The University of Manchester

Molecular Dynamics Study of the DNA Melting Mechanism using a New Hydrogen Bond Definition

A thesis submitted to The University of Manchester for the degree of M.Phil in the Faculty of Engineering and Physical Sciences

2011

Emily Viney

School of Chemistry/ Theoretical Chemistry

List of Contents

Total Word Count 20,081

List of Figures

List of Tables

The University of Manchester
Faculty of Engineering and Physical Sciences

ABSTRACT OF THESIS submitted by **Emily Viney**

For the degree of **M.Phil in Theoretical Chemistry** and entitled **Molecular Dynamics Study of the DNA Melting Mechanism using a New Hydrogen Bond Definition**

Date of submission: **30/09/2011**

A detailed melting mechanism for DNA at the atomistic level is fundamental to the understanding of much of biochemistry and biotechnology. There are few molecular dynamics studies of this process to date. Previous work has proposed a two state mechanism for the unfolding but more recent studies have suggested multiple intermediates. Here molecular dynamics simulations of homo-oligomeric 4-mer and 8-mer duplexes were performed. Melting was induced by heating the systems to 373 K. The hydrogen bonding array was analysed using a newly developed definition which makes use of intermolecular potentials. The definition gives each hydrogen atom donating to only one acceptor atom at all times. The melting mechanism proposed goes via repeated fraying of terminal base pairs and shifting of the strands towards the 5' direction. This study identifies that there is a preference for unpaired bases formed to be at the 5' ends rather than the 3' ends. The fraying and shifting processes are thought to be induced by loss of planarity between base pairs. This mechanistic information is limited by the length and the homo-oligomeric nature of the strands considered. However, the proposed mechanism still gives an indication of the melting mechanism for more complex and biologically important DNA structures.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgements

I would like to thank my supervisor, Dr. Richard Henchman for his contribution to and guidance with this project. Of particular note was his input towards the various coding aspects involved. I would also like to thank Ian Cotton and Mike Croucher for their invaluable help with computational resources by establishing AMBER on the University Condor system.

1. Introduction

1.1. Overview of Simulation

Computational simulation is a useful and ever expanding technique allowing for the study of a plethora of diverse chemical systems. It enables the effective study of systems which may be too dangerous, expensive or microscopic to look at experimentally.[1] There are a broad range of computational techniques available and deciding which to use often involves a compromise between accuracy and computational cost. Current computational methods can generally be classified as either quantum mechanics or molecular modelling.[2]

Molecular modelling techniques are based on the Born-Oppenheimer approximation which allows for the separation of electronic and nuclear motions. The distinctly smaller mass of electrons enables them to react almost instantaneously to sudden changes in nuclear positions. Thus the potential energy of a system can be considered as a function of the nuclear coordinates only. Molecular modelling techniques include molecular dynamics (MD) and Monte Carlo methods.

1.2. Project Outline

The melting process is loosely defined as the unfolding of double stranded DNA to form two single strands. This process is illustrated in Figure 1.1.

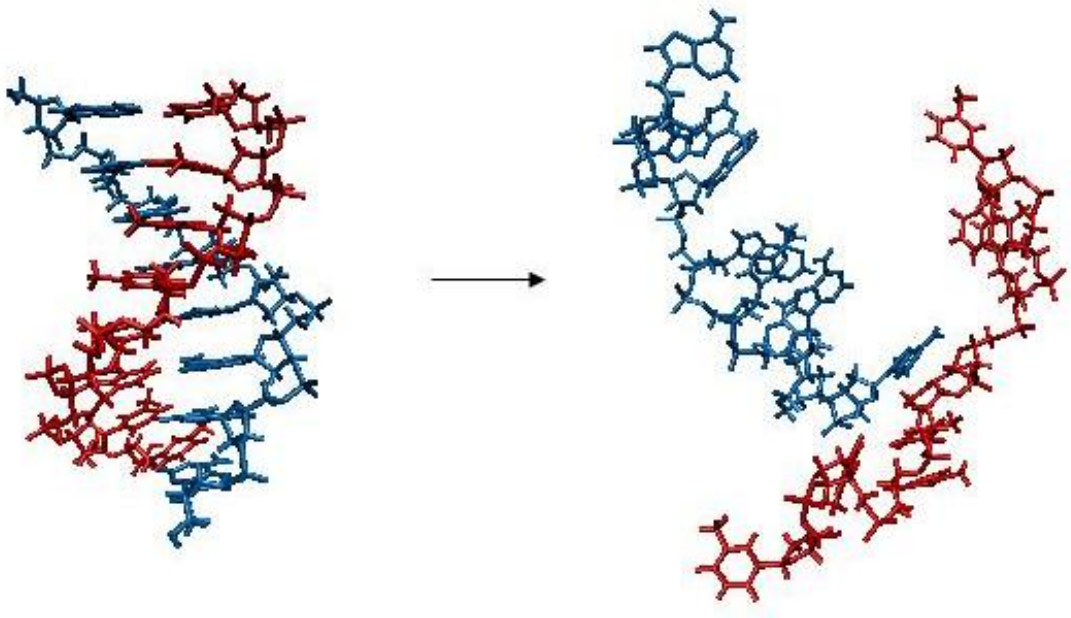


Figure 1.1. Conformations of d(A_8)d(T_8) DNA at the start and end of the melting process. The A-strand is represented in red and the T-strand in blue. The 5' end of the A-strand is located at the top of each image.

The main focus of this project was to determine at the atomistic level a mechanism for the melting of DNA. This was achieved by looking at MD simulations of DNA of varying length and sequence. MD trajectories were analysed using a newly developed hydrogen bonding definition. The definition avoids the used of arbitrary cut-off values; such as angle and distance restraints, whilst also avoiding over complication which can be introduced by using charge density parameters. The new definition was scripted into the hydrogen bonding analysis belonging to the Ptraj module of AMBER and further analysis and characterisation was carried out using shell and Perl scripts.

A clear understanding of the mechanism could aid the design of DNA microarrays used in gene expression, drug design and DNA-based computing.[3] The results of this

study could be further applied to understand the structure and stability of the DNA replication fork.

2. Computer Simulation

2.1. Molecular Mechanics

2.1.1. Force Fields

Molecular mechanics methods use force fields and so are often labelled force field techniques. They look at systems in terms of nuclear positions only; as such they can be much less computationally expensive than quantum mechanics and can be applied to larger systems.[2] Molecular mechanics for this reason cannot be used to look at properties which are reliant on the electron distribution. Force fields can be used in energy minimisation and molecular dynamics (MD). Force fields calculate the overall potential energy of a system in terms of atomic positioning. They are generally referred to as empirical techniques.[4] Force fields model the potential energy as the sum of the individual energy components. The specific functional form varies depending on the family of force fields in question. The generalised form of the force field equation expresses the potential energy (U) at position (r) as:

$$U(r) = U_{bonds} + U_{angles} + U_{dihedrals} + U_{electrostatics} + U_{vanderWaals}$$

The force field parameters - force constants, equilibrium bond lengths, equilibrium angles and charges - can be derived either from experiment or *ab initio* quantum calculations using computational software such as Gaussian.[5] Parameters are commonly ascertained for a small number of systems. They can then be applied to a much larger range with the assumption that the parameters are transferable.

Choosing the correct force field is dependent on the type of simulation being attempted and the type of molecules concerned. Some of the more common force

field families are CHARMM (Chemistry at Harvard Molecular Mechanics),[6] GROMOS (Groningen Molecular Simulation Packages),[7] OPLS (Optimised Potential for Liquid Simulations)[8] and AMBER (Assisted Model Building and Energy Refinement).[9] The most accurate force fields and hence computationally expensive are deemed 'all atom' where each atom is modelled separately. For larger systems, such as proteins, atoms are sometimes grouped together to reduce computational cost. These are known as coarse-grain methods.[10]

2.1.2. AMBER (Assisted Model Building and Energy Refinement)

AMBER force fields are commonly used force fields for the simulation of nucleic acids, proteins and other biomolecules. They have shown consistent reproducibility for many properties of nucleic acid systems when using explicit solvent effects.[1,2] In particular they have accurately depicted hydrogen bonding and stacking interactions.[3-5] AMBER is also the name of the software package used to simulate AMBER force fields. The AMBER force fields all have the following generalised functional form:[11]

$$U(r^N) = \sum_{bonds} \frac{k_r}{2}(r - r_o)^2 + \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_o)^2 + \sum_{dihedrals} \frac{V_n}{2}\{1 + \cos(n\theta - \gamma)\}$$

$$+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \varepsilon_{i,j} \left[\left(\frac{r_{o_{ij}}}{r_{ij}} \right)^{12} - 2\left(\frac{r_{o_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} \right\}$$

$U(r^N)$ comprises of five separate energy terms and is the potential energy as a function of positions r for a total of N particles. The first term is the bonding term which describes bond stretching energies using a harmonic potential, where k_r is the bond force constant, r is bond length and r_o is equilibrium bond length.[12] Next is the

angle term describing bond bending energies also using a harmonic potential where k_θ is the angle force constant, θ is the angle and θ_o is the equilibrium angle value. The shape of these harmonic potentials can be seen in Figure 2.1.

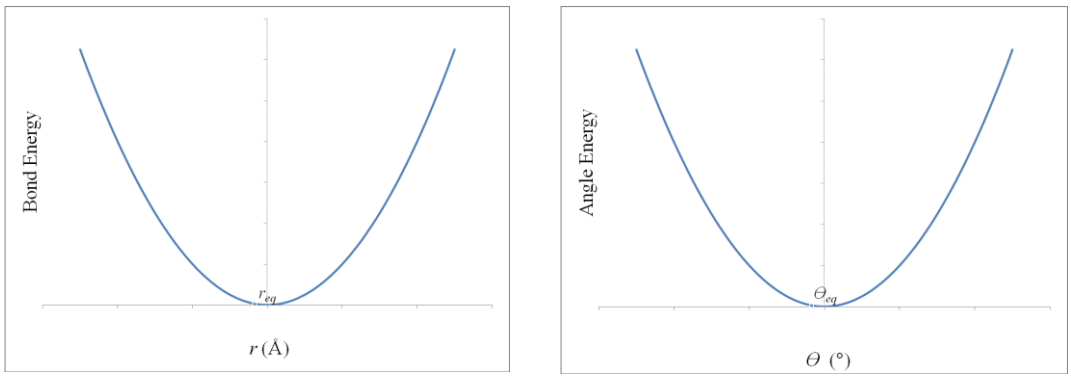


Figure 2.1. Bond (left) and angle (right) harmonic potentials.

The dihedral term is represented by the energy required to twist a bond (torsion) and is dependent on bond order. V_n is the barrier to rotation, γ is the phase factor and θ is the torsion angle. Each bond can be represented by multiple dihedral terms and so the overall torsional term is represented by a Fourier series, this represents the periodic dihedral function as an oscillating cosine function. [2] The final term represents the total non-bonded energy between all atom pairs, i and j. The first part of the non-bonded expression describes the van der Waals energy using the 6-12 Lennard-Jones potential (see Figure 2.2), where N represents the number of point charges in the molecule, r_o is the collision diameter and ε_o is the well depth.[13]

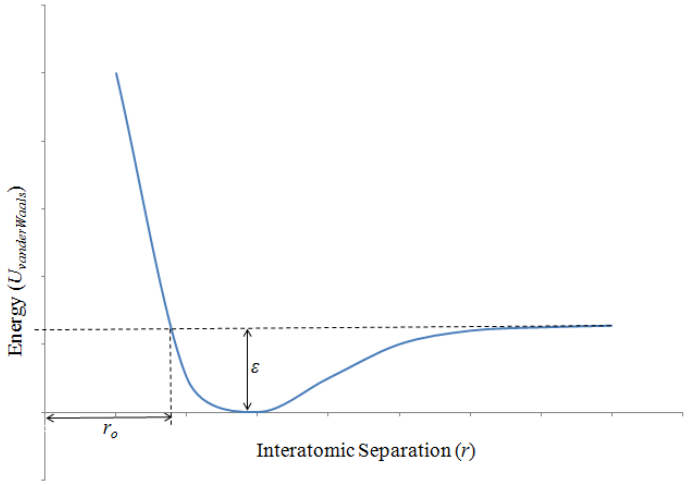


Figure 2.2. The Lennard-Jones 6-12 potential.

The second part of the non-bonded potential represents the electrostatic interactions between the pairwise atoms where q_i and q_j represent their respective charges. The functional form of the force field is constant for all AMBER force fields and the parameters will vary depending on the specific force field.

A possible issue with the AMBER force fields is that they ignore polarisation and charge transfer effects. This is true of even the more recently developed force fields.[6] A CHARMM polarisable force field for nucleic acids based on the classical Drude oscillator model has recently been developed.[7] A singularly charged (Drude) particle is attached to all the atoms, except the hydrogens, by a harmonic spring to allow for polarisability. The force field was developed so that it could accurately reproduce crystal-phase thermodynamic and structural data in addition to structural and energetic properties for several different hydrogen bonded nucleic acid base pairs.

2.1.3. Force Field Simulations of Water

Computational efficiency is key when simulating water due to the sheer number of water molecules present in a system. The most commonly used water models - SPC,[14] SPC/E,[15] TIP3P[16] and TIP4P - model each water molecule rigidly with the non-bonding interactions being described by pairwise coulumbic and Lennard-Jones expressions. Both TIP3P and SPC models use three interaction sites with slightly differing geometries, charges and Lennard-Jones parameters. The SPC/E model is the more recent version of the SPC model with improved geometries, hydrogen charges and Lennard-Jones parameters. The TIP4P model differs from the TIP3P model in the location of charges; it moves the negative charge from the oxygen to the bisecting point of the HOH angle, see Figure 2.3.

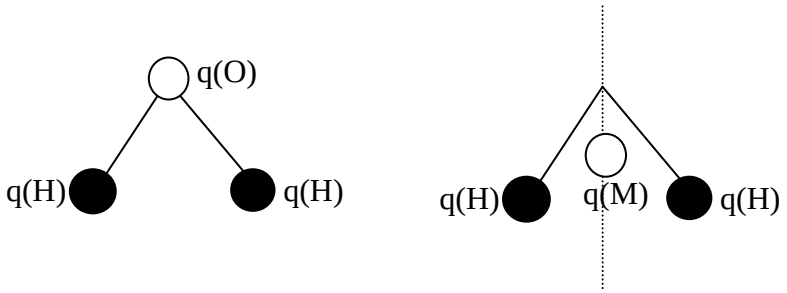


Figure 2.3. Left: Location of point charges for SPC, SPC/E and TIP3P water models. Right: Location of point charges for the TIP4P water model.

2.1.4. Energy Minimisation

One of the main uses of molecular mechanics is for energy minimisation. It involves the examination of changes in energy with atomic coordinates represented by a potential energy surface or hypersurface. The lowest points of the potential energy

surface represent stable system states. Algorithms are utilised to search the system's energy surface; these minimum points are referred to as local minima.[17] Mathematically minima can be defined as the point when the first derivative of the surface is equal to zero and the second derivative is positive.

1st derivative:

$$\frac{\partial U}{\partial x_i} = 0$$

2nd derivative:

$$\frac{\partial^2 U}{\partial x_i^{\,2}} > 0$$

Where U is the potential energy and x_i is the position on the energy surface. The overall lowest minimum point is called the global minimum. Finding any minima requires the use of algorithms and there are a number of different options. They generally involve making gradual changes to the atomic coordinates in order to move the system to the nearest minimum point. Derivative methods directly use the mathematical expressions above. First order minimisation methods use only the first derivative expression.[2] Second order minimisation methods utilise both the gradient and the second derivative expression. The second derivative gives details of the function's curvature indicating whether the point is a maximum or minimum. In this work only first derivative methods will be explored in further detail. Two such methods are the steepest descents method and the conjugate gradients minimisation procedure.

The steepest descents method involves taking arbitrary steps along the direction parallel to the net force which equates to moving along the steepest negative gradient.[18]

$$S_k = \frac{-g_k}{|g_k|}$$

S_k is the 3N dimensional unit vector and g is the gradient at point k. If the first step leads to a reduction in energy then the step size is increased and the process repeats. If a step leads to an increase in energy then the step size is reduced. The steepest descents method works well initially, removing the largest strains in the system; this is especially useful when an initial structure is a long distance from a minimum. A disadvantage of the algorithm is that it can oscillate close to minima leading to it constantly correcting itself.

One way to avoid the oscillatory behaviour of the steepest descents method is to implement the conjugate gradients minimisation once the large initial strains are removed.[19] The direction of movement in the conjugate gradients method is given by:

$$V_k = -g_k + y_k V_{k-1}$$

V_k is the new direction vector, g is the gradient at point k, V_{k-1} is the previous direction vector and y_k is the scalar constant. y_k can be calculated by:

$$y_k = \frac{{g_k}^2}{{g_{k-1}}^2}$$

Energy minimisation techniques have been extensively used to study protein folding interactions.[20-22] For instance it is possible to quantify the relative probabilities of an amino acid residue being encompassed into a protein or being exposed to solvent based only on the amino acid side chain.[23] Energy minimisation is limited because it

only considers minima and ignores the rest of the structures. This means entropy effects are not taken into account, so calculations such as the comparison of equilibrium between local minima are not possible.

If all minimum states can be identified then partition functions derived from statistical mechanics can be used to calculate useful thermodynamic properties. Atomic level information at the atomic level; this information can then be translated into macroscopic properties using the principles of statistical mechanics. There are many types of ensemble with different properties of the system remaining constant or varying throughout the simulation ensemble. Common ensembles include the canonical ensemble where the number of particles N, volume V and temperature T of the system remain constant; the microcanonical ensemble with constant N, V and energy E; the isobaric-isothermal ensemble with constant N, T and pressure P and the grand canonical ensemble where V, T and chemical potential μ are constant. However this is only possible for small molecules in the gas phase.[24] For larger more complex systems other techniques such as Monte Carlo and MD are employed.

2.2. Molecular Dynamics

2.2.1. Newton's Laws of Motion

Classical MD calculates the instantaneous forces acting on particles using force fields. Other MD methods use quantum chemical models or a mixture of force fields and quantum chemistry, known as quantum mechanical or molecular mechanical methods (QM/MM).[8] Using MD simulations certain thermodynamic values such as

internal energy, enthalpy, Helmholtz free energy and Gibbs free energy can be calculated.[25]

MD monitors how a particle's position and velocity change over time. The procedure is based on Newton's Laws of Motion which are:

1st: A body continues to move in a straight line at constant velocity unless a force acts upon it.

2nd: Force is equal to the mass m times acceleration a:

$$F = ma$$

By expressing the acceleration in terms of change in position and time Newton's Second Law can be written as:

$$\frac{\partial^2 x_i}{\partial t^2} = \frac{F_{xi}}{m_i}$$

x is the position of particle *i* in the *x* coordinate direction and *t* is the time.

3rd: Every action has an equal and opposite reaction.

By integrating Newton's Second Law of Motion the change in atomic positions and velocities with time can be ascertained, these then comprise the MD trajectory.

2.2.2. Finite Difference Methods

There are several different methods for integrating Newton's Second Law, the most commonly used of which are the finite difference methods. They utilise continuous potential models which calculate the change in force on each particle from the particle's change in position and the change in position of any interacting particles.[26] The integration process is divided into stages separated in time by constant time step δt. The force acting on each particle at each time is calculated by summing together

all its interactions with other particles. The forces, positions and velocities are then accumulated at time t and used to calculate new positions and velocities at time $t+\delta t$. It is assumed that the force over the time step δt remains constant. From the newly obtained positions and velocities the corresponding forces can be calculated and the process continued.

Finite difference methods use a variety of different algorithms to integrate Newton's Second Law, the most extensively used of which is the Verlet.[27] The Verlet algorithm uses positions and acceleration from time t and time $t-\delta t$ (the previous time step) to calculate the positions and acceleration at the new time step, $t+\delta t$. The new positions can be calculated from the old positions and accelerations such that:

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \delta t^2 a(t)$$

Here r represents the positions of the atoms and a represents their accelerations. Some accuracy is lost when summing together $\delta t^2 a(t)$ term and the larger $2r(t)$ and $v(t-\delta t)$ terms. New velocities can be calculated from the difference in positions at times $t+\delta t$ and $t-\delta t$ by:

$$v(t) = \frac{\left[r(t+\delta t) - r(t-\delta t)\right]}{\delta t}$$

Additionally velocities have to be calculated separately which is not possible until the positions have been calculated. The Verlet method requires minimal computational time and storage space.[28]

A variant on the Verlet algorithm is leapfrog integration[29] where positions and velocities are calculated at interspersed points. For example if the velocity is calculated at time t then the position is calculated at time $t+\delta t$. The leapfrog method has two distinct advantages over Verlet integration. Firstly, there is no loss of precision when summing and, more importantly, velocities are included in the main calculation such that:

$$v(t) = \frac{1}{2} v(t + \frac{1}{2}\delta t) + v(t - \frac{1}{2}\delta t)$$

A major disadvantage of the leapfrog method is positions and velocities are not known for the same time step and hence it is not possible to calculate the kinetic energy at all time steps.

2.2.3. Time Step

The time step δt is the interval over which forces acting on the particles are assumed to be constant. Choosing the correct time step is essential to the success of the simulation. If the time step is too small then not enough of the hypersurface is explored; if it is too large it can lead to inaccuracies during the integration process. The largest possible time step must be less than the smallest time interval over which the forces acting on the particles remain approximately constant. However, this may not be known. The smaller the time step, the longer the simulation and thus the higher computational cost.

Selecting a time step an order of magnitude smaller than the shortest motion in the system is the standard procedure. Generally the shortest motions correspond to the vibrations of hydrogen atoms and so by constraining these hydrogens to their equilibrium values using the SHAKE algorithm[30] a larger time step can be selected and computational cost reduced.

2.2.4. Periodic Boundary Conditions

Treating boundaries appropriately is an important consideration for MD; poor treatment prohibits the accurate calculation of macroscopic properties. Periodic

boundary conditions replicate the three dimensional solvent shape infinitely in all directions so that even the atoms at the edges of the solvent cage exhibit forces as though they were in the bulk.[7] If at any point during the MD a particle leaves its respective box, then an image particle enters the box from the opposite side, thus maintaining a constant number of particles in each box. Specific formulae can then be used to translate the particles back into the central box.

The most commonly used shapes for periodic boundary conditions are cubic and truncated octahedral. For DNA the truncated octahedron is more appropriate as rectangular solvation can lead to problems with solute rotation and cubic solvation adds unnecessary water molecules and so increases computational cost.[31] Problems can arise when using periodic boundary conditions if the size of the box, referred to as the cut-off, is not large enough to encompass all long range electrostatic interactions.

2.2.5. Particle-Mesh Ewald

Particle-Mesh Ewald (PME) can be used to treat long range electrostatics so that the periodic boundary cut-off does not have to be ridiculously large. It is a variant of the Ewald summation method[32] and utilises a Fourier transform algorithm in order to treat a system's electrostatics more accurately. PME accounts for all interactions between a particle and its neighbours in its box and all of its images in the periodic array. PME reduces the simulation timescale by ignoring atom pairs for which the interatomic distance exceeds a certain cutoff.[33] The appropriate cut-off depends on the type system that being modelled but standard values for DNA are 8.0 Å or 10.0

Å.[34-37] PME does enable the use of a smaller unit cell but implementation is still moderately expensive.

The PME method has been used to model many different biological systems including proteins, lipid bilayers and DNA.[9,38]

2.3. Quantum Mechanics

Quantum mechanical techniques are used for the calculation of properties which are explicitly based on electron distributions. The two major areas of computational quantum mechanics are *ab initio* methods and semi-empirical methods.[39]

2.3.1. *Ab initio* Methods

Ab initio techniques are used to calculate the locations of transition states, thermodynamic and structural properties. They are also used for the derivation of parameters used in molecular force fields.[40] *Ab initio* can be translated as 'from first principles', so the only inputs into the calculations are physical constants such as electron mass.[41] Calculations are used to obtain the electronic wavefunctions, these represent all possible states and all geometries of these states. From the wavefunctions potential energy functions can be derived and from these molecular properties calculated.

Commonly *ab initio* calculations make use of the Hartree-Fock and Roothaan-Hall[42,43] equations:

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda=1}^{k}\sum_{\sigma}^{k} \rho_{\lambda\sigma}\left[(\mu\nu\left|\lambda\sigma\right.) - \frac{1}{2}(\mu\lambda\left|\nu\sigma\right.)\right]$$

Here F is the Fock matrix, H^{core} is the core Hamiltonian operator, k is the number of basis functions, ρ is the charge density matrix and $\lambda, \sigma, \mu, \nu$ are the basis functions. An important use of these calculations is in the determination of electric multipole moments.[44] The most commonly used computer software for *ab initio* calculations is Gaussian, created by John Pople in 1970.[45]

2.3.2. Semi-Empirical Methods

Ab initio techniques are extremely demanding computationally and therefore computationally expensive. As such it is sometimes necessary to include parameters which have been derived from experiment in order to reduce costs. These are known as semi-empirical quantum mechanical techniques and have been known to provide more accurate results than *ab initio* techniques.[46] Semi-empirical methods consider only valence electrons grouping the core electrons with the nucleus. Common computational software used to perform semi-empirical calculations are the MOPAC (molecular orbital package) and AMPAC[47,48] programs. Major inaccuracies can be introduced due to the use of certain fixed values and so the reduced computational cost can come at a penalty.

3. The Structure and Stability of DNA

3.1. The Structure of DNA

The structure of deoxyribose nucleic acid (DNA) as discovered by Watson and Crick[49] in 1953 is a double helix comprising of two sugar phosphate backbone strands with bases between the strands. The two backbones are held together by hydrogen bonds (HBs) between bases on different strands. Base pair stacking interactions between adjacent bases on the same strand stabilise the structure.[50] Standard Watson and Crick (WC) pairing of DNA bases is adenine (A) with thymine (T) and guanine (G) with cytosine (C), see Fig 3.1.

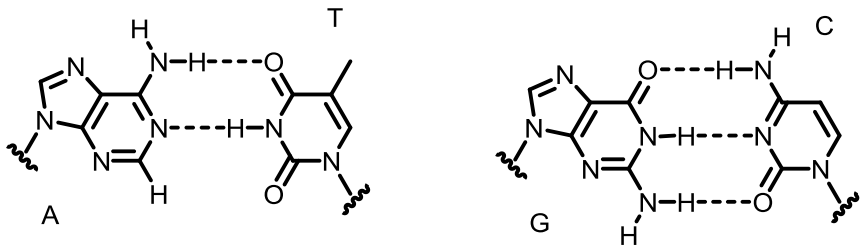


Figure 3.1. The Watson and Crick AT (left) and GC (right) base pairs.

The GC base pair consists of three Watson and Crick hydrogen bonding interactions while the AT pair has two.[51] The detailed sequence of base pairs is of particular importance due to the base contribution to the transcription and translation processes which are key stages in the replication of genes.

Base pairs within the backbone are not always planar in geometry with some HBs able to form out of plane of the purine and pyrimidine ring systems.[52-54] Nonplanarity is brought about partially by sp^3 hybridisation of the base amino groups.[55,56]

3.2. The Replication Fork

The DNA replication fork is formed when double stranded DNA (duplex) unwinds. The structure has two branches: the leading strand template and the lagging strand template.[57] Primers, short complementary strands of RNA, are able attach to the templates. DNA polymerase then uses these primers as starting sequences to build new complementary strands onto the templates. DNA can only extend in the 5'→3' direction and so this replication process is only efficient for the leading strand where replication proceeds 5'→3'. For the lagging strand, multiple short sections of DNA are synthesised in the 5'→3' direction, opposite to the direction of the replication fork movement.

This process leads to the newly formed double helix having an overhang at the 3' end which can cause loss of genetic material.[58] Chromosomal DNA contains repeating sequences of DNA at the end of each terminus known as telomere repeats. These repeats act as buffers enabling replication to proceed 5'→3' without loss of genetic information. However, after a number of replication cycles, chromosomal DNA begins to shorten as the telomere repeats get used up. Due to the integral role of the replication fork in transcription it is important to fully understand its structural properties.

3.3. Hydrogen Bonding

3.3.1. Role of the Hydrogen Bond

Steiner and Saenger[59] define a HB as "any cohesive interaction X-H--Y, where H carries a positive charge and Y a negative (partial or full) charge, and the charge on X is more negative than on H." Hydrogen bonding is key in the determination of the specific secondary, tertiary and quaternary structures of many biological molecules, namely proteins and nucleic acids.[60] It contributes unique properties to both chemical and biochemical stability and introduces conformational diversity.[61] Detection and evaluation of HB strength is particularly important for the development of new procedures in molecular recognition and supramolecular complex analysis.[62] High accuracy of these analysis procedures is reliant upon a robust HB definition. For the purposes of nucleic acid analysis there have been several well defined descriptions for what constitutes a HB.

3.3.2. Summary of Current Hydrogen Bonding Definitions in DNA

Both base stacking interactions[63-65] and interstrand HBs[35,66-73] have previously been studied using a wide spectrum of computational techniques. The definition of hydrogen bonding varies depending on the study and the information required. The DNA HB can be computationally difficult to study due to the vastness of biomolecular systems. These can involve thousands of potential donor and acceptor atoms which need to be considered. Coarse-grained methods can be used if there is a requirement for longer simulations times as they are much less computationally

demanding. The identification of hydrogen bonding criteria enables the detection of minima and therefore a more detailed exploration of the potential energy surface.[74] Generally the HB can be classified according to geometry criteria or charge density criteria.

Current HB analysis methods were researched before deciding a new definition needed to be developed. The new definition detects HBs based on the strongest formed intermolecular forces between atoms. It was important to understand the pros and cons of the existing analysis methods in order to successfully develop the new HB definition. Some previous analysis programs which were considered are summarised.

3.3.3. Ptraj

The Ptraj module of the AMBER software package[75] is a program used to process MD coordinates and trajectory files. Some of the uses of Ptraj include: calculating angles, computing atomic fluctuations, obtaining average structures, clustering trajectory frames, calculating the root mean squared standard deviation (RMSD) with time and analysing intermolecular attractions. In particular Ptraj contains a hydrogen bonding facility which can be used to track hydrogen bonding interactions. Ptraj identifies HBs according to two geometry parameters; a distance cut-off and an angle cut-off. The default selection is to have a D-A-H angle cut-off of 120°, where D is the donor atom, A is the acceptor and H the hydrogen. The default A-H distance cut-off is 2.5 Å. A list of potential donor and acceptor atoms needs to be compiled by the user; a measure implemented to reduce computational time.[76] The script can also be modified to include or exclude hydrogen bonding interactions with solvent

molecules. Ptraj gives a list of all the interactions which fulfil the specified parameters along with their D-A-H, A-H values and percentage occupancy. Percentage occupancy as a function of the trajectory can also be calculated in order to analyse hydrogen bonding as a function of time. A major issue with Ptraj HB analysis is HBs only slightly outside the cut-offs are omitted. Some preliminary HB analysis was run using Ptraj in this study. This was prior to the development of the new HB definition and all analysis using Ptraj was repeated using the new definition.

3.3.4. HBPLUS, HBexplore and MPlot

HBPLUS[77] is a software package which was designed specifically for the exploration of HBs in biological molecules. HBs are identified by five geometrical criteria, two distance restraints and three angle restraints, see Figure 3.2.

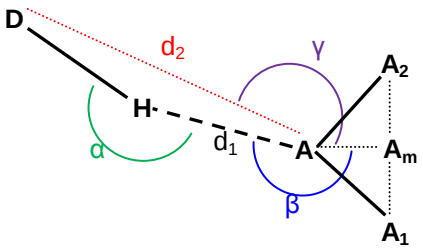


Figure 3.2. Geometrical criteria for selecting HBs used by HBPLUS. Showing the donor (D), acceptor (A), hydrogen (H), distance d_1 (H-A), distance d_2 (D-H-A), angle α (D-H-A), angle β (H-A-A_m) and angle γ (D-A-A_m).

HBexplore[78] was evolved from HBPLUS to produce an algorithm which offered superior analysis of nucleic acids; the main application of HBPLUS being the

analysis of proteins. The HBexplore algorithm is comprised of three steps; firstly the coordinates of the hydrogen atoms are generated and then there are two successive HB searches. Each of the searches is based on a different set of geometrical criteria and an interaction is deemed to be hydrogen bonding only if both sets of criteria are met. The first set of geometrical criteria is the same as that utilised by HBPLUS. The second includes the directionality of the hybrid orbital of atom A in addition to distance restraints, d_1 and d_2, and the D-H-A angle restraint. The directionality of the acceptor hybrid orbital (ε) is defined as the angle which forms between the linear d_1 bond and the hybrid orbital on A. It must construct an angle of less than 60°, see Figure 3.3.

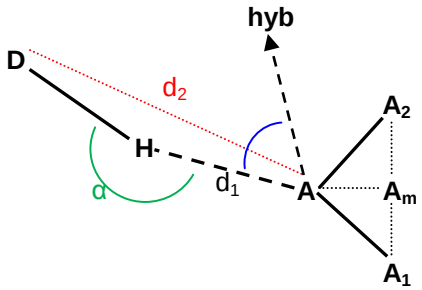


Figure 3.3. 2nd set of geometrical criteria for selecting HBs used by the HB explore program. Showing donor (D), acceptor (A), hydrogen atom (H), directionality of the acceptor hybrid orbital (ε), distance d_1 (H-A), distance d_2 (D-H-A) and angle α (D-H-A).

HBexplore enables not only the identification of HBs but also the calculation of the number of HBs per residue and the relative percentages of donor and acceptor atoms in the overall network. The program can be utilised in the detailed exploration of structures containing multiple bases and has been implemented during the analysis

of both RNA and DNA structures.[79-81] However, it presents the same problems with the use of cut-offs as Ptraj.

Mplot[82] is a web server that can be used for exploration of hydrogen bonded networks, the analysis of helix crossing angles and van der Waals contacts. The software selects potential HBs according to the HBexplore algorithms. This web integrated version of the HBexplore software then enables the specifications of the maximum distances and maximum angles between donor and acceptor atoms, avoiding some of the issues that arise when using preset cut-offs. These geometry cut-offs would give some broken hydrogen bonds between DNA base pairs and as such deemed inappropriate for this study.

3.3.5. 3DNA

The 3DNA program[83] is capable of the general analysis of antiparallel and parallel double helices, single stranded nucleic acid structures, triplexes, quadruplexes and other complex tertiary folding motifs in both DNA and RNA. It is able to classify double helix character for specific base pair steps. The hydrogen bonding array is defined in terms of spatial displacement and orientation of the interacting bases. The six parameters used to describe the relative positions of the interacting base pairs are shear, buckle, stretch, propeller, stagger and opening. Shear and stretch define the distances between base origins, opening defines the angle between the *x* axes of the two bases and buckle, propeller and stagger are used to describe base pair non-planarity. Using spatial displacement as hydrogen bonding criteria gives the same previously mentioned issues with broken base pair HBs and so 3DNA was not used in this study.

3.3.6. Electron Density Parameters

SHB interactions[70] is a program based on the extended Hückel calculation which utilises Muliken overlap populations as quantitative hydrogen bonding identification criteria. The use of Muliken analysis enables electron density to be partitioned between nuclei so that the atomic charge on each nucleus can be calculated. The more positive the electronic population overlap distribution between two atoms, the more it contributes towards their intermolecular interaction energy. SHB interactions enable an estimate to be made of individual HB's contribution to the overall stabilisation of DNA and DNA-drug complexes. It is possible to differentiate between different types of HBs and also to compare HBs of the same type.

Hydrogen bonding definitions based on charge density criteria can often involve more complex restraints. In a recent paper by Dominiak *et al.*[73]electron density analysis was performed on WC B-DNA base pairs using the quantum theory of atoms in molecules (AIMs). They utilised a charge density definition for HB analysis which consisted of eight charge density parameters.[84,85] They applied this definition to nearly a hundred crystallographic structures to examine the effect of conformational variability on intermolecular interaction energies. This would require consideration of electrons not possible using MD and would be highly computationally demanding.

3.4. Melting of DNA

3.4.1. Role in Gene Expression

The DNA double helix unwinds during transcription as a precursor to gene expression.[86] Transcription is the process in which two single DNA strands are used as templates to create two new antiparallel complementary RNA strands; the process is aided by DNA polymerase enzymes. Gene expression is the overall process of going from a gene to a gene specific product, usually protein or RNA. Discovering the DNA unfolding mechanism - the process of going from double stranded to single stranded DNA - is an important step in understanding the whole expression process. It is the formation of these two single DNA strands which allows transcription to take place. *In vivo* the unwinding of double stranded DNA utilises helicases. These are enzymes which use energy obtained from ATP hydrolysis to separate DNA into single strands. DNA melting as referred to in this thesis does not make use of helicases and alternative methods for driving DNA separation are discussed.

3.4.2. Defining the Melting Point

The melting point of a DNA duplex is loosely defined by the state at which the interstrand HBs between base pairs are destroyed and the duplex transforms into two single strands. More precise definitions of the melting point exist. Feng and Ping[87] define the melting point in terms of the motion of the hydrogen atom. If the average value of the mean square of the displacement of the hydrogen atom, $<\mu^2>$ is greater than zero melting has occurred. A higher value for $<\mu^2>$ means a greater melting

effect. For the homo-oligomeric duplex poly(dG)-poly(dC) under physiological conditions melting occurs when $<\mu^2>$ is 0.04 $Å^2$.

Baiesi *et al.*[88] identify melting by the extent of separation of the backbone sugar-phosphate strands. The exact melting point occurs when the minimum distance between the strands exceeds certain predefined threshold values. Threshold values of 10 Å and 20 Å were used as standard with the 20 Å value producing only marginally longer unwinding times.

Chen and Prohofsky[89] characterise DNA melting statistically where the melting temperature equates to base-pair opening probability nearing a half. The probability is calculated from the product of all the individual bond disruption probabilities. These individual probabilities can in turn be calculated from the maximum base pair separation before disruption and the mean square vibrational amplitude of the HBs. These are derived theoretically from the eigenvectors and eigenvalues of the system. For canonical B form DNA at physiological conditions the opening probability for a WC AT base pair was found to be between 10^{-3} and 10^{-5} and for a WC GC base pair 10^{-6}.

Self-consistent microscopic theory defines DNA melting as the point at which half the base pairs are separated, where separation of a single base pair corresponds to breakage of all the standard WC HBs.[90] To be classified as broken the HB must cross a saddle point on the potential energy curve induced by vibrational motion of the atoms. In real terms this saddle point crossing corresponds to a base pair separation of around 4 Å.

3.4.3. Thermodynamics of Melting

The melting of DNA into a single strand is a conformational change which is thermodynamically a first order phase transition. The process is accompanied by an increase in DNA volume and an increase in entropy.[91] The free energy required to melt a GC base pair is 3.5 kcal mol^{-1} and for an AT base pair is 1 kcal mol^{-1}.[92] During transcription, conditions surrounding DNA are physiological and the energy required to melt DNA is provided by the hydrolysis of adenosine triphosphate (ATP). During computational studies it is necessary to initiate melting using other methods.

Melting can be induced by increasing the temperature of the system to above the melting point, by increasing the salt concentration, changing the cosolvent or by varying the salt present. The DNA melting temperature is directly proportional to the external salt concentration, so it is possible to denature DNA at room temperature by increasing the salt concentration. The exact concentration required for full DNA melting is dependent on the composition of GC and AT base pairs. Less extreme changes in the external conditions surrounding DNA can lead to conformational changes such as conversion from B-DNA to A-DNA.[89]

3.4.4. Timescale of Melting

The unfolding process is found to occur on the microsecond timescale[93] - a possible limitation for MD studies. It has also been observed that DNA can undergo bond disruption and base pair opening at temperatures that are considerably lower than the melting point.[94-96] Even at 0°C the average base pair is open a small percentage of

the time. Distorted or open forms are the result of interbase HBs breaking and forming HBs with water. The double helix structure is generally considered stable as long as the rate constant for the formation of the closed state is larger than the rate constant for the open state. To properly access the differences between these fluctuations and full denaturation for real systems, it may be necessary to have MD timescales of 100s of milliseconds.

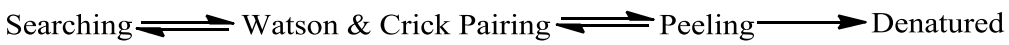
3.4.5. Previous Studies of Melting

Kinetics studies propose that for short chain lengths melting goes via a two state mechanism with unfolding originating at the AT base pairs and in the major grooves.[87,97-99] The Peyrad and Bishop model[100] uses two degrees of freedom for each base pair which retain to their equilibrium displacement along the interstrand HBs. HBs are approximated using a Morse potential and stacking interactions by anharmonic coupling. Force constants for intact HBs can be obtained by integrating the second derivative of the Morse potential weighted by a vibrational distribution function.[89] The melting mechanism is then explained by the transfer of force between WC HBs. When the force supplied to the imino (NH) group HBs exceeds a certain threshold, the excess force is transferred to the amino (NH_2) group HBs. Because amino group HBs have an additional degree of freedom, the extra stress is distributed in a non-uniform way and this application of excess force leads to melting. The model has also been extended to include helical degrees of freedom.[101] Coarse-grained models have been used to represent each nucleotide by two sites: the base and the sugar-phosphate backbone. This has enabled longer simulations but the model lacks atomistic detail.[102,103]

3.4.6. Molecular Dynamics Studies of Melting

Despite computational limitations it is necessary to use MD simulations to provide atomistic detail. This detail is important for understanding more complex scenarios such as protein induced DNA melting. Currently there are a limited number of studies concerning the MD of DNA melting.

MD studies do not agree with a two state mechanism but propose the formation of multiple compact intermediates. Wong and Pettitt[104] induced unwinding of the homo-oligomeric 12-basepair B-DNA complex $d(A_{12})d(T_{12})$ using explicit salt water and heating to 400 K. Sections of trajectory where the root mean-square deviations (RMSDs) between configurations were similar, were grouped together to simplify them and the analysis of different modes of motion was performed using principle component analysis (PCA). The 3DNA software package[83] was used to analyse hydrogen bonding patterns. Three major motions and conformational states were observed and are defined as searching, fraying and peeling. Searching is the incremental swapping of WC HBs, fraying the loss of WC hydrogen bonding at the ends of the duplex and peeling the process of bases forming HBs with the backbone. They proposed the following schematic to describe the unfolding process:



A more recent study from Perez and Orozco[105] used MD to look at the unwinding of the Dickerson's dodecamer, d(CGCGAATTCGCG).[106] Melting was induced using the chemical denaturant pyridine and heating to 373 K. The Ptraj module of AMBER with default geometry restraints was used to analyze the bonding array. The analysis revealed transient states corresponding to the partially unfolded

conformations. It was concluded that unfolding occurred via a "fraying peeling" mechanism whereby the ends of the double helix become disrupted, thus enabling opened bases to move towards the centre of the duplex via the major grooves. Repeated fraying-peeling is thought to be the driving force for conversion to the unfolded state.

3.4.7. Role of Water in Melting

Previous work has shown that base pair HBs can be mediated by water molecules. The water mediated interactions were generally found to be strong with inserted water molecules being integral to the base pairings.[107] By looking at the geometry of the base pair and adding increasing numbers of water molecules, the populations of stacked versus hydrogen bonded bases were compared.[108] It was observed that for isolated gas-phase AT base pairs a small number of explicit water molecules can cause hydrogen bonded base pairs to convert to stacked base pairs. There has also been a study looking at the solvent environment around RNA during melting using MD[109] but to date there appears to be no detailed MD studies on the effects of solvent on the melting mechanism.

4. Computational Details

4.1. Development of Initial Model Structures

4.1.1. Nucleic Acid Builder

Single bases, base pairs and DNA polymers of 2, 3, 4, 6, 8 and 16 base pairs in length were simulated using differing compositions of WC AT and GC base pairs. Initial DNA structures were created using the Nucleic Acid Builder (NAB)[110] module of AmberTools[111] and the *fd_helix* routine to create files in protein data bank (pdb) format. DNA was modelled in the Arnott B-DNA conformation because B-DNA is the most commonly occurring and biologically important.[112] Labelling of the bases was as follows: DA, DA3', DA5', DAN, DT, DT3', DT5', DTN, DG, DG3', DG5', DGN, DC, DC3', DC5', DCN. Where A represents an adenine base, T thymine, G guanine and C cytosine. The D prefix indicates a deoxyribose base and the 5' and 3' suffixes refer to 5' and 3' terminal residues respectively. The 5' and 3' ends of the helix are capped by OH groups. The labels without a suffix refer to non terminal residues and these contain leading phosphate groups. The N suffix signifies neutral residues capped by OH groups; these were implemented for the single base simulations, see Figure 4.1.

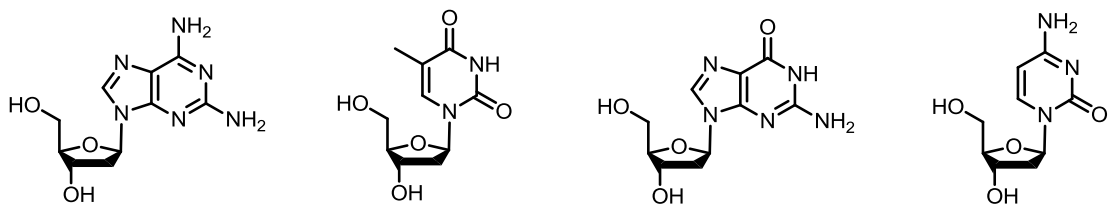


Figure 4.1. Charge neutral base structures from left to right: adenine (DAN), thymine (DTN), guanine (DGN) and cytosine (DCN).

4.1.2. LEaP

The structures were loaded into the LEaP module of AmberTools individually. LEaP stands for Link, Edit and Parm and was used to read in the topology, coordinate and force field information. Files required to run molecular mechanics calculations were then created. The default AMBER force field is the Cornell *et al.*[113] (FF94), but the FF99bsc0[114] is recommended for nucleic acid simulations so parameters from the FF99bsc0 force field were loaded. FF99bsc0 has improved psi-psi angle parameters; these represent the backbone torsional term and lead to distortions in the α/γ term. Where α is the O-P-O-C dihedral of the phosphate group and γ is the O-C-C-C between the sugar ring and the phosphate group.

The pdb files created in NAB were loaded and explicit net neutralising conditions were created by the addition of sodium ions. Single bases required no sodium ions as the bases were already in a charge-neutral state. For solvation an explicit solvent model was chosen over an implicit model. Implicit solvation expresses the solvent as a continuum and behaviour is averaged. Explicit solvation models water molecules separately and hence is more accurate but more costly. Comparative studies of five different explicit water models, namely SPC, SPC/E, TIP3P, TIP4P and TIP4P-Ew have shown that the most effective for biological simulations is the SPC/E model.[115,116] Therefore the SPC/E water model was implemented with a cut-off of 8.0 Å. Previous work has showed that the experimental radial distribution functions doesn't show any features beyond this distance.[117] The next step was choosing the solvation shape. Due to the cylindrical shape of DNA, cubic solvation adds unnecessary water molecules and so computational cost. Rectangular solvation can be problematic as solute rotation can lead to the longest dimension of DNA

lying across the smallest dimension of the box. A truncated octahedron shape was selected for the solvation as this seemed the least problematic and cost efficient. Table 4.1 shows the composition of the DNA created using NAB and solvation and charge data added in LEaP for homo-oligomeric strands produced.

	Composition or Sequence	No. Na+ Ions Added	No. Water Molecules Added	DNA Conc (mM)	Na+ Conc (M)
DAN	Single A base	0	457	92.13	0.0000
DTN	Single T Base	0	442	93.24	0.0000
DGN	Single G Base	0	724	61.41	0.0000
DCN	Single C Base	0	439	94.21	0.0000
AT	A	1	1240	36.17	0.0362
GC	G	1	1232	36.55	0.0366
$d(A_2)(T_2)$	AA	2	1417	31.49	0.0630
$d(G_2)(C_2)$	GG	2	1524	29.58	0.0592
$d(A_3)(T_3)$	AAA	4	1381	31.36	0.1250
$d(G_3)(C_3)$	GGG	4	1706	30.83	0.1230
$d(A_4)(T_4)$	AAAA	6	1669	25.91	0.1550
$d(G_4)(C_4)$	GGGG	6	2021	21.7	0.1300
$d(A_5)(T_5)$	AAAAA	8	1472	27.8	0.2220
$d(G_5)(C_5)$	GGGGG	8	1933	22.05	0.1760
$d(A_8)(T_8)$	AAAAAAAA	14	3055	14.25	0.1990
$d(G_8)(C_8)$	GGGGGGGG	14	2714	15.78	0.2210
$d(A_{16})(T_{16})$	AAAAAAAAAAAAAAAA	30	9289	4.98	0.1490
$d(G_{16})(C_{16})$	GGGGGGGGGGGGGGGG	30	9136	5.06	0.1520

Table 4.1. Data for structures created using NAB and LEaP with the FF99bsc0 force field, the SPC/E explicit water model and explicit net neutralising conditions.

For the 2-mers and 4-mers structures for all of the different possible base combinations were created with the following sequences:

2-mers: at, gc, ga and gt.

4-mers: aaat, aaag, aaac, aata, aaga, aaca, ggga, gggt, gggc, ggag, ggtg, ggcg, atta, atat, aatt, gccg, gcgc, ggcc, aagg, aacc, aagc, aacg.

For each of the structures the pdb, amber topology (prmtop) and coordinate (inpcrd) files were saved.

4.2. Energy Minimisation

4.2.1. Aim of Minimisation

Prior to running MD, structures were minimized using the SANDER (Stimulated Annealing with NMR-Derived Energy Restraints) module of AMBER. SANDER can be used for many types of simulation unrelated to NMR refinement including energy minimisation and MD. Energy minimisation can be used to find the system's stable states represented by minima on the potential energy surface. Here minimisation was used to remove the system's largest strains and not to reach the global minimum. Newly added water molecules have not adjusted to their environment prior to minimization. The presence of sodium ions and solute atoms could cause the equilibrated waters to induce gaps between solute, solvent or the box edges. This can cause major instabilities in subsequent MD calculations. Therefore the minimisation process was carried out in two steps; the first allows the solvent molecules and ions to reach equilibrium and the second equilibrates the whole system.

4.2.2. First Minimisation Stage

For the initial minimisation stage PME with an 8 Å cut-off was implemented for the treatment of electrostatics. Constant volume periodic boundaries were also implemented. To allow only the solvent molecules to equilibrate positional restraints were applied to the DNA atoms. The atoms were restrained to their starting positions using a harmonic potential which has the general form:

$$F = -kx$$

x is displacement of the atom from the original position, F is the restoring force and k is the force constant. By utilising a large force constant, in this case 500 kcal mol^{-1} A^{-2}, the restoring force becomes large and so there is little deviation from the original positions.

The initial minimization procedure was 1000 steps in total comprising of 500 steps of the steepest decents algorithm followed by 500 steps of the conjugate gradients algorithm. The steepest decents algorithm was used first because it is the best algorithm for quickly removing a system's largest strains. However, it tends to converge slowly to the minima and so it was necessary to switch to the conjugate gradients method after a period. The input coordinates were those produced using NAB.

4.2.3. Second Minimisation Stage

The second minimisation stage was to minimise of the whole system and so position restraints on the DNA atoms were removed. In total 2,500 steps of minimisation were run during the second minimisation stage. This was on the generous side but because minimisation is much quicker than MD it is better to remove as much instability as possible before carrying out MD. The 2,500 steps comprised of an initial 1000 steps of steepest decents minimisation followed by 1,500 steps of conjugate gradients minimisation. The input coordinates were the final coordinates obtained from the first stage of the minimisation procedure.

4.2.4. Checking Minimisation

Minimization was checked by comparison of the minimized structures with the original structures. The original and minimized structures were aligned and the root mean squared standard deviation (RMSD) between the corresponding backbone atoms was calculated using visual molecular dynamics software (VMD)[118] by the following formula:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{i=N}\delta_i^2}$$

N is the total number of backbone atom pairs and δ is the distance between each of the pairs.

The aligned original (green) and minimised (blue) structures are compared in Figure 4.2 for the homo-oligomeric 8-mers d(A_8)d(T_8) and d(G_8)d(C_8).

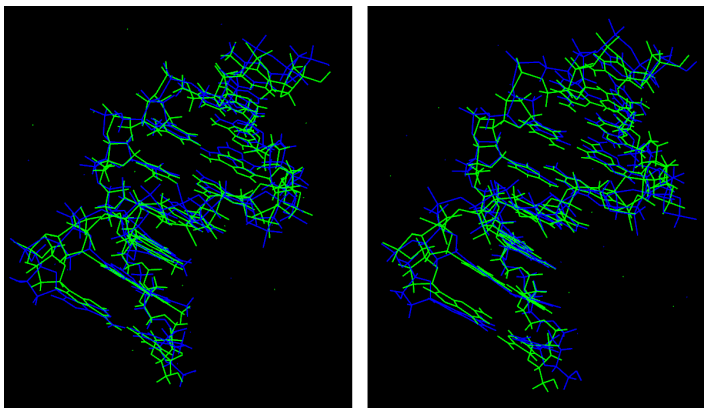


Figure 4.2. Comparison of original (green) and minimised (blue) structures for d(A_8)d(T_8), left and d(G_8)d(C_8), right. Solvent molecules are omitted for clarify and the DNA is represented in stick form.

The calculated RMSD values for the all the homo-oligomeric DNA strands, single bases and base pairs can be seen in Table 4.2.

	Composition or Sequence	Minimization RMSD (Å)
DAN	Single A base	0.135
DTN	Single T Base	0.193
DGN	Single G Base	0.144
DCN	Single C Base	0.153
AT	A	0.175
GC	G	0.39
$d(A_2)(T_2)$	AA	0.586
$d(G_2)(C_2)$	GG	0.446
$d(A_3)(T_3)$	AAA	1.07
$d(G_3)(C_3)$	GGG	0.972
$d(A_4)(T_4)$	AAAA	0.824
$d(G_4)(C_4)$	GGGG	1.037
$d(A_5)(T_5)$	AAAAA	1.027
$d(G_5)(C_5)$	GGGGG	0.829
$d(A_8)(T_8)$	AAAAAAAA	1.078
$d(G_8)(C_8)$	GGGGGGGG	1.121
$d(A_{16})(T_{16})$	AAAAAAAAAAAAAAAA	1.012
$d(G_{16})(C_{16})$	GGGGGGGGGGGGGGGG	1.035

Table 4.2. Minimisation RMSD values for homo-oligomeric DNA strands, single bases and base pairs.

The larger the RMSD value the more the DNA backbone atoms have moved from their original positions. Values are much lower for single bases and base pairs than they are for longer DNA strands. This is as expected as the smaller systems contain fewer torsions and unfavourable atom interactions which need alleviating. There does not appear to be a trend as to whether the AT or GC base pairs lead to higher RMSD values during minimisation.

4.3. Molecular Dynamics

4.3.1. Purpose of MD

MD was run so that the time-dependent nature of the DNA melting mechanism could be examined at the molecular level. To induce melting it was necessary to put energy into the system. This energy was provided by heating to 373 K. MD was carried out in two stages the same as minimisation.

4.3.2. First Stage of MD

During the initial MD phase the system was heated from 0 K to 298 K. To avoid any large fluctuations position restraints were imposed on the DNA atoms. A force constant value of 10 kcal mol^{-1} $Å^{-2}$ was used; much weaker than the force constant used during minimisation. Additionally the system's hydrogen atoms were constrained using the SHAKE algorithm. The system temperature was set using the Langevin temperature equilibrium scheme; a more reliable scheme than the commonly used Berendsen thermostat. The collision frequency used for the Langevin scheme was $1.0ps^{-1}$.

Input coordinates for this stage were the final coordinates obtained from the second minimisation stage. Initial velocities were generated at random from a Boltzmann distribution. The initial temperature of the system, 0 K was calculated from the kinetic energy. The system was quickly heated to 298 K and maintained at this temperature by adjustment of the velocities. The initial MD stage was run for 20 ps comprising of 1000 total steps with a 2 fs time. The short 2 fs time step was selected

in order to maintain a good level of accuracy and stability for the molecular dynamics calculations. Keeping the time step small means calculations are preformed frequently reducing the chance of major fluctuations in the data. A smaller time step would have unnecessarily increased computational time.

Data was written to the output file every 100 steps (200 fs), to the trajectory file also every 100 steps (200 fs) and to the restart file every 1000 steps (2 ps). The 200 fs trajectory writing frequency was selected as a compromise. The fast exchange rate between HBs meant that selecting too large a time step could have led to by-passing important switches contributing towards the melting mechanism. The half life of HBs between water molecules is around 3 ps and using a 200 fs writing frequency was used so that the breaking and formation of water hydrogen bonds could be monitored. Given the length of the MD being run it was important to keep file sizes as small as possible. Hence writing to the output, trajectory and restart files was not more frequent.

Output files contain state information and are formatted so that the information can be easily understood. Trajectory files contain all the coordinate sets over the whole trajectory and so require some post-processing. The restart files contain coordinate, velocity and box information so can be used if the simulation terminates unexpectedly and needs restarting.

4.3.3. Second Stage of MD (Equilibrium Phase)

For the equilibrium phase, the temperature was maintained at 298 K. The weak restraints on the DNA atoms were removed so that the DNA was allowed to equilibrate. The initial coordinates used were the final coordinates produced from

the initial MD stage. Velocity and box information generated from the initial MD stage was also input. Constant volume periodic boundaries were no longer used. To allow the water molecules to relax constant pressure periodic boundary conditions were used. Isotropic position scaling was used to maintain atmospheric pressure (101,325 Pa) with a relaxation time of 2 ps.

It was necessary to implement the use of the iwrap command. This causes the coordinates in the restart files and trajectory files to be returned back to the primary box. The use of iwrap maintains the solvent shape and so makes the trajectories look better when visualizing but it also prevents the numbers in the restart files from getting too large and causing SANDER to terminate. Iwrap has no effect on energy or force but can effect diffusion though that was not an issue for these simulations.

Due to limitations on computational time, cost and resources it was not possible to carry out long enough MD simulations to allow all the aforementioned different sequences and lengths to unwind. Therefore for the purposes of examining the melting mechanism only the homo-oligomeric 4-mers and 8-mers were looked at. For each structure; $d(A_4)d(T_4)$, $d(G_4)d(C_4)$, $d(A_8)d(T_8)$ and $d(G_8)d(C_8)$, eleven different equilibrium times were used. This created eleven different starting structures for each of $d(A_4)d(T_4)$, $d(G_4)d(C_4)$, $d(A_8)d(T_8)$ and $d(G_8)d(C_8)$ for the heating phase. The eleven equilibrium (eq.) times used were 0 ns, 0.1 ns, 0.2 ns, 0.3 ns, 0.4 ns, 0.5 ns, 0.6 ns, 0.7 ns, 0.8 ns, 0.9 ns and 1ns. The structures will be identified by their equilibrium time length for the rest of this paper such as $d(A_4)d(T_4)$ 0 ns eq. and $d(G_4)d(C_4)$ 0.1ns eq. Etc. Having a variety of equilibrium times also enabled the determination of any correlation between equilibrium time length and melting time.

4.3.4. Third Stage of MD (Heating Phase)

For the third stage of MD the temperature was quickly raised from 298 K to 373 K. Initial coordinates, velocity and box information was input from the equilibrium MD phase. All other parameters remained the same as the second MD stage.

For the eleven heating simulations carried out for $d(A_4)d(T_4)$, a heating length of 15 ns was adequate for all the double helices to unwind. The $d(G_4)d(C_4)$ simulations had not unwound after 15 ns and so the simulations were extended to 45 ns. $d(A_8)d(T_8)$ required an even longer 50 ns and the $d(G_8)d(C_8)$ simulations were extended to 60 ns but were not completely unwound at this point.

4.3.5. Control MD Simulations

To check the stability of the structures they were each simulated at room temperature for 5 ns. Simulations of single bases; A, T, G and C, were carried out for 5 ns at both 298 K and 373 K in order to establish the individual base stability. Single base pair simulations were also carried out for AT and GC for 5 ns at 298 K and 373 K to give an idea of how the bases acted independently without the presence of the DNA backbone.

4.3.6. Restrained MD Simulations

In addition to the 8-mer simulations already mentioned, some MD simulations were carried out on $d(A_8)d(T_8)$ and $d(G_8)d(C_8)$ using position restraints. The purpose of these simulations was to establish how the melting mechanism would proceed if the

terminal base pairs could not separate. Previous studies[104,105] have shown fraying to be a key step in the unfolding mechanism and so it was thought that restraining the ends might make other processes involved clearer.

To prevent the ends of the double helix from fraying, Watson and Crick harmonic constraints were placed on the terminal bases. The harmonic function acts as though a spring has been placed between the WC atom pairs. Three distance restraints were placed on the GC base pairs and two on the AT pairs, see Figure 4.3 for atom numbering.

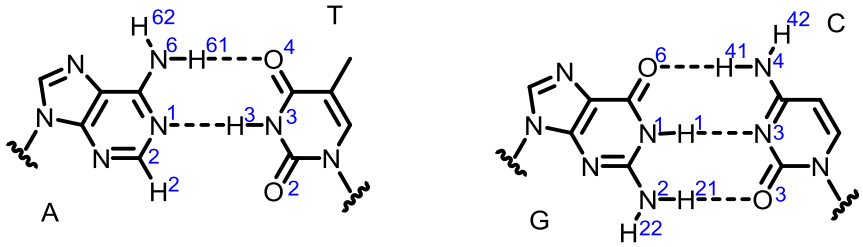


Figure 4.3. Atom labelling of the AT (left) and GC (right) base pairs.

Restraints were placed between the hydrogen atom and donor atom involved in WC HBs. From the topology file the equilibrium distance between these atoms were found and the lower bound distance was set at 0.10 Å below this and the upper bound distance restraint 0.10 Å above. These restraints values can be seen in Table 4.3 and the atom Figure 4.3 shows the atom labelling.

Donor Atom	Hydrogen Atom	Lower Bound	Upper Bound
ADE N1	THY H3	1.87	2.07
THY O4	ADE H61	1.77	1.97
GUA O6	CYT H41	1.84	2.04
CYT N3	GUA H1	1.85	2.05
CYT O2	GUA H21	1.72	1.92

Table 4.3. WC upper bound and lower bound distance restraints. Abbreviations: adenine (ADE), thymine (THY), guanine (GUA) and cytosine (CYT).

Values for the harmonic force constants, rk2 and rk3 were 10 kcal mol^{-1} $Å^{-2}$. This created a flat bottomed harmonic potential whereby the energy penalty increased exponentially before and after the lower bound and upper bound restraints. These restrained simulations were run for 10 ns each. These simulations showed little change over the 10 ns heating trajectory and due to limited computational time they were considered no further.

5. Results and Analysis

5.1. Analysis of Output Files

A Perl script created by the makers of AMBER[75] was used to extract the changes in certain properties with time. Extracted properties were temperature, pressure, density, volume, kinetic energy, potential energy and total energy. This data is already contained in the output file and so running the script simply makes the data easier to plot. Once the initial DNA restraints are removed and the system's temperature brought up to 298K these properties remain constant. Once the system is heated to 373K they change but again stabilise.

The temperature remains constant once the system is brought to 373 K. This shows the Langevin temperature equilibrium scheme to be effective. No density, pressure or volume information is provided for the first 20 ps of MD as this is when constant volume periodic boundary conditions are in place. Density of the system drops below that of water once the system is heated to 373 K. The pressure, volume, kinetic energy, potential energy and total energies all increase with heating to 373 K. Figure 5.1 shows the kinetic energy, potential energy and total energies for $d(A_4)d(T_4)$ 0 ns eq. for the first 20 ps with DNA restraints and the first 5 ns of heating.

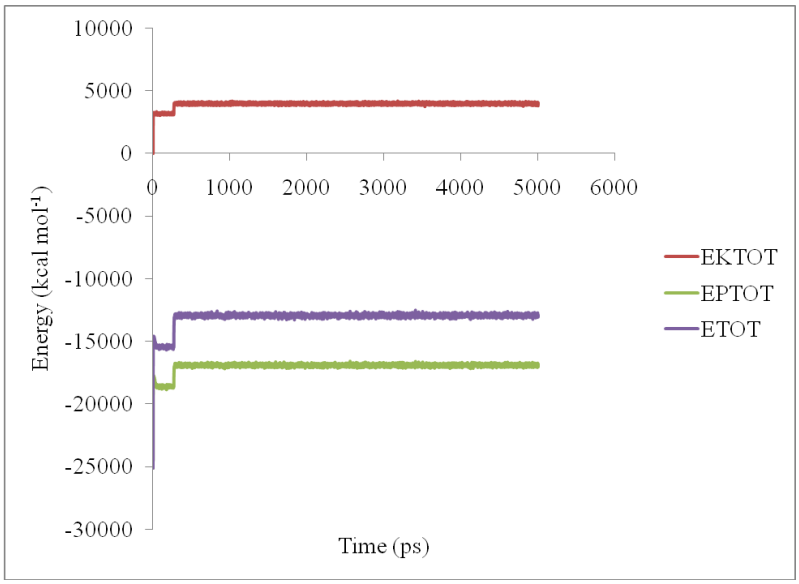


Figure 5.1. The kinetic (EKTOT), potential (EPTOT) and total (ETOT) energies for the first 5 ns of the d(A_4)d(T_4) 0 ns eq. trajectory. The energy is measured in kcal mol^{-1}.

For this particular trajectory there was no equilibrium phase (0 ns eq.) and the temperature was set to 373 K after 20ps when the initial restraints were removed. It can be seen that the energy increases rapidly upon heating but then stabilises for the remainder of the trajectory. The equivalent trajectory for d(G_4)d(C_4) has slightly lower energy values, indicating the increased stability of the GC base pair. This analysis was used as a check that no excessive fluctuations had occurred and so is not considered in any more detail.

5.2. Reimaging

Implementation of the iwrap command was to avoid post processing of the trajectory data but after visualising in VMD it was clear that some reimaging would still be required. Iwrap kept the solvent molecules ordered and in truncated octahedral

formation but the DNA did not remain centred in the truncated octahedron. Some of the DNA split instantaneously with half of the double helix appearing at the edge of the box.

In order to correct the imaging issues Ptraj was used and a script created to format the trajectory files. Firstly one of the DNA strands was centred by mass origin; for 4-mers this meant centring on bases 1-4 and for 8-mers bases 1-8 (bases 5-8 and 9-16 could also have been centred for the same effect). The trajectory was then reimaged using the origin as the centre point. Secondly the whole DNA double helix was centred; bases 1-8 for 4-mers and bases 1-16 for 8-mers. The trajectory was reimaged again using the origin as centre point. This solved the issue the DNA splitting and kept the DNA located in the centre of the truncated octahedron.

5.3. RMSD

Ptraj was used to calculate the change in the backbone RMSD with time. The six main backbone atoms were considered; a phosphate, two oxygens and three carbons, see Figure 5.2.

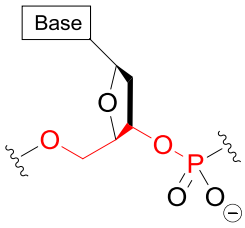


Figure 5.2. Segment of the DNA backbone showing the six atoms in red to which the RMSD fit was applied.

For AT structures the RMSD rose much more quickly; the $d(A_4)d(T_4)$ structures had RMSD values of up to 12 Å in the first 3.4 ns whereas the equivalent $d(G_4)d(C_4)$

structures only reached RMSD values of 7 Å during the same time period (see Figure 5.3).

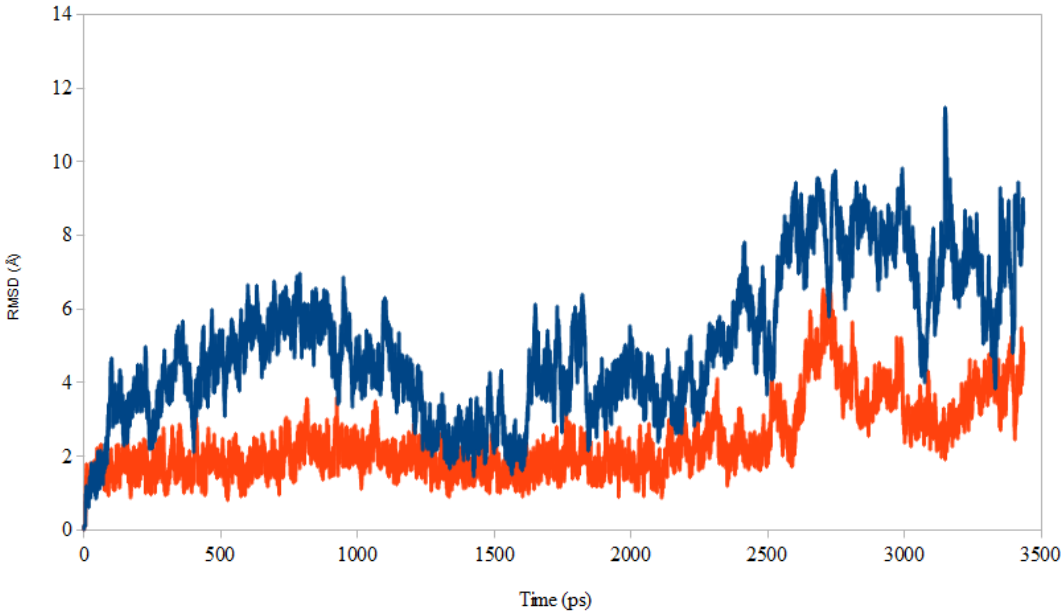


Figure 5.3. RMSD values for the first 3.4 ns of trajectory with no equilibrium phase for d(A_4)d(T_4) and d(G_4)d(C_4) structures.

This reiterates the relevance of the WC HBs; the major difference between these structures is the additional hydrogen bond between the GC pairs. This additional HB makes the GC complex considerably more stable.

5.4. HB Analysis

5.4.1. New HB Definition

To analyse the HBs arrays a new definition was developed which eliminated the requirement for cut-off values. The definition states that a HB is established between a hydrogen atom and a non-covalently bonded acceptor atom when the intermolecular force between these atoms is strongest, and each hydrogen atom may only be hydrogen bonded to one acceptor atom at any point in time. By analysing the hydrogen bonding patterns in this way, arbitrarily defined HBs are avoided and strained HBs are not disqualified. The issue of bifurcated hydrogens, for which hydrogen atoms are assigned to multiple oxygens, is also eliminated. Bifurcated hydrogens can arise when the cut-off selected is too generous.

The new definition means that each hydrogen atom donates to the acceptor atom giving the strongest intermolecular force. The hydrogen switches only if the interaction with a new acceptor leads to a stronger intermolecular force. There is no limitation on the HB and so each hydrogen donates to only one acceptor except at the exact point at which the switch is occurring.

In order to conserve computational time not all hydrogens and all acceptor atoms within the system were analysed. Only the hydrogen atoms on the A, T, G and C bases were selected. Acceptor atoms were limited to the most electronegative, these are base oxygen and nitrogen atoms, the backbone oxygen atoms and the phosphate group oxygen atoms. For adenine this meant three donors and one acceptor: H61, H62, H2, N1 for thymine one donor and two acceptors: H3; O2, O4 for guanine three donors and one acceptor: H1, H21, H22, O6 and for cytosine two donors and

one acceptor: H41, H42, N3 and O2. For atom labelling see Figure 4.3. The backbone oxygen and phosphate group oxygen atoms were also selected and HBs to solvent molecules were included in the search.

5.4.2. Categorisation of HBs

The unformatted output file for this novel analysis method produced a list of all selected hydrogen atoms and the donor atom with which they had the strongest intermolecular force at each time frame. Considering the length of the trajectories and the number of bases in each system this produced a large amount of complex data. The data was reduced by reprinting with omission of the frames over which the HBs remained constant. This produced a list of the frames over which HB switching occurred.

Next the HBs were categorised by looking at the most commonly occurring atom pairs and visualising in VMD. Six major types of HB were identified: intrastrand (I_{intra}), interstrand (I_{inter}), Watson and Crick (WC), interstrand backbone (B_{inter}), intrastrand backbone (B_{intra}) and solvent (S). Intrastrand HBs are those formed between different bases belonging to the same polynucleotide strand. An interstrand interaction encompasses all HBs between nucleotide bases on adjacent strands which are not WC. WC interactions total two for each AT base pair and three for each GC base pair; see Figure 4.3. Intrastrand backbone accounts for HBs formed between a base and backbone atom on the same strand and interstrand backbone for interactions with backbone atoms on the other strand. Solvent included all HBs formed to solvent molecules. An additional classification of none (N) was included in case there were any HB types not included in this list.

To identify which base pairs were forming the HBs a number was used. 0 corresponds to HBs between the originally aligned bases in the topology file. For WC this means base pairing such that base 1 is paired with 8 (BP1-8), 2 with 7 (BP2-7), 3 with 6 (BP3-6) and 4 with 5 (BP4-5). For interstrand backbone HBs this corresponds to the backbone attached to the originally aligned base and for intrastrand backbone interactions the base's own backbone. 1 represents a HB formed with a base that is 1 base away from perfect matching and positive and negative signs were introduced in order to differentiate between the directions. +1WC signifies HBs between bases 2-8, 3-7 and 4-6 as oppose to 1-8, 2-7, 3-6 and 4-5 for 0WC. Intrastrand HBs donating towards the 3' direction are denoted +1 so that for the purine bases 1 donates to 2, 2 to 3 and 3 to 4 and for the pyrimidine bases 5 donates to 6, 6 to7 and 7 to 8. -1WC signifies donating towards the 5' direction in each case. Figure 5.4 shows schematics of each of the possible WC structures which could arise if all HBs were of the same type. In reality many of the structures formed would be intermediate between these. The adenine strands are on the left of each structure.

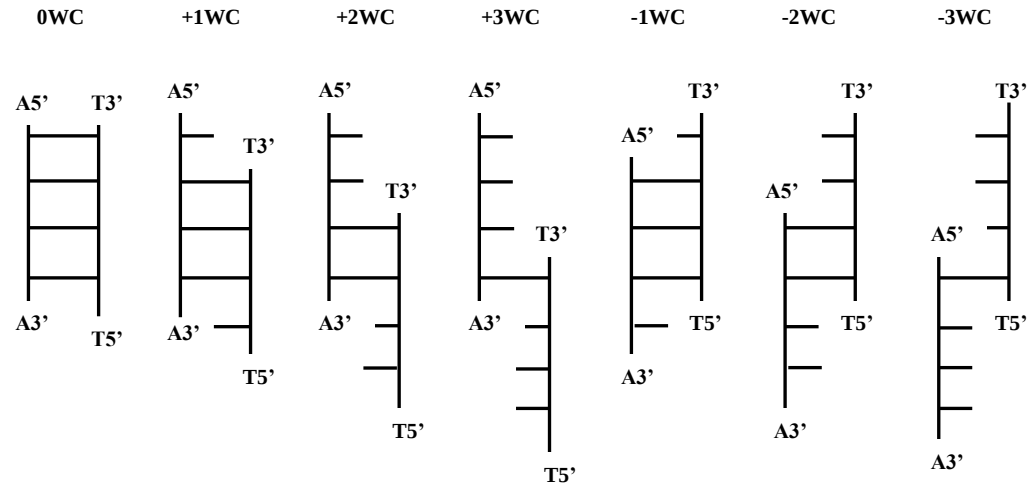


Figure 5.4. Schematics of possible Watson and Crick structures for a 4-mer. Twisting of the double helices are omitted for simplicity. Hydrogen bonded base

pairs are represented by a complete ‘ladder rung’ whereas separated base pairs by half rungs.

Base numbering goes from 1 to 4 for A5’ to A3’ and then from 5 to 8 for T5’ to T3’. For homo-oligomeric GC structures there would be guanine bases in place of the adenine bases and cytosine bases in place of thymine bases. The 8-mers have base numbering going from 1 to 8 A5’ to A3’ and 9 to 16 for T5’ to T3’. Thus there is a total of 14 non-standard (shifted) WC HBs with up to +/-7WC. However not all of these necessarily occurred during each trajectory.

This classification system was simplified because it was found that many of the categories contained very low numbers or zero values. Interactions were condensed into seven major groups; 0WC, NWC, I_{inter}, I_{intra}, B_{inter}, B_{intra} and S. 0WC represents perfect WC base pairings as they appeared in the original topology file. I interactions accounts for all base intrastrand HBs, I_{inter} all base interstrand, B_{inter} and B_{intra} all backbone interstrand and intrastrand HBs respectively.

NWC includes all other WC HBs, also referred to as shifted structures.The shifted structures are the result of what Perez and Orozo refer to as a ‘fraying-peeling’ mechanism.[105] They describe this mechanism as the separation of terminal base pairs followed by movement of these unpaired bases towards the centre of the duplex. They propose that this movement of unpaired bases goes via the major grooves. They also suggest that it is the ‘fraying-peeling’ which eventually causes DNA to unfold.

5.4.3. HB Trajectories Averaged Every ns

The number of each interaction type was averaged every 1000 frames (1 ns in simulation time). The trajectories for the eleven different starting structures used for the heating were compared by considering only the heating part of the trajectory, the last 15 ns of trajectory for $d(A_4)d(T_4)$ structures. This meant ignoring the initial 20 ps with restraints on the DNA along with the variable equilibrium periods. Plotting graphs of the data showed the melting time to be highly variable across the eleven structures. A longer equilibrium period might give a more stable starting structure as the structures have had longer to eliminate unfavourable interactions. As such it was thought that the melting time might increase with the equilibrium time, however this was not always the case. Figure 5.5 shows the number of HBs of each type during the heating phase for four different starting structures of $d(A_4)d(T_4)$.

Initially the 0WC interaction (the original base alignment) is the most commonly occurring type of HB. The 0WC HBs then decline and the solvent-base HBs dominate. After the initial decline of 0WC HBs shifted Watson and Crick HBs (NWC) begin to appear. These NWC interactions seem to delay the increase in solvent-base HBs causing them to plateau slightly. Eventually both the OWC and NWC disappear completely thus indicating complete strand separation has occurred.

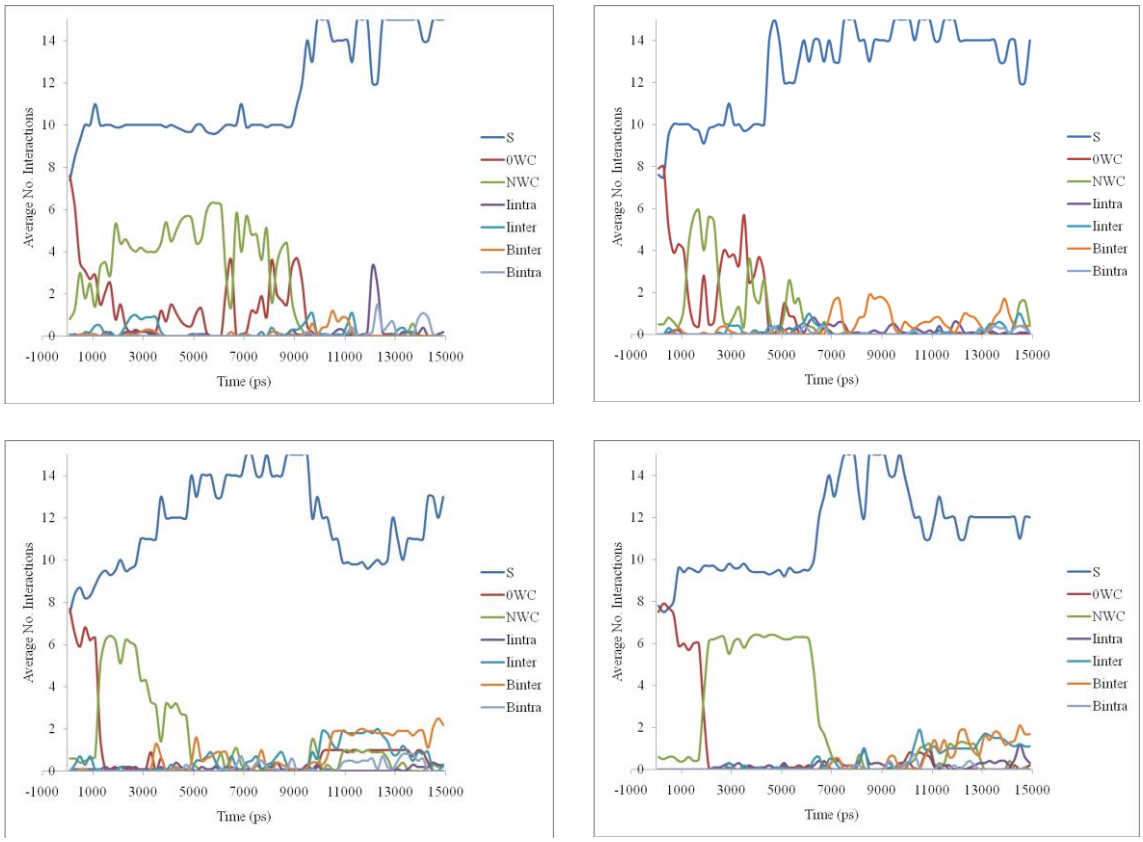


Figure 5.5. Number of S, 0WC, NWC, I_{intra}, I_{inter}, B_{inter} and B_{intra} HBs over 15 ns trajectory for d(A_4)d(T_4) structures. Temperature: 373 K, data averaged every 1 ns. Equilibrium time lengths (temperature: 298K) were as follows; top left 0.1 ns eq. top right 0.3 ns eq. bottom left 0.4 ns eq. and bottom right 0.6 ns eq.

It is the point at which both the 0WC and NWC first both reach 0 that the melting time is defined. Figure 5.6 shows these melting times as a function of equilibrium time for each of the eleven d(A_4)d(T_4) structures.

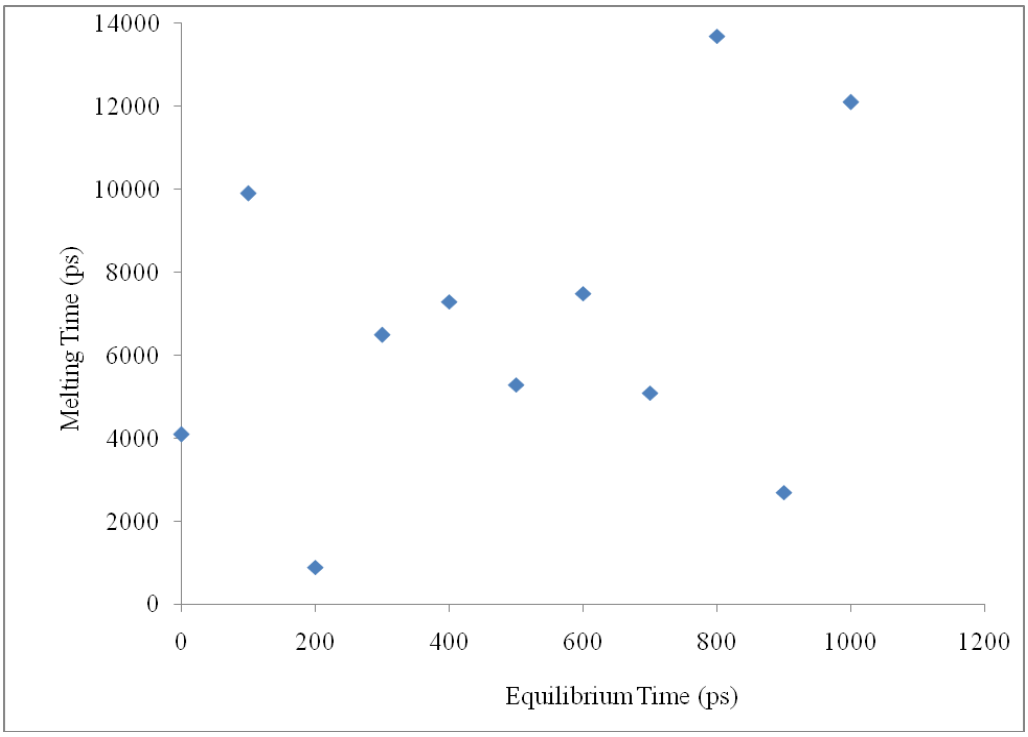


Figure 5.6. Correlation between equilibrium time and melting time for $d(A_4)d(T_4)$ structures.

Correlation between equilibrium length and melting time for the $d(A_4)d(T_4)$ structures produces an R^2 value of 0.133. Hence it is deduced that there is no direct correlation between the equilibrium time length and melting time for these structures.

After melting the I_{inter}, I_{intra}, B_{inter}, B_{intra} and S HBs are all present; the specific numbers of each varies greatly between trajectories. The strands are almost completely separated and so HBs could be the result of random association of the strands and may not have any particular pattern. A plot was created showing the numbers of each HB as an average over the heating phase for the eleven starting structures. This can be seen in Figure 5.7.

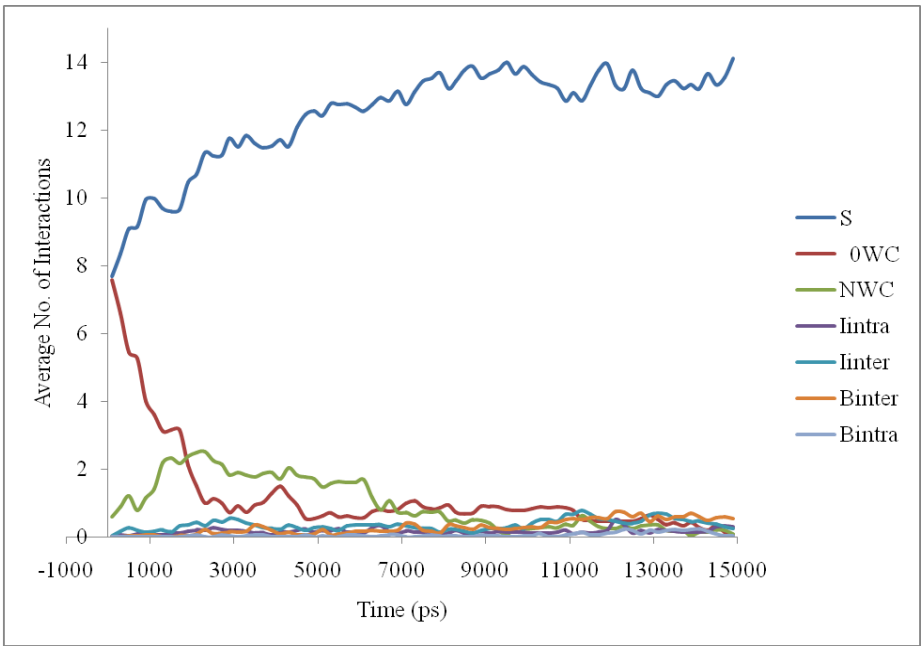


Figure 5.7. Graph showing the average number of each interaction type for the eleven different starting structures of d(A_4)d(T_4) over 15 ns at 373 K.

This produces a smoother, less erratic looking graph. Over the course of the trajectory, solvent HBs increase, 0WC HBs decrease and NWC HBs increase prior to the melting time. The great variety in melting trajectories and the small number of structures considered meant it was not possible to get an accurate value for the melting time. As such it seemed appropriate to look at the melting mechanism in terms of state rather than time.

5.4.4. Removal of Time Dependency

The number of remaining Watson and Crick HBs seemed a good indication as to the stage of unwinding of the double helix. The state is defined by the total number of standard Watson and Crick HBs (0WC) and the total number of shifted Watson and Crick HBs (NWC). The maximum WC state for d(A_4)d(T_4) is 8, for d(G_4)d(C_4) 12,

for d(A_8)d(T_8) 16 and for d(G_8)d(C_8) 24. The minimum in each case is 0. The total numbers of each of the seven interaction types (0WC, NWC, I_{intra}, I_{inter}, B_{intra}, B_{inter} and S) which occurred at each of the WC states was determined. In addition the number of each 1-8, 2-7, 3-6 and 4-5 base pair 0WC HBs occurring at each WC state was monitored.

Information about the type of HB exchanges or switches which occurred was also produced. This was to determine what HB exchanges were possible.

5.5. Mechanism as a Function of WC State

5.5.1. d(A_4)d(T_4)

For each AT base pair formed there are two possible WC HBs; ADE H61-THY 04 and ADE N1-THY H3 (Figure 4.3 for atom labelling). For a 4-mer this means 8 maximum WC HBs. This translates into 9 possible WC states: 0-8. The specific types of WC HBs within these states are considered equivalent. The numbers of each type of HB were normalised by dividing through by the total number of HBs at that state. This gave the numbers as probabilities. Figure 5.8 shows the probability of each interaction type at each WC state.

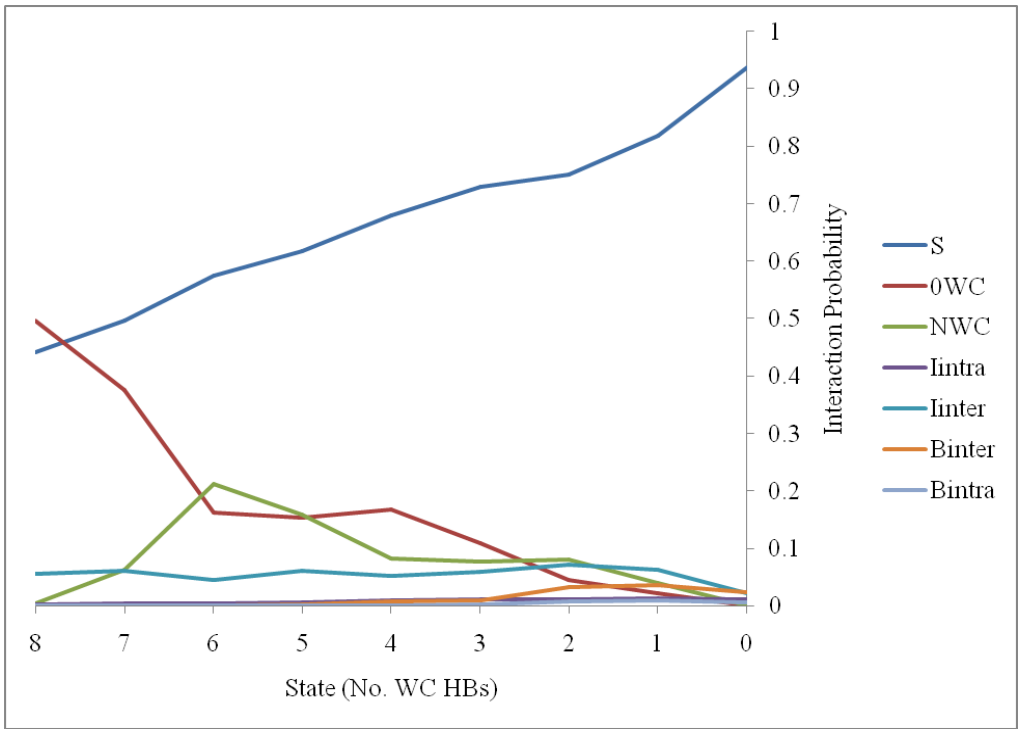


Figure 5.8. A graph showing the probability of each interaction type averaged over 11 trajectories of $d(A_4)d(T_4)$ as a function of state.

As WC state decreases (double helix unwinding) the probability of a solvent HB tends to 1 and the probability of a standard WC HB (0WC) tends to 0. The number of shifted WC (NWC) and standard WC has corresponding peaks and troughs. The specific probability of getting a NWC at any given state varies greatly between the trajectories but is generally at its highest for WC states 6-4. The probability of getting a non WC base pair HB (I_{inter}) remains fairly consistent for WC states 8-2 but then drops for the lowest WC states: 1-0WC, when the solvent probability is nearing 1. The probability of intrastrand backbone HBs (B_{intra}) is negligible for all states for the majority of trajectories. The probability of an interstrand backbone HB (B_{inter}) remains low through WC states 8-3 but then increases slightly for WC states 2-0.

The non WC base pair HBs (I_{inter}) seem to be a result of extreme base non-planarity causing the distance of the WC HBs to become further than other pairings. As the

strands separate these interactions become more likely as the degrees of freedom increase. B_{inter} interactions increase with unwinding because the strands are able to twist more so that bases are no longer facing towards the centre of the helix. For a strand to twist on itself so that a base is facing its own backbone (B_{intra} HBs) there must be a high energy barrier, represented by the low probability of B_{intra} HBs throughout the trajectory.

By looking at the HBs which contribute to the NWC term in more detail it is clear that the +/-1WC structures make up most of this total. For the averaged trajectory these two interactions combined are responsible for over 95% of the shifted HBs. This is not altogether surprising because as the shift number increases from 1 the number of WC HBs holding the strands together is decreasing. The 1WC shifted structures have a maximum of 6 WC HBs whereas the 2WC shifted structures have a maximum of 4 WC HBs and the 3WC shifted structures only 2WC HBs. The probabilities of each of the possible shifts for each trajectory and for the averaged trajectory can be seen in Table 5.1.

	P(-3WC)	P(-2WC)	P(-1WC)	P(+1WC)	P(+2WC)	P(+3WC)
0 ns eq.	0.007	0.013	0.829	0.127	0.019	0.005
0.1 ns eq.	0.000	0.001	0.003	0.976	0.020	0.000
0.2 ns eq.	0.007	0.155	0.136	0.662	0.015	0.024
0.3 ns eq.	0.003	0.063	0.054	0.827	0.049	0.004
0.4 ns eq.	0.001	0.003	0.015	0.745	0.234	0.003
0.5 ns eq.	0.000	0.006	0.295	0.694	0.005	0.000
0.6 ns eq.	0.003	0.000	0.011	0.851	0.120	0.014
0.7 ns eq.	0.001	0.017	0.453	0.410	0.092	0.027
0.8 ns eq.	0.002	0.017	0.278	0.702	0.001	0.000
0.9 ns eq.	0.121	0.146	0.073	0.644	0.007	0.009
1 ns eq.	0.006	0.013	0.536	0.425	0.018	0.002
Avg	**0.005**	**0.015**	**0.177**	**0.774**	**0.027**	**0.002**

Table 5.1. Probabilities of the six possible shift interactions for $d(A_4)d(T_4)$ for the eleven different starting structures.

From this data it is also clear that the +1WC shift is more favourable than the -1WC shift with an average occurrence of 77.4%. The +1WC shift corresponds to the 5' adenine and thymine being unpaired. However by looking more closely at the individual trajectories it can be seen that this is not always the case. For the 0 ns eq., 0.7 ns eq. and 1 ns eq. trajectories the -1WC shift has the highest probability and for the case of 0ns eq. this probability of the -1WC is considerably higher than +1WC. The analysis of a greater number of trajectories is required before a definite conclusion can be reached on the preference of the +1WC or -1WC shift.

The specifics of the 0WC interaction were examined more closely by looking at the probabilities of 1-8, 2-7, 3-6 and 4-5 base pairings with WC state change. It can be seen that the terminal base pair HBs; 1-8 and 4-5, decrease more rapidly than the central base pair HBs with decreasing WC state. See Figure 5.9.

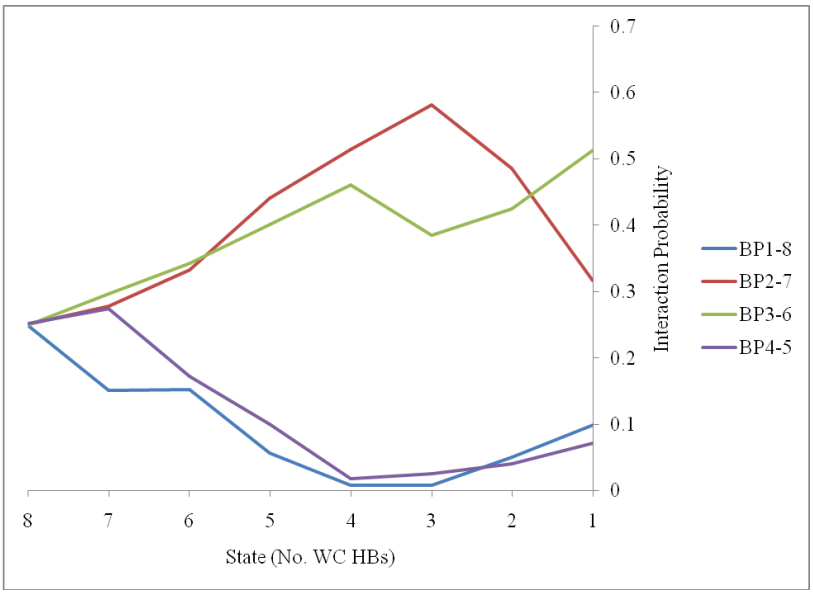


Figure 5.9. The probabilities of HBs formed between the 4 standard base pairings for an averaged trajectory for d(A_4)d(T_4).

When the WC state is 8 the probability of each base pair interaction is 0.25, there is no other choice for this state. Each base pair has 2 HBs and there are 8 HBs in total

and so the probability of getting a HB belonging to a specific base pair is 2/8 = 0.25. As the WC state decreases and the strands begin to separate the probabilities of the terminal base pair interactions decrease in approximately equal amounts. This is a result of the ends of the double helix fraying. At WC states 2 and 1 the probability of the terminal base interactions increases and the probabilities of the central base pair HBs are no longer equal. At this point in the melting the strands are held together by only one WC HB and so there is little structure left to the double helix.

5.5.2. d(G_4)d(C_4)

Each GC base pair can form up to 3 WC HBs; GUA 06-CYT H41, GUA H1-CYT N3 and GUA H21-CYT 03 (see Figure 4.3). For a 4-mer homo-oligomeric GC complex there are 13 possible WC states: 0-12 HBs. The number of each interaction type at each state was normalised to give a probability. An average trajectory was produced for the 11 different equilibrium times (0 ns eq., 0.1ns eq., 0.2 ns., 0.3 ns eq., 0.4 ns eq., 0.5 ns eq., 0.6 ns eq., 0.7 ns eq., 0.8 ns eq., 0.9 ns eq. and 1 ns eq.). Similarly to the d(A_4)d(T_4) mechanism the solvent probability tends to 1 with decreasing WC state as the 0WC probability tends to 0, see Figure 5.10.

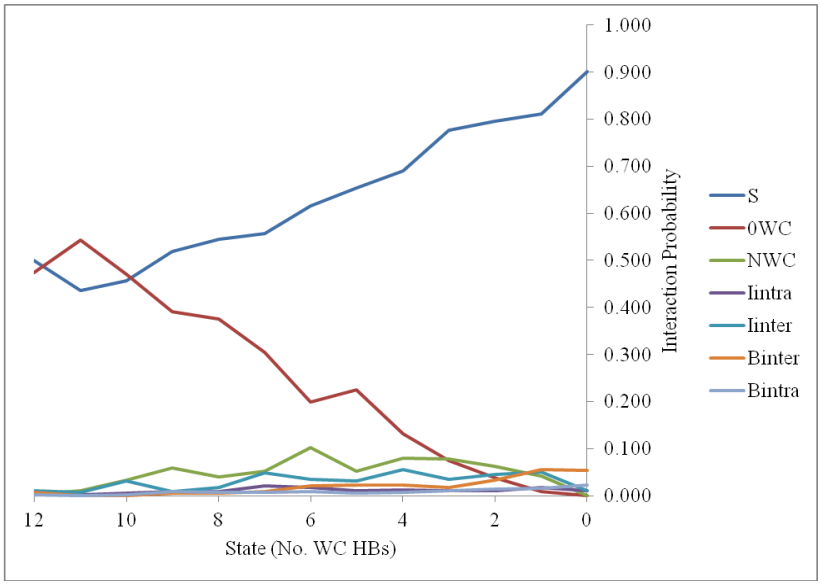


Figure 5.10. The probability of each interaction type averaged over 11 different trajectories for $d(G_4)d(C_4)$ as a function of WC state.

For $d(G_4)d(C_4)$ the shifted interactions (NWC) seem to be slightly less dominant than for $d(A_4)d(T_4)$. For the trajectory average the probability of an NWC HB never became higher than that for a 0WC HB. When the trajectories are considered individually this is not always the case, see Figure 5.11. This indicates multiple possible melting mechanisms are occurring.

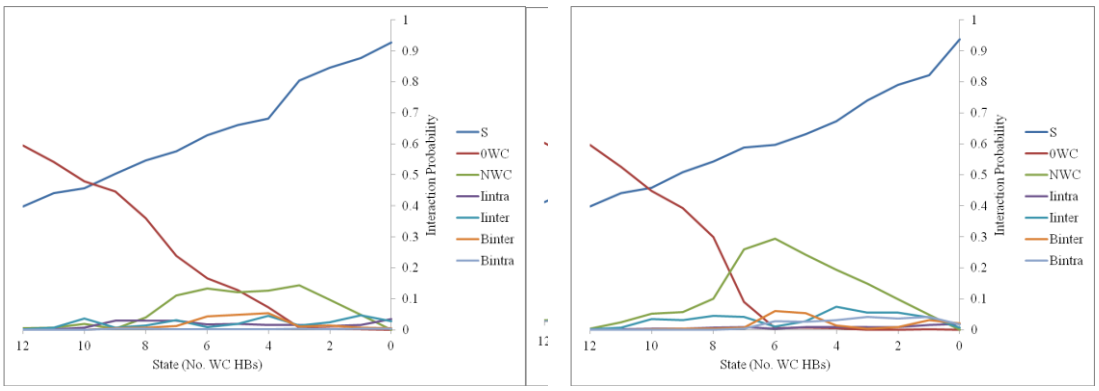


Figure 5.11. The probability of each interaction type for the 0.1 ns eq. trajectory (left) and the 0.8 ns eq. trajectory (right) for $d(G_4)d(C_4)$ as a function of WC state.

The probability of getting a shifted structure peaks after the 0WC disappears around WC states 3 and 6 for the 0.1 ns eq. and 0.8 ns eq. Trajectories respectively. Looking at the individual contributions to the NWC term provides information about the direction and extent of the shifting (see Table 5.2).

	P(-3WC)	P(-2WC)	P(-1WC)	P(+1WC)	P(+2WC)	P(+3WC)
0 ns eq.	0.000	0.031	0.708	0.228	0.032	0.000
0.1 ns eq.	0.001	0.003	0.036	0.851	0.108	0.000
0.2 ns eq.	0.006	0.028	0.626	0.249	0.024	0.067
0.3 ns eq.	0.005	0.000	0.037	0.956	0.002	0.000
0.4 ns eq.	0.000	0.000	0.876	0.071	0.000	0.053
0.5 ns eq.	0.000	0.000	0.428	0.572	0.000	0.000
0.6 ns eq.	0.092	0.597	0.110	0.128	0.069	0.004
0.7 ns eq.	0.006	0.325	0.358	0.275	0.015	0.022
0.8 ns eq.	0.009	0.008	0.016	0.785	0.111	0.071
0.9 ns eq.	0.000	0.004	0.145	0.847	0.003	0.000
1 ns eq.	0.038	0.022	0.111	0.744	0.084	0.000
Avg	**0.014**	**0.093**	**0.314**	**0.519**	**0.041**	**0.020**

Table 5.2. The probabilities of the six possible shift interactions for $d(G_4)d(C_4)$ for eleven different starting structures.

Probabilities of +/-2WC and +/-3WC shifted structures are generally very low. The two exceptions to this are for the 0.6 ns eq. and 0.7 ns eq. trajectories. Here the probability of the -2WC interaction is more substantial. The average shows that the majority of shifted interactions come from 5' shifting (positive shifts). These contribute over 68% of the shifts in total. This correlates with what was seen for $d(A_4)d(T_4)$ though for the $d(A_4)d(T_4)$ trajectories this number is slightly higher with 80% of the shifts coming from 5' shifting.

For $d(G_4)d(C_4)$ there are the same number of possible 0WC base pairings as $d(A_4)d(T_4)$ which is 4, base pairs 1-8, 2-7, 3-6 and 4-5. By looking at the probability of each of these BP interactions at each WC state the order in which they separated was determined. Comparing the averaged trajectories for these BP interactions for

d(A_4)d(T_4) and d(G_4)d(C_4) it can be seen that the order of separation is very similar, see Figure 5.12.

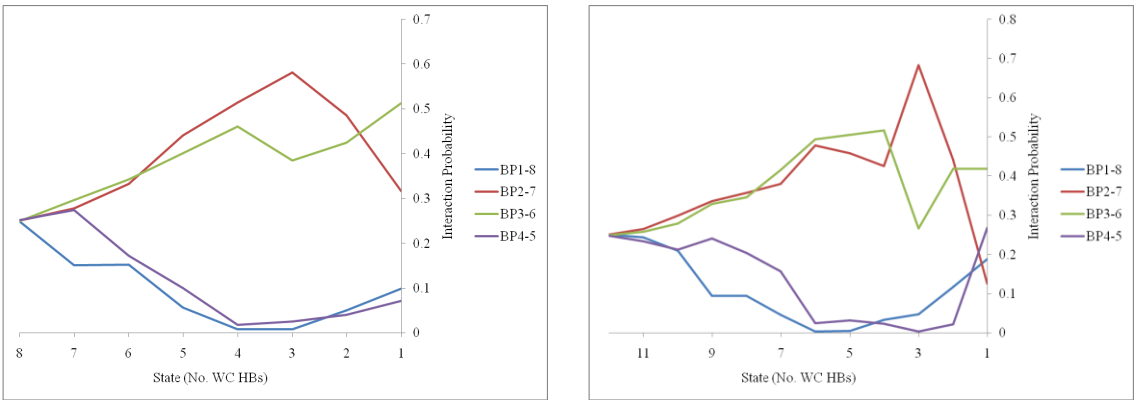


Figure 5.12. The probabilities of a HB between the 4 standard base pairings for an averaged trajectory of d(A_4)d(T_4) , left and d(G_4)d(C_4), right.

In both cases the terminal base pair HBs disappear the most quicky with decreasing WC state. For both d(A_4)d(T_4) and d(G_4)d(C_4) it is BP1-8 HBs which fall off quickest followed by BP4-5 HBs. The order of dissociation between the terminal base pairs fluctuates between trajectories. More trajectories would need to be analysed before there was definite proof that BP1-8 is the first to dissociate. At high WC states, the BP2-7 and BP3-6 HBs fall off at a similar rate. At WC state 3 the probability of BP3-6 HBs drops dramatically for both d(A_4)d(T_4) and d(G_4)d(C_4) structures. This is followed by a decrease in BP2-7 HBs the probability of which stays low through to WC state 1. Here the probability of a BP3-6 interaction becomes the more likely. The BP3-6 HBs are longer lasting than BP2-7 because BP3-6 is next to BP4-5, the longer lasting of the terminal base pairs. Separation of BP4-5 base means more degrees of rotational freedom so the BP2-7 HBs break before the BP3-6 HBs.

5.5.3. d(A_8)d(T_8)

d(A_8)d(T_8) strucutures have 17 possible WC states; 0-16. It was not possible within the given time frame to allow all eleven structures to unfold completely. Within the 50 ns of heating MD at 373K the different d(A_8)d(T_8) strucutres unfolded varying amounts. The 0.4 ns eq. trajectory reached WC state 2, the 0.7 ns eq., 0.8 ns eq. and 1 ns eq. trajectories reached WC state 3 and the 0.5 ns eq. and 0.9 ns eq. reached WC state 4. The other trajectories; 0 ns eq., 0.1 ns eq., 0.3 ns eq. and 0.6 ns eq. all reached WC state 0. All the trajectories were still included when calculating averages. This meant that at the lower WC states the HB probabilities were averaged over fewer trajectories and so were less representative. d(A_8)d(T_8) followed similar unfolding trends to d(A_4)d(T_4) and d(G_4)d(C_4) structures. The average trajectory (Figure 5.13) shows that the solvent probability tends to 1 with decreasing WC state. The probability of 0WC remains higher than solvent until WC state 15 where solvent becomes higher, 0WC then drops lower than 0.1% beyond WC state 5. Here the probability of NWC HBs becomes higher than 0WC HBs. This continues up until WC state 1 where intrastrand and interstrand interactions dominate over 0WC.

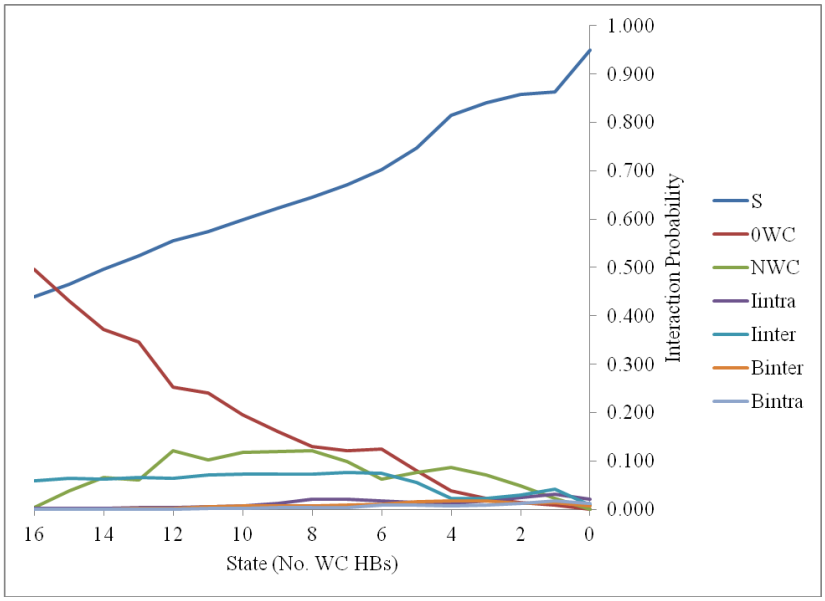


Figure 5.13. The probability of each interaction type averaged over 11 different trajectories for d(A_8)d(T_8) as a function of WC state.

The non-WC interstrand HBs (I_{inter}) remain quite high for the majority of the trajectory. These I_{inter} HBs include ADE H2-THY O2 HB (Figure 4.3) which are not considered WC due to their longer length and hence weaker intermolecular potential. This explains why for both the 4-mer and 8-mer AT simulations the percentage of non WC interstrand interactions seems much higher than for the GC simulations. The most likely HB to be formed at WC state 0, other than solvent, is intrastrand. Here the double helix is completely unwound and the strongest HBs the bases can form are with solvent molecules and in a few cases other bases on the same strand (I_{intra}).

HBs formed with the backbone are also present but not until the WC state becomes extremely low. Backbone HBs (B_{inter} and B_{intra}) seem not to be active in the actual melting mechanism but are sometimes formed once the strands separate. The 0.4 ns eq. trajectory (Figure 5.14) shows that formation of shifted structures can be reversible. Perfect matching of the base pairs (0WC) reoccurs after the NWC peak.

However, the formation of shifted structures does seem to play an important role in the melting mechanism. With all three structures considered ($d(A_4)d(T_4)$, $d(G_4)d(C_4)$ and $d(A_8)d(T_8)$) there is an increase in shifted HBs at mid to low WC states.

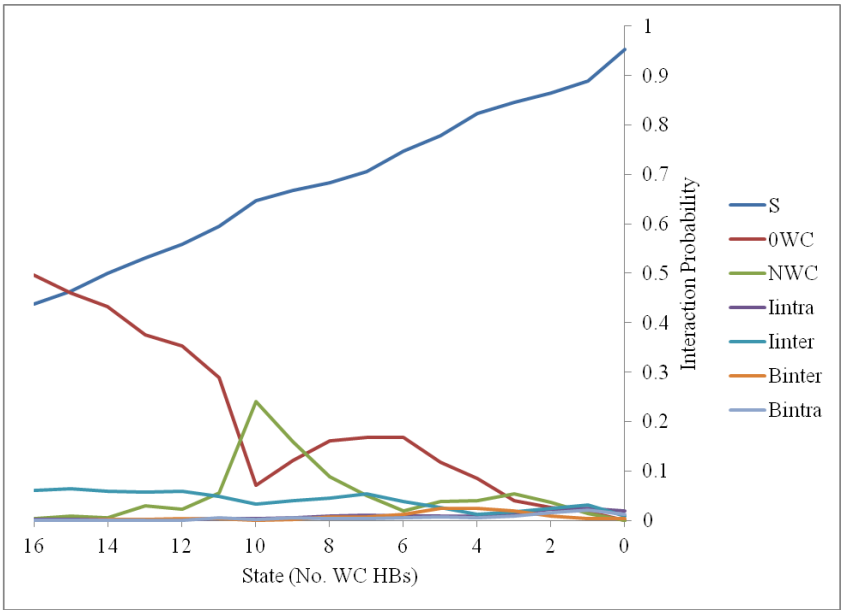


Figure 5.14. The probability of each interaction type given as a function of WC state over the heating phase for a $d(A_8)d(T_8)$ trajectory with 0.1 ns equilibrium time.

For 8-mers there are 14 different possible classifications for the shift HBs. These consist of shifts up to 7 BPs in each direction, +/-7WC. $d(A_4)d(T_4)$ and $d(G_4)d(C_4)$ showed preference for 5' shifts (positive shifts). The $d(A_8)d(T_8)$ trajectories confirm this preference giving nearly 97% of the shifts as 5', see Tables 5.3 and Table 5.4.

	(-7WC)	(-6WC)	(-5WC)	(-4WC)	(-3WC)	(-2WC)	(-1WC)
0 ns eq.	0.000	0.001	0.000	0.000	0.000	0.000	0.007
0.1 ns eq.	0.000	0.000	0.005	0.014	0.040	0.034	0.138
0.2 ns eq.	0.000	0.001	0.000	0.000	0.000	0.000	0.007
0.3 ns eq.	0.000	0.000	0.000	0.000	0.000	0.001	0.002
0.4 ns eq.	0.000	0.000	0.000	0.000	0.000	0.001	0.030
0.5 ns eq.	0.023	0.001	0.005	0.000	0.000	0.000	0.126
0.6 ns eq.	0.000	0.001	0.001	0.006	0.039	0.004	0.024
0.7 ns eq.	0.002	0.002	0.000	0.000	0.000	0.000	0.040
0.8 ns eq.	0.000	0.000	0.000	0.000	0.000	0.000	0.001
0.9 ns eq.	0.000	0.000	0.000	0.000	0.000	0.000	0.038
1 ns eq.	0.000	0.002	0.003	0.007	0.002	0.000	0.085
Avg	**0.001**	**0.000**	**0.001**	**0.001**	**0.003**	**0.002**	**0.022**

Table 5.3. Probabilities of the seven possible negative shift interactions for $d(A_8)d(T_8)$ for the eleven (0-1ns eq.) different trajectories.

	(+1WC)	(+2WC)	(+3WC)	(+4WC)	(+5WC)	(+6WC)	(+7WC)
0 ns eq.	0.912	0.080	0.000	0.000	0.000	0.000	0.000
0.1 ns eq.	0.737	0.029	0.004	0.000	0.000	0.000	0.000
0.2 ns eq.	0.912	0.080	0.000	0.000	0.000	0.000	0.000
0.3 ns eq.	0.605	0.312	0.072	0.007	0.000	0.000	0.000
0.4 ns eq.	0.968	0.000	0.000	0.000	0.000	0.000	0.000
0.5 ns eq.	0.842	0.003	0.000	0.000	0.000	0.000	0.000
0.6 ns eq.	0.865	0.056	0.003	0.000	0.000	0.000	0.000
0.7 ns eq.	0.950	0.006	0.001	0.000	0.000	0.000	0.000
0.8 ns eq.	0.255	0.704	0.041	0.000	0.000	0.000	0.000
0.9 ns eq.	0.961	0.001	0.000	0.000	0.000	0.000	0.000
1 ns eq.	0.885	0.000	0.015	0.000	0.000	0.000	0.000
Avg	**0.663**	**0.280**	**0.025**	**0.001**	**0.000**	**0.000**	**0.000**

Table 5.4. Probabilities of the seven possible positive shift interactions for $d(A_8)d(T_8)$ for the eleven (0-1ns eq.) different trajectories.

The most commonly occurring shift is +1WC accounting for 66%. The +2WC and +3WC shifts both account for more of the shifted total than the -1WC shift.

Standard (0WC) base pairings fall off as previously seen, with the terminal base pair HBs being lost almost completely after WC state 12 (see Figure 5.15).

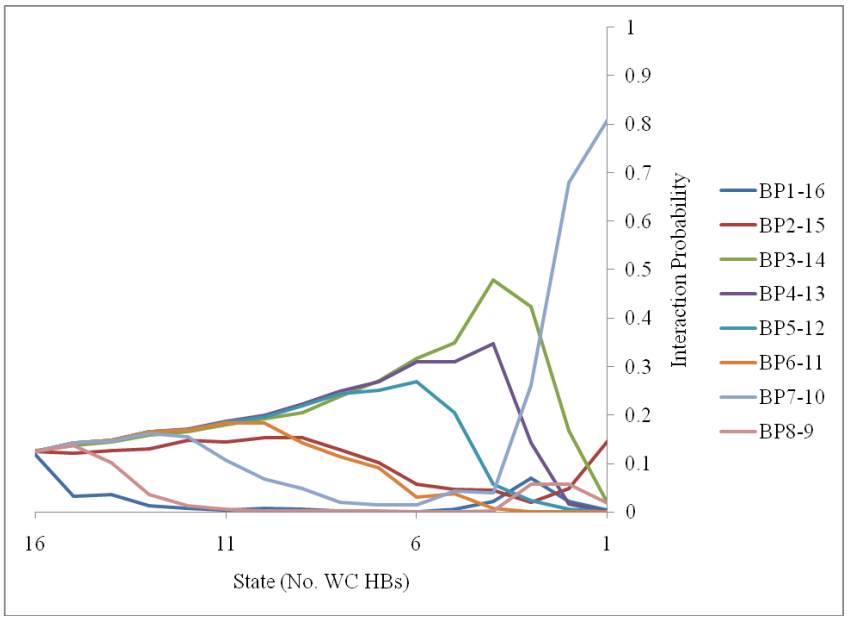


Figure 5.15. Probabilities of the 0WC base pairings for d(A_8)d(T_8) averaged over eleven different trajectories (0-1 ns eq.).

It is BP1-16 which dissociates slightly more quickly out of the two terminal bases. The remaining central BPs separate in differing orders depending on the particular trajectory being looked at.

5.5.4. d(G_8)d(C_8)

d(G_8)d(C_8) structures have 25 possible WC states: 0-24. The MD heating phase was run for 60 ns at 373 K but the strucutures failed to unfold completely within this time frame. Data for the 0.3 eq. trajectory was omitted from the analysis as the full 60 ns heating MD had not completed at the time of writing. The 0.1 ns eq., 0.6 ns eq. and 0.7 ns eq. all reached WC state 8, 0.2 eq., 0.4 eq. and 0.9 ns eq. reached WC state 9, 0 ns eq. and 1 ns eq. WC state 11 and 0.8 ns reached WC state 13.

The trajectories were analysed in the same way as for d(A_4)d(T_4), d(G_4)d(C_4) and d(A_8)d(T_8) and the numbers of each HB type converted to probabilities. The same trends can be seen with decline of standard WC HBs and increase in solvent HBs as WC state decreases. The averaged trajectory (Figure 5.16) shows the presence of some shifted structures.

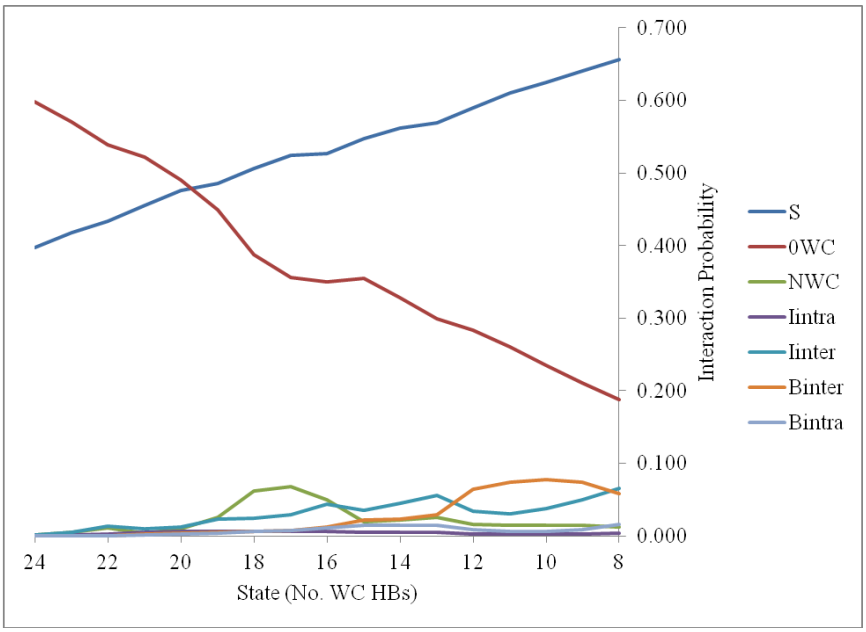


Figure 5.16. The probability of each interaction type averaged over 10 different trajectories for d(G_8)d(C_8) as a function of WC state.

The probability of obtaining a shifted structure (NWC) was much lower than previously seen. Unfolding had not occurred to the same extent here, the lowest WC state being 8. The lower presence of the shifted structures reinforces the key role shifting plays in the melting mechanism. A more detailed breakdown of the shifting HBs can be seen in Tables 5.6 and 5.7.

	(-7WC)	(-6WC)	(-5WC)	(-4WC)	(-3WC)	(-2WC)	(-1WC)
0ns	0.000	0.000	0.000	0.001	0.002	0.008	0.087
0.1ns	0.000	0.000	0.000	0.001	0.001	0.001	0.549
0.2ns	0.000	0.000	0.000	0.000	0.000	0.000	0.524
0.3ns	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.4ns	0.000	0.000	0.011	0.012	0.007	0.000	0.269
0.5ns	0.000	0.000	0.000	0.000	0.000	0.000	0.515
0.6ns	0.000	0.000	0.000	0.000	0.000	0.001	0.149
0.7ns	0.000	0.000	0.000	0.000	0.000	0.000	0.620
0.8ns	0.000	0.000	0.000	0.000	0.000	0.000	0.451
0.9ns	0.000	0.000	0.000	0.000	0.000	0.002	0.595
1ns	0.000	0.000	0.000	0.000	0.000	0.000	0.007
Avg	0.000	0.000	0.000	0.001	0.000	0.001	0.109

Table 5.5. Probabilities of the seven possible negative shift interactions for $d(G_8)d(C_8)$ for ten different trajectories.

	(+1WC)	(+2WC)	(+3WC)	(+4WC)	(+5WC)	(+6WC)	(+7WC)
0ns	0.902	0.000	0.000	0.000	0.000	0.000	0.000
0.1ns	0.436	0.000	0.000	0.001	0.010	0.000	0.000
0.2ns	0.475	0.000	0.000	0.000	0.000	0.000	0.000
0.3ns	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.4ns	0.699	0.001	0.000	0.000	0.000	0.000	0.000
0.5ns	0.482	0.003	0.000	0.000	0.000	0.000	0.000
0.6ns	0.847	0.003	0.000	0.000	0.000	0.000	0.000
0.7ns	0.380	0.000	0.000	0.000	0.000	0.000	0.000
0.8ns	0.549	0.000	0.000	0.000	0.000	0.000	0.000
0.9ns	0.204	0.054	0.135	0.009	0.000	0.000	0.000
1ns	0.991	0.001	0.000	0.000	0.000	0.001	0.000
Avg	0.881	0.003	0.004	0.000	0.000	0.000	0.000

Table 5.6. Probabilities of the seven possible positive shift interactions for $d(G_8)d(C_8)$ for ten different trajectories.

The 5’ (positive) shift dominates, contributing 89 % of the total shifted HBs. This correlates with what was observed for the $d(A_4)d(T_4)$, $d(G_4)d(C_4)$ and $d(A_8)d(T_8)$ structures.

Fraying of terminal base pairs occurs with complete loss of BP1-16 and BP8-9 WC HBs for all trajectories no later than WC state 15. There is little difference in the breakage order of these terminal bases. The longest lasting of the standard base

pairings are the central two bases BP4-13 and BP5-12. The probabilities of these bases remain fairly similar throughout the trajectory. Only the terminal base pair HBs completely disappear during the 60 ns of heating trajectory, see Figure 5.17.

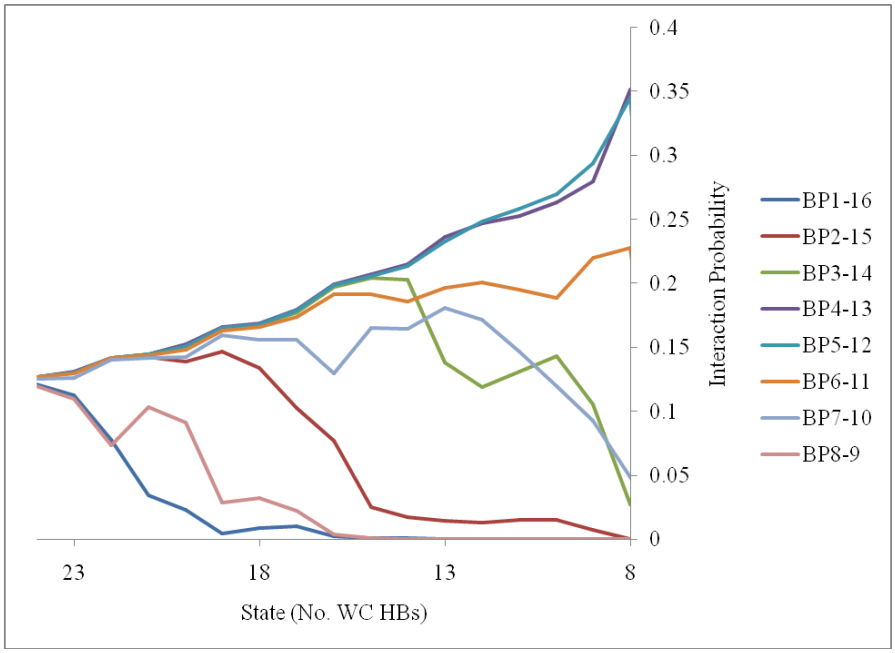
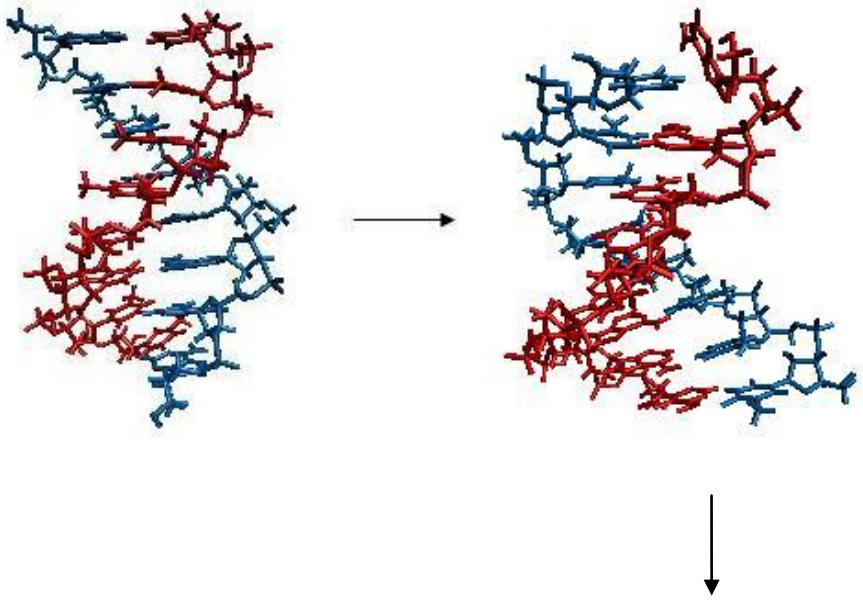


Figure 5.17. Probabilities of the 0WC base pairings for $d(G_8)d(C_8)$ averaged over ten different trajectories.

5.6. Overall Mechanism

There appears not to be an exact sequence of melting events and intermediates formed but a variety of possible routes. From the trajectories analysed it is clear that some of these processes are more dominant than others. As such a generalised mechanism is proposed. The first step of this generalised unfolding procedure is fraying of the terminal base pairs. This means loss of BP1-8 and BP4-5 WC HBs for 4-mers and loss of BP1-16 and BP8-9 WC HBs for 8-mers. This is facilitated by the loss of base pair planarity and temporary formation of non-WC HBs (I_{inter}). For the majority of trajectories analysed fraying occurred at the 5' purine (A or G) end first.

For 4-mers this meant loss of WC HBs between BP1-8 and for 8-mers BP1-16. Once terminal base pair hydrogen bonding is lost the strands shift in the 5' direction (positive WC) and WC base pairing reforms. This is also facilitated by loss of base base planarity and formation of non-WC HBs. The duplex then begins to unfold via a combination of base distortions, 5' shifts and fraying. Figure 5.18 exemplifies this general mechanism using a trajectory of $d(A_8)d(T_8)$ with 0 ns eq. time. The first structure shows complete matching of base pairs and full WC pairing (WC state 16). This snapshot was taken at time 0 ns of the heating phase. The second structure occurred after 3.5 ns of heating and shows the fraying of the terminal base pair at the A5' end. The third structure, at 29.5 ns of heating MD, is a combination of 5' shifting (+1WC, Figure 5.4) and fraying. The structure has an overhang of one unpaired base at each 5' terminus and fraying of one base pair at each end of the duplex. The fourth structure at 33 ns shows the same 5' shifted (+1WC) structure as structure three but the fraying has increased by an additional base pair at the A3' terminus. The final structure at 49.8 ns shows the double helix completely unwound with no WC HBs remaining.



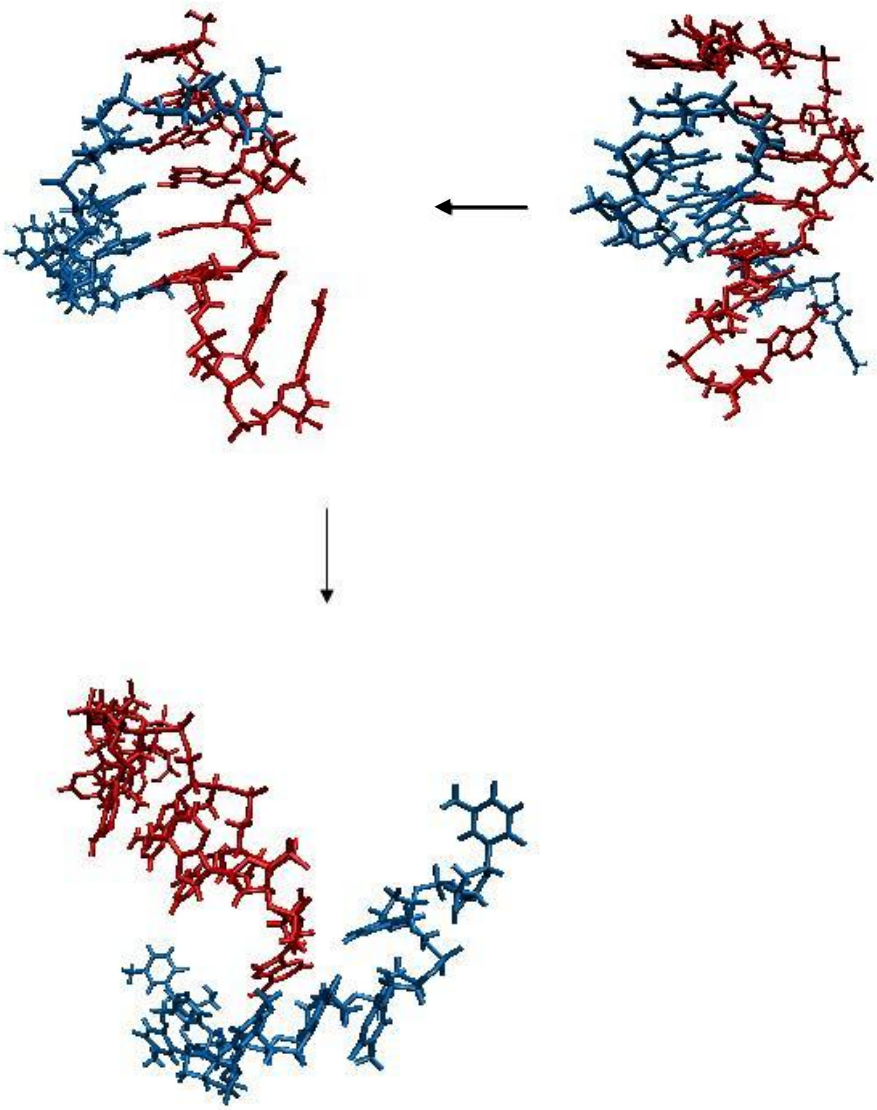


Figure 5.18. Conformations of d(A_8)d(T_8) DNA at different stages of the melting process. The A-strand is represented in red and the T-strand in blue. The 5' end of the A-strand is located at the top of each image.

Figure 5.19 shows the structures from Figure 5.18 represented as a schematic. The fraying and shifting can be more easily visualised without twisting of the double helix. The structures are orientated so that the top left corner is the A5' terminus.

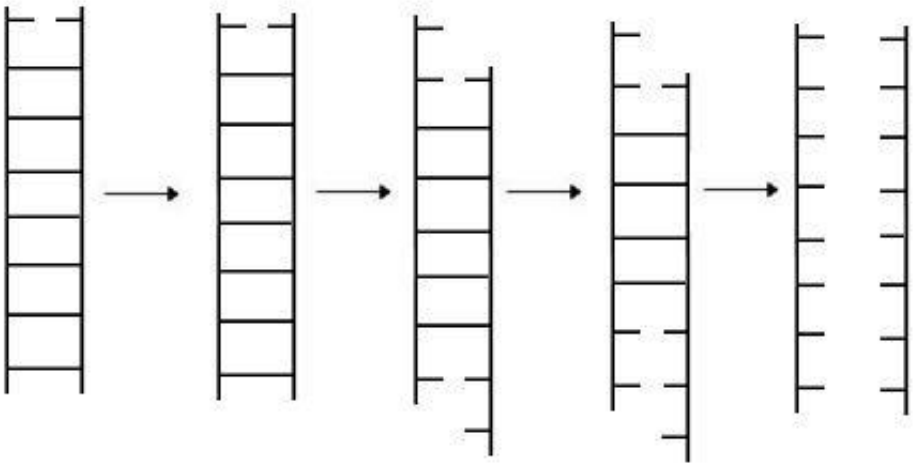


Figure 5.19. Schematic of $d(A_8)d(T_8)$ DNA at different stages of the melting process. The 5’ end of the A-strand is located at the top left corner of each image.

Representing this mechanism as a more generalised scheme gives:

Standard WC Pairing $\rightleftharpoons$ Base Tilting $\rightleftharpoons$ Fraying $\rightleftharpoons$

Base Tilting $\rightleftharpoons$ 5' WC Shifting $\rightleftharpoons$ Strand Separation

Base tilting accounts for loss of base pair planarity and the formation of non-WC HBs. Complete strand separation occurs via repeated fraying and 5’ shifting. The exact sequence of shifting and fraying will differ for different sequences, lengths and structures. Many of the steps will also be reversible.

6. Conclusions and Further Work

The proposed mechanism for the melting of double stranded DNA is based primarily on data for homo-oligomeric 4-mers and 8-mers. Eleven different starting structures were utilised for the heating MD phase in each case. This gave a total of 44 different heating trajectories for analysis. Melting was induced by heating the systems to 373 K for MD time lengths of 15, 45, 50 and 60 ns. The proposed mechanism was considered as a function of state rather than time. State is defined by the number of WC HBs at any given frame.

The most probable route to unfolding is summarised in the following scheme:

Standard WC Pairing $\rightleftharpoons$ Base Tilting $\rightleftharpoons$ Fraying $\rightleftharpoons$

Base Tilting $\rightleftharpoons$ 5' WC Shifting $\rightleftharpoons$ Strand Separation

Standard WC pairing represents the original alignment of the two nucleotide strands with no bases unpaired. Fraying represents loss of WC base pairing at the ends of the duplex. This occurs firstly in the terminal base pairs. Base tilting is the loss of base pair planarity as a result of backbone distortion and also leads to the formation of some non-WC HBs. 5’ shifting (positive shifting, see Figure 5.4) leads to overhang of unpaired bases at the 5’ ends. Shifted HBs are numbered by the number of bases by which the structure has shifted. This number increases during unfolding and this combined with fraying causes the strands to separate. The final state in unfolding is two single DNA strands completely surrounded by solvent molecules. This is just one plausible unfolding pathway. 3’ shifting was seen in many of the trajectories but with much lower probabilities. It is therefore the 5’ shifting which is thought to be the dominant process in the unfolding.

The preference for 5'shifting as opposed to 3' shifting could arise because of incomplete alignment of the backbone strands. When all base pairs are matched this causes the 3' ends to stick out slightly more than the 5' ends. Unpaired bases at the 3' ends would protrude more than unpaired bases at the 5' ends. The preference for 5' shifting may be to minimise protruding of the unpaired bases. This is one possible explanation but there could be any number of other factors which cause shifting to occur in the 5' direction. 5' shifting could be related to the reverse mechanism, folding of the duplex. Close to equilibrium it is possible that the melting process can be reversed. Here complementary strands would bind to the 3' ends first, leaving the 5' ends unpaired. The strands could easily replicate from the overhanging 5' ends as replication typically proceeds 5'→3'.

HBs were defined using a newly developed definition. This HB definition assigns each base hydrogen atom a HB to one acceptor atom at all times. The acceptor atom selected is the one which forms the strongest intermolecular force with the hydrogen atom. Use of this new definition means that there are no bifurcated hydrogen atoms. It also avoids arbitrarily defined HBs and does not disqualify strained HBs. HBs can therefore be classified more simply. There are no intermediate HB states and so each hydrogen atom moves from one HB straight to another. Weak or strained hydrogen bonds would sometimes be excluded using some of the previous HB analysis programs such as Ptraj. This could result in intermediate structures important to the unfolding being overlooked.

Solvation plays an important role in unfolding. The probability of a HB to a solvent molecule approaches 1 with decreasing WC state. A previous study[107] stated that water-mediated base pairings were strong, with the water molecules being integral to the pairing. The formation of water-mediated pairings could be an important step in

unfolding. To further investigate the role of solvent it would be necessary to examine the locations of individual solvent molecules. For the purposes of this study all HBs formed to solvent molecules were classified as S. Hence HB exchanges occurring between solvent molecules were not monitored.

This study has confirmed the increased stability of the GC base pair over AT. For 4-mer strands, all AT duplexes were unwound after 15 ns heating MD whereas equivalent GC duplexes took up to 45 ns to unwind. Doubling the duplex length from 4-mers to 8-mers also increased the unwinding time. Many of the 8-mer AT duplexes were still not completely unwound after 50 ns and none of the 8-mer GC complexes completely unwound after 60 ns. Given more time the 8-mer simulations would have been run until unfolding had completed. For this study unfolding was induced by increasing the system's temperature to 373 K. To reduce the length of MD other conditions could be implemented. The chemical denaturant pyrimidine could be added, the salt concentration raised or a higher temperature used. Each of these methods would present their own unique issues. Heating above the boiling point of water (373 K) would give non-physical denaturing conditions.

The systems considered were simplistic. In reality, a DNA strand is likely to be a mixture of AT and GC base pairs. Further work could be done on this mechanism by using a wider variety of base sequences. Due to computational limitations in time and resources the DNA strands considered were also shorter than a more realistic system. Lengths of 16 base pairs or longer would provide a useful insight into the melting mechanism. To gain accurate melting times, many more trajectories would need to be analysed. This would require the use of a supercomputer such as HECToR (high end computing terascale resource). Accuracy of the simulations could be improved by consideration of polarisation effects. Currently AMBER does

not support this possibility but there has been a recently developed drude oscillator model for CHARMM for the simulation of nucleic acids.[6]

Even for the relatively short homo-oligomeric strands of DNA considered in this study, it is clear that the DNA melting process is extremely complex. There are many different intermediate structures and several possible routes via which melting can occur. Some of the possible mechanistic intermediates have been identified and discussed. It is likely that there are other intermediate structures which occur in more complex systems.

6. References

1. Smith, S. J.; Sutcliffe, B. T. Revs Comput Chem 1997, 70, 271-316.
2. Leach, A. R. Molecular Modelling: Principles and Applications; Pearson Education Ltd.: Harlow, 2001.
3. Marenduzzo, D.; Bhattacharjee, S. M.; Maritan, A.; Orlandini, E.; Seno, F. Phys Rev Lett 2001, 88, 028102.
4. Mackerell, A. D. J Comput Chem 2004, 25, 1584-1604.
5. Ponder, J. W.; Case, D. A. AdvProtein Chem 2003, 66, 27-85.
6. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. J Comput Chem 2009, 30, 1545-1614.
7. Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; Van Gunsteren, W. F. J Comput Chem 2005, 26, 1719-1751.
8. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. J Phys Chem B 2001, 105, 6474-6487.
9. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. J Comput Chem 2005, 26, 1668-1688.
10. Periole, X.; Cavalli, M.; Marrink, S. J.; Ceruso, M. A. J Chem Theory Comput 2009, 5, 2531-2543.
11. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Proteins 2006, 65, 712-725.
12. Keseru, G. M.; Kolossvary, I. Molecular Mechanics & Conformational Analysis in Drug Design; Wiley-Blackwell: Cornwell, 1999.
13. Reichi, L. E. A Modern Course in Statistical Physics; Wiley & Sons: Morlenbach, 2009.
14. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. V.; Hermes, J. Interaction Models for Water in Relation to Protein Hydration; D. Reidel Publishing Company: Dordrecht, 1981.
15. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. J Phys Chem 1987, 91, 6269-6271.
16. Jorgensen, W. L.; Madura, J. D. J Am Chem Soc 1983, 105, 1407-1413.
17. Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. Rev Mod Phys 1992, 64, 1045-1097.
18. Fedoryuk, M. V. Method of Steepest Descent; Kluwer Academic Publishers, 2001.
19. Knyazev, A. V.; Lashuk, I. SIAM J Matrix Anal Appl 2007, 29, 1267-1280.
20. Larriva, M.; de Sancho, D.; Rey, A. Physica A 2006, 371, 449-462.
21. Li, H.; Zhou, Y. J Bioinform Comput Biol 2005, 3, 1151-1170.
22. Zhang, C.; Liu, S.; Zhou, H.; Zhou, Y. Protein Sci 2004, 13, 400-411.
23. de Sancho, D.; Rey, A. J Comput Chem 2008, 29, 1684-1692.

24. Allan, N. L.; Barrera, G. D.; Purton, J. A.; Sims, C. E.; Taylor, M. B. Physical Chemistry Chemical Physics 2000, 2, 1099-1111.
25. Mesirov, J. P.; Schulten, K.; Summers, D. W. Pursing Laplate's Vision on Modern Computers; Springer-Verlag: New York, 1996.
26. Pirozzoli, S. In Annual Review of Fluid Mechanics, Vol 43; Annual Reviews: Palo Alto, 2011, p 163-194.
27. Verlet, L. Phys Rev 1967, 159, 98.
28. Sadus, R. J. Molecular Simulation of Fluids: Theory, Algorithms and Object-Orientation; Elsevier Science: Amsterdam, 2002.
29. Huang, W. Z.; Leimkuhler, B. SIAM J Sci Comput 1997, 18, 239-256.
30. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. J Comput Phys 1977, 23, 327-341.
31. Smith, W. CCP5 Quarterly 1983, 10, 37-42.
32. Ewald, P. P. Ann Phys 1921, 369, 253-287.
33. Darden, T.; York, D.; Pedersen, L. J Chem Phys 1993, 98, 10089-10092.
34. Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. Chem Soc Rev 2003, 32, 350-364.
35. Cheatham, T. E. Curr Opin Struct Biol 2004, 14, 360-367.
36. Beveridge, D. L.; McConnell, K. J. Curr Opin Struct Biol 2000, 10, 182-196.
37. Giudice, E.; Lavery, R. Accounts Chem Res 2002, 35, 350-357.
38. de Azevedo, W. F. Curr Med Chem 2011, 18, 1353-1366.
39. Cancès, E.; Defranceschi, M.; Kutzelnigg, W.; Le Bris, C.; Maday, Y.; Bris, C. L. In Handbook of Numerical Analysis; Elsevier, 2003, p 3-270.
40. Jensen, F. Introduction to Computational Chemistry; John Wiley & Sons: Chichester, England, 2007.
41. Chen, T. C. J Chem Phys 1955, 23, 2200-2201.
42. Hall, G. G. P R Soc London A Mat 1951, 205, 541-552.
43. Roothaan, C. C. Revs Modern Phys 1951, 23, 69-89.
44. Delhalle, J.; Piela, L.; Bredas, J.-L.; Andre, J.-M. Phys Rev B 1980, 22, 6254-6267.
45. Hehre, W. J.; Ditchfie.R; Pople, J. A. J Chem Phys 1972, 56, 2257-2261.
46. Dewar, M. J. S.; Jie, C.; Yu, J. Tetrahedron 1993, 49, 5003-5038.
47. Stewart, J. J. P. J Mol Modeling 2007, 13, 1173-1213.
48. Rzepa, H. S.; Yi, M. Y.; Karelson, M. M.; Zerner, M. C. J Chem Soc-Perkin Trans 2 1991635-637.
49. Crick, F. H. C.; Watson, J. D. Nature 1953, 171, 737-738.
50. Monajjemi, M.; Chahkandi, B.; Zare, K.; Amiri, A. Biochemistry-Moscow 2005, 70, 366-376.
51. Powell, B. M.; Martel, P. Biophys J 1981, 34, 311-323.
52. Alhambra, C.; Luque, F. J.; Gago, F.; Orozco, M. J Phys Chem B 1997, 101, 3846-3853.
53. Sponer, J.; Leszczynski, J.; Hobza, P. Biopolymers 2001, 61, 3-31.
54. Sponer, J.; Sabat, M.; Burda, J. V.; Leszczynski, J.; Hobza, P. J Phys Chem B 1999, 103, 2528-2534.
55. Li, X.-Q.; Fan, P. J Theor Biol 2010, 266, 374-379.
56. Gorb, L.; Podolyan, Y.; Dziekonski, P.; Sokalski, W. A.; Leszczynski, J. J Am Chem Soc 2004, 126, 10119-10129.
57. Toueille, M.; Hubscher, U. Chromosoma 2004, 113, 113-125.
58. Hamdan, S. M.; van Oijen, A. M. J Biol Chem 2010, 285, 18979-18983.

59. Steiner, T.; Saenger, W. Acta Crystallogr Sect B-Struct Commun 1994, 50, 348-357.
60. Berg, J. M.; Tymoczko, J. L.; Stryer, L. Biochemistry: New York, 2007.
61. Qiu, Z. M.; Wang, H. J.; Xia, Y. M. Struct Chem 2010, 21, 931-937.
62. Mohajeri, A.; Nobandegani, F. F. J Phys Chem A 2008, 112, 281-295.
63. Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. Nucleic Acids Res 2006, 34, 564-574.
64. Sponer, J.; Jurecka, P.; Marchan, I.; Luque, F. J.; Orozco, M.; Hobza, P. Chem-Eur J 2006, 12, 2854-2865.
65. Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A. J Comput Chem 2010, 31, 1644-1655.
66. Baker, C. M.; Anisimov, V. M.; MacKerell, A. D. J Phys Chem B 2010, 115, 580-596.
67. Asensio, A.; Kobko, N.; Dannenberg, J. J. J Phys Chem A 2003, 107, 6441-6443.
68. Grunenberg, J. J Am Chem Soc 2004, 126, 16310-16311.
69. Sfyrakis, K.; Provata, A.; Povey, D. C.; Howlin, B. J. J Mol Model 2004, 10, 185-197.
70. Bendic, C.; Enache, M.; Volanschi, E. J Mol Graph 2005, 24, 10-16.
71. Brandhorst, K.; Grunenberg, J. ChemPhysChem 2007, 8, 1151-1156.
72. Sedlak, R.; Jurecka, P.; Hobza, P. J Chem Phys 2007, 127, 3.
73. Czyznikowska, Z.; Gora, R. W.; Zalesny, R.; Lipkowski, P.; Jarzembska, K. N.; Dominiak, P. M.; Leszczynski, J. J Phys Chem B 2010, 114, 9629-9644.
74. Babcock, M. S.; Pednault, E. P. D.; Olson, W. K. J Biomol Struct Dyn 1993, 11, 597-628.
75. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, L.; Yang, C.; Tan, J.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. In AMBER; University of California, S. F., Ed., 2008.
76. Kaukinen, U.; Lonnberg, H.; Perakyla, M. Org Biomol Chem 2004, 2, 66-73.
77. McDonald, I. K.; Thornton, J. M. J Mol Biol 1994, 238, 777-793.
78. Lindauer, K.; Bendic, C.; Suhnel, J. Comput Appl Biosci 1996, 12, 281-289.
79. Dailey, M. M.; Miller, M. C.; Bates, P. J.; Lane, A. N.; Trent, J. O. Nucleic Acids Res 2010, 38, 4877-4888.
80. Sorin, E. J.; Pande, V. S. Biophys J 2005, 88, 2472-2493.
81. Shen, L. X.; Tinoco, J. I. J Mol Biol 1995, 247, 963-978.
82. Rose, A.; Goede, A.; Hildebrand, P. W. Nucleic Acids Res 2010, 38, W602-W608.
83. Lu, X. J.; Olson, W. K. Nucleic Acids Res 2003, 31, 5108-5121.
84. Koch, U.; Popelier, P. L. A. J Phys Chem 1995, 99, 9747-9754.
85. Bader, R. F. W. J Phys Chem A 1998, 102, 7314-7323.
86. Bar, A.; Kafri, Y.; Mukamel, D. Phys Rev Lett 2007, 98, 038103.
87. Xiao-Feng, P.; Ynan-Ping, F. J Biomol Struct Dyn 2008, 25, 435-451.
88. Baiesi, M.; Barkema, G. T.; Carlon, E.; Panja, D. J Chem Phys 2010, 133, 154907-154911.
89. Chen, Y. Z.; Prohofsky, E. W. Phys Rev E 1993, 48, 3099-3106.
90. Chen, Y. Z.; Prohofsky, E. W. Biopolymers 1993, 33, 351-362.

91. Wenner, J. R.; Williams, M. C.; Rouzina, I.; Bloomfield, V. A. Biophys J 2002, 82, 3160-3169.
92. Currie, J. F.; Krumhansl, J. A.; Bishop, A. R.; Trullinger, S. E. Phys Rev B 1980, 22, 477.
93. Ma, H.; Wan, C.; Wu, A.; Zewail, A. H. PNAS 2007, 104, 712-716.
94. Marco, B.; Roberto, L. J Phys A-Math Theor 2009, 42, 082003.
95. H.Kunz; Livi, R.; Suto, A. J Stat Mach-Theory E 2007, 2007, P06004.
96. Novotny, T.; Pedersen, J. N.; Ambjornsson, T.; Hansen, M. S.; Metzler, R. Europhys Lett 2007, 77, 48001-48008.
97. Ying, L. M.; Wallace, M. I.; Klenerman, D. Chem Phys Lett 2001, 334, 145-150.
98. Kuznetsov, S. V.; Shen, Y. Q.; Benight, A. S.; Ansari, A. Biophys J 2001, 81, 2864-2875.
99. Ansari, A.; Kuznetsov, S. V.; Shen, Y. Q. PNAS 2001, 98, 7771-7776.
100. Peyrard, M.; Bishop, A. R. Phys Rev Lett 1989, 62, 2755-2758.
101. Kabak; ccedil; iogbrevelu, A.; Orlandini, E.; Mukamel, D. Phys Rev E 2009, 80, 010903.
102. Drukker, K.; Wu, G. S.; Schatz, G. C. J Chem Phys 2001, 114, 579-590.
103. Drukker, K.; Schatz, G. C. J Phys Chem B 2000, 104, 6108-6111.
104. Wong, K.-Y.; Pettitt, B. M. Biophys J 2008, 95, 5618-5626.
105. Perez, A.; Orozco, M. Angew Chem Int Ed 2010, 49, 4805-4808.
106. Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. PNAS 1981, 78, 2179-2183.
107. Imoto, T. Biochim Biophys Acta 1977, 475, 409-416.
108. Kabelac, M.; Ryjacek, F.; Hobza, P. Phys Chem Chem Phys 2000, 2, 4906-4909.
109. Auffinger, P.; Westhof, E. Biophys Chem 2002, 95, 203-210.
110. Brown, R. A.; Case, D. A. J Comput Chem 2006, 27, 1662-1675.
111. Crowley, M. F.; Williamson, M. J.; Walker, R. C. International Journal of Quantum Chemistry 2009, 109, 3767-3772.
112. Ghosh, A.; Bansal, M. Acta Crystallogr D 2003, 59, 620-626.
113. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179-5197.
114. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Biophys J 2007, 92, 3817-3829.
115. Hess, B.; van der Vegt, N. F. A. J Phys Chem B 2006, 110, 17616-17626.
116. Cheatham, T. E.; Kollman, P. A. Molecular dynamics simulation of nucleic acids in solution: How sensitive are the results to small perturbations in the force field and environment?; Adenine Press: Schenectady, 1998.
117. Tieleman, D. P.; Berendsen, H. J. C. J Chem Phys 1996, 105, 4871-4880.
118. Humphrey, W.; Dalke, A.; Schulten, K. J Mol Graph 1996, 14, 33-38.