DEPARTMENT OF PHYSICS
UNIVERSITY OF JYVÄSKYLÄ
RESEARCH REPORT No. 1/2006

DYNAMIC MODELING OF DNA

BY
RAIMO LOHIKOSKI

Academic Dissertation
for the Degree of
Doctor of Philosophy

To be presented, by permission of the
Faculty of Mathematics and Natural Sciences
of the University of Jyväskylä,
for public examination in Auditorium FYS - 1 of the
University of Jyväskylä on January 21, 2006
at 12 o'clock noon

Jyväskylä, Finland
January 2006

ISBN 978-951-39-3166-7
URN:ISBN:978-951-39-3166-7

Kopijyvä Oy, Jyväskylä 2006

Preface

This thesis is based on studies carried out at the Department of Physics in the University of Jyväskylä and at the Arrhenius laboratory in the Stockholm University.

I wish to give all my greatest thanks to Professor Aatto Laaksonen in Stockholm. The accomplishment of my DNA works as this thesis is very much credit to him. We have had a long and educative trip till this point.

In Jyväskylä professors Jouko Korppi-Tommola and Jussi Timonen are the persons to acknowledge. In particular, the latter took long care of my prerequisities.

Financial support provided by the The Finnish Society of Science and Letters via Magnus Ehrnrooth foundation, and the Finnish Academy of Science and Letters from the Vilho, Yrjö and Kalle Väisälä foundation deserve also attention for the finalization of this work.

Jyväskylä, January 2006

Raimo Lohikoski

Abstract

Progress in biomolecular science is fast. Biologists, physicists, chemists, mathematicians and algorithm-informaticians are pushing the frontier regarding methods, observations, and theories continuously deeper and deeper into the secrets of life.

This thesis gives a touch on a few sides of this huge field. I have collected here some of my results of computationally revealed dynamical properties of the biological DNA in solution. I have used molecular mechanical modeling combined with the molecular dynamics simulation methodology and analysis tools devised for the DNA. Testing and application of these methods have existed hand by hand, leading at the end to some predictions of the dynamical behaviour of the DNA molecule in its natural environment.

The highlight of the presented works is the computer aided 'single molecule mechanical manipulation' of the DNA by stretching, extended for more than twice of its equilibrium length, performed as an all-atom molecular dynamical simulation in the natural-like environment. As a result, a novel structure was found, which consists of a complete unwinding of the two strands, in conjunction with stacking of the base-pairs collectively on the major groove side. The observed force-strain curve during this process is qualitatively very similar to the ones obtained experimentally from single-DNA. These simulations may consequently offer first understanding of atomistic processes taking place during single-molecule manipulations.

List of Publications

This thesis is based on the following publications:

I Raimo A. Lohikoski and Aatto Laaksonen, *Simple Models for Nonlinear States of Double Stranded DNA*, In 'Modern methods for theoretical physical chemistry of biopolymers', Edited by S. Tanaka, J. Lewis, and E. Starikow, Elsevier, (2005)

II Raimo A. Lohikoski, Jussi Timonen, Alexander P. Lyubartsev and Aatto Laaksonen, *Internal Structure and Dynamics of the Decamer* $d(ATGCAGTCCAG)_2$ *in* Li^+ - H_2O *solution: A Molecular Dynamics Simulation Study*, Molecular Simulation **29**, 47-62 (2003)

III Raimo Lohikoski, Jussi Timonen, Aatto Laaksonen, *Molecular dynamics simulation of single DNA stretching reveals a novel structure*, Chem. Phys. Lett. **407**, 23-29 (2005)

IV Raimo A. Lohikoski, Martin Dahlberg and Aatto Laaksonen, *Computer Simulations of DNA Stretching*, In 'Modern methods for theoretical physical chemistry of biopolymers', Edited by S. Tanaka, J. Lewis, and E. Starikow, Elsevier, (2005)

As the author I have written the drafts of the above manuscripts and been refining them with the other contributors. The Appendix of paper IV was written by M. Dahlberg. The modeled DNA trajectory I studied in paper II is by A. P. Lyubartsev. Otherwise all actual planning, development, coding, analysis, data set care and also partially computers management is by myself.

This thesis benefits also from the following works:

V R.A. Lohikoski, *Excitation Transport in Helical Proteins: Dynamics of the Davydov-Scott Equations in the Adiabatic Approximation*, Licentiate Thesis, Department of Physics, University of Jyväskylä (Laboratory Report 3/1992)

VI V. Helenius, J. Korppi-Tommola, S. Kotila, J. Nieminen, R. Lohikoski and J. Timonen, *Anomalous temperature dependence of the IR spectrum of polyalanine*, Chem. Phys. Lett. 280, 325-332 (1997)

VII Juraj Bunta, Martin Dahlberg, Leif A. Eriksson, Nikolai Korolev, Aatto Laaksonen, Raimo Lohikoski, Alexander Lyubartsev, Miroslav Pinak and Patrik Schyman, *Solvating, manipulating, damaging and repairing DNA in a computer, In Press*, International Journal of Quantum Chemistry for the proceedings from the Theoretical Biophysics Symposium, Örebro, Sweden, June 28 - July 1, 2005

Contents

1 DNA

All living things are composed of *cells*. Most cells have a *nucleus*. Each nucleus includes a DNA molecule, typically packed in a compact form called *chromatin*. The DNA molecule affects every part of all living systems, also those parts whose cells do not include the DNA itself.

Basic processes of normal DNA are preservation, transcription and replication of genetic information. The word *processes* has to be emphasized because these events, at the current level of knowledge, are cooperative processes, where DNA has its role or roles and other participating molecules their own. Seeing only one of these entities does not make us wise enough. We need to look at the whole system, and this is the state where structure-function oriented science is going today.

Term *preservation* means trusted storage or survival of the information content of the genetic information in the DNAs molecular structure. In *transcription* DNA is 'read' or it 'gives' a piece of this genetic information, typically for protein building processes. And, in *replication* occurring nearer the state of the cell division phase semiconservative 'division' makes two new DNAs of the old one, with the same genetic code or information content, offering possibilities for the continuing existence of our diverse biosystems. Molecular machinery seems to have own subprocesses for all of these processes governed by a some kind of determinism combined with randomness.

All molecules are physical and chemical compositions. Molecules appearing in biological environments are called *biomolecules*. In the following I shall give the basics and terminology relating quite tightly to the object of this work, the double stranded DNA. Much more detailed discussions and, e.g. closely related RNA facts, can be found from Refs. [1, 2, 3, 4, 5].

1.1 Structure of DNA

The current view of DNA mentions it as a linear molecule having double helical form during most of its life cycle. It is composed of nucleotides, which are obtained by esterification of the 3' or 5' hydroxyl group of nucleosides with phosphoric acid. Nucleosides in DNA are deoxyribosyl derivatives of certain pyrimidine (Y) and purine (R) bases.

1.1.1 Bases

In the DNA these bases are thymine (T or THY) and cytosine (C or CYT) pyrimidines and adenine (A or ADE) and guanine (G or GUA) purines. Their structure is a heterocycle composed of one six atom ring in pyrimidines and ring duplex of five and six atoms in purines. Figure 1 illustrates these planar structures.

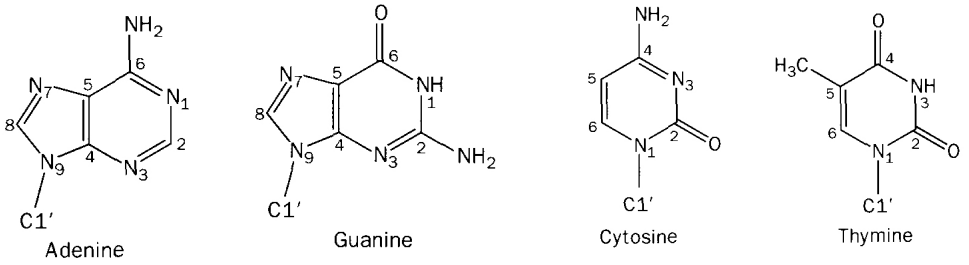

Figure 1.1: The bases of a DNA [4, Fig. 2.1]

1.1.2 Sugars

The deoxyribosyl part of DNA is called sugar when exact chemical terminology is not necessary. Atoms counted towards the sugar group in DNA are by convention marked with an apostrophe (') or a star (*).

DNA sugars are basically flat. However, above their normal modes they undergo phenomenon called *sugar puckering*. In relation to it five endocyclic torsion angles are used to define the sugar conformation, see Fig. 1.2. Sugar pucker is an *envelope* if one ring atom is above the plane of the rest four and a *twist* if two atoms pucker on opposite sides with respect to the plane of the other three atoms.

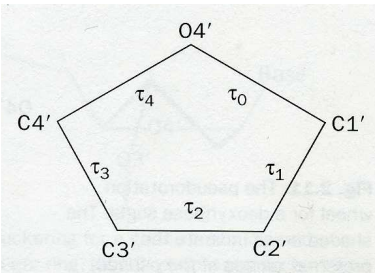

Figure 1.2: Internal torsion angles of sugar group [4, Fig. 2.10]

These five torsion angles are used to define the pseudo-rotation phase angle P:

$$\tan(P) = \frac{(\tau_4 + \tau_1) - (\tau_3 + \tau_0)}{2\tau_2(\sin(36°) + \sin(72°))}, \tag{1.1}$$

and the maximum degree of the pucker:

$$\tau_m = \frac{\tau_2}{\cos(P)}. \tag{1.2}$$

Pseudo-rotation wheel is proved to be useful and convenient way to represent P, Fig. 1.3.

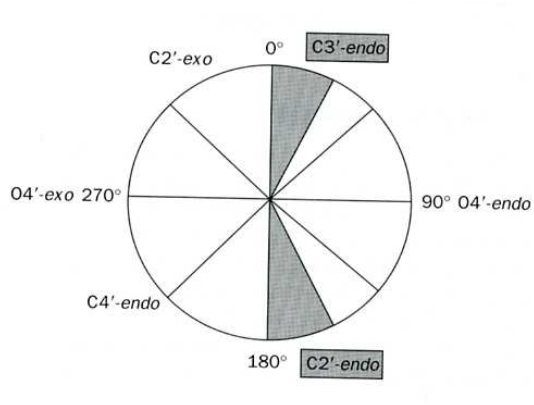

Figure 1.3: Pseudorotation wheel [4, Fig. 2.11]

The C3'-*endo* and C2'-*endo* in Fig. 1.3 indicate the typical observed puckers. Term *endo* refers to the appearance of the (major) pucker on the same side as the base and C4'-C5' bond; otherwise it is termed *exo*.

When unbound, purines favor C2'-*endo* and pyrimidines C3'-*endo* conformations. In coupled structures things change. The overall picture is still incomplete. In DNA sugar puckering correlates with many backbone conformational variables.

Bond between sugar and its coupled base is called *glycosidic bond*. Ref. [4] emphasizes its stereochemical property. In natural nucleic acids this bond is always above the sugar plane on the 5' side (β).

The atom chains O4'-C1'-N9-C4 in purines and O4'-C1'-N1-C2 in pyrimidines define the only parameter (variable using basic course language). It is a torsion angle denoted by χ. It's *syn* and *anti* regions are shown in Fig. 1.4.

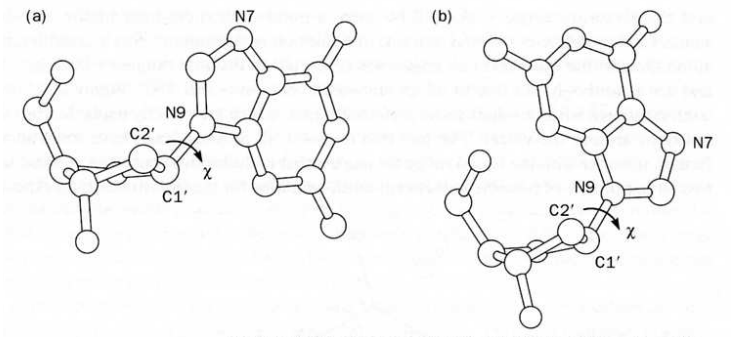

Figure 1.4: Adenine nucleoside visualizing the, a) *syn* and, b) *anti* conformations [4, Fig. 2.13]

1.1.3 Phosphates

The third submolecule in DNA is phosphate, the P-atom surrounded with four oxygens, see Fig. 1.5. Sometimes it is marked with "p".

One nucleotide or basic block of DNA with atomic information is shown in Fig. 1.5. Ideal coordinates of all the bases are obtainable from the Nucleic Acid Database [6]. A lot of structural data can also be found from the Cambridge Structural Database [7].

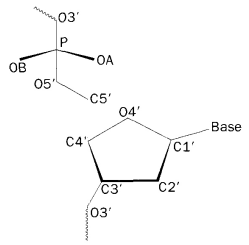

Figure 1.5: One nucleotide of a DNA [4, Fig. 2.2]

1.1.4 Backbone - strands - conformation parameters

DNA strands or polynucleotide chains are obtained putting nucleotides on top of each others like LEGO-blocks. The O3' atom of the (previous) phosphate binds always to C3'-atom of the (next) sugar.

Convention in single strand numbering goes from the 5' end to the 3' end. This means that for a strand or combination of strands in a solution the last nucleotide is cut

after the O5'-atom, and terminated by e.g. hydrogen. The other end, i.e. O3'-atom, is terminated 'naturally' with a suitable chemical group.

Most typical textual notation of a chain is without p's or phosphates, giving the sequence of bases, i.e. ATGGC is given instead of ApTpGpGpC. Usually this notation means double stranded DNA whose other strand is evident according to the base pairing rules. Many times texts contain notation like $d(ATGGC)_2$. The 'd' emphasizes the DNA ('deoxy') character of the molecule and want make difference with the RNA ('ribo')-family. Subscript emphasises the double helical character of the given sequence. Notation like poly(dG) refers to synthetic homopolymer composed of G nucleotides. Synthetic heteropolymer could be e.g. poly(dA-dT) for alternating sequences, and poly(dA, dT) for randomly distributed nucleotides. Complexes formed of homologous sequences may carry notation like poly(G) • 4 poly(C). Unfortunately completele rules for notations are not known by all scientists, and variations occur occasionally.

Physico-chemically the nucleic acid backbone is a highly charged polyelectrolyte: the phosphate region of the structure is negatively charged. This charged character is one important fact making structure variable with respect to differing conditions and environments. As an acid it is deprotonated in water.

In the case of DNA backbone the early analyses of 1^{st} and 2^{nd} order interactions - forbidden interactions or steric hindrances depending on change of only one torsion angle or two of such - has reduced the DNA strand structural representation to the use of *backbone torsion angles* [8]. These, $\alpha, \beta, \gamma, \delta, \epsilon$ and ζ and the glycosidic parameter form the *conformational parameters* of the double stranded DNA structures. All these are shown in Fig. 1.6a.

Their preferred values in single nucleotides are shown in Fig.1.6b. Often terms *gauche*$^+$ (g^+), *gauche*$^-$ (g^-) and *trans* (t) are used for these when their values are around 60°, −60°, and 180°, respectively. In the Watson-Crick DNA structures the α and ζ can only have the g^-g^- conformation. Torsion angle β is mostly in *trans* position. For γ the right-handed double helices seem have g^+. The δ relating to the sugar puckering has values around 75° for $C3' - endo$ and 150° for $C2' - endo$ puckers.

One set of nomenclatures for the torsion angles from the organic chemistry is the *Klyne-Prelog* system[3]. In it *syn* (s) designation is used for torsion angles near 0° and *anti* (a) for those near 180°. Values between these carry nominations $\pm$*synclinal* ($\pm sc$) - around $\pm$60° - and $\pm$*anticlinal* ($\pm ac$) - around $\pm$120°.

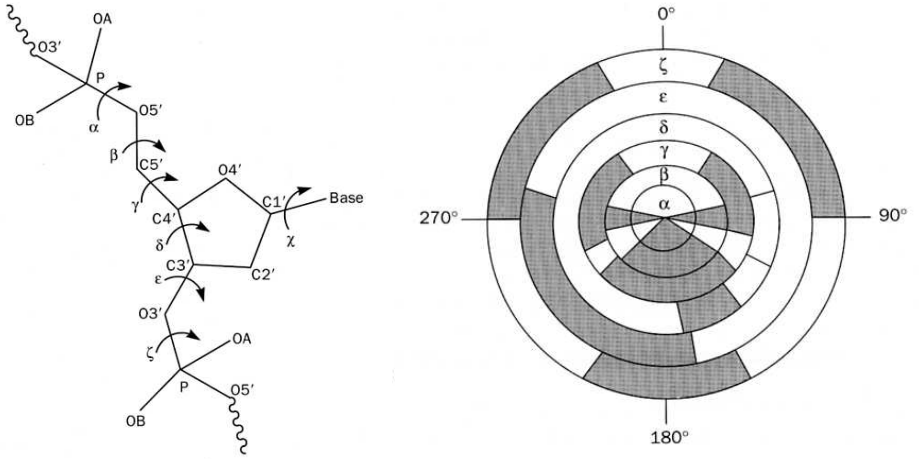

Figure 1.6: a) DNA's conformational parameters [4, Fig. 2.14], and b) their allowed ranges of in nucleosides, nucleotides, deoxyoligonucleotides and deoxypolynucleotides [4, Fig. 2.15].

1.2 Conformations of DNA

While the biological form of DNA ('native DNA') has the famous right-handed, double-stranded structure, there are many different structures, composed of nucleotides, that are grouped under the name 'DNA'.

Geometric information of DNA and its constituents is based on X-ray crystallography and nuclear magnetic resonance (NMR). High resolution crystallography (resolution 0.7-0.9 Å) also gives information of electron-density distribution in molecules and, consequently, partial charges of the atoms.

1.2.1 DNA double helices

In nature the DNA bases form pairs A-T and G-C with the hydrogen bonding mechanism as proposed by Watson and Crick in 1953/4 [9, 10], and shown in Fig. 1.7.

Adenine and thymine are bonded by two hydrogen bonds, guanine and cytosine have three coupling hydrogen bonds.

In favourable solution conditions monomeric nucleic acids always and spontaneously form these hydrogen bonded pairs [1]. Instead, in solid state mixtures only monomers G and C make the same, but A and T never [1]. Solvated bases are hydrophobic and attractive in the Watson-Crick style. Further, experiments on solvated single

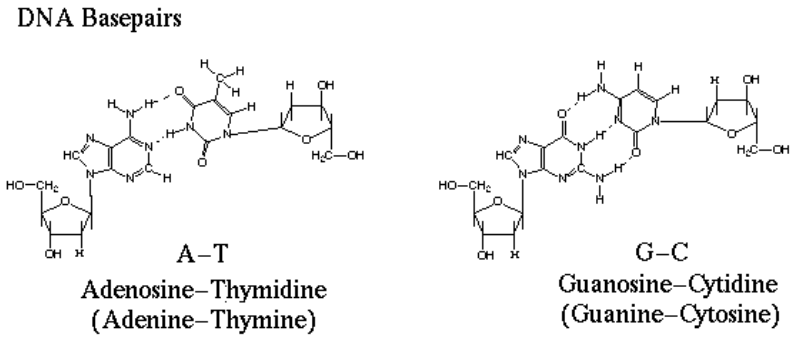

Figure 1.7: Base pairs in the double stranded DNAs [11]

stranded polynucleotides have shown that bases have stacking tendency, explainable by the connective π-electron clouds on the flat sides of the planar structure. All this together with the classical potential energy calculations [3] make the double stranded helical DNA structure understandable. Further, Ref. [4] mentions the fact that sugar groups links each base on the same side as the keys in the structure of dsDNA. The C1' - C1' distance in both types of base pairs is about 10.6 Å.

The helicoidal structures of the 'normal' DNA is shown in Fig. 1.11.

Before going into the DNA structures I give here the usual structural concepts:

- The base pair *axis system*. The x-axis (shorter axis) points in the direction of the major groove along what would be the pseudo-dyad axis of an ideal Watson-Crick base pair. The y-axis runs along the long axis of the idealized base pair in the direction of the sequence strand, parallel to the C1' - C1' vector, and displaced so as to pass through the intersection on the x-axis of the vector connecting the pyrimidine C6 and purine C8 atoms. The 3^{rd} axis follows the usual right-handed rule.
- *Helix sense* refers to the handedness of the double helix.
- *Pitch* (per turn) P_h is the distance along the helical axis in one full turn.
- Number of residues per turn n_b is the number of base pairs in one P_h.
- *Axial rise* h is the axial distance between two base pairs.
- Unit *twist* or rotation per residue $= \Omega = 360°n_b$ is the rotation of DNA axis from one base pair to the next one.
- Helix *diameter* measures the diameter of the cylindrical shape of DNA.

- The wider and narrower indentations in the double strand structure are called respectively *major* and *minor grooves*, distinguishable by the fact that the base-sugar bonds are on the minor-groove sides. The perpendicular distance between opposite phosphorous atoms on opposite strands deducted by van der Waals diameter of a phosphate group (5.8 Å) gives usually the width of the grooves. Depths of the grooves are typically determined as differences in radii between phosphorous and GUA N2 atom (minor groove), and phosphorous and ADE N6 (major groove).

Helicoidal parameters

Besides the previous geometric definitions for the nucleic acid helices there is a set of *local* and a set of *global* base pair related variables which commonly are called helicoidal parameters. Those names, symbols, and sign conventions were made in 1989 [12] in the meeting carrying name 'Cambridge Accord'. The 'Tsukuba Accord' in 1999 [13] corrected deficiencies occurred specifically in the mathematical definitions. Due to these historical incoherencies all structural parameters having same names are not comparable until exact equations are known.

The local variables specify orientation between successive base pairs. The global parameters reflecting a best linear or overall curved molecular axis, are not necessarily comparable with these values.

The axis-base pair parameters related to each base pair are:

- *Displacement* X (XDP) indicate the translation of a base pair in its mean plane in x-direction. Positive XDP means displacement towards the major groove.
- *Displacement* Y (YDP) indicate the translation of a base pair in its mean plane in y-direction. YDP is positive when movement is towards the 1^{st} strand.
- *Inclination* (INC; η) is the angle formed by the longer axis of the base pair and plane perpendicular to the helix axis.
- TIP (θ) is the angle between the base pair's shorter axis and the (local) helical axis z.

The helical twist (Ω) and helical rise (h) are sometimes includes in this category.

The complementary base-pair parameters are:

- *Shear* (SHR)

- *Stretch* (STR)
- *Stagger* (STG)
- *Buckle* (BKL; κ) is the angle that the bases form around their short axis when PRP is zero. Sign of BKL is positive when distortion is convex in direction 5' to 3' in strand 1.
- *Opening* (OPN; σ)
- *Propeller* (π; PRP) is defined as the angle between base plane's normals. Viewed along the base pairs long axis this angle is positive if nearer base is clockwise rotated.

The base pair step parameter set is composed of the following:

- *Shift* (SHF; Dx)
- *Slide* (SLD; Dy) tells the relative displacement of successive base pairs towards the strand 1. Here the midpoints of the longer axes are the reference points.
- *Rise* (RIS; Dz)
- *Tilt* (TLT; τ) is the same as ROL but along the x- or short axis.
- *Roll* (ROL; ρ) indicates the rotation of one base pair around its longer axis with respect to its neighbour base pair. This is positive when base pair step opens towards the minor groove.
- *Twist* (TWS; ω) measures angle between consecutive base pairs as the change in the C1'-C1' vectors direction and down the helix axis.

All these are depicted in Figs. 1.8 and 1.9 with division in translational and rotational groups.

A-, B-, Z- and sub-families

The original Watson-Crick structure model of DNA based on diffraction patterns obtained at high humidity (92 %). This structure got later name B-DNA, the natural DNA because this is the structure DNA has in its natural molecular environment. A-DNA forms at 65 to 75 % relative humidity. Z-DNA is the third one to be mentioned because the A, B and Z forms are the stardard or canonical structures [1]. Factors

[1] in this context in the meaning 'generally accepted' or ideal

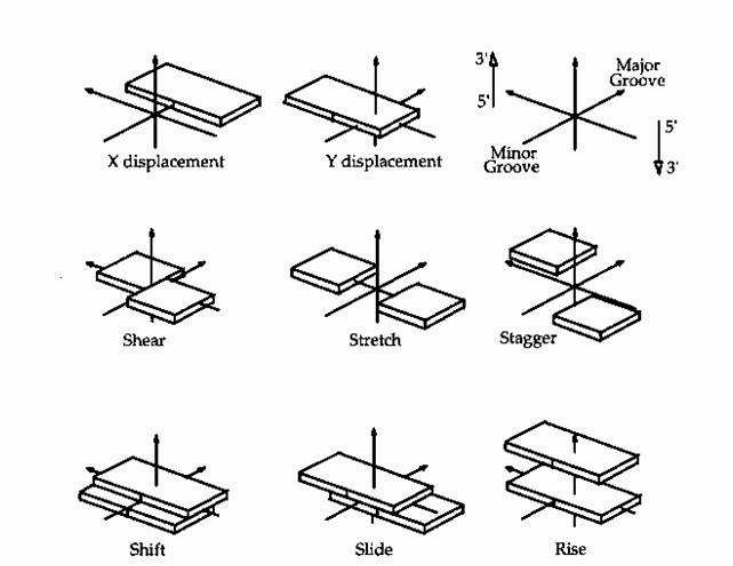

Figure 1.8: Translational helicoidal parameters [14]

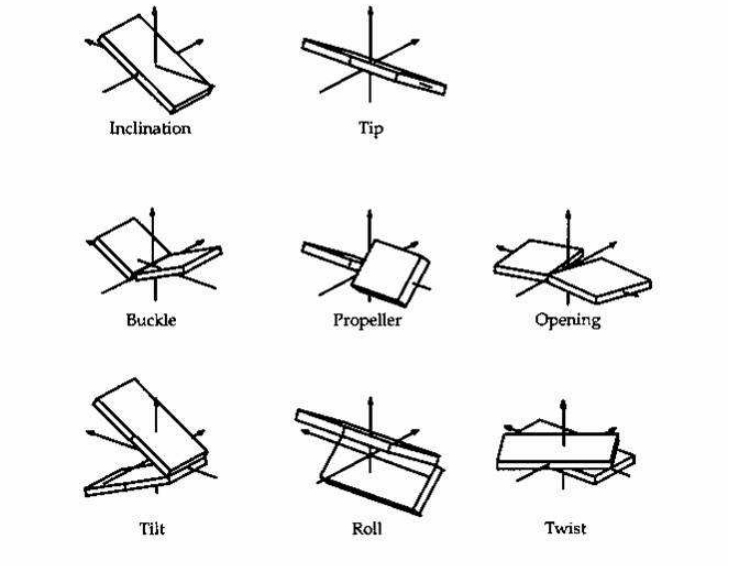

Figure 1.9: Rotational helical parameters [14]

affecting on structural form of DNA contain humidity, temperature, pH and sequence composition.

The canonical structures are shown in Figures 1.10, 1.11, and 1.12. Structural data are given in tables 1.1,1.2, 1.3 and 1.4.

	Unit repeat	RIS	TWS	PRP	ROL	INC
Form		(Å)	(°)	(°)	(°)	(Å)
A	11	2.54	32.7	-10.5	0.0	22.6
B	10	3.38	36.0	-15.1	0.0	2.8
C	28/3	3.31	38.6	-1.8	0.0	-8.0
D	8	3.01	45.0	-21.0	0.0	-13.0
Z (C)	6	7.25	-49.3	8.3	5.6	0.1
Z (G)	6	7.25	-10.3	8.3	-5.6	0.1

Table 1.1: Structural parameters of known DNA forms

Form	α	β	γ	δ	ε	ζ	χ
	(°)	(°)	(°)	(°)	(°)	(°)	(°)
A	-52	175	42	79	-148	-75	-157
B	-30	136	31	143	-141	-161	-98
C	-37	-160	37	157	161	-106	-97
D	-59	156	64	145	-163	-131	-102
Z (C)	-140	-137	51	138	-97	82	-154
Z (G)	52	179	-174	95	-104	-65	59

Table 1.2: Conformational parameters of DNA's

Form	Major groove		Minor groove	
	Width (Å)	Depth (Å)	Width (Å)	Depth (Å)
A	2.2	13.0	11.1	2.6
B	11.6	8.5	6.0	8.2
C	10.5	7.6	4.8	7.9
D	9.6	6.2	0.8	7.4
Z	8.8	3.7	2.0	13.8

Table 1.3: Polymorphism of DNA structures

Parameter	Symbol	A-DNA	B-DNA
Complementary base-pair parameters			
Buckle (°)	κ	0.1 (7.8)	0.5 (6.7)
Propeller (°)	π	-11.8 (4.1)	-11.4 (5.3)
Opening (°)	σ	0.6 (2.8)	0.6 (3.1)
Shear (Å)	Sx	0.01 (0.23)	0.00 (0.21)
Stretch (Å)	Sy	-0.18 (0.10)	-0.15 (0.12)
Stagger (Å)	Sz	0.02 (0.25)	0.09 (0.19)
Base-pair step parameters			
Tilt (°)	τ	0.1 (2.8)	-0.1 (2.5)
Roll (°)	ρ	8.0 (3.9)	0.6 (5.2)
Twist (°)	ω	31.1 (3.7)	36.0 (6.8)
Shift (Å)	Dx	0.00 (0.54)	-0.02 (0.45)
Slide (Å)	Dy	-1.53 (0.34)	0.23 (0.81)
Rise (Å)	Dz	3.32 (0.20)	3.32 (0.19)
Local helical parameters			
Inclination (°)	η	14.7 (7.3)	2.1 (9.2)
Tip (°)	y	θ 0.1 (5.2)	0.0 (4.3)
Helical twist (°)	Ω_h	32.5 (3.8)	36.5 (6.6)
x-displacement (Å)	dx	-4.17 (1.22)	0.05 (1.28)
y-displacement (Å)	dy	0.01 (0.89)	0.02 (0.87)
Helical rise (Å)	h	2.83 (0.36)	3.29 (0.21)

Table 1.4: Helicoidal parameter values for the A- and B-DNAs [13, Table 3]

The main characteristics of the canonical forms can be listed shortly:

A-DNA:

- 11 bp/ 1 full turn
- Hollow cylinder
- Very narrow major groove

B-DNA:

- 10 bp/ 1 full turn
- Nearly perfect verticality between helical axis and the bases

Z-DNA:

- The pyrimidine nucleosides have *anti* glycosidic angles with $-145° < \chi < -160°$, mean 152°, and C2'-*endo* sugar pucker. α, γ conformation is $trans, gauche^+$ and $gauche^+$ for ζ.
- The purine nucleosides have *syn* glycosidic angles with $55° < \chi < 80°$, mean 60° and C3'-*endo* puckers. α, γ conformation is $gauche^+, trans$ and $gauche^-$ for ζ.
- Consequently repeating unit is CpG dinucleoside.
- Helix does not have any major groove and only a very narrow minor groove.

The Z-DNA characteristics are for the standard type denoted by Z_I. In the other form, Z_{II}, purines have $gauche^+$ ζ angle, and $gauche^-$ for ε instead of *trans*.

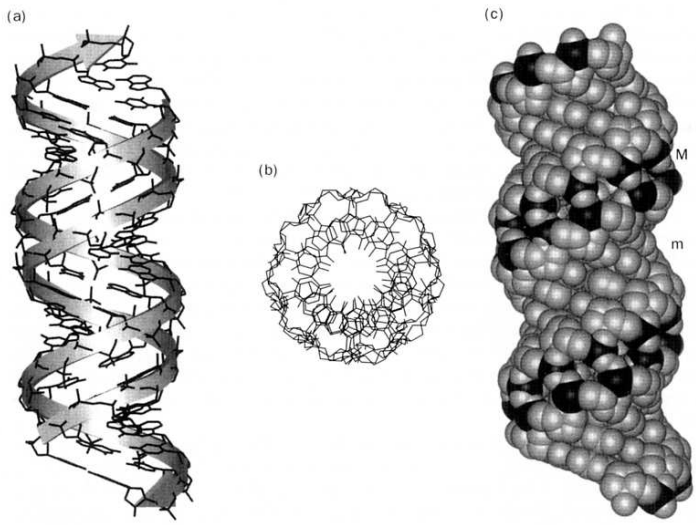

Figure 1.10: A-DNA [4, Fig. 3.5]

The Z-DNA differs from all other structural forms by its left-handedness. It is mentioned to be a sequence of poly(dC-dG) • poly(dC-dG) and, obviously, a structure which needs high salt solution, over 2.5 M NaCl, in order to form.

The *Dickerson-Drew dodecamer*, d(CGCGAATTCGCG), was the first structure independently found to reproduce the Watson-Crick DNA model structure in 1979 [4]. It also revealed fine details, like very restructured grooves in the 5'-AATT region (12.7 Å major groove vs. 3.2 Å minor groove) as well as 'spine of hydration', local base pair

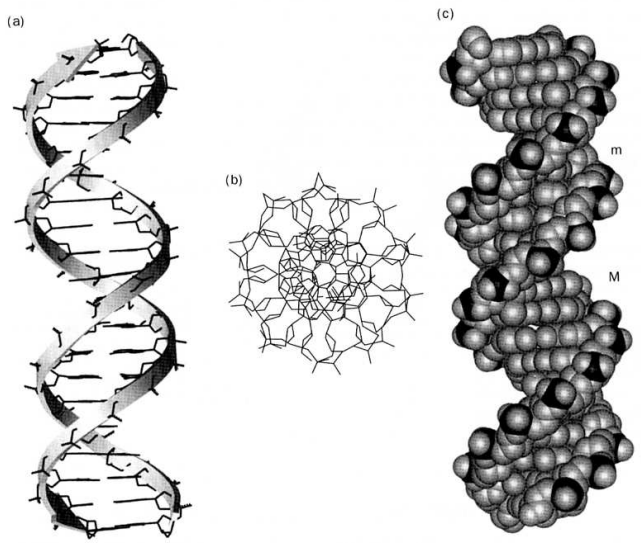

Figure 1.11: B-DNA [4, Fig. 3.3]

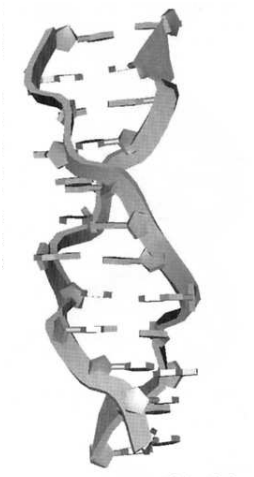

Figure 1.12: Ribbon representation of the Z-DNA [4, Fig. 3.7]

parameter deviations, especially in the central region of A-T pairs, wide distribution of values for the other parameters, the B_I/B_{II} structural detail of p arising from the $trans, gauche^-$ vs. $gauche^-, trans$ domain change in ε and ζ when nucleotide is followed by a purine, and *bent* structure.

Calladine rules [3] originate from Calladines notion that PRP while improving stacking interactions in double stranded structures also may lead to steric clashes between neighbouring purines in opposite strands. In purine-(3',5')-pyrimidine sequences interaction of GUA O6 and ADE N6 in the major groove side is the reason for this clash. In pyrimidine-(3',5')-purine helices clashes emerge in the minor groove side due to both GUA N3 and N2 and ADE N3 atoms. Effect of clashes is twice as large in the latter case as in the former. Most notably, these are in mixed purine/pyrimidine sequences though they exist in homopurine or homopyrimidine sequences. The Calladine rules tell how these are avoidable: reduce local PRP, increase ROL, increase SLD to pull purine out of the base stack and decrease local TWS to minimize stacking.

The structural studies made so far with differing double strand sequences have given a vast amount of structural fine details. These relate both to experimental conditions as well as to specificity of the sequences considered. Because the purpose of this chapter is to give the terminology relating to the DNA research and the main lines on double stranded DNAs any finer survey is beyond the scope of this introduction.

1.2.2 Other 'DNA' structures

Other DNA structures are other forms composed of the same nucleotides or molecules obtained from the double stranded structures. Most of these are synthetic products.

*Single-strand*ed nucleotide chains and helices exist. One way to make these is to unzip double helix.

Triple helix is formed from an A-DNA when it accommodates an extra strand in its major groove. This bonds by *Hoogsten* mechanism there, see Fig. 1.13. In Hoogsten and reverse Hoogsten base pairing N6 and N7 atoms of the ADE are hydrogen bonded.

Guanine may form *quadruplexes*, Fig. 1.14.

Interactions with other macromolecules produce largely differing structures. DNA may also be curved, bent, have mismatches, lone pairs, nicks and double strands may have *Holliday junctions*, Fig. 1.15. All these are properties of the structures, not own structures with current knowledge. These atom-level structures have higher level forms. E.g. the normal DNA may form plectonemes, and positively and negatively supercoiled (solenoidic) structures.

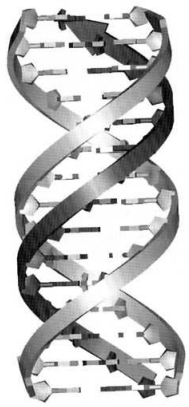

Figure 1.13: A triple helix [4, Fig. 4.7]

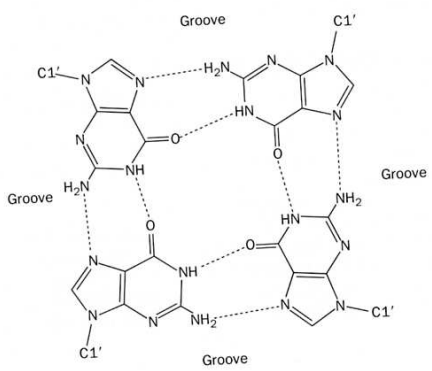

Figure 1.14: Schematic head view of the GUA quadruplex [4, Fig. 4.19]

Figure 1.15: Holliday junction between four strands [4, Fig. 4.33]. Interestingly it has the ability to slidi along the double stranded stuctures.

1.2.3 Tools to compute DNA parameters

There exists different software to compute various properties of the double stranded DNA. Some of the most common are given here:

- CURVES computes global helical axis optimizing it in to the underlaying atom position data[15, 16]. It produces both global and local parameters. This is the software used to compute the structural parameters in the works contained in this thesis.
- FREEHELIX, follower of NEWHELIX[17], creates optimal linear helical axis[18] in order to compute the structural parameters.
- RNA is a newer program based on the Tsukuba Accord[19, 20, 21]
- CEHS [22]
- CompDNA [23, 24]
- NUPARM [25]
- 3DNA [26]
- SCHNAaP [27]

1.3 Single Molecule Mechanical Manipulations

The year 1992 was a remarkable cornerstone in the era of the DNA research. Smith, Finzi, and Bustamante gave out their first publication of direct method to measure elasticity properties of a single DNA molecule [28]. Multimers of 48502 base pairs containing λ-DNA were stretched with forces between 0.01 pN and 10 pN. Four years later the same group and a French collaboration struck the pay-gravel [29, 30]. By still developed methods they were able to use forces up to 80 pN and 160 pN, respectively.

Both of these force ranges were wide enough to make it possible to observe a bizarre force-extension behaviour of the individual, bare DNA molecule. Experiments revealed a huge elongation caused by a few pN increase in pulling force after about 65 pN force was obtained. The relative length (r.l.) changed from 1.1 to 1.7.

The French collaboration also made molecular mechanical modelings by JUMNA [31], another reduced variable software. These simulations revealed a B -> S transition (S after Stretch) indicated by a reduced helical diameter and a strong base pair inclination. According to these modelings [32] base pairing and stacking reside at least to the relative length of 2.0.

Today, individual DNA-molecules can be handled mechanically by the receding meniscus (molecular combing) [33, 34], optical tweezers [35, 29, 36], magnetic tweezers [37], microneedles (microfibers) [30], atomic force microscopy (AFM) [38, 39] and using flow based [40, 41] or electrokinetic setups [42]. Ref. [43] gives a summary of many of these methods. One must also note that the purpose of the experiment - stretch, twist, unzip or a combination of these - determines the physical principles in the setup.

Fig.1.16 shows essential parts of the observed polymeric properties of double stranded DNA: Essentially the stretching curve for a freely unwinding DNA. The inset structure is JUMNA-modeled over-wound and stretched DNA [44]. It is nominated as P-DNA after the alike proposal to DNA structure by Linus Pauling in 1953 [45].

1.4 Physico-Chemical Properties of DNA

Chemically DNA is quite stable. It generally resists in acid and alkali solutions. However, in mild acid solutions - at pH 4 - purine bases (N7 of GUA, N3 of ADE) become protonated. These are good leaving groups for the hydrolysis. Once this happens, the depurinated sugar can easily isomerize into the open-chain form and in this form the depurinated DNA is susceptible to cleavage by hydroxyl ions. Enzymes (deoxyribonucleases) also hydrolyze DNA. They may digest a DNA strand from the end(s)

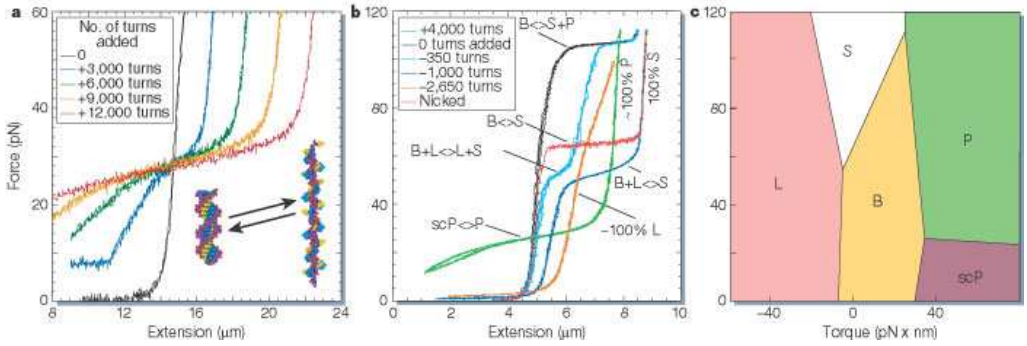

Figure 1.16: a) Force-extension curves for a 44.4-kilobase DNA molecule overwound by successively larger numbers of turns. The curves cross at an 'isosbestic point' marking the force at which B-DNA and P-DNA have equal extensions. b) Multiple plateaux occur in force-extension curves for torsionally constrained DNA. c) a 'phase diagram' for DNA under torque and tension. Coloured regions represent conditions under which pure phases occur; lines indicate conditions for phase coexistence within a molecule. S, overstretched; P, Pauling, sc, supercoiled and L for an average left-handed twist. [46, Fig. 3]

(exonucleases) or internally (endonucleases).

Physically the two polynucleotide chains of double-helical DNA can be separated under certain conditions, typically by raising the temperature. The transition from double-stranded DNA (dsDNA) to a single-stranded random coil DNA is called a *helix − coil* transition. Typical terms that describe the change from dsDNA to ssDNA are: *melting*, *denaturation*, and *strandseparation*. In the other direction, in the change from ssDNA to dsDNA they are *annealing*, *renaturation*, and, in certain contexts, *hybridization*.

The stability of dsDNA depends on a number of factors. These contain

- Base composition: The three hydrogen bonds containing G - C pairs need more energy to break than the A - T pairs. The same applies to stacking interactions between GC/GC and AT/AT base pairs.
- Salt concentration: Unless neutralised the close being p-groups repel each others. In cells neutralization is achieved through salt ions, polyamines and special DNA-binding proteins. When part of the DNA is in a single stranded form special proteins maintain it in this ('unstable') form.
- pH: Low pH i.e. acidic solutions denature DNA.

The denaturation of double stranded DNA is easily followed spectroscopically. The purine and pyrimidine bases of DNA absorb ultraviolet (UV) light maximally at a

wavelength near 260 nm. In double-stranded DNA, however, the absorption is decreased due to base-stacking interactions. When DNA is denatured, these interactions are disrupted and an increase in absorbance (hyperchromic effect) is seen. The extent of the effect can be monitored as a function of temperature.

The temperature at which the DNA molecules are 50% denatured is the melting temperature. DNA samples that are high in AT content have relatively lower melting temperatures while DNA samples that are high in GC content have relatively higher melting temperatures. Fig. 1.17 exemplifies this 'difficulty' in determining the 'exact' melting temperature of a double stranded DNA. Partial denaturation of DNA can be observed using electron microscopy. This has offered a means to find out A - T and G- C- rich regions in long DNA sequences.

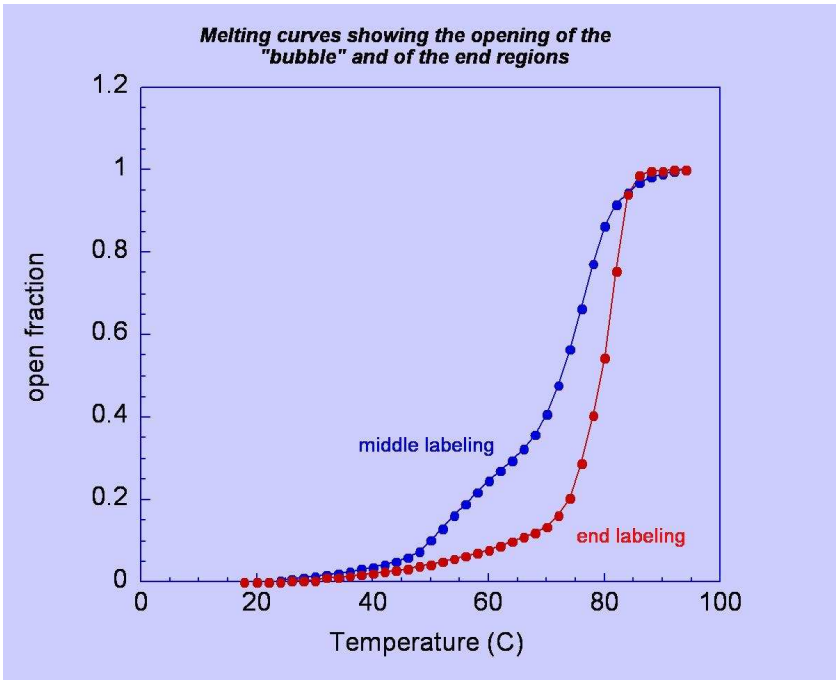

Figure 1.17: DNA's melting properties [47, 48]. 'Middle labeling' is in A - T rich region and 'end labeling' in G - C rich part of the hairpin DNA.

2 Dynamical Modeling of Molecules

Molecular Mechanics based *Molecular Dynamics* Simulation Method (MM-MD) is the central method used in this research. It has been used to produce the DNA data trajectories via computer runs. This data are dynamic. The ultimate goal has been to mimic DNA's internal evolution in full interaction in its natural environments as well as possible within the limits of the method and technical modeling facilities.

MM-MD is one of the many computer assisted modeling methods among the MD simulation methods family. In attempts to find the fundamental understanding in questions posed by Nature the quantum mechanical principles based *Ab Initio molecular dynamics* (AI-MD) would be the ultimate choice, but, so far, technical limitations hinder its use e.g. in molecular solution studies.

2.1 Basics

The physical roots of the MD methods are in the Newtonian Physics. The first one is Newton's second law of motion of mechanics. Using the language of Mathematics it is stated as

$$\vec{F}_{tot} = m\vec{a}. \tag{2.1}$$

Symbolism in Eqn. 2.1 glues the total force on a particle ($\vec{F}_{tot}$) to the product of its mass (m) and acceleration ($\vec{a}$). Vector notation '$\vec{}$' only means the need to take into account all independent dimensions of the space. The second principle is relation between the force and potential energy

$$\vec{F}_{tot} = -\vec{\nabla}U = -\left(\frac{\partial U}{\partial x}\hat{i} + \frac{\partial U}{\partial y}\hat{j} + \frac{\partial U}{\partial z}\hat{k}\right). \tag{2.2}$$

The $\frac{\partial}{\partial(x,y,z)}$ is the partial derivative operator and accent '$\hat{}$' denotes unit vector. The basic thought in MM-MD methods is that interaction between particles is expressed by a potential energy function. Force affecting on each particle is computed using Eqn. 2.2; Dynamics of each particle follows from Eqn. 2.1 by knowing that the acceleration of a particle is the first derivative of its velocity and second derivative of its position: $\vec{a} = \dot{\vec{v}} = \ddot{\vec{r}}$.

The Newtonian mechanics covers only part of the basics. Therefore the fundamentals are given more generally by Hamiltonian formalism of the general mechanics

$$H = E_k + U = \sum_i \frac{\vec{p}_i^2}{2m_i} + U(\vec{p}_i, \vec{r}_i), \tag{2.3}$$

where the Hamilton function (H) is sum of kinetic (E_k) and potential energy (U) contributions and $\vec{p}_i$ and $\vec{r}_i$ are the momentum and position vectors forming the phase space. Developing the equations of motion from 2.3 give [49, 50] the equations of motion

$$\begin{aligned} \frac{\partial H}{\partial \vec{p}_i} &= \frac{\vec{p}_i}{m_i} = \dot{\vec{r}}_i \\ \frac{\partial H}{\partial \vec{r}_i} &= -\dot{\vec{p}}_i = \vec{F}_i = -\vec{\nabla} U, \end{aligned} \tag{2.4}$$

for Cartesian coordinates. Here the conservation of forces $\vec{F}$, conservation of total energy E and form of the Hamiltonian H are not the limiting factors. These equations are time-reversible, a fact emphasized by some method developers [51, 52].

For all living organisms non-zero temperatures (in Kelvins) are important. In connection with MD methods the temperature of a particle follows the equipartition principle of statistical physics,

$$E_k^1 = \frac{1}{2} k_B T, \tag{2.5}$$

stating that the kinetic energy of one particle in one dimension equals to its thermodynamic energy.

2.2 MM-MD for Biomolecules

Currently typical biomolecular simulation system consists of target molecules and solution molecules. The mutual interaction between all these is in the potential energy function U, which traditionally in physics is called *potential energy* but in chemical sciences *force-field*. The latter is very understandable because of the 'field' all the molecules or atoms of the system create between themselves.

In biomolecular modeling or simulation field the basic force-field function for DNA (and proteins) is currently of the form

$$\begin{aligned} U = &\sum_{bonds} k_b (r - r_{eq})^2 + \sum_{angles} k_a (\theta - \theta_{eq})^2 + \sum_{torsions} k_t (1 - cos(m\phi))^2 + \\ &\sum_{i<j} \left\{ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\}. \end{aligned} \tag{2.6}$$

Components in this function describe bond stretching, angle bending, and rotation barriers around the bonds, the torsional angles. In addition, there are both short-range van der Waals and long range electrostatic interactions between the atoms. Covalent bonds are harmonic taking care of preservation of average distance between directly bonded atoms. Similarly angles in the central atom of three covalently bonded atoms have harmonic 'restriction' function. The torsion angle potential term is for the hindered rotation of groups about a bond. The nonbonded interactions are, of course, for the directly non-bonded but interacting atoms. The electrostatic term comes from the charged nature of the molecules.

The Lennard-Jones 12-6 potential in Eqn.(2.6) is due to its mathematical and computational convenience. Its cross-term parameters σ and ε are normally calculated by the Lorentz-Berthelot combination rules:

$$\varepsilon_{i,j} = (\varepsilon_i \varepsilon_j)^{1/2}, \ \sigma_{i,j} = (\sigma_i + \sigma_j)/2. \tag{2.7}$$

Atoms of a molecular system are non-bonded if they are really in different molecules, or in the same molecule with separation of three or more covalent bonds. Sometimes atoms separated exactly three covalent bonds are considered as '1-4' interactions. Of the other incredients the Urey-Bradley and improper torsion angle terms may be worth of mentioning. The former allows for a better fit to experimental data and the latter keeps planar species flat when sp and sp2 hybridized atoms are present.

Often force-fields also carry the names of the modeling software package, like CHARMM[1], AMBER[2], GROMOS[3] or OPLS[4], where it is as one of the core substances; And whose developers maintain and develop these force-fields.

The molecular MM-MD force-field consists of two important ingredients: the functional form and the associated parameter set. The latest larger step in the MM-MD force-field development for biomolecular simulations was made in 1995. The 'Second Generation Force Field's were published both for the CHARMM [53] and AMBER [54, 55] force-fields. This renewal was targeted especially on the parameters of Eqn. 2.6. Minor improvements have published later [56]. And, GROMOS got not until 2005 new nucleic acid parameters [57].

Water, which is one of the most typical solvent molecule, has many own force-field models. Important in the modeling is that the chosen water model produces essential properties in the modeling. Although there are many water models none of them are able to produce all the observable properties. In biomolecular and DNA context most

[1] Chemistry at HARvard Macromolecular Mechanics
[2] Assisted Model Building with Energy Refinement
[3] GROningen MOlecular Simulation
[4] Optimized Potential for Liquid Simulations

used models are the TIP3P[5] [58] and SPC[6] models. Currently the best resource of water knowledge and e.g. of its models is the web-site [59]. A recent dedicated reference is Ref. [60].

The important consequence of the second generation force-fields was that they made explicit solution modelings trustworthy.

2.3 MM-MD Practice

The roots of MM-MD are in the mechanical understanding of molecules as static structures. Then appeared the idea to put mechanics alive, in other words the MD. Consequence of this was that observations from these simulations should be comparable to experiments (obtained from multimolecule samples), which maybe led to theoretical idea that deterministic run in time has coupling with the environment where it occurs. This leads to considerations of statistical mechanics/physics *ensembles* and their changes when certain experimental observations and theoretically found results are to be compared.

Observations are the fundamental basics of the sciences verifying finally all features of the theories build with limited knowledge of underlying processes the simulations are one way to test the theories. Therefore, it has been important to get basis of the simulations to match with the experimental situations. The Newtonian mechanics based MM-MD preserves total energy of the simulation system. This is mentioned to be the link with realistic systems. Namely, the total energy is the quantity which also appears in statistical physical equations, which again describes the real world in which we make observations. Using statistical physics terminology the output from Newtonian MM-MD belongs to the NVE *ensemble*, i.e. number of particles, simulation volume and total energy are the constants of motion. Typically in experiments conditions are such that temperature or pressure can be kept constant and observations are made in NVT or NPT *ensembles*. Chemical potential, enthalpy, etc. may also be constants. In principle it is always possible to choose an *ensemble* that best desrcibes the experiments. In practice, the use of other *ensembles* that NVE leads to a modification of the Newtonian equation for the force, Eqn. 2.1. In Refs. [51] the equations of motions for the molecular dynamics simulations in other than NVE *ensemble* have been derived from the Hamiltonian principles. This is to say that various situations have differing set of difference equations governing just each *ensemble*. The MM-MD field knows many algorithms to generate the trajectories in their phase-spaces in correct *ensembles*, therefore I do not present here any details of those.

[5] Transferable Intermolecular Potentials

[6] Simple Point Charge

The phases of a MM-MD simulation can be divided in five parts: system generation, equilibration, production, analysis and reporting.

When a certain object in a solution or a whole system is thought to be modeled, the goals of this study determine the *ensemble* in which the system evolves. Varying *ensembles* may also be needed during one study, i.e. the *ensemble* is not necessarily fixed throughout the simulation. E.g. in the DNA stretching simulation of this thesis, NPT and NVT *ensembles* were used in the equilibration phase.

System generation means creation of the simulation system geometry: Choice of the simulation volume, whose shape is periodically repeatable, and positioning of the molecules and atoms in this volume. E.g. in the DNA simulation canonical B-DNA can be created by NAB[7] [61], filtered in a suitable format with parameters by VEGA [62], and these result can be incorporated in M.Dyna*Mix* as an initial data, which itself has a generator to create the solution molecules around it. Known structures are also available from the Nucleic Acid Database [6] and Cambridge Structural Database [7]. Initial velocities of the atoms can be taken according to the Maxwell distribution mimicking thermodynamic equilibrium state or set to zero and perform the thermalization this way. Simulation system is kept stationary by subtracting the induced momentum relatively from all particles momentum.

In the equilibration phase, the simulated system should obtain its equilibrium state. In principle this may contain three conditions: (i) all the "bonds" connecting the atoms fluctuate around their equilibrium values, (ii) the energy content of the systems' degrees of freedom is in a thermodynamical equilibrium with the environment and (iii) the system is in a chemical equilibrium with the environment . When simulating a macromolecule in a solution this would include build-up of the molecular environment, e.g. hydration layers around DNA. Depending on type of the simulation the equilibration phase uses tools by which the wanted result is obtained. Generally, equilibrium indicators can be taken from the physical fact that physical quantities normally approach their known equilibrium values. In the NVE or NVT simulation atomistic velocities are scaled after a certain amount of time-steps to get the correct total energy or temperature. The approach to the structural equilibrium is seen as stabilisation of the root mean squared deviation of the atoms. The density of the system can be modified by changing the volume (box or cell size). The equilibrium phase and the associated technical manipulations can be viewed as means to transfer the system from one state to another, after which the "Newtonian mechanics" is again operative.

After the equilibrium is obtained the production phase can be started, and relevant data collected for further analysis, if not analyzable during the simulation runs. Analysis and reporting are both in a close relationship with themselves but also with the

[7] Nucleic Acid Builder

production part. Namely, it is the phase when quite a lot of the obtainable results is determined, meaning that one has to know what one wants take out of the simulation or modeling effort. Biomolecular simulations at the classical level, not to mention the *ab initio* efforts, are so time consuming that becoming afterwards aware of a set of 'useful' facts, is too late. Of course everything is repeatable but weeks to months cpu-time may involve in practice much more wall-clock time. Other side is the available storage capacity. Gigabytes of space can easily be used. Nearly always a dynamical modeling production is composed of a series of runs. In order to make this possible all system data has to be saved in a suitable form to be used as the initial data for the next start.

What can we take out of the simulations? The all-embracing answer could be: all that your imagination produces and which is in the methodological limits of the simulation procedure and principles behind them. A good collection of what have been gained by the various simulations can be obtained by making a large review of all simulated systems. Unfortunately the amount of these is so vast that all finer details are never accessible, broadly speaking. Fortunately there is a 'standard' set of tools or principles, which are very often the adequate set to apply in each specific problem.

Root mean squared deviation (RMS/RMSD) is typically used in the equilibration phase because it measures the average arrival of atoms (molecules, etc.) to their balance positions in the system under the affecting global force-field. But, it also can be used in analysis purposes, e.g. indicating transitions in the modeled system or its sub-system, as was made in the longest DNA simulation of its time in 1994 [63]. Especially in molecular simulations two-dimensional RMS-maps are often used to analyze conformational changes and families of structures in a trajectory. In DNA suitable candidates for this would be bases, sugars and phosphates. Slope of the RMSD with time gives the diffusion constant e.g. for the water molecules in a DNA in solution simulation, when it only is calculated for these.

Computation on various correlations among the modeled system is one of the basic tools. Time correlation functions are important due to their relation to transport coefficients and spectra via linear response theory. Correlation quantities can be divided into the real space and reciprocal or inverse space quantities. Usual real space quantities are pair-correlation function or radial distribution function, $g(r)$. For it the volume size effects has to be kept in mind and make cut-off at a suitable radius. In the reciprocal space static structure factor, $S(k)$, can be computed. In practice, because $g(r)$ and $S(k)$ are in relation, and computation of the latter is faster, this way is used to compute $g(r)$. When other distributions are needed the same procedure can be used.

Visual inspections may include many kind of monitoring of the system. However, it is one of the best ways to make a state of the system concrete. In the equilibration phase

it may easily indicate possible serious problems in the modeling. In the production phase snapshots of the system give clear presentations of the system state. Recently developed three-dimensional techniques are very useful for molecular systems. Especially CAVE[8] like environments when system details during the dynamical evolution are of interest.

Special feature of the biomolecular force-field modelings is their inability to chemical reactivity, bonds fission and fusion. This precludes use of caloric curve, E(T), which in other type of simulations reveals phase-transitions.

Because molecular dynamics simulation method is completely deterministic sequential procedure, its results can be compared with experimental spectroscopic results. The space and time results are Fourier-transformed with respect to both quantities and result turns in reciprocal and wavevector space.

Simulations of this thesis have been performed by the molecular dynamics package M.Dyna*Mix*[9] [64]. M.Dyna*Mix* is a general purpose scalable parallel MD simulation package having the standard form of force-field as its core. It also has many additional potential forms to be used with the basic formula or as such.

M.Dyna*Mix* includes solution algorithms for the standard NVE MM-MD applying Verlet Leap-frog algorithm, NVT Nose-Hoover MM-MD, and NPT MM-MD. For the flexible molecular models the double time-step algorithm of Tuckerman et al. [51] is available. Its implementation is mentioned to be close the Martyna et al. [65] realization.

The use of the multistep algorithm is another method to make simulations faster. In this two-step procedure the fast intramolecular as well as the fast nearest short-range non-bonded interactions are calculated at each time step while the slow motions in the molecules and due to long-range non-bonded interactions are computed only at every long time step. Typical values for the short and long time steps in DNA like (biomolecular) systems are 0.2 fs and 2.0 fs.

M.Dyna*Mix* has two possibilities to handle the long-range electrostatic interactions; The Ewald method or the reaction field. In the Ewald method used in the DNA modeling of this work the Coulombic forces are divided into long-range and short-range components. The long-range part is calculated in reciprocal space. The short-range part is computed together with the Lennard-Jones forces.

M.Dyna*Mix* offers four simulation cell possibilities. Of these the hexagonal cell is used with the DNA simulations. Periodic boundary conditions and minimum image convention are used.

[8] Immersive virtual reality environment with environment control

[9] Molecular Dynamics of Mixtures

3 Simple Models

During late 1980's and early 1990's, intensive research on the dynamical properties of DNA by using simple non-linear models was carried out world-wide. Special emphasis was put on finding traces of *solitary excitations* relating transcription and replication of DNA as well as the connected mechanisms for DNA melting. Review in paper I is devoted to these nonlinear physics models.

Interestingly, one of these model has persisted until today. The Peyrard and Bishop initiated a model which was developed later also by Dauxois (the DPB model). It has attracted great attention among many researchers till these days. However, like the other very simplified models for flexible macromolecules, it has the feature of being a specific, event related, model.

We used the McMoldyn MM-MD-package [66] (predecessor of the M.Dyna*Mix*) in order to reveal these kind of excitations in DNA structure. At that time one time-step algorithms were in use. NVT ensemble and SHAKE [67] algorithm were employed. Simulations were typically tens of picoseconds long, in some cases few hundred picoseconds.

In these simulations the DNA in ionic water solution were first equilibrated as usual. Then initial excitation were created in the double strand, e.g. by turning one or both of the bases certain amount aside towards the minor and major grooves. When dynamical forcing had caused large enough effect the turning force was switched off and systems evolution was followed under the NVT conditions. The general result of these modelings was that base or both bases either were left in the positions where they were turned or they gradually turned back into their usual positions. Changes were very local.

Travelling vibrations were also tried to initiate kicking DNA backbone atoms while bases were turned, or pulled/pushed along base pair axis. No results in the structure were seen.

These attempts were time-consuming, and gave so little information that no published reports were ever written. Generally, only a very few works in these directions have been reported in the literature[68, 69, 70, 71]. After the born of new millennium base flipping has got attention a lot. The references [70, 72, 73, 74, 75, 76] form only a small subset of the all published.

4 Modeled Equilibrium of DNA

Natural starting point for a study of any system is the equilibrium state. What comes to the real DNA immediate question is what is its equilibrium state? Are the DNA samples found from frozen mammoth's remnants somewhere in tundra in equilibrium compared for example to fresh DNA samples from a human blood ? Naive maybe, but extremely physical answer is that both are in equilibrium in their status quo. In biological and living context the equilibrium state is intuitively anything but not well defined state due to the simple fact that generally thought everything is changing.

The other aspect is methodological. How do we know that a certain physical method, which should get the virtual object in a computer in its equilibrium, really does its task and correctly?

4.1 Equilibrium

In Paper II [77]) I have analyzed simulated DNA from the perspective of equilibrium properties. This analysis is both methodological testing and DNA sequence testing. Results are compared with the existing observed data and other recent simulations of DNA.

In that study decamer ATGCATGCAG in Li^+ - water solution were simulated in typical physiological conditions using the second-generation CHARMM force-field and M.Dyna*Mix* simulation software. Parameters of the modeling corresponded predominately a 'pre-transitional' region from the B-form to the C-form [78].

In the thermodynamically stable state around 300 K we tested what the kinetic part of Boltzmann's H function reveals of the equilibrium state. H is referred as a measure to find the turning point from equilibration to equilibrium [79].

The relevant equation in numerical treatment with underlying Maxwell distribution is

$$H_i = \frac{1}{2} + \sum_i \frac{N_i}{N} ln\left(\frac{N_i}{N}\right), \tag{4.1}$$

where N_i is the count number in velocity class i and 0 refers to zero-velocity class.

The effects of class width, cutting distribution tails, and overall quality of the distribution were examined.

In the case of atomistic velocity distributions of the MM-MD simulation the classification of velocities and H have a complicated behaviour. When the number of classes is large (narrow classes) H is clearly negative. When class width is increased, H gradually becomes positive, and approaches the zero limit of Maxwell distribution just at the limit of disappearing fluctuations in the distribution. When there are only three nonzero classes in the distribution, H begins again to increase.

Tests with the Gaussian distribution form showed that a truncated Gaussian form is easily recognizable with H values. Those show large deviations from the ideal value.

These H tests with momentary and only DNA atom data in use were not encouraging for continuous monitoring. Things can be made better by collecting the distribution from a certain amount of consecutive time steps (of an equilibrium modeling). However, large systems would be a more natural 'environment' for this indicator.

The structural stability was examined by the standard tool, RMS deviation measure, over all the DNA atoms. Fig. 2a of Ref. [77] shows the normal behaviour of it. When equilibrium has been reached the RMSD stops its increase. In Fig. 2b of Ref. [77] a test is made with respect to the 'initial' structure, i.e. the structure with respect to the comparison in RMSD is made. The chosen snapshot from equilibrium state dynamics gives evidence about the nature of the equilibrium state - a fluctuating state.

A partial structural analysis for the equilibrated DNA contained analysis of the conformational and helicoidal parameters both as distributions and dynamically. This analysis is based on the Cambridge Accord and Curves program [16] version 5.2. The conformational parameters show values typical both for the A- and B-forms of DNA. Individual distributions, composed often of two (or even three) subdistributions, show conformation parameters hoppings during the evolution. The positions of helicoidal parameter distributions show similar heterogeneity with respect to canonical values. Visually the distribution profiles of TIP, STR, ROL and TLT gave hints towards B-DNA.

The dynamical features of individual conformational parameters can be condensed into words: transient jumps, drift, stable and/or bi-stable. Basically the variable behaviour of these parameters comes from the energy landscape of the dynamical system. Conditions change and allow structural modifications when other than the existing potential well becomes possible and available. Clearly correlated motions seemed to have the ε - ζ and α - β -pairs.

The dynamical trajectories of the helicoidal parameters resembled random like motion. Within the time scale of this simulation (2.07 ns) few parameters showed strong drifting leading to skewed distribution form.

We studied occurrence of concerted motions in DNA by correlation analyses of the conformational and helicoidal parameters. Tables II - V of Ref. [77] show the most relevant correlations. The amount is not small. Comparison with the observed data revealed missing correlations in the modeling. Especially those lacking for the γ are striking. This comparison also gave structural tendency towards the B-form.

The foreground value of this DNA simulation is the *dynamical stability*, too often forgotten subject in scientific articles. Many sophisticated studies produce new 'excellent' structures but the fact remains that as often they are 'snaphots' of the existing forms. The 'observed' novelties here and there in details may be flaws of the measurement apparatus and sample preparations - in the same fashion as incomplete theories or computational force-fields produce their own results.

After the above manuscript was made new books have been appeared [80, 4, 81] and new studies have been published, see e.g. citations of Ref. [82].

5 Modeled Stretching of DNA

Papers III and IV contain results from the MM-MD modeling of a DNA stretching.

In this simulation we used 22mer, GTCTGAATTCTAATGTAGTATA, placed in a long hexagonal cell filled with water and Na^+ ions. With respect to the equilibrium state modeling slightly refined CHARMM force-field were used for the DNA [56]. For water we used the SPC model [83]. Sodium parameters were from Ref. [84].

In the beginning of the production run the O3' and O5' atoms at the other end of the helix were fixed in space while at the opposing end the same atoms were subjected to an external or "virtual" axially stretching force. This force was linearly coupled with time to guarantee continuous but gradual stretching. The technical details of the simulation can be found from paper IV.

The relative length (r.l.) vs. external force or time is shown in Fig. 1a of Ref. [85]. Experimentally the initially completely tension free DNA sequences have shown transitions around 1.65 r.l. when force is about 70 pN. Our simulation shows transition around 2.1 r.l. The force at this point is about 130 pN. However, one has to keep in mind that the force is the pulling force and obviously does not correspond to the forces in the experiments. These are made with very much slower pulling speeds offering DNA and its surrounding solution develop in a 'relaxed' way. The computationally found r.l. vs. force curve resembles surprisingly well the stretching experiment result from the early DNA research, from the year 1955, shown in Fig. 1 of paper IV.

The energetics of DNA within the force-field frames is illustrated in Fig. 1b of paper III. The 'spring loading' mechanism seen clearly in the integrated energy curve of bond, angle, and torsion angle components explains well the r.l. - force -curve. In the beginning, small force is enough to stretch DNA considerably until especially internal torsion angle contribution turns to very restrictive. When its effect ends, the harmonic bond and angle contributions are becoming more prominent keeping further extension hard. The backbone parameters α, γ, and the glycosidic torsion angle χ show structural changes at the turning point, see Figs. 2 b,c, and d of paper III.

A slight touch on the question of force-field validity in this kind of non-equilibrium modeling case has been made by evaluating the deviations of bonds angles in the extended state to their equilibrium dynamics values. Mean deviations are relatively quite small. But much remains to be done in the future to consider these limits

properly.

Because the stretching is a complicated process, its analysis requires various complementary visual tools. The first one is simply the visual inspection of the trajectory by using a suitable animation program. This can be done in 2D or 3D where the stereo-image 3D visualization naturally yields a better insight into the 3D dynamics. The most realistic insight can be gained by using the so-called immersive virtual reality (VR) environment where one can literally "walk inside the molecule". In this work, a one-wall VR environment and usual computer monitors with related softwares [86, 87, 88] have been used to visualize the 3D DNA. 3D animation revealed that the transition from the normal like DNA to the new base-stacked structure contains base pairs breakings, analogous to nucleation phenomena in e.g. solid state melting, trace of the much speculated S state [30] and temporary merging until the formation of the final structure begins. The found *base-stacked* form of double stranded DNA is not previously known by any means. Future scientific advances will show what kind of a role this or similar structure has in the basic biological processes. The current nanoscience and nanotechnology may find many uses for the forced forms besides the nano-springs [89].

6 Conclusions and Outlook

According to this work the simple Hamiltonian/Lagrangian physics models developed mainly in the non-linear theoretical physics to describe DNA's vibration dynamical features are able to give qualitative understanding of some of the observed properties in DNA's structure. The literature shows that many models have been created but they can be characterised best by the word 'proposal'. One exception is the model 'line' initiated by Peyrard and Bishop (paper [I]). It shows the richness these simple models may have.

The equilibrium state of a DNA decamer in explicit solution with counter-ions has been studied by the MM-MD modeling methodology (paper [II]). DNA's overall structural stability has been confirmed by different RMSDs. The conformational and helicoidal parameters have been analyzed statistically and dynamically. Dynamical analysis reveals the existence of dynamical sub-states, which typically appear as abrupt changes from a mean level to another in the value of parameter. In statistical analysis these sub-states cannot often be detected. Correlation analysis shows concerted motions among the many nearby structural parameters. The applicability of the kinetic part of the Boltzmann's H function as a measure of thermodynamic equilibrium was also tested. The mere DNA trajectory data proved to be statistically too small to find H as a good measure. However, for larger datasets it's suitability should be studied further.

The MM-MD modeling of the DNA stretching with the CHARMM force-field revealed a new structure for the DNA (paper[III,IV]). This forms from the B-DNA especially via the unwinding of the double helix. Relatively over about 1.7 times stretched DNA has a structure where the backbone strands are parallel and the bases are stacked on the major groove side. This is the first reported computational mechanical manipulation 'experiment' with all-atom force-fields, and having explicit solution. Methodological limits have been discussed. Our work is one of the first ones where the applicabability of force fields, that have originally been developed for reproducing known equilibrium structures, have been studied in a strongly non-equilibrium state. Some biological connections of the simulation results are also indicated in paper [IV].

What comes to the DNA and the knowledge of it and its place in biology, the main lines related to its structure and functions seems to be known. But what comes to its structural or nanoscopic details much remains to be revealed. The analysis of equilibrium dynamics of DNA revealed that the stability of macromolecules in their natural

environments does not mean the same as the stability of solids in their 'natural' environments, but includes even large transient changes in the molecules' conformation without any seemingly clear reasons. Based on these experiences of the equilibrium state of DNA, one may expect many 'not-so-evident' findings of DNA in the intermediate states when its goes through its life cycle and the related events with all the other molecules during the cell cycles.

The very big challenge for the physical theoretical sciences seems to be in developing methods which are not only for undertanding of the observed findings but are able to go beyond this and give predictions for the experiments and seeds to make tools for the applications. This creates a further need to develop both the few variable methods and the MM-MD touched in this compilation.

For the experimentalists the here observed stretched structure verification forms surely one challenge as well as the substructures along the path to this.

"Structure - function - genetics" is an old phrase but still the truth, giving us big challenges for a long time to come.

References

[1] C. R. Cantor and P. R. Schimmel. **Biophysical Chemistry, Part I: The Conformation of Biological Macromolecules**. W. H. Freeman and Company, 1980.

[2] C. R. Cantor and P. R. Schimmel. **Biophysical Chemistry, Part III: The Behaviour of Biological Macromolecules**. W. H. Freeman and Company, 1980.

[3] W. Saenger. **Principles of Nucleic Acid Structure**. Springer Advanced Texts in Chemistry. Springer-Verlag, 1984.

[4] S. Neidle. **Nucleic Acid Structure and Recognition**. Oxford University Press, 2002.

[5] **IUPAC Compendium of Chemical Terminology** . http://www.iupac.org/goldbook/, 1997.

[6] **Nucleic Acids Database**. http://ndbserver.rutgers.edu/.

[7] F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartright, A. Doubleday, and H. Higgs. **The Cambridge Crystallographic Data Centre: Computer Based Search, Retrieval, Analysis and Display of Information**. *Acta Crystallogr. Sect. B*, 35:2331–2339, 1979.

[8] W. K. Olson and P. J. Flory. **Steric Configurations of Polynucleotide Chains, I: Steric Interactions in Polynucleotides: A Virtual Bond Model**. *Biopolymers*, 11(1):1, 1972.

[9] J. Watson and F. Crick. **Molecular Structure of Nucleic Acids: A Structure of Deoxyribose Acid**. *Nature*, 171:737–738, 1953.

[10] F. H. C. Crick and J. D. Watson. **The Complementary Structure of Deoxyribonucleic Acid**. *Proc. Roy. Soc. (London) Ser. A*, 223:80–96, 1954.

[11] **MIT Biology Hypertextbook** . http://web.mit.edu/esgbio/www/7001main.html.

[12] R. E. Dickerson, M. Bansal, C. R. Calladine, S. Diekmann, W. N. Hunter, and O. Kennard. **Definitions and Nomenclature of Nucleic Acid Structure Parameters**. *EMBO J.*, 8:1–4, 1989.

[13] W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X.-J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C.-S. Tung, E. Westhof, C. Wolberger, and H. M. Berman. **A Standard Reference Frame for the Description of Nucleic Acid Base-pair Geometry**. *J. Mol. Biol.*, 313:229–237, 2001.

[14] R. Monasson. **DNA: Structure, Dynamics and Recognition**. In *Summer School: Multiple Aspects of DNA and RNA: from Biophysics to Bioinformatics*, Les Houches, France, 2-27 August 2004. Marie Curie Training Course of the European Union organised by D. Chatenay and S. Cocco and O. Krichevsky and R. Monasson and D. Thieffry.

[15] R. Lavery and H. Sklenar. **The Definition of Generalized Helicoidal Parameters and of Axis Curvature for Irregular Nucleic Acids**. *J. Biol. Struct. Dyn.*, 6:63, 1988.

[16] R. Lavery and H. Sklenar. **Defining the Structure of Irregular Nucleic Acids: Conventions and Principles**. *J. Biol. Struct. Dyn.*, 6:655, 1989.

[17] A. V. Fratini, M. L., Kopka, H. R. Drew, and R. E. Dickerson. **Reversible Bending and Helix Geometry in a B-DNA Dodecamer: CGCGAATTCGCG**. *J. Biol. Chem.*, 257:14686–14707, 1982.

[18] R. E. Dickerson. **DNA Bending: the Prevalence of Kinkiness and the Virtues of Normality**. *Nucl. Acids Res.*, 26:1906–1926, 1998.

[19] M. S. Babcock, E. P. D. Pednault, and W. K. Olson. **Nucleic Acid Structure Analysis: A Users Guide to a Collection of New Analysis Programs**. *J. Biomol. Struct. Dyn.*, 11:597–628, 1993.

[20] M. S. Babcock, E. P. D. Pednault, and W. K. Olson. **Nucleic Acid Structure Analysis. Mathematics for Local Cartesian and Helical Structure Parameters that are Truly Comparable Between Structures**. *J. Mol. Biol.*, 237:125–156, 1994.

[21] M. S. Babcock and W. K. Olson. **The Effect of Mathematics and Coordinate System on Comparability and Dependencies of Nucleic Acid Structure Parameters**. *J. Mol. Biol.*, 237:98–124, 1994.

[22] M. A. El Hassan and C. R. Calladine. **The Assessment of the Geometry of Dinucleotide Steps in 236 Standard Nucleic Acid Reference Frame Double-Helical DNA: A New Local Calculation Scheme with an Appendix**. *J. Mol. Biol.*, 251:648–664, 1995.

[23] A. A. Gorin, V. B. Zhurkin, and W. K. Olson. **B-DNA Twisting Correlates with Base-Pair Morphology**. *J. Mol. Biol.*, 247:34–48, 1995.

[24] K. M. Kosikov and A. A. Gorin V. B. Zhurkin anb W. K. Olson. **DNA Stretching and Compression: Large-scale Simulations of Double Helical Structures**. *J. Mol. Biol.*, 289:1301–1326, 1999.

[25] M. Bansal, D. Bhattacharyya, and B. Ravi. **NUPARM and NUCGEN: Software for Analysis and Generation of Sequence Dependent Nucleic Acid Structures**. *Comput. Appl. Biosci.*, 11:281–287, 1995.

[26] X.-J. Lu and W. K. Olson. **3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-dimensional Nucleic Acid Structures**. *Nucl. Acids Res.*, 31:5108–5121, 2003.

[27] X.-J. Lu, M. A. El Hassan, and C. A. Hunter. **Structure and Conformation of Helical Nucleic Acids: Analysis Program (SCHNAaP)**. *J. Mol. Biol.*, 273:668–680, 1997.

[28] S.B. Smith, L. Finzi, and C. Bustamante. **Direct Mechanical Measurements of Elasticity of Single DNA Molecules Using Magnetic Beads**. *Science*, 258:1122–1126, 1992.

[29] S.B. Smith, Y. Cui, and C. Bustamante. **Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules**. *Science*, 271:795–799, 1996.

[30] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, and F. Caron. **DNA: An Extensible Molecule**. *Science*, 271:795–799, 1996.

[31] R. Lavery, K. Zakrzewska, and H. Sklenar. **JUMNA: Junction Minimization of Nucleic Acids**. *Comp. Phys. Comm.*, 91:135:158, 1995.

[32] A. Lebrun and R. Lavery. **Modelling Extreme Stretching of DNA**. *Nucl. Acids Res.*, 24:2260–2267, 1996.

[33] A. Bensimon, A. Simon, A. Chiffaudel, V. Croquette, F. Heslot, and D. Bensimon. **Alignment and Sensitive Detection of DNA by Moving Interface**. *Science*, 265:2096–2098, 1994.

[34] D. Bensimon, A. Simon, V. Croquette, and A. Bensimon. **Stretching DNA with a Receding Meniscus: Experiment and Models**. *Phys. Rev. Lett.*, 74:4754–4757, 1995.

[35] A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and S. Chu. **Observation of a Single-Beam Gradient Force Optical Trap for Dielectric Particles**. *Opt. Lett.*, 11:288, 1986.

[36] C. G. Baumann, V. A. Bloomfield, S. B. Smith, C. Bustamante, M. D. Wang, and S. M. Block. **Stretching of Single Collapsed DNA Molecules**. *Biophys. J.*, 78:1965–1978, 2000.

[37] C. Gosse and V. Croquette. **Magnetic Tweezers: Micromanipulation and Force Measurements at the Molecular Level**. *Biophys. J.*, 82:3314–3329, 2002.

[38] T. Thundat, D.P. Allison, and R.J. Warmack. **Stretched DNA Structures Observed with Atomic Force Microscopy**. *Nucl. Acid Res.*, 22:4224–4228, 1994.

[39] C. Stroh, H. Wang, R. Bash, B. Ashcroft, J. Nelson, H. Gruber, D. Lohr, S. M. Lindsay, and P. Hinterdorfer. **Single-Molecule Recognition Imaging Microscopy**. *Proc. Natl. Acad. Sci. USA*, 101:12505, 2004.

[40] T. T. Perkins, D. E. Smith, and S. Chu. **Single Polymer Dynamics in an Elongational Flow**. *Science*, 276:2016, 1997.

[41] E. Y. Chan, N. M. Goncalves, R. A. Haeusler, A. J. Hatch, J. W. Larson, A. M. Maletta, G. R. Yantz, E. D. Carstea, M. Fuchs, G. G. Wong, S. R. Gullans, and R. Gilmanshin. **DNA Mapping Using Microfluidic Stretching and Single-Molecule Detection of Fluorescent Site-Specific Tags**. *Genome Res.*, 14:1137, 2004.

[42] N. Kaji, M. Ueda, and Y. Baba. **Molecular Stretching of Long DNA in Agarose Gel Using Alternating Current Electric Fields**. *Biophys. J.*, 82:335, 2002.

[43] U. Bockelmann. **Single-Molecule Manipulation of Nucleic Acids**. *Curr. Opin. Struc. Biol.*, 14:368–373, 2004.

[44] J.F. Allemand, D. Bensimon, R. Lavery, and V. Croquette. **Stretched and Overwound DNA Forms a Pauling-like Structure with Exposed Bases**. *Proc. Natl. Acad. Sci. USA*, 95:14152–14157, 1998.

[45] L. Pauling and R. B. Corey. **A Proposed Structure for the Nucleic Acids**. *Proc. Natl Acad. Sci. USA*, 39:84–97, 1953.

[46] C. Bustamante, Z. Bryant, and S.B. Smith. **Ten Years of Tension: Single-Molecule DNA Mechanics**. *Nature*, 421:423–427, 2003.

[47] O. Krichevsky. **Fluorescence Correlation Spectroscopy and Its Applications to DNA Dynamics**. In *Summer School: Multiple Aspects of DNA and RNA: from Biophysics to Bioinformatics*, Les Houches, France, 2-27 August 2004. Marie Curie Training Course of the European Union organised by D. Chatenay and S. Cocco and O. Krichevsky and R. Monasson and D. Thieffry.

[48] G. Altan-Bónnet, A. Libchaber, and O. Krichevsky. **Bubble Dynamics in Double-Stranded DNA**. *Phys. Rev. Lett.*, 90:138101, 2003.

[49] H. Goldstein. **Classical Mechanics**. Pearson Higher Education, second edition, 1980.

[50] M. P. Allen and D. J. Tildesley. **Computer Simulations of Liquids**. Clarendon, Oxford, 1987.

[51] M. Tuckerman and B. J. Berne. **Reversible Multiple Time Scale Molecular-Dynamics**. *J. Chem. Phys.*, 97:1990, 1992.

[52] M. E. Tuckerman and G. J. Martyna. **Understanding Modern Molecular Dynamics: Techniques and Applications**. *J. Phys. Chem. B*, 104:159–178, 2000.

[53] A. D. MacKerell, J. Wiorkiewicz-Kuczera, and M. Karplus. **An All-Atom Empirical Energy Function for the Simulation of Nucleic Acids**. *J. Am. Chem. Soc.*, 117:11946, 1995.

[54] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson Jr., D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. **A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules**. *J. Am. Chem. Soc.*, 117:5179, 1995.

[55] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould K. M. Merz, D. M. Ferguson, D.C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. **A second generation force field for the simulation of proteins, nucleic acids, and organic molecules**. *J. Am. Chem. Soc.*, 118:2309, 1996.

[56] N. Foloppe and A. D. MacKerell Jr. **All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimisation based on Small Molecule and Condensed Phase Macromolecular Target Data**. *J. Comp. Chem.*, 21:86, 2000.

[57] T. A. Soares, P. H. Hünenberger, M. A. Kastenholz, V. Kräutler, T. Lenz, R. D. Lins, C. Oostenbrink, and W. F. Van Gunsteren. **An Improved Nucleic Acid Parameter Set for the GROMOS Force Field**. *J. Comp. Chem.*, 26:725–737, 2005.

[58] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. **Comparison of Simple Potential Functions for Simulating Liquid Water**. *J. Chem. Phys.*, 79:926–935, 1983.

[59] Martin Chaplin. **Water Structure and Behaviour** . http:/www.lsbu.ac.uk/water/index.html.

[60] W. L. Jorgensen and J. Tirado-Rives. **Potential Energy Functions for Atomic-level Simulations of Water and Organic and Biomolecular Systems**. *P. Natl. Acad. Sci. USA*, 102:6665, 2005.

[61] T. Macke and D.A. Case. **Molecular Modeling of Nucleic Acids**, pages 379–393. Americal Chemical Society, 1998.

[62] **VEGA**. http:/users.unimi.it/ddl/vega/index2.html.

[63] K. J. McConnell, R. Nirmala, M. A. Young, G. Ravishanker, and D. L. Beveridge. **A Nanosecond Molecular Dynamics Trajectory for a B-DNA Double Helix: Evidence for Substates**. *J. Am. Chem. Soc.*, 116:4461, 1994.

[64] A. P. Lyubartsev and A. Laaksonen. **Molecular Dynamics Simulations of DNA in Solution with Different Counter-ions**. *J. Biomol. Struct. Dyn.*, 16:579, 1998.

[65] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein. **Explicit Reversible Integrators for Extended Systems Dynamics**. *Mol. Phys. 87*, page 1117, 1996.

[66] A. Laaksonen. **Computer-Simulation Package for Liquids and Solids with Polar Interactions .I. McMoldyn** H_20 **- Aqueous Systems**. *Comp. Phys. Comm.*, 42:271–300, 1986.

[67] J. P. Ryckaert. **Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes**. *J. Comp. Phys.*, 23:327, 1977.

[68] Y. Z. Chen, V. Mohan, and R. H. Griffey. **Effect of Backbone Zeta Torsion Angle on Low Energy Single Base Opening in B-DNA Crystal Structures** . *Chem. Phys. Lett.*, 287:570–574, 1998.

[69] Y. Z. Chen, V. Mohan, and R. H. Griffey. **The Opening of a Single Base without Perturbations of Neighboring Nucleotides: A Study on Crystal B-DNA Duplex** $d(CGCGAATTCGCG)_(2)$. *J. Biol. Str. Dyn.*, 15:765, 1998.

[70] Y. Z. Chen, V. Mohan, and R. H. Griffey. **Spontaneous base flipping in DNA and its possible role in methyltransferase binding** . *Phys. Rev. E*, 62:1133–1137, 2000.

[71] Y. Z. Chen, V. Mohan, and R. H. Griffey. **Base opening in RNA and DNA duplexes: Implication for RNA stability**. *Phys. Rev. E*, 61:5640–5645, 2000.

[72] E. Giudice, P. Varnai, and R. Lavery. **Energetic and Conformational Aspects of A:T Base-pair Opening within the DNA Double Helix** . *ChemPhysChem*, 2:673, 2001.

[73] N. K. Banavali and A. D. MacKerell. **Free Energy and Structural Pathways of Base Flipping in a DNA GCGC Containing Sequence**. *J. Mol. Biol.*, 319:141–160, 2002.

[74] N. Huang and A. D. MacKerell. **Atomistic view of base flipping in DNA**. *Philos. T. Roy. Soc. A*, 362:1439–1460, 2004.

[75] X. L. Cheng, C. Kelso, V. Hornak, C. de los Santos, A. P. Grollman, and C. Simmerling. **Dynamic Behavior of DNA Base Pairs Containing 8-Oxoguanine**. *J. Am. Chem. Soc.*, 127:13906–13918, 2005.

[76] D. Coman and I. M. Russu. **A Nuclear Magnetic Resonance Investigation of the Energetics of Basepair Opening Pathways in DNA**. *Biophys. J.*, 89:3285–3292, 2005.

[77] R. Lohikoski, J. Timonen, A. Lyubartsev, and A. Laaksonen. **Internal Structure and Dynamics of the Decamer** $d(ATGCAGTCAG)_2$ **in** $Li^+ - H_2O$ **Solution: A Molecular Dynamics Simulation Study**. *Mol. Sim.*, 29:47, 2003.

[78] S. A. Lee, S. M. Lindsay, J. W. Powell, T. Weidlich, N. J. Tao, G. D. Lewen, and A. Rupprecht. **A Brillouin-scattering study of the hydration of Li-DNA and Na-DNA films**. *Biopolymers*, 26:1637, 1987.

[79] J. M. Haile. **Molecular Dynamics Simulation**. John Wiley & Sons, New York, 1992.

[80] S. Neidle, editor. **Oxford Handbook of Nucleic Acid Structure**. Oxford University Press, 1999.

[81] T. Schlick. **Molecular Modeling and Simulation, An Interdisciplionary Guide**. Interdisciplinary Applied Mathematics: Mathematical Biology. Springer, 2002.

[82] A. D. MacKerell Jr. **Empirical Force Fields for Biological Macromolecules: Overview and Issues**. *J. Comp. Chem.*, 25:1584–1604, 2004.

[83] K. Toukan and A. Rahman. **Molecular-dynamics study of atomic motions in water**. *Phys. Rev. B*, 31:2643, 1985.

[84] K. Heinzinger. **Computer simulations of aqueous electrolyte solutions**. *Physica B*, 131:196–216, 1985.

[85] R. Lohikoski, J. Timonen, and A. Laaksonen. **Molecular Dynamics Simulation of Single DNA Stretching Reveals a Novel Structure**. *Chem. Phys. Lett.*, 407:23, 2005.

[86] L. Laaksonen. **A Graphics Program for the Analysis and Display of Molecular Dynamics Trajectories**. *J. Mol. Graph.*, 10:33–34, 1992.

[87] D. L. Bergman, L. Laaksonen, and A. Laaksonen. **Visualization of Solvation Structures in Liquid Mixtures**. *J. Mol. Graph. Model.*, 15:301–306, 1997.

[88] W. Humphrey, A. Dalke, and K. Schulten. **VMD – Visual Molecular Dynamics**. *Journal of Molecular Graphics*, 14:33–38, 1996.

[89] F. C. Simmel and W. U. Dittmer. **DNA Nanodevices**. *SMALL*, 1:284–299, 2005.