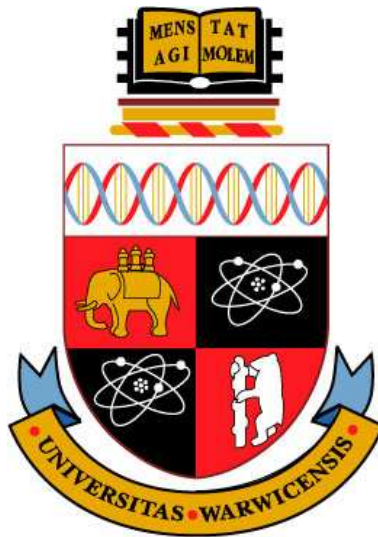




Electronic Transport in DNA

by

K. Daphne Klotso

Thesis

Submitted to the University of Warwick

for the degree of

MSc in Physics, by Research

Department of Physics

October 2004

THE UNIVERSITY OF
WARWICK

Contents

List of Tables	iv
List of Figures	v
Acknowledgments	x
Declarations	xi
Abstract	xii
Chapter 1 Introduction	1
1.1 The DNA Molecule	1
1.2 The Link Between DNA and Physics	5
1.3 Previous Studies and Results on the Electronic Properties of DNA	8
Chapter 2 The Disorder Perspective	11
2.1 Electronic Transport in the Quantum Regime	11
2.2 Localisation and Disordered Systems	13
2.3 Scaling Hypothesis of Localisation	14
2.4 The Transfer Matrix Method	17
2.5 Tight-Binding Models for the Electronic Transport Along DNA .	21
2.5.1 The One-channel “Fishbone Model”	22
2.5.2 The Two-channel “Ladder Model”	24

Chapter 3 Numerical Details	27
3.1 DNA Sequences	27
3.2 Fishbone and Ladder Models Input Files	29
3.3 The Computation Process	31
3.4 The Sliding Window Technique of the Ladder Model	33
3.5 Correlation Functions Between the Localisation Lengths at Different Parts of the DNA Sequence	35
3.6 Programmes and Computational Techniques	37
3.7 The SPARTAN Package for Semi-empirical Molecular Orbital Calculations	38
Chapter 4 Results	39
4.1 Results for the Fishbone model	39
4.1.1 Varying the energy	39
4.1.2 Varying the Backbone Disorder	42
4.2 Results for the Ladder model	44
4.2.1 Varying the energy	45
4.2.2 Varying the backbone disorder	45
4.3 Sliding Window Technique Results: Correlation Functions Between Localisation Lengths	49
4.4 Semi-empirical Molecular Orbital Results	55
Chapter 5 Discussion	60
Appendix A	64
Appendix B	65
Bibliography	67

List of Tables

3.1	The four DNA sequences that we have used, poly(dG)-poly(dC), telomeric, random-ATGC, λ -DNA and scrambled λ -DNA; with some characteristics, composition of bases, the way these are arranged, the length (in bases) and additional comments.	29
3.2	An example of the input file that contains all the fixed parameters that are fed into the programme, for the ladder model. The input file for the fishbone model is a little simpler.	30
4.1	The HOMO's (Highest Occupied Molecular Orbitals), the LUMO's (Lowest Unoccupied Molecular Orbitals) and the gap between them are shown for different components of the DNA molecule, excluding the backbone, using the semi-empirical molecular orbital calculations that the molecular modelling programme SPARTAN provides [1].	55
4.2	The HOMO's (Highest Occupied Molecular Orbitals), the LUMO's (Lowest Unoccupied Molecular Orbitals) and the gap between them are shown for different components of the DNA molecule, including the backbone, using the semi-empirical molecular orbital calculations that were obtained using the molecular modelling programme SPARTAN [1].	56

List of Figures

1.1	The chemical composition of DNA with the four bases Adenine, Thymine, Cytosine, Guanine and the backbone. The backbone is made of phosphorylated sugars shown in yellow and brown. . . .	2
1.2	The double-helix structure of DNA and the way the four bases A,T,G,C and the backbone are arranged in the molecule.	3
1.3	The way DNA is arranged with the histone proteins to form the chromosome [2].	4
2.1	Function β versus the natural logarithm of the dimensionless conductance. For $\beta > 0$ the behaviour is metallic whereas for $\beta < 0$ it is insulating. For a two-dimensional sample disorder can induce a metal insulator transition (MIT).	16
2.2	The quasi-one dimensional strip of length N and width $M = 3$. The wavefunction at each site N can be distributed in space in M different ways. Transfer matrix method makes sure the initial orthonormal vectors are reorthonormalised every few (usually 10) sites.	21
2.3	The fishbone model for electronic transport along DNA. The model corresponds to the Hamiltonian given in Equation 2.27. It was first used as a model for DNA by Cuniberti <i>et al.</i> [3]	23

2.4	The ladder model for electronic transport along DNA. The model corresponds to the Hamiltonian given in Equation 2.36.	25
3.1	A schematic view of the application of how the correlation function between the localisation lengths is calculated. The sequence length is $m = 10$, the window size is $l = 3$ and thus localisation lengths can be computed for $n = m - (l - 1)$ intervals. k is the distance from each localisation length to all the others in the sequence, measured from the starting base of the window they correspond to.	36
4.1	Dependence of the localisation lengths upon the system's energy, with no additional disorder, using the fishbone model of electronic transport along DNA, for four different sequences: poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA. The first two sequences seem to have extended states within the minibands whereas the latter two are rather localised over a few tens of bases.	40
4.2	Plot of the inverse localisation lengths versus the Fermi energy, for four DNA sequences as in Figure 4.1.	41
4.3	The localisation lengths as a function of binary onsite disorder, at the backbone sites, at $E = 0$, using the fishbone model for the DNA sequences: poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA.	43
4.4	The natural logarithm of the localisation lengths of the DNA sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, plotted as a function of the binary onsite disorder at the backbone sites, at $E = 3$, using the fishbone model.	43

4.5	The localisation lengths of the DNA sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, plotted as a function of the uniform onsite disorder at the backbone sites, at $E = 0$, using the fishbone model.	44
4.6	Dependence of the localisation lengths upon the energy, with no additional disorder, using the ladder model of electronic transport along DNA, for four different sequences: poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA.	46
4.7	The inverse localisation lengths versus energy with no additional disorder, is plotted here, for poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, using the ladder model.	46
4.8	The localisation lengths of the DNA sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, plotted as a function of the binary onsite disorder at the backbone sites, at $E = 0$, using the ladder model.	47
4.9	The localisation lengths of the DNA sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, plotted as a function of the binary onsite disorder at the backbone sites, at $E=3\text{eV}$, using the ladder model.	48
4.10	The localisation lengths of the DNA sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, plotted as a function of uniform onsite disorder at the backbone sites, at $E=0\text{eV}$, using the ladder model.	48

4.11	The correlation functions plotted as a function of the DNA sequence position, i.e. the number of the base, using the window technique of the ladder model for all window lengths 10, 30, 100, 200, 500, 1000 bases, for λ -DNA. The inset shows the same correlation function versus the sequence position <i>divided by</i> the window length for each window size.	50
4.12	The correlation functions plotted as a function of the DNA sequence position using the window technique of the ladder model for all window lengths 10, 30, 100, 200, 500, 1000 bases, for scrambled λ -DNA. The inset shows the same correlation function versus the sequence position <i>divided by</i> the window length for each window size.	51
4.13	Comparing the correlation functions of biological λ -DNA and scrambled λ -DNA as a function of the sequence position, using the sliding window technique, for a window size of 30 bases. The inset shows a zoom into the first 90 bases and what we observe is that the correlations disappear at the point where the window frame ends, i.e. at the <i>30th</i> base.	52
4.14	Comparing the correlation functions of biological λ -DNA and scrambled λ -DNA as a function of the sequence position, using the sliding window technique, for a window size of 100 bases. The inset shows a zoom into the first 400 bases and what we observe is that the correlations disappear at the point where the window frame ends, i.e. at the <i>100th</i> base.	53

- 4.15 Comparing the correlation functions of biological λ -DNA and scrambled λ -DNA as a function of the sequence position, using the sliding window technique, for a window size of 500 bases. The inset shows a zoom into the first 1000 bases and what we observe is that the correlations disappear at the point where the window frame ends, i.e. at 500 bases. 54
- 4.16 Comparing the correlation functions of biological λ -DNA and scrambled λ -DNA as a function of the sequence position, using the sliding window technique, for a window size of 1000 bases. The inset shows a zoom into the first 1500 bases and what we observe is that the correlations disappear at the point where the window frame ends, i.e. at 1000 bases. 54

Acknowledgments

Firstly I would like to thank my supervisors Rudolf A. Römer and Matthew S. Turner for their continuous support and patience. John Grime for his work on the window technique and Peter Taylor for his help with the SPARTAN package. Also Christoph, Mat and Grok. I would like to thank the people who have helped me get here and have supported me all the way; my parents and step-parents, Vasso, Glenn and Yago. Finally, I wish to thank Pam Thomas who has supported me throughout all my years in Warwick.

Declarations

I hereby declare that this thesis represents my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree at The University of Warwick or any other educational institution, except where the acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at The University of Warwick or elsewhere, is explicitly acknowledged in the thesis.

Parts of the thesis have been submitted for publication:

Klotsa, Daphne, K. and Roemer, Rudolf, A. and Turner, Matthew, S. (2004) Electronic Transport in DNA - the Disorder Perspective. In Proceedings 27th International Conference on the Physics of Semiconductors(Q5 129), pages 328, Flagstaff, Arizona.

Abstract

We are studying the electronic properties of DNA theoretically by looking at two tight-binding models and four sequences of DNA. The electronic transport is studied in terms of localisation lengths as a function of the energy and the backbone disorder. Onsite potentials and different types of disorder account for different real environments. We have performed calculations on poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA. We find that poly(dG)-poly(dC) and telomeric DNA are conducting when no additional disorder is present whereas random-ATGC and λ -DNA are insulating. Furthermore we have looked at the correlations between the coding and non-coding parts of biological λ -DNA but find no systematic differences in their behaviour. Finally, the correlation functions between the localisation lengths of λ -DNA and scrambled λ -DNA do not yield any obvious regularity.

Chapter 1

Introduction

1.1 The DNA Molecule

In biology the unit that defines a living organism is the cell. Its size is of the order of micrometers, $\sim 10\mu\text{m}$ for animal cells and $\sim 100\mu\text{m}$ for plant cells. Its central component is called the *cell nucleus* and is between $3 - 10\mu\text{m}$ large in its diameter depending on the cell. Apart from the nucleus' size, which covers about $1/4$ to $1/3$ of the whole cell, the nucleus is the home base of a number of fundamental operations, including cell division. Deoxyribo-Nucleic Acid, DNA, the material that carries the genetic information and the basic constituent of the chromosome is found in the nucleus.

DNA is a very complex molecule in every possible way, structurally, chemically, energetically and is also very sensitive to its environment; the temperature, the solution it is in etc. In this Chapter, we review some important properties of DNA as a biological supermolecule as well as the mysteries that it still conceals.

Chemically DNA is a polymer of nucleotides, which consist of bases and phosphorylated sugars. There are four bases, the purines Adenine (A) and Guanine (G) and the pyrimidines Thymine (T) and Cytosine (C). The bases are organic compounds that apart from Carbon, contain the following; Adenine: 3N , NH ,

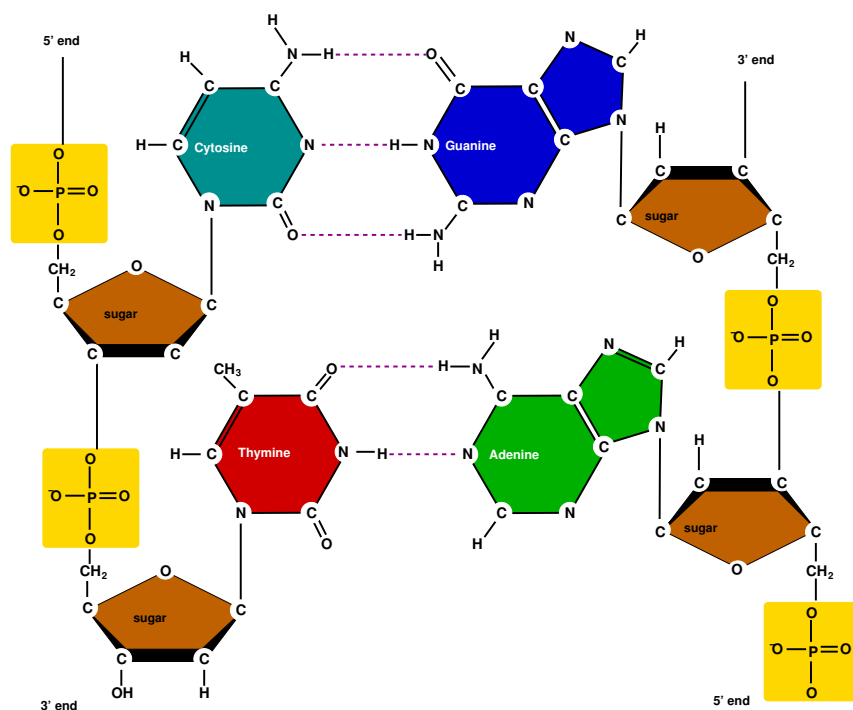


Figure 1.1: The chemical composition of DNA with the four bases Adenine, Thymine, Cytosine, Guanine and the backbone. The backbone is made of phosphorylated sugars shown in yellow and brown.

NH₂, Guanine: 2N, NH, NH₂, HN, O, Thymine: NH, HN, 2O, CH₃, Cytosine: N, NH, NH₂, O, as shown in Figure 1.1.

Each base is linked to 2'-deoxyribose, which belongs to the sugars group. One Nitrogen from each base bonds with a 2'-deoxyribose molecule and that is the only link between the base and the sugar, which together construct a nucleoside. The structure is further complicated with the addition of a phosphodiester bond between the 3'hydroxyl group of one nucleotide sugar and the 5' phosphate group of the consecutive one as shown in Fig. 1.1, the sugar in brown colour and the phosphate group in yellow. The next base in the sequence is then linked to this

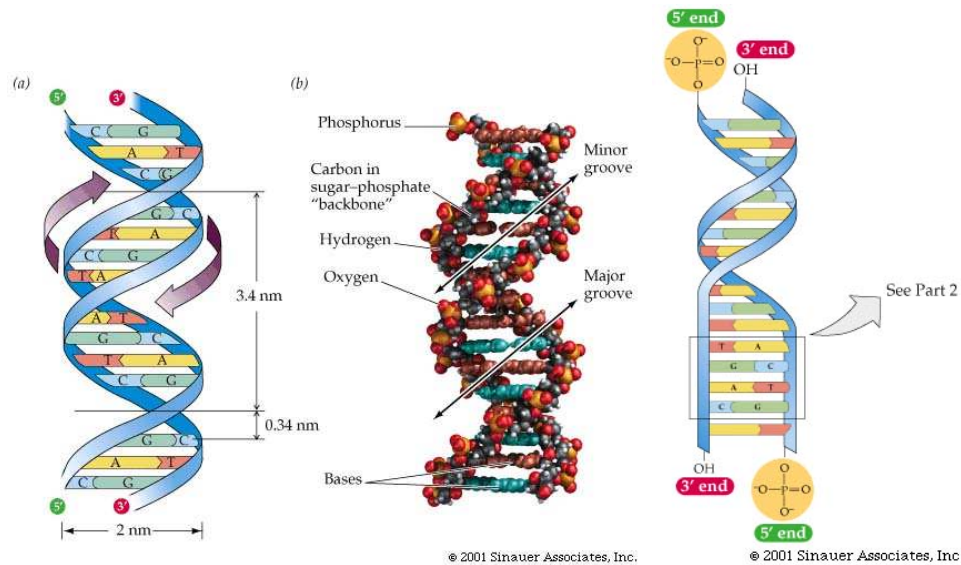


Figure 1.2: The double-helix structure of DNA and the way the four bases A,T,G,C and the backbone are arranged in the molecule.

sugar. This structure now consists of a chain of unconnected bases linked to the sugars which are interconnected through phosphate groups [4].

The polynucleotide chain has a sense of direction with one end being the 5' phosphate group and the other the 3' hydroxyl group. For each chain like this there is a complementary one with the opposite sense of direction. The two chains are linked only through the matching bases; Adenine always pairs with Thymine and Guanine with Cytosine with two and three Hydrogen bonds, respectively. These two chains are wound around one another to form the famous double-helix structure of DNA, shown in Figure 1.2. The phosphorylated sugars form the backbone of DNA which resembles the side rails of a helical ladder. Due to the negatively charged backbone, the electrostatic interactions and the Hydrogen bonds that between the hydrophobic bases, hold the DNA together to its double-helix structures at all temperatures below the melting point.

In its natural environment, DNA is wrapped around one histone protein twice in a spiral manner and then to another protein and so on as shown in Figure

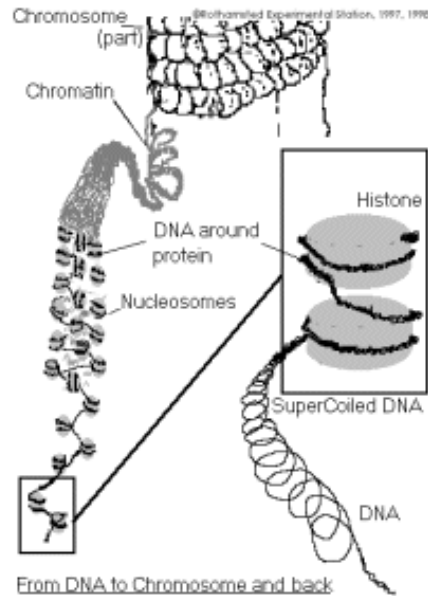


Figure 1.3: The way DNA is arranged with the histone proteins to form the chromosome [2].

1.3. This very closely packed structure is arranged in the shape of a solenoid or a spring; the DNA aggregate and the protein form the chromatin which is the constituent of chromosomes. Human DNA consists of about 3 billion bases forming a 2 meter long helical biopolymer.

Evidently, DNA is neither simple as a molecule alone nor in the way it arranges itself within living organisms. However, there are some characteristics of the double-helix structure that remain as such in most cases, and we will assume so in this report. The diameter of the helix is $20 \text{ \AA}$ and the vertical distance between consecutive bases $3.4 \text{ \AA}$. One turn of the helix consists of 10 bases i.e. $30.4 \text{ \AA}$ and the twist angle is normally 36° .

The order or the sequence of the DNA bases conveys most of the information that is required for the operations it has to process. The bases are the alphabet of the DNA code and slightly altered arrangements could transform a perfectly healthy cell into a cancerous one. The DNA sequence, the exact way in

which these four bases are aligned, is a determining and crucial factor to the type of job that DNA molecule has to carry out.

1.2 The Link Between DNA and Physics

Especially after the discovery of its molecular structure in 1953 by Watson and Crick [5], which contributed so much to our understanding, DNA has been a compelling subject for research in biological sciences. Yet quite a mysterious one, too. Although biologists have made numerous and remarkable discoveries about the nature, properties, behaviour and function of DNA, there are still many unanswered questions. An issue of scientific concern and a key one to this report is the exact way in which DNA and other biomolecules such as proteins signal one another in order to carry out the required operations. Both to scientists and to the untrained eye it seems fascinating how DNA-binding proteins find their way across thousands of bases in a DNA sequence and know exactly where to attach themselves to, i.e. their target, which could be a fragment as short as a few tens of bases. There is a signalling method, not apparent to us, that guides processes like these. Additional complexity is introduced when one considers the shape the DNA strands adopt as part of a living organism. As explained in the previous Section, this long molecule is coiled and coiled again so tightly that the 2 meters of length that comprise it, can fit into a few micrometers within the cell nucleus. The movement of DNA and RNA (Ribose-Nucleic Acid), this folding and unfolding that make them function more efficiently and the factors that initiate these processes are crucial and central for the deeper understanding the scientific community requires.

Despite the automatic association of DNA with biological sciences only, statistical mechanics, non-linear physics and mathematics have already embraced these issues [6,7]. The twists and twirls, the bending, the precise manner in which it can fit into restricted space and the motion that governs all these are aspects of

DNA that can further be studied through the physics and mathematics lens. All this deals with observing the way DNA behaves and trying to describe it mathematically. In this study, we are mostly concerned about the exact, direct signalling strategies these complex molecules use in order to perform their functions in living systems rather than their movement. The nature of the communication between proteins and DNA is not known and could be lying in the realms of chemistry, physics, biology or most likely of all of them combined. This interdisciplinary combination makes an already complex system more complex altogether.

More specifically, we are looking at the possibility that DNA interacts with other biomolecules through some kind of electron exchange or transfer, as has already been reported [8]. The fact that there is no established scientific explanation or experimental consistency could be expected considering the difficulty of detecting these effects in a living organism and the spontaneity and speed with which they would take place. There are experimental studies where cancerous DNA on the skin repairs itself through the aid of electron transfer [9, 10]. (Skin DNA serves as an excellent subject to experiments because it is much easier to observe and interact with compared to DNA in inner organs). Moreover, it has been observed that damaged DNA strands show significantly different electronic behaviour to healthy DNA if an electric current is passed through [11]. Damaged DNA means that some of the bases or whole fragments of the original sequence are missing so that not all the bonds between the bases of the two wound chains that make the double-helix are present. As a result the molecule is single stranded at places which have lost their complementary part. With this observation in mind, the same group [11] showed that by detecting the damaged regions of DNA by means of passing an electric current through, which can flow in healthy regions but not in damaged ones, one can direct specific “healing” proteins to these regions and repair the damage.

Apart from the biological and medical implications that such a discovery

entails, the same issue, i.e. the capability of DNA to conduct or not, has been approached from a technological point of view. If one assumes that DNA does use some kind of electronic transfer to contact other molecules and that healthy double-stranded DNA lets current pass through, then one can picture numerous ways where this unexpected property of DNA could be utilized. Even if no breakthroughs concerning how these biological systems operate are ever made, using DNA's electronic properties separately presents a whole new possibility with immense applications. The biologists' insight that led them look into electronic transfer along the DNA molecule sparked off a lot of initiative in the nanoscience communities.

With the continuous and rapid progress of technology and nanotechnology, the need for size miniaturisation in electronic devices is becoming more and more palpable. The requirements that need to be met in terms of device size lie at the nanometer scale, the inevitable outcome of which is that the physical laws governing the processes that take place are not the same as in longer scales. The quantum nature of the components and thus of the electronic transport itself is the effect that rules the physics at this level. Molecular electronics is a very promising route and a persistently growing field in nanotechnology. Individual molecules used as switches, rectifiers, wires or memories are currently under the microscope waiting to mark the nanoscience era with their exceptional properties. Carbon nanotubes, fullerenes and various biomolecules are being investigated as potential molecules for the purpose of nanoelectronics. DNA constitutes an excellent candidate biomolecule for usage in nanotechnology, as long as its electronic properties allow so. As explained by Porath *et al.* [12], as a biological molecule, DNA has two unique properties that would be extremely valuable if successfully employed into nanodevices: recognition and self-assembly. Recognition is the property of the molecule to store information and respond accordingly when interacting with other biomolecules, for example proteins. Self-assembly is the capability of molecules to

assemble spontaneously, based on their chemical and structural properties, until a stable aggregate of minimum energy is reached. It is clear how these properties could be effectively utilised in nanodevice implementation. Both recognition and self-assembly could drive the construction of circuits from individual components due to their ability to store information, select the right component and organise the different molecules into more complex configurations. Thus increased efficiency within reduced-in-size devices is achieved.

Although the electronic properties of DNA are still a mystery, the wish to use such a molecule in current technology has driven scientists to try and force the desired properties out of DNA anyway. More specifically, DNA could be chemically engineered, for instance doped with suitable elements, in order to *become* e.g a conductor, a semiconductor etc. regardless of what its natural behaviour would be [13]. If we could chemically control the electronic properties of DNA then the applications in nanotechnology would be guaranteed.

However, there is one crucial question that is inevitably posed when DNA applications in nanotechnology are considered. Would the molecule, once integrated into circuits and devices or having been chemically doped, actually maintain these properties that are present when operating in a biological system i.e. recognition and self-assembly, these same properties that made DNA a good candidate in the first place?

1.3 Previous Studies and Results on the Electronic Properties of DNA

In the research where DNA's electronic properties are under investigation, one can distinguish two categories of experimental activity; when the DNA is in solution, mostly performed by chemistry and biophysics communities and when it is dry where we find direct conductivity measurements, mostly carried out by physicists.

In solution experiments, DNA is chemically processed to host a donor and an acceptor molecule at different sites along its long axis. Photoinduced charge transfer rates are then measured whilst the donor/acceptor molecules, the distance and the sequence of DNA that lies between them are varied. The reactions depend on the type of DNA used, the intercalation, the integrity of the intervening base pair stack but not so much upon the molecular distance. Probably the most active group in this area is that of J. K. Barton and a large number of papers have been published explaining these photoinduced reactions that include electron transfer [11,14–17].

In the past few years there have been many endeavours to directly measure the conductivity of dry DNA by applying a voltage between two metal, typically golden, contacts, as prescribed by solid state experiments. However, the results of these experiments show remarkable diversity. The problem focuses on the fact that each is performed under very different conditions. More specifically, the alignment and drying techniques, the types and lengths of DNA, the apparatus and devices used to measure the conductivity are all factors that induce further confusion and inconsistency in an already complex field of research. As far as the characterisation of its electronic behaviour is concerned, there are reports on DNA being an insulator [18], an ohmic conductor [13,19–22] and a semiconductor with bandgaps of 2V [23] and a few hundreds of millivolts [13].

Different alignment techniques include molecular combing [21], printing on a film [22] and “gluing” the sticky ends of the DNA molecule onto surface-bound oligomers [13]. Types of DNA that have been used both experimentally and theoretically are amongst others λ -DNA [13,19,20,24], poly(dG)-poly(dC) DNA [23,25–27] and also engineered forms of doped DNA [13] of lengths between tens to hundreds of nanometers.

Similar diversity characterises the theoretical understanding and models for charge transport in DNA. We find interpretations of chemistry solution experiments

for DNA-mediated charge migration as single-step superexchange [28] and multi-step hole hopping [29]. There are also several theoretical models accounting for the direct conductivity measurement experiments. DNA has been viewed as a 1-D disordered system where a variable range hopping mechanism and a dependence of localisation lengths upon temperature is proposed [30]. A fishbone-like structure, in a quantum mechanical regime, with backbone disorder, has been considered [3] which is in agreement with the semiconductor experimental results of Porath *et al.* [23] and the same model has been used to suggest a transition from an insulating to a conducting state above a certain threshold backbone disorder [31]. Quantum mechanical approaches to the problem that use standard one-dimensional (1-D) tight-binding models, [25–27, 32, 33] have also been proposed. Both theoretical and experimental studies have shown that the environment of the DNA i.e. the solution, the temperature etc. has significant effects mostly on the backbone which in other words means that the structure and hence the properties of the DNA that one might be measuring such as the conductivity are also affected [30, 34, 35].

Finally, a number of excellent review papers have recently been published, describing the history, milestones and current theoretical and experimental progress in the field of DNA electronics [12, 36, 37].

Chapter 2

The Disorder Perspective

2.1 Electronic Transport in the Quantum Regime

In solid state physics, one of the most important and well-studied subjects is that of electronic transport. Electricity itself but in fact all technological paraphernalia make use of the discoveries of solid state physics with materials like semiconductors and superconductors and of course conductors and insulators. Localisation is a property of the states in a random quantum mechanical system and the theory based on Anderson's findings in 1958 [38] and Mott's developments ten years later [39] explains the reduction in the observed macroscopic conductivity in disordered solids.

The system is quantum mechanical and this means that the particles, i.e. the electrons, are described in terms of their wavefunctions, their wavefunction amplitudes and probabilities. The term disorder is a macroscopic measure of the entropy of a quantum mechanical system. A theoretical crystal lattice is perfect but in reality one finds impurities, vacancies and dislocations that affect the system's transport properties and thus need to be accounted for. Disorder is a measure of all these defects which otherwise ideal crystals display. In addition, disorder can be viewed to be connected with the amount of randomness in the positions

of the atoms in a crystal lattice regardless of their strength. A highly disordered material would be one with all its atoms at totally random positions, like glasses for example. Modelling certain types of disorder theoretically is quite a different issue and will be discussed in Section 3.2.

According to localisation theory, a quantum state in a solid can be characterised to be extended or localised in space which means exactly the degree to which the wavefunction is spatially extended before it decays completely. The square of the amplitude of the electron's wavefunction is of course proportional to the probability of the electron being in that region in space. Localised particles are confined in a region in space and extended ones are not. This will be easier to understand once the following two extremes are described; insulators and conductors.

Insulators have localised states at $T = 0$ and the dc. conductivity vanishes. Hence, insulators cannot conduct at zero temperature. However, for $T > 0$, thermal activation of the charge carriers results in giving finite values for the dc. conductivity. At the limit where $T \rightarrow 0$, the thermal processes have more or less faded away and conduction is possible via the hopping of charge carriers between localised states [40].

Metallic behaviour is a little different. At $T = 0$, conductors have extended states and the dc. conductivity is finite. For $T > 0$ and opposite to insulators, the dc. conductivity in metals decreases with increasing temperature. The thermal processes that contribute to the transport in insulators have the exact opposite effect in metals; they add entropy to an otherwise already conducting system. Ideal metallic behaviour with maximum conductivity is observed at the absolute zero of temperature. One can summarise the above by making the useful distinction where:

- if at $T = 0$, states near the Fermi energy of the system are localised $\Rightarrow$ we have insulating behaviour.

- if at $T = 0$, states near the Fermi energy of the system are extended $\Rightarrow$ we have metallic behaviour.

2.2 Localisation and Disordered Systems

A definition of the effect of localisation would be: the amplitudes of the one-particle wavefunctions in macroscopic disordered quantum systems at the absolute zero of temperature are exponentially decaying at infinity [39]. The significance of this theory lies in the fact that it connects the system's disorder to the effective localisation of the particles' quantum states in space and macroscopically to the observed reduced conductivity. More specifically, the presence of disorder induces the localisation of quantum mechanical wavefunctions due to self-interference; as explained in the last paragraph this implies insulating behaviour. When the initial state is metallic, i.e. the wavefunctions are extended, and the disorder is introduced, then the wavefunctions eventually become localised and what we observe is a metal-insulator transition (MIT). The transition describes the switch in the electronic transport properties from metallic to insulating as a result of disorder. Energetically the edge between localised and extended states is called the mobility edge. In the extended states regime, when conduction takes place, the appropriate quantity to measure/calculate is the dc. conductivity σ_{dc} ; whereas in the localised regime, the actual transport is diminished and conductivity as a quantity no longer makes sense. The concept of the localisation length is thus introduced as a measure, describing the extent of the localisation of the particle's wavefunction and is given by

$$\psi(x) \sim \exp\left(-\frac{|x - x_0|}{\lambda}\right) \quad (2.1)$$

where x_0 is the center of the localisation envelope and λ is the localisation length [41].

2.3 Scaling Hypothesis of Localisation

From Ohm's law we know that the current that flows through a conductor is proportional to the voltage applied at its ends; and the constant of proportionality is the resistance, $V = IR$. We define the conductance of a material to be the reciprocal of its resistance, i.e. $G = 1/R$. Similarly, the conductivity of a material is defined as the reciprocal of its resistivity, $\sigma = 1/\rho$. The resistivity itself in a three-dimensional wire with cross sectional area A and of length L and resistance R , is given by

$$\rho = \frac{RA}{L} \quad (2.2)$$

and generalising this for dimensions d , we have

$$\rho = RL^{d-2} \Leftrightarrow \rho = \frac{L^{d-2}}{G} \Rightarrow G = \sigma L^{d-2} \quad . \quad (2.3)$$

Using the fact that the conductance is quantised in multiples of e^2/h , the dimensionless conductance g can be introduced

$$G = g \frac{e^2}{h} \quad (2.4)$$

where e is the electronic charge and h is Planck's constant. Equation 2.3 can be written as

$$g = \left(\frac{h}{e^2} \sigma \right) L^{d-2} \quad . \quad (2.5)$$

Evidently this relationship applies to metals only or, in this context, to the metallic, extended states regime. In other words Equation 2.5 gives the relation between the maximum conductance and maximum conductivity. In the localised regime, on the other hand, where the most suitable quantity is the localisation length λ and not the conductivity, we will have

$$g \sim \exp - (L/\lambda) \quad . \quad (2.6)$$

In order to investigate further, we need to define $\beta(g)$

$$\beta(g) = \frac{d(\ln g)}{d(\ln L)} \quad . \quad (2.7)$$

If we just look at the behaviour of this function, we can see that for $\beta > 0$, or the slope of the function $\ln g(L)$ being positive, $\ln g(L)$ is a monotonously increasing function, which implies that $\ln g$ increases with increasing $\ln L$. More qualitatively, the conductance increases as the length of the wire increases, which is what we expect for Ohmic behaviour. From Equation 2.5, which characterises the Ohmic conductor, we can find the maximum value of $\beta(g)$; and that only holds for the metallic part of the spectrum. So taking the natural logarithm of Equation 2.5

$$\ln g = \ln \left(\frac{h}{e^2} \sigma \right) + (d - 2) \ln L \quad (2.8)$$

and in order to obtain $\beta(g)$, differentiate Equation 2.8 with respect to $\ln L$

$$\frac{d(\ln g)}{d(\ln L)} = d - 2 \quad (2.9)$$

since σ and h/e^2 are constant. Therefore for β positive, g increases with L until β reaches its maximum at $\beta = d - 2$ indicating we are in the metallic-behaviour regime. For $\beta < 0$, the slope of the function $\ln g(L)$ is negative, and that implies that $\ln g(L)$ decreases with increasing $\ln L$; or the conductance decreases as the length of the sample is increased. This behaviour is certainly not Ohmic and thus we conclude that it must be in the insulating regime. Therefore we must look at Equation 2.6 or most appropriately take the natural logarithm of both sides in order to link with the β function

$$\ln g \sim \left| -\frac{L}{\lambda} \right| \quad (2.10)$$

$$L \sim \lambda \ln g \quad (2.11)$$

Taking the natural logarithm once more on both sides, we get

$$\ln L \sim \ln \lambda \ln(\ln g) \quad (2.12)$$

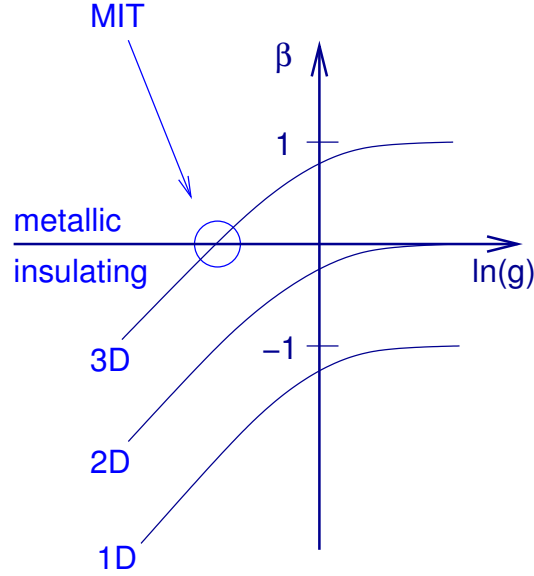


Figure 2.1: Function β versus the natural logarithm of the dimensionless conductance. For $\beta > 0$ the behaviour is metallic whereas for $\beta < 0$ it is insulating. For a two-dimensional sample disorder can induce a metal insulator transition (MIT).

and differentiating this with respect to $\ln g$

$$\frac{d(\ln L)}{d(\ln g)} \sim 0 + \frac{1}{\ln g} \quad (2.13)$$

or

$$\frac{d(\ln g)}{d(\ln L)} = \ln g \quad . \quad (2.14)$$

Hence, in the localised regime, the conductance g decreases with increasing length of sample L until β reaches the value $\ln g$, as shown in Equation 2.14. At the critical point where $\beta = d(\ln g)/d(\ln L) = 0$, we have a stationary point, where the behaviour of the β function, or the slope of $\ln g(L)$ switches from being positive into being negative, and that is exactly where the MIT occurs.

In this thesis the basic concepts of localisation theory are used to describe the electronic transport in DNA in a quantum mechanical regime. The calculated localisation lengths give us a qualitative measure of the tendencies of this system.

For example, if the localisation lengths tend to infinity, then the states are extended and thus transport is very probable.

2.4 The Transfer Matrix Method

In the previous section we saw how localisation theory expands our understanding of the electronic properties of solids. Moreover, concepts such as disorder and the localisation length λ were introduced. However, the relations/equations of the last section can only be applied to crystals that deviate a little from the ideal crystalline structure. In these cases, the defects are treated as perturbations and in order to apply perturbation theory, *the perturbation must be relatively small*. But if one wants to look at more randomly disordered materials such as glasses, one cannot use perturbation theory or in fact study the problem analytically whilst including the localisation phenomenon of the quantum states.

Here, a numerical approach to this issue will be explained, which is also the one we have used to perform our calculations, called the transfer-matrix method (TMM). TMM is a way of studying random systems by making use of statistical analyses of the eigenvalues and eigenstates of random operators. Firstly, since the system under consideration is a quantum mechanical system, we start off from Schrödinger's equation

$$-\frac{1}{2m}\nabla^2\psi(x) + V(x)\psi(x) = E\psi(x) \quad (2.15)$$

where $\psi(x)$ is the wavefunction of the particle at position x , m is the mass of the particle, $V(x)$ is the potential energy at position x and E is the energy. When studying crystal lattices it is more convenient to use the site representation of Schrödinger's equation [40], which for a one-dimensional system at site n is

$$t\psi_{n+1} + t\psi_{n-1} + \varepsilon_n\psi_n = E\psi_n \quad (2.16)$$

The connection with the original Schrödinger's equation, is that now the position

x is replaced with the site position n , through the assumption that $x = na$, where a is the lattice constant. The coefficients t denote the transfer probability amplitudes from a site n to its neighbouring sites $n - 1$ and $n + 1$. The onsite potential parameters ε_n equal the potential energy term $V(x)$, which are often chosen from a box distribution [40]

$$P(\varepsilon_n) = W^{-1}\Theta(W/2 - |\varepsilon_n|) \quad (2.17)$$

where W is the width of the distribution and a measure of the disorder in the system and Θ is a step function. The disorder W reflects the extent of the fluctuations in the potential energy terms, corresponding to the number/strength of impurities in the lattice. Equation 2.16 can be written in matrix form

$$\begin{pmatrix} \psi_{n+1} \\ \psi_n \end{pmatrix} = \begin{pmatrix} \frac{E-\varepsilon_n}{t} & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \psi_n \\ \psi_{n-1} \end{pmatrix} = \mathbf{T}_n \begin{pmatrix} \psi_n \\ \psi_{n-1} \end{pmatrix} \quad (2.18)$$

It is worth noting how the wavefunctions are expressed as vectors grouped in two pairs, depending on their positions in the lattice. So the left-hand side contains the future (ψ_{n+1}) and present (ψ_n) wavefunctions which can be obtained from the right-hand side, where a matrix is multiplied with the present (ψ_n) and past (ψ_{n-1}) wavefunctions. $\mathbf{T}_n$ is called the transfer matrix. From the way Equation 2.18 is arranged, one can see that at each site we look at, we need to know information from the previous two, so in order to obtain ψ_5 , we write it in terms of ψ_4 and ψ_3 . In order to get ψ_4 we write it in terms of ψ_3 and ψ_2 and so on. Every time this iteration is carried out an extra transfer matrix is multiplied into the right-hand side of the equation. Let us look at an example with $n=3$. We have

$$\begin{pmatrix} \psi_4 \\ \psi_3 \end{pmatrix} = \begin{pmatrix} \frac{E-\varepsilon_3}{t} & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \psi_3 \\ \psi_2 \end{pmatrix} = \mathbf{T}_3 \begin{pmatrix} \psi_3 \\ \psi_2 \end{pmatrix} \quad (2.19)$$

Now ψ_3 and ψ_2 are given by

$$\begin{pmatrix} \psi_3 \\ \psi_2 \end{pmatrix} = \mathbf{T}_2 \begin{pmatrix} \psi_2 \\ \psi_1 \end{pmatrix} \quad (2.20)$$

So that finally one gets

$$\begin{pmatrix} \psi_4 \\ \psi_3 \end{pmatrix} = \mathbf{T}_3 \times \mathbf{T}_2 \times \mathbf{T}_1 \times \begin{pmatrix} \psi_1 \\ \psi_0 \end{pmatrix} \quad (2.21)$$

Let us define the product matrix T_N as

$$T_N = \prod_{n=1}^N \mathbf{T}_n \quad (2.22)$$

For this product matrix Oseledec's theorem [40] states the existence of a limiting matrix

$$\Gamma \equiv \lim_{N \rightarrow \infty} (T_N T_N^\dagger) \quad (2.23)$$

with eigenvalues $\exp(\gamma_m)$, where γ_m are called the Lyapunov exponents and $T_N^\dagger$ the adjoint of T_N , which in this case is the same as the transpose of T_N because the matrix is real. The significance of the theorem lies in the fact that we can know the relationship between the initial states wavefunction and the wavefunction at *any later site* only by calculating the product matrix. The localisation length is connected to the two Lyapunov exponents γ_0 and γ_1 with

$$\lambda \equiv -\gamma_0^{-1} \equiv \gamma_1^{-1} \quad (2.24)$$

$\gamma_0 = -\gamma_1$ due to symplecticity [40]. So, the larger the Lyapunov exponent, the more localised the wavefunction, the smaller the localisation length.

In order to apply the TMM to two and three dimensional systems, we need to extend the above calculation to quasi-1D geometries. This means that we have a strip of length N representing the sites, as before, but the strip has also width M , where $N \gg M$. At each site N there are M ways to distribute the wavefunction ψ_n . If we look at the diagramme of Figure 2.2, at a site N and for $M = 3$, ψ_N can be at $\psi_{N,1}$, $\psi_{N,2}$ or $\psi_{N,3}$. So each amplitude ψ_n in the 1-D system, here is represented by a column vector. For the case $M = 3$, at site $n = N$, the amplitude

matrix would be

$$\begin{pmatrix} \psi_{N,1} \\ \psi_{N,2} \\ \psi_{N,3} \end{pmatrix} \quad (2.25)$$

With this in mind, one needs to extend Schrödinger's equation, given in Equation 2.16, to the quasi-1D system. For example, for the system under consideration, shown in Figure 2.2 with $M = 3$, the first term in Equation 2.18 would become for $N = n$

$$\begin{pmatrix} \psi_{n+1} \\ \psi_n \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} \psi_{N+1,1} \\ \psi_{N+1,2} \\ \psi_{N+1,3} \end{pmatrix} \\ \begin{pmatrix} \psi_{N,1} \\ \psi_{N,2} \\ \psi_{N,3} \end{pmatrix} \end{pmatrix} \quad (2.26)$$

Hence the present and future amplitude vector is a $2M \times 1$ matrix. Following the rules of matrix multiplication, $\mathbf{T}_n$ has got to be a $2M \times 2M$ matrix and so does T_N . The product matrix will thus have $2M$ eigenvalues, which due to symplecticity, appear in pairs of $\exp(\gamma_m)$ and $\exp(-\gamma_m)$. The smallest positive Lyapunov exponent gives the largest characteristic length, the localisation length λ_M .

Numerically, the eigenvalues of the product matrix are calculated by consecutive multiplications of the transfer matrices on a set of M orthogonal initial amplitudes. However, when this is carried out starting with only one initial vector we find that the series of multiplications alter the result to yield the largest eigenvalue whereas we need the smallest positive Lyapunov exponent in order to determine the localisation length. A way out from this problem is to preserve orthogonality throughout the iteration process by forcing the vectors to become orthogonal, after a number of multiplications each time.

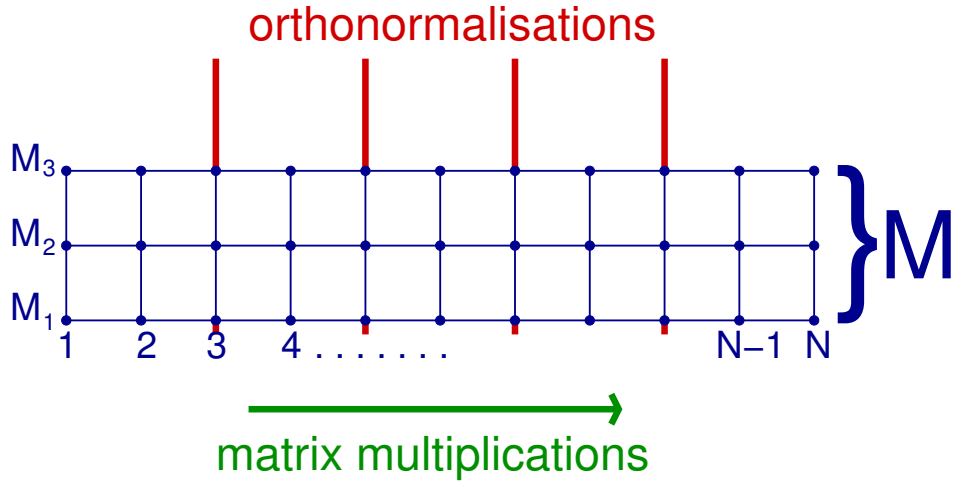


Figure 2.2: The quasi-one dimensional strip of length N and width $M = 3$. The wavefunction at each site N can be distributed in space in M different ways. Transfer matrix method makes sure the initial orthonormal vectors are reorthonormalised every few (usually 10) sites.

2.5 Tight-Binding Models for the Electronic Transport Along DNA

In this study, both of our DNA models use standard tight-binding Hamiltonians and nearest-neighbour interactions for electronic transport of non-interacting particles in one direction. Although the 1-channel model is widely used in solid state physics, it was firstly proposed as a model to describe transport in DNA by Cuniberti *et al.* [3] and is shown in Figure 2.3. The 2-channel model, see Figure 2.4, is an extension of the previous one but with a biologically more relevant structure considering the arrangement of the components of the DNA molecule, as in Figure 1.2.

The models are designed so that charge migration takes place in one direction, the long axis of the DNA molecule, and only via consecutive bases. This is a consequence of a very important assumption that has been made, which is that the bases display overlap in their electronic π -orbitals. Density Functional Theory, (DFT), calculations have demonstrated this observation [12, 24], where

the bases, especially Guanine, provide the most likely way for electrons to travel along a DNA fragment. So far all studies, experimental and theoretical, to the best of our knowledge, seem to agree on that assumption apart from one [42], where the transport is via the sugar-phosphate backbone.

2.5.1 The One-channel “Fishbone Model”

The 1-channel model is arranged in a way such that there is only one branch for transport in the forward direction and it is shown in Figure 2.3. The sites on the conduction branch are interconnected and each represents a base pair. For example, site i could be an A-T pair and site $i+1$ a G-C pair or vice versa. Each base pair site is further linked to an upper and lower site, representing the upper and lower backbone. The backbone sites are not interconnected since we have assumed that they do not serve as a path for electron migration. Each site has a potential and every link between sites implies the presence of a hopping amplitude. The Hamiltonian for the 1-channel model is given by

$$H_F = - \sum_{\langle i,j \rangle}^L t_{ij} |i\rangle \langle j| - \sum_{q=\uparrow,\downarrow} \sum_{i=1}^L t_i^q |q_i\rangle \langle i| + h.c. + \sum_{i,q} \varepsilon_i |i\rangle \langle i| + \varepsilon_i^q |q_i\rangle \langle q_i| \quad (2.27)$$

where t_{ij} is the hopping between nearest-neighbour sites i, j along the central branch, t_i^q with $q = \uparrow, \downarrow$ gives the hopping from each site on the central branch to the upper and lower backbone respectively; ε_i denotes the onsite energy at each site along the central branch and ε_i^q , with $q = \uparrow, \downarrow$, gives the onsite energy at the sites of the backbone. L is the number of sites/bases in the sequence, $h.c.$ is the hermitian conjugate and each hopping amplitude is assumed to act onto the corresponding wavefunction of each site. Writing Schrödinger's equation in lattice form as in Equation 2.16, for the i th site in the central branch, the upper and lower backbone we have

$$t_{i+1}\psi_{i+1} + t_i\psi_{i-1} + t_i^\uparrow\psi_i^\uparrow + t_i^\downarrow\psi_i^\downarrow + \varepsilon_i\psi_i = E\psi_i \quad (2.28)$$

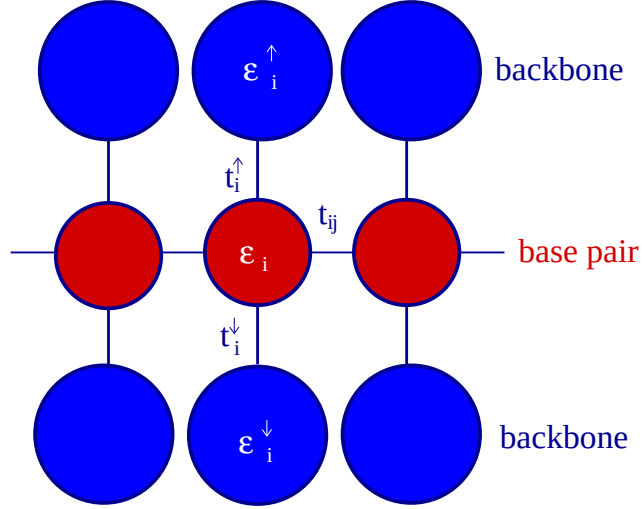


Figure 2.3: The fishbone model for electronic transport along DNA. The model corresponds to the Hamiltonian given in Equation 2.27. It was first used as a model for DNA by Cuniberti *et al.* [3]

$$t_i^\uparrow \psi_i + \varepsilon_i^\uparrow \psi_i^\uparrow = E \psi_i^\uparrow \quad , \quad (2.29)$$

$$t_i^\downarrow \psi_i + \varepsilon_i^\downarrow \psi_i^\downarrow = E \psi_i^\downarrow \quad , \quad (2.30)$$

respectively. Expressing $\psi_i^\uparrow$ and $\psi_i^\downarrow$ in terms of ψ_i

$$\psi_i^\uparrow = \frac{t_i^\uparrow}{E - \varepsilon_i^\uparrow} \psi_i \quad , \quad \psi_i^\downarrow = \frac{t_i^\downarrow}{E - \varepsilon_i^\downarrow} \psi_i \quad (2.31)$$

Substituting these into Equation 2.28 and rearranging yields

$$t_{i+1} \psi_{i+1} + \left(\frac{(t_i^\uparrow)^2}{E - \varepsilon_i^\uparrow} + \frac{(t_i^\downarrow)^2}{E - \varepsilon_i^\downarrow} + \varepsilon_i - E \right) \psi_i + t_i \psi_{i-1} = 0 \quad (2.32)$$

which is arranged so that: $FUTURE + PRESENT + PAST = 0$. Defining $\tilde{\varepsilon}$

$$\tilde{\varepsilon} = \varepsilon_i + \frac{(t_i^\uparrow)^2}{E - \varepsilon_i^\uparrow} + \frac{(t_i^\downarrow)^2}{E - \varepsilon_i^\downarrow} \quad (2.33)$$

and substituting into Equation 2.32 and we have

$$\psi_{i+1} = \frac{(E - \tilde{\varepsilon}) \psi_i - t_i \psi_{i-1}}{t_{i+1}} \quad . \quad (2.34)$$

We then express Equation 2.34 in matrix form to get the recursive formula shown below

$$\begin{pmatrix} \psi_{i+1} \\ \psi_i \end{pmatrix} = \begin{pmatrix} \frac{E-\tilde{\epsilon}}{t_{i+1}} & -\frac{t_i}{t_{i+1}} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \psi_i \\ \psi_{i-1} \end{pmatrix} \quad (2.35)$$

Equation 2.35 is then used to calculate the amplitude of the wavefunction at each site. It links the present and future amplitude to the past and present amplitude through the 2×2 transfer matrix shown on the right hand side.

2.5.2 The Two-channel “Ladder Model”

The ladder model is a planar projection of the structure of DNA with its double-helix unwound, as shown in Figure 2.4. The helical structure of DNA is very important and certainly plays an important role in the molecule’s properties. However, the complexity of the structure makes the study extremely complicated - so much so that we believe it is unnecessary and misleading at this early stage. In the 2-channel model, each base occupies a site, so that there are *two* central conduction branches. The sites on the two central branches are connected horizontally and vertically and further linked to the upper and lower backbone sites as in the 1-channel model; again the backbone sites are not interconnected. A sketch of the ladder model can be seen in Figure 2.4. The Hamiltonian for the 2-channel model is given by

$$\begin{aligned} H_L = & \sum_{i=1}^L \left(\sum_{\tau=1,2} t_{i,\tau} |i, \tau\rangle \langle i+1, \tau| + \varepsilon_{i,\tau} |i, \tau\rangle \langle i, \tau| \right. \\ & + \sum_{q=\uparrow, \downarrow} t_i^q |i, q\rangle \langle i, q| + t_{1,2} |i, 1\rangle \langle i, 2| + \varepsilon_i^q |i, q\rangle \langle i, q| \Big) \\ & + h.c. \end{aligned} \quad (2.36)$$

where $t_{i,\tau}$, $\tau = 1, 2$, is the hopping amplitude from one site to the next along each branch τ , t_i^q is the hopping amplitude to the upper ($q = \uparrow$) and lower ($q = \downarrow$)

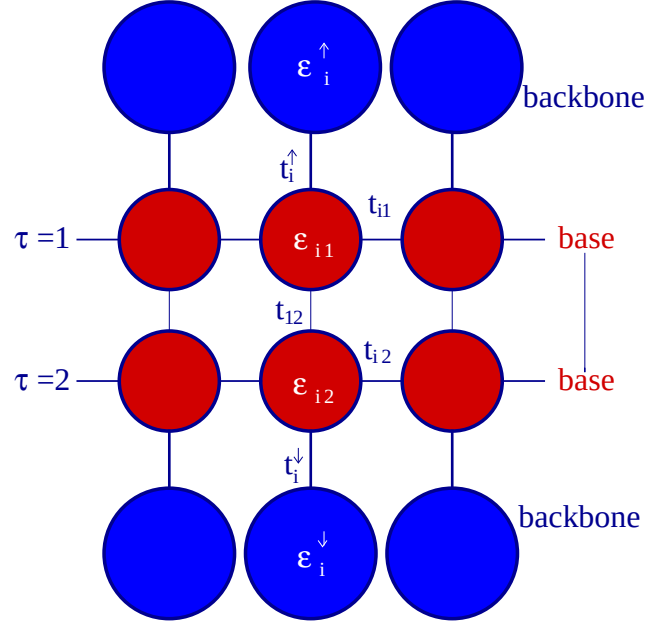


Figure 2.4: The ladder model for electronic transport along DNA. The model corresponds to the Hamiltonian given in Equation 2.36.

backbone, t_{12} represents the hopping between the two central branches (which is perpendicular to the direction of conduction) i.e. an electron going from site $|i, 1\rangle$ to site $\langle i, 2|$ or vice versa; $\varepsilon_{i,\tau}$ is the onsite potential energy on each site along the two central branches and ε_i^q gives the onsite energy at the sites of the backbone.

What used to be the central branch in the 1-channel model is now split into two and we will therefore have one extra equation. If we write down Schrödinger's equation in site representation, we get the following four equations for site i on branches $\tau=1$ and $\tau=2$ and for the upper and lower backbone respectively

$$t_{i+1}^{(1)}\psi_{i+1}^{(1)} + t_i^{(1)}\psi_{i-1}^{(1)} + t_i^\uparrow\psi_i^\uparrow + t_{12}\psi_i^{(2)} + \varepsilon_i^{(1)}\psi_i^{(1)} = E\psi_i^{(1)} \quad (2.37)$$

$$t_{i+1}^{(2)}\psi_{i+1}^{(2)} + t_i^{(2)}\psi_{i-1}^{(2)} + t_i^\uparrow\psi_i^\uparrow + t_{12}\psi_i^{(1)} + \varepsilon_i^{(2)}\psi_i^{(2)} = E\psi_i^{(2)} \quad (2.38)$$

$$t_i^\uparrow\psi_i + \varepsilon_i^\uparrow\psi_i^\uparrow = E\psi_i^\uparrow \quad (2.39)$$

$$t_i^\downarrow\psi_i + \varepsilon_i^\downarrow\psi_i^\downarrow = E\psi_i^\downarrow \quad (2.40)$$

Equations 2.39 and 2.40 are the same as Equations 2.29 and 2.30. We already have $\psi_i^\uparrow$ and $\psi_i^\downarrow$ in terms of ψ_i from Equation 2.31. Substitute these in Equations 2.37 and 2.38 and rearrange to get for the upper branch, $\tau = 1$

$$t_{i+1}^{(1)}\psi_{i+1}^{(1)} = \left[E - \left(\frac{(t_i^\uparrow)^2}{E - \varepsilon_i^\uparrow} + \varepsilon_i^{(1)} \right) \right] \psi_i^{(1)} - t_{12}\psi_i^{(2)} - t_i^{(1)}\psi_{i-1}^{(1)} \quad , \quad (2.41)$$

and for the lower branch $\tau = 2$

$$t_{i+1}^{(2)}\psi_{i+1}^{(2)} = \left[E - \left(\frac{(t_i^\downarrow)^2}{E - \varepsilon_i^\downarrow} + \varepsilon_i^{(2)} \right) \right] \psi_i^{(2)} - t_{12}\psi_i^{(1)} - t_i^{(2)}\psi_{i-1}^{(2)} \quad . \quad (2.42)$$

Let us define the effective disorder for each branch:

$$\tilde{\varepsilon}_i^{(1)} = \varepsilon_i^{(1)} + \frac{(t_i^\uparrow)^2}{E - \varepsilon_i^\uparrow} \quad , \quad (2.43)$$

$$\tilde{\varepsilon}_i^{(2)} = \varepsilon_i^{(2)} + \frac{(t_i^\downarrow)^2}{E - \varepsilon_i^\downarrow} \quad . \quad (2.44)$$

So Equations 2.41 and 2.42 become

$$t_{i+1}^{(1)}\psi_{i+1}^{(1)} = \left(E - \tilde{\varepsilon}_i^{(1)} \right) \psi_i^{(1)} - t_{12}\psi_i^{(2)} - t_i^{(1)}\psi_{i-1}^{(1)} \quad , \quad (2.45)$$

$$t_{i+1}^{(2)}\psi_{i+1}^{(2)} = \left(E - \tilde{\varepsilon}_i^{(2)} \right) \psi_i^{(2)} - t_{12}\psi_i^{(1)} - t_i^{(2)}\psi_{i-1}^{(2)} \quad . \quad (2.46)$$

These two equations can be contracted into one, if written in matrix form and the wavefunctions written as vectors. The 2-channel model is actually a quasi-1D system, in the sense that numerically we have a strip of length N , which is the number of bases in the sequence and width $M = 2$, accounting for the two central branches. We have extended the theory of section 2.4 in order to include the backbone sites. Thus following the TMM, the amplitude of the wavefunction at each site can be computed.

One difference with the 1-channel model is that depending on which branch the particle is, there are hopping amplitudes for either the upper or the lower backbone but not both. The backbone hopping term that is missing is replaced by the perpendicular hopping between the two branches.

Chapter 3

Numerical Details

3.1 DNA Sequences

In the various parts that this project involves, five specific sequences of DNA have been used: poly(dG)-poly(dC), telomeric, random-ATGC, λ -DNA and scrambled λ -DNA.

In biology there are millions of identified DNA sequences, parts of which have certain functions and parts that seem to have no purpose, but which are most probably essential for the maintenance and function of the whole molecule. One can also generate random or correlated sequences numerically on a computer. In this theoretical study of DNA's electronic properties we have aimed at using sequences that facilitate the comparison with experimental results and previous theoretical studies. In this section we explain why the above sequences were used; a schematic summary is shown in Table 3.1.

Poly(dG)-poly(dC), in this report, is a computer generated random arrangement of the bases G and C only, over a total length of 10,000 bases. It was generated using Mathematica, the code is shown in Appendix A. Sequences that contain only G and C bases have been extensively used in experimental and theoretical studies concerning electronic transfer [43, 44], as also described in Section

1.3. The reason is that, theoretically, consecutive Guanine bases demonstrate the largest π -orbital overlap and thus serve as a better conduit for electronic transfer, [12,24]. Although extended versions of this sequence are not found anywhere in nature, they are easily synthesised in a laboratory. However, the sequences containing only G and C bases that are reported in other studies are not identical to the poly(dG)-poly(dC) sequence that is used here. We believe that this slight inconsistency is not significant enough to affect the comparison between our calculations and previously reported ones. (We do not use the term random for this sequence; its randomness is reduced because only two bases are included).

One example of a Guanine rich sequence in nature is telomeric DNA, which is a real biological fragment found at the ends of the human chromosome and consists of a sequence of bases with the pattern TTAGGG repeated many times up to many thousand bases; the telomeric sequence we use consists of 6,000 bases. Telomeric DNA contains all four bases but its behaviour is expected to be affected by the fact that there are three Guanine bases, grouped together, every other three bases in the sequence. The sequence of one strand of telomeric DNA looks like this: TTAGGGTTAGGGTTAGGG...

λ -DNA is also a biological form of DNA, actually a virus called bacteriophage, that infects bacteria. λ -DNA is a commonly used sequence in experiments and theory because it is biologically well characterised. Most of the 48,502 bases that it contains serve functions that are well understood by biologists. The significance of this in the current study is that we can compare the localisation lengths of the parts that have specific operations and the remaining “junk” that is ubiquitous in real DNA.

Finally, we introduce and study two computer generated random sequences, which we call random-ATGC and scrambled λ -DNA. Random-ATGC is a 10,000 bases long sequence that contains all four bases A, T, G, C in equal amounts, arranged randomly. There are $4^{10,000}$ ways of arranging a 10,000 position se-

Name	Bases	Sequenced	Length	Comment
poly(dG)-poly(dC)	G,C	randomly	10,000 bases	computer generated
telomeric DNA	TTAGGG repeated	periodic	6,000 bases	found at ends of human chromosome
random-ATGC	A,T,G,C	randomly	10,000 bases	computer generated
λ -DNA	A,T,G,C	-	48,502 bases	virus
scrambled λ -DNA	A,T,G,C	-	48,502 bases	computer generated

Table 3.1: The four DNA sequences that we have used, poly(dG)-poly(dC), telomeric, random-ATGC, λ -DNA and scrambled λ -DNA; with some characteristics, composition of bases, the way these are arranged, the length (in bases) and additional comments.

quence of 4 different bases. However, what we call random-ATGC DNA is one randomly arranged sequence that was generated *once* and this same one has been used throughout this project. Scrambled λ -DNA is also a computer generated sequence; it is a 48,502 long sequence that has the same composition of bases as the bacteriophage λ -DNA but distributed randomly. The reason for making and using these two random sequences is so that we can compare with real DNA and look at differences concerning their electronic behaviour.

3.2 Fishbone and Ladder Models Input Files

The programme that has been used for the calculations is designed in a way so that there is a large number of fixed parameters that can be modified and a number of variables whose behaviour over an interval can be computed. In this study we have not made use of all the possible variations that the programme allows us to, due to time limitations. An example of the parameter-input file that is fed into the programme containing all the information on the parameters is shown in Table 3.2; a more precise description of which will be given later. In Ref. [3], it has been argued that $t_{i,\tau} \approx t_i^q/2 \sim 0.5\text{eV}$ can describe experimental results in poly(dG)-

NOOrtho = 10	DNAHopDis0 = 0.0	HopDN = 2.0
NOGamma = 1	DNAHopDis1 = 0.1	HopTA = 1.0
IKeepFlag = 0	dDNAHopDis = 0.0	HopCA = 0.5
IWriteFlag = 0	DNAHopPPP = 0	HopGA = 0.5
IRandomFlag = 0	BBSiteDis0 = 0.0	HopCT = 0.5
lInitFlag = 1	BBSiteDis1 = 0.01	HopGT = 0.5
ISortFlag = 2	dBBSiteDis = -0.01	HopGC = 1.0
IFluxFlag = 4	BBSitePPP = 2	HopAA = 1.0
IWidth = 2	BBHopDis0 = 0.0	HopTT = 1.0
ISite0 = 0	BBHopDis1 = 0.0	HopCC = 1.0
ISite1 = 6	dBBSiteDis = 0.5	HopGG = 1.0
dISite = 1	BBHopPPP = 0	OnSiteA = 0.0
DNASiteDis0 = 0.0	Energy0 = -5.0	OnSiteT = 0.0
DNASiteDis1 = 0.8	Energy1 = 5.0	OnSiteC = 0.0
dDNASiteDis = -0.2	dEnergy = 0.07	OnSiteG = 0.0
DNASitePPP = 0	HopUP = 2.0	OnSiteBB = 0.0
Epsilon = 0.5E-5	DNAHopperp = 0.1	

Table 3.2: An example of the input file that contains all the fixed parameters that are fed into the programme, for the ladder model. The input file for the fishbone model is a little simpler.

poly(dC) DNA for the fishbone model. Our semi-empirical calculations on DNA base pairs using the SPARTAN package have shown that the wavefunction overlap between the individual bases of a base pair is weak and therefore we take it to be $t_{1,2} = t_{i,\tau}/10$. We make the further assumption that the wavefunction overlap between consecutive bases along the DNA strand is weaker between unlike and non-matching bases, for which we thus choose $t_{i,\tau} = t_i^q/4$, whereas $t_{i,\tau} = t_i^q/2$ is chosen between identical and matching bases (i.e. AT/TA, GC/CG, AA, TT, GG, CC). Moreover, in the same paper [3], Cuniberti *et al.* obtain the numerical result that the hopping between Guanine bases is 0.37 eV. Our calculations have been made relative to one another using $t = 0.37$ eV as the unit for all the (energy) quantities.

For example when we say the energy is 5 - it is assumed it is $5t$, i.e. 5×0.37 eV. Therefore the fixed parameters, which are the hopping amplitudes

between consecutive bases are given the value 1.0 for same and matching bases and 0.5 for different and non-matching bases. The hopping terms to the upper and lower backbone have the constant value 2.0.

In vivo and most experimental situations, DNA is exposed to diverse environments. The solution, the thermal effects and the available space (causing the DNA to bend) are factors that alter the structure and properties that one is measuring [30,34]. Here, we model this by introducing various types of disorder into the hopping and onsite-energy parameters. In our calculations we only vary the onsite disorder at the backbone sites of either model. This onsite disorder is basically a potential that varies between the values W and $-W$ according to a distribution that can be uniform, Gaussian or binary. The binary onsite disorder is chosen in order to model the random adhesion of saline solvents. It is an onsite potential on the backbone that can only take the values W_{binary} and $-W_{binary}$ at each step with a 50% probability each, accounting for random partial absorption of positive solutions ions by the negatively charged sugar-phosphate backbone. Uniform disorder is also considered in our calculations.

3.3 The Computation Process

We have carried out simulations calculating the localisation lengths as a function of the system's energy and different types of backbone disorder for four different DNA sequences; poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA.

For each model, each of the DNA sequences is fed into the programme, as an input file that contains the bases of one strand of the sequence, ATTGC-CTAA...etc, for instance. In the fishbone model this is sufficient because each site represents a whole base pair. In the ladder model, however, this needs to be accounted for and thus the code has been written in a way so that the complementary strand of the sequence-input file is generated every time so as to assign

each base to a distinct site. Then, each distinct site is linked with its consecutive one and with the upper/lower backbone sites (and the lower/upper branch in the case of the ladder model) through the hopping amplitudes of the parameter-input file of Table 3.2.

In the calculations where the system's energy is varied, the backbone onsite disorder is fixed at zero value. Actually, there is no disorder in the system of any type and all the disorder parameters in Table 3.2 are set to zero. Thus, the energy calculations are done for an ordered system where only the hopping terms that connect the individual sites govern the transport. The energy range is $[-5, 5]$ (in units of t) in steps of $0.07t$. At each step, one localisation length value, for the whole sequence, is computed, depending on the hopping amplitudes that connect the sites of the model. So at the end of one run of calculations we have $10/0.07$ rows of data; a localisation length with its error, at each energy step.

When the backbone disorder is varied the system's energy has a fixed value. Although we have performed calculations at many different constant energies, the most representative ones were chosen at $E = 0$ and at $E = 3$ because of the conduction bands that are observed and will be shown in Chapter 4. As explained previously, the type of disorder can vary as well. We have used uniform and binary backbone onsite disorder. The strength of the disorder is in units of t and starts from zero up to a maximum of 10 in steps of 0.07. Similar to the energy calculations, there are $10/0.07$ values of localisation lengths for each step in disorder.

Every time a calculation of a localisation length is done, an error is also calculated. In the current project it has been set as the absolute difference between the calculated localisation lengths that is less than 0.5×10^{-5} . (It is the term called "epsilon" in the parameter-input file Table 3.2).

The parameter-input file of Table 3.2 will be explained here. The terms in bold are the most relevant ones in this report; the rest deal with computational

methods or introduce additional variables and parameters that we have not used here. The relevant terms will be briefly explained. “BBSiteDis” stands for backbone site disorder; the interval of which is determined by “BBSiteDis0” as the starting value and “BBSiteDis1” as the maximum value in steps of “dBBSiteDis”; the type of disorder uniform(0)/Gaussian(1)/binary(2) is determined by the term “BBSitePPP”. The energy interval is defined through the terms “Energy0” and “Energy1” in steps of “dEnergy”. Which of the two quantities is varied, is determined through the parameter “lFluxFlag” that varies the backbone disorder when set at the value of 2 and the energy when set at the value of 4. When one of these two quantities is varied, the other one is fixed; then the first value of the interval is assumed i.e. “BBSiteDis0” when energy is varied and “Energy0” when the backbone disorder is varied. “HopUP” and “HopDN” give the hopping amplitudes to the upper and lower backbone. “DNAHopperp” is the hopping term $t_{1,2}$ that connects the sites between the two branches in the ladder model. The rest of the terms “HopTA”, “HopGC”, “HopCC”, etc. are basically the hopping amplitudes between all the combinations of bases. The programme recognizes that “HopTA” is the same as “HopAT” and so on, which is embedded in the code, so needs not be determined in the parameter-input file. To sum up, we have performed calculations varying the energy with no added disorder, varying the binary backbone onsite disorder at fixed energies $E = 0$ and $E = 3$, and uniform backbone onsite disorder, at $E = 0$, for four sequences of DNA, poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA. The results obtained are shown in the next chapter.

3.4 The Sliding Window Technique of the Ladder Model

The sliding window technique is a numerically different approach to the calculation of the electronic properties of DNA applied to the ladder model. In the initial cal-

culations, as described in the last paragraph, the localisation lengths are computed at each energy/backbone disorder step as an average over the whole length of the sequence. The interpretation is that the larger the localisation lengths the more likely it is for the particular sequence to act as a medium for charge particle migration. However, there is a possibility that less global effects are present within a sequence and overshadowed by the averaging process. Local variations that make the localisation lengths change rapidly over only a few base pairs, would not be detected during the initial computations using the ladder and fishbone models.

The programme for the ladder model has been modified in order to include more local phenomena. A sliding window frame is designed to shift through each sequence computing the localisation length over a given number of bases, fragments of the full array. A brief example might illustrate this better. Suppose we have a DNA sequence that consists of $m = 20$ bases and the window size is set at $l = 10$ bases and that the bases are treated as nothing more than elements of a number sequence. The code is written so that the programme always begins by looking at the first interval that starts with the first base of the sequence. Hence the programme is initially considering the interval $[1, 10]$ and calculates one value for the localisation length. Then the frame *shifts* to exactly the next base, i.e. the *2nd* one and computes a localisation length for the interval $[2, 11]$, since the window frame remains constant. In what follows one value for the localisation length is calculated for each interval, $[3, 12]$, $[4, 13]$ and so on until the interval $[11, 20]$ is considered. The window size remains constant throughout one round of calculations that corresponds to shifting once through each whole sequence. If the total number of bases in a sequence is m and the window length is l , then the last localisation length is computed for the frame whose starting position is at the $m - (l - 1)$ base.

Finally, the question on how long this frame ought to be in order to reveal local effects, should be put forward. We have tried to capture a wide range of

window sizes and therefore we have performed calculations on two sequences λ -DNA and scrambled λ -DNA for window lengths of 10, 30, 50, 100, 200, 500 and 1,000 bases. This way we expect to be in a position that allows us to observe and compare the local electronic behaviour of biological and scrambled DNA. All the calculations for the window technique described here, are done at a fixed energy $E = 3$.

The actual data that are obtained show the window starting position and the corresponding localisation length with its error; since the window size is fixed each time, we basically have one localisation value assigned to each interval. In order to evaluate this set of data, we have studied the correlations between these localisation lengths for either sequence.

3.5 Correlation Functions Between the Localisation Lengths at Different Parts of the DNA Sequence

The correlation function is a quantity that describes the amount of correlation that is present between the elements of a sequence of numbers. It takes the maximum value of 1 for a fully correlated sequence and vanishes at zero when there is no correlation at all. The concept of the correlation length follows naturally and is nothing but the gradient of the plot of the natural logarithm of the correlation function against the position of the sequence and expresses the distance over which the elements are correlated with one another, on average.

In the current project we have used these concepts to look at the correlations between the values of the localisation lengths each of which correspond to a specific window frame. Although we have a number overlapping window intervals as opposed to a number of discrete elements, the important quantity here is the

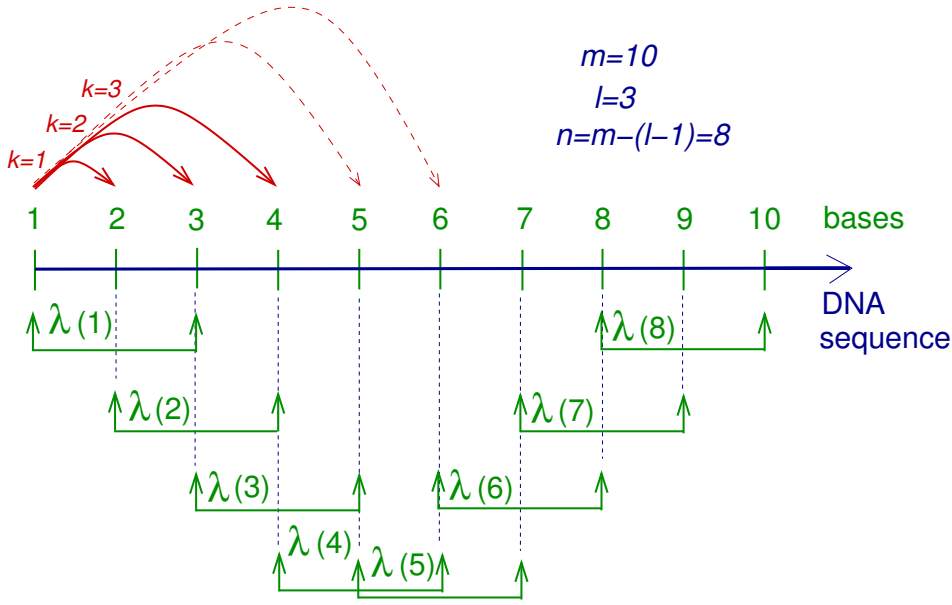


Figure 3.1: A schematic view of the application of how the correlation function between the localisation lengths is calculated. The sequence length is $m = 10$, the window size is $l = 3$ and thus localisation lengths can be computed for $n = m - (l - 1)$ intervals. k is the distance from each localisation length to all the others in the sequence, measured from the starting base of the window they correspond to.

localisation length. The idea is that we have a sequence of quantities, in this case the localisation lengths λ , whose positions in the sequence are defined through i and k , where i represents the site and k the distance from that site. In other words we want to determine the relationship between the quantity at site i and the same quantity a distance k from the site i ; i.e. look at $\lambda(1)$ and $\lambda(1 + 2)$, for $i = 1$ and $k = 2$.

For a given window size l , that shifts through a sequence of m bases, where each time a localisation length has been computed, the correlation function can be calculated. Firstly the mean localisation length $\bar{\lambda}$ needs to be defined

$$\bar{\lambda} = \frac{\sum_{i=1}^n \lambda_i}{n} \quad (3.1)$$

where i is the current sequence position, which is the same as the window starting

position and $n = m - (l - 1)$, is the total number of bases minus the number of intervals implied by the window size, accounting for the end of the sequence. The correlation function $s(i, k)$ is then given by

$$s(i, k) = \frac{\sum_{k=1}^{n-1} \sum_{i=1}^{n-k} (\lambda_i - \bar{\lambda})(\lambda_{i+k} - \bar{\lambda})}{\sum_{i=1}^n (\lambda_i - \bar{\lambda})^2} . \quad (3.2)$$

Basically each localisation length is defined through the starting position of the window interval it is associated with, but since every time the window size remains fixed the whole interval will not be referred to explicitly nor used again after the correct n has been calculated. So for a sequence of $m = 10$ bases where the window size $l = 3$, we get $m - (l - 1) = 8$ intervals over which the localisation lengths have been computed. The first localisation length $\lambda(1)$ starts at base 1 and extends to base 3; the last one $\lambda(8)$ starts at base 8 and extends to base 10. Then the correlation between each element with all the other elements of the sequence is determined through the use of the distance k that starts from 1 and gets the maximum value $n - 1$. In the above example, $\lambda(1)$ can be correlated with $\lambda(2)$ when $k = 1$, $\lambda(3)$ when $k = 2, \dots$, $\lambda(8)$ when $k = 7$. But $\lambda(8)$ is the last value for the localisation length. So $k_{max} = n - 1$.

3.6 Programmes and Computational Techniques

Fortran 90 has been used for the code of all models and the calculation of localisation lengths. The fishbone and ladder models and the modification of the ladder model for the sliding window technique were ran on the Task Farm in the Centre for Scientific Computing, University of Warwick. The modification from the code of the ladder to include the sliding window technique was done by John Grime. The non-biological sequences were generated using Mathematica, the code of which is shown in Appendix A. For scrambled DNA, Mathematica was used as well, where the command `RandomPermutation` was applied to the existent λ -DNA sequence. If we call the original λ -DNA sequence `DNaseq` then the command was

`newDNaseq = RandomPermutation[DNaseq]`. The correlation functions were calculated using Mathematica, the code is shown in Appendix B. All the Figures that are not referenced were produced on Xfig and all the graphs on xmgrace.

3.7 The SPARTAN Package for Semi-empirical Molecular Orbital Calculations

We have used the SPARTAN package to calculate the Highest Occupied Molecular Orbitals HOMOs and the Lowest Unoccupied Molecular Orbitals LUMOs of certain DNA fragments, by means of semi-empirical calculations. `Spartan5.1` is a molecular modelling programme. It performs molecular mechanics and quantum chemical calculations; in particular semi-empirical molecular orbital, *ab initio* and density functional calculations. For the purposes of this project, the nucleotide bases along with the phosphorylated sugars are built into the package. Due to time limitation we have not made calculations for more than 2 base pairs (i.e. 4 bases altogether).

We have performed semi-empirical calculations using the SPARTAN package, in order to get an estimate for the hopping amplitude between the two conduction branches in the ladder model, t_{12} . The actual results are given in Chapter 4 where they are also discussed. This part of the project has been developed with the help of P. Taylor in the Chemistry Department of the University of Warwick.

Chapter 4

Results

4.1 Results for the Fishbone model

In order to be able to assess the resulting graphs, one needs to keep in mind the nature of the quantities that are being computed. Firstly, the localisation lengths show how fast in space the wavefunction of the electron wears off; how extended the electron is in space in the particular system. Large localisation lengths mean that the electron could be anywhere in that region because its wavefunction is extended there and thus there is enhanced transfer probability across that length. The opposite applies for small localisation lengths. In our model the localisation length is measured in units of the base-to-base distance, $3.4 \text{ \AA}$ (along the long axis). Initially, all small onsite potential fluctuations due to differences in base-ionization energies are ignored and all energy measurements are in units of $t = 0.37\text{eV}$.

4.1.1 Varying the energy

Initially, the fishbone model has been considered and there is no additional disorder to the system; all the onsite potentials are set to zero and the hopping amplitudes are the only terms that control the transport. In Figure 4.1 the localisation lengths

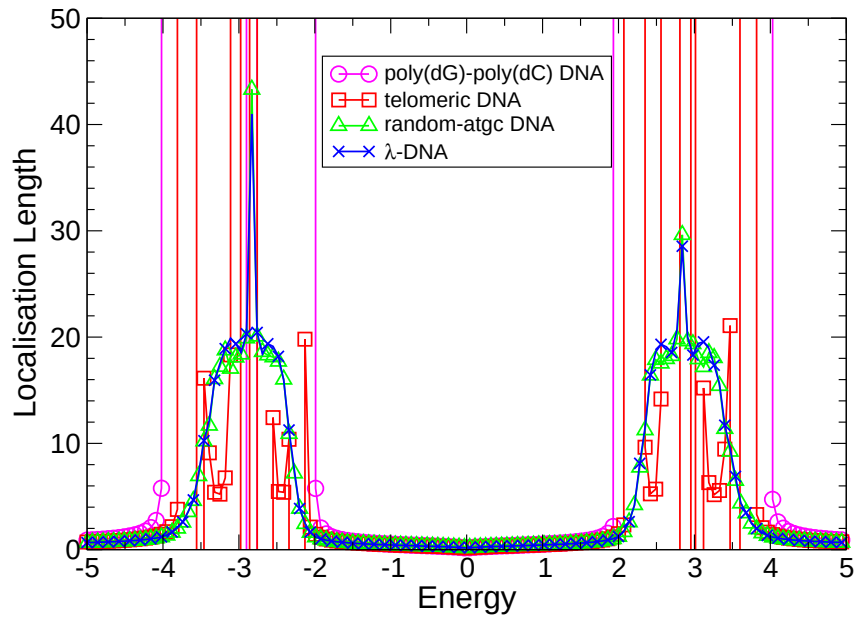


Figure 4.1: Dependence of the localisation lengths upon the system's energy, with no additional disorder, using the fishbone model of electronic transport along DNA, for four different sequences: poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA. The first two sequences seem to have extended states within the minibands whereas the latter two are rather localised over a few tens of bases.

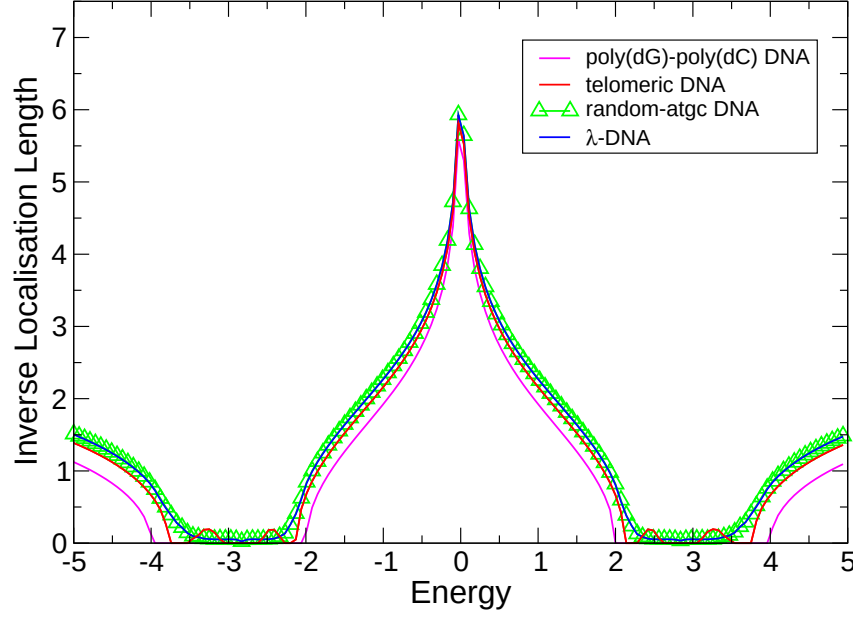


Figure 4.2: Plot of the inverse localisation lengths versus the Fermi energy, for four DNA sequences as in Figure 4.1.

are shown as a function of the Fermi energy for the four sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA. On this graph we can see there is symmetry about the $E = 0$ line and two minibands are formed on either side roughly between $(-4, -2)$ and $(2, 4)$. Within the energy spectrum $(-5, 5)$, the localisation lengths are finite only over the minibands and vanish everywhere else. More specifically, poly(dG)-poly(dC) and telomeric DNA have infinite localisation lengths whereas random-ATGC and λ -DNA seem to have rather localised states. Their localisation lengths do not exceed 20 bases apart from two spikes at the centre of each band ~ 30 -40 bases. These peaks resemble the Dyson singularities [45, 46] that are intrinsic to random hopping models [47]. The Dyson singularities are in principal symmetric but do not appear so in the graph because the energy steps are not symmetric $dE = 0.07$ about $E = 0$. In Figure 4.2, the inverse localisation lengths are plotted against energy, for the same system. For poly(dG)-poly(dC) the minibands are very well defined by the inverse localisation lengths

which remain zero exactly within the intervals $(-4, -2)$ and $(2, 4)$. As was already expected from Figure 4.1 telomeric DNA's behaviour is very similar to poly(dG)-poly(dC)'s. However, two prominent differences can be observed: the first is that the bands are slightly smaller covering the intervals $(-3.7, -2.2)$ and $(2.2, 3.7)$ and the second is that there are two tiny and smooth peaks within each band, reaching values a little less than 0.5, where we would have imagined everything to be zero. These states also display symmetry about the $E = 0$ line.

4.1.2 Varying the Backbone Disorder

Continuing our research with the fishbone model, we have computed the localisation lengths as a function of different types of disorder. The first system we looked at, is at fixed energy $E = 0$ with binary disorder, W_{binary} applied at the backbone sites. Although localisation theory tells us that disorder leads to a localisation of states, see Section 2, it is clear from Figure 4.3 that this does not seem to be the case here; the localisation lengths *increase* with increasing disorder. Random-ATGC and λ -DNA are almost identical throughout with a nearly linear behaviour up to about $W_{binary} = 5$. Telomeric DNA follows them closely up to $W_{binary} = 5$ as well but then deviates with a larger slope reaching localisation lengths of about 12 bases at $W_{binary} = 10$ as opposed to 8 bases of random-ATGC and λ -DNA. Poly(dG)-poly(dC) is the sequence whose localisation lengths increase the fastest reaching localisation lengths of 27 bases long at $W_{binary} = 10$. The same calculations have been made at fixed energy $E = 3$, i.e. within the minibands, and as illustrated in the plot of Figure 4.4, the overall behaviour has thoroughly changed. Here we have plotted the logarithmic localisation lengths versus disorder and they are clearly decreasing as the disorder gets stronger. At $W_{binary} = 0$ poly(dG)-poly(dC) and telomeric DNA have infinite localisation lengths whereas random-ATGC and λ -DNA have $\ln(\lambda) = 3$. For all four sequences as the disorder increases, the localisation lengths drop fast, hitting zero at $W_{binary} = 2$ and re-

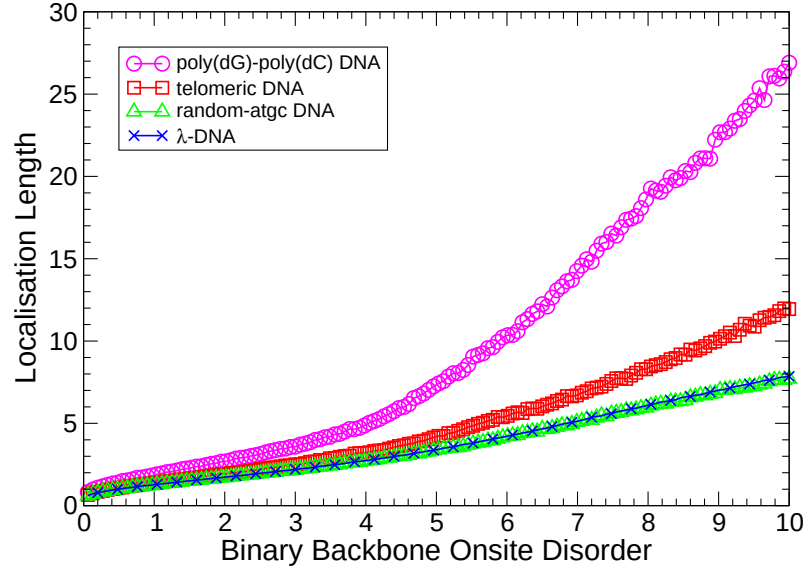


Figure 4.3: The localisation lengths as a function of binary onsite disorder, at the backbone sites, at $E = 0$, using the fishbone model for the DNA sequences: poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA.

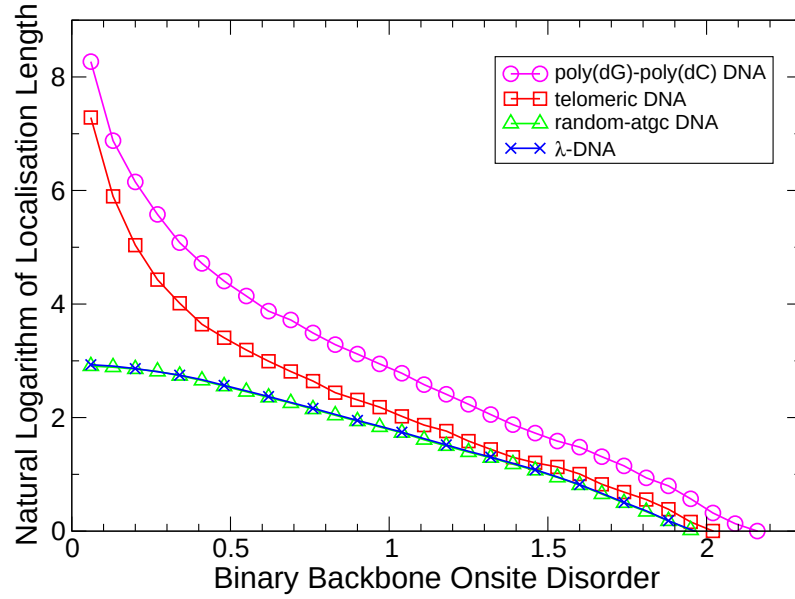


Figure 4.4: The natural logarithm of the localisation lengths of the DNA sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, plotted as a function of the binary onsite disorder at the backbone sites, at $E = 3$, using the fishbone model.

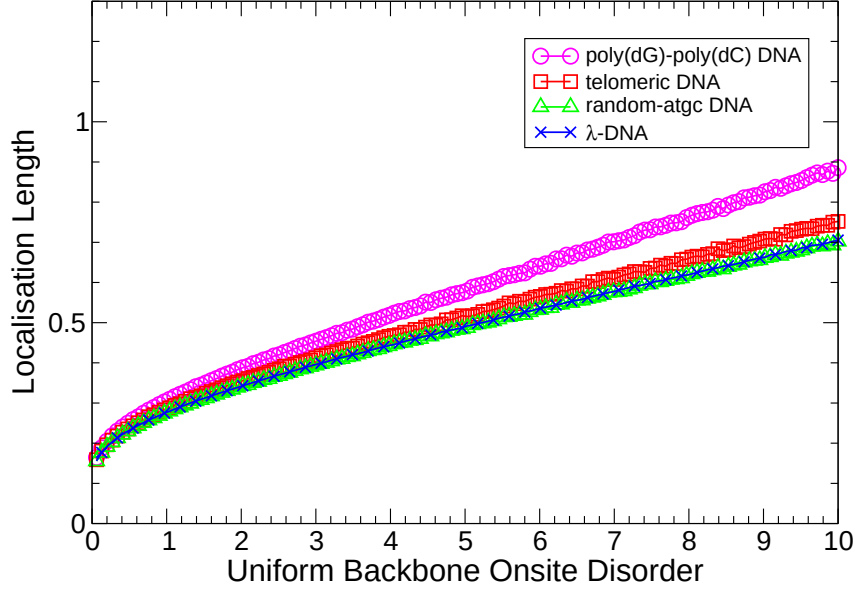


Figure 4.5: The localisation lengths of the DNA sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, plotted as a function of the uniform onsite disorder at the backbone sites, at $E = 0$, using the fishbone model.

maintaining there for $W_{binary} > 2$. Finally, the distribution of the disorder has been changed from binary to uniform and the localisation lengths as a function of the disorder, at $E = 0$, have been computed; the plot is shown in Figure 4.5. At first glance we see the same effect as in Figure 4.3, where the disorder seems to induce partial delocalisation of states. Indeed the localisation lengths increase as the disorder increases but the scale is much smaller. In fact, it is clear that the localisation lengths do not even reach one whole base of length. We believe that this counterintuitive effect is closely connected with the application of *binary* onsite disorder only and will be discussed in Chapter 5.

4.2 Results for the Ladder model

The two models yield similar results, with a slight enhancement of the localisation lengths in the ladder model. This is not a surprising outcome considering the fact

that in the ladder model the particle can choose to travel in either of the two channels, both enhancing the transport in the forward direction.

4.2.1 Varying the energy

In Figure 4.6 we show the behaviour of the localisation lengths as a function of the Fermi energy, with no additional disorder, computed for the four DNA sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA. The two minibands where localisation lengths are finite, are slightly shifted compared to the fishbone model and are shown more clearly in Figure 4.7, where the inverse localisations lengths are plotted for the same system. Poly(dG)-poly(dC) and telomeric DNA have infinite localisation lengths within the minibands which roughly extend over the intervals $(-3.3, -1.2)$ and $(1.2, 3.3)$. Random-ATGC and λ -DNA are again more localised, with localisation lengths of the order of 20 bases which reach a maximum of ~ 30 at the Dyson singularities. The minibands observed for those are identical and their range is $(-1.5, 2.7)$ and $(1.5, 2.7)$. Finally, the same non-zero states within the bands, where the localisation length has two tiny smooth peaks, that were observed for telomeric DNA in the fishbone model, are also observed here.

4.2.2 Varying the backbone disorder

When we apply binary onsite disorder at the backbone sites to the ladder model, we observe the exact same peculiar effect where the localisation lengths increase as the disorder gets stronger, as shown in Figure 4.8. Random-ATGC and λ -DNA display a slow, almost linear increase whereas poly(dG)-poly(dC) and telomeric DNA increase rapidly. For telomeric DNA the behaviour is very similar to that of random-ATGC and λ -DNA up to $W_{binary} \sim 5$ above which it deviates from them with a much larger slope. Poly(dG)-poly(dC) demonstrates the most abrupt

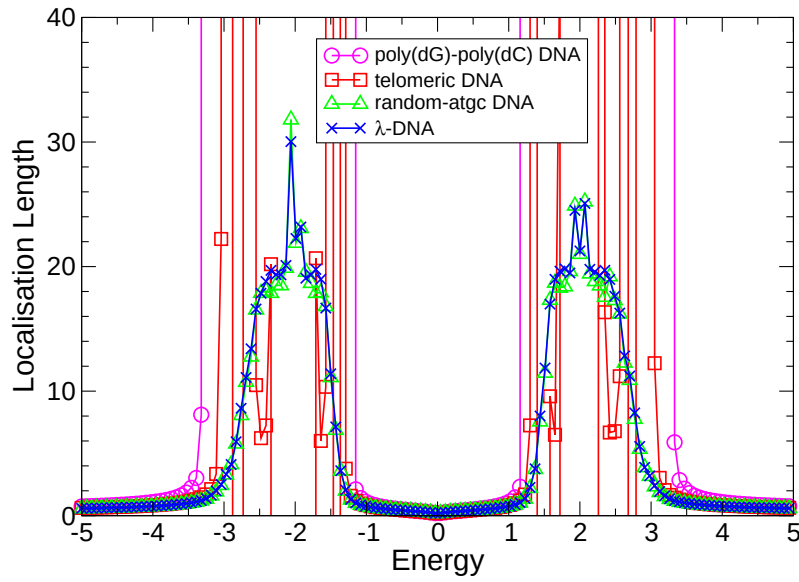


Figure 4.6: Dependence of the localisation lengths upon the energy, with no additional disorder, using the ladder model of electronic transport along DNA, for four different sequences: poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA.

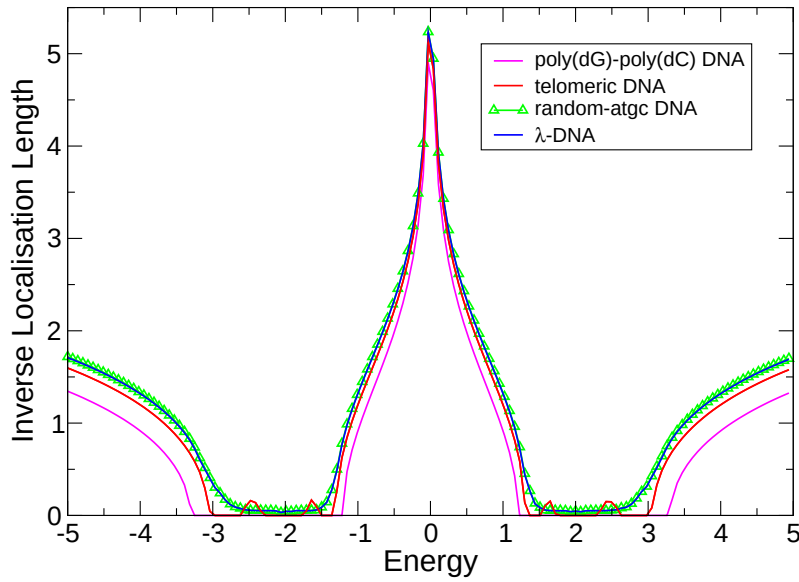


Figure 4.7: The inverse localisation lengths versus energy with no additional disorder, is plotted here, for poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, using the ladder model.

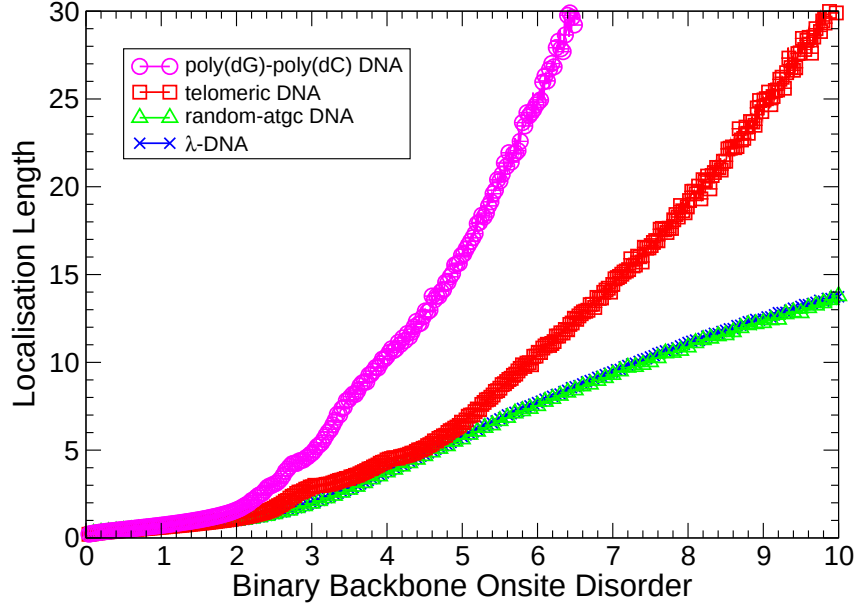


Figure 4.8: The localisation lengths of the DNA sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, plotted as a function of the binary onsite disorder at the backbone sites, at $E = 0$, using the ladder model.

increase and has already diverged from the other three at $W_{binary} = 2$. At the maximum disorder $W_{binary} = 10$ of our spectrum, poly(dG)-poly(dC) has reached localisation lengths as high as 80 bases long (see Appendix c); telomeric DNA has reached 30 bases and random-ATGC and λ -DNA have only got localisation lengths of the order of ~ 14 bases. When the same binary disorder is applied at a fixed energy $E = 3$ within the miniband the localisation lengths decay as the disorder gets stronger, as shown in Figure 4.9. At $W_{binary} = 0$ poly(dG)-poly(dC) and telomeric DNA have infinite localisation lengths that drop rapidly hitting zero at $W_{binary} = 3$. Random-ATGC and λ -DNA have localisation lengths of the order of only 3 bases at $W_{binary} = 0$ and as the disorder increases, the localisation lengths drop and finally vanish when $W_{binary} = 3$. For $W_{binary} > 3$ all sequences have the same behaviour; localised around one base. We have performed the same calculations for a uniform distribution of the disorder at a fixed energy $E = 0$. Similar to the fishbone model, also here we can see the effect where the localisation

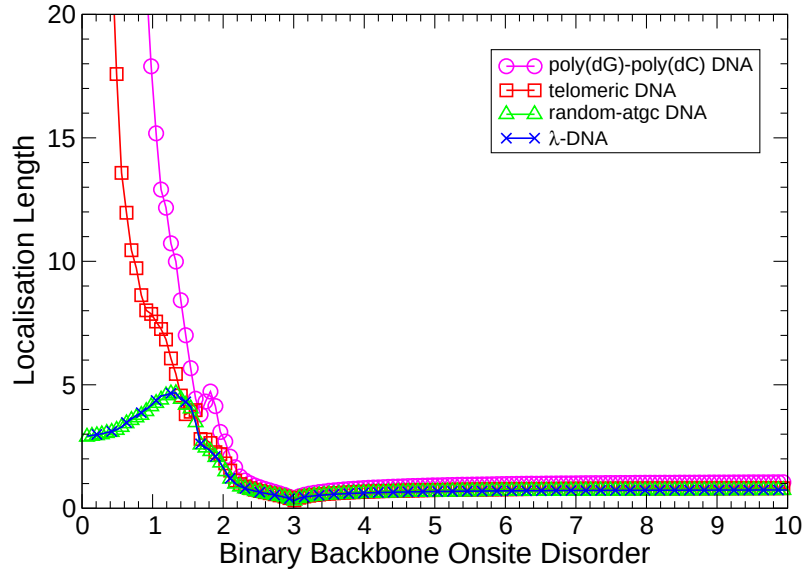


Figure 4.9: The localisation lengths of the DNA sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, plotted as a function of the binary onsite disorder at the backbone sites, at $E=3\text{eV}$, using the ladder model.

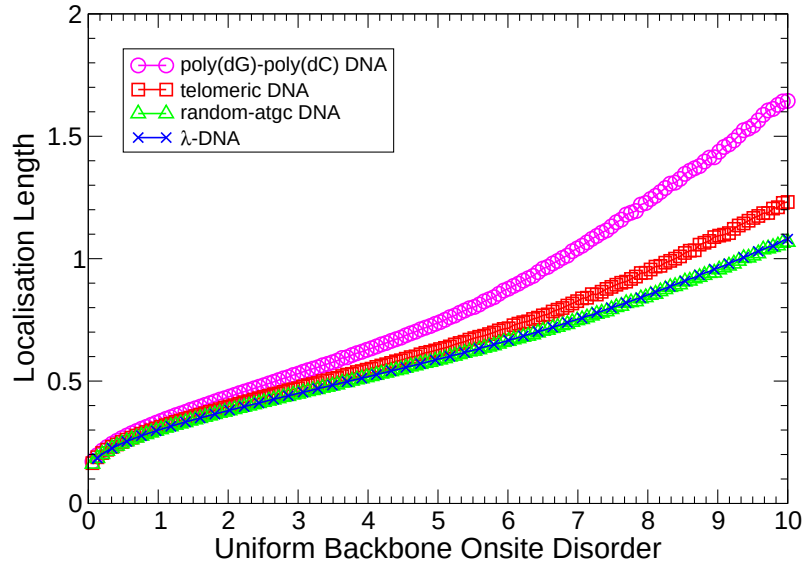


Figure 4.10: The localisation lengths of the DNA sequences poly(dG)-poly(dC), telomeric, random-ATGC and λ -DNA, plotted as a function of uniform onsite disorder at the backbone sites, at $E=0\text{eV}$, using the ladder model.

lengths get larger as the disorder gets stronger. Again, this increase is so slow and smooth that the effect becomes insignificant altogether. Poly(dG)-poly(dC) demonstrates the quickest increase having localisation length ~ 1.5 at disorder $W_{uniform} = 10$. Telomeric DNA increases only slightly faster than random-ATGC and λ -DNA reaching localisation lengths of 1.2 bases at $W_{uniform} = 10$; random-ATGC and λ -DNA manage to reach localisation lengths of the order of only 1 base.

4.3 Sliding Window Technique Results: Correlation Functions Between Localisation Lengths

In Section we introduced the basic concepts of correlation between the elements of a sequence and furthermore linked those to the current study by considering the correlation functions between the localisation lengths of each window frame. Here, we present the results of the sliding window technique analysed in terms of correlations. We have calculated how the correlation function between the localisation lengths of each frame of the sequence changes as we browse through the sequence for various window sizes. (The x -axis called the sequence position coincides with the window frame starting position, which is the reference position for the localisation lengths as well).

In Figures 4.11 and 4.12 the correlation function versus the sequence position for window lengths of 10, 30, 100, 200, 500 and 1000 bases for λ -DNA and scrambled λ -DNA respectively, have been plotted. The insets include the same data as the large graphs apart from the x -axis which now represents the sequence position *divided* by the window length corresponding to each set of data.

As we can see the two plots are quite similar in behaviour. In the large graphs, we see fluctuations about the $y = 0$ line, which increase with the window size. In fact, as the window frame gets larger, the line seems to fold around the

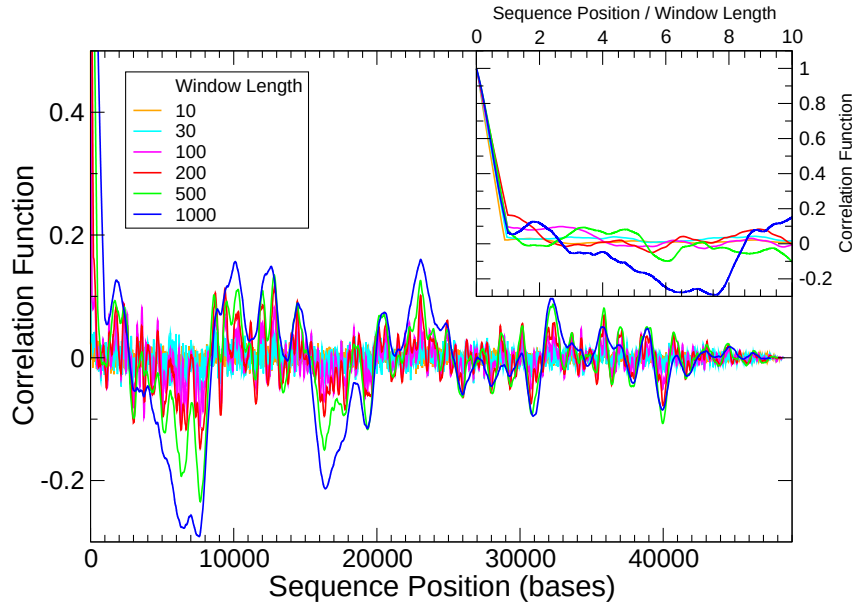


Figure 4.11: The correlation functions plotted as a function of the DNA sequence position, i.e. the number of the base, using the window technique of the ladder model for all window lengths 10, 30, 100, 200, 500, 1000 bases, for λ -DNA. The inset shows the same correlation function versus the sequence position *divided by* the window length for each window size.

plots of the smaller window lengths. If we look at the small window lengths for example up to 100 bases, it seems almost impossible to distinguish any peaks or any consistent behaviour besides small probably random fluctuations about $y = 0$. But as the window length increases the behaviour becomes more interesting in the sense that the fluctuations are strong and even negative for large window frames. In the insets we see the correlation drops from its maximum value of 1 to hit zero correlation at the point where the first window frame has been surpassed, i.e. when $\text{sequence position} / \text{window length} = 1$. For scrambled λ -DNA the inset shows two peaks for the two largest window sizes 500 and 1000 bases which are not present in λ -DNA.

In the following graphs Figures 4.13-4.16, we have plotted the correlation function versus the sequence position at each window frame 30, 100, 500, 1000

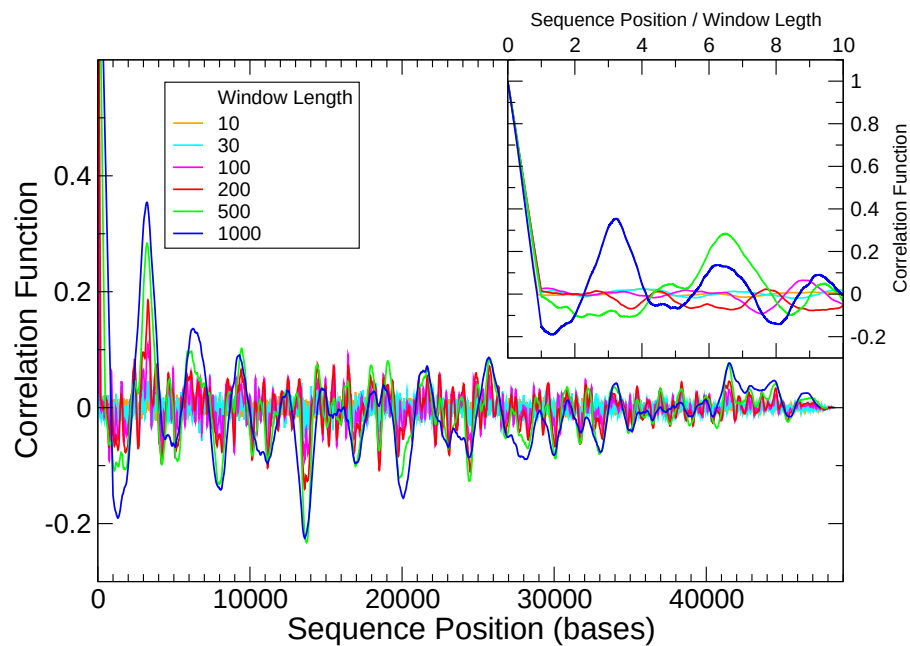


Figure 4.12: The correlation functions plotted as a function of the DNA sequence position using the window technique of the ladder model for all window lengths 10, 30, 100, 200, 500, 1000 bases, for scrambled λ -DNA. The inset shows the same correlation function versus the sequence position *divided by* the window length for each window size.

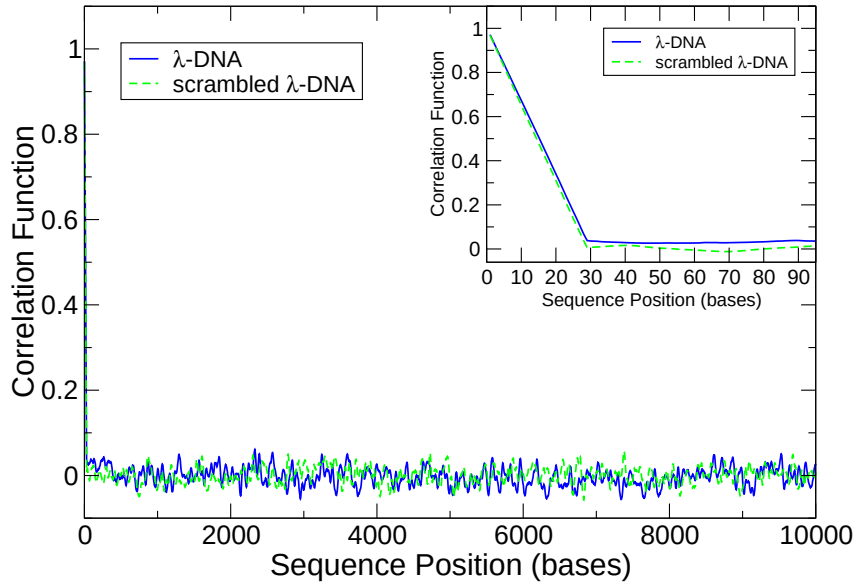


Figure 4.13: Comparing the correlation functions of biological λ -DNA and scrambled λ -DNA as a function of the sequence position, using the sliding window technique, for a window size of 30 bases. The inset shows a zoom into the first 90 bases and what we observe is that the correlations disappear at the point where the window frame ends, i.e. at the 30th base.

separately, for either sequence. For the smallest window frame of 30 bases, shown in Figure 4.13, we see no particular behaviour or any significant differences between the two sequences. The correlation function drops from its maximum value of 1 into small fluctuations around $y = 0$. In the inset, which is a zoom of the first 90 bases of the larger graph, it is clear that the correlations disappear as soon as the sequence position reaches 30 bases and slightly fluctuates for the remaining part of the sequence.

The correlations of the localisation lengths for λ -DNA and scrambled λ -DNA for a window length of 100 bases is shown in Figure 4.14. The fluctuations here seem slightly larger and the negative values for correlation become more apparent. The two sequences do not seem to have any significant differences. In the inset, the correlations disappear at the 100th base of the sequence.

In Figure 4.15 the correlation function versus the sequence position is shown

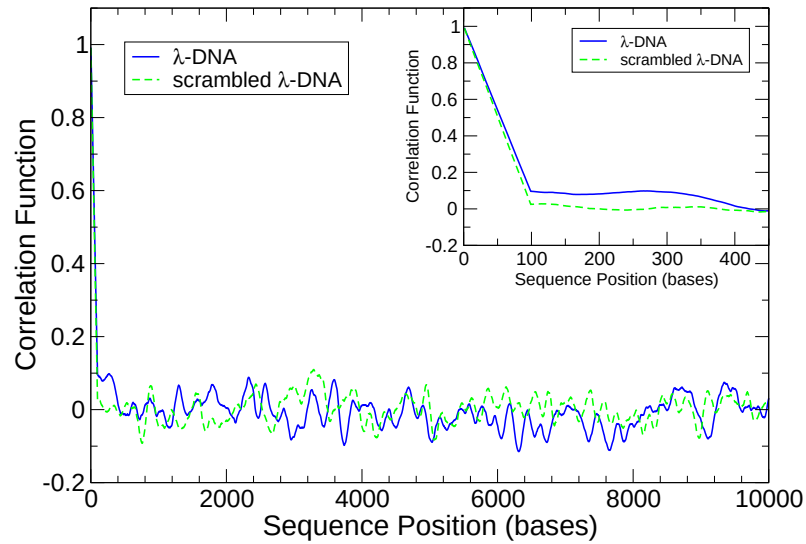


Figure 4.14: Comparing the correlation functions of biological λ -DNA and scrambled λ -DNA as a function of the sequence position, using the sliding window technique, for a window size of 100 bases. The inset shows a zoom into the first 400 bases and what we observe is that the correlations disappear at the point where the window frame ends, i.e. at the 100th base.

for a window length of 500 bases. Here the fluctuations have become significantly larger and the two sequences seem to start diverging from one another in their behaviour. The largest fluctuations occur in the first half of the sequence. In the inset we can see again how the correlations disappear at the point that corresponds to the end of the window i.e. at 500 bases. Finally, in Figure 4.16, we can see the plot of the correlation function as a function of the sequence position for a window size of 1000 bases. The correlations are stronger throughout the whole range and the two sequences clearly demonstrate different behaviour. The correlations hit a minimum of -0.3 and a maximum of 0.2 for λ and -0.2 and 0.3 respectively for scrambled λ -DNA. In the inset the correlations disappear at the end of the first window frame, i.e. at the 1000th base.

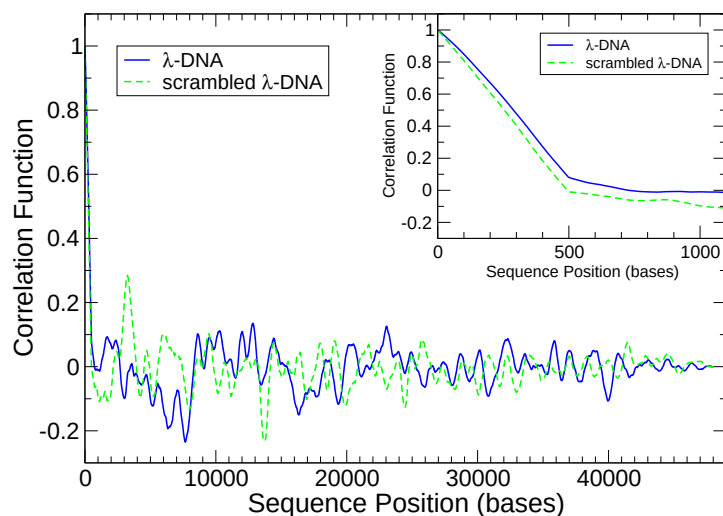


Figure 4.15: Comparing the correlation functions of biological λ -DNA and scrambled λ -DNA as a function of the sequence position, using the sliding window technique, for a window size of 500 bases. The inset shows a zoom into the first 1000 bases and what we observe is that the correlations disappear at the point where the window frame ends, i.e. at 500 bases.

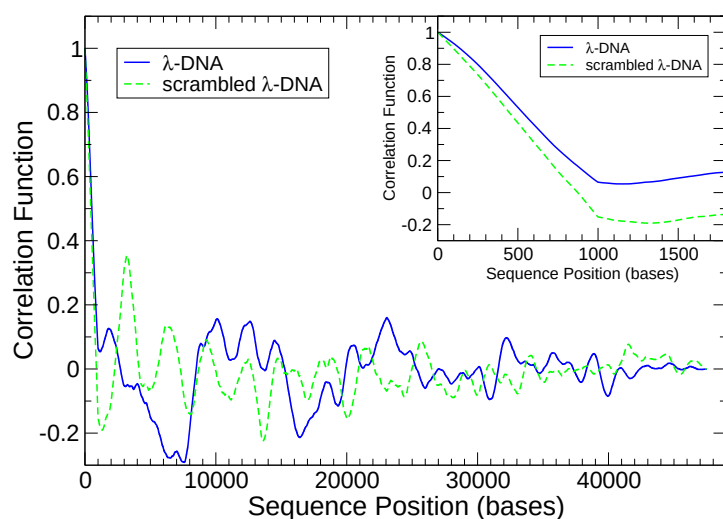


Figure 4.16: Comparing the correlation functions of biological λ -DNA and scrambled λ -DNA as a function of the sequence position, using the sliding window technique, for a window size of 1000 bases. The inset shows a zoom into the first 1500 bases and what we observe is that the correlations disappear at the point where the window frame ends, i.e. at 1000 bases.

Component (excluding backbone)	HOMO (eV)	LUMO (eV)	Gap (eV)
Adenine	-0.932	5.865	6.798
Thymine	-1.447	5.849	7.296
Guanine	-1.167	5.836	7.003
Cytosine	-1.180	6.308	7.488
A-T pair	-1.552	2.549	4.101
G-C pair	-1.805	3.198	5.003
AA/TT sequence	-1.734	2.229	3.963
CC/GG sequence	-2.099	2.375	4.474
AT/TA sequence	-0.692	1.189	1.881
GC/CG sequence	-0.780	1.239	2.020
AC/CA/TG/GT sequence	-1.235	1.786	3.021
AG/GA/TC/CT sequence	-1.874	2.451	4.326

Table 4.1: The HOMO's (Highest Occupied Molecular Orbitals), the LUMO's (Lowest Unoccupied Molecular Orbitals) and the gap between them are shown for different components of the DNA molecule, excluding the backbone, using the semi-empirical molecular orbital calculations that the molecular modelling programme SPARTAN provides [1].

4.4 Semi-empirical Molecular Orbital Results

We have performed semi-empirical calculations, using the SPARTAN package, that give us the HOMO and the LUMO energies of single bases, base pairs and sequences consisting of two bases with their complementary ones, including and excluding the sugar-phosphate backbone. The first column shows the component under consideration. The sequences are written so that only the base of one of the two strands is stated because there is only one complementary for it. So for example what is referred to as a GT sequence in Tables 4.1 and 4.2 is a molecule that consists of G and T consecutive bases that are paired with a C and an A base respectively. As a result the sequence GT and the sequence CA are in fact identical. All energies are given in 3 decimal places and in units of eV. In Table 4.2 exactly the same calculations have been performed but with the backbone included. The addition of the sugar-phosphate backbone makes the calculation a lot slower and complicated as some of the calculations actually failed for no apparent

reason. The time limit did not allow us to look further into this problem. Now

Component (including backbone)	HOMO (eV)	LUMO (eV)	Gap (eV)
Adenine	-5.511	1.589	7.101
Thymine	-5.529	1.529	7.058
Guanine	-5.393	1.851	7.244
Cytosine	-5.394	1.483	6.877
A-T pair	-5.370	-3.332	2.037
G-C pair	-6.285	-2.318	3.967
AA/TT sequence	-	-	-
CC/GG sequence	-6.959	-2.971	3.988
AC/CA/TG/GT sequence	-	-	-
AG/GA/TC/CT sequence	-	-	-

Table 4.2: The HOMO's (Highest Occupied Molecular Orbitals), the LUMO's (Lowest Unoccupied Molecular Orbitals) and the gap between them are shown for different components of the DNA molecule, including the backbone, using the semi-empirical molecular orbital calculations that were obtained using the molecular modelling programme SPARTAN [1].

we will discuss why the semi-empirical calculations were performed and how they were useful. We developed a theoretical approach to the system consisting of a base pair and we show how the HOMO's and the LUMO's can help us obtain an order of magnitude for the hopping term between the two central branches' sites, t_{12} , in the ladder model.

Suppose we have a system of two bases A and T which are not coupled, with energies γ_1 and γ_2 respectively. If a hopping term t is introduced these energies will be affected by the perturbation and the eigenvalue, Γ , will be the energy of the coupled system. The Hamiltonian is given by

$$H = \begin{pmatrix} \gamma_1 & t \\ t & \gamma_2 \end{pmatrix} \quad (4.1)$$

Using Schrödinger's equation $\hat{H}\psi = \Gamma\psi$ and diagonalising we get

$$(\gamma_1 - \Gamma)(\gamma_2 - \Gamma) - t^2 = 0 \quad , \quad (4.2)$$

and solving Equation 4.2 for Γ we get

$$\Gamma_{1,2} = \frac{\gamma_1 + \gamma_2}{2} \pm \frac{1}{2} [(\gamma_1 - \gamma_2)^2 + 4t^2]^{1/2} \quad (4.3)$$

which gives two possible values for t

$$t^2 = \Gamma_1^2 + \gamma_1\gamma_2 - \Gamma_1(\gamma_1 + \gamma_2) \quad , \quad (4.4)$$

$$t^2 = -\Gamma_2^2 + \gamma_1\gamma_2 - \Gamma_2(\gamma_1 + \gamma_2) \quad . \quad (4.5)$$

The second value of t in Equation 4.5 will be rejected for giving a negative t^2 , the calculation will be shown later where actual values are put in for all the other parameters are put in. In order to calculate t we need Γ , γ_1 and γ_2 . Γ is the energy of the coupled system A-T. In chemistry terms this is going to be the energy of the Highest Occupied Molecular Orbital (HOMO) of the A-T base pair. Similarly, γ_1 and γ_2 are the individual energies of the two bases A and T respectively, right before the coupling is introduced. So, γ_1 is the HOMO of base A and γ_2 is the HOMO of base T. Hence t can be obtained.

In using SPARTAN for the calculation of HOMO and LUMO energies, certain parameters need to be set before the programme runs one of which is whether an electron is added to the system or whether an electron excited from the system. In what follows we just show the modifications that would need to be made to the above calculation if an electron is added to the system.

Suppose we have a system of two bases A and T, which are not coupled, with energies γ_1 and γ_2 respectively, just as previously. If an electron is added into this system, it will either go to A or T and that gives two possibilities for the overall energy of the system: If the electron sits on site A, it will modify it and the energy will be $\gamma_1' + \gamma_2$. Similarly, if the electron initially chooses site T, the overall energy will be $\gamma_1 + \gamma_2'$.

Therefore in the case of adding an electron to the system, we get a new starting point from which the coupling is then introduced. Actually it gives us two

starting points depending on which site the electron occupies. So if the electron goes to A, the initial energies will be γ'_1 and γ_2 . This will give a final hopping term

$$t^2 = \Gamma_1^2 + \gamma'_1\gamma_2 - \Gamma(\gamma'_1 + \gamma_2) \quad . \quad (4.6)$$

If the electron sits on site T instead, t will be

$$t^2 = \Gamma_1^2 + \gamma_1\gamma'_2 - \Gamma(\gamma_1 + \gamma'_2) \quad . \quad (4.7)$$

Note that the case where there is no added electron remains the same, and the hopping term is given by Equation 4.4.

In the uncoupled system, γ_1 and γ_2 are the energies of the individual bases A and T respectively with no added electron and are given by γ_1 =energy of HOMO of A, with no added electron and γ_2 =energy of HOMO T, with no added electron. γ'_1 =energy of base A, with an added electron and is equal to the energy of the HOMO of A with an added electron. γ'_2 =energy of base T, with an added electron and is equal to the energy of the HOMO of T with an added electron.

In the case where the system is coupled, Γ =the energy of the coupled system A-T base pair, and is equal to the energy of the HOMO of A-T, with no added electron. t is the coupling term between the two bases A-T and is calculated from Equations 4.4 4.6 and 4.7 depending on the initial conditions; namely no added electron, added electron occupying one site of the base-pair and added electron occupying the other site, in the example A and T, respectively.

In the following calculations we are using Equation 4.4, since that corresponds to the parameter of "no added electron" that was set on SPARTAN, and the values of Tables 4.1 and 4.2 in order to calculate a value for the hopping term between the two branches of the ladder model t_{12} . First do the calculation for an A-T pair excluding the backbone, i.e. $\gamma_1 = -0.932$, $\gamma_2 = -1.447$ and $\Gamma = -1.552$.

$$t^2 = (-1.552)^2 + (-0.932)(-1.447) - (-1.552)(-0.932 - 1.447) \Rightarrow t = 0.26 \quad .$$

Do the same calculation for a G-C pair excluding the backbone, i.e. $\gamma_1 = -1.167$, $\gamma_2 = -1.180$ and $\Gamma = -1.805$.

$$t^2 = (-1.805)^2 + (-1.167)(-1.180) - (-1.805)(-1.167 - 1.180) \Rightarrow t = 0.63 \quad .$$

Now calculating t for base pair A-T including the backbone, i.e. $\gamma_1 = -5.511$, $\gamma_2 = -5.529$ and $\Gamma = -5.370$.

$$t^2 = (-5.370)^2 + (-5.511)(-5.529) - (-5.370)(-5.511 - 5.529) \Rightarrow t = 0.15 \quad .$$

For the base pair G-C including backbone, i.e. $\gamma_1 = -5.393$, $\gamma_2 = -5.394$ and $\Gamma = -6.285$.

$$t^2 = (-6.285)^2 + (-5.393)(-5.394) - (-6.285)(-5.393 - 5.394) \Rightarrow t = 0.89 \quad .$$

Although there results for the hopping term t are different and give an average of ~ 0.5 , our initial guess of 0.1 is in the right order of magnitude. If more calculations on the ladder model were made these values here could be used instead.

Finally, we show why Equation 4.5 is rejected. Use the values of the A-T pair without backbone just as an example although it holds for all the cases.

$$t^2 = -(-1.552)^2 + (-0.932)(-1.447) - (-1.552)(-0.932 - 1.447) = -4.75 \quad .$$

Chapter 5

Discussion

The two models, the fishbone and the ladder model, give very similar results for the localisation properties of DNA. We observe a behaviour with two minibands with a finite number of states when the localisation lengths versus Fermi energy are plotted. This is also in agreement with Ref. [3]. However we see that even in the ordered regime, random-ATGC and λ -DNA display localisation lengths of the order of $\sim 20 - 30$ bases only. Migration over $\sim 20 - 30$ bases out of the 10,000 and 48,502 bases that random-ATGC and λ -DNA, respectively, have in total, is not encouraging for any long-scale electron transport. Thus we find that these sequences are insulating and become even more so when additional disorder is introduced.

On the other hand, poly(dG)-poly(dC) and telomeric DNA have infinite localisation lengths in the ordered case. For the former, this is expected because of the hopping amplitudes that are $t_i = 1$ for same or matching bases and poly(dG)-poly(dC) only consists of G and C bases. The model was designed so that the conduction pathway is due to π -orbital overlap and this result only confirms what was assumed in the first place. The similar behaviour of telomeric DNA seems reasonable considering the fact that it has a sequence of GGG and TT consecutive bases in its periodic arrangement TTAGGG. We also see that there are two tiny

smooth peaks in the graph of Figure 4.7 on either side of the y -axis. These two extra minibands are observed within the two main conduction bands for telomeric DNA only and we believe are due to the periodic repetition similar to a Peierls instability [48] where extra gaps are induced. This interpretation is supported by the fact that such peaks have not been observed for any other sequence but only for the periodic telomeric DNA.¹

The two sharp peaks that are present at the centre of the bands for random-ATGC and λ -DNA, shown in the graphs of Figures 4.1 and 4.6, we believe, are Dyson singularities [45]. Dyson singularities are seen in theoretical studies of the Anderson model when there is only random hopping between the sites and no onsite disorder. The overall symmetry $\lambda(E) = \lambda(-E)$ is the result of the absence of a potential term in the Hamiltonian. The splitting into two regions, the two minibands, is in contrast to the usual 1-D Anderson model of localisation where only a single band around $E = 0$ is present and therefore can only be a consequence of the backbone. In fact this is confirmed if the backbone sites are removed completely the two symmetric minibands collapse into one in the centre of the graph. The presence of the backbone sites explicitly has a significant effect on the overall properties of our models.

As for the backbone disorder we find a very interesting result. In particular we see that when we apply binary disorder to the backbone sites of either model we induce partial delocalisation. This can be understood by looking at Equations 2.43 and 2.44 where we defined the effective disorder for each branch.

$$\tilde{\varepsilon}_i^{(1)} = \varepsilon_i^{(1)} + \frac{(t_i^\dagger)^2}{E - \varepsilon_i^\dagger}$$

We first note that by adding disorder at the backbone sites, states start to appear at $E = 0$. This has been checked explicitly by computing the density of states in

¹The negative localisation lengths that are observed in Figures 4.1, 4.6 have no physical meaning and are only a computational imprecision. The numbers are too big for the finite-precision arithmetic of the machine when the calculations have been performed and they are sometimes automatically taken as negative.

this situation by Greens function techniques [49]. There is no site potential on the central branches so $\varepsilon_i^{(1,2)} = 0$ and the hopping to the backbone sites is $t_i^\dagger = 2$. In the case of binary backbone onsite disorder $\varepsilon_i^\dagger$, can take the values of $\pm W_{binary}$. So at $E = 0$ the relationship is reduced to

$$\tilde{\varepsilon}_i^{(1)} = \frac{(t_i^\dagger)^2}{-\varepsilon_i^\dagger}, \quad (5.1)$$

and we see that increasing binary disorder W_{binary} decreases the effective 1-D onsite disorder $\tilde{\varepsilon}_i^{(1,2)}$. We emphasize this outcome of delocalisation through additional disorder because it is in complete contrast to 1-D Anderson localisation.

In the calculations for the correlations of the localisation lengths, as shown in Section 4.3 we see no significant correlations within the biological λ -DNA sequence beyond sizes exceeding the length of the particular window chosen. Initially, we put forward the hypothesis that coding fragments of biological DNA and non-coding parts would give different localisation behaviour. However, this is not the case within the accuracy of our calculations. Furthermore, when the plots are compared to the scrambled λ -DNA sequence, we see no systematic differences between them. Basically, for both sequences, we find that the correlation function starts at 1 due to normalisation and steadily drops to nearly zero when the window size is reached. From then on, the behaviour becomes complicated and no systematic tendency can be extracted. For large window sizes more regularity seems to be present but a separate study focusing on the differences of the coding versus the non-coding “junk” of λ -DNA has not produced a clear picture either [49]. Again with respect to the scrambled sequence the deviations appear arbitrary. Nevertheless, it might be that there is some periodicity in the way these correlations appear or vanish that would also differentiate the behaviour between the biological and scrambled sequences.

In this last paragraph, we will point out the results we regard as having the most important implications. Firstly, we saw that all non-periodic sequences of DNA, i.e. random-ATGC and λ -DNA, give localised behaviour which indicates that

they are insulating at $T = 0$. Telomeric DNA is a very encouraging and interesting sequence because it gives infinite localisation lengths in the ordered regime which are not significantly reduced when little disorder is introduced; whilst being a real biological sequence. As far as the correlations within biological λ -DNA are concerned, we see no difference between the localisation properties of the coding and non-coding parts. Finally, when backbone onsite disorder is introduced to the system, new states appear in the gap and in particular for increasing binary onsite disorder, we see more delocalisation.

Appendix A

```
SetDirectory["C:\\Documents and Settings\\Rudolf A Roemer\\My Documents\\physics projects\\tmseidDNA\\DNA"]
C:\\Documents and Settings\\Rudolf A Roemer\\My Documents\\physics projects\\tmseidDNA\\DNA

$Log = False;

General::spell1 : Possible spelling error: new symbol name "$Log" is similar to existing symbol "Log".
```

Random DNA

```
len = 10000;
Bases = {"a", "t", "g", "c"};

General::spell1 : Possible spelling error: new symbol name "Bases" is similar to existing symbol "Cases".

filename = StringJoin["DNA-R-", StringJoin[Bases], "-", ToString[len], ".dna"];

DNAseq = StringJoin[Table[ Bases[[Random[Integer, {1, Length[Bases]}]]], {i, 1, len}]];

If[$Log, Print[ {filename, DNAseq}]]

Export[filename, DNAseq, "Lines"]

DNA-R-atgc-10000.dna
```

Random GC-DNA

```
len = 10000;
Bases = {"g", "c"};

filename = StringJoin["DNA-R-", StringJoin[Bases], "-", ToString[len], ".dna"];

DNAseq = StringJoin[Table[ Bases[[Random[Integer, {1, Length[Bases]}]]], {i, 1, len}]];

If[$Log, Print[ {filename, DNAseq}]]

Export[filename, DNAseq, "Lines"]

DNA-R-gc-10000.dna
```

Appendix B

```
<<Statistics`DescriptiveStatistics`
<<Statistics`ContinuousDistributions`
<<Graphics`MultipleListPlot`
<<Statistics`MultiDescriptiveStatistics`

SetDirectory["C:\Documents and Settings\user\My Documents\tmse2dWinDNA\tmse2dWinDNA project"]
C:\Documents and Settings\user\My Documents\tmse2dWinDNA\tmse2dWinDNA project

$LOG = False;

allfiles = FileNames["*0.raw"]

{energyzero00048503-00000001_00048503_00000001-00000100.raw, energyzero00048503-00000001_00048503_(

cortab = {};
Do[
  If[MemberQ[allfiles, {filename = allfiles[[ifile]]}],
    StylePrint["working on " <> filename, "Subsection"];
    Get[filename];
    raw = Transpose[Drop[data, -1]];
    gamma = raw[[9]];
    loclength = 1/gamma;

    mean = Mean[loclength];
    variance = Variance[loclength];
```

```

n = Length[loclength];
cor = Table[Divide[Sum[(loclength[[i]] - mean) * (loclength[[i + k]] - mean),
    {i, 1, n - k}], Sum[(loclength[[i]] - mean)^2, {i, 1, n}]]
    ,
    {k, 1, n - 1}
];

cor = Transpose[{Table[m, {k, 1, n - 1}], cor}];
If[$LOG, Print[{ifile, cor}]];

AppendTo[cortab, cor];

ListPlot[cor, AxesLabel -> {"k", "cor"}, PlotLabel -> allfiles[[ifile]], PlotRange -> All];
],
{ifile, 1, 1Length[allfiles]}
1

```

Bibliography

- [1] Wavefunction Inc., 18401 Von Karman Ave., Suite 370 Irvine, CA 92612. *SPARTAN version 5.0, User's Guide*.
- [2] <http://www.rothamsted.bbsrc.ac.uk/notebook/courses/guide/chromo.htm>.
- [3] G. Cuniberti, L. Craco, D. Porath, and C. Dekker. Backbone-induced semi-conducting behavior in short DNA wires. *Phys. Rev. B*, 65:241314–4, 2002.
- [4] G. M. Cooper. *The Cell, a Molecular Approach*. ASM Press and Sinauer Associates, Inc., Washington, D. C., 2000.
- [5] J. D. Watson and F. H. C. Crick. A structure for deoxyribose nucleic acid. *Nature*, 171:737, 1953.
- [6] S. Cuenda and A. Sanchez. Disorder and fluctuations in nonlinear excitations in dna. 2004. ArXiv: q-bio.BM/0403003v1.
- [7] M. Peyrard. Nonlinear dynamics and statistical physics of dna. *Nonlinearity*, 17:R1–R40, 2004.
- [8] H. Park, S. Kim, A. Sancar, and J. Deisenhofer. Crystal structure of DNA photolyase from *escherichia coli*. *Science*, 268:1866–1872, 1995.
- [9] P. O'Neil and E. M. Fielden. Primary free radical processes in DNA. *Adv. Radiat. Biol.*, 17:53–120, 1993.

- [10] J. Retel, B. Hoebee, J. E. F. Braun, J. T. Lutgerink, E. Van der Akker, H. Wanamarta, H. Joenje, and M. V. M. Lafleur. Mutational specificity of oxidative DNA damage. *Mutation Res.*, 299:165, 1993.
- [11] E. Boon, A. Livingston, N. Chmiel, S. David, and J. Barton. DNA-mediated charge transport for DNA repair. *PNAS*, 100:12543, 2003.
- [12] D. Porath, G. Cuniberti, and R. Di Felice. Charge transport in dna-based devices. *Topics in Current Chemistry*, 237:183, 2004.
- [13] A. Rakitin, P. Aich, C. Papadopoulos, Y. Kobzar, A.S. Vendeneev, J.S. Lee, and J.M. Xu. Metallic conduction through engineered DNA: DNA nanoelectric building blocks. *Phys. Rev. Lett.*, 86:3670–3673, 2001.
- [14] Y. A. Berlin, A. L. Burin, and M. A. Ratner. On the long-range charge transfer in dna. *The Journal of Physical Chemistry*, 104:443–445, 2000.
- [15] M. A. O’Neil, C. Dohno, and J. K. Barton. Direct chemical evidence for charge transfer between photoexcited 2-aminopurine and guanine in duplex dna. *Journal of the American Chemical Society Communications*, 126:1316–1317, 2004.
- [16] S. Delaney and J. K. Barton. Long-range dna charge transport. *J. Org. Chem.*, 68:6475–6483, 2003.
- [17] C. R. Treadway, M. G. Hill, and J. K. Barton. Charge transport through a molecular π -stack: Double helical dna. *Chemical Physics*, 281:409–428, 2002.
- [18] E. Braun, Y. Eichen, U. Sivan, and G. Ben-Yoseph. DNA-templated assembly and electrode attachment of a conducting silver wire. *Nature*, 391:775–778, 1998.

- [19] H-W. Fink and C. Schonenberger. Electrical conduction through DNA molecules. *Nature*, 398:407–410, 1999.
- [20] H. Nakao, M. Gad, S. Sugiyama, K. Otobe, and T. Ohtani. Transfer-printing of highly aligned DNA nanowires. *J. Am. Chem. Soc.*, 125:7162–7163, 2003.
- [21] A. Asabaeva and M. Tang. Electrical conductivity in oriented DNA. *National Nanofabrication*, pages 56–57, 2000.
- [22] Y. Okahata, T. Kobayashi, K. Tanaka, and M. Shimomura. Anisotropic electric conductivity in an aligned DNA cast film. *Journal of the American Chemical Society*, 120:6165–6166, 1998.
- [23] D. Porath, A. Bezryadin, S. Vries, and C. Dekker. Direct measurement of electrical transport through DNA molecules. *Nature*, 403:635–638, 2000.
- [24] P. J. Pablo, F. Moreno-Herrero, J. Colchero, J. Gomez Herrero, P. Hererro, P. Baro, A. M. an Ordejon, J. M. Soler, and E. Artacho. Absence of dc-conductivity in λ -DNA. *Phys. Rev. Lett.*, 85:4992–4995, 2000.
- [25] H. Wang, J. P. Lewis, and O. F. Sankey. Band-gap tunneling states in dna. *Phys. Rev. Lett.*, 93:016401, 2004.
- [26] W. Zhang and S. E. Ulloa. Structural and dynamical disorder and charge transport in dna. *Microelectronics Journal*, 35:23–26, 2004.
- [27] S. Roche. Sequence dependent dna-mediated conduction. *Phys. Rev. Lett.*, 91:108101, 2003.
- [28] C. J. Murphy, M. A. Arkin, Y. Jenkins, N. D. Ghatlia, S. Bossman, N. J. Turro, and J. K. Barton. Long-range photoinduced electron transfer through a DNA helix. *Science*, 262:1025–1029, 1993.

- [29] M. Bixon, B. Giese, S. Wessely, T. Langenbacher, M. E. Michel-Beyerle, and J. Jortner. Long-range charge hopping in DNA. *PNAS*, 96:11713, 1999.
- [30] Z.G. Yu and X. Song. Variable range hopping and electrical conductivity along the DNA double helix. *Phys. Rev. Lett.*, 86:6018–6021, 2001.
- [31] J. Zhong. Electronic transport in DNA molecules with backbone disorder. 2003. private communication.
- [32] W. Zhang and S. E. Ulloa. Extended states in disordered systems: Role of off-diagonal correlations. *Phys. Rev. B*, 69:153203, 2004.
- [33] S. Roche, D. Bicout, E. Macia, and E. Kats. Long range correlations in dna: Scaling properties and charge transfer efficiency. *Phys. Rev. Lett.*, 91:228101, 2003.
- [34] R. N. Barnett, C. L. Cleveland, A. Joy, U. Landman, and G. B. Schuster. Charge migration in DNA: ion-gated transport. *Science*, 294:567–571, 2001.
- [35] R. Bruinsma, G. Gruner, M. R. D’Orsogna, and J. Rudnick. Fluctuation-facilitated charge migration along DNA. *Phys. Rev. B*, 85:4393–4396, 2000.
- [36] C. Dekker and M. A. Ratner. Electronic properties of dna. *Physics World*, 14:29, 2001.
- [37] R. G. Endres, D. L. Cox, and R. P. Singh. The quest for high-conductance dna. *Rev. Mod. Phys.*, 76:195–214, 2004.
- [38] P. W. Anderson. Absence of diffusion in certain random lattices. *Phys. Rev.*, 109:1492–1505, 1958.
- [39] B. Kramer and A. MacKinnon. Localization: theory and experiment. *Rep. Prog. Phys.*, 56:1469–1564, 1993.

- [40] M. Schreiber (Eds.) K. H. Hoffmann. *Computational Statistical Physics: From Billiards to Monte Carlo*. Springer-Verlag, Berlin, 2002.
- [41] R. A. Römer. From localization to delocalization in disordered systems. 1999. Habilitationsschrift, Technische Universität Chemnitz.
- [42] N. R. Walet and W. J. Zakrzewski. A simple model of the charge transfer in DNA-like substances. 2004. ArXiv: cond-mat/0402059v1.
- [43] I. L. Garzon, E. Artacho, M. R. Beltran, A. Garcia, J. Junquera, K. Michaelian, P. Ordejon, C. Rovira, D. Sanchez-Portal, and J. M. Soler. Hybrid DNA-gold nanostructured materials: an ab-initio approach. *Nanotechnology*, 12:126–131, 2001.
- [44] H. Yamada, E. B. Starikov, D. Hennig, and J. F. R. Archilla. Localization properties of electronic states in polaron model for poly(dg)-poly(dc) and poly(da)-poly(dt) dna polymers. 2004. ArXiv: cond-mat/0407148v1.
- [45] F. J. Dyson. The dynamics of a disordered linear chain. *Phys. Rev.*, 92:1331–1338, 1953.
- [46] P. Biswas, P. Cain, R. A. Römer, and M. Schreiber. Off-diagonal disorder in the Anderson model of localization. *phys. stat. sol. (b)*, 218:205–209, 2000. ArXiv: cond-mat/0001315.
- [47] A. Eilmes, R. A. Römer, and M. Schreiber. The two-dimensional Anderson model of localization with random hopping. *Eur. Phys. J. B*, 1:29–38, 1998.
- [48] P. Y. Yu and M. Cardona. *Fundamentals of Semiconductors*. Springer-Verlag, Berlin, 1999.
- [49] R. A. Römer. private communication.