q(R)/C^N$$

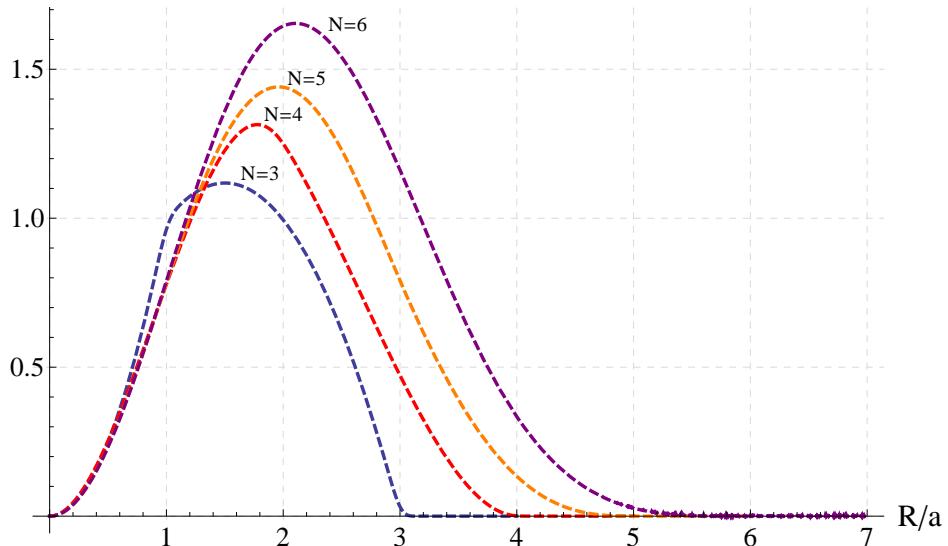


Figure 5.10: A plot of normalised partition functions for an extendable FJC in three dimensions. The plot shows distributions for  $N = 3, 4, 5, 6$  with  $p = 0.1a$ .

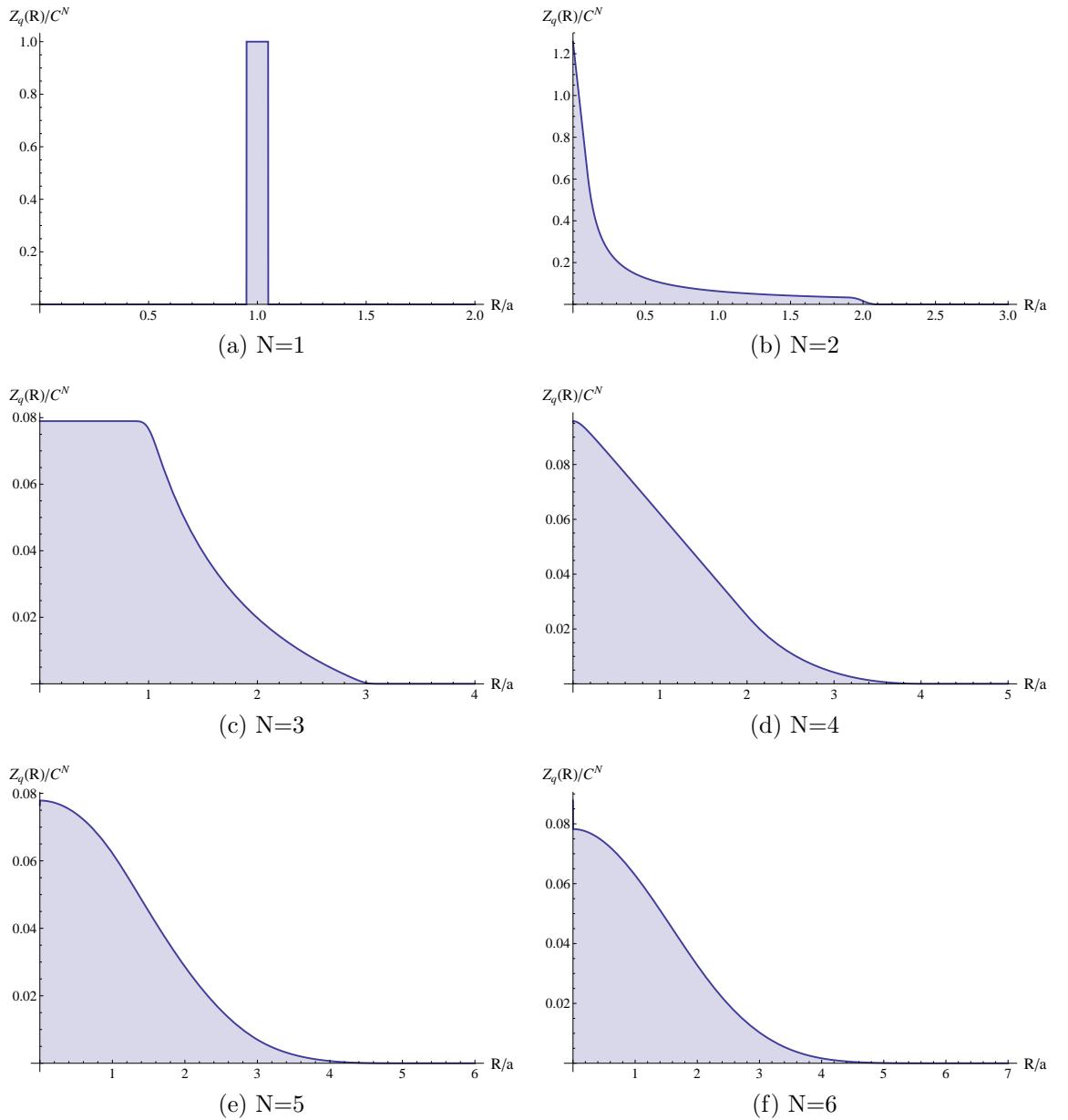


Figure 5.11: Plots showing the partition function for an extendable FJC in three dimensions with  $p = 0.1a$ .

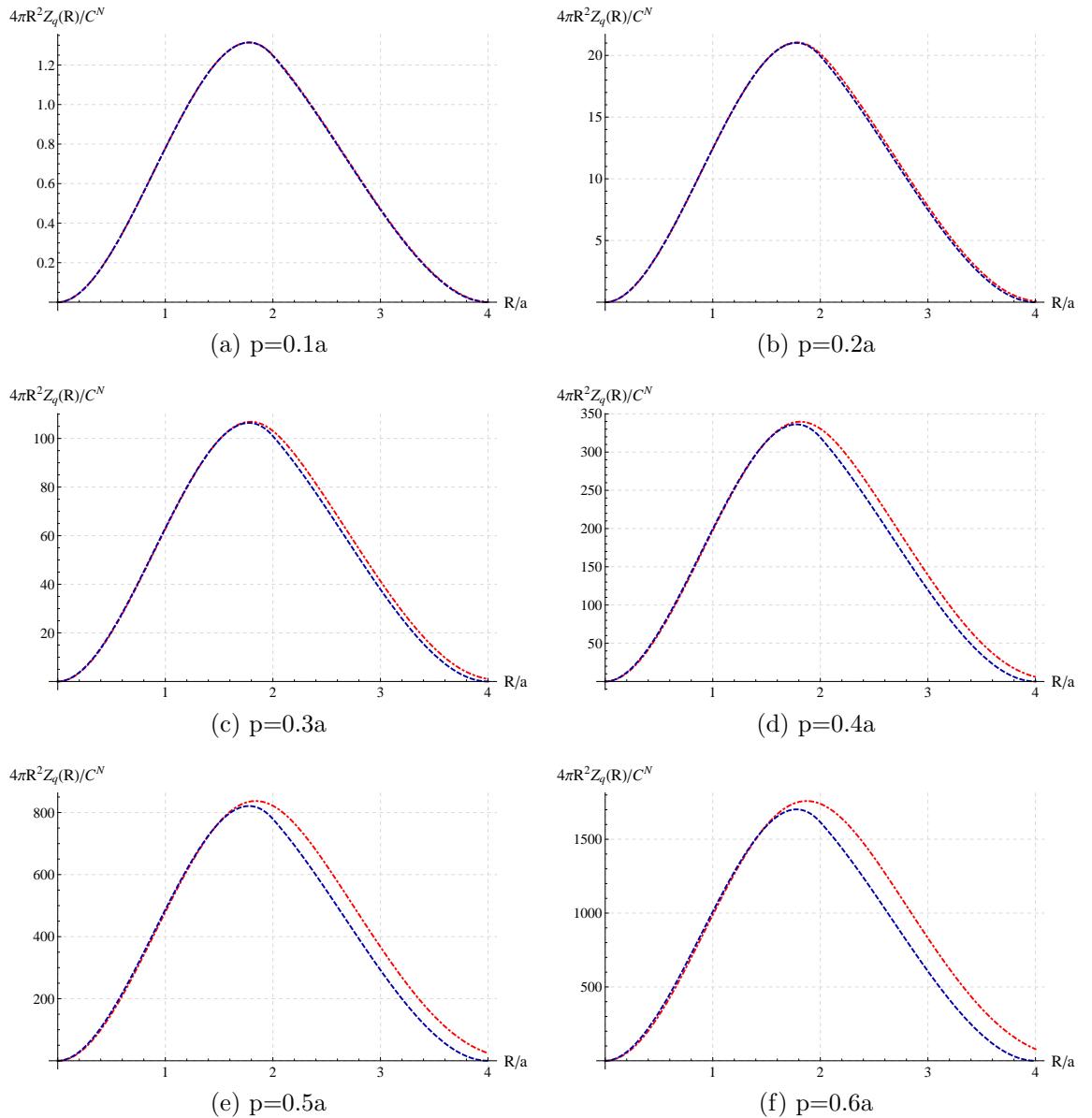


Figure 5.12: These plots show a comparison of the probability distribution for the FJC Eq. 5.96 (blue curve) and the EFJC Eq. 5.100 (red curve) for different values of  $p$  with  $N = 4$ .

## 5.5 Conclusion

Using the Fourier Transform Integral method we have presented a model which gives a partition function to describe an Extendable Freely Jointed Chain as an improvement to the non-extendable Freely Jointed Chain. The EFJC improves the standard FJC polymer model by allowing each link in the chain to extend and contract independently to emulate the mechanical behaviour of intermolecular bonds within polymers.

In one dimension, the partition function for EFJC was calculated using two top-hat functions as the weighting function in the Fourier Transform integral. This potential allowed each link to move freely in one dimension that included a weighting function to provide linear extensibility. Taking a small approximation of  $p$  in the EFJC model transformed the partition function to that of a FJC which was calculated using delta functions as the weighting functions.

In Fig. 5.12 we see a plot which compares the results of  $Z_q$  with  $N=4$  for the FJC and the EFJC with different  $p$ . It is interesting to note that for extensions where the range of  $p$  is less than  $0.2a$  the results are almost identical. Above this range we see that the partition function distribution becomes larger in the EFJC model.

With the negligible differences in the partition function distributions for small  $p$  we can conclude that results from the EFJC can be well represented by those from the FJC.

# Chapter 6

## DNA Model

The complex structure of DNA allows external forces to be applied in various ways. In cells, DNA is constantly twisted, bent and stretched by numerous proteins for biological processes and therefore to create a mathematical model that encompasses all these actions would be extremely difficult.

Our DNA model focuses on a shearing force applied to DNA. It specifically looks at the behaviour of the hydrogen bonding between the base pairs under strain and evaluates the free energy of different breakage patterns. The statistical model is in the form of a partition function Eq. 1.1 where the multi-dimensional integral over phase-space is evaluated by the use of the Transfer Matrix Method. We begin by analysing the Hamiltonian for our DNA model.

### 6.1 Construction of the Hamiltonian

To form a tractable mathematical model using the transfer matrix method we simplify the double helix structure to a one dimensional ladder structure allowing us to introduce a suitable Hamiltonian. This basic structure is similar to the one first used by de Gennes [75]. The ball and stick diagram of the ladder structure consists of two rows of regularly spaced particles connected by elastic springs Fig. 6.1. The springs along the two rows which represent the backbone strands in DNA have a spring constant  $\kappa$ , and the springs perpendicular to the backbones, representing the base pairs, are harmonic for small extensions with a spring constant  $\kappa'$ . In this model we will neglect the torsional effects in DNA. The load on the ladder structure will be applied by keeping one end of one of the backbones fixed while the other end of the other backbone is extended axially by a distance  $u$ .

Each particle in the ladder structure represents a nucleotide. Each stick con-

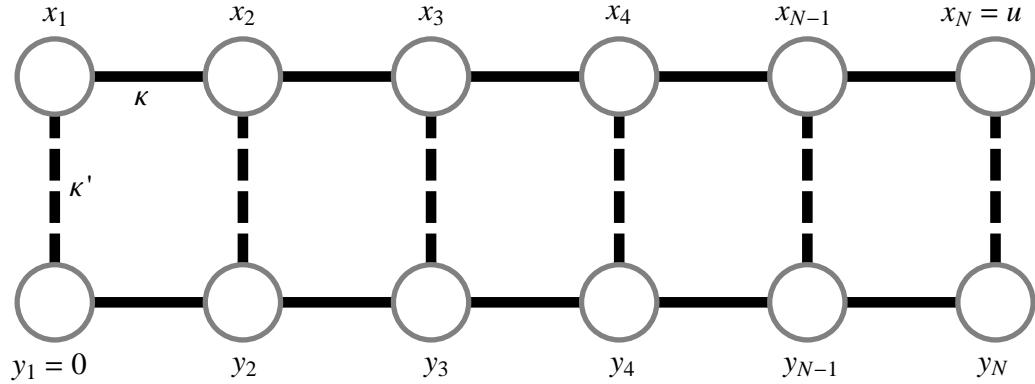


Figure 6.1: Schematic diagram of the two dimensional ladder model for DNA.

necting the base pairs represents the 2-3 hydrogen bonds, and each stick along the backbone represents a peptide bond connecting the nucleotides. The axial displacements of particles in each backbone are labelled  $x_i$  and  $y_i$ , and the partners forming the  $i^{th}$  base pair interact through a potential  $V_i$  which is dependent on the axial pair separation  $(x_i - y_i)$ . For a model that has  $N$  base pairs we can write the Hamiltonian as

$$H = H_x + H_y + H_{bp} \quad (6.1)$$

where

$$H_x = \sum_{i=1}^{N-1} \frac{1}{2} \kappa (x_{i+1} - x_i)^2 \quad (6.2)$$

$$H_y = \sum_{i=1}^{N-1} \frac{1}{2} \kappa (y_{i+1} - y_i)^2 \quad (6.3)$$

$$H_{bp} = \sum_{i=1}^N V_i (x_i - y_i) \quad (6.4)$$

To eliminate the cross terms in the Hamiltonian we change the co-ordinate system of variables  $x_i$  and  $y_i$  to convert the Hamiltonian to a solvable linear system. The following substitutions transform the system to a new co-ordinate system in  $\xi$  and  $\eta$ ,

$$\xi_i = x_i + y_i \quad (6.5)$$

$$\eta_i = x_i - y_i \quad (6.6)$$

The new Hamiltonian is now a function of  $\xi$  and  $\eta$

$$H(\xi, \eta) = H_\xi + H_\eta \quad (6.7)$$

where

$$H_\xi = \sum_{i=1}^{N-1} \frac{\kappa}{4} (\xi_{i+1} - \xi_i)^2 \quad (6.8)$$

$$H_\eta = \sum_{i=1}^{N-1} \frac{\kappa}{4} (\eta_{i+1} - \eta_i)^2 + \sum_{i=1}^N V(\eta_i) \quad (6.9)$$

The base pairs will experience a restoring force brought about by the extension  $u$  but when the bond separation goes beyond  $\eta_B$  the force is set to zero, to begin with, signifying broken bonds as illustrated in Fig. 6.3.

$$\begin{aligned} V_i(\eta_i) &= \frac{1}{2} \kappa' \eta_i^2 && \text{when } \eta_i \leq \eta_B \\ &= \frac{1}{2} \kappa' \eta_B^2 && \text{when } \eta_i > \eta_B \end{aligned} \quad (6.10)$$

Variations on Eq. 6.10 can have the base pairs experiencing either a stronger or weaker restoring force after a certain separation  $\eta_B$ , for a given extension.

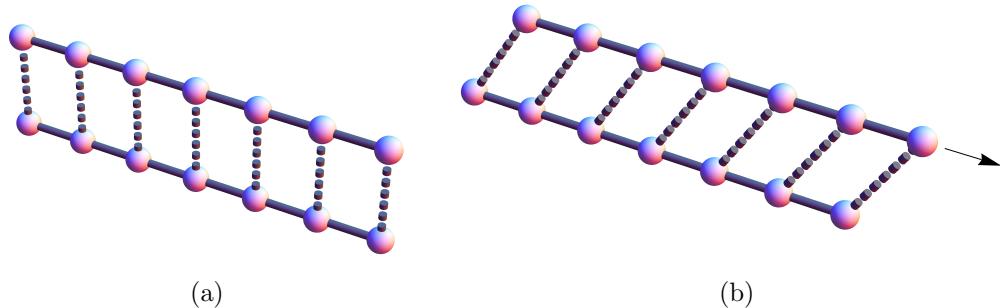


Figure 6.2: A three-dimensional illustration of the DNA ladder model. In Fig. 6.2a we have an intact ladder structure that hasn't been influenced by axial extension. When an extension is applied to the ladder structure from one end as illustrated in Fig. 6.2b the base pairs tilt towards the displacement end, experiencing a restoring force defined by the base pair potential.

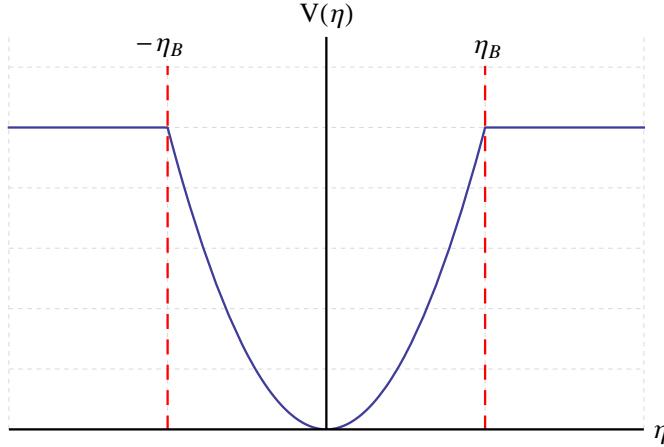


Figure 6.3: A plot of the base pair potential as described by Eq. 6.10.

## 6.2 Partition Function Integral

The separation of variables in the Hamiltonian allows us to evaluate the partition function for various breakage patterns. We can neglect the momentum part of the Hamiltonian as we already shown in the EFJC that this only contributes a constant factor. The partition function integral is evaluated over phase-space  $d\tau = d\xi d\eta$  per base pair. For a  $N$  base pair system we have a multi-dimensional partition function integral

$$Z = \int \prod_{i=1}^N d\xi_i d\eta_i \exp(-\beta H(\xi_i, \eta_i)) \quad (6.11)$$

Where  $\beta = \frac{1}{k_B T}$ . We simplify the Hamiltonian Eq. 6.7 further by converting  $\xi$  and  $\eta$  into dimensionless quantities such that  $\xi \rightarrow \xi \left(\frac{\kappa\beta}{4}\right)^{-\frac{1}{2}}$ , and  $\eta \rightarrow \eta \left(\frac{\kappa\beta}{4}\right)^{-\frac{1}{2}}$ . The Hamiltonian becomes

$$\beta H = \sum_{i=1}^{N-1} (\xi_{i+1} - \xi_i)^2 + \sum_{i=1}^{N-1} (\eta_{i+1} - \eta_i)^2 + \sum_{i=1}^N V(\eta_i) \quad (6.12)$$

and the base pair potential Eq. 6.10 becomes a function of dimensionless  $\eta$

$$\begin{aligned} V(\eta) &= \frac{2\kappa'}{\kappa} \eta^2 && \text{when } \eta \leq \eta_B \\ &= \frac{2\kappa'}{\kappa} \eta_B^2 && \text{when } \eta > \eta_B \end{aligned} \quad (6.13)$$

Applying the transformation to the partition function Eq. 6.11 creates a constant  $\left(\frac{\kappa\beta}{4}\right)^N$  in front of the integral which is ignored from the calculation. Like the mo-

mentum part of the partition function integral it doesn't affect the behaviour of the free energy curves, and therefore has no impact on the force-extension calculations.

### 6.3 Transfer Matrices

To express  $Z$  in terms of transfer matrices we focus our attention on the exponential term in the partition function. We insert Eq. 6.12 into Eq. 6.11 and make an expansion of the  $\exp(-\beta H)$  term. We are then able to group like terms together and adjust the exponentials such that they are symmetric in both  $\xi$  and  $\eta$ :

$$\begin{aligned} \exp(-\beta H) &= \prod_{i=1}^{N-1} \exp(-(\xi_{i+1} - \xi_i)^2) \\ &\times \prod_{j=1}^{N-1} \exp\left(-( \eta_{j+1} - \eta_j)^2 + \left(\frac{V(\eta_{i+1}) + V(\eta_i)}{2}\right)\right) \\ &\times \exp\left(-\frac{V(\eta_1) + V(\eta_N)}{2}\right) \end{aligned} \quad (6.14)$$

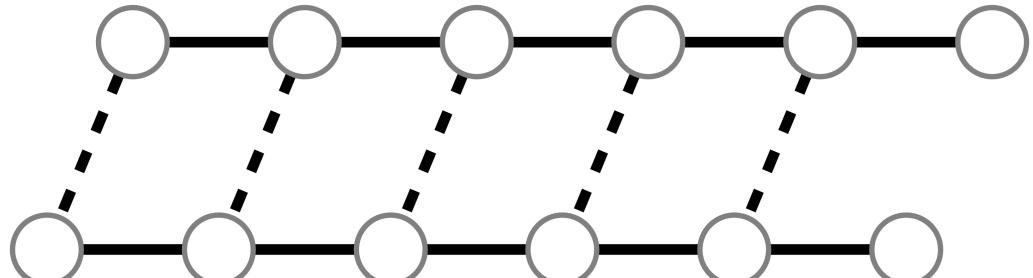
We can now introduce a set of transfer matrices for both  $\xi$  and  $\eta$

$$T(\xi_{i+1}, \xi_i) = \exp(-(\xi_{i+1} - \xi_i)^2) \quad (6.15)$$

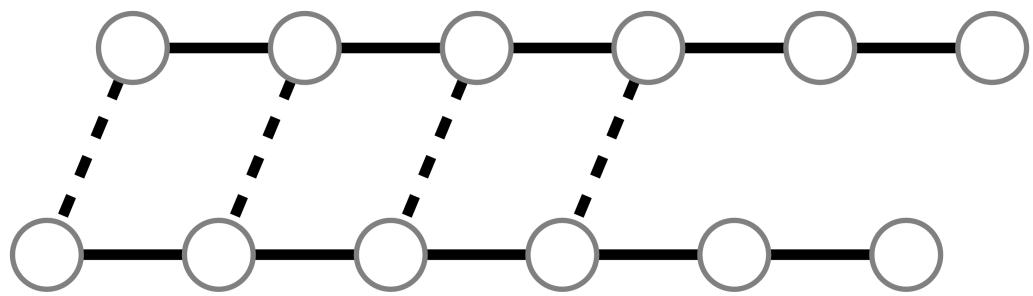
$$\hat{T}(\eta_{j+1}, \eta_j) = \exp\left(-(\eta_{j+1} - \eta_j)^2 - \left(\frac{V(\eta_{i+1}) + V(\eta_i)}{2}\right)\right) \quad (6.16)$$

When we talk about the breakage patterns in DNA we are actually referring to patterns of broken base pairs. Two types of breakage patterns occur in DNA, fraying and bubbling. Fraying occurs when the base pairs break from the ends of the DNA molecule Fig. 6.4, and bubbling occurs when base pairs break from within the DNA molecule Fig. 6.5. To calculate the free energies of various breakage patterns we need to incorporate these breakage patterns into the DNA ladder model by a modification of the transfer matrices. We can do this by introducing a bond breakage criterion embodied in a  $g$  factor,  $g_\mu$ . Here the label  $\mu$  indicates whether a base pair is either intact ( $\mu = 0$ ), or broken ( $\mu = 1$ ). Since these breakage patterns only apply to the base pairs we can insert the following factors as a function of  $\eta$  into the transfer matrices,  $\hat{T}$ . For an intact base pair

$$\begin{aligned} g_0(\eta) &= 1 \quad \text{if} \quad |\eta| \leq \eta_B \\ &= 0 \quad \text{Otherwise} \end{aligned} \quad (6.17)$$

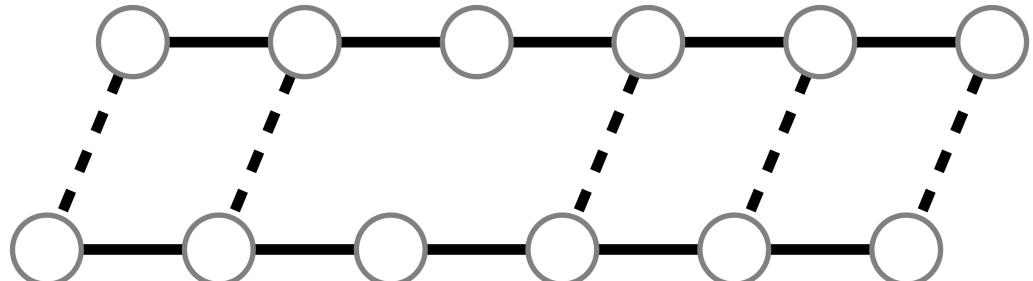


$$(a) \{\mu\} = \{0, 0, 0, 0, 0, 1\}$$

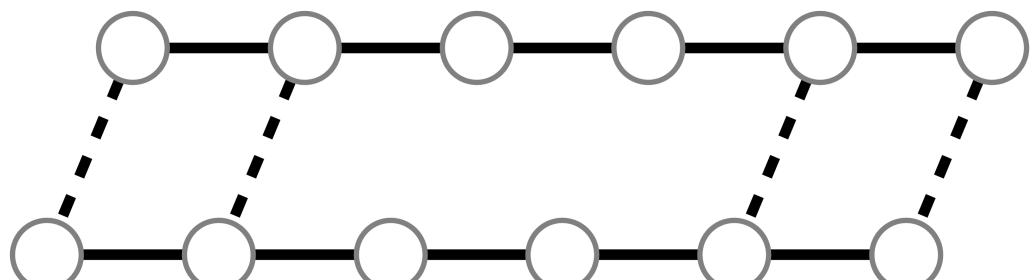


$$(b) \{\mu\} = \{0, 0, 0, 0, 1, 1\}$$

Figure 6.4: Different frayed states in DNA for  $N = 6$ .



$$(a) \{\mu\} = \{0, 0, 1, 0, 0, 0\}$$



$$(b) \{\mu\} = \{0, 0, 1, 1, 0, 0\}$$

Figure 6.5: Different bubbled states in DNA for  $N = 6$ .

For a broken base pair the  $g$  factor becomes,

$$\begin{aligned} g_1(\eta) &= 1 \quad \text{if} \quad |\eta| > \eta_B \\ &= 0 \quad \text{Otherwise} \end{aligned} \quad (6.18)$$

Inserting these factors makes the following transformation of the transfer matrix,  $\hat{T} \rightarrow \hat{T}_{\mu_i, \mu_{i+1}}$ , such that

$$\hat{T}_{\mu_i, \mu_{i+1}}(\eta_i, \eta_{i+1}) = g_{\mu_i}(\eta_i) g_{\mu_{i+1}}(\eta_{i+1}) \exp \left( -(\eta_{i+1} - \eta_i)^2 - \left( \frac{V(\eta_{i+1}) + V(\eta_i)}{2} \right) \right) \quad (6.19)$$

The partition function integral can now be expressed as a product of transfer matrices

$$\begin{aligned} Z &= \int \prod_{i=1}^N d\xi_i \left( \prod_{i'=1}^{N-1} T(\xi_{i'+1}, \xi_{i'}) \right) \\ &\times \int \prod_{j=1}^N d\eta_j \left( \prod_{j'=1}^{N-1} \hat{T}_{\mu_{j'}, \mu_{j'+1}}(\eta_{j'}, \eta_{j'+1}) \right) \exp \left( -\frac{V(\eta_1) + V(\eta_N)}{2} \right) \end{aligned} \quad (6.20)$$

This partition function constructed out of transfer matrices can be simplified if we somehow provide eigenfunctions that allow the expression to be contracted. We next look at how these eigenfunctions arise from the boundary conditions of the system.

## 6.4 Boundary Conditions

The boundary conditions in the ladder structure are applied to the partition function through a set of delta functions: the particle at the position  $y_1$  is fixed by the insertion of  $\delta(y_1)$  such that it remains in equilibrium, and the particle at the pulled end is constrained to take the extension  $u$  from its original position  $\delta(x_N - u)$ . In the new co-ordinate system the boundary conditions are transformed by Eq. 6.8 and Eq. 6.9 to

$$\delta(y_1) \delta(x_N - u) = \delta(\xi_1 - \eta_1) \delta(\xi_N + \eta_N - 2u) \quad (6.21)$$

## 6.5 Eigenvalue Equations

The transfer integral method simplifies the partition function calculation of the DNA model to the solutions of a set of eigenvalue equations. We saw in the Ising model in chapter 3 that the transfer operators contain all the information needed

to describe how the spins were distributed and correlated with their neighbours. In the DNA model the transfer operators provides information on whether a base pair is broken or intact, and the effects of the potential at a given extension. The eigenfunctions needed to contract the transfer matrices come from the expansion of the delta functions that impose the boundary conditions.

$$\delta(\xi_1 - \eta_1) = \sum_t \hat{\psi}_t^*(\xi_1) \hat{\psi}_t(\eta_1) \quad (6.22)$$

$$\delta(\xi_N + \eta_N - 2u) = \sum_s \psi_s^*(2u - \xi_N) \psi_s(\eta_N) \quad (6.23)$$

We introduce  $\psi_s$ , eigenfunctions of the transfer matrix Eq. 6.15, and  $\hat{\psi}_t$ , the eigenfunctions of the transfer matrix Eq. 6.16. The eigenfunctions satisfy the eigenvalue equations for the backbone  $T$ , intact base pair  $\hat{T}_{00}$ , and a broken base pair  $\hat{T}_{11}$  transfer matrices by

$$\int d\xi' T(\xi, \xi') \psi_s(\xi') = \lambda_s \psi_s(\xi) \quad (6.24)$$

$$\int d\eta' \hat{T}_{00}(\eta, \eta') \hat{\psi}_t(\eta') = \hat{\lambda}_t \hat{\psi}_t(\eta) \quad (6.25)$$

$$\int d\eta' \hat{T}_{11}(\eta, \eta') \tilde{\psi}_v(\eta') = \tilde{\lambda}_v \tilde{\psi}_v(\eta) \quad (6.26)$$

respectively. Using these eigenvalue equations we are able to contract the expression for  $Z$  transfer matrices leaving a sum involving eigenvalues and eigenvectors. Numerically the eigenfunctions can be calculated if we discretise the transfer matrix; however, in certain cases analytical solutions can be found.

### 6.5.1 Analytical solutions for eigenvectors and eigenvalues of $T$

If we take a closer look at Eq. 6.15 we find that  $T$  is just a Gaussian function, and thus Eq. 6.24 can be solved analytically to give solutions for  $\psi_s$  and  $\lambda_s$ . The integral equation becomes

$$\int d\xi' \exp\left(-(\xi' - \xi)^2\right) \psi_s(\xi') = \lambda_s \psi_s(\xi) \quad (6.27)$$

By assuming the eigenfunction is of the form  $\psi_s(\xi) = A \exp(is\xi B)$  where  $A$  and  $B$  are constants the integral equation becomes

$$\int d\xi' \exp(-(\xi - \xi')^2) A \exp(is\xi' B) = \lambda_s \psi_s(\xi) \quad (6.28)$$

Then making a substitution  $\gamma = \xi - \xi'$  we get

$$-A \exp(is\xi B) \int d\gamma \exp(-\gamma^2) \exp(-is\gamma B) = \lambda_s \psi_s(\xi) \quad (6.29)$$

and we find that the integral is simply a Fourier transform of the Gaussian function

$$\int_{-\infty}^{\infty} \exp(-\alpha x^2) \exp(-ivx) dx = \sqrt{\frac{\pi}{\alpha}} \exp\left(-\frac{v^2}{4\alpha}\right) \quad (6.30)$$

Using this transformation we get

$$-A \exp(is\xi B) \left( \sqrt{\pi} \exp\left(-\frac{B^2 s^2}{4}\right) \right) = \lambda_s \psi_s(\xi) \quad (6.31)$$

Hence

$$\lambda_s = \sqrt{\pi} \exp\left(-\frac{B^2 s^2}{4}\right) \quad (6.32)$$

All that remains is to find the constant  $B$ . Here we impose periodicity such that we can represent the  $\psi_s$  eigenfunction on a finite spatial range  $L$ . At this point we also take  $s$  to be an integer label. We impose

$$\psi_s(\alpha + L) = \psi_s(\alpha) \quad (6.33)$$

such that

$$\exp(isB(\alpha + L)) = \exp(isB\alpha) \quad (6.34)$$

or

$$\exp(isBL) = 1 \quad (6.35)$$

Since  $\exp(is2\pi) = 1$ , we can substitute that on the RHS, and take the logarithmic of both sides. This result gives

$$B = \frac{2\pi}{L} \quad (6.36)$$

To find the constant  $A$  we simply use the normalisation condition. Integrating over finite range from  $-\frac{L}{2}$  to  $\frac{L}{2}$ , we get

$$\int_{-\frac{L}{2}}^{\frac{L}{2}} |\psi_s(\xi)|^2 d\xi = 1 \quad (6.37)$$

$$|A|^2 L = 1 \quad (6.38)$$

Hence

$$A = L^{-\frac{1}{2}} \quad (6.39)$$

Thus the  $T$  transfer matrix has eigenfunctions that are expressed as

$$\psi_s(\xi) = L^{-\frac{1}{2}} \exp\left(\frac{is2\pi\xi}{L}\right) \quad (6.40)$$

with corresponding eigenvalues

$$\lambda_s = \pi^{\frac{1}{2}} \exp\left(-\frac{\pi^2 s^2}{L^2}\right) \quad (6.41)$$

We are now able to make some simplifications to the delta function expansion of Eq. 6.23 where the analytical form of the eigenfunction  $\psi_s$  allows us to take out extension parameter  $u$  to become

$$\psi_s^*(2u - \xi_N) = \exp\left(-\frac{i4\pi su}{L}\right) \psi_s(\xi_N) \quad (6.42)$$

The delta function expansion of Eq. 6.23 can now be expressed as

$$\delta(\xi_N + \eta_N - 2u) = \sum_s \exp\left(-\frac{i4\pi su}{L}\right) \psi_s(\xi_N) \psi_s(\eta_N) \quad (6.43)$$

### 6.5.2 Analytical solutions for eigenvectors and eigenvalues of $\hat{T}$

The  $g$  factors in Eq. 6.19 make it inconvenient to find analytical solutions for  $\hat{\psi}_t$  and  $\hat{\lambda}_t$  except for a special case when  $\eta_b \rightarrow \infty$  for intact base pairs. In this limit the  $g$  factors become unity reducing the transfer matrix equation  $\hat{T}_{00}$  to  $\hat{T}$ , as given in Eq. 6.16. The simplification of the transfer matrix means we are able to solve the

eigenvalue equation by inserting  $\hat{T}$  into Eq. 6.25.

$$\int d\eta' \exp \left( -(\eta - \eta')^2 - \frac{\kappa'}{\kappa} (\eta^2 + \eta'^2) \right) \hat{\psi}_t(\eta') = \hat{\lambda}_t \hat{\psi}_t(\eta) \quad (6.44)$$

This integral is very similar to a Gaussian integral which therefore indicates that  $\hat{\psi}_t$  might must be a Gaussian in some form. We assume that

$$\hat{\psi}_t(\eta') = A \exp \left( -\frac{B}{2} \eta'^2 \right) \quad (6.45)$$

where A and B are unknown constants. If we collect the like terms together our eigenvalue equation becomes

$$A \int d\eta' \exp \left( -\eta^2 \left( \frac{\kappa + \kappa'}{\kappa} \right) - 2\eta\eta' - \eta'^2 \left( 1 + \frac{2\kappa' + \kappa(B+2)}{2\kappa} + \frac{B}{2} \right) \right) = \hat{\lambda}_t \hat{\psi}_t(\eta) \quad (6.46)$$

which is a Gaussian integral of the form

$$\int_{-\infty}^{\infty} k \exp(-fx^2 + gx + h) = k \sqrt{\frac{\pi}{f}} \exp \left( \frac{g^2}{4f} + h \right) \quad (6.47)$$

The integration gives us the following result

$$A \left( \frac{\pi\kappa}{\kappa + 2B\kappa'} \right)^{\frac{1}{2}} \exp \left( -\frac{D\eta^2}{2} \right) = \hat{\lambda}_t \hat{\psi}_t(\eta) \quad (6.48)$$

where

$$D = 2 + \frac{2\kappa'}{\kappa} - \frac{4\kappa}{(2+B)\kappa + 2\kappa'} \quad (6.49)$$

The function Eq. 6.45 is only an eigenfunction of  $\hat{T}$  if  $B = D$ , therefore from Eq. 6.49

$$B = \left( \frac{4\kappa'(2\kappa + \kappa')}{\kappa^2} \right)^{\frac{1}{2}} \quad (6.50)$$

and the first eigenvalue is simply

$$\hat{\lambda}_0 = \left( \frac{\pi\kappa}{\kappa + 2B\kappa'} \right)^{\frac{1}{2}} \quad (6.51)$$

This is consistent with the analysis of Parisi who calculated the first analytical eigenfunction and eigenvalue for a similar eigenvalue equation [41]. His solutions

take the form

$$\hat{\psi}_0(\eta) = \exp\left(-\frac{c}{2}\eta^2\right) \quad (6.52)$$

$$\hat{\lambda}_0 = \frac{2\pi^{\frac{1}{2}}}{[(\mu^2 + 2\mu\beta)^{\frac{1}{2}} + \mu + \beta]^{\frac{1}{2}}} \quad (6.53)$$

where  $c = (\mu^2 + 2\beta\mu)^{\frac{1}{2}}/2$ . If we substitute  $\beta = 4$  and  $\mu = 4\kappa'/\kappa$  into Eq. 6.52 and Eq. 6.53 we return the same result for the eigenvalue and eigenfunction. Following Parisi [41], we find the remaining eigenvector and eigenvalues are

$$\begin{aligned} \hat{\psi}_t(\eta) &\propto H_t(b\eta) \exp\left(-\frac{c\eta^2}{2}\right) \\ \hat{\lambda}_n &= \hat{\lambda}_0 \exp(-mn) \\ b &= \frac{1}{2} \left( \frac{\Delta^2 - \beta^2}{\Delta} \right)^{\frac{1}{2}} \\ m &= \ln\left(\frac{\Delta}{\beta}\right) \\ \Delta &= 2c + \mu + \beta \end{aligned} \quad (6.54)$$

Where  $H_t$  is a Hermite polynomial. To complete solution the we use the following orthogonality relation for Hermite polynomials

$$\int_{-\infty}^{\infty} H_m(x) H_n(x) \exp(-x^2) dx = \sqrt{\pi} 2^n n! \delta_{nm} \quad (6.55)$$

and the eigenfunctions of the  $\hat{T}$  transfer matrix become

$$\hat{\psi}_t(\eta) = \left(\frac{\pi}{c}\right)^{-\frac{1}{4}} \left(\frac{1}{2^t t!}\right)^{\frac{1}{2}} H_t(b\eta) \exp\left(-\frac{c\eta^2}{2}\right) \quad (6.56)$$

## 6.6 Transfer Matrix Method

Inserting the eigenfunction representation of the delta function into the partition function integral we not only transform the expression to a function of the extension parameter  $u$ , but also enable the contraction of many of the transfer matrices using

the eigenvalue equations. We start from

$$\begin{aligned} Z(u) &= \sum_s \exp\left(-\frac{i4\pi su}{L}\right) \psi_s(\xi_N) \psi_s(\eta_N) \int \prod_{i=1}^N d\xi_i \left( \prod_{i'=1}^{N-1} T(\xi_{i'+1}, \xi_{i'}) \right) \\ &\quad \times \sum_t \hat{\psi}_t^*(\eta_1) \hat{\psi}_t(\xi_1) \int \prod_{j=1}^N d\eta_j \left( \prod_{j'=1}^{N-1} \hat{T}_{\mu,\mu'}(\eta_{j'+1}, \eta_{j'}) \right) \exp\left(-\frac{V(\eta_1) + V(\eta_N)}{2}\right) \end{aligned} \quad (6.57)$$

The transfer matrix  $T$  has no dependence on  $\eta$  and therefore does not depend on the base pair breakage pattern, and consequently we are able to proceed with the contraction of these transfer matrices. Using  $\psi_s(\xi_N)$  with the eigenvalue equation Eq. 6.24 we propagate the eigenfunction down to  $\psi_s(\xi_1)$  leaving an eigenvalue for each contraction. The delta function Eq. 6.22 also allows us to replace  $\eta_1$  with  $\xi_1$  in the potential term to give

$$\begin{aligned} Z(u) &= \sum_{s,t} \lambda_s^{N-1} \exp\left(-\frac{i4\pi su}{L}\right) \int d\xi_1 \psi_s(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \hat{\psi}_t(\xi_1) \quad (6.58) \\ &\quad \times \int \prod_{j=1}^N d\eta_j \left( \prod_{j'=1}^{N-1} \hat{T}_{\mu,\mu'}(\eta_{j'+1}, \eta_{j'}) \right) \hat{\psi}_t^*(\eta_1) \exp\left(-\frac{V(\eta_N)}{2}\right) \psi_s(\eta_N) \end{aligned}$$

From this equation we will explore different intact, frayed, and bubble states of DNA using various contraction methods.

### 6.6.1 Intact DNA

For an intact ladder structure all base pairs are considered intact setting all the  $\mu$  indices in the transfer matrix  $\hat{T}$  to be  $\mu = 0$ . Starting from Eq. 6.58 our partition function looks like

$$\begin{aligned} Z(u) &= \sum_{s,t} \lambda_s^{N-1} \exp\left(-\frac{i4\pi su}{L}\right) \int d\xi_1 \psi_s(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \hat{\psi}_t(\xi_1) \\ &\quad \times \int \prod_{j=1}^N d\eta_j \left( \prod_{j'=1}^{N-1} \hat{T}_{00}(\eta_{j'+1}, \eta_{j'}) \right) \hat{\psi}_t^*(\eta_1) \exp\left(-\frac{V(\eta_N)}{2}\right) \psi_s(\eta_N) \quad (6.59) \end{aligned}$$

To contract the  $\hat{T}_{00}$  transfer matrices we use the eigenvalue equation Eq. 6.25 to propagate  $\hat{\psi}_t^*(\eta_1)$  up to  $\hat{\psi}_t^*(\eta_N)$ . After the contractions our partition function for

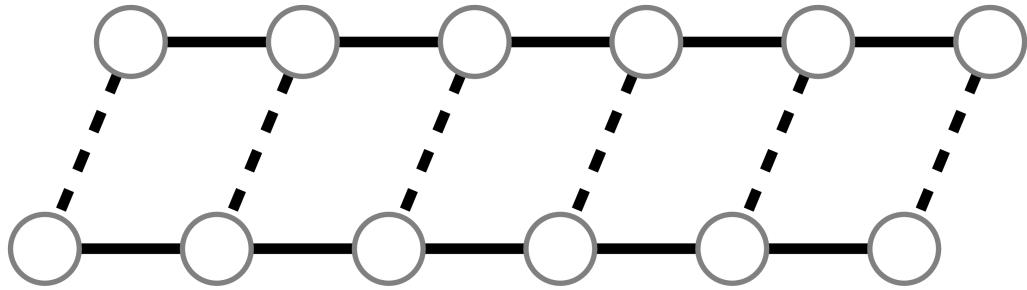


Figure 6.6: An intact DNA structure where  $\{\mu\} = \{0, 0, 0, 0, 0, 0\}$  for  $N = 6$ .

an intact state becomes

$$\begin{aligned} Z(u) &= \sum_{s,t} \lambda_s^{N-1} \hat{\lambda}_t^{N-1} \exp\left(-\frac{4\pi i s u}{L}\right) \int d\xi_N \psi_s(\xi_N) \exp\left(-\frac{V(\xi_N)}{2}\right) \hat{\psi}_t^*(\xi_N) \\ &\quad \times \int d\eta_1 \psi_s(\eta_1) \exp\left(-\frac{V(\eta_1)}{2}\right) \hat{\psi}_t(\eta_1) \end{aligned} \quad (6.60)$$

Notice that this leaves just two integrations, as well as two summations over the eigenvector label.

### 6.6.2 Frayed DNA

In a frayed configuration there is a single transfer matrix that links a broken base pair and an intact base pair, either  $\hat{T}_{01}$  and  $\hat{T}_{10}$ . We have not considered the eigenvalue equation for this matrix but we can alter the way we contract the matrices, leaving the single matrix in one of the remaining integrals. Starting from Eq. 6.58 for  $b$  broken base pairs the partition function looks like

$$\begin{aligned} Z(u) &= \sum_{s,t} \lambda_s^{N-1} \exp\left(-\frac{i4\pi s u}{L}\right) \int d\xi_1 \psi_s(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \hat{\psi}_t(\xi_1) \\ &\quad \times \int \prod_{j=1}^N d\eta_j \hat{T}_{00}(\eta_1, \eta_2) \dots \hat{T}_{01}(\eta_{N-b-1}, \eta_{N-b}) \dots \hat{T}_{11}(\eta_{N-1}, \eta_N) \\ &\quad \times \hat{\psi}_t^*(\eta_1) \exp\left(-\frac{V(\eta_N)}{2}\right) \psi_s(\eta_N) \end{aligned} \quad (6.61)$$

Similar to the intact state we can use the  $\hat{\psi}_t^*(\eta_1)$  eigenfunction to contract the  $\hat{T}_{00}$  transfer matrices up to  $\eta_{N-b-1}$  to give

$$\begin{aligned} Z(u) &= \sum_{s,t} \lambda_s^{N-1} \hat{\lambda}_t^{N-b} \exp\left(-\frac{i4\pi su}{L}\right) \int d\xi_1 \psi_s(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \hat{\psi}_t(\xi_1) \\ &\times \int \prod_{j=N-b-1}^N d\eta_j \hat{\psi}_t^*(\eta_{N-b-1}) \hat{T}_{01}(\eta_{N-b-1}, \eta_{N-b}) \dots \hat{T}_{11}(\eta_{N-1}, \eta_N) \\ &\times \exp\left(-\frac{V(\eta_N)}{2}\right) \psi_s(\eta_N) \end{aligned} \quad (6.62)$$

The eigenfunction  $\psi_s(\eta_N)$  is unsuitable for contracting the remaining transfer matrices. To obtain an appropriate eigenfunction we expand  $\psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right)$  in the set of eigenfunctions of  $\hat{T}_{11}$ :

$$\psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right) = \sum_v c_v^s \tilde{\psi}_v(\eta_N) \quad (6.63)$$

where

$$c_v^s = \int d\eta'_N \psi_s(\eta'_N) \exp\left(-\frac{V(\eta'_N)}{2}\right) \tilde{\psi}_v^*(\eta'_N) \quad (6.64)$$

Inserting these equations into Eq. 6.62 we get

$$\begin{aligned} Z(u) &= \sum_{s,t,v} \lambda_s^{N-1} \hat{\lambda}_t^{N-b} c_v^s \exp\left(-\frac{i4\pi su}{L}\right) \int d\xi_1 \psi_s(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \hat{\psi}_t(\xi_1) \\ &\times \int \prod_{j=N-b-1}^N d\eta_j \hat{\psi}_t^*(\eta_{N-b-1}) \hat{T}_{01}(\eta_{N-b-1}, \eta_{N-b}) \dots \hat{T}_{11}(\eta_{N-1}, \eta_N) \tilde{\psi}_v(\eta_N) \end{aligned} \quad (6.65)$$

We are now able to contract the remaining transfer matrices using  $\tilde{\psi}_v(\eta_N)$  and eigenvalue equation Eq. 6.26. The residual transfer matrix  $\hat{T}_{01}$  is left in the integral to be evaluated numerically. The expression for the partition function becomes

$$\begin{aligned} Z(u) &= \sum_{s,t,v} \lambda_s^{N-1} \hat{\lambda}_t^{N-b} \tilde{\lambda}_v^{b-1} c_v^s \exp\left(-\frac{i4\pi su}{L}\right) \int d\xi_1 \psi_s(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \hat{\psi}_t(\xi_1) \\ &\times \iint d\eta_{N-b-1} d\eta_{N-b} \hat{\psi}_t^*(\eta_{N-b-1}) \hat{T}_{01}(\eta_{N-b-1}, \eta_{N-b}) \tilde{\psi}_v(\eta_{N-b}) \end{aligned} \quad (6.66)$$

In the above equation the residual transfer matrix indicates that the broken base pairs lie on the right hand side of the structure as shown in Fig. 6.4, but the symmetric nature of the structure means that this equation would give the same results

for broken base pairs from left side of the ladder structure.

### 6.6.3 Double Frayed DNA

The symmetric behaviour of the system suggests that the base pairs corresponding to  $\eta_1$  and  $\eta_N$  should break at the same time. With the remaining base pairs intact we can consider this as a double frayed state such that base pairs break at both ends of the ladder structure. Using  $b$  and  $b'$  to denote the number of broken base pairs at each end we employ a different expansion of the delta function boundary conditions where

$$\delta(\xi_1 - \eta_1) = \sum_v \tilde{\psi}_v(\xi_1) \tilde{\psi}_v^*(\eta_1) \quad (6.67)$$

$$\delta(\xi_N + \eta_N - 2u) = \sum_s \psi_s^*(2u - \xi_N) \psi_s(\eta_N) \quad (6.68)$$

Using eigenfunctions of the  $\hat{T}_{11}$  transfer matrices, the contractions of the  $T$  transfer matrices are not affected by this new expansion. The partition function integral becomes

$$\begin{aligned} Z(u) = & \sum_{s,v} \lambda_s^{N-1} \exp\left(-\frac{4\pi i s u}{L}\right) \int d\xi_1 \psi_s(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \tilde{\psi}_v(\xi_1) \int \prod_{j=1}^N d\eta_j \\ & \times \psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right) \tilde{\psi}_v^*(\eta_1) \hat{T}_{11}(\eta_1, \eta_2) \dots \hat{T}_{10}(\eta_b, \eta_{b+1}) \hat{T}_{00}(\eta_{b+1}, \eta_{b+2}) \dots \\ & \dots \hat{T}_{00}(\eta_{N-b'-2}, \eta_{N-b'-1}) \hat{T}_{01}(\eta_{N-b'}, \eta_{N-b'+1}) \dots \hat{T}_{11}(\eta_{N-1}, \eta_N) \end{aligned} \quad (6.69)$$

We make an expansion of the functions in  $\eta_N$  in terms of eigenfunctions of the  $\hat{T}_{11}$  transfer matrix:

$$\psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right) = \sum_{v'} D_{v'}^s \tilde{\psi}_{v'}(\eta_N) \quad (6.70)$$

where

$$D_{v'}^s = \int d\eta'_N \psi_s(\eta'_N) \exp\left(-\frac{V(\eta'_N)}{2}\right) \tilde{\psi}_{v'}^*(\eta'_N) \quad (6.71)$$

Inserting this expression into the partition function we get

$$\begin{aligned} Z(u) &= \sum_{s,v,v'} \lambda_s^{N-1} D_{v'}^s \exp\left(-\frac{4\pi i s u}{L}\right) \int d\xi_1 \psi_s(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \tilde{\psi}_v(\xi_1) \\ &\times \int \prod_{j=1}^N d\eta_j \tilde{\psi}_v^*(\eta_1) \hat{T}_{11}(\eta_1, \eta_2) \dots \hat{T}_{10}(\eta_b, \eta_{b+1}) \hat{T}_{00}(\eta_{b+1}, \eta_{b+2}) \dots \\ &\dots \hat{T}_{00}(\eta_{N-b'-2}, \eta_{N-b'-1}) \hat{T}_{01}(\eta_{N-b'}, \eta_{N-b'+1}) \dots \hat{T}_{11}(\eta_{N-1}, \eta_N) \tilde{\psi}_{v'}(\eta_N) \end{aligned} \quad (6.72)$$

After making the relevant contractions using the properties of the  $\tilde{\psi}_v$  and  $\tilde{\psi}_{v'}$  eigenfunctions we get

$$\begin{aligned} Z(u) &= \sum_{s,v,v'} \lambda_s^{N-1} \tilde{\lambda}_v^{b-1} \tilde{\lambda}_{v'}^{b'-1} D_{v'}^s \exp\left(-\frac{4\pi i s u}{L}\right) \int d\xi_1 \psi_s(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \tilde{\psi}_v(\xi_1) \\ &\times \int \prod_{j=b}^{N-b'+1} d\eta_j \tilde{\psi}_v^*(\eta_b) \hat{T}_{10}(\eta_b, \eta_{b+1}) \hat{T}_{00}(\eta_{b+1}, \eta_{b+2}) \dots \\ &\dots \hat{T}_{00}(\eta_{N-b'-2}, \eta_{N-b'-1}) \hat{T}_{01}(\eta_{N-b'}, \eta_{N-b'+1}) \tilde{\psi}_{v'}(\eta_{N-b'+1}) \end{aligned} \quad (6.73)$$

To contract the remaining  $\hat{T}_{00}$  transfer matrices we introduce a pair of appropriate eigenfunctions by inserting the following representation of unity:

$$\int_{-\infty}^{\infty} \delta(\phi - \eta_{N-b'}) d\phi = \int_{-\infty}^{\infty} d\phi \sum_t \hat{\psi}_t(\phi) \hat{\psi}_t^*(\eta_{N-b'}) \quad (6.74)$$

where  $\phi$  is a dummy variable. After further contractions the partition function becomes

$$\begin{aligned} Z(u) &= \sum_{s,v,v',t} \lambda_s^{N-1} \tilde{\lambda}_v^{b-1} \tilde{\lambda}_{v'}^{b'-1} \tilde{\lambda}_t^{N-b-b'-1} D_{v'}^s \exp\left(-\frac{4\pi i s u}{L}\right) \\ &\times \int d\xi_1 \psi_s(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \tilde{\psi}_v(\xi_1) \iint d\eta_b d\phi \tilde{\psi}_v^*(\eta_b) \hat{T}_{10}(\eta_b, \phi) \hat{\psi}_t(\phi) \\ &\times \iint d\eta_{N-b'} d\eta_{N-b'+1} \hat{\psi}_t^*(\eta_{N-b'}) \hat{T}_{01}(\eta_{N-b'}, \eta_{N-b'+1}) \tilde{\psi}_{v'}(\eta_{N-b'+1}) \end{aligned} \quad (6.75)$$

The remaining expression involves five integrals and four summations.

#### 6.6.4 Bubbling DNA

Bubbling in dsDNA occurs when broken base pairs exist within the structure, not at the ends. This requires us to consider two regions of the ladder structure that have

intact base pairs and a region in the middle with broken base pairs. If we consider a configuration where we have  $l$  intact base pairs followed by  $b$  broken base pairs with the remaining base pairs intact such that  $l + b < N$ , the partition function integral would be of the form

$$\begin{aligned} Z(u) = \sum_{s,t} \lambda_s^{N-1} \exp\left(-\frac{4\pi i s u}{L}\right) \int d\xi_N \psi_s(\xi_N) \exp\left(-\frac{V(\xi_N)}{2}\right) \hat{\psi}_t^*(\xi_N) \int \prod_{j=1}^N d\eta_j \\ \times \psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right) \hat{\psi}_t^*(\eta_1) \hat{T}_{00}(\eta_1, \eta_2) \dots \hat{T}_{00}(\eta_{l-1}, \eta_l) \hat{T}_{01}(\eta_l, \eta_{l+1}) \hat{T}_{11}(\eta_{l+1}, \eta_{l+2}) \dots \\ \dots \hat{T}_{11}(\eta_{l+b-1}, \eta_{l+b}) \hat{T}_{10}(\eta_{l+b}, \eta_{l+b+1}) \hat{T}_{00}(\eta_{l+b+1}, \eta_{l+b+2}) \dots \hat{T}_{00}(\eta_{N-1}, \eta_N) \end{aligned} \quad (6.76)$$

The eigenfunction  $\hat{\psi}_t^*$  is available to contract the transfer matrices for intact base pairs from  $\eta_1 \rightarrow \eta_l$ . However, for the transfer matrices from  $\eta_N \rightarrow \eta_{l+b+1}$  we need to expand  $\psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right)$  in the set of eigenfunctions of  $\hat{T}_{00}$ ,

$$\psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right) = \sum_{t'} d_{t'}^s \hat{\psi}_{t'}(\eta_N) \quad (6.77)$$

where

$$d_{t'}^s = \int d\eta'_N \psi_s(\eta'_N) \exp\left(-\frac{V(\eta'_N)}{2}\right) \hat{\psi}_{t'}^*(\eta'_N) \quad (6.78)$$

By inserting these expressions into Eq. 6.76 we can contract all the  $\hat{T}_{00}$  transfer matrices to give

$$\begin{aligned} Z(u) = \sum_{s,t,t'} \lambda_s^{N-1} \hat{\lambda}_t^{l-1} \hat{\lambda}_{t'}^{N-l-b-1} d_{t'}^s \exp\left(-\frac{4\pi i s u}{L}\right) \int d\xi_N \psi_s(\xi_N) \exp\left(-\frac{V(\xi_N)}{2}\right) \hat{\psi}_t^*(\xi_N) \\ \times \int \prod_{j=l}^{N-l-b} d\eta_j \hat{\psi}_t^*(\eta_j) \hat{T}_{01}(\eta_l, \eta_{l+1}) \hat{T}_{11}(\eta_{l+1}, \eta_{l+2}) \dots \hat{T}_{11}(\eta_{l+b-1}, \eta_{l+b}) \dots \\ \times \hat{T}_{10}(\eta_{l+b}, \eta_{l+b+1}) \hat{\psi}_{t'}(\eta_{l+b+1}) \end{aligned} \quad (6.79)$$

To contract the remaining  $\hat{T}_{11}$  transfer matrices we introduce a pair of appropriate eigenfunctions by inserting the following representation of unity:

$$\int_{-\infty}^{\infty} \delta(\phi - \eta_{l+b}) d\phi = \int_{-\infty}^{\infty} d\phi \sum_v \tilde{\psi}_v^*(\phi) \tilde{\psi}_v(\eta_{l+b}) \quad (6.80)$$

where  $\phi$  is a dummy variable. After the final contractions the partition function becomes

$$\begin{aligned} Z(u) = & \sum_{s,t,t',v} \lambda_s^{N-1} \hat{\lambda}_t^{l-1} \hat{\lambda}_{t'}^{N-l-b-1} \tilde{\lambda}_v^{b-1} d_{t'}^s \exp\left(-\frac{4\pi i s u}{L}\right) \\ & \times \int d\xi_N \psi_s(\xi_N) \exp\left(-\frac{V(\xi_N)}{2}\right) \hat{\psi}_t^*(\xi_N) \\ & \times \int d\eta_l \int d\phi \hat{\psi}_{t'}(\eta_l) \hat{T}_{01}(\eta_l, \phi) \tilde{\psi}_v^*(\phi) \\ & \times \int d\eta_{l+b} \int d\eta_{l+b+1} \tilde{\psi}_v(\eta_{l+b}) \hat{T}_{10}(\eta_{l+b}, \eta_{l+b+1}) \hat{\psi}_t(\eta_{l+b+1}) \end{aligned} \quad (6.81)$$

once again involving five integrations and four summations.

## 6.7 Mean Axial Displacement

We now employ the partition function to determine mean axial displacement patterns for various patterns of breakage. Increasing the extension  $u$  in the DNA ladder structure increases the stress across each of the base pairs. The redistribution of the forces in the ladder structure means that the mean axial displacement for each base pair will not be the same. We can calculate the average of the structural variables  $\xi$  and  $\eta$  from the knowledge of the partition function. These mean axial displacements with the boundary conditions labelled as  $C(u)$  are

$$\langle \xi_\alpha \rangle = \frac{\int \prod_{j=1}^N d\xi_j d\eta_j \xi_\alpha \exp(-\beta H(\xi_j, \eta_j)) C(u)}{\int \prod_{j=1}^N d\xi_j d\eta_j \exp(-\beta H(\xi_j, \eta_j)) C(u)} \quad (6.82)$$

$$\langle \eta_\alpha \rangle = \frac{\int \prod_{j=1}^N d\xi_j d\eta_j \eta_\alpha \exp(-\beta H(\xi_j, \eta_j)) C(u)}{\int \prod_{j=1}^N d\xi_j d\eta_j \exp(-\beta H(\xi_j, \eta_j)) C(u)} \quad (6.83)$$

where the denominator is the partition function Eq. 6.11. Recombining these quantities using Eq. 6.5 and Eq. 6.6 we are able to find the average positions  $\langle x_\alpha \rangle$  and  $\langle y_\alpha \rangle$ .

### 6.7.1 Calculation of average $\eta_\alpha$

In the intact state of DNA the calculation of average  $\eta_\alpha$  requires the evaluation of the following integral using the transfer integral method

$$Z_{\eta_\alpha}(u) = \int \prod_{j=1}^N d\xi_j d\eta_j \eta_\alpha \exp(-\beta H(\xi_j, \eta_j)) C(u) \quad (6.84)$$

This integral only affects the contraction of transfer matrices in  $\eta$ , therefore we can continue from a modified version of Eq. 6.58 that includes  $\eta_\alpha$

$$\begin{aligned} Z_{\eta_\alpha}(u) &= \sum_{s,t} \lambda_s^{N-1} \exp\left(-\frac{4\pi i s u}{L}\right) \int d\xi_N \psi_s(\xi_N) \exp\left(-\frac{V(\xi_N)}{2}\right) \hat{\psi}_t^*(\xi_N) \\ &\times \int \prod_{j=1}^N d\eta_j \psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right) \hat{\psi}_t^*(\eta_1) \eta_\alpha \left( \prod_{j'=1}^{N-1} \hat{T}_{\mu_{j'}, \mu_{j'+1}}(\eta_{j'+1}, \eta_{j'}) \right) \end{aligned} \quad (6.85)$$

Setting  $\mu = 0$  for all the base pairs we are left with  $\eta_\alpha$  within a set of transfer matrices. For a given  $\eta_\alpha$  the partition function becomes

$$\begin{aligned} Z_{\eta_\alpha}(u) &= \sum_{s,t} \lambda_s^{N-1} \exp\left(-\frac{4\pi i s u}{L}\right) \int d\xi_N \psi_s(\xi_N) \exp\left(-\frac{V(\xi_N)}{2}\right) \hat{\psi}_t^*(\xi_N) \\ &\times \int \prod_{j=1}^N d\eta_j \psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right) \hat{\psi}_t^*(\eta_1) \\ &\times \hat{T}_{00}(\eta_1, \eta_2) \dots \hat{T}_{00}(\eta_{\alpha-1}, \eta_\alpha) \eta_\alpha \hat{T}_{00}(\eta_\alpha, \eta_{\alpha+1}) \dots \hat{T}_{00}(\eta_{N-1}, \eta_N) \end{aligned} \quad (6.86)$$

We now contract the matrices using the relevant eigenvalue equations starting from  $\eta_1$ , and downwards from  $\eta_N$ , leaving  $\eta_\alpha$  left in the integral.

$$\begin{aligned} Z_{\eta_\alpha}(u) &= \sum_{s,t,t'} \lambda_s^{N-1} \hat{\lambda}_t^{\alpha-1} \hat{\lambda}_{t'}^{N-\alpha} d_{t'}^s \exp\left(-\frac{i4\pi s u}{L}\right) \\ &\times \int d\xi_1 \psi_s(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \hat{\psi}_t(\xi_1) \int d\eta_\alpha \hat{\psi}_t^*(\eta_\alpha) \eta_\alpha \hat{\psi}_{t'}(\eta_\alpha) \end{aligned} \quad (6.87)$$

where  $d_{t'}^s$  is given by Eq. 6.78.

### 6.7.2 Calculation of average $\xi_\alpha$

To calculate the  $Z_{\xi_\alpha}$  integral we begin from a modified version of Eq. 6.58 that includes the extra term  $\xi_\alpha$  and use  $\hat{\psi}_t^*(\eta_1)$  to contract all the  $\hat{T}$  transfer matrices up

to  $\eta_N$

$$\begin{aligned} Z_{\xi_\alpha}(u) &= \sum_{s,t} \hat{\lambda}_t^{N-1} \exp\left(-\frac{i4\pi su}{L}\right) \\ &\times \int d\eta_N \psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right) \hat{\psi}_t^*(\eta_N) \\ &\times \int \prod_{i=1}^N d\xi_i \left( \prod_{i'=1}^{N-1} T(\xi_{i'+1}, \xi_{i'}) \right) \xi_\alpha \hat{\psi}_t(\xi_1) \psi_s(\xi_N) \exp\left(-\frac{V(\xi_1)}{2}\right) \end{aligned} \quad (6.88)$$

We then use  $\psi_s$  to contract the first set of  $T$  transfer matrices from  $\xi_N$  down to  $\xi_\alpha$

$$\begin{aligned} Z_{\xi_\alpha}(u) &= \sum_{s,t} \lambda_s^{N-\alpha} \hat{\lambda}_t^{N-1} \exp\left(-\frac{i4\pi su}{L}\right) \\ &\times \int d\eta_N \psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right) \hat{\psi}_t^*(\eta_N) \\ &\times \int \prod_{i=1}^\alpha d\xi_i \left( \prod_{i'=1}^{\alpha-1} T(\xi_{i'+1}, \xi_{i'}) \right) \psi_s(\xi_\alpha) \xi_\alpha \hat{\psi}_t(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) \end{aligned} \quad (6.89)$$

To contract the remaining set of transfer matrices we perform an expansion of the  $\hat{\psi}_t$  eigenfunction in  $\psi_s$

$$\hat{\psi}_t(\xi_1) \exp\left(-\frac{V(\xi_1)}{2}\right) = \sum_{s'} h_{s'}^t \psi_{s'}(\xi_1) \quad (6.90)$$

where

$$h_{s'}^t = \int d\xi'_1 \psi_{s'}^*(\xi'_1) \exp\left(-\frac{V(\xi'_1)}{2}\right) \hat{\psi}_t(\xi'_1) \quad (6.91)$$

Inserting this into the partition function we are now able to contract the remaining  $T$  transfer matrices to get

$$\begin{aligned} Z_{\xi_\alpha}(u) &= \sum_{s,s',t} \lambda_{s'}^{\alpha-1} \lambda_s^{N-\alpha} \hat{\lambda}_t^{N-1} \exp\left(-\frac{i4\pi su}{L}\right) h_{s'}^t \\ &\times \int d\eta_N \psi_s(\eta_N) \exp\left(-\frac{V(\eta_N)}{2}\right) \hat{\psi}_t^*(\eta_N) \int d\xi_\alpha \psi_{s'}(\xi_\alpha) \xi_\alpha \psi_s(\xi_\alpha) \end{aligned} \quad (6.92)$$

Next we perform numerical calculations using these expressions.

## 6.8 Results for DNA eigensystem calculations

The numerical calculation of the eigenvalue problem was discretised into  $2m+1$  steps over an interval range  $[-\frac{L}{2}, \frac{L}{2}]$  to give a step size  $\Delta = \frac{L}{2m+1}$ . The accuracy of the eigenvalue calculations is dependent on the  $L$  and  $m$  variables and the associated determination of  $\Delta$  in the numerical calculation. There is a balance to be found between the discretisation of the eigenvalue problem, and the range in  $\eta$  that can be included in the integrals. For the  $T$  and  $\hat{T}$  transfer matrices in the DNA model, and for a given  $\Delta$ , we sampled a variety of different  $L$  and  $m$  combinations in order to determine their effect on the accuracy of the numerical calculations. The eigenvalues for  $\lambda_s$  are shown in Figs. 6.7-6.9 for a range of  $\Delta$ , and the  $\hat{\lambda}_t$  eigenvalues that correspond to the Parisi eigenvalues of the  $\hat{T}$  transfer matrix where  $\eta_B = L/2$  are shown in Figs. 6.10-6.11.

Analysis of the  $\lambda_s$  eigenvalues of the  $T$  transfer matrix indicates that it is possible to find a combination of  $L$  and  $m$  that produces sufficient agreement between numerical and analytical values for all  $\Delta$ . From the results shown in Fig. 6.7d, Fig. 6.8d, and Fig. 6.9d we are able to observe that this is brought about by increasing the integral range parameter  $L$ . Another example of this effect can be seen by comparing the results in Fig. 6.8a, and Fig. 6.7b. Both these plots have  $m = 10$  and we note that as we increase  $L$  from 5 to 10 we find the agreement between the analytical and numerical results improves. Expecting the increase of  $L$  to have the same effect on the  $\hat{\lambda}_t$  eigenvalues of the  $\hat{T}$  transfer matrix, we discover that for all combinations of  $L$  and  $m$  there are excellent agreements between the numerical and analytical calculations as shown in Fig. 6.10 and Fig. 6.11.

The eigenfunction plots of  $\hat{\psi}_t$  in Fig. 6.12, and Fig. 6.13 for the highest eigenvalues corresponding to  $t = 1, 2, 3, 4$  show that every eigenfunction has a limited range in  $\eta$ , and therefore by choosing a small  $L$  we are able to capture the non-zero parts of the integral equation Eq. 6.25 when calculating the eigenvalues. The eigenfunctions for  $\psi_s$  are periodic and therefore do not contain this limited range in  $\eta$ .

The eigenfunctions in Fig. 6.12, and Fig. 6.13 also demonstrate that by reducing  $\Delta$  we increase the resolution of the numerically generated eigenfunctions. This is essential for an accurate free energy calculation by numerical methods as it increases the number of data points in the discrete representation of the eigenfunction.

Turning our attention to the eigenfunctions  $\hat{\psi}$  and  $\tilde{\psi}$  of the  $\hat{T}_{00}$  and  $\hat{T}_{11}$  transfer matrices, respectively, we are able to observe the effect of finite  $\eta_B$  on the eigenfunctions through the  $g$  factors we imposed in Eq. 6.17, and Eq. 6.18 . The eigenfunction  $\hat{\psi}_t$  shown in Fig. 6.14 is non-zero only for  $|\eta| \leq \eta_B$  and the opposite behaviour of

the  $\tilde{\psi}_t$  eigenfunction is shown in Fig. 6.15.

### 6.8.1 Conclusion

Analysis of the eigenvalue calculations gives an insight into the range of parameters that would be suitable to perform partition function calculations. We have found that a balance of choosing  $L$  and  $m$  is needed in order to perform integrations over a suitable range, as well as providing enough data points within the eigenfunction to calculate the partition function accurately.

The plots for both the eigenvalues and eigenvectors of the  $T$  and  $\hat{T}$  transfer matrices have a high degree of accuracy when we compare the numerical and analytical results. For the eigenvalues of  $\lambda_s$ , the plots converge to a perfect fit as  $L$  is increased to 50, and from all the plots we find that by having a sufficient integral range over  $\eta$  we can encapsulate the complete eigenfunction. Considering the discretisation parameters we have used, the results are accurate when  $L$  is chosen in the range of 25 to 50, and  $m$  in a range of 50 to 100.

It is possible to increase the accuracy of the eigenvalues and eigenfunctions further by choosing larger discretisation parameters, however, as we can see in Fig. 6.9d this effect is minimal if we compare it to Fig. 6.9c. A disadvantage of increasing  $m$  is that it results in a significant increase in computation time.

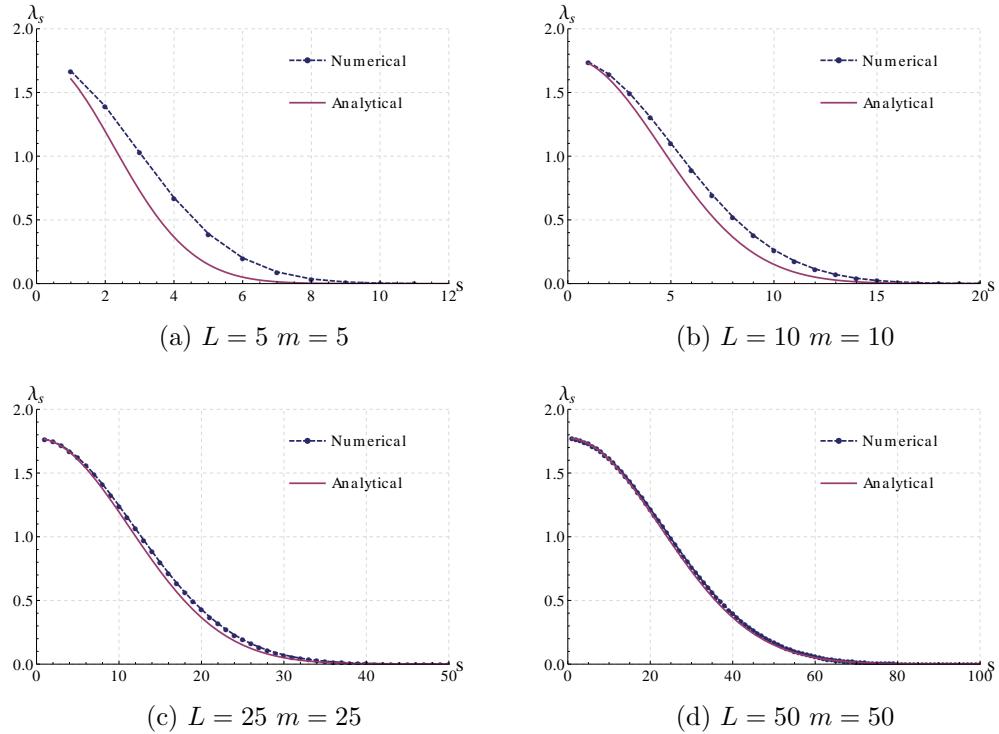
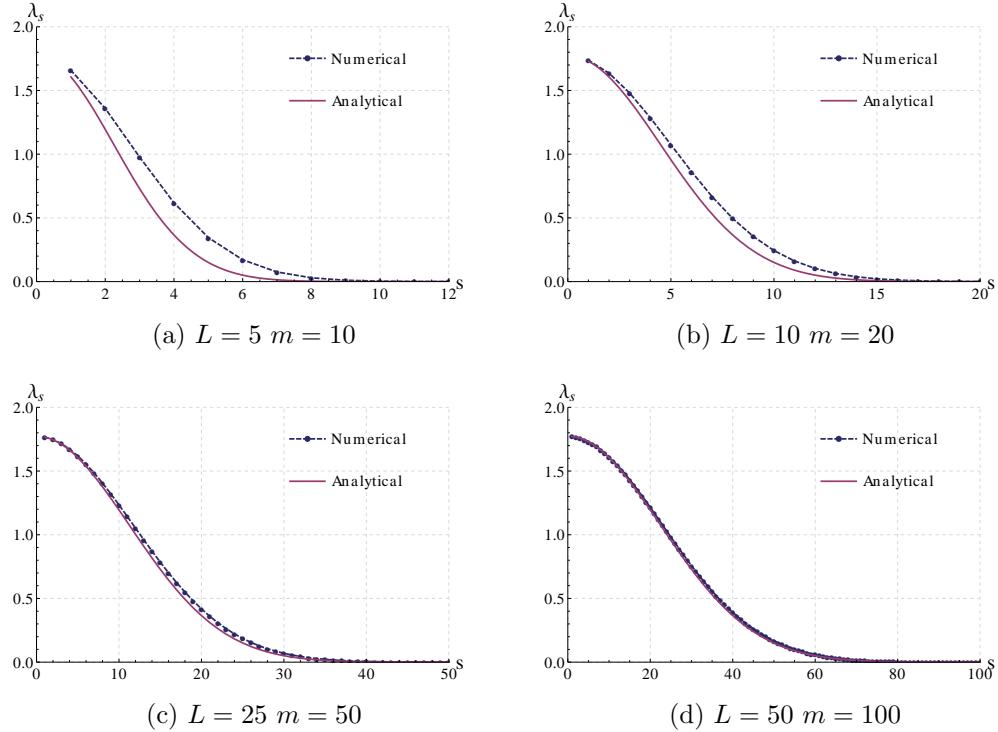
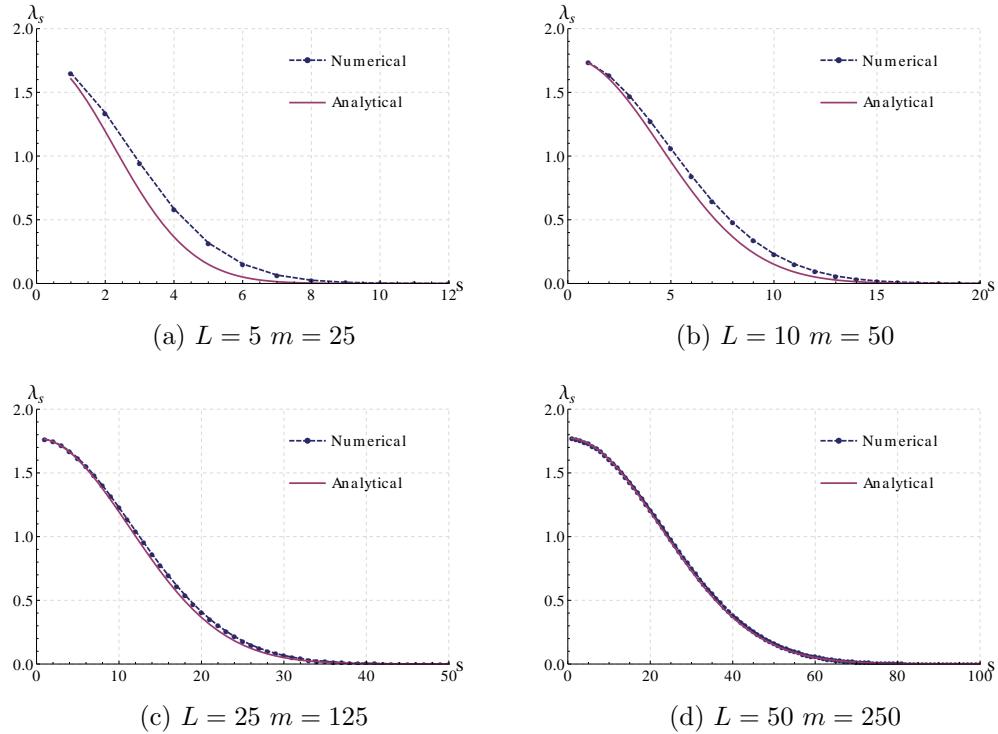


Figure 6.7: Numerical and analytical eigenvalue plots for  $\lambda_s$  where the  $2m + 1$  eigenvalues are shown in descending order. Different combinations of  $L$  and  $m$  are used such that  $\Delta = 0.5$  in these calculations.

Figure 6.8: Numerical and analytical eigenvalue plots for  $\lambda_s$  where  $\Delta = 0.25$ .Figure 6.9: Numerical and analytical eigenvalue plots for  $\lambda_s$  where  $\Delta = 0.1$ .

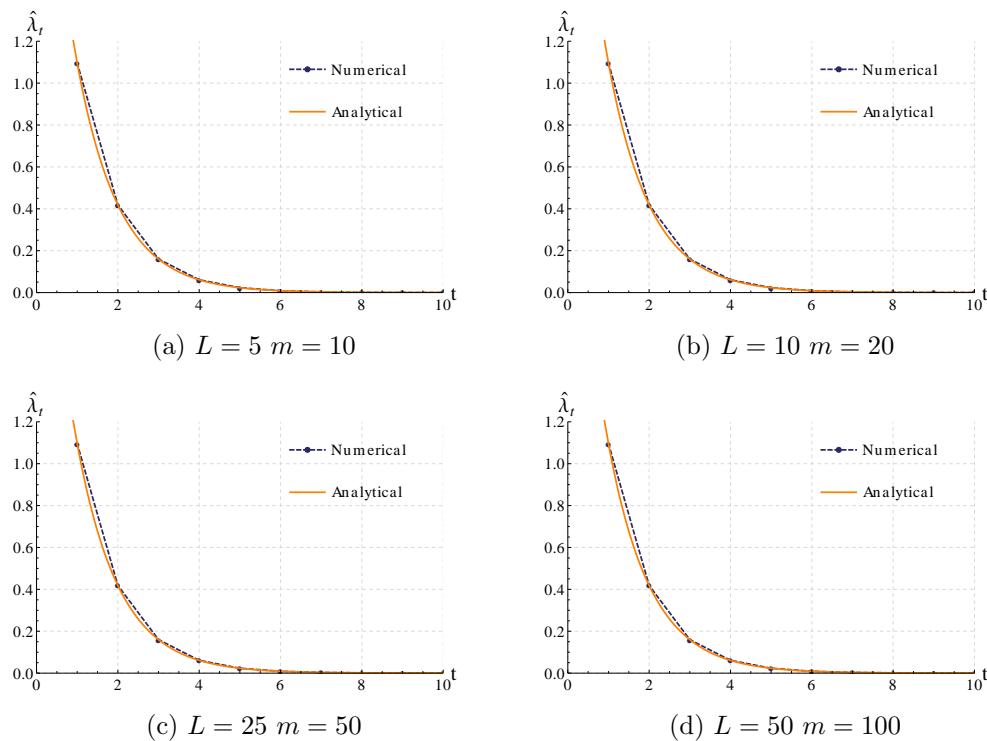


Figure 6.10: Numerical and analytical eigenvalue plots of  $\hat{\lambda}_t$  where  $\Delta = 0.25$ .

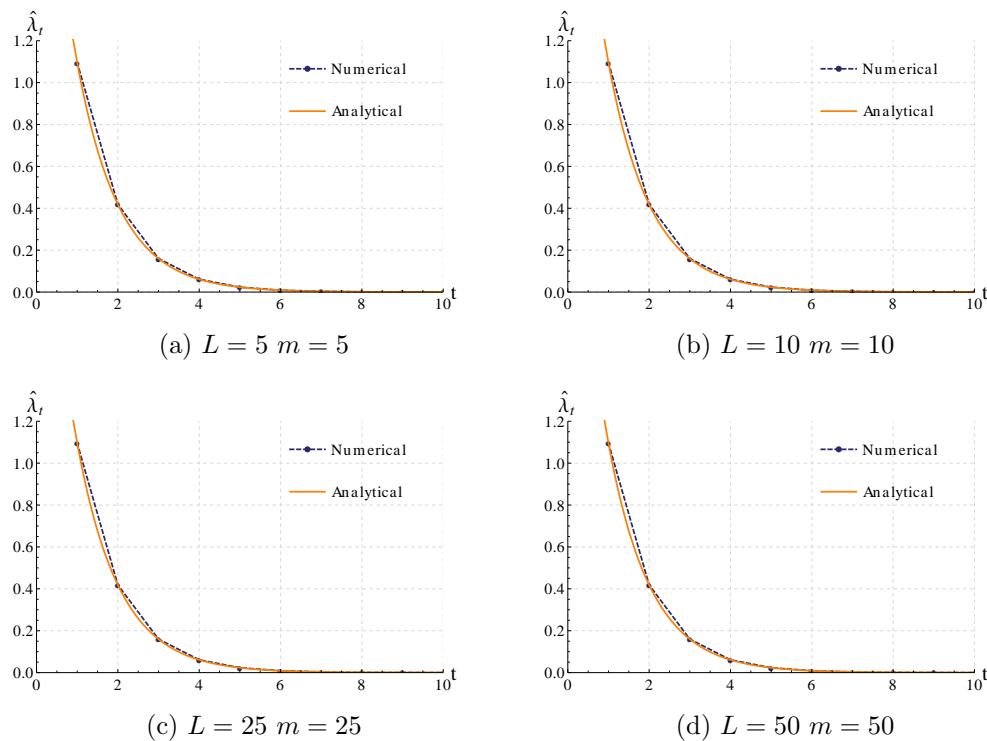


Figure 6.11: Numerical and analytical eigenvalue plots of  $\hat{\lambda}_t$  where  $\Delta = 0.5$ .

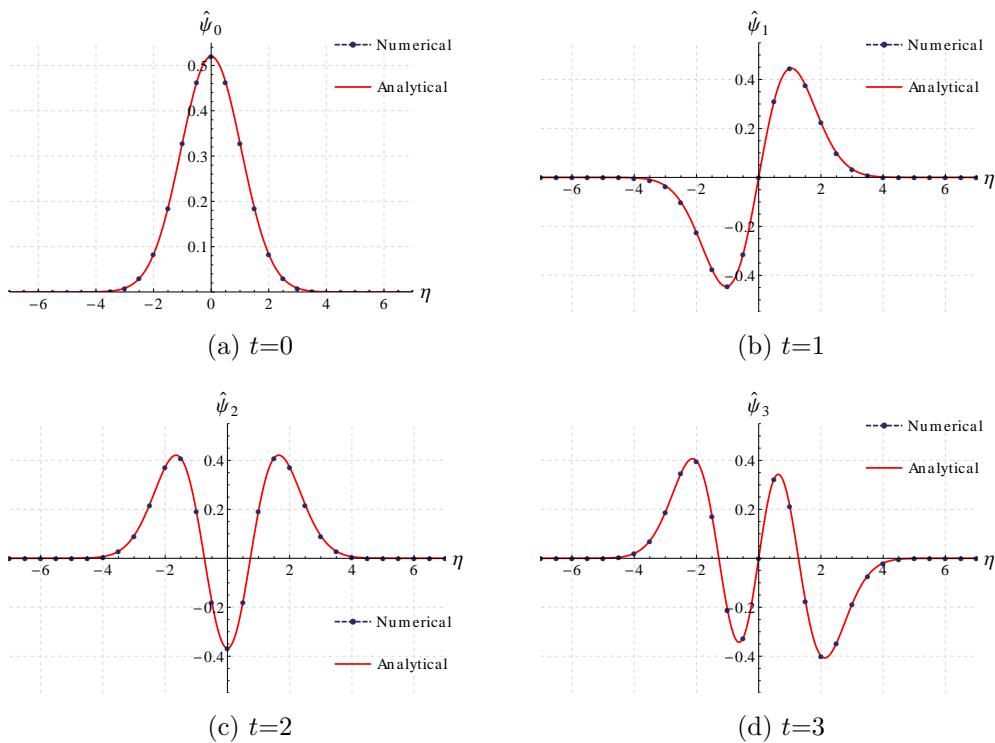


Figure 6.12: Eigenfunction plots of  $\hat{\psi}_t$  with  $\Delta = 0.5$  ( $L=50$  and  $m=50$ ).

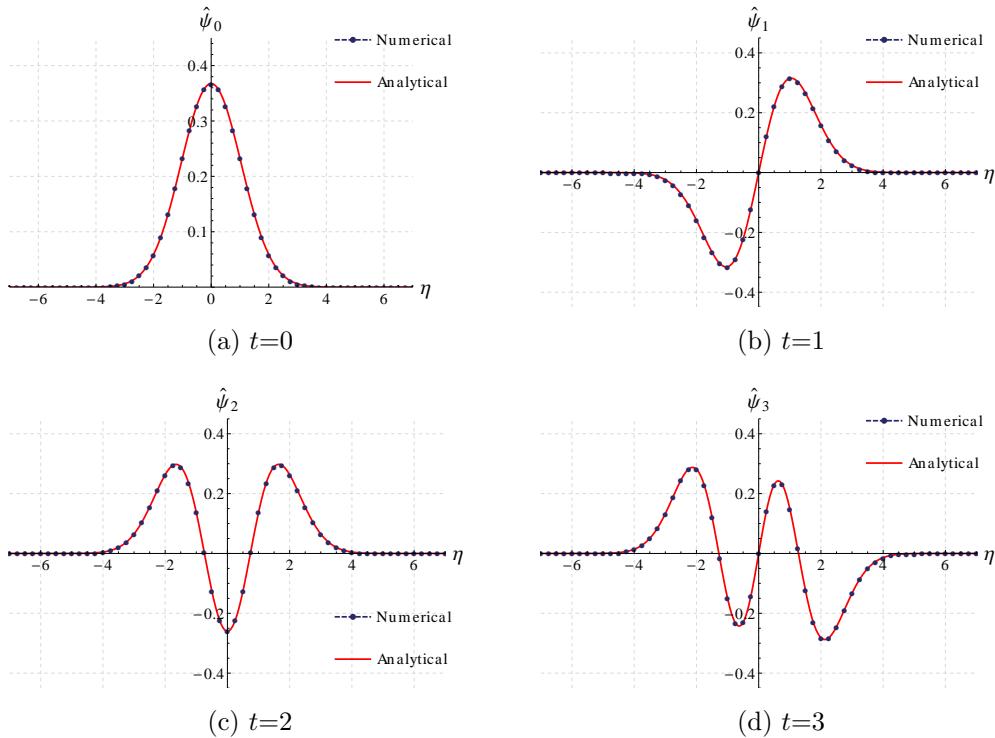


Figure 6.13: Eigenfunction plots of  $\hat{\psi}_t$  with  $\Delta = 0.25$  ( $L=50$  and  $m=100$ ).

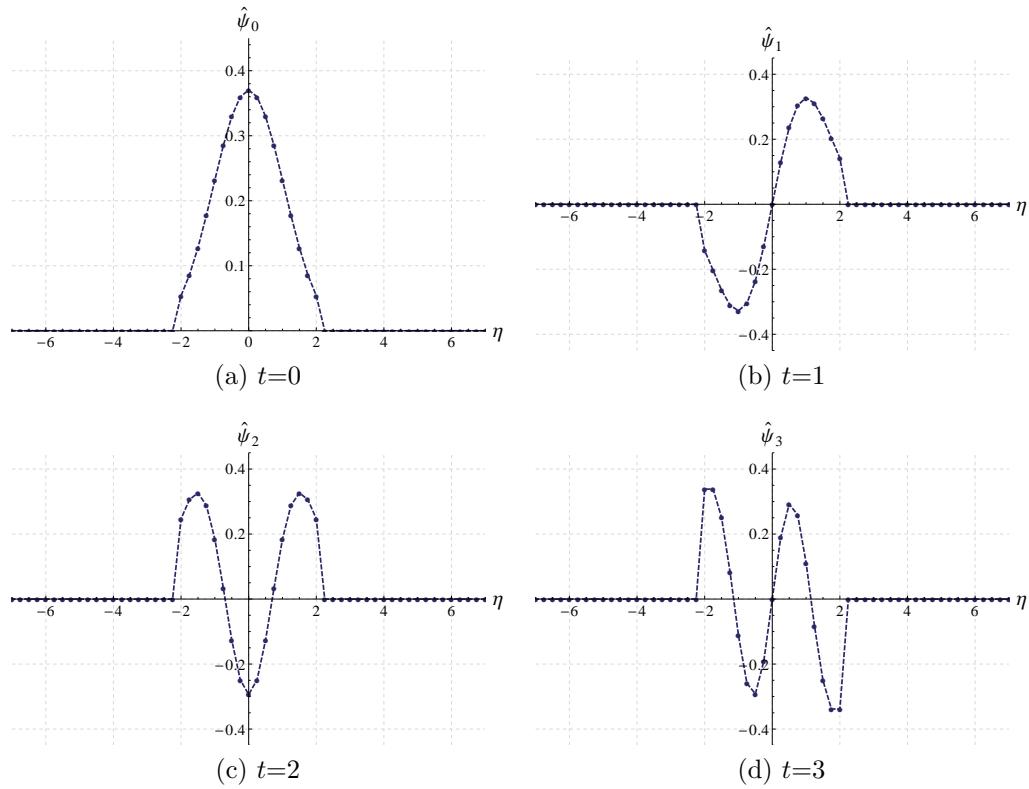


Figure 6.14: Eigenfunction plots of  $\hat{\psi}_t$  with  $\Delta = 0.25$  ( $L=50$  and  $m=100$ ) and  $\eta_B = 2$ .

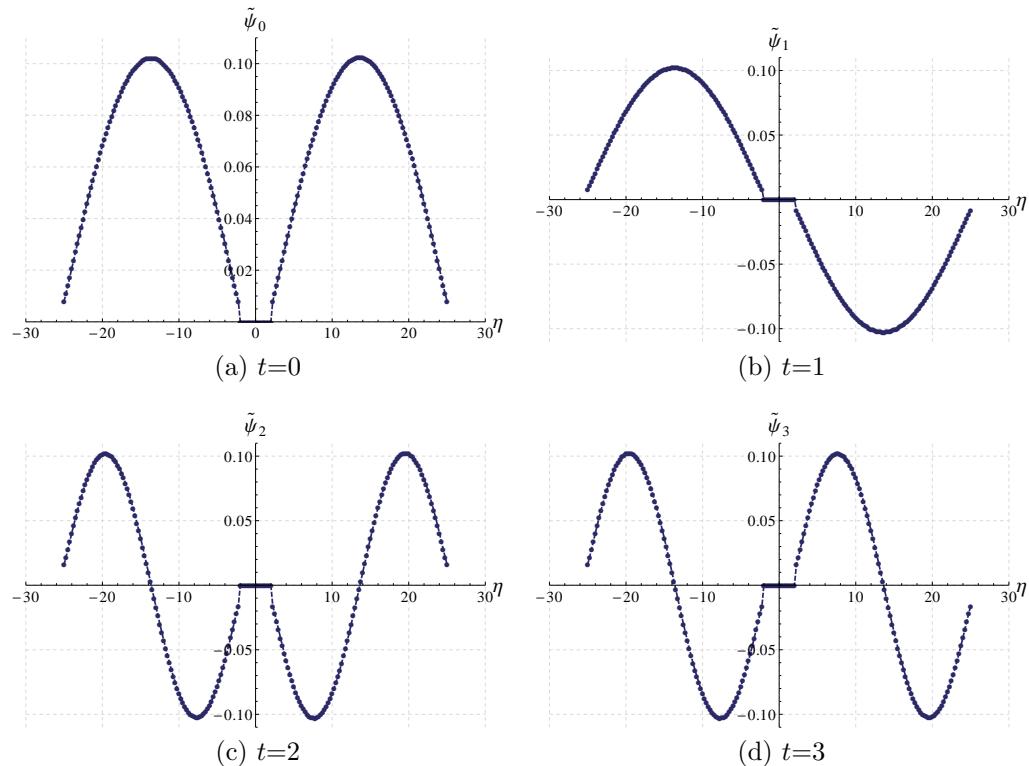


Figure 6.15: Eigenfunction plots of  $\tilde{\psi}_t$  with  $\Delta = 0.25$  ( $L=50$  and  $m=100$ ) and  $\eta_B = 2$ .

## 6.9 Results for DNA Model Calculations

Results calculated for an intact state of DNA using Eq. 6.60 are shown in Fig. 6.16. Observing a ladder structure that has all its base pairs intact we calculate free energies for a test case of  $N = 5$  by using a range of  $\eta_B$  and  $\kappa/\kappa'$  ratios in order to understand their effect on the ladder structure. Even though a DNA molecule with 5 base pairs is not thermodynamically stable we can still use this small structure to determine the general behaviour of the model. With  $\kappa$  being the spring constant of the backbone and  $\kappa'$  being the stiffness constant of the base pairs, we change the stiffness constant ratio by varying  $\kappa'$  and keeping  $\kappa$  fixed. The parameter  $\eta_B$  still represents the point at which the base pair bond breaks.

$$F(u) = -\frac{1}{\beta} \ln Z(u) \quad (6.93)$$

The effect of the parameter  $\eta_B$  on the intact state calculations can be seen in Figs. 6.16a-6.16c by the change in the curvature of free energy curves. As we increase the  $\eta_B$  limit we are effectively increasing the range of  $\eta$  at which a restoring force is felt by the base pair. When the extension  $u$  is at the  $\eta_B$  limit the bonds break and the restoring force exhibited by the potential vanishes.

From the dimensionless base pair potential Eq. 6.13, the stiffness constant ratio between the backbone and base pair quantifies the free energy or work needed to apply an extension  $u$  to the ladder structure. For  $\kappa/\kappa' \approx 1$  we the structure is rather rigid requiring a lot of work to extend the structure, whereas for a high stiffness constant ratio  $\kappa/\kappa' \approx 100$  the stiffness constant of the base pairs is much softer, requiring less work for the structure to extend. All of the free energy calculations in Fig. 6.16 show that reducing  $\kappa'$  lowers the free energy curve, this demonstrates that less work was required to extend the ladder structure.

The free energy calculations for different frayed states of the ladder structure are shown in Fig. 6.17. Compared with an intact state, we plot the four different frayed states of a ladder structure with  $N = 5$ , and for various  $\eta_B$ . Frayed ends of the ladder structure cause an increase of the initial free energy at  $u = 0$ ; as we increase  $\eta_B$  we find that this increase becomes more noticeable before falling to a minimum. This behaviour can be explained by the compression of the backbone bonds adjacent to the broken base pairs in the ladder structure such that at small  $u$  energy is stored in these bonds. For larger  $\eta_B$  more work is required to compress the backbone bonds resulting in more energy being stored. This is illustrated well in Fig. 6.17a through to Fig. 6.17f where we see the increase from  $\eta_B = 0.5$  to  $\eta_B = 1.25$ . As the structure

is extended the broken-pairs relax stabilising at a minimum free energy.

Bubble states in Fig. 6.18 have a larger increase in free energies at a given  $u$  than frayed states of a similar extent. We attribute this effect to the higher energy cost needed to break the bonds within the ladder structure than at the ends, due to the greater associated distortion of the backbones.

The mean axial displacement calculations in Fig. 6.19 highlight the symmetry that exists within the ladder structure. Although we are extending the structure from the  $y$  backbone, the equal and opposite reaction force acting on the  $x$  backbone results in a symmetry in the calculation of  $\langle \eta_j \rangle$  shown in Fig. 6.19a, Fig. 6.19c and Fig. 6.19e. For all extensions we find that the outermost base pairs extend more than the base pairs within the structure. As we increase the  $\eta_B$  parameter in Fig. 6.19 we allow the base pairs to extend over a larger range in  $\eta$  before breaking, and this results in a larger mean axial displacement for each base pair.

The calculation of average  $\xi_j$  shown in Fig. 6.19b, Fig. 6.19d and Fig. 6.19f draws our attention to the behaviour of the backbones in the ladder structure which is independent of the  $\eta_B$  parameter. Using both  $\langle \xi_j \rangle$  and  $\langle \eta_j \rangle$  we calculate the average backbone positions  $x_j$  and  $y_j$  shown in Fig. 6.20. For each extension we correctly observe the constraints applied by the delta functions with  $y_1 = 0$  and  $x_N = u$ .

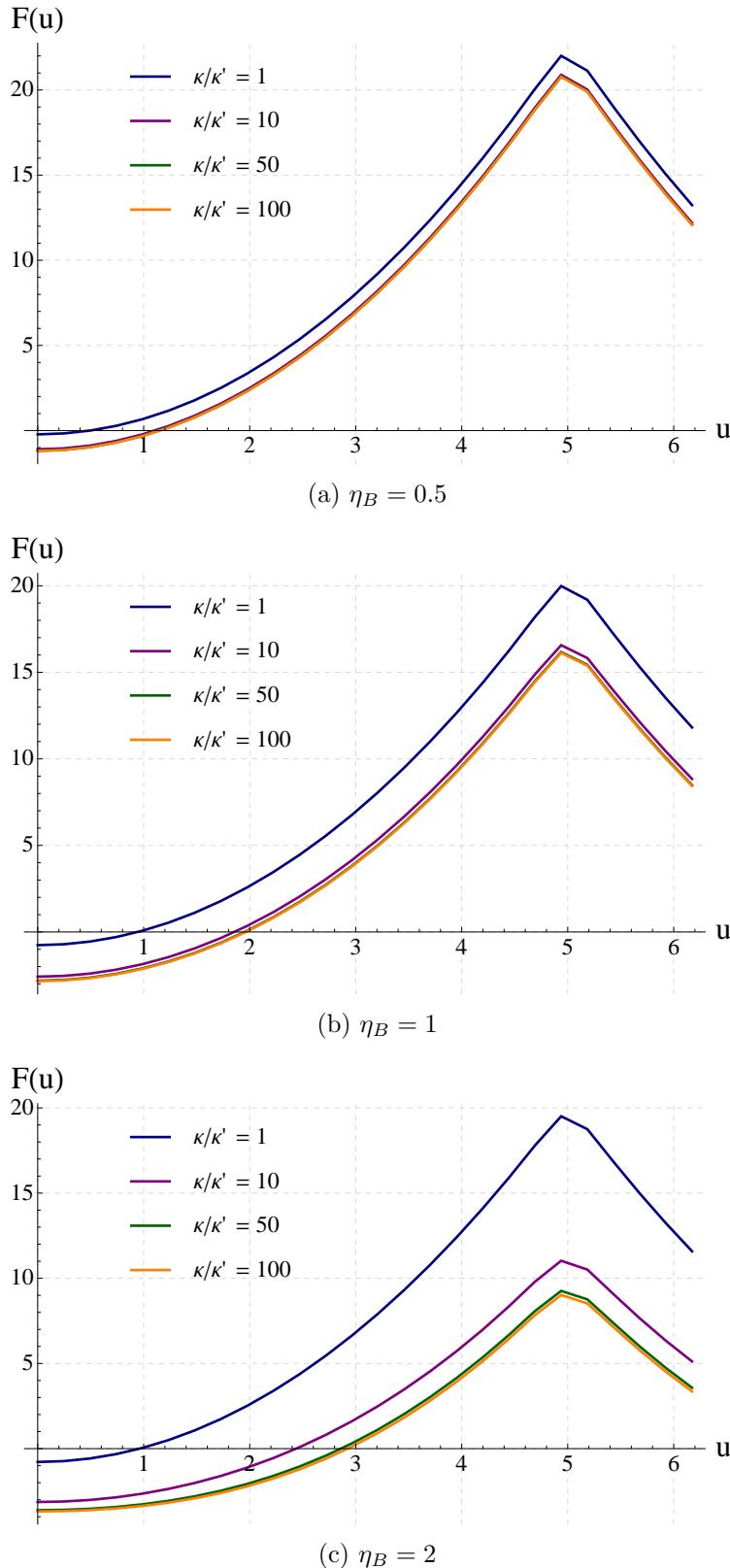


Figure 6.16: DNA Model simulations for an intact state of the ladder structure where  $N = 5$ . Simulations are run for a range of cases that vary the  $\eta_B$  limit, and the ratios of stiffness constants between the base pair and backbone ( $L = 20$ ). The peak in the free energy arises from the periodicity of the extension term in the partition function Eq. 6.60

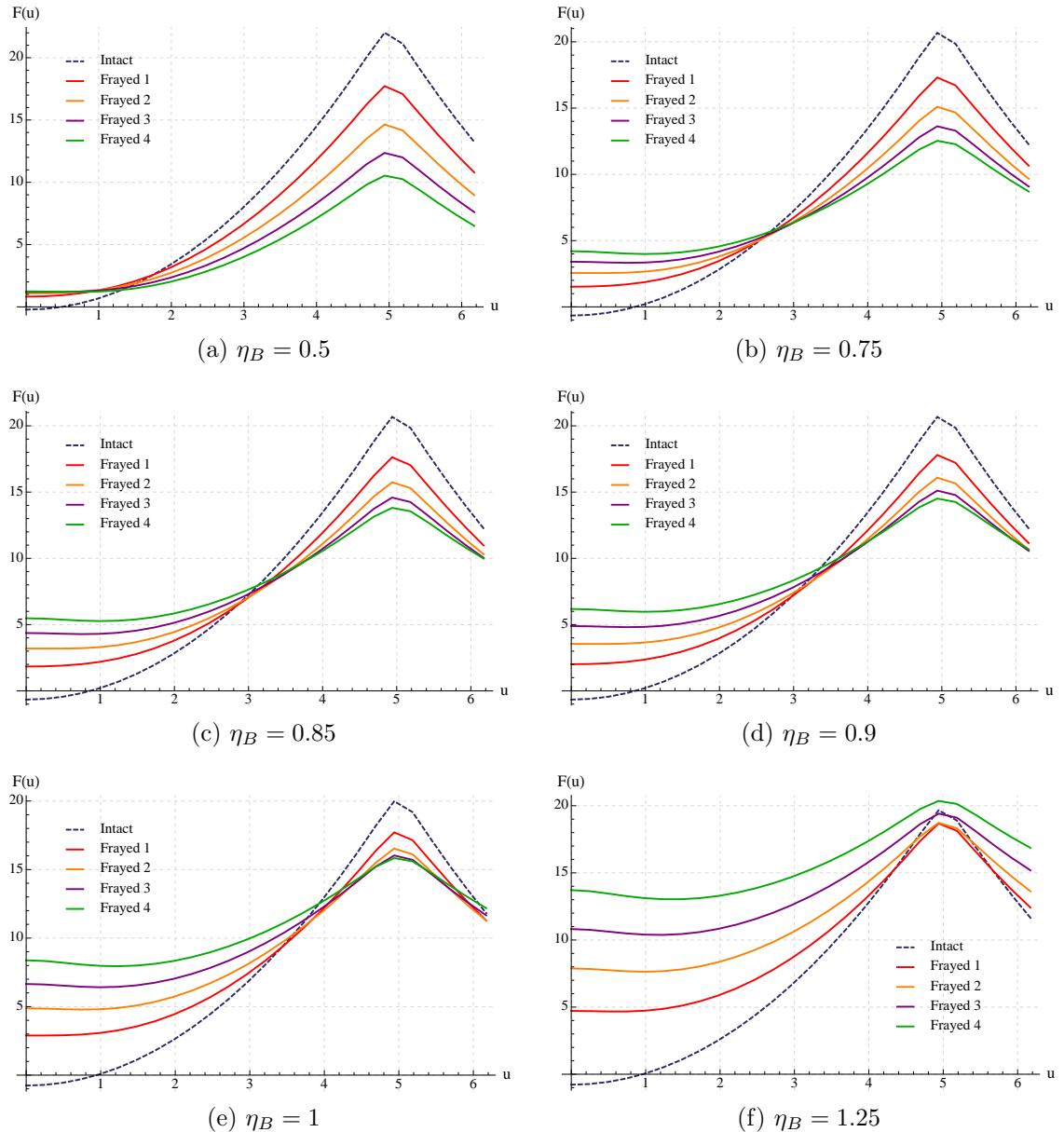


Figure 6.17: DNA free energy calculations for frayed states of the ladder structure where  $N = 5$ . The label for each frayed state indicates the number of broken base pairs from one end of the ladder structure. Simulations are run for a range of cases that vary the  $\eta_B$  limit.

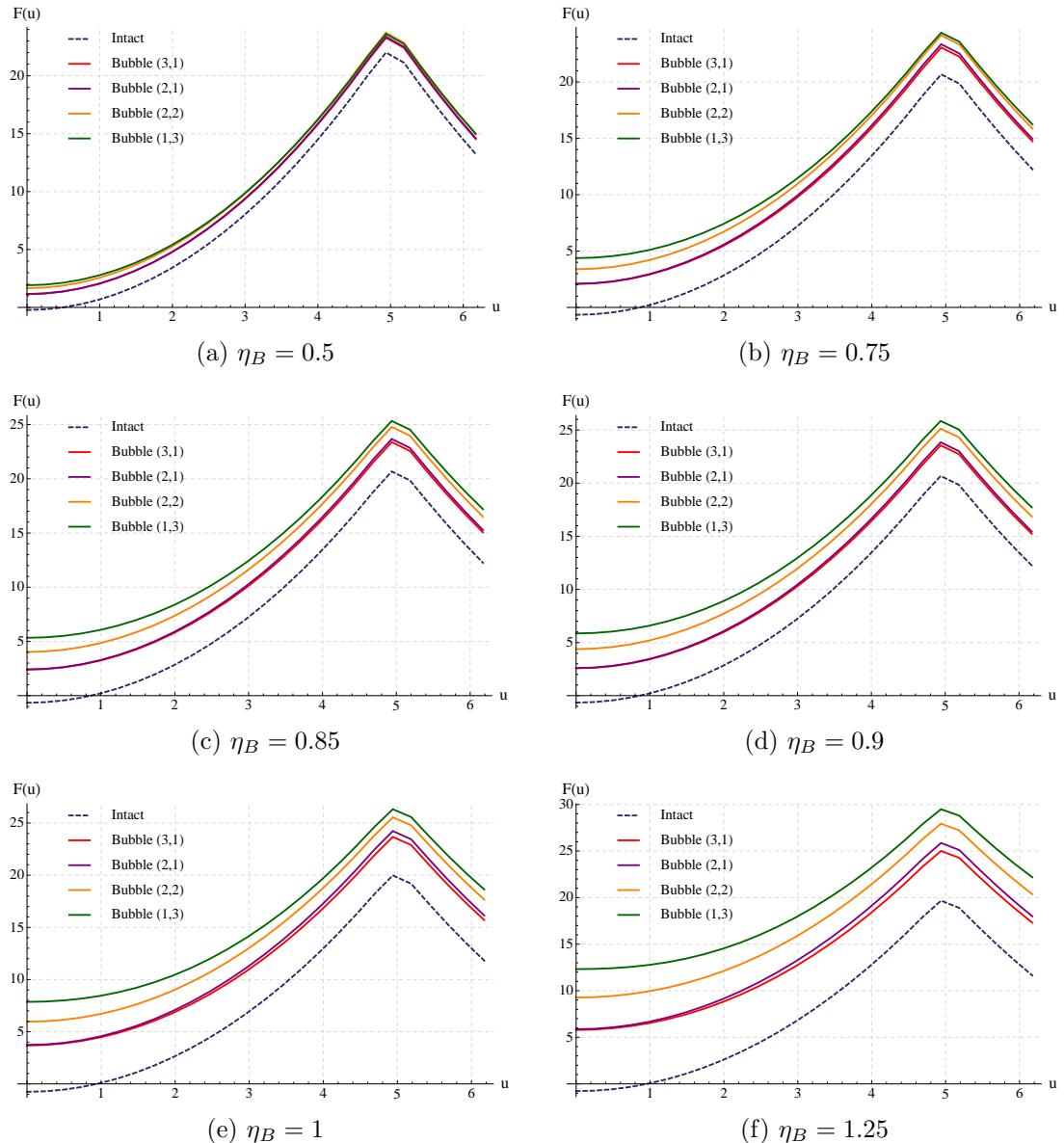


Figure 6.18: DNA free energy calculations for a bubble states of the ladder structure where  $N = 5$ . The label for each bubbled state indicates the number of intact base pairs, followed by the number of broken base pairs (*intact, broken*). For all the bubble states the last remaining base pair is always intact. Calculations are performed for a range of cases that vary the  $\eta_B$  limit.

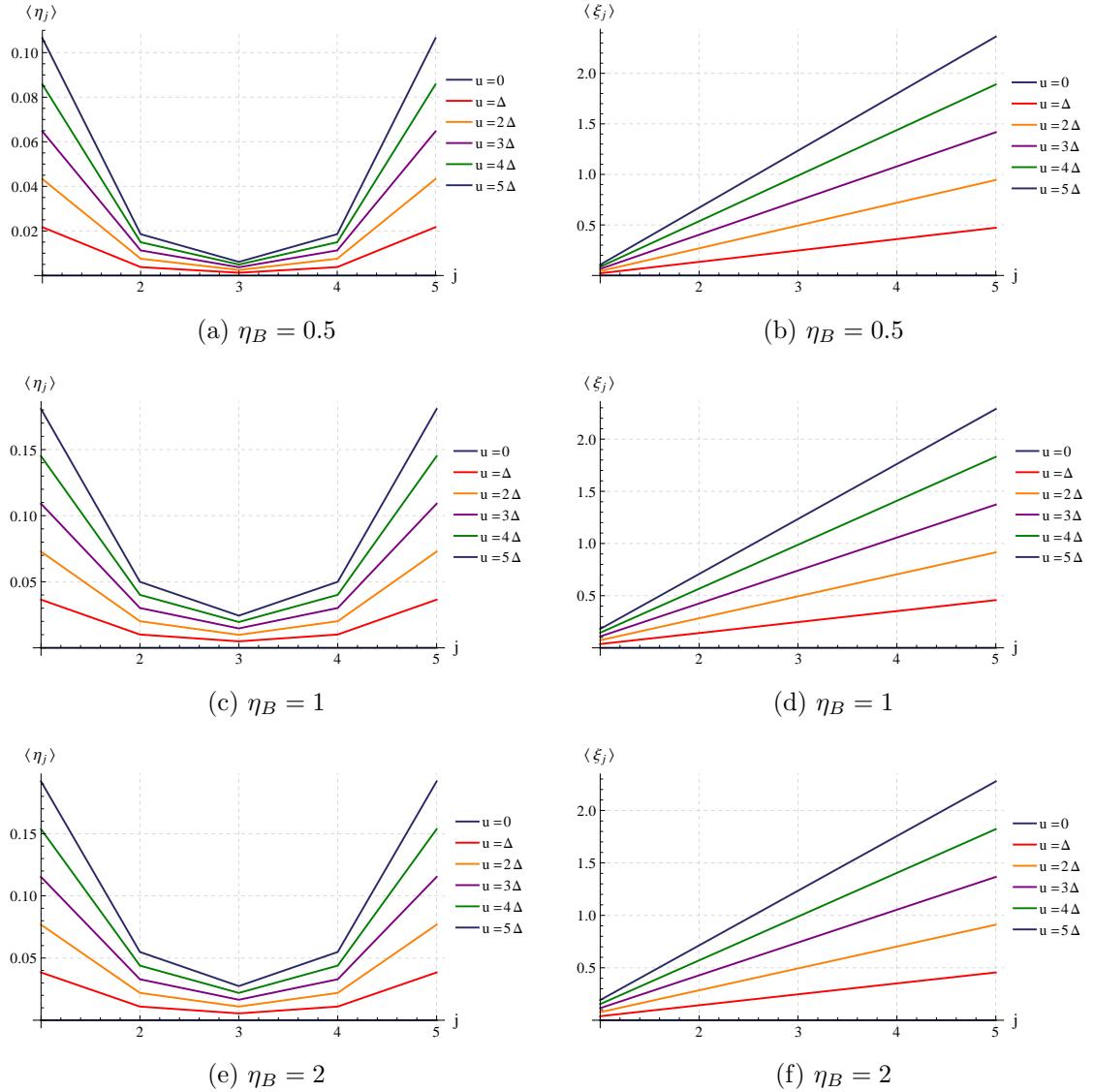


Figure 6.19: Mean axial displacement calculations for the average base pair separation when we apply an extension  $u$  on the ladder structure. Simulations are run for an intact state with  $\kappa/\kappa' = 1$  at different  $\eta_B$  limits ( $\Delta \approx 0.25$ ).

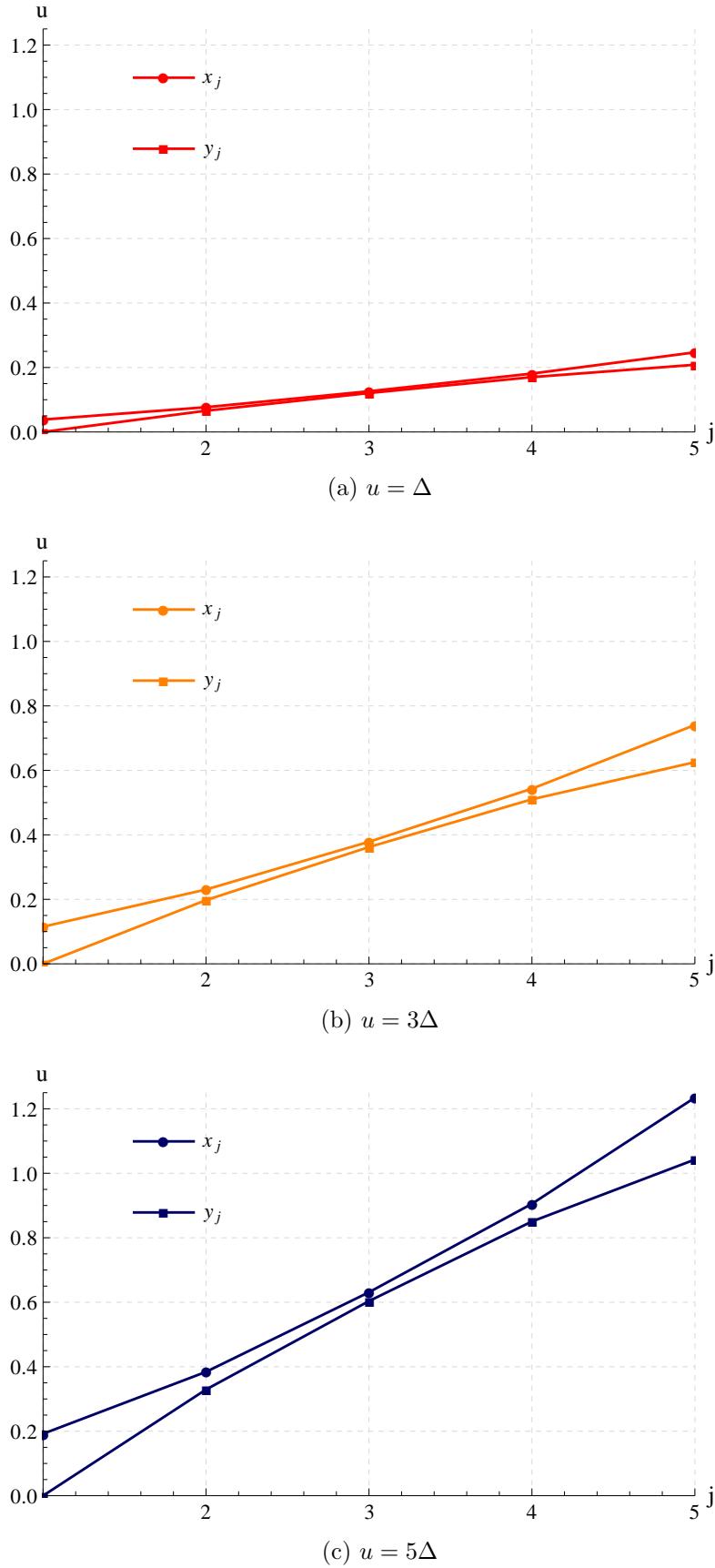


Figure 6.20: Mean axial displacement calculations for the average backbone displacements when we apply an extension  $u$  on the ladder structure. Calculations are performed for an intact state with  $\kappa/\kappa' = 1$  at different extensions ( $\Delta \approx 0.25$ ).

### 6.9.1 Conclusion

From studying the  $N = 5$  ladder structure we can draw some conclusions on how an applied extension affects the free energy for intact, frayed and bubble states. In all instances applying an extension to the ladder structure moves it away from its equilibrium position due to the work done on the thermodynamic system. This increases the internal energy and consequently the free energy. As we reduce  $\kappa'$  less work is required to extend the base pairs within the ladder structure lowering the free energy curve. Here the stiffness constant ratio increases since  $\kappa$  is always kept constant.

Broken base pairs in the frayed states move the equilibrium position (minimum free energy) of the ladder structure from  $u = 0$ . As calculations of  $\langle \eta \rangle$  have shown, mean base pair separations are not uniform within the structure, and therefore to allow the relaxation of the backbone adjacent to broken base pairs a large enough extension is needed. As the number of broken base pairs increases a larger extension is needed to relax the structure.

The crossover behaviour in free energy curves shown in Fig. 6.17c, Fig. 6.17d and Fig. 6.17e indicates separate phase transitions for a structure that unzips from one end. Initially, in an intact state at equilibrium, applying an extension from one end disturbs the structure causing the free energy to increase. The base pairs begin to separate with the largest displacements being experienced at the end base pairs. As the extension increases, the intact free energy curve will eventually intersect the first frayed free energy curve implying a phase transition to a frayed structure with a single broken base pair. This process will continue to happen for subsequent frayed states where more base pairs are broken. Following a path of lowest free energy we are able to observe how the base pairs break in a ladder structure from an intact state as the extension  $u$  increases.

The shift in higher free energy curves for bubble states arises from the distorted backbone around each end of the region of broken base pairs within the structure. The shift can be related to the number of broken base pairs that exist. No phase transitions exist for the bubble states as  $u$  increases.

## 6.10 Results for anharmonic DNA calculations

In our formulation of the partition function we chose a potential that mechanically described DNA as a harmonic ladder structure. Each base pair experiences a restoring force proportional to a given extension and eventually breaks once the base pair separation reaches the  $\eta_B$  limit, at which point the restoring force vanishes. Following the recent work of Prakash et al. who uses the same ladder structure to model DNA we find that their potential on the molecular scale gives a more realistic representation of the base pair potentials. This potential contains a long range attractive interaction and short range harmonic attraction that resembles our simpler base pair potential when the displacement is small. Expanding the Prakash potential  $V_p(y_i)$  in ascending powers of  $y$  one gets [77],

$$V_p(y_i) = -\epsilon + \frac{1}{2} \left( \frac{12\epsilon}{\sigma^2} \right) y_i^2 \quad (6.94)$$

$$\kappa' = \frac{12\epsilon}{\sigma^2} \quad (6.95)$$

where  $\epsilon$  is the potential depth,  $\sigma$  is the diameter of DNA and  $\kappa'$  is the stiffness constant of the base pair. In the limit of  $\eta_B \rightarrow \infty$  our potential remains harmonic. To model DNA we adopt Prakash's potential in our partition function where  $\eta_B$  now indicates where the base pair bond breaks as it is distorted.

The results from the first set of calculations compares the shearing force of different molecular lengths with experimental data obtained by Hatch et al. [79]. We use a method first employed by Prakash et al. [77] where for an intact structure the shear force  $f_s$  is determined when the average separation of the end base pairs,  $\eta_1$  and  $\eta_N$ , are equal to our fitting parameter  $\eta_B$ , this is also known as the shearing separation of a single base pair. At the applied extension we are able to determine the dimensionless force  $f$  from the partition function by taking the derivative of the free energy with respect to  $u$ . We find that

$$f(u) = \frac{1}{\beta} \frac{d \ln Z(u)}{du} \quad (6.96)$$

which can be transformed into a dimensional force by multiplying by a scaling factor  $(\frac{\kappa k_B T}{4})^{\frac{1}{2}}$ . The model is dependent on two fitting parameters,  $\eta_B$  and the base pair stiffness constant  $\kappa'$ , which we are able to adjust to produce a fit with experimental data. Using a backbone stiffness constant of  $\kappa = 0.10 \text{ eV}/\text{\AA}^2$  [77] we can estimate an initial value for  $\eta_B$  and  $\kappa'$  from data already obtained by Hatch et al.; a stiffness

constant ratio of  $\kappa/\kappa' = 90$  gave a base pair stiffness constant of  $\kappa' = 0.0011 \text{ eV}/\text{\AA}^2$ . The shearing separation  $\eta_B$  was calculated from the shearing force of a single base pair  $f_{bp}$  by taking a differential of the base pair potential at small separations. We find that at room temperature ( $T = 300$ ),

$$\eta_B = \frac{f_{bp}}{\kappa'} \left( \frac{\kappa}{4k_B T} \right)^{\frac{1}{2}} \simeq 2.2 \quad (6.97)$$

where  $f_{bp} = 4 \text{ pN}$  [79]. The dimensionless quantity  $\eta_B$  is converted into real dimensions by the scaling factor  $(\frac{4k_B T}{\kappa})^{\frac{1}{2}}$  to give  $\eta'_B = \eta_B (\frac{4k_B T}{\kappa})^{\frac{1}{2}} \simeq 2.25 \text{\AA}$  which is within the acceptable range found in [77]. In the DNA model these initial fitting parameters did not agree well with experimental data and therefore had to be adjusted to produce a better fit. The results in Fig. 6.21 show the shear forces for different molecular lengths that produced the best fit with experimental data. These calculations had a  $\kappa/\kappa'$  ratio of approximately 150 with a shear force base pair separation of  $\eta_B = 3.75$ . A fit of de Gennes's formula Eq. 4.22 to Hatch's data [79] gave a  $\kappa/\kappa' \approx 180$  and  $f_{bp} \approx 3.24 \text{ pN}$  ( $\eta'_B \approx 3 \text{\AA}$ ).

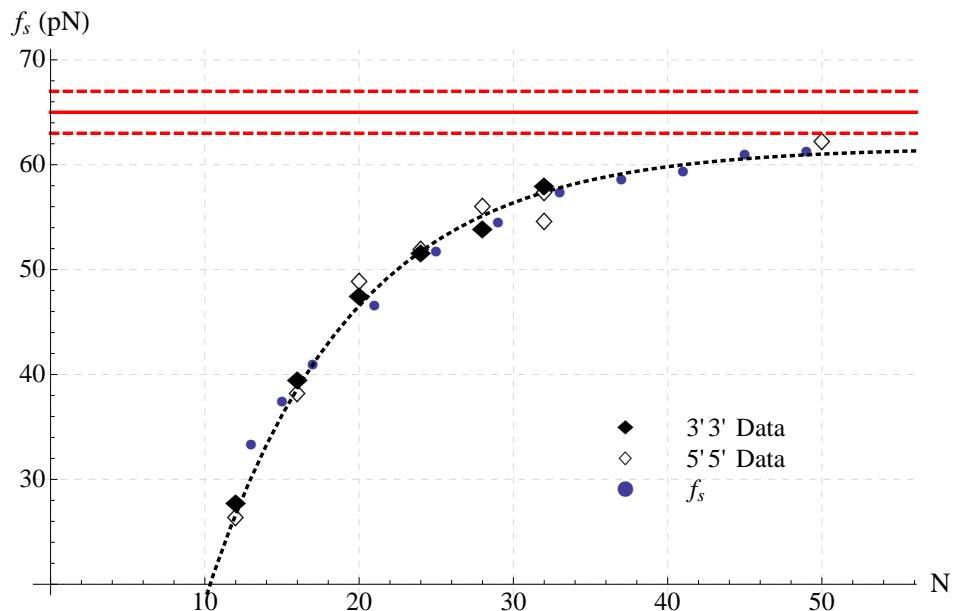
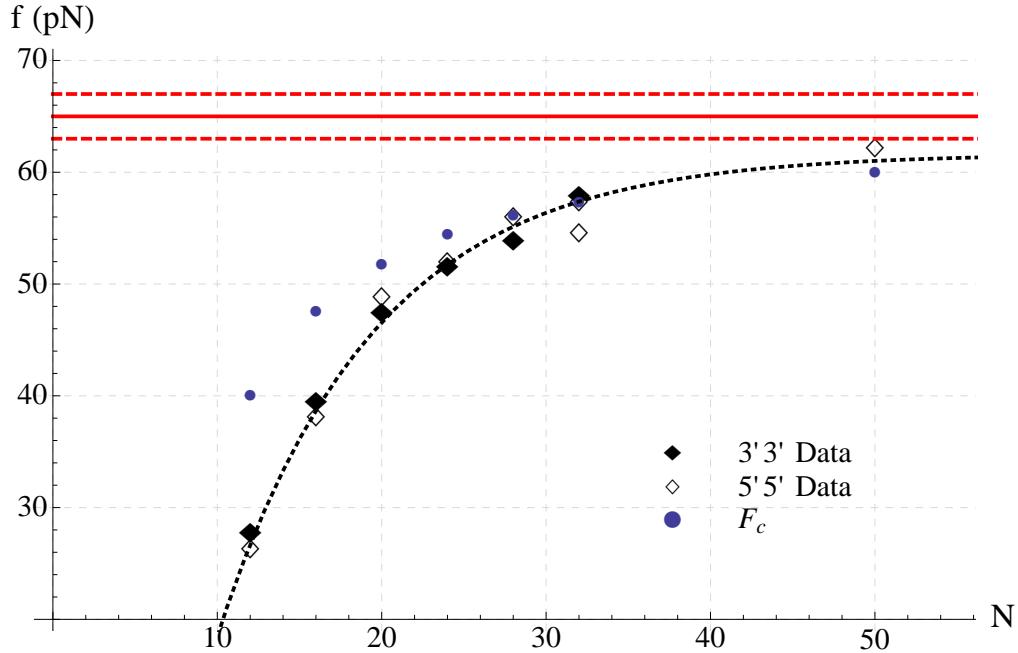


Figure 6.21: The shear force calculated for various molecular lengths plotted with experimental data. The parameters for these free energy calculations used a stiffness constant ratio of 150 with  $\eta_B = 3.75$ . A fit of de Gennes's model (dotted curve) is made with the experimental data. The centre red horizontal line represents the critical force for infinitely long polymers and the upper and lower dashed red lines correspond to 110% and 90% of the overstretching transition threshold [79].

In the second set of results we use the first phase transition, the transition from

an intact state to a frayed 1 state, to determine the shear force of DNA. Here the  $g$  factors within the transfer matrices set the pattern of breakage allowing the mechanics to determine the maximum sustainable shear force of the structure. The free energy curves shown for a sample of different molecular lengths in Fig. 6.22b and Fig. 6.22d show the intersection of different free energy curves, highlighting the pathways of breakages that occur in our model, corresponding to the path of minimum free energy as described in section 6.9.1. The results shown in Fig. 6.22a and Fig. 6.23 show some of the best fits with experimental data for a range of molecular lengths  $N = 12, 16, 20, 24, 28, 32, 50$  when we consider the point at which DNA shears to be the first phase transition. These results have a better consistency with the stiffness constant ratios found by Hatch [79] even though the results for short DNA do not agree so well with experimental data.



(a) The shear force calculated for various molecular lengths plotted with experimental data. Parameters for these free energy calculations include a stiffness constant ratio of  $\approx 65$  with  $\eta_B = 3.75$ . A plot of de Gennes model (dotted curve) is made, together with the experimental data.

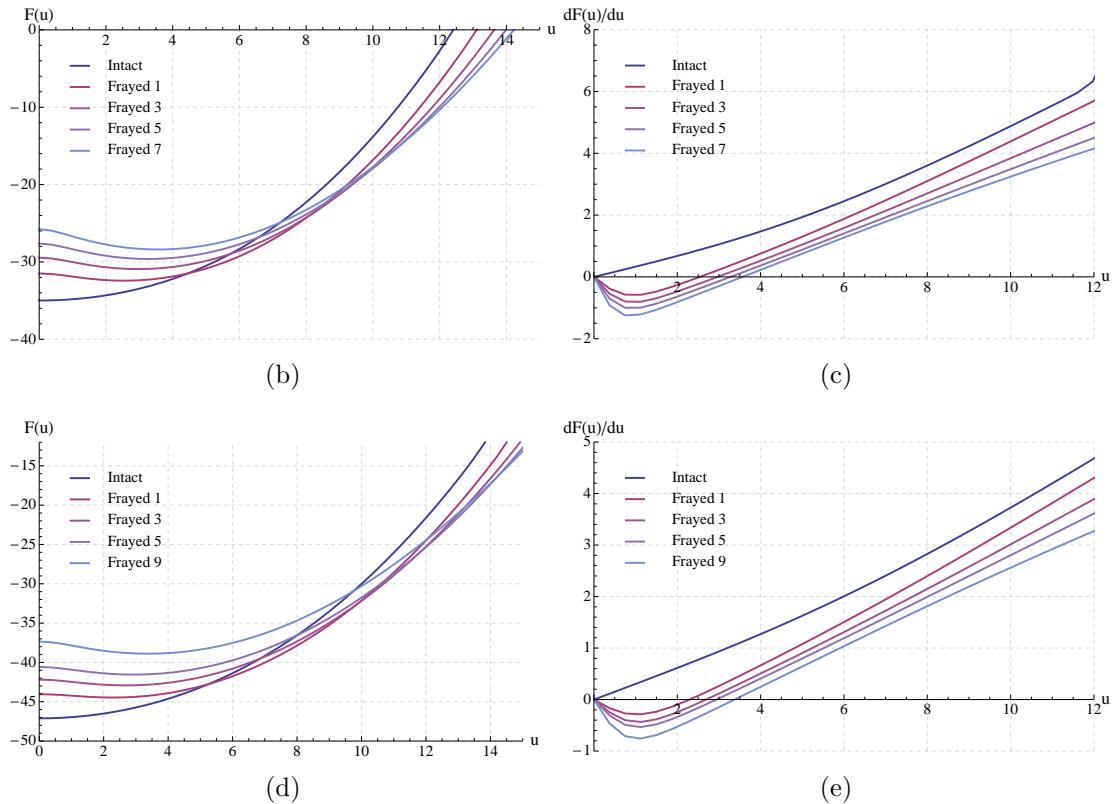


Figure 6.22: Free energy curves for the intact state and first few frayed states focusing on specific data points corresponding to  $N = 12$  and  $N = 16$ . A plot of the dimensionless force corresponding to each free energy curve is illustrated on the right.

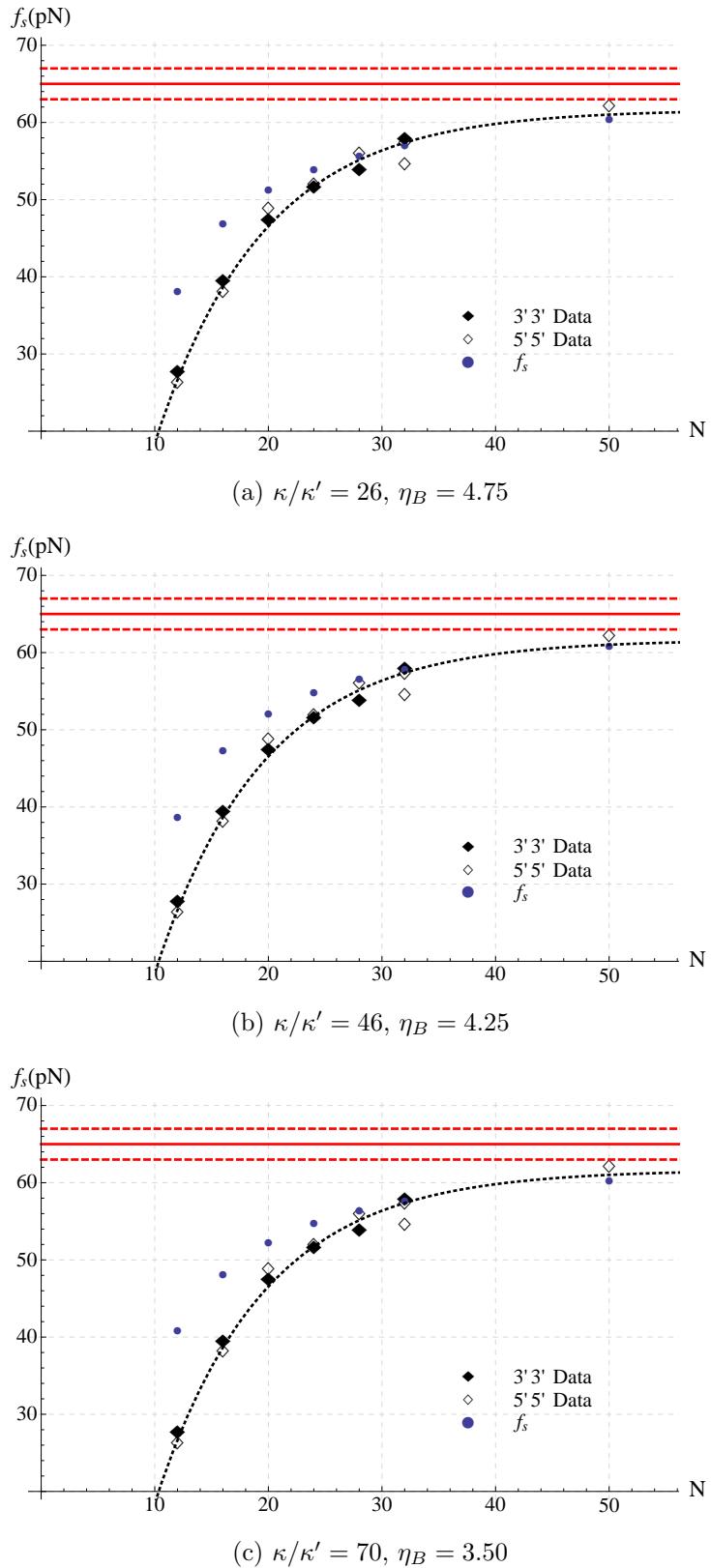


Figure 6.23: Shear force plots for some of the parameters that have close agreements with experimental data. The shear force was calculated from the first phase transition.

# Chapter 7

## Collagen Model

Having explored the use of the transfer integral method to model DNA, we now turn our attention to the collagen molecule. As with the model of DNA, the collagen model will focus on the shearing force applied to collagen molecule, and study the behaviour of the residue pairs under strain. A major structural difference between the DNA and collagen model is the greater dimensionality that arises from the addition of a third backbone in collagen. Our methodology in calculating the partition function will be similar to that of DNA, hence there will be some common features.

### 7.1 Collagen Model with a general residue potential

Starting with a Collagen Model that has a general residue potential we begin with the construction of a Hamiltonian.

#### 7.1.1 Construction of the Hamiltonian

The architecture of the collagen molecule is fairly complex with its triple helix structure. In order to mathematically express this as a Hamiltonian we simplify this structure to one that resembles a triangular prism. This structure retains the important composition of collagen such as the three backbones and residue pairs linking them. Once again we neglect the torsional effects in the triple helix structure and use a potential that is harmonic for small extensions in both the backbone and the pair potentials between the residues. The simplified structure consists of three rows of regularly spaced particles (residues) connected by elastic springs Fig. 7.1. The

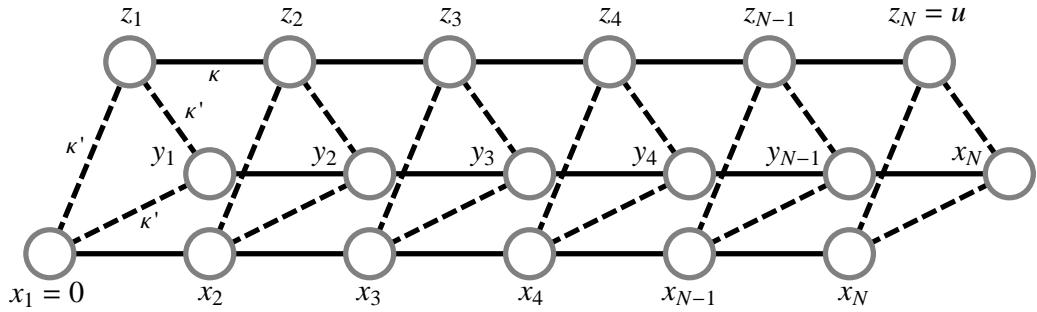


Figure 7.1: Schematic diagram of the collagen model. The thick lines represent the backbones with high spring constants  $\kappa$ , the dashed lines represent the residue pairs of low spring constant  $\kappa'$ .

backbone interactions have a spring constant  $\kappa$  whereas the residue pair potentials have a spring constant  $\kappa'$ . The load on the triangular prism structure is applied axially by pulling on one end of the structure while keeping the other end fixed in the following way. Each of the three backbones are labelled by  $x$ ,  $y$  and  $z$ , and the positions of particles in each backbone are labelled by  $x_i$ ,  $y_i$ , and  $z_i$ . Specifically, we fix the  $x$ -backbone at  $x_1 = 0$  with the opposite end of the structure being pulled on the  $z$ -backbone by an extension  $z_N = u$ . The  $y$ -backbone is free from any external constraints. The partners forming the  $i^{th}$  residue pair interact through a general potential  $V_r(\alpha)$  which is dependent on the axial pair separations of  $(x_i - y_i)$ ,  $(x_i - z_i)$  and  $(z_i - y_i)$ . For a model with residues at  $N$  axial positions the Hamiltonian can be written as

$$H = H_{bb_x} + H_{bb_y} + H_{bb_z} + H_r \quad (7.1)$$

where

$$H_{bb_x} = \sum_{i=1}^{N-1} \frac{1}{2} \kappa (x_{i+1} - x_i)^2 \quad (7.2)$$

$$H_{bb_y} = \sum_{i=1}^{N-1} \frac{1}{2} \kappa (y_{i+1} - y_i)^2 \quad (7.3)$$

$$H_{bb_z} = \sum_{i=1}^{N-1} \frac{1}{2} \kappa (z_{i+1} - z_i)^2 \quad (7.4)$$

$$H_r = \sum_{i=1}^N V_r(x_i - y_i) + V_r(x_i - z_i) + V_r(z_i - y_i) \quad (7.5)$$

The separation of variables in the interaction terms is a common 3-body problem in physics where transformations to different co-ordinate systems are used extensively.

Drawing similarities to our own Hamiltonian we can apply a transformation that decouples the residue interaction terms in a new co-ordinate system of  $R$ ,  $\rho$ , and  $\lambda$ . We make the following transformations

$$R_i = \frac{1}{\sqrt{3}}(x_i + y_i + z_i) \quad (7.6)$$

$$\rho_i = \frac{1}{\sqrt{2}}(x_i - y_i) \quad (7.7)$$

$$\lambda_i = \frac{1}{\sqrt{6}}(x_i + y_i - 2z_i) \quad (7.8)$$

In terms of the original co-ordinate system we have

$$x_i = \frac{R_i}{\sqrt{3}} + \frac{\rho_i}{\sqrt{2}} + \frac{\lambda_i}{\sqrt{6}} \quad (7.9)$$

$$y_i = \frac{R_i}{\sqrt{3}} - \frac{\rho_i}{\sqrt{2}} + \frac{\lambda_i}{\sqrt{6}} \quad (7.10)$$

$$z_i = \frac{R_i}{\sqrt{3}} - \sqrt{\frac{2}{3}}\lambda_i \quad (7.11)$$

Hence the new Hamiltonian may be written as a function of  $\vec{R}$ ,  $\vec{\rho}$  and  $\vec{\lambda}$

$$H = \sum_{i=1}^{N-1} \frac{\kappa}{2} (R_i - R_{i+1})^2 + \frac{\kappa}{2} (\rho_i - \rho_{i+1})^2 + \frac{\kappa}{2} (\lambda_i - \lambda_{i+1})^2 + \sum_{i=1}^N V_r \left( \sqrt{2}\rho_i \right) + V_r \left( \frac{\rho_i - \sqrt{3}\lambda}{\sqrt{2}} \right) + V_r \left( \frac{\sqrt{3}\lambda - \rho_i}{\sqrt{2}} \right) \quad (7.12)$$

### 7.1.2 Partition Function Integral

The partition function is evaluated using the same partition function integral method already employed in the DNA Model. As we have transformed the co-ordinate system to the new co-ordinates  $R$ ,  $\rho$  and  $\lambda$  we can ensure that integration over phase-space is also transformed correctly by calculating the Jacobian matrix of Eq. 7.9, Eq. 7.10 and Eq. 7.11. Transforming the integration over phase-space we have

$$\int d\tau = \iiint dx dy dz = \iiint J dR d\rho d\lambda \quad (7.13)$$

where

$$J(R, \rho, \lambda) = \begin{vmatrix} \frac{\partial x}{\partial R} & \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \lambda} \\ \frac{\partial y}{\partial R} & \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \lambda} \\ \frac{\partial z}{\partial R} & \frac{\partial z}{\partial \rho} & \frac{\partial z}{\partial \lambda} \end{vmatrix} = \begin{vmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \sqrt{\frac{2}{3}} \end{vmatrix} = 1 \quad (7.14)$$

With new variables converted into dimensionless quantities using

$$\begin{aligned} R &\rightarrow R \left( \frac{\kappa \beta}{2} \right)^{-\frac{1}{2}} \\ \rho &\rightarrow \rho \left( \frac{\kappa \beta}{2} \right)^{-\frac{1}{2}} \\ \lambda &\rightarrow \lambda \left( \frac{\kappa \beta}{2} \right)^{-\frac{1}{2}} \end{aligned} \quad (7.15)$$

the Hamiltonian becomes

$$\begin{aligned} \beta H &= \sum_{i=1}^{N-1} (R_i - R_{i+1})^2 + (\rho_i - \rho_{i+1})^2 + (\lambda_i - \lambda_{i+1})^2 \\ &+ \sum_{i=1}^N \beta V_r \left( \frac{2\rho_i}{(\beta \kappa)^{\frac{1}{2}}} \right) + \beta V_r \left( \frac{\rho_i + \sqrt{3}\lambda_i}{(\beta \kappa)^{\frac{1}{2}}} \right) + \beta V_r \left( \frac{\sqrt{3}\lambda_i - \rho_i}{(\beta \kappa)^{\frac{1}{2}}} \right) \end{aligned} \quad (7.16)$$

Making the variables dimensionless introduces a constant factor to the partition integral which we can ignore. The partition function integral before the imposition of boundary constraints becomes

$$Z = \int \prod_{i=1}^N dR_i d\rho_i d\lambda_i \exp(-\beta H(R_i, \rho_i, \lambda_i)) \quad (7.17)$$

### 7.1.3 Transfer Matrices

The transfer matrices are found by rewriting the exponential terms of the partition function integral. Inserting the Hamiltonian equation Eq. 7.16 into Eq. 7.17, the exponential terms become

$$\begin{aligned} \exp(-\beta H) &= \prod_{i=1}^{N-1} \exp(-(R_{i+1} - R_i)^2) \prod_{j=1}^{N-1} \exp(-(\rho_{j+1} - \rho_j)^2) \\ &\times \prod_{k=1}^{N-1} \exp(-(\lambda_{k+1} - \lambda_k)^2) \prod_{l=1}^N \exp(-V(\rho_l, \lambda_l)) \end{aligned} \quad (7.18)$$

Where

$$V(\rho_l, \lambda_l) = \beta V_r \left( \frac{2\rho_l}{(\beta\kappa)^{\frac{1}{2}}} \right) + \beta V_r \left( \frac{\rho_l + \sqrt{3}\lambda_l}{(\beta\kappa)^{\frac{1}{2}}} \right) + \beta V_r \left( \frac{\sqrt{3}\lambda_l - \rho_l}{(\beta\kappa)^{\frac{1}{2}}} \right) \quad (7.19)$$

The combination of the two variables in the potential  $V_r$  requires us to create a transfer matrix that is a function of both  $\rho$  and  $\lambda$ . By factorising the products and keeping the terms symmetric in  $R$ ,  $\rho$  and  $\lambda$ , the exponential terms can be written as

$$\begin{aligned} \exp(-\beta H) &= \prod_{i=1}^{N-1} \exp(-(R_{i+1} - R_i)^2) \\ &\times \prod_{j=1}^{N-1} \left[ \exp(-(\rho_{j+1} - \rho_j)^2) \exp(-(\lambda_{j+1} - \lambda_j)^2) \exp\left(-\frac{V(\rho_j, \lambda_j)}{2}\right) \right. \\ &\quad \left. \times \exp\left(-\frac{V(\rho_{j+1}, \lambda_{j+1})}{2}\right) \right] \exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \exp\left(-\frac{V(\rho_N, \lambda_N)}{2}\right) \end{aligned} \quad (7.20)$$

The similarities between the exponentials in Eq. 6.14 and Eq. 7.20 mean that we can use Eq. 6.15 for the transfer matrices in  $R$ . The two variable potential  $V(\rho, \lambda)$  requires a different type of transfer matrix that might be termed a general transfer hyper-matrix.

$$T_R(R, R') = \exp(-(R - R')^2) \quad (7.21)$$

$$\hat{T}_{\rho\lambda}(\rho, \lambda, \rho', \lambda') = \exp(-(\rho - \rho')^2) \exp(-(\lambda - \lambda')^2) \exp\left(-\left(\frac{V(\rho, \lambda) + V(\rho', \lambda')}{2}\right)\right) \quad (7.22)$$

The four variable  $\hat{T}_{\rho\lambda}$  transfer hyper-matrix arises because of the inseparable nature of the two variables  $\rho$  and  $\lambda$  when we consider a general potential  $V_r(\rho, \lambda)$ . We demonstrate in the following chapter that the terms do separate when we use a harmonic potential. Using these transfer matrices the partition function integral for a general residue pair potential becomes

$$\begin{aligned} Z &= \int \prod_{k=1}^N dR_k d\rho_k d\lambda_k \prod_{i=1}^{N-1} T_R(R_i, R_{i+1}) \prod_{j=1}^{N-1} \hat{T}_{\rho\lambda}(\rho_j, \lambda_j, \rho_{j+1}, \lambda_{j+1}) \\ &\quad \times \exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \exp\left(-\frac{V(\rho_N, \lambda_N)}{2}\right) \end{aligned} \quad (7.23)$$

Residue pair breakage descriptions introduced as  $g$  factors into the transfer matrices will not be included into the collagen model. We find that doing so makes the mathematics for contracting the transfer matrices extremely difficult. Instead we follow Peyrard et al. and describe breakage by using an anharmonic residue pair potential. The collagen model will be developed for an intact state.

### 7.1.4 Boundary Conditions

Delta functions are used to impose the external constraints on the collagen backbones. Because we have an ancillary  $y$  backbone in the model, a refinement is needed to the delta function at the fixed end where  $x_1 = 0$  that correctly describes the behaviour of a shearing force. In order to exploit the symmetry of the collagen structure both the  $x$ , and  $z$  backbones are extended by  $u/2$  using the delta functions  $\delta(x_1 + \frac{u}{2})$  and  $\delta(z_N - \frac{u}{2})$ .

$$\delta\left(x_1 + \frac{u}{2}\right) \delta\left(z_N - \frac{u}{2}\right) = \delta\left(\frac{R_1}{\sqrt{3}} + \frac{\rho_1}{\sqrt{2}} + \frac{\lambda_1}{\sqrt{6}} + \frac{u}{2}\right) \delta\left(\frac{R_N}{\sqrt{3}} - \sqrt{\frac{2}{3}}\lambda_N - \frac{u}{2}\right) \quad (7.24)$$

An additional integral over a delta function  $\int d\rho'_N \delta(\rho'_N - \rho_N)$  is introduced into the partition integral to aid the contraction of the  $\hat{T}_{\rho\lambda}$  transfer matrices. The full set of inserted boundary conditions now become

$$\int d\rho'_N \delta\left(\frac{R_1}{\sqrt{3}} + \frac{\rho_1}{\sqrt{2}} + \frac{\lambda_1}{\sqrt{6}} + \frac{u}{2}\right) \delta\left(\frac{R_N}{\sqrt{3}} - \sqrt{\frac{2}{3}}\lambda_N - \frac{u}{2}\right) \delta(\rho'_N - \rho_N) \quad (7.25)$$

which is to be inserted into the integrand in the expression for the partition function.

### 7.1.5 Eigenvalue Equations

From the Hamiltonian Eq. 7.12 we can draw some parallels with the DNA model. The backbone in both models have same harmonic behaviour in each strand determined by a stiffness constant  $\kappa$  leading to identical backbone transfer matrices. This allows us to use the same form of the eigenfunction used to contract the backbone transfer matrices in DNA. Differences arise in general potential terms where different eigenfunctions for the residue pairs need to be calculated. However, we will see later that for a harmonic residue pair potential we can in fact use the same form of the base-pair eigenfunction from DNA for the special case where  $\eta_b \rightarrow \infty$  for an intact structure.

The boundary condition delta functions are expanded as a set of normalised eigenfunctions of the transfer matrices satisfying equations Eq. 7.21, and Eq. 7.22. The expansion gives

$$\begin{aligned} \delta\left(\frac{R_1}{\sqrt{3}} + \frac{\rho_1}{\sqrt{2}} + \frac{\lambda_1}{\sqrt{6}} + \frac{u}{2}\right) &= \sum_{l_R} \psi_{l_R} \left( \frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}} \right) \psi_{l_R} \left( R_1 + \frac{u}{2}\sqrt{3} \right) \\ &= \sum_{l_R} \psi_{l_R} \left( \frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}} \right) \psi_{l_R}(R_1) \exp\left(\frac{il_R\sqrt{3}\pi u}{L}\right) \end{aligned} \quad (7.26)$$

$$\delta(\rho'_N - \rho_N) \delta\left(\frac{R_N}{\sqrt{3}} - \sqrt{\frac{2}{3}}\lambda_N - \frac{u}{2}\right) = \sum_{l_{\rho\lambda}} \hat{\psi}_{l_{\rho\lambda}}^*(\rho_N, \lambda_N) \hat{\psi}_{l_{\rho\lambda}} \left( \rho'_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}} \right) \quad (7.27)$$

where the eigenvalue equations for  $\psi_{l_R}$  and  $\hat{\psi}_{l_{\rho\lambda}}$  are

$$\int dR' T_R(R, R') \psi_{l_R}(R') = \Lambda_{l_R} \psi_{l_R}(R) \quad (7.28)$$

$$\iint d\rho' d\lambda' \hat{T}_{\rho\lambda}(\rho, \lambda, \rho', \lambda') \hat{\psi}_{l_{\rho\lambda}}(\rho', \lambda') = \hat{\Lambda}_{l_{\rho\lambda}} \hat{\psi}_{l_{\rho\lambda}}(\rho, \lambda) \quad (7.29)$$

For the eigenfunctions  $\psi_{l_R}$  the transfer matrix  $T_R(R, R')$  is identical to Eq. 6.15, and therefore we can use the analytical eigenfunctions calculated already,  $\psi_{l_R}(\alpha) = L^{-\frac{1}{2}} \exp(is2\pi\alpha/L)$  (Eq. 6.40). The transfer matrix in Eq. 7.29 is in fact an eigenproblem expressed in terms of an operator acting on a matrix rather than on a vector.

### 7.1.6 Transfer Matrix Method

Inserting the boundary conditions Eq. 7.26 and Eq. 7.27 to the partition function integral Eq. 7.23 we get

$$\begin{aligned}
Z(u) = & \int \prod_{i=1}^N dR_i d\rho_i d\lambda_i \prod_{j=1}^{N-1} T_R(R_j, R_{j+1}) \prod_{k=1}^{N-1} \hat{T}_{\rho\lambda}(\rho_k, \lambda_k, \rho_{k+1}, \lambda_{k+1}) \\
& \times \exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \exp\left(-\frac{V(\rho_N, \lambda_N)}{2}\right) \\
& \times \int d\rho'_N \sum_{l_R} \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right) \psi_{l_R}(R_1) \exp\left(\frac{il_R\sqrt{3}\pi u}{L}\right) \\
& \times \sum_{l_{\rho\lambda}} \hat{\psi}_{l_{\rho\lambda}}^*(\rho_N, \lambda_N) \hat{\psi}_{l_{\rho\lambda}}\left(\rho'_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right)
\end{aligned} \tag{7.30}$$

We can easily contract the  $T_R$  transfer matrices with the eigenfunctions  $\psi_{l_R}(R_1)$  up to  $\psi_{l_R}(R_N)$  to give

$$\begin{aligned}
Z(u) = & \sum_{l_R, l_{\rho\lambda}} \Lambda_{l_R}^{N-1} \exp\left(\frac{il_R\sqrt{3}\pi u}{L}\right) \int \prod_{i=1}^N d\rho_i d\lambda_i \prod_{j=1}^{N-1} \hat{T}_{\rho\lambda}(\rho_j, \lambda_j, \rho_{j+1}, \lambda_{j+1}) \\
& \times \iint dR_N d\rho'_N \psi_{l_R}(R_N) \exp\left(-\frac{V(\rho'_N, \lambda_N)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}\left(\rho'_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \\
& \times \hat{\psi}_{l_{\rho\lambda}}^*(\rho_N, \lambda_N) \exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right)
\end{aligned} \tag{7.31}$$

We can use the eigenfunction  $\hat{\psi}_{l_{\rho\lambda}}^*(\rho_N, \lambda_N)$  to contract the  $\hat{T}_{\rho\lambda}$  transfer matrices. The delta function introduced in Eq. 7.27 allow us to write  $\rho_N = \rho'_N$  and  $\lambda_N = \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}$  in  $V(\rho_N, \lambda_N)$ .

$$\begin{aligned}
Z(u) = & \sum_{l_R, l_{\rho\lambda}} \Lambda_{l_R}^{N-1} \hat{\Lambda}_{l_{\rho\lambda}}^{N-1} \exp\left(\frac{il_R\sqrt{3}\pi u}{L}\right) \\
& \times \iint dR_N d\rho'_N \psi_{l_R}(R_N) \exp\left(-\frac{V\left(\rho'_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}\left(\rho'_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \\
& \times \iint d\rho_1 d\lambda_1 \hat{\psi}_{l_{\rho\lambda}}^*(\rho_1, \lambda_1) \exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right)
\end{aligned} \tag{7.32}$$

Here we make a change of variable such that

$$\tilde{R}_N = \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}} \quad (7.33)$$

Ignoring the constants outside the sum the partition function now becomes

$$\begin{aligned} Z(u) &= \sum_{l_R, l_{\rho\lambda}} \Lambda_{l_R}^{N-1} \hat{\Lambda}_{l_{\rho\lambda}}^{N-1} \exp\left(\frac{il_R 2\sqrt{3}\pi u}{L}\right) \\ &\times \iint d\tilde{R}_N d\rho'_N \psi_{l_R}(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{V(\rho'_N, \tilde{R}_N)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}(\rho'_N, \tilde{R}_N) \\ &\times \iint d\rho_1 d\lambda_1 \hat{\psi}_{l_{\rho\lambda}}^*(\rho_1, \lambda_1) \exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right) \end{aligned} \quad (7.34)$$

### 7.1.7 Mean Axial Displacement

The calculation of the mean axial displacement of various components in the collagen model is rather different to DNA where it was the average of the base-pair ( $x_i - y_i$ ) separation. The third backbone in the collagen model adds another variable to the mean axial displacement calculation where the average axial separation is now the difference between the mid-point of the  $xz_i$  residue pair and the position of the  $y_i$  residue in the  $i^{th}$  segment of the structure, that is,  $\frac{x_i+z_i}{2} - y_i$ . Let us define the mean axial displacement to be  $\zeta$ , where

$$\langle \zeta_\alpha \rangle = \langle x_\alpha + z_\alpha - 2y_\alpha \rangle \quad (7.35)$$

and  $\alpha$  is a segment label in the range 1 to  $N$ . Transforming this to the  $R\rho\lambda$  coordinate system we use Eq. 7.9, Eq. 7.10 and Eq. 7.11 to get

$$\langle \zeta_\alpha \rangle = \frac{-\sqrt{3}\langle \lambda_\alpha \rangle + 3\langle \rho_\alpha \rangle}{\sqrt{\beta\kappa}} \quad (7.36)$$

We therefore need to calculate  $\langle \lambda_\alpha \rangle$  and  $\langle \rho_\alpha \rangle$  to find the mean axial displacements for the collagen model. These equations are similar to Eq. 6.83 where

$$\langle \rho_\alpha \rangle = \frac{\int \prod_{i=1}^N dR_i d\rho_i d\lambda_i \rho_\alpha \exp(-\beta H(R_i, \rho_i, \lambda_i)) B_c}{\int \prod_{i=1}^N dR_i d\rho_i d\lambda_i \exp(-\beta H(R_i, \rho_i, \lambda_i)) B_c} \quad (7.37)$$

and

$$\langle \lambda_\alpha \rangle = \frac{\int \prod_{i=1}^N dR_i d\rho_i d\lambda_i \lambda_\alpha \exp(-\beta H(R_i, \rho_i, \lambda_i)) B_c}{\int \prod_{i=1}^N dR_i d\rho_i d\lambda_i \exp(-\beta H(R_i, \rho_i, \lambda_i)) B_c} \quad (7.38)$$

The boundary conditions that apply constraints to the collagen structure from Eq. 7.25 are contained within  $B_c$ . With the partition function already calculated we are left with evaluating the following two integrals to find the mean axial displacement

$$Z_{\rho_\alpha} = \iiint \prod_{i=1}^N dR_i d\rho_i d\lambda_i \rho_\alpha \exp(-\beta H(R_i, \rho_i, \lambda_i)) B_c \quad (7.39)$$

$$Z_{\lambda_\alpha} = \iiint \prod_{i=1}^N dR_i d\rho_i d\lambda_i \lambda_\alpha \exp(-\beta H(R_i, \rho_i, \lambda_i)) B_c \quad (7.40)$$

### 7.1.8 Calculation of average $\rho_\alpha$ and $\lambda_\alpha$

We begin with  $Z_{\lambda_\alpha}$ , and continue from Eq. 7.30 where the transfer matrices in  $R$  have already been contracted. A modification of this partition function is needed to include the variable  $\lambda_\alpha$

$$\begin{aligned} Z_{\lambda_\alpha}(u) &= \sum_{l_R, l_{\rho\lambda}} \Lambda_{l_R}^{N-1} \exp\left(\frac{i l_R 2\sqrt{3}\pi u}{L}\right) \iint \prod_{i=1}^N d\rho_i d\lambda_i \hat{T}_{\rho\lambda}(\rho_1, \lambda_1, \rho_2, \lambda_2) \dots \\ &\times \hat{T}_{\rho\lambda}(\rho_{\alpha-1}, \lambda_{\alpha-1}, \rho_\alpha, \lambda_\alpha) \lambda_\alpha \hat{T}_{\rho\lambda}(\rho_\alpha, \lambda_\alpha, \rho_{\alpha+1}, \lambda_{\alpha+1}) \dots \hat{T}_{\rho\lambda}(\rho_{N-1}, \lambda_{N-1}, \rho_N, \lambda_N) \\ &\times \iint d\tilde{R}_N d\rho'_N \psi_{l_R}(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{V(\rho'_N, \tilde{R}_N)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}(\rho'_N, \tilde{R}_N) \\ &\times \hat{\psi}_{l_{\rho\lambda}}^*(\rho_N, \lambda_N) \exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right) \end{aligned} \quad (7.41)$$

We can start from the eigenfunction  $\hat{\psi}_{l_{\rho\lambda}}^*(\rho_N, \lambda_N)$  to contract the  $\hat{T}_{\rho\lambda}$  transfer matrices down to  $\hat{\psi}_{l_{\rho\lambda}}^*(\rho_\alpha, \lambda_\alpha)$

$$\begin{aligned} Z_{\lambda_\alpha}(u) &= \sum_{l_R, l_{\rho\lambda}} \Lambda_{l_R}^{N-1} \Lambda_{l_{\rho\lambda}}^{N-\alpha} \exp\left(\frac{i2\sqrt{3}l_R\pi u}{L}\right) \\ &\times \iint d\tilde{R}_N d\rho'_N \psi_{l_R}(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{V(\rho'_N, \tilde{R}_N)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}(\rho'_N, \tilde{R}_N) \\ &\times \iint \prod_{i=1}^{\alpha} d\rho_i d\lambda_i \hat{T}_{\rho\lambda}(\rho_1, \lambda_1, \rho_2, \lambda_2) \dots \hat{T}_{\rho\lambda}(\rho_{\alpha-1}, \lambda_{\alpha-1}, \rho_\alpha, \lambda_\alpha) \lambda_\alpha \hat{\psi}_{l_{\rho\lambda}}^*(\rho_\alpha, \lambda_\alpha) \\ &\times \exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right) \end{aligned} \quad (7.42)$$

In order to contract the remaining transfer matrices we use

$$\exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right) = \sum_{l'_{\rho\lambda}}^N \Xi_{l'_{\rho\lambda}}^{l_R} \hat{\psi}_{l'_{\rho\lambda}}(\rho_1, \lambda_1) \quad (7.43)$$

where

$$\Xi_{l'_{\rho\lambda}}^{l_R} = \iint d\rho'_1 d\lambda'_1 \hat{\psi}_{l'_{\rho\lambda}}^*(\rho'_1, \lambda'_1) \exp\left(-\frac{V(\rho'_1, \lambda'_1)}{2}\right) \psi_{l_R}\left(\frac{\lambda'_1 + \sqrt{3}\rho'_1}{\sqrt{2}}\right) \quad (7.44)$$

using the orthonormality of the  $\hat{\psi}_{l_{\rho\lambda}}(\rho_1, \lambda_1)$  Eq. 7.87 becomes

$$\begin{aligned} Z_{\lambda_\alpha}(u) &= \sum_{l_R, l_{\rho\lambda}, l'_{\rho\lambda}} \Xi_{l'_{\rho\lambda}}^{l_R} \Lambda_{l_R}^{N-1} \Lambda_{l_{\rho\lambda}}^{N-\alpha} \exp\left(\frac{il_R 2\sqrt{3}\pi u}{L}\right) \\ &\times \iint d\tilde{R}_N d\rho'_N \psi_{l_R}(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{V(\rho'_N, \tilde{R}_N)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}(\rho'_N, \tilde{R}_N) \\ &\times \iint \prod_{i=1}^{\alpha} d\rho_i d\lambda_i \hat{\psi}_{l'_{\rho\lambda}}(\rho_1, \lambda_1) \hat{T}_{\rho\lambda}(\rho_1, \lambda_1, \rho_2, \lambda_2) \dots \hat{T}_{\rho\lambda}(\rho_{\alpha-1}, \lambda_{\alpha-1}, \rho_\alpha, \lambda_\alpha) \lambda_\alpha \hat{\psi}_{l_{\rho\lambda}}^*(\rho_\alpha, \lambda_\alpha) \end{aligned} \quad (7.45)$$

Contracting the remaining transfer matrices we get the final result for  $Z_{\lambda_\alpha}$

$$\begin{aligned} Z_{\lambda_\alpha}(u) &= \sum_{l_R, l_{\rho\lambda}, l'_{\rho\lambda}} \Xi_{l'_{\rho\lambda}}^{l_R} \Lambda_{l_R}^{N-1} \Lambda_{l_{\rho\lambda}}^{N-\alpha} \Lambda_{l'_{\rho\lambda}}^{\alpha-1} \exp\left(\frac{il_R 2\sqrt{3}\pi u}{L}\right) \\ &\times \iint d\tilde{R}_N d\rho'_N \psi_{l_R}(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{V(\rho'_N, \tilde{R}_N)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}(\rho'_N, \tilde{R}_N) \\ &\times \iint d\rho_\alpha d\lambda_\alpha \hat{\psi}_{l'_{\rho\lambda}}(\rho_\alpha, \lambda_\alpha) \lambda_\alpha \hat{\psi}_{l_{\rho\lambda}}^*(\rho_\alpha, \lambda_\alpha) \end{aligned} \quad (7.46)$$

Since  $\rho$  and  $\lambda$  appear similarly in the hyper-matrix we can interchange  $\lambda_\alpha$  and  $\rho_\alpha$  to give the result for  $Z_{\rho_\alpha}$

$$\begin{aligned} Z_{\rho_\alpha}(u) &= \sum_{l_R, l_{\rho\lambda}, l'_{\rho\lambda}} \Xi_{l'_{\rho\lambda}}^{l_R} \Lambda_{l_R}^{N-1} \Lambda_{l_{\rho\lambda}}^{N-\alpha} \Lambda_{l'_{\rho\lambda}}^{\alpha-1} \exp\left(\frac{il_R 2\sqrt{3}\pi u}{L}\right) \\ &\times \iint d\tilde{R}_N d\rho'_N \psi_{l_R}(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{V(\rho'_N, \tilde{R}_N)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}(\rho'_N, \tilde{R}_N) \\ &\times \iint d\rho_\alpha d\lambda_\alpha \hat{\psi}_{l'_{\rho\lambda}}(\rho_\alpha, \lambda_\alpha) \rho_\alpha \hat{\psi}_{l_{\rho\lambda}}^*(\rho_\alpha, \lambda_\alpha) \end{aligned} \quad (7.47)$$

Using Eq. 7.47 and Eq. 7.46 we are now able to calculate the mean values of  $\lambda_\alpha$  and  $\rho_\alpha$ , and ultimately the mean axial displacements, through

$$\langle \lambda_\alpha \rangle = \frac{Z_{\lambda_\alpha}(u)}{Z(u)} \quad (7.48)$$

$$\langle \rho_\alpha \rangle = \frac{Z_{\rho_\alpha}(u)}{Z(u)} \quad (7.49)$$

### 7.1.9 Average residue positions and the calculation of average $R_\alpha$

The transformations used to in Eq. 7.6 to decouple the cartesian variables requires the additional calculation of  $\langle R_\alpha \rangle$  to evaluate the average positions of each residue. We begin with the calculation of the partition function integral

$$Z_{R_\alpha} = \iiint \prod_{i=1}^N dR_i d\rho_i d\lambda_i R_\alpha \exp(-\beta H(R_i, \rho_i, \lambda_i)) B_c \quad (7.50)$$

which is a modification of Eq. 7.30 that includes the extra term  $R_\alpha$ . Using the delta function Eq. 7.24 we also make  $\lambda_N = \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}$  in the exponential potential term to give

$$\begin{aligned}
Z_{R_\alpha}(u) &= \int \prod_{i=1}^N dR_i d\rho_i d\lambda_i \prod_{k=1}^{N-1} \hat{T}_{\rho\lambda}(\rho_k, \lambda_k, \rho_{k+1}, \lambda_{k+1}) \\
&\times T_R(R_1, R_2) \dots T_R(R_{\alpha-1}, R_\alpha) R_\alpha T_R(R_\alpha, R_{\alpha+1}) \dots T_R(R_{N-1}, R_N) \\
&\times \int d\rho'_N \sum_{l_R} \psi_{l_R} \left( \frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}} \right) \exp \left( -\frac{V(\rho_1, \lambda_1)}{2} \right) \psi_{l_R}(R_1) \exp \left( \frac{il_R\sqrt{3}\pi u}{L} \right) \\
&\times \sum_{l_{\rho\lambda}} \exp \left( -\frac{V \left( \rho_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}} \right)}{2} \right) \hat{\psi}_{l_{\rho\lambda}} \left( \rho'_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}} \right) \hat{\psi}_{l_{\rho\lambda}}^*(\rho_N, \lambda_N)
\end{aligned} \tag{7.51}$$

Here we are able to contract the  $\hat{T}_{\rho\lambda}$  transfer matrices using the eigenfunction  $\hat{\psi}_{l_{\rho\lambda}}^*(\rho_N, \lambda_N)$ .

$$\begin{aligned}
Z_{R_\alpha}(u) &= \sum_{l_R, l_{\rho\lambda}} \Lambda_{l_{\rho\lambda}}^{N-1} \exp \left( \frac{il_R\sqrt{3}\pi u}{L} \right) \\
&\times \int \prod_{i=1}^N dR_i T_R(R_1, R_2) \dots T_R(R_{\alpha-1}, R_\alpha) R_\alpha T_R(R_\alpha, R_{\alpha+1}) \dots T_R(R_{N-1}, R_N) \\
&\times \int d\rho'_N \psi_{l_R}(R_1) \exp \left( -\frac{V \left( \rho_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}} \right)}{2} \right) \hat{\psi}_{l_{\rho\lambda}} \left( \rho'_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}} \right) \\
&\times \iint d\rho_1 d\lambda_1 \psi_{l_R} \left( \frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}} \right) \exp \left( -\frac{V(\rho_1, \lambda_1)}{2} \right) \hat{\psi}_{l_{\rho\lambda}}^*(\rho_1, \lambda_1)
\end{aligned} \tag{7.52}$$

The first set of contractions of the  $T_R$  transfer matrices can be easily accomplished up to  $R_\alpha$  using  $\psi_{l_R}(R_1)$ .

$$\begin{aligned}
Z_{R_\alpha}(u) &= \sum_{l_R, l_{\rho\lambda}} \Lambda_{l_R}^{\alpha-1} \Lambda_{l_{\rho\lambda}}^{N-1} \exp\left(\frac{il_R\sqrt{3}\pi u}{L}\right) \\
&\times \int \prod_{i=\alpha}^N dR_i \psi_{l_R}(R_\alpha) R_\alpha T_R(R_\alpha, R_{\alpha+1}) \dots T_R(R_{N-1}, R_N) \\
&\times \int d\rho'_N \exp\left(-\frac{V\left(\rho_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}\left(\rho'_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \\
&\times \iint d\rho_1 d\lambda_1 \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}^*(\rho_1, \lambda_1) \quad (7.53)
\end{aligned}$$

To contract the remaining  $T_R$  transfer matrices we impose  $\rho = \rho'$  as allowed by the delta function Eq. 7.27, and proceed with an expansion of the  $\hat{\psi}_{l_{\rho\lambda}}$  eigenfunction in terms of  $\psi_{l_R}$ . We find that

$$\int d\rho'_N \exp\left(-\frac{V\left(\rho'_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}\left(\rho'_N, \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) = \sum_{l'_R} \Theta_{l'_R}^{l_{\rho\lambda}}(u) \psi_{l'_R}^*(R_N) \quad (7.54)$$

where

$$\Theta_{l'_R}^{l_{\rho\lambda}}(u) = \iint d\rho'_N dR'_N \psi_{l'_R}(R'_N) \exp\left(-\frac{V\left(\rho'_N, \frac{R'_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}\left(\rho'_N, \frac{R'_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \quad (7.55)$$

To evaluate this integral we make a change of variable such that

$$\tilde{R}_N = \frac{R'_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}} \quad (7.56)$$

to give

$$\begin{aligned}
\Theta_{l'_R}^{l_{\rho\lambda}}(u) &= \exp\left(\frac{il'_R\sqrt{3}\pi u}{L}\right) \iint d\rho'_N dR'_N \psi_{l'_R}(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{V\left(\rho'_N, \tilde{R}_N\right)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}\left(\rho'_N, \tilde{R}_N\right) \\
&= \tilde{\Theta}_{l'_R}^{l_{\rho\lambda}} \exp\left(\frac{il'_R\sqrt{3}\pi u}{L}\right) \quad (7.57)
\end{aligned}$$

Combining Eq. 7.54 and Eq. 7.57 with Eq. 7.53 the partition function integral becomes

$$\begin{aligned} Z_{R_\alpha}(u) &= \sum_{l_R, l_{\rho\lambda}, l'_R} \Lambda_{l_R}^{\alpha-1} \Lambda_{l_{\rho\lambda}}^{N-1} \exp\left(\frac{i l_R \sqrt{3}\pi u}{L}\right) \tilde{\Theta}_{l'_R}^{l_{\rho\lambda}} \exp\left(\frac{i l'_R \sqrt{3}\pi u}{L}\right) \\ &\times \int \prod_{i=\alpha}^N dR_i \psi_{l_R}(R_\alpha) R_\alpha T_R(R_\alpha, R_{\alpha+1}) \dots T_R(R_{N-1}, R_N) \psi_{l'_R}^*(R_N) \\ &\times \iint d\rho_1 d\lambda_1 \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}^*(\rho_1, \lambda_1) \quad (7.58) \end{aligned}$$

The contraction of the remaining transfer matrices gives us a final expression of  $Z_{R_\alpha}$

$$\begin{aligned} Z_{R_\alpha}(u) &= \sum_{l_R, l_{\rho\lambda}, l'_R} \Lambda_{l'_R}^{N-\alpha} \Lambda_{l_R}^{\alpha-1} \Lambda_{l_{\rho\lambda}}^{N-1} \tilde{\Theta}_{l'_R}^{l_{\rho\lambda}} \exp\left(\frac{i(l_R + l'_R)\sqrt{3}\pi u}{L}\right) \\ &\times \int dR_\alpha \psi_{l_R}(R_\alpha) R_\alpha \psi_{l'_R}^*(R_\alpha) \\ &\times \iint d\rho_1 d\lambda_1 \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(-\frac{V(\rho_1, \lambda_1)}{2}\right) \hat{\psi}_{l_{\rho\lambda}}^*(\rho_1, \lambda_1) \quad (7.59) \end{aligned}$$

This expression for  $Z_{R_\alpha}$  allows us to evaluate the average separation of each residue pair in the collagen structure for a general residue pair potential. Using these average quantities we are also able to illustrate how each residue distorts with an applied extension by transforming the variables back into their original *xyz* co-ordinate system. We now consider a simpler case where the residue pair potential is harmonic in the collagen structure.

## 7.2 Collagen Model with a Harmonic Potential

The calculation of the partition function in DNA highlighted a special case when using a harmonic base-pair potential in the limit of  $\eta_B \rightarrow \infty$ . We were able to successfully separate the cross terms in the Hamiltonian for the backbones and the base-pair potentials. By employing a similar harmonic potential for the residue pairs,  $V(\alpha_i) = \frac{1}{2}\kappa'\alpha_i^2$ , we find that all the cross terms separate in Hamiltonian for the collagen model. Starting from Eq. 7.16, we focus on the residue pair interaction

of the Hamiltonian  $H_r$  and insert the Harmonic residue potential such that  $V_r = V$ ,

$$\begin{aligned}\beta H_r &= \sum_{i=1}^N \beta V \left( \frac{2\rho_i}{(\beta\kappa)^{\frac{1}{2}}} \right) + \beta V \left( \frac{\rho_i + \sqrt{3}\lambda_i}{(\beta\kappa)^{\frac{1}{2}}} \right) + \beta V \left( \frac{\sqrt{3}\lambda_i - \rho_i}{(\beta\kappa)^{\frac{1}{2}}} \right) \\ &= \sum_{i=1}^N \frac{2\kappa'}{\kappa} \rho_i^2 + \frac{\kappa'}{2\kappa} \left[ (\sqrt{3}\lambda_i - \rho_i)^2 + (\sqrt{3}\lambda_i + \rho_i)^2 \right] \\ &= \sum_{i=1}^N \frac{3\kappa'}{\kappa} \rho_i^2 + \frac{3\kappa'}{\kappa} \lambda_i^2\end{aligned}\quad (7.60)$$

Expressing the terms in this part of the Hamiltonian back as a function of  $V$ , the full Hamiltonian may be written as

$$\begin{aligned}\beta H(R, \rho, \lambda) &= \sum_{i=1}^{N-1} (R_i - R_{i+1})^2 + (\rho_i - \rho_{i+1})^2 + (\lambda_i - \lambda_{i+1})^2 \\ &\quad + \sum_{j=1}^N \frac{6V(\rho_j)}{\kappa} + \frac{6V(\lambda_j)}{\kappa}\end{aligned}\quad (7.61)$$

and the partition function including the boundary conditions from Eq. 7.25 indicated by the parameter  $B_c$  becomes

$$Z = \iiint \prod_{i=1}^N dR_i d\rho_i d\lambda_i \exp(-\beta H(R_i, \rho_i, \lambda_i)) B_c \quad (7.62)$$

### 7.2.1 Transfer Matrices

In the partition function integral we consider the exponential of the Hamiltonian and group like terms together, in such a way that they are symmetric in  $R$ ,  $\rho$  and  $\lambda$ . We have

$$\begin{aligned}\exp(-\beta H) &= \prod_{i=1}^{N-1} \exp(-(R_{i+1} - R_i)^2) \\ &\quad \times \prod_{j=1}^{N-1} \exp \left( -(\rho_{j+1} - \rho_j)^2 - \left( \frac{3V(\rho_{j+1}) + 3V(\rho_j)}{\kappa} \right) \right) \\ &\quad \times \prod_{k=1}^{N-1} \exp \left( -(\lambda_{k+1} - \lambda_k)^2 - \left( \frac{3V(\lambda_{k+1}) + 3V(\lambda_k)}{\kappa} \right) \right) \\ &\quad \times \exp \left( -\frac{3V(\rho_1) + 3V(\rho_N)}{\kappa} \right) \exp \left( -\frac{3V(\lambda_1) + 3V(\lambda_N)}{\kappa} \right)\end{aligned}\quad (7.63)$$

The similarities between the exponentials in Eq. 6.14 and Eq. 7.63 allow us to use Eq. 6.15 for the transfer matrices in  $R$ , and introduce transfer matrices that resemble Eq. 6.16 for  $\rho$  and  $\lambda$ . The transfer matrices are

$$T_R(R, R') = \exp\left(-(R - R')^2\right) \quad (7.64)$$

$$\hat{T}_\rho(\rho, \rho') = \exp\left(-(\rho - \rho')^2\right) \exp\left(-\frac{3V(\rho) + 3V(\rho')}{\kappa}\right) \quad (7.65)$$

$$\hat{T}_\lambda(\lambda, \lambda') = \exp\left(-(\lambda - \lambda')^2\right) \exp\left(-\frac{3V(\lambda) + 3V(\lambda')}{\kappa}\right) \quad (7.66)$$

Even for a harmonic case the mathematics involved in the collagen model does not easily allow us to introduce  $g$  factors to the transfer matrices in  $\rho$ , and  $\lambda$  to study base-pair breakage. We will continue for a collagen model that is Intact for a harmonic potential. Representing the exponentials with transfer matrices in Eq. 7.63 the partition function integral becomes

$$\begin{aligned} Z = & \int \prod_{i=1}^N dR_i \left( \prod_{i=1}^{N-1} T_R(R_i, R_{i+1}) \right) \\ & \times \int \prod_{j=1}^N d\rho_j \left( \prod_{j=1}^{N-1} \hat{T}_\rho(\rho_j, \rho_{j+1}) \right) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \exp\left(-\frac{3V(\rho_N)}{\kappa}\right) \\ & \times \int \prod_{k=1}^N d\lambda_k \left( \prod_{k=1}^{N-1} \hat{T}_\lambda(\lambda_k, \lambda_{k+1}) \right) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(-\frac{3V(\lambda_N)}{\kappa}\right) B_c \end{aligned} \quad (7.67)$$

### 7.2.2 Eigenvalue Equations

The boundary condition delta functions are expanded as a set of eigenfunctions of the transfer matrices satisfying the transfer matrix equations Eq. 7.64, Eq. 7.65 and Eq. 7.66. The expansion gives

$$\begin{aligned} \delta\left(\frac{R_1}{\sqrt{3}} + \frac{\rho_1}{\sqrt{2}} + \frac{\lambda_1}{\sqrt{6}} + \frac{u}{2}\right) &= \sum_{l_R} \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right) \psi_{l_R}\left(R_1 + \frac{u}{2}\sqrt{3}\right) \\ &= \sum_{l_R} \psi_{l_R}\left(\frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}}\right) \psi_{l_R}(R_1) \exp\left(\frac{il_R\sqrt{3}\pi u}{L}\right) \end{aligned} \quad (7.68)$$

$$\delta\left(\frac{R_N}{\sqrt{3}} - \sqrt{\frac{2}{3}}\lambda_N - \frac{u}{2}\right) = \sum_{l_\lambda} \hat{\psi}_{l_\lambda}^*(\lambda_N) \hat{\psi}_{l_\lambda}\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \quad (7.69)$$

$$\delta(\rho_N - \rho'_N) = \sum_{l_\rho} \hat{\psi}_{l_\rho}^*(\rho'_N) \hat{\psi}_{l_\rho}(\rho_N) \quad (7.70)$$

where the eigenvalue equations for  $R, \rho$  and  $\lambda$  are

$$\int dR' T_R(R, R') \psi_{l_R}(R') = \lambda_{l_R} \psi_{l_R}(R) \quad (7.71)$$

$$\int d\rho' \hat{T}_\rho(\rho, \rho') \hat{\psi}_{l_\rho}(\rho') = \hat{\lambda}_{l_\rho} \hat{\psi}_{l_\rho}(\rho) \quad (7.72)$$

$$\int d\lambda' \hat{T}_\lambda(\lambda, \lambda') \hat{\psi}_{l_\lambda}(\lambda') = \hat{\lambda}_{l_\lambda} \hat{\psi}_{l_\lambda}(\lambda) \quad (7.73)$$

For the eigenfunctions  $\psi_{l_R}$  the transfer matrix  $T_R(R, R')$  is identical to Eq. 6.15 and therefore we can use the analytical eigenfunction already calculated Eq. 6.40. For the identical  $\hat{T}_\rho(\rho, \rho')$  and  $\hat{T}_\lambda(\lambda, \lambda')$  transfer matrices we can use the eigenfunctions of the  $\hat{T}$  matrix already calculated for DNA in the limit of  $\eta_B \rightarrow \infty$ . Using these equations we now have analytical expressions for the eigenvalues and eigenfunctions for the transfer matrix equations in collagen.

### 7.2.3 Transfer Matrix Method

Inserting the boundary conditions into the partition function integral Eq. 7.67 we are able to use the eigenvalue equations to contract the transfer matrices in  $R$ ,  $\rho$  and  $\lambda$ . We start from the partition function integral with the explicit representation of the boundary conditions

$$\begin{aligned} Z(u) &= \int \prod_{i=1}^N dR_i \left( \prod_{i=1}^{N-1} T_R(R_i, R_{i+1}) \right) \\ &\times \int \prod_{j=1}^N d\rho_j \left( \prod_{j=1}^{N-1} \hat{T}_\rho(\rho_j, \rho_{j+1}) \right) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \exp\left(-\frac{3V(\rho_N)}{\kappa}\right) \\ &\times \int \prod_{k=1}^N d\lambda_k \left( \prod_{k=1}^{N-1} \hat{T}_\lambda(\lambda_k, \lambda_{k+1}) \right) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(-\frac{3V(\lambda_N)}{\kappa}\right) \\ &\times \sum_{l_R} \psi_{l_R} \left( \frac{\lambda_1 + \sqrt{3}\rho_1}{\sqrt{2}} \right) \psi_{l_R}(R_1) \exp\left(\frac{il_R\sqrt{3}\pi u}{L}\right) \sum_{l_\lambda} \hat{\psi}_{l_\lambda}^*(\lambda_N) \hat{\psi}_{l_\lambda} \left( \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}} \right) \\ &\times \int d\rho'_N \sum_{l_\rho} \hat{\psi}_{l_\rho}^*(\rho'_N) \hat{\psi}_{l_\rho}(\rho_N) \end{aligned} \quad (7.74)$$

Starting with the transfer matrices in  $R$  we have

$$\begin{aligned}
Z(u) &= \sum_{l_R, l_\lambda, l_\rho} \exp\left(\frac{il_R\pi\sqrt{3}u}{L}\right) \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(\frac{il_R2\pi\lambda_1}{\sqrt{2}L}\right) \int d\rho'_N \hat{\psi}_{l_\rho}^*(\rho'_N) \hat{\psi}_{l_\rho}(\rho_N) \\
&\times \int \prod_{i=1}^N dR_i \psi_{l_R}(R_1) T_R(R_1, R_2) T_R(R_3, R_4) \dots T_R(R_{N-1}, R_N) \hat{\psi}_{l_\lambda}\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \\
&\times \iint \prod_{i=1}^N d\rho_i d\lambda_i \left( \prod_{i=1}^{N-1} \hat{T}_\rho(\rho_i, \rho_{i+1}) \right) \left( \prod_{i=1}^{N-1} \hat{T}_\lambda(\lambda_i, \lambda_{i+1}) \right) \hat{\psi}_{l_\lambda}^*(\lambda_N) \\
&\times \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(-\frac{3V(\rho_N)}{\kappa}\right) \exp\left(-\frac{3V(\lambda_N)}{\kappa}\right)
\end{aligned} \tag{7.75}$$

Notice how we use the full analytical solution of  $\psi_{l_R}$  to separate the  $\rho_1$  and  $\lambda_1$  terms in the eigenfunction  $\psi_{l_R}^*$  of Eq. 7.74. Following from Eq. 7.75, using Eq. 7.71 we contract the  $T_R$  transfer matrices with the eigenfunction  $\psi_{l_R}(R_1)$  up to  $\psi_{l_R}(R_N)$  leaving an evaluation of a single integral. The partition function integral becomes

$$\begin{aligned}
Z(u) &= \sum_{l_R, l_\lambda, l_\rho} \Lambda_{l_R}^{N-1} \exp\left(\frac{il_R\pi\sqrt{3}u}{L}\right) \int dR_N \psi_{l_R}(R_N) \hat{\psi}_{l_\lambda}\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \\
&\times \iint \prod_{i=1}^N d\rho_i d\lambda_i \prod_{i=1}^{N-1} \hat{T}_\rho(\rho_i, \rho_{i+1}) \hat{\psi}_{l_\rho}(\rho_N) \prod_{i=1}^{N-1} \hat{T}_\lambda(\lambda_i, \lambda_{i+1}) \hat{\psi}_{l_\lambda}^*(\lambda_N) \\
&\times \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(-\frac{3V(\rho_N)}{\kappa}\right) \exp\left(-\frac{3V(\lambda_N)}{\kappa}\right) \\
&\times \int d\rho'_N \hat{\psi}_{l_\rho}^*(\rho'_N) \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(\frac{il_R2\pi\lambda_1}{\sqrt{2}L}\right)
\end{aligned} \tag{7.76}$$

To contract the transfer matrices in  $\lambda$  we use the eigenfunction  $\hat{\psi}_{l_\lambda}^*(\lambda_N)$

$$\begin{aligned}
 Z(u) &= \sum_{l_R, l_\lambda, l_\rho} \Lambda_{l_R}^{N-1} \exp\left(\frac{il_R\pi\sqrt{3}u}{L}\right) \int dR_N \psi_{l_R}(R_N) \hat{\psi}_{l_\lambda}\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \\
 &\quad \times \int \prod_{i=1}^N d\lambda_i \hat{T}_\lambda(\lambda_1, \lambda_2) \hat{T}_\lambda(\lambda_2, \lambda_3) \dots \hat{T}_\lambda(\lambda_{N-1}, \lambda_N) \hat{\psi}_{l_\lambda}^*(\lambda_N) \\
 &\quad \times \int \prod_{i=1}^N d\rho_i \prod_{i=1}^{N-1} \hat{T}_\rho(\rho_i, \rho_{i+1}) \int d\rho'_N \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(\frac{il_R 2\pi \lambda_1}{\sqrt{2}L}\right) \hat{\psi}_{l_\rho}^*(\rho'_N) \hat{\psi}_{l_\rho}(\rho_N) \\
 &\quad \times \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(-\frac{3V(\rho_N)}{\kappa}\right) \exp\left(-\frac{3V(\lambda_N)}{\kappa}\right)
 \end{aligned} \tag{7.77}$$

The eigenvalue equation Eq. 7.73 propagates the eigenfunction down to  $\lambda_1$  leaving the remaining transfer matrices in  $\rho$ . We recast the  $\lambda_N$  dependence of the potential in Eq. 7.77 by using the boundary condition Eq. 7.69 to set  $\lambda_N = \frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}$

$$\begin{aligned}
 Z(u) &= \sum_{l_R, l_\lambda, l_\rho} \Lambda_{l_R}^{N-1} \hat{\Lambda}_{l_\lambda}^{N-1} \exp\left(\frac{il_R\pi\sqrt{3}u}{L}\right) \\
 &\quad \times \int dR_N \psi_{l_R}(R_N) \exp\left(-\frac{3V\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right)}{\kappa}\right) \hat{\psi}_{l_\lambda}\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \\
 &\quad \times \int d\lambda_1 \hat{\psi}_{l_\lambda}^*(\lambda_1) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(\frac{il_R 2\pi \lambda_1}{\sqrt{2}L}\right) \int \prod_{i=1}^N d\rho_i \prod_{i=1}^{N-1} \hat{T}_\rho(\rho_i, \rho_{i+1}) \\
 &\quad \times \int d\rho'_N \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \exp\left(-\frac{3V(\rho_N)}{\kappa}\right) \hat{\psi}_{l_\rho}^*(\rho'_N) \hat{\psi}_{l_\rho}(\rho_N)
 \end{aligned} \tag{7.78}$$

The final set of transfer matrices are contracted using Eq. 7.73 in

$$\begin{aligned}
Z(u) = & \sum_{l_R, l_\lambda, l_\rho} \Lambda_{l_R}^{N-1} \hat{\Lambda}_{l_\lambda}^{N-1} \exp\left(\frac{il_R\pi\sqrt{3}u}{L}\right) \\
& \times \int dR_N \psi_{l_R}(R_N) \exp\left(-\frac{3V\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right)}{\kappa}\right) \hat{\psi}_{l_\lambda}\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \\
& \times \int d\lambda_1 \hat{\psi}_{l_\lambda}^*(\lambda_1) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(\frac{il_R 2\pi \lambda_1}{\sqrt{2}L}\right) \\
& \times \int \prod_{i=1}^N d\rho_i \prod_{i=1}^{N-1} \hat{T}_\rho(\rho_1, \rho_2) \hat{T}_\rho(\rho_2, \rho_3) \dots \hat{T}_\rho(\rho_{N-1}, \rho_N) \hat{\psi}_{l_\rho}(\rho_N) \\
& \times \int d\rho'_N \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \exp\left(-\frac{3V(\rho_N)}{\kappa}\right) \hat{\psi}_{l_\rho}^*(\rho'_N) \quad (7.79)
\end{aligned}$$

propagating the eigenfunction  $\hat{\psi}_{l_\rho}(\rho_N)$  down to  $\hat{\psi}_{l_\rho}(\rho_1)$ , having used the delta function we introduced in Eq. 7.70 to set  $\rho_N = \rho'_N$ . Inside the potential and the eigenfunction  $\hat{\psi}_{l_\lambda}$  we use Eq. 7.56 to make a change of variable. Applying this to the partition function integral we get

$$\begin{aligned}
Z(u) = & \sum_{l_R, l_\lambda, l_\rho} \Lambda_{l_R}^{N-1} \hat{\Lambda}_{l_\lambda}^{N-1} \hat{\Lambda}_{l_\rho}^{N-1} \exp\left(\frac{il_R\pi 2\sqrt{3}u}{L}\right) \\
& \times \int d\tilde{R}_N \psi_{l_R}\left(\sqrt{2}\tilde{R}_N\right) \exp\left(-\frac{3V(\tilde{R}_N)}{\kappa}\right) \hat{\psi}_{l_\lambda}(\tilde{R}_N) \\
& \times \int d\rho'_N \exp\left(-\frac{3V(\rho'_N)}{\kappa}\right) \hat{\psi}_{l_\rho}^*(\rho'_N) \\
& \times \int d\lambda_1 \hat{\psi}_{l_\lambda}^*(\lambda_1) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(\frac{il_R 2\pi \lambda_1}{\sqrt{2}L}\right) \\
& \times \int d\rho_1 \hat{\psi}_{l_\rho}(\rho_1) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \quad (7.80)
\end{aligned}$$

#### 7.2.4 Mean Axial Displacement

We need to calculate  $Z_{\rho_\alpha}$ , and continue from Eq. 7.74 where the transfer matrices in  $R$  and  $\lambda$  have already been contracted. A modification of this partition function

is constructed to include the variable  $\rho_\alpha$

$$\begin{aligned}
 Z_{\rho_\alpha}(u) = & \sum_{l_R, l_\lambda, l_\rho} \Lambda_{l_R}^{N-1} \hat{\Lambda}_{l_\lambda}^{N-1} \exp\left(\frac{il_R\pi 2\sqrt{3}u}{L}\right) \\
 & \times \int d\tilde{R}_N \psi_{l_R}(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{3V(\tilde{R}_N)}{\kappa}\right) \hat{\psi}_{l_\lambda}(\tilde{R}_N) \\
 & \times \int d\lambda_1 \exp\left(-\frac{il_R 2\pi \lambda_1}{\sqrt{2}L}\right) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \hat{\psi}_{l'_\lambda}^*(\lambda_1) \\
 & \times \int \prod_{i=1}^N d\rho_i \hat{\psi}_{l_\rho}(\rho_N) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \\
 & \times \int d\rho'_N \exp\left(-\frac{3V(\rho'_N)}{\kappa}\right) \hat{\psi}_{l_\rho}^*(\rho'_N) \\
 & \times \hat{T}_\rho(\rho_1, \rho_2) \hat{T}_\rho(\rho_2, \rho_3) \dots \hat{T}_\rho(\rho_{\alpha-1}, \rho_\alpha) \rho_\alpha \hat{T}_\rho(\rho_\alpha, \rho_{\alpha+1}) \dots \hat{T}_\rho(\rho_{N-2}, \rho_{N-1}) \hat{T}_\rho(\rho_{N-1}, \rho_N)
 \end{aligned} \tag{7.81}$$

Using the eigenfunction  $\hat{\psi}_{l_\rho}(\rho_N)$  we contract the transfer matrices on the right side of  $\rho_\alpha$ , propagating the function down to  $\rho_\alpha$  from  $\rho_N$

$$\begin{aligned}
 Z_{\rho_\alpha}(u) = & \sum_{l_R, l_\lambda, l_\rho} \Lambda_{l_R}^{N-1} \hat{\Lambda}_{l_\lambda}^{N-1} \hat{\Lambda}_{l_\rho}^{N-\alpha} \exp\left(\frac{il_R\pi 2\sqrt{3}u}{L}\right) \\
 & \times \int d\tilde{R}_N \psi_{l_R}(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{3V(\tilde{R}_N)}{\kappa}\right) \hat{\psi}_{l_\lambda}(\tilde{R}_N) \\
 & \times \int d\lambda_1 \exp\left(-\frac{il_R 2\pi \lambda_1}{\sqrt{2}L}\right) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \hat{\psi}_{l'_\lambda}^*(\lambda_1) \\
 & \times \int \prod_{i=1}^{\alpha-1} d\rho_i \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \\
 & \times \int d\rho'_N \exp\left(-\frac{3V(\rho'_N)}{\kappa}\right) \hat{\psi}_{l_\rho}^*(\rho'_N) \\
 & \times \hat{T}_\rho(\rho_1, \rho_2) \hat{T}_\rho(\rho_2, \rho_3) \dots \hat{T}_\rho(\rho_{\alpha-1}, \rho_\alpha) \rho_\alpha \hat{\psi}_{l_\rho}(\rho_\alpha)
 \end{aligned} \tag{7.82}$$

To contract the remaining transfer matrices from  $\rho_1$  to  $\rho_\alpha$  we expand the eigenfunctions of  $\psi_{l_R}$ , in terms of  $\rho_1$ , as eigenfunctions of  $\hat{\psi}_{l_\rho}$ . The expansion we make is

$$\exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) = \sum_{l'_\rho} \Omega_{l'_\rho}^{l_R} \hat{\psi}_{l'_\rho}(\rho_1) \tag{7.83}$$

where

$$\Omega_{l_\rho}^{l_R} = \int d\rho'_1 \psi_{l_R} \left( \frac{\sqrt{3}\rho'_1}{\sqrt{2}} \right) \exp \left( -\frac{3V(\rho'_1)}{\kappa} \right) \hat{\psi}_{l_\rho}^*(\rho'_1) \quad (7.84)$$

Inserting this equation into the partition function Eq. 7.81 we are able to contract the remaining transfer matrices to give

$$\begin{aligned} Z_{\rho_\alpha}(u) &= \sum_{l_R, l_\lambda, l_\rho, l'_\rho} \Omega_{l_\rho}^{l_R} \Lambda_{l_R}^{N-1} \hat{\Lambda}_{l_\lambda}^{N-1} \hat{\Lambda}_{l_\rho}^{N-\alpha} \hat{\Lambda}_{l'_\rho}^{\alpha-1} \exp \left( \frac{i l_R \pi 2 \sqrt{3} u}{L} \right) \\ &\times \int d\tilde{R}_N \psi_{l_R} \left( \sqrt{2} \tilde{R}_N \right) \exp \left( -\frac{3V(\tilde{R}_N)}{\kappa} \right) \hat{\psi}_{l_\lambda}(\tilde{R}_N) \\ &\times \int d\lambda_1 \exp \left( -\frac{i l_R 2 \pi \lambda_1}{\sqrt{2} L} \right) \exp \left( -\frac{3V(\lambda_1)}{\kappa} \right) \hat{\psi}_{l_\lambda}^*(\lambda_1) \\ &\times \int d\rho'_N \exp \left( -\frac{3V(\rho'_N)}{\kappa} \right) \hat{\psi}_{l_\rho}^*(\rho'_N) \int d\rho_\alpha \hat{\psi}_{l_\rho}(\rho_\alpha) \rho_\alpha \hat{\psi}_{l_\rho}(\rho_\alpha) \end{aligned} \quad (7.85)$$

All that remains is the calculation of  $Z_{\lambda_\alpha}(u)$ . Starting again from a modified version of Eq. 7.74 to include the variable  $\lambda_\alpha$ , we contract all the transfer matrices in  $R$  and  $\rho$ . We are left with a partition function integral that looks like

$$\begin{aligned} Z_{\lambda_\alpha}(u) &= \sum_{l_R, l_\lambda, l_\rho} \Lambda_{l_R}^{N-1} \hat{\Lambda}_{l_\rho}^{N-1} \exp \left( \frac{i l_R \pi 2 \sqrt{3} u}{L} \right) \\ &\times \int d\tilde{R}_N \psi_{l_R} \left( \sqrt{2} \tilde{R}_N \right) \exp \left( -\frac{3V(\tilde{R}_N)}{\kappa} \right) \hat{\psi}_{l_\lambda}(\tilde{R}_N) \\ &\times \int d\rho_1 \hat{\psi}_{l_\rho}(\rho_1) \exp \left( -\frac{3V(\rho_1)}{\kappa} \right) \psi_{l_R} \left( \frac{\sqrt{3}\rho_1}{\sqrt{2}} \right) \\ &\times \int d\rho'_N \exp \left( -\frac{3V(\rho'_N)}{\kappa} \right) \hat{\psi}_{l_\rho}^*(\rho'_N) \\ &\times \int \prod_{i=1}^N d\lambda_i \hat{\psi}_{l_\lambda}^*(\lambda_N) \exp \left( -\frac{3V(\lambda_1)}{\kappa} \right) \exp \left( -\frac{i l_R 2 \pi \lambda_1}{\sqrt{2} L} \right) \\ &\times \hat{T}_\lambda(\lambda_1, \lambda_2) \hat{T}_\lambda(\lambda_2, \lambda_3) \dots \hat{T}_\lambda(\lambda_{\alpha-1}, \lambda_\alpha) \lambda_\alpha \hat{T}_\lambda(\lambda_\alpha, \lambda_{\alpha+1}) \dots \hat{T}_\lambda(\lambda_{N-2}, \lambda_{N-1}) \hat{T}_\lambda(\lambda_{N-1}, \lambda_N) \end{aligned} \quad (7.86)$$

We then use the eigenfunction  $\hat{\psi}_{l_\lambda}^*(\lambda_N)$  to contract the transfer matrices down from  $\lambda_N$  to  $\lambda_\alpha$

$$\begin{aligned}
 Z_{\lambda_\alpha}(u) &= \sum_{l_R, l_\lambda, l_\rho} \Lambda_{l_R}^{N-1} \hat{\Lambda}_{l_\rho}^{N-1} \hat{\Lambda}_{l_\lambda}^{N-\alpha} \exp\left(\frac{i l_R \pi 2 \sqrt{3} u}{L}\right) \\
 &\quad \times \int d\tilde{R}_N \psi_{l_R}(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{3V(\tilde{R}_N)}{\kappa}\right) \hat{\psi}_{l_\lambda}(\tilde{R}_N) \\
 &\quad \times \int d\rho_1 \hat{\psi}_{l_\rho}(\rho_1) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \\
 &\quad \times \int d\rho'_N \exp\left(-\frac{3V(\rho'_N)}{\kappa}\right) \hat{\psi}_{l_\rho}^*(\rho'_N) \\
 &\quad \times \int \prod_{i=1}^{\alpha-1} d\lambda_i \exp\left(-\frac{3V(\lambda_i)}{\kappa}\right) \exp\left(-\frac{i l_R 2\pi \lambda_1}{\sqrt{2}L}\right) \hat{\psi}_{l_\lambda}^*(\lambda_\alpha) \\
 &\quad \times \hat{T}_\lambda(\lambda_1, \lambda_2) \hat{T}_\lambda(\lambda_2, \lambda_3) \dots \hat{T}_\lambda(\lambda_{\alpha-1}, \lambda_\alpha) \lambda_\alpha
 \end{aligned} \tag{7.87}$$

Without a suitable eigenfunction to contract the remaining transfer matrices we expand the terms in  $\lambda_1$  as eigenfunctions of  $\hat{\psi}_{l'_\lambda}$ . The expansion looks like

$$\exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(-\frac{i l_R 2\pi \lambda_1}{\sqrt{2}L}\right) = \sum_{l'_\lambda} \Xi_{l'_\lambda}^{l_R} \hat{\psi}_{l'_\lambda}(\lambda_1) \tag{7.88}$$

where

$$\Xi_{l'_\lambda}^{l_R} = \int d\lambda'_1 \exp\left(-\frac{i l_R 2\pi \lambda'_1}{\sqrt{2}L}\right) \exp\left(-\frac{3V(\lambda'_1)}{\kappa}\right) \hat{\psi}_{l'_\lambda}^*(\lambda'_1) \tag{7.89}$$

Inserting this equation into the partition function integral we get

$$\begin{aligned}
 Z_{\lambda_\alpha}(u) &= \sum_{l_R, l_\lambda, l'_\lambda, l_\rho} \Lambda_{l_R}^{N-1} \Lambda_{l_\rho}^{N-1} \Lambda_{l_\lambda}^{N-\alpha} \Lambda_{l'_\lambda}^{\alpha-1} \Xi_{l'_\lambda}^{l_R} \exp\left(\frac{i l_R \pi 2 \sqrt{3} u}{L}\right) \\
 &\quad \times \int d\tilde{R}_N \psi_{l_R}(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{3V(\tilde{R}_N)}{\kappa}\right) \hat{\psi}_{l_\lambda}(\tilde{R}_N) \int d\lambda_\alpha \hat{\psi}_{l'_\lambda}(\lambda_\alpha) \lambda_\alpha \hat{\psi}_{l_\lambda}^*(\lambda_\alpha) \\
 &\quad \times \int d\rho_1 \hat{\psi}_{l_\rho}(\rho_1) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \int d\rho'_N \exp\left(-\frac{3V(\rho'_N)}{\kappa}\right) \hat{\psi}_{l_\rho}^*(\rho'_N)
 \end{aligned} \tag{7.90}$$

Using Eq. 7.85 and Eq. 7.90 we are able to calculate mean values of  $\lambda_\alpha$  and  $\rho_\alpha$ , and the mean axial displacement through

$$\langle \lambda_\alpha \rangle = \frac{Z_{\lambda_\alpha}(u)}{Z(u)} \quad (7.91)$$

$$\langle \rho_\alpha \rangle = \frac{Z_{\rho_\alpha}(u)}{Z(u)} \quad (7.92)$$

### 7.2.5 Calculation of $\langle R \rangle$ and average residue positions for an applied extension

Average residue positions in terms of  $x_i$ ,  $y_i$  and  $z_i$  require the calculation of average  $R$ . Starting from Eq. 7.75 we insert  $R_\alpha$ , and contract the  $\hat{T}_\rho$  and  $\hat{T}_\lambda$  transfer matrices using the appropriate eigenfunctions.

$$\begin{aligned} Z_R(u) &= \sum_{l_R, l_\lambda, l_\rho} \Lambda_{l_\rho}^{N-1} \Lambda_{l_\lambda}^{N-1} \exp\left(\frac{i l_R \pi \sqrt{3} u}{L}\right) \\ &\times \int \prod_{i=1}^N dR_i \psi_{l_R}(R_1) T_R(R_1, R_2) \dots T_R(R_{\alpha-1}, R_\alpha) R_\alpha T_R(R_\alpha, R_{\alpha+1}) \dots \\ &\dots T_R(R_{N-1}, R_N) \exp\left(-\frac{3V\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right)}{\kappa}\right) \hat{\psi}_{l_\lambda}\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \\ &\times \int d\rho'_N \hat{\psi}_{l_\rho}^*(\rho'_N) \exp\left(-\frac{3V(\rho'_N)}{\kappa}\right) \int d\rho_1 \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \hat{\psi}_{l_\rho}(\rho_1) \\ &\times \int d\lambda_1 \hat{\psi}_{l_\lambda}^*(\lambda_1) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(-\frac{i l_R 2\pi \lambda_1}{\sqrt{2}L}\right) \end{aligned} \quad (7.93)$$

After contracting the transfer matrices up to  $R_\alpha$  using  $\psi_{l_R}(R_1)$  we get

$$\begin{aligned}
 Z_R(u) = & \sum_{l_R, l_\lambda, l_\rho} \Lambda_{l_R}^{\alpha-1} \Lambda_{l_\rho}^{N-1} \Lambda_{l_\lambda}^{N-1} \exp\left(\frac{i l_R \pi \sqrt{3} u}{L}\right) \\
 & \times \int \prod_{i=1}^N dR_i \psi_{l_R}(R_\alpha) R_\alpha T_R(R_\alpha, R_{\alpha+1}) \dots \\
 & \dots T_R(R_{N-1}, R_N) \exp\left(-\frac{3V(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}})}{\kappa}\right) \hat{\psi}_{l_\lambda}\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \\
 & \times \int d\rho'_N \hat{\psi}_{l_\rho}^*(\rho'_N) \exp\left(-\frac{3V(\rho'_N)}{\kappa}\right) \int d\rho_1 \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \hat{\psi}_{l_\rho}(\rho_1) \\
 & \times \int d\lambda_1 \hat{\psi}_{l_\lambda}^*(\lambda_1) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(-\frac{i l_R 2\pi \lambda_1}{\sqrt{2}L}\right)
 \end{aligned} \tag{7.94}$$

To contract the remaining transfer matrices we make the following expansion

$$\exp\left(-\frac{3V(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}})}{\kappa}\right) \hat{\psi}_{l_\lambda}\left(\frac{R_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) = \sum_{l'_R} G_{l'_R}^{l_\lambda}(u) \psi_{l'_R}(R_N) \tag{7.95}$$

where

$$G_{l'_R}^{l_\lambda}(u) = \int dR'_N \psi_{l'_R}^*(R'_N) \exp\left(-\frac{3V(\frac{R'_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}})}{\kappa}\right) \hat{\psi}_{l_\lambda}\left(\frac{R'_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}}\right) \tag{7.96}$$

The calculation of this integral can be simplified by relabelling the terms with  $\frac{R'_N - \frac{u}{2}\sqrt{3}}{\sqrt{2}} = \tilde{R}_N$  to give

$$\begin{aligned}
 G_{l'_R}^{l_\lambda}(u) &= \exp\left(-\frac{i l'_R \sqrt{3} \pi u}{L}\right) \int d\tilde{R}_N \psi_{l'_R}^*(\sqrt{2}\tilde{R}_N) \exp\left(-\frac{3V(\tilde{R}_N)}{\kappa}\right) \hat{\psi}_{l_\lambda}(\tilde{R}_N) \\
 &= \exp\left(-\frac{i l'_R \sqrt{3} \pi u}{L}\right) \tilde{G}_{l'_R}^{l_\lambda}
 \end{aligned} \tag{7.97}$$

Combining these terms together in the partition function integral we get

$$\begin{aligned}
Z_R(u) = & \sum_{l_R, l_\lambda, l_\rho, l'_R} \Lambda_{l_R}^{\alpha-1} \Lambda_{l_\rho}^{N-1} \Lambda_{l_\lambda}^{N-1} \tilde{G}_{l'_R}^{l_\lambda} \exp\left(\frac{i l_R \sqrt{3}\pi u}{L}\right) \exp\left(-\frac{i l'_R \sqrt{3}\pi u}{L}\right) \\
& \times \int \prod_{i=1}^N dR_i \psi_{l_R}(R_\alpha) R_\alpha T_R(R_\alpha, R_{\alpha+1}) \dots T_R(R_{N-1}, R_N) \psi_{l'_R}(R_N) \\
& \times \int d\rho'_N \hat{\psi}_{l_\rho}^*(\rho'_N) \exp\left(-\frac{3V(\rho'_N)}{\kappa}\right) \int d\rho_1 \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \hat{\psi}_{l_\rho}(\rho_1) \\
& \times \int d\lambda_1 \hat{\psi}_{l_\lambda}^*(\lambda_1) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(-\frac{i l_R 2\pi \lambda_1}{\sqrt{2}L}\right)
\end{aligned} \tag{7.98}$$

which allows us to contract the remaining  $N - \alpha$  transfer matrices. The final form of the partition function integral becomes

$$\begin{aligned}
Z_R(u) = & \sum_{l_R, l_\lambda, l_\rho, l'_R} \Lambda_{l_R}^{\alpha-1} \Lambda_{l_\rho}^{N-1} \Lambda_{l_\lambda}^{N-1} \Lambda_{l'_R}^{N-\alpha} \tilde{G}_{l'_R}^{l_\lambda} \exp\left(\frac{i(l_R - l'_R)\sqrt{3}\pi u}{L}\right) \\
& \times \int dR_\alpha \psi_{l_R}(R_\alpha) R_\alpha \psi_{l'_R}(R_\alpha) \int d\lambda_1 \hat{\psi}_{l_\lambda}^*(\lambda_1) \exp\left(-\frac{3V(\lambda_1)}{\kappa}\right) \exp\left(-\frac{i l_R 2\pi \lambda_1}{\sqrt{2}L}\right) \\
& \times \int d\rho'_N \hat{\psi}_{l_\rho}^*(\rho'_N) \exp\left(-\frac{3V(\rho'_N)}{\kappa}\right) \int d\rho_1 \psi_{l_R}\left(\frac{\sqrt{3}\rho_1}{\sqrt{2}}\right) \exp\left(-\frac{3V(\rho_1)}{\kappa}\right) \hat{\psi}_{l_\rho}(\rho_1)
\end{aligned} \tag{7.99}$$

We have found that using a harmonic potential in the general Hamiltonian we were able to make simplifications to the residue potential terms and avoid transfer hypermatrix calculations needed for the eigenvalue equations. Instead we were able to write the expression for the partition function as a sum over eigenfunctions already detailed in the DNA chapter.

With a simplified semi-analytical partition function expression we can now numerically test the expression for the general partition function when we use a harmonic potential. This will be a verification of the hypermatrix calculations.

### 7.3 Results for collagen eigensystem calculations

When a harmonic potential is used for the residue pair potential the transfer hyper-matrix factorises, avoiding complicated transfer hyper-matrix calculations. The similarities with the DNA ladder structure allows us to use the analytical eigenfunction  $\hat{\psi}_t$  for both the  $\rho$  and  $\lambda$  contributions to the eigenfunctions. Since these eigenvalues and eigenfunctions have already been discussed in the DNA eigensystem results we will focus primarily on the numerical calculations of the transfer hyper-matrix, and their validity.

In DNA the two variable transfer matrix equation produced matrices where the dimension was equal to  $2m + 1$  elements set by the discretisation parameter  $m$ . The eigensystem analysis of this type of matrix can be readily calculated numerically. Problems arise in the four variable transfer hyper-matrix equation since it adds additional dimensions to the eigenvalue problem where standard numerical methods do not exist. To overcome this problem we project the hyper-matrix  $H_{abcd}$  on to a larger matrix  $M_{ij}$  of dimension  $(2m + 1)^2$ . The projection can be expressed in terms of the indices as

$$M_{ij} = H_{abcd} \quad (7.100)$$

where

$$i = (a - 1)(2m + 1) + b \quad (7.101)$$

$$j = (c - 1)(2m + 1) + d \quad (7.102)$$

where we are now able to numerically evaluate the eigensystem for the matrix  $M_{ij}$ . A definitive test of these projections was achieved by using a harmonic potential in the hyper-matrix calculations to compare results with semi-analytical harmonic collagen model. The numerically calculated hyper-eigenvalues and hyper-eigenfunctions should be able to be reproduced from the analytical eigenvalues and eigenfunctions.

We begin with the analysis of the hyper-eigenvalues of the matrix  $M_{ij}$  labelled by  $\hat{\Lambda}_{l\rho\lambda}$  with  $l$  being the index of the hyper-eigenvalue. To be able to generate these eigenvalues from the analytical eigenvalues we need to take notice of the degeneracy of the eigenvalues shown in Fig. 7.2. By associating them to a degenerate set  $d_i$ , where  $i$  indicates the number of degenerate eigenvalues, we find the first few

eigenvalues grouped as

$$d_1 = \{\hat{\Lambda}_{1_{\rho\lambda}}\} \quad (7.103)$$

$$d_2 = \{\hat{\Lambda}_{2_{\rho\lambda}}, \hat{\Lambda}_{3_{\rho\lambda}}\} \quad (7.104)$$

$$d_3 = \{\hat{\Lambda}_{4_{\rho\lambda}}, \hat{\Lambda}_{5_{\rho\lambda}}, \hat{\Lambda}_{6_{\rho\lambda}}\} \quad (7.105)$$

Since the analytical eigenvalues  $\hat{\lambda}_{l_\rho}$  and  $\hat{\lambda}_{l_\lambda}$  are equal we find that when we multiply certain analytical eigenvalues together we reproduce degenerate hyper-eigenvalues. An example of this can be observed from the first two analytical eigenvalues where  $\hat{\lambda}_{1_\rho}\hat{\lambda}_{2_\lambda} = \hat{\lambda}_{2_\rho}\hat{\lambda}_{1_\lambda}$  which correspond to the hyper-eigenvalues in Eq. 7.104. We can assign the product of two eigenvalues with an index  $l$  and  $l'$  to a degenerate set of eigenvalues by using the relation

$$\hat{\lambda}_{l_\rho}\hat{\lambda}_{l'_\lambda} \in d_{l+l'-1} \quad (7.106)$$

such that

$$d_1 = \{\hat{\lambda}_{1_\rho}\hat{\lambda}_{1_\lambda}\} \quad (7.107)$$

$$d_2 = \{\hat{\lambda}_{1_\rho}\hat{\lambda}_{2_\lambda}, \hat{\lambda}_{2_\rho}\hat{\lambda}_{1_\lambda}\} \quad (7.108)$$

$$d_3 = \{\hat{\lambda}_{1_\rho}\hat{\lambda}_{3_\lambda}, \hat{\lambda}_{2_\rho}\hat{\lambda}_{2_\lambda}, \hat{\lambda}_{3_\rho}\hat{\lambda}_{1_\lambda}\} \quad (7.109)$$

The comparison in Fig. 7.2 demonstrates a perfect agreement between the numerical and analytical results.

If we follow the same methodology with the eigenfunctions of the transfer hyper-matrix shown in Fig. 7.3, we can demonstrate the replication of these eigenfunction solutions by a combination of two analytical eigenfunctions  $\hat{\psi}_{l_\rho}$ , and  $\hat{\psi}_{l_\lambda}$ . The single variable eigenfunction  $\hat{\psi}_{1_{\rho\lambda}}$  plotted in Fig. 7.3a is a product of the two eigenfunctions  $\hat{\psi}_{1_\rho}$ , and  $\hat{\psi}_{1_\lambda}$  as shown in Fig. 7.4a. Similarly, the second and third hyper-matrix eigenfunctions plotted in Fig. 7.3b, and Fig. 7.3c are a product of the first two eigenfunctions of the  $\hat{T}$  transfer matrix, both these combinations produce the same eigenfunction  $\hat{\psi}_{1_\rho}\hat{\psi}_{2_\lambda}$ , and  $\hat{\psi}_{2_\rho}\hat{\psi}_{1_\lambda}$ .

Although there is a lack of agreement between second, third, and fourth hyper-eigenfunctions a better comparison can be made if we transform the single variable hyper-eigenfunctions back into matrix form. Fig. 7.5 is for the numerically calculated hyper-eigenfunctions, and Fig. 7.6 is for the hyper-eigenfunctions calculated from the product of the two analytical eigenfunctions  $\hat{\psi}_{t_\rho}$  and  $\hat{\psi}_{t_\lambda}$ . A comparison clearly demonstrates the excellent agreement between the numerical and analytical hyper-

eigenfunction solutions.

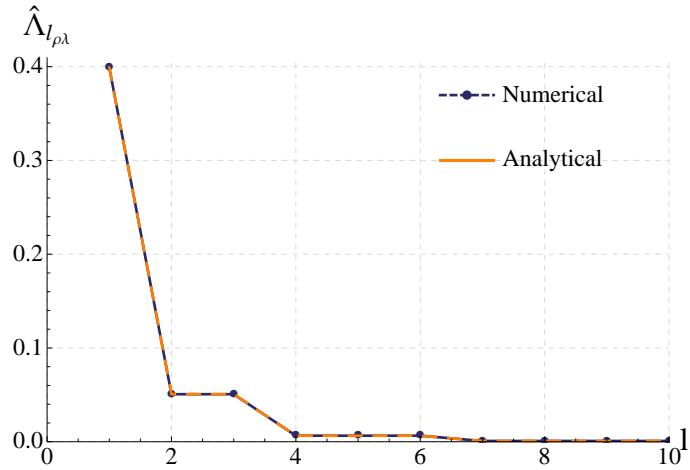


Figure 7.2: Eigenvalues calculated numerically from the transfer hyper-matrix plotted with the analytical eigenvalues calculated by combining  $\hat{\lambda}_{l_\rho}$  and  $\hat{\lambda}_{l_\lambda}$ . The excellent agreement between the numerical and analytical results suggests that we are able to generate the eigenfunctions of the hyper-matrix from the analytical eigenfunctions  $\hat{\psi}_{l_\rho}$  and  $\hat{\psi}_{l_\lambda}$ . These eigenvalues are calculated with  $\Delta = 0.125$ , where  $L = 5$  and  $m = 20$ .

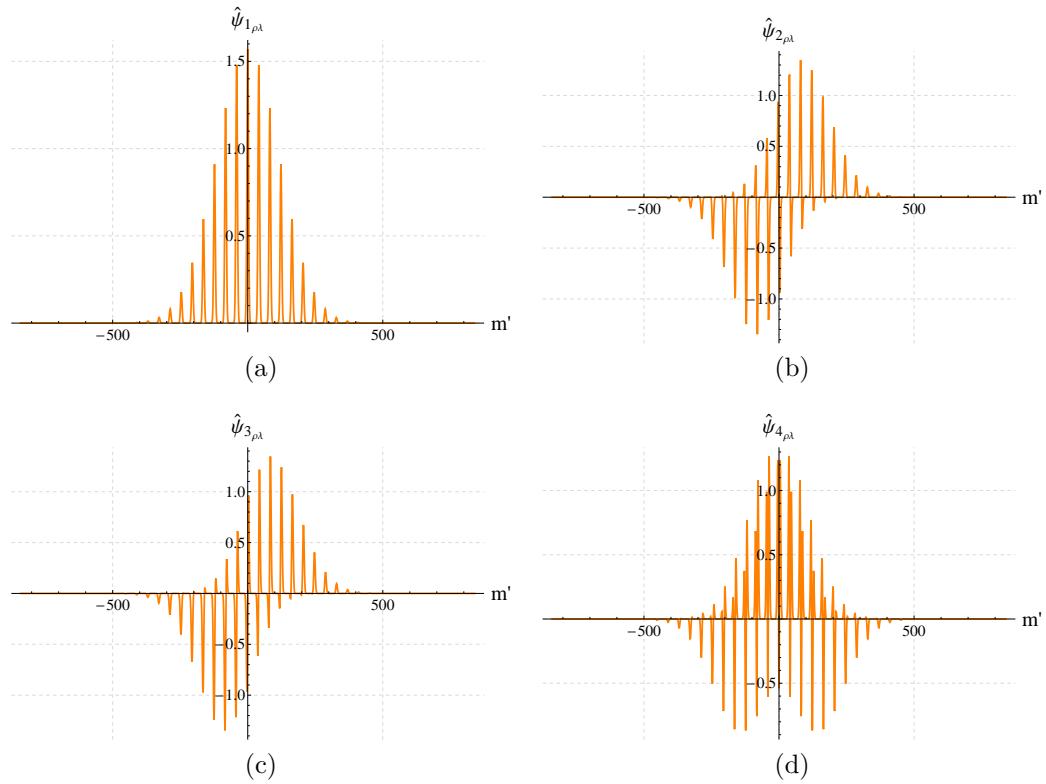


Figure 7.3: A plot of the first four eigenfunctions of the transfer hypermatrix. The hypermatrix was converted to a normal  $N \times N$  matrix before the numerical calculation. These eigenfunctions are calculated with  $N = (2m + 1)^2$  and  $\Delta = 0.125$ , where  $L = 5$  and  $m = 20$ .

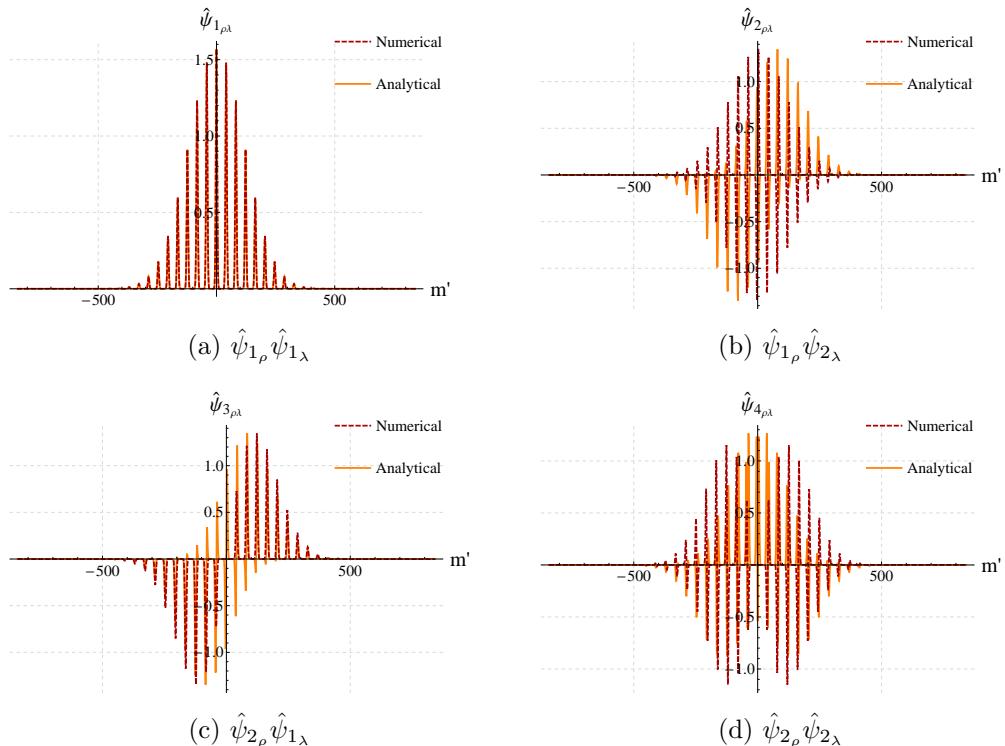


Figure 7.4: In addition to results shown in Fig. 7.3 we plot eigenfunctions that were reproduced by the analytical eigenfunctions  $\hat{\psi}_{l,\rho}$  and  $\hat{\psi}_{l,\lambda}$ .

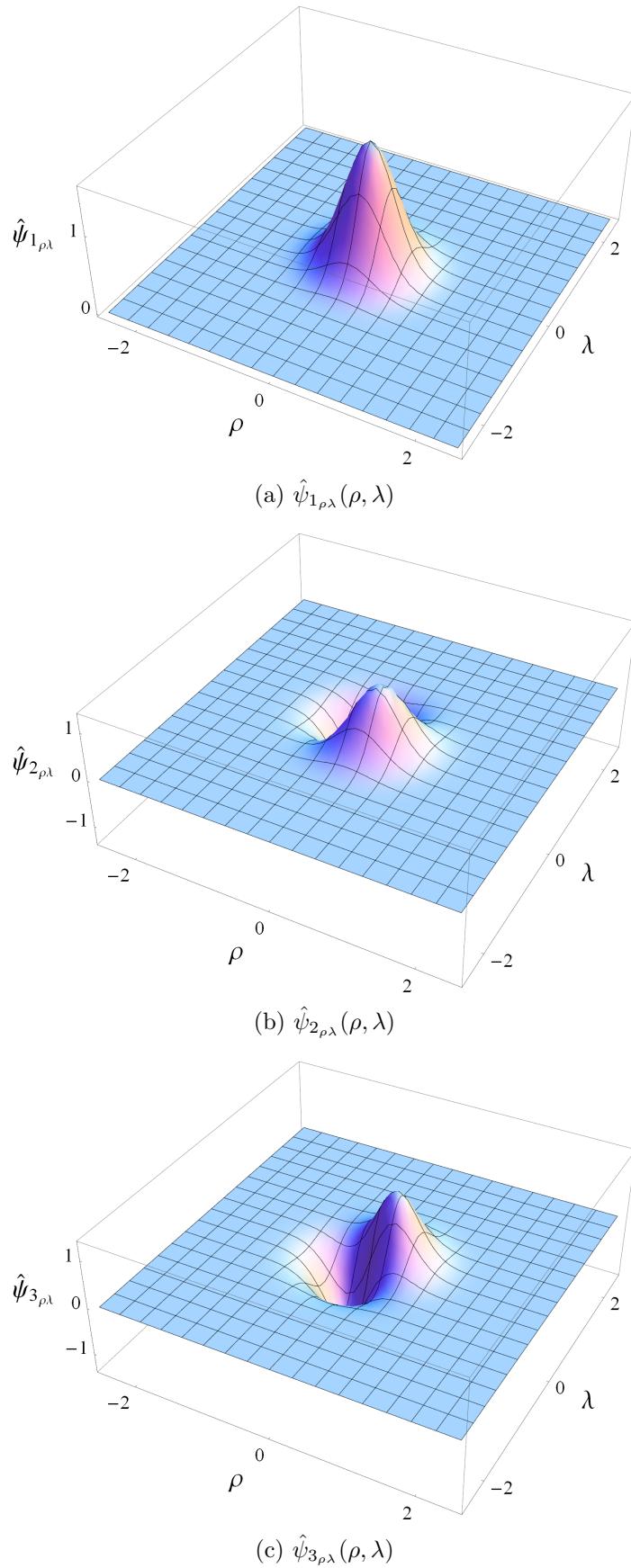


Figure 7.5: 2D plots of the first 3 hypermatrix eigenfunctions when converted back into two variable eigenfunctions. These eigenfunctions are calculated with  $\Delta = 0.125$ , where  $L = 5$  and  $m = 20$ .

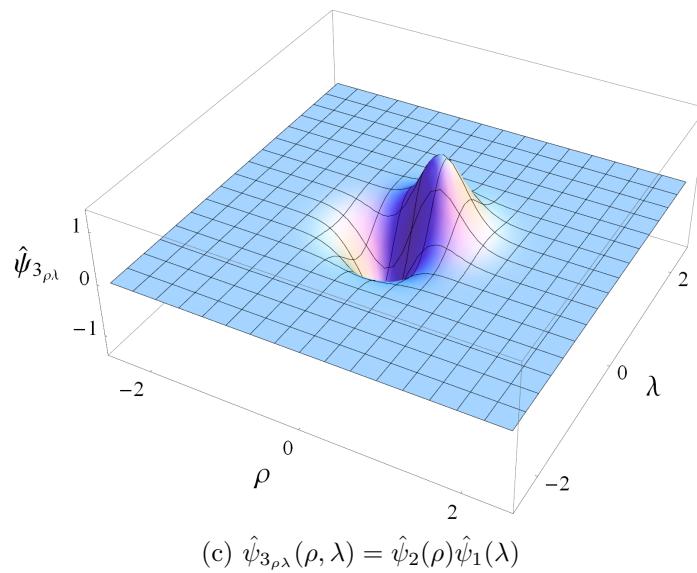
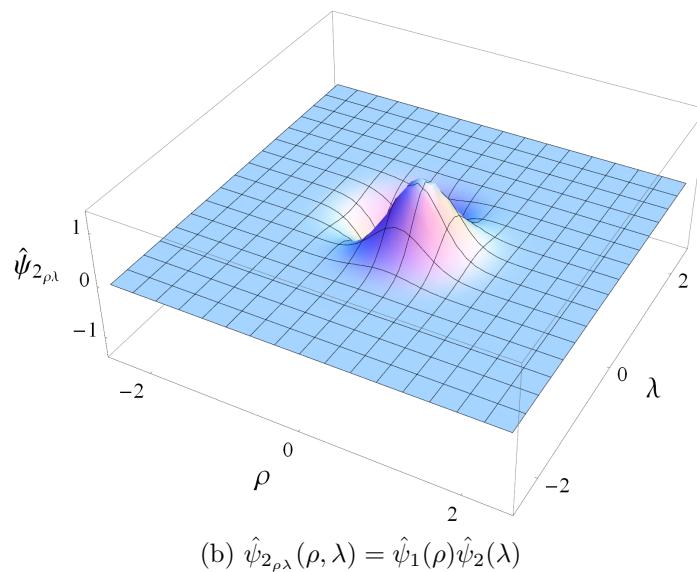
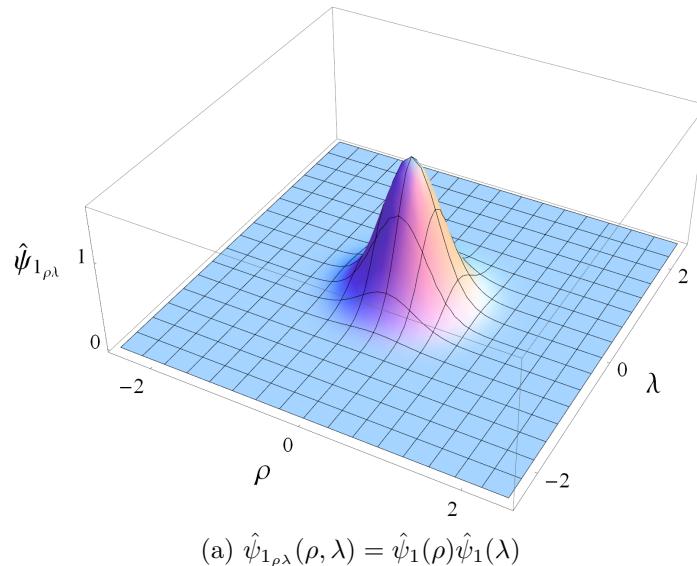


Figure 7.6: 2D plots created from the analytical eigenfunctions  $\hat{\psi}_{l_\rho}$  and  $\hat{\psi}_{l_\lambda}$  that demonstrates the reproduction of the hypermatrix eigenfunctions. In order to match the second and third eigenfunctions we had to apply a rotation of  $\pi/4$  about the  $z$ -axis.

## 7.4 Results for collagen model calculations

Our analysis of the simplified collagen structure gives us two descriptions of the collagen model. The first collagen model is for a general potential  $V$  representing the residue pairs in the collagen structure that uses the hypermatrix analysis to evaluate the partition function through semi-analytical eigenfunctions. In the second collagen model, the harmonic collagen model, a harmonic potential was used early in the analysis of the partition function. This greatly simplified the calculations to express the partition function through the full sets of analytical eigenfunctions that we obtained in DNA for  $\eta_B = \infty$ . In both these models we now go beyond the free energy calculations to describe the force extension behaviour for a given  $u$ .

The models use different calculation methods to evaluate the partition function, but by using a harmonic potential for the residue pair we should return similar results for the partition function. The results shown in Fig. 7.7 compares the two models when we use a harmonic potential in the general collagen model. The excellent agreement between the two results validates the hyper-matrix calculations in the general model. Both collagen models show that for a harmonic potential there is a constant restoring force for an applied extension. The periodicity of the force-extension curves at  $u \approx 2.25$  arises from the periodic constraints on  $\psi_{l_R}$ .

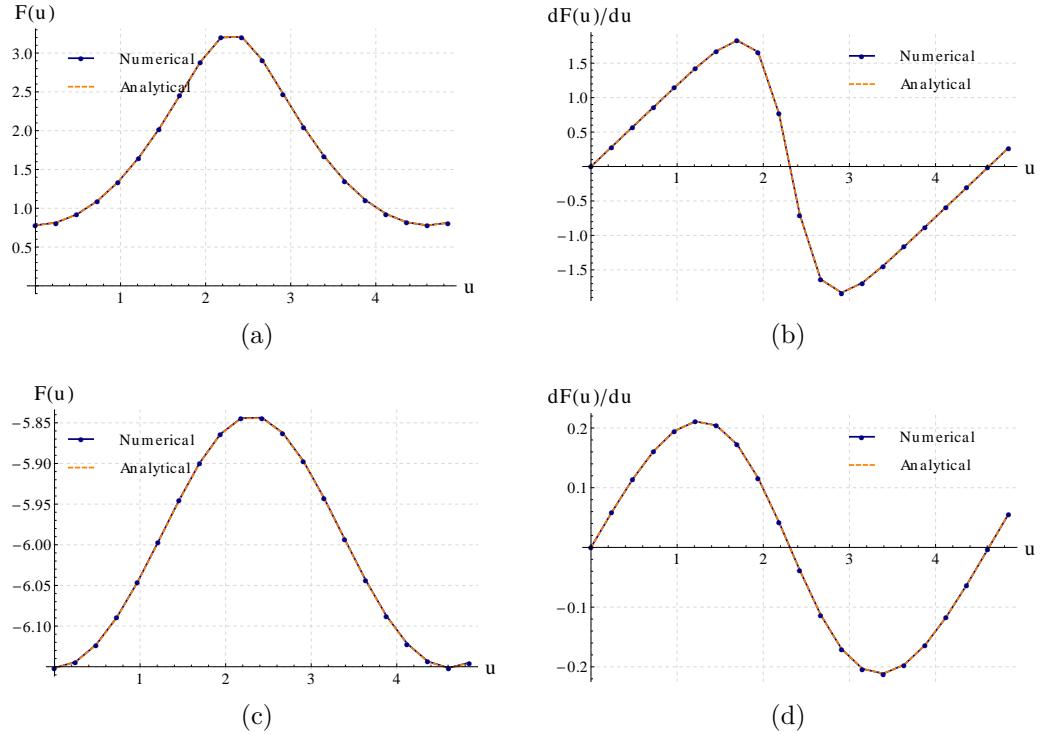


Figure 7.7: A comparison of the results generated by the numerical and semi-analytical calculations of the free energy for harmonic collagen, and its differential with respect to  $u$ . The excellent agreement between the two results is found for different  $\kappa/\kappa'$  ratios as shown in Figs. 7.7a-7.7b for  $\kappa/\kappa' = 1$ , and Figs. 7.7c-7.7d for  $\kappa/\kappa' = 25$ . These calculations were run with  $L = 8$ , and  $m = 16$ .

Using the general collagen model we explore different force-extension curves by applying different potentials for the residue pair interactions. In Fig. 7.8 we demonstrate a comparison of two such potentials with a pure harmonic interaction that has a constant stiffness of  $\kappa'_{ref} = 1 \text{ eV}/\text{\AA}^2$ . They are described as follows:

1. **Soft-Hard** - is a two step anharmonic potential described by two different stiffness constants in the residue pair potential where there is a transition from  $\kappa'_{soft} \rightarrow \kappa'_{hard}$  as the extension increases above a threshold. The term "Soft" indicates a potential with a low stiffness constant  $\kappa'_{soft} = 0.0625 \text{ eV}/\text{\AA}^2$  whereas "Hard" refers to a high stiffness constant  $\kappa'_{hard} = 8 \text{ eV}/\text{\AA}^2$ , hence  $\kappa'_{soft} < \kappa'_{ref} < \kappa'_{hard}$ .
2. **Hard-Soft** - is an opposite description of the Soft-Hard potential. It is a two step anharmonic potential where the stiffness constant in the residue pair potential makes a transition from  $\kappa'_{hard} \rightarrow \kappa'_{soft}$  as the extension increases above a threshold. In this potential  $\kappa'_{soft} = 0.0625 \text{ eV}/\text{\AA}^2$  and  $\kappa'_{hard} = 16 \text{ eV}/\text{\AA}^2$ .

The free energy calculations, and force-extension curves for these residue pair potentials are shown in Fig. 7.9 and Fig. 7.10. A large difference between the free energy curves is primarily due to the low entropy of the structure corresponding to the Hard-to-Soft potential. Alternatively, the Soft-to-Hard residue potential gives the structure a higher entropy state. The gradient of the free energy curve relates to the restoring force of the structure when an extension  $u$  is applied. In the Hard-Soft potential the residue pairs are initially Hard resulting in a larger restoring force. Similarly, the Soft-Hard starts from a weaker restoring force due to the Soft stiffness constant.

Mean axial displacement calculations for the various potentials are shown in Fig. 7.11 for extensions up to  $u = 0.48$ . Again, using the harmonic mean axial displacement calculations as a comparison we can identify the features that correspond to the Soft-Hard, and Hard-Soft potentials. In Fig. 7.11c we notice larger distortions of the mean axial displacements for each residue triplet due to the low stiffness constant in the Soft-Hard potential. The distortion also spreads further from the ends of the structure making the mean axial displacements for each  $\alpha$  larger on average when compared with the harmonic case. For the Hard-to-Soft potential we see the opposite effect in Fig. 7.11b, here, the high stiffness constant reduces the average mean axial displacements and concentrates most of the distortion at the ends of the structure.

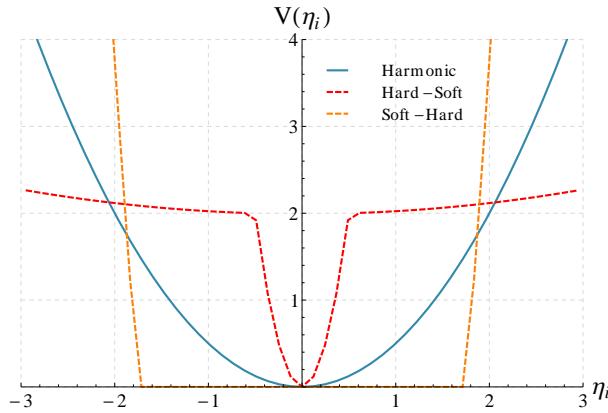


Figure 7.8: A potential plot of the harmonic potential and two anharmonic potentials. The potentials are a function of residue pair separations ( $x_i - y_i$ ), ( $x_i - z_i$ ) and ( $z_i - y_i$ ) which we have labelled  $\eta_i$ .

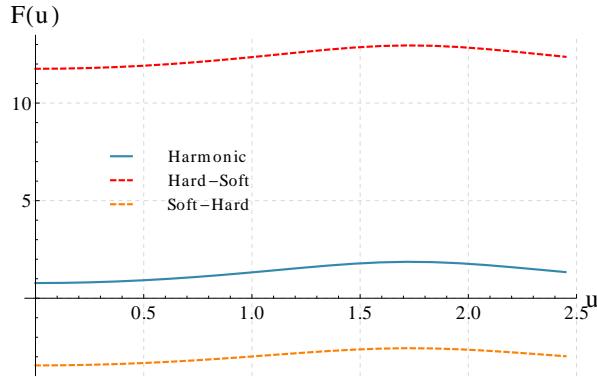


Figure 7.9: A plot of the resulting free energy calculations of the various potentials.

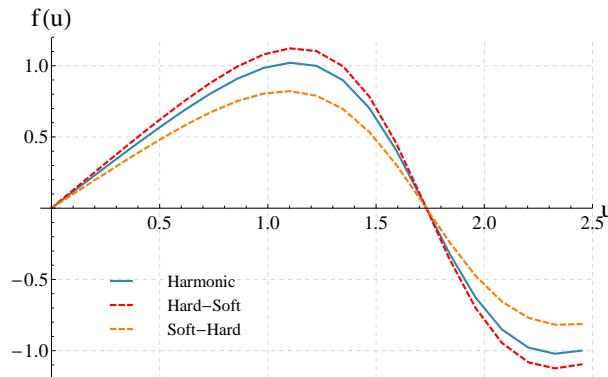


Figure 7.10: The force-extension plots for the various potentials clearly show the effects of the transitions in the two anharmonic potentials. We include the force-extension of the harmonic potential that has a constant restoring force for an applied extension.

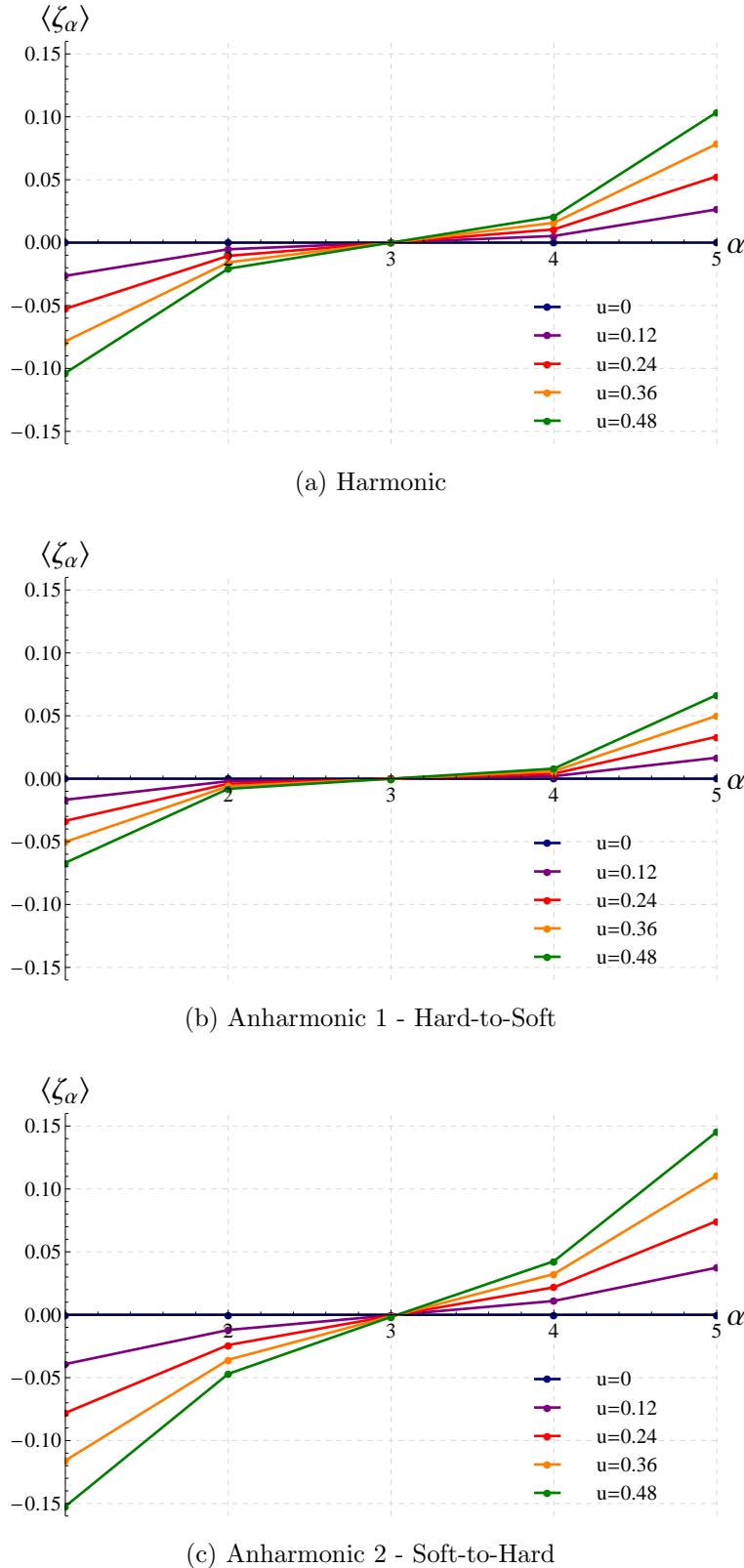


Figure 7.11: Mean axial displacement plots for the Harmonic, Soft-to-Hard and Hard-to-Soft potentials up to an extension of  $u = 0.48$ . Unlike DNA where the mean axial displacement is the average separation of the base-pair, in collagen it is the average separation of adjacent residue triplets given by Eq. 7.36.

## 7.5 Results for collagen calculations

Starting from the same potential to model the base pairs in DNA, we adjust the stiffness to take into account only a single hydrogen bond between a residue pair in collagen. Reducing the spring constant by a factor of 2.5 we determine  $\kappa' \approx 2.7 \times 10^{-4} \text{ eV}/\text{\AA}^2$ .

Several views can be taken to determine the backbone spring constant of collagen. The simplest case is to view the collagen backbone as a collection of covalent bonds identical to DNA and this gives  $\kappa/\kappa' \approx 370$ . MD simulations in [93] provided data for backbone stretching in phase II of the force-extension plot Fig. 4.17 which we are able to use to fit the backbone spring constant in our model. Since these MD simulations had constraints applied to each of the three carbon atoms at the ends of the collagen molecule, in our coarse-grained structure, we calculated a backbone spring constant of  $\kappa \approx 6 \text{ eV}/\text{\AA}^2$ . This gave an extremely high ratio of  $\kappa/\kappa' \approx 22300$ .

Atomistic calculations on a single polypeptide molecule using a reactive force field gave a backbone spring constant of  $\kappa \approx 1.2 \text{ eV}/\text{\AA}^2$  [83] and a more accurate estimation of the backbone spring constant using MD simulations was calculated by the extended MARTINI force-field to give  $\kappa \approx 0.16 \text{ eV}/\text{\AA}^2$  [99]. This value fits much closer to the backbone stiffness of DNA used by Prakash et al. as well as giving a more reasonable backbone to residue pair stiffness ratio of  $\kappa/\kappa' \approx 500$  for collagen.

With the known quantities of the backbone and residue pair stiffness constants we attempt to reproduce shearing force data for different molecular lengths of collagen. Using a method employed by Prakash et al. [77], we determine the shear force  $f_s$  when the separation of the end  $xz$  residue pairs in the structure,  $\eta_1$  and  $\eta_N$ , is equal to our fitting parameter  $\eta_B$ . We express  $\eta$  as follows

$$\eta_i = x_i - z_i \quad (7.110)$$

where  $i$  refers to the index of the residue triplet. This is also known as the shearing separation of a residue pair. To simplify the calculation we use a harmonic approximation of Prakash's base pair potential for the residue pairs; this allows us to use the semi-analytical expression of the partition function. Setting the shearing distance of a single hydrogen bond  $\eta_B$  to be around 2-3 Å, we calculate the shear forces against molecular lengths for various backbone spring constants Fig. 7.12. The results clearly demonstrate the relationship between larger shearing forces and the increase in the backbone stiffness. It is understood that for a given extension, a larger  $\kappa$  increases the amount of energy required to shear the molecule, thus in-

creasing the shear force.

In comparing Fig. 7.12a to Fig. 7.12c we observe clearly that the breakage becomes insensitive to the length of the molecule, suggesting that the distortion only penetrates a certain distance into the structure from the ends of at the extension where failure occurs. By increasing  $\kappa$  we increase the stiffness ratio between the residue pair and the backbone  $\kappa/\kappa'$  allowing more energy to be stored in the residue pairs. This increases the penetrating length of distortions within the molecular structure at an applied extension.

The results shown in Fig. 7.12a are considered to be a reasonable prediction of the response of to the collagen molecule. Though no experimental studies have been done to give the relationship between the molecular length and shearing force we find that the backbone stiffness constant, and the backbone residue pair stiffness ratio, are within a reasonable estimates and compatible with what we have learned from studies of DNA. Compared to DNA other values of  $\kappa$  were very large and gave extremely high  $\kappa/\kappa'$  ratios.

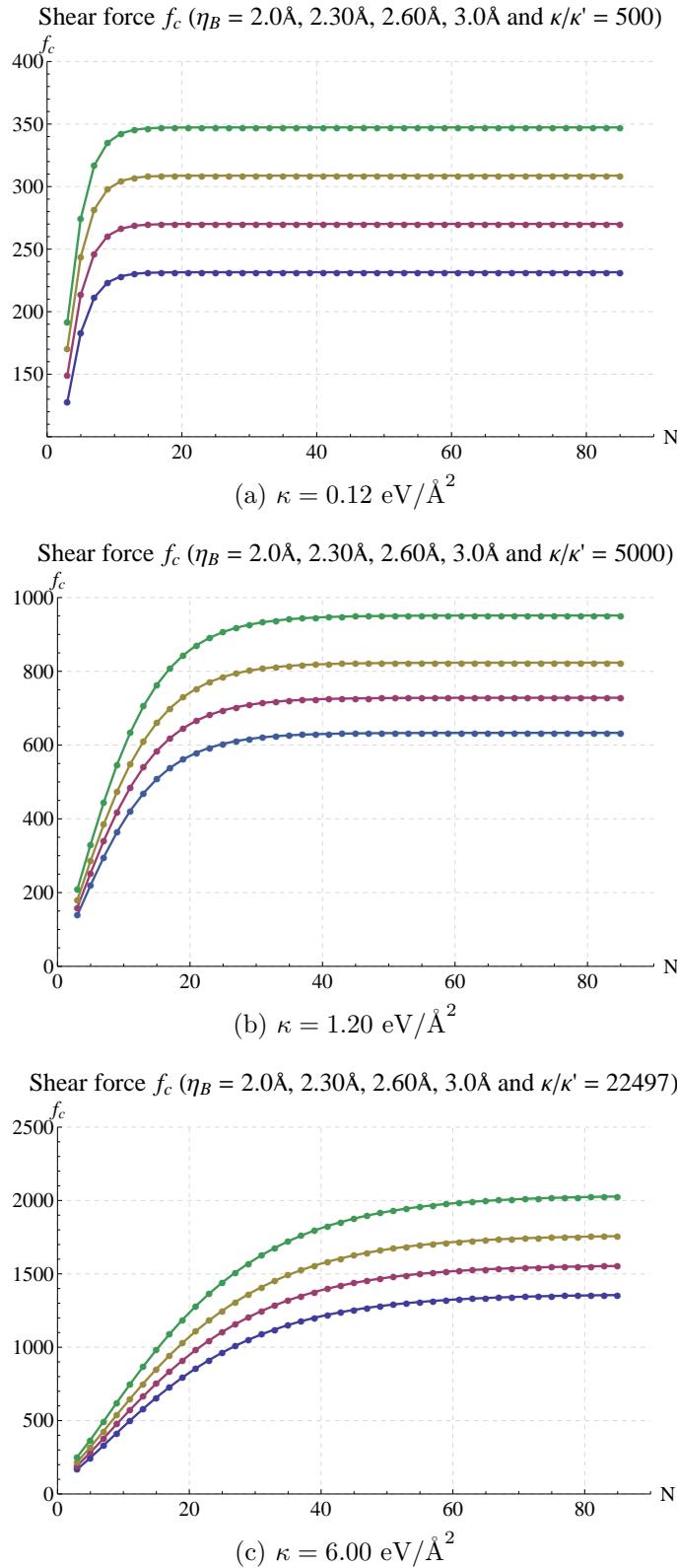


Figure 7.12: Calculations of the shear force against molecular length. In each plot we determine the shear force curve for different  $\eta_B$  values having the range between 2-3 Å. The results for three different values of  $\kappa$  are shown in Fig. 7.12a, Fig. 7.12b and Fig. 7.12c.

# Chapter 8

## Conclusion

It has been shown in this thesis that by simplifying the structure of a multi-stranded biomolecule we are able to find an expression for the partition function needed to evaluate the free energies as a function of axial extension. The mathematical framework of using the transfer integral method to study shearing of different biomolecules has been central to our investigations. The framework was first used to model denaturation of DNA in the early 90's, and we have been able to extend the method to investigate the axial shearing of DNA and progress to study a triple stranded biomolecule. Though we have looked at two different multi-stranded biomolecules the generalisation to more strands is possible though numerically costly.

As an introduction to the partition function calculations we began with a toy model description for a Extendable Freely Jointed Chain as an improvement to the non-extendable Freely Jointed Chain. Using two top-hat weighting functions in the partition function we allowed each link in the chain to extend and contract independently.

In chapter 6 we began with applying the transfer integral method to the DNA molecule. By simplifying the geometrical structure we were able to express the exponentiated Hamiltonian as a product of transfer matrices. By applying relevant contractions we were able to find expressions for the partition function as a function of molecular extension for cases where the DNA structure is in an intact, frayed and bubbled state. Mean base pair axial displacement calculations for the intact state gave a quantitative insight into the mechanical behaviour of the molecule under axial shear stress.

For DNA we used two distinct methods to compare shear force data with experimental results. In the first method we used intact state calculations to determine the force when the average separation of the two end base-pairs reached a shear-

ing length determined by  $\eta_B$ . Determining the relationship between the molecular length and the shear force we found that the results agreed well with the experimental measurements made by Hatch et al. [79]. The values for the stiffness ratio between the backbone and base-pair that we used to fit our model to experimental data was nearly double to what was measured in [79], but agreed well with the stiffness constant ratio fitted in the de Gennes model.

In the second method we used the first phase transition between an intact and frayed state to determine the shear force for failure. These results produced good agreements with experimental shear force data for large molecular lengths. The stiffness ratio that we used to fit our model with the experimental data was also close to the stiffness ratio measured from experiment. However, the lack of agreement for smaller molecular lengths can be attributed to the instability of base-pairs for short DNA molecules causing large errors in the shearing force from experiment.

In Chapters 7 & 8 we introduce the transfer integral method to the collagen molecule. By simplifying its triple helix structure to a triangular prism we study the mechanical behaviour of residue pairs under a shearing force arising from axial extension. Only studying the intact state of the molecule we formulated a general expression of the partition function using numerical hypermatrix calculations valid for all residue pair potentials as well as a harmonic approximation expression. Using data from MD simulations of collagen molecules we produced results showing the relationship between the shear force for failure and molecular length of collagen. Although no axial experiments on tropocollagen have been attempted so far the shear forces are close in magnitude to the properties other biomolecules and we gain insight into the modes of distortion of the structure. Further experimental work would be needed to test these results.

An improvement to the one dimensional models of DNA and collagen could be made by including an additional dimension that accounts for the helical structure around the axis of the biomolecule. In this thesis it was a major macroscopic feature that was largely ignored to simplify the free energy calculations, but is essentially needed to include the effects of torsion, and helical twist as we apply an extension to the structure.

The complexity of biological systems make it extremely difficult to have a single model to determine all of its behaviours and physical properties. In reality different models study different properties of the system, usually dependent on the property that is being measured, and therefore adds to the greater understanding of the system being studied.

# Bibliography

- [1] F. Crick and J. Watson, “Molecular structure of nucleic acids,” *Nature*, **171**, 737–38, (1953).
- [2] C. Kittel, “X-ray diffraction from helices: Structure of DNA,” *American Journal of Physics*, **36**, 610–6, (1968).
- [3] D. Makarov and Z. Wang, “On the interpretation of force extension curves of single protein molecules,” *The Journal of Chemical Physics*, **116**, 7760–65, (2002).
- [4] M. Cheung, L. Chavez, and J. Onuchic, “The energy landscape for protein folding and possible connections to function,” *Polymer*, **45**, 547–55, (2004).
- [5] H. Clausen-Schaumann and M. Seitz, “Force spectroscopy with single biomolecules,” *Current Opinion in Chemical Biology*, **4**, 524–30, (2000).
- [6] D. Beveridge and F. DiCapua, “Free energy via molecular simulation: applications to chemical and biomolecular systems,” *Annual Review of Biophysics and Chemistry*, **18**, 431–92, (1989).
- [7] P. Kollman, “Free energy calculations: applications to chemical and biochemical phenomena,” *Chemical Reviews*, **93**, 2395–417, (1993).
- [8] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry*. W. H. Freeman, 5th ed., (2010).
- [9] M. Karplus, T. Ichiye, and B. Pettitt, “Configurational entropy of native proteins,” *Biophysical Journal*, **52**, 1083–5, (1987).
- [10] M. Prevost, S. J. Wodak, B. Tidor, and M. Karplus, “Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the Ile-96 to Ala mutation in barnase,” *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 10880–4, (1991).

- [11] a. L. Lee and a. J. Wand, “Microscopic origins of entropy, heat capacity and the glass transition in proteins,” *Nature*, **411**, 501–4, (2001).
- [12] B. Trebbi, F. Dehez, P. W. Fowler, and F. Zerbetto, “Favorable entropy of aromatic clusters in thermophilic proteins,” *The Journal of Physical Chemistry B*, **109**, 18184–8, (2005).
- [13] M. J. Stone, “NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding,” *Accounts of Chemical Research*, **34**, 379–88, (2001).
- [14] J. Carlsson and J. Aqvist, “Absolute and relative entropies from computer simulation with applications to ligand binding,” *The Journal of Physical Chemistry B*, **109**, 6448–56, (2005).
- [15] S. W. Homans, “Probing the binding entropy of ligand-protein interactions by NMR.,” *ChemBioChem : A European Journal of Chemical Biology*, **6**, 1585–91, (2005).
- [16] X. Daura, “Molecular Dynamics Simulation of Peptide Folding,” *Theoretical Chemistry Accounts*, **116**, 297–306, (2005).
- [17] D. J. Bicout and A. Szabo, “Entropic barriers, transition states, funnels, and exponential protein folding kinetics: a simple model,” *Protein Science*, **9**, 452–65, (2000).
- [18] L. L. Chavez, J. N. Onuchic, and C. Clementi, “Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates,” *Journal of the American Chemical Society*, **126**, 8426–32, (2004).
- [19] C. Chipot, “Free Energy Calculations in Biological Systems. How Useful Are They in Practice?,” *Springer Lecture Notes in Computational Science and Engineering*, **49**, (2006).
- [20] C. Chipot and A. Pohorille, *Free Energy Calculations: Theory and Applications in Chemistry and Biology*. Springer, (2007).
- [21] R. Sinden, *DNA Structure and Function*. Academic Press, (1994).
- [22] M. Williams and L. M. III, *Biophysics of DNA-protein Interactions: From Single Molecules to Biological Systems*. Springer, (2010).

- [23] N. Sasaki and S. Odajima, "Stress-strain curve and Young's modulus of a collagen molecule as determined by the X-ray diffraction technique," *Journal of Biomechanics*, **29**, 655–58, (1996).
- [24] L. Bozec and M. Horton, "Topography and mechanical properties of single molecules of type I collagen using atomic force microscopy," *Biophysical Journal*, **88**, 4223–31, (2005).
- [25] M. P. E. Wenger, L. Bozec, M. a. Horton, and P. Mesquida, "Mechanical properties of collagen fibrils," *Biophysical Journal*, **93**, 1255–63, (2007).
- [26] P. J. Flory, "Spatial Configuration of Macromolecular Chains," *Science*, **188**, 1268–76, (1975).
- [27] P. J. Flory, *Principles of Polymer Chemistry*. Cornell University Press, (1971).
- [28] L. Sperling, *Introduction to Physical Polymer Science*. John Wiley & Sons, fourth ed., (2006).
- [29] M. Plischke, *Equilibrium Statistical Physics*. World Scientific Pub Co Inc, 3rd ed., (2006).
- [30] P.-G. de Gennes, *Scaling Concepts in Polymer Physics*. Cornell University Press, (1979).
- [31] M. Doi and S. Edwards, *The Theory of Polymer Dynamics*. Oxford University Press, USA, (1988).
- [32] Y. Chan, R. G. Haverkamp, and J. M. Hill, "Force-extension formula for the worm-like chain model from a variational principle," *Journal of Theoretical Biology*, **262**, 498–504, (2010).
- [33] O. Kratky and G. Porod, "Diffuse small-angle scattering of X-rays in colloid systems," *Journal of Colloid Science*, **4**, 35–70, (1949).
- [34] J. V. D. Maarel, *Introduction to Biopolymer Physics*. World Scientific, (2008).
- [35] M. Peyrard, "Nonlinear dynamics and statistical physics of DNA," *Nonlinearity*, **17**, R1–R40, (2004).
- [36] J. Yeomans, *Statistical Mechanics of Phase Transitions*. Clarendon Press, (1992).

- [37] B. H. Zimm, "Theory of "Melting" of the Helical Form in Double Chains of the DNA Type," *The Journal of Chemical Physics*, **33**, 1349–56, (1960).
- [38] R. Wartell and A. Benight, "Thermal denaturation of DNA molecules: a comparison of theory with experiment," *Physics Reports*, **2**, 67–107, (1985).
- [39] D. Poland and H. A. Scheraga, "Occurrence of a Phase Transition in Nucleic Acid Models," *The Journal of Chemical Physics*, **45**, 1464–9, (1966).
- [40] E. Carlon, A. Dkhissi, M. Malki, and R. Blossey, "Stability domains of actin genes and genomic evolution," *Physical Review E*, **76**, 051916, (2007).
- [41] G. Parisi, *Statistical Field Theory*. Perseus Books, (1998).
- [42] H. Kramers and G. Wannier, "Statistics of the two-dimensional ferromagnet. Part II," *Physical Review*, **60**, 252–62, (1941).
- [43] L. Onsager, "Crystal statistics. I. A two-dimensional model with an order-disorder transition," *Physical Review*, **65**, 117–49, (1944).
- [44] S. Brush, "History of the Lenz-Ising model," *Reviews of Modern Physics*, **39**, 883–93, (1967).
- [45] R. Baxter, *Exactly Solved Models in Statistical Mechanics*. Academic Press, (2008).
- [46] G. Newell and E. W. Montroll, "On the theory of the Ising model of ferromagnetism," *Reviews of Modern Physics*, **25**, 353–89, (1953).
- [47] T. D. Schultz, "Two-Dimensional Ising Model as a Soluble Problem of Many Fermions," *Review of Modern Physics*, **36**, 856–71, (1964).
- [48] A. Klug, "The Discovery of the DNA Double Helix," *Journal of Molecular Biology*, **335**, 3–26, (2004).
- [49] M. Wang, H. Yin, R. Landick, and J. Gelles, "Stretching DNA with optical tweezers," *Biophysical Journal*, **72**, 1335–46, (1997).
- [50] J. van Mameren, P. Gross, G. Farge, P. Hooijman, M. Modesti, M. Falkenberg, G. J. L. Wuite, and E. J. G. Peterman, "Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging," *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 18231–6, (2009).

- [51] J. Marko and S. Cocco, “The micromechanics of DNA,” *Physics World*, 37–41, (2003).
- [52] R. Lavery, A. Lebrun, J.-F. Allemand, D. Bensimon, and V. Croquette, “Structure and mechanics of single biomolecules: experiment and simulation,” *Journal of Physics: Condensed Matter*, **14**, R383–R414, (2002).
- [53] T. R. Strick, M. N. Dessainges, G. Charvin, N. H. Dekker, J. F. Allemand, and D. Bensimon, “Stretching of macromolecules and proteins,” *Reports on Progress in Physics*, **66**, 1–45, (2003).
- [54] M. Karplus and G. Petsko, “Molecular dynamics simulations in biology,” *Nature*, **347**, 631–39, (1990).
- [55] S. A. Harris, Z. A. Sands, and C. A. Laughton, “Molecular dynamics simulations of duplex stretching reveal the importance of entropy in determining the biomechanical properties of DNA.,” *Biophysical Journal*, **88**, 1684–91, (2005).
- [56] T. Dauxois, M. Peyrard, and A. R. Bishop, “Entropy-driven DNA denaturation,” *Physical Review E*, **47**, R44–7, (1993).
- [57] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, “Structural transitions and elasticity from torque measurements on DNA,” *Nature*, **424**, 338–41, (2003).
- [58] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, “Single-molecule studies of DNA mechanics,” *Current Opinion in Structural Biology*, **10**, 279–85, (2000).
- [59] J. Watson, *Molecular Biology of the Gene*. Pearson/Benjamin Cummings, 5th ed., (2003).
- [60] S. A. Harris, “The physics of DNA stretching,” *Contemporary Physics*, **45**, 11–30, (2004).
- [61] S. B. Smith, L. Finzi, and C. Bustamante, “Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads,” *Science*, **258**, 1122–26, (1992).
- [62] J. F. Marko and E. D. Siggia, “Stretching DNA,” *Macromolecules*, **28**, 8759–70, (1995).

- [63] C. Bustamante, J. F. Marko, and E. D. Siggia, “Entropic elasticity of lambda-phage DNA,” *Science*, **265**, 1599–600, (1994).
- [64] C. Bouchiat, M. D. Wang, J.-F. Allemand, T. Strick, S. M. Block, and V. Croquette, “Estimating the persistence length of a worm-like chain molecule from force-extension measurements,” *Biophysical Journal*, **76**, 409–13, (1999).
- [65] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, and J. Viovy, “DNA: an extensible molecule,” *Science*, **271**, 792–4, (1996).
- [66] S. B. Smith, Y. Cui, and C. Bustamante, “Overstretching B-DNA : The elastic response of individual double-stranded and single- stranded DNA molecules,” *Science*, **271**, 795–9, (1996).
- [67] C. Bustamante, Z. Bryant, and S. Smith, “Ten years of tension: single-molecule DNA mechanics,” *Nature*, **421**, 423–7, (2003).
- [68] M. C. Williams, I. Rouzina, and M. J. McCauley, “Peeling back the mystery of DNA overstretching,” *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 18047–8, (2009).
- [69] H. Fu, H. Chen, J. F. Marko, and J. Yan, “Two distinct overstretched DNA states,” *Nucleic Acids Research*, **38**, 5594–600, (2010).
- [70] M. Peyrard and A. Bishop, “Statistical mechanics of a nonlinear model for DNA denaturation,” *Physical Review Letters*, **62**, 2755–8, (1989).
- [71] T. Dauxois, M. Peyrard, and A. Bishop, “Dynamics and thermodynamics of a nonlinear model for DNA denaturation,” *Physical Review E*, **47**, 684–95, (1993).
- [72] A. Campa, “Experimental tests of the Peyrard-Bishop model applied to the melting of very short DNA chains,” *Physical Review E*, **58**, 3585–8, (1998).
- [73] D. Cule and T. Hwa, “Denaturation of heterogeneous DNA,” *Physical Review Letters*, **79**, 2375–8, (1997).
- [74] N. Theodorakopoulos, “Melting of genomic DNA: Predictive modeling by nonlinear lattice dynamics,” *Physical Review E*, **82**, 021905, (2010).
- [75] P.-G. de Gennes, “Maximum pull out force on DNA hybrids,” *Comptes Rendus de l'Académie des Sciences*, **2**, 1505–8, (2001).

- [76] B. Chakrabarti and D. R. Nelson, "Shear unzipping of DNA," *The Journal of Physical Chemistry B*, **113**, 3831–6, (2009).
- [77] S. Prakash and Y. Singh, "Shear unzipping of double-stranded DNA," *Physical Review E*, **84**, 031905, (2011).
- [78] R. K. Mishra, G. Mishra, M. S. Li, and S. Kumar, "Effect of shear force on the separation of double-stranded DNA," *Physical Review E*, **84**, 032903, (2011).
- [79] K. Hatch, C. Danilowicz, V. Coljee, and M. Prentiss, "Demonstration that the shear force required to separate short double-stranded DNA does not increase significantly with sequence length for sequences longer than 25 base pairs," *Physical Review E*, **78**, 011920, (2008).
- [80] M. D. Shoulders and R. T. Raines, "Collagen structure and stability," *Annual Review of Biochemistry*, **78**, 929–58, (2009).
- [81] P. Fratzl, *Collagen*. Springer, (2008).
- [82] J. Brinckmann, H. Notbohm, and P. Müller, *Collagen: Primer in Structure, Processing, and Assembly*. Springer, (2005).
- [83] M. J. Buehler, "Nature designs tough collagen: explaining the nanostructure of collagen fibrils," *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12285–90, (2006).
- [84] L. Bozec, G. van der Heijden, and M. Horton, "Collagen fibrils: nanoscale ropes," *Biophysical Journal*, **92**, 70–5, (2007).
- [85] V. Ottani, M. Raspanti, and A. Ruggeri, "Collagen structure and functional implications," *Micron*, **32**, 251–60, (2001).
- [86] M. J. Buehler, "Atomistic and continuum modeling of mechanical properties of collagen: elasticity, fracture, and self-assembly," *Journal of Materials Research*, **21**, 1947–61, (2006).
- [87] G. Ramachandran and G. Kartha, "Structure of collagen," *Nature*, **174**, 269–70, (1954).
- [88] D. Nelson and M. Cox, *Lehninger Principles of Biochemistry*. W. H. Freeman, 4th ed., (2004).

- [89] A. Bhattacharjee and M. Bansal, “Collagen structure: the Madras triple helix and the current scenario.,” *IUBMB life*, **57**, 161–72, (2005).
- [90] G. Ramachandran and G. Kartha, “Structure of collagen,” *Nature*, **176**, 593–5, (1955).
- [91] A. Rich and F. Crick, “The structure of collagen,” *Nature*, **176**, 915–6, (1955).
- [92] A. Perczel and V. Palfi, “Stability of the hydration layer of tropocollagen: A QM study,” *Journal of Computational Chemistry*, **31**, 764–77, (2010).
- [93] M. J. Buehler and S. Y. Wong, “Entropic elasticity controls nanomechanics of single tropocollagen molecules,” *Biophysical Journal*, **93**, 37–43, (2007).
- [94] Y.-L. Sun, Z.-P. Luo, A. Fertala, and K.-N. An, “Stretching type II collagen with optical tweezers,” *Journal of Biomechanics*, **37**, 1665–9, (2004).
- [95] Y. L. Sun, Z. P. Luo, and K. N. An, “Stretching short biopolymers using optical tweezers,” *Biochemical and Biophysical Research Communications*, **286**, 826–30, (2001).
- [96] Y.-L. Sun, Z.-P. Luo, A. Fertala, and K.-N. An, “Direct quantification of the flexibility of type I collagen monomer,” *Biochemical and Biophysical Research Communications*, **295**, 382–6, (2002).
- [97] T. Gutsmann, G. E. Fantner, J. H. Kindt, M. Venturoni, S. Danielsen, and P. K. Hansma, “Force spectroscopy of collagen fibers to investigate their mechanical properties and structural organization,” *Biophysical Journal*, **86**, 3186–93, (2004).
- [98] A. C. Lorenzo and E. R. Caffarena, “Elastic properties, Young’s modulus determination and structural stability of the tropocollagen molecule: a computational study by steered molecular dynamics,” *Journal of Biomechanics*, **38**, 1527–33, (2005).
- [99] A. Gautieri, A. Russo, and S. Vesentini, “Coarse-grained model of collagen molecules using an extended MARTINI force field,” *Journal of Chemical Theory and Computation*, **6**, 1210–18, (2010).
- [100] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, “The MARTINI force field: coarse grained model for biomolecular simulations,” *The Journal of Physical Chemistry B*, **111**, 7812–24, (2007).

- [101] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink, “The MARTINI coarse-grained force field: extension to proteins,” *Journal of Chemical Theory and Computation*, **4**, 819–34, (2008).
- [102] F. Reif, *Fundamentals of statistical and thermal physics*. McGraw-Hill, (1965).
- [103] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical methods in physics and engineering*. Cambridge University Press, (2002).
- [104] I. N. Bronshtein, K. A. Semendyayev, and G. Musiol, *Handbook of Mathematics*. Springer, 5th ed., (2007).