UPTEC X07 011
Examensarbete 20 p
Feb 2007

A study of DNA base dimers using quantum mechanical ab initio methods

Göran Wallin

UPPSALA UNIVERSITET

Molecular Biotechnology Programme

Uppsala University School of Engineering

UPTEC X 07 011	**Date of issue 2007-01**

Author

Göran Wallin

Title

A study of DNA base dimers using quantum mechanical *ab initio* methods

Abstract

Ab initio quantum chemical calculations have been performed on DNA nucleobases adenine, guanine, cytosine and thymine, together with stacked dimers of guanine and cytosine to determine the effect of destacking upon rotation about the helical axis. The predicted geometries and orbital energies of the single bases were seen to correspond well with respect to experimental photoelectron spectra and averaged high resolution X-ray crystallographic structures. The potential energy along the helical rotation of the dimers showed qualitative agreement with observed B-DNA structures, where the totally stacked geometries corresponded to energetically unfavourable conformations in every studied case, with associated energy lowerings upon destacking in the order of 100-200 meV per dimer. Disagreements with the predicted dipole-dipole potential energies suggest that these lowerings may be mainly quantum mechanical in nature, and consequently caused by electron correlation and mechanisms similar to the *pseudo Jahn-Teller* effect.

Keywords

Base stacking, π-π interaction, *ab initio*, MP2, DNA

Supervisors

Prof. Leif Karlsson
Dept. of Physics, Uppsala University

Scientific reviewer

Prof. Sten Lunell
Dept. of Quantum Chemistry, Uppsala University

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **82**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

A study of DNA base dimers using quantum mechanical *ab initio* methods

Göran Wallin

Sammanfattning

Den genetiska informationen i våra celler lagras i DNA-molekylen — en spiralformad polymer som består av en socker-fosfatryggrad och fyra olika baser, adenin, guanin, cytosin och tymin.

Fokus med det här arbetet har varit att studera den direkta interaktionen mellan olika kombinationer av baserna guanin och cytosin, som de uppträder i DNA, för att se om den kvantmekaniska orbitalinteraktionen kan vara en bidragande orsak till dess karakteristiska spiralform.

Resultatet av beräkningarna visade att de här baserna gärna ställer sig i den formation vi hittar i biologiskt DNA på grund av huvudsakligen kvantmekaniska effekter.

Även om vägen är lång tills vi kan beskriva DNA i sin helhet med kvantmekaniska metoder, och det här arbetet bara skrapar på ytan till problemet, har de här resultaten antydit att kvantmekaniken kan avslöja hittills okända mekanismer bakom DNA-molekylens struktur.

Examensarbete 20 p i Molekylär bioteknikprogrammet

Uppsala Universitet, Februari 2007

Contents

Acknowledgements

I would like to thank my supervisor professor **Leif Karlsson** for suggesting the importance of the orbital interactions in the DNA double helix which culminated in this master's thesis, and for helping me see this through, giving advice and encouragment when needed. *Thank God for professors!*

I would also like to thank my scientific reviewer professor **Sten Lunell**, at Quantum Chemistry, for help and guiding lines on everything from the quantum mechanical description of dispersion forces to the incompleteness of the Gaussian basis sets, as well as for the helpful comments on this document.

Thank you Dr. **Raimund Feifel** for the encouragement, the discussions and for the reviewal of this document.

None of this would have been possible if it were not for **Martin Agback**, also at Quantum Chemistry, to whom I am greatly indebted for helping me with the many hardware, `Linux` and `Gaussian` issues during this project.

I am very grateful for the computing time at the `Ra` opteron cluster[1], which was graciously made available through the *Uppsala Multidisciplinary Center for Advanced Computational Science* (UPPMAX)[2]. Thank you **Jukka Kominaho**, at UPPMAX, for helping me sort out the `Sun Grid Engine` issues.

Further **Torbjörn Rander** and **Andreas Lindblad** for providing a computing environment during the initial trials and timings.

Also **Lage Hedin**, **Kjell Pernestål**, **Tommy Andersson**, **Sonja Droschke** and all others at the Physics Department for making this such an enjoyable experience.

I would also like to thank my family and **Marie**, for supporting me throughout this project.

[1] `http://www.uppmax.uu.se/systems/ra/ra-cluster`.
[2] See `http://www.uppmax.uu.se/`.

1 Introduction

It has hardly escaped anyone's notice that the bearer of genetical information in all living beings, *the DNA molecule*, is helical. However, what has eluded severe scientific scrutiny, essentially since its revelation in 1953 by Rosalind Franklin [22], James Watson and Francis Crick [63], are consistent probable causes behind this preferred helicity.

The reason for this may lie in the fundamental quantum mechanical nature of DNA, which has been, and still is, intractable to probe in its entirety.

Although the grand question is currently beyond the cababilities of our finest computer systems, and indeed the scope of a master's thesis, the helicity of DNA will be studied here on a dimeric nucleobase level using quantum mechanical methods.

Also, the methodology is developed in such a way that the problem can be scaled up for future studies, where we have in mind that since the processing capabilities of computers effectively doubles every second year [40], even the grand problem may become solvable within a not too distant future.

1.1 Purpose

The purpose of this thesis is to verify what conformations stacked dimers of the nucleobases guanine and cytosine will prefer when subjected to a rotation about the B-DNA helical axis, using a purely quantum mechanical treatment.

1.2 Outline

The strategy that will be followed to fulfill this purpose can be subdivided into the following four steps.

i. Obtain quantitatively accurate quantum mechanical descriptions of the DNA bases adenine, guanine, cytosine and thymine as compared with experimental results.

ii. Subject dimers of the guanine and cytosine bases from these descriptions to a rotation about the helical axis.

iii. Calculate the quantum mechanical potential curves along the path of these transformations.

iv. Compare these potential curves with X-ray crystallographic data on B-DNA.

In this report, the realization of this strategy is described within the following main subjects.

Background. Since the model only encompasses the bases themselves, this implies a rather drastic simplification of the grand problem. To remedy this, the expected impact of the factors that are not dealt with in this model system is briefly described in the background.

In addition, to be able to rotate about the helical axis, it is required that this axis is defined. For this purpose *the Standard Reference Frame* of helical coordinates [41] will be followed, and is consequently also reviewed.

Theory. In this chapter, the underlying equations behind the most important quantum mechanical methods are derived, along with a few interpretations and notes. For swift examination, a non-technical abridgement is also available at the end of the chapter.

Methods and Results. The methods and results have been separated into two chapters, the first describing the implementation of the Standard Reference Frame, with the modest statistical analysis of X-ray crystal structures that ensues, and the second describing the actual computations.

At the beginning of the computational chapter, the results on the model system Ne_2 is presented, which is subsequently used to evaluate to what extent the quantum mechanical *electron correlation* describes the chemical *van der Waals* interaction for second row atoms. Here, we remember that rare gases are by definition non-interacting, save for dispersion attraction, which is the actual motivation for this choice of model system together with the abundance of available experimental data from cluster experiments, as introduced in [32].

The remainder of the chapter is devoted to the computations on the individual bases, which are compared with experimental results from photoelectron spectroscopy recently obtained by this group [60], followed by the dimer calculations and their resulting potential curves.

Discussion. In the discussion, the potential curves are interpreted vis-à-vis electrostatics, electron correlation and finally the direct orbital interaction as given by the stationary non-correlating state.

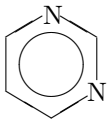
While having in mind that this particular model system is expected to describe single stranded DNA, a comparison with the dinucleotide data from double stranded B-DNA, as given by the modest statistical analysis, is also made.

At the end of the discussion, the conclusions that are drawn from these considerations are summarized, followed by suggestions on future improvements.

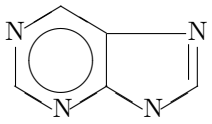
2 Background

> Wir müssen wissen,
> wir werden wissen.
>
> David Hilbert

Deoxyribonucleic acid (DNA) is made up of a sugar-phosphate backbone onto which nucleic acids are affixed. There are four kinds of DNA bases, namely adenine, guanine, cytosine and thymine, where adenine and guanine are purine derivatives, whereas cytosine and thymine are substituted pyrimidines.

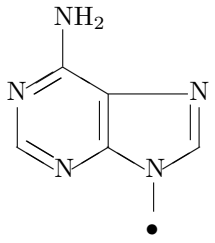


Pyrimidine

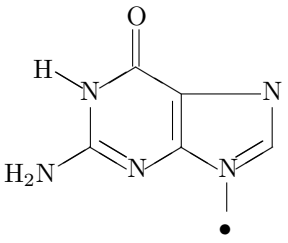


Purine

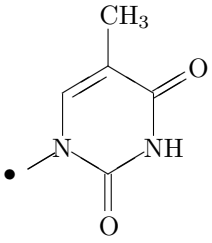
The chemical structures of the DNA bases are shown below, where their connection to the backbone deoxyribose is marked by a bullet (•).



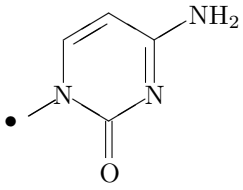
Adenine



Guanine

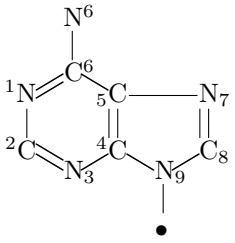


Thymine

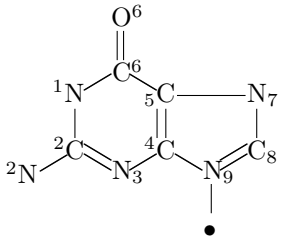


Cytosine

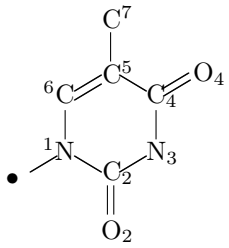
The atomic numbering of these bases will here follow the ordinary IUPAC nomenclature [20] as given below[1],



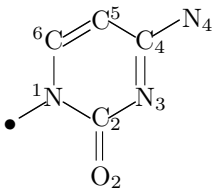
Adenine



Guanine



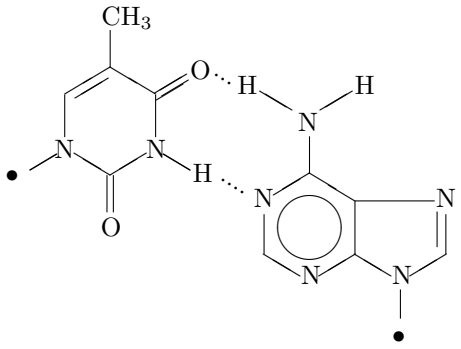
Thymine



Cytosine

By convention, the orientation of the DNA polymer is in the direction of the 5′ to the 3′ carbon of the sugar, usually written $5' \rightarrow 3'$.

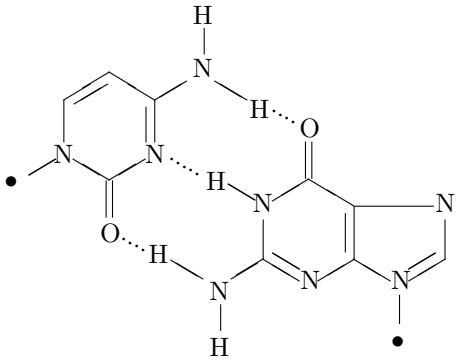
Chemical formulae describing the Watson-Crick base pairing [63, 22] are shown below. We notice that whereas the G–C base pair has three hydrogen bonds, the A–T only harbors two[2].



Thymine – Adenine

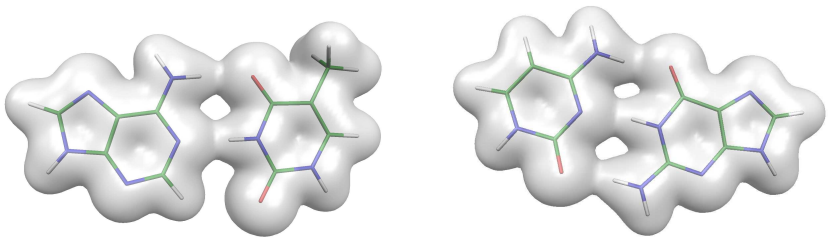
[1] Note that the subscripts and superscripts are merely for clarity in the chemical formulae. The atoms are referred to by the chemical symbol followed by a number, e.g. N6 for the sixth nitrogen.

[2] Interestingly, Watson and Crick themselves originally only envisaged two hydrogen bonds between guanine and cytosine [62].



Cytosine – Guanine

By using the molecular quantum mechanical techniques that will be outlined later on, the result of a rather modest calculation will already confirm this hydrogen bonding scheme. Specifically, figures 2.1(a) and 2.1(b) show the total electron density at the Hartree-Fock level of the two Watson-Crick basepairs.



(a) Adenine-Thymine basepair (b) Cytosine-Guanine basepair

FIGURE 2.1: Electron probability surfaces for the Watson-Crick basepairs calculated with modest *ab inito* methods (HF/6-31G). The calculations confirm the predicted hydrogen bonds between the molecules, two for the A-T and three for the G-C basepair. The near transparent surfaces represent the locations where the probability of finding electrons is $p = 0.01$.

There are three major families of DNA, namely A- , B- [23] and Z-DNA [61] that will, at least in some cases, reversibly transform into each other given certain circumstances.

Actually, the first two single crystal X-ray structures of DNA at atomic resolution were Z-DNA [61, 15], whereas the first full turn B-DNA was announced in 1980 by Richard Dickerson and his group [14]. It held the twelve-nucleotide sequence $5'-$CGCGAATTCGCG$-3'$ and was hence termed *the Dickerson dodecamer*. This article was followed by a comprehensive study on that structure [17], the specific influence of the base sequence on the helix conformation [12] and the details of the water geometry[13].

Many single crystal structures of DNA have been resolved since then, and in much the same way that the *Brookhaven Protein Data Bank* (PDB) [6] is the central repository for protein structures, the *Nucleic Acids Database* (NDB) [5] currently contains over 3000 nucleic acids structures.

2.1 DNA Conformation

We can outline at least five major contributing factors to the conformation of the double helix.

i. The base sequence and vertical base interactions.

ii. The sugar-phosphate backbone.

iii. Hydration and counter-ions.

iv. Temperature.

v. Crystal packing effects and lattice interactions.

Crystal packing Beginning with the least significant factor, the dilemma of whether we are seeing the overall optimal conformation in a crystal structure or simply an artificial local minimum due to intermolecular interaction within the crystal lattice is termed *the tyranny of the lattice*, or the effect of crystal packing.

A way of gauging such effects is by crystallizing the same sequence in different space groups, and by such methods it has been shown that the crystal environment will indeed have a significant impact on the DNA conformation, but more so on a global rather than local scale, primarily affecting DNA bending [10], although this view has been challenged [31].

Temperature The resolution of identical nucleic acids structures in temperatures ranging from $16\,K$ to room temperature reveals a slight impact of thermal motion on DNA conformation but much less so than from structures in solution [16].

Hydration and Counter-ions Cations are mostly seen to interact with the negatively charged sugar-phosphate backbone or to coordinate water molecules [56] perhaps even in the minor groove [52], and probably alleviate the charge repulsion between the two strands, hence bringing them closer together and permitting conformational states that otherwise would have been energetically unfavourable. The most prominent example is perhaps the previously mentioned Z-DNA conformation that arises in high salt solutions of $(\mathrm{CG})_n$ sequence motifs *in vitro* [61] and presumably by negative torsional strain *in vivo* [64].

Water molecules are often seen coordinated in *the spine of hydration* that runs within the minor groove of B-DNA [13]. It is composed of four layers shaped in an hexagonal pattern [52, 53], which in its lowest layer interacts directly with the bases, specifically the N2 or H2 atoms of guanine, the O2 atom of cytosine and thymine and the N3 atom of adenine, as reported already in the Dickerson dodecamer [17].

Hypothetically, *water bridges* — that reach between the strands and consequently interconnect them — may be an appropriate metaphor to describe the role of water in the minor groove [13], but recent neutron diffraction experiments reveal a much more complicated picture [2], which may be the subject of future *ab initio* investigations.

The Sugar-phosphate Backbone Relatively wide ranges of backbone conformations are accessible in DNA [50] but the internal sugar-phosphate coordinates were early known to show statistical bivariate correlation [18, 24]. More recently, a number of multivariate statistical analyses have shown significant clustering of DNA subclasses [4, 54] and are often accompanied with molecular dynamics or Monte Carlo calculations [19, 44].

Although a few important degrees of freedom of the sugar-phosphate backbone are consistently revealed in such analyses, the precise causal relations remain somewhat elusive. Nevertheless, one of these studies provides us with a counter-hypothesis, which is that

> Twist is insensitive to the base stacking interactions and is determined solely by the constraints of a relatively rigid fixed length backbone.

as quoted from Packer *et al.* [44], a point of view which is further elaborated in [42] and [43].

The Base Sequence and Vertical Base Interactions Vertical sequence specific base interactions, or *base stacking*, is the focus of this report, and although plentifully reported elsewhere, e.g. [26], a modest statistical analysis of a dinucleotide dataset derived from NDB will be performed. This will subsequently serve as a starting point for a base stacking energy calculation of a few bases with *ab initio* methods.

2.2 Energies and Conformations

At this point, it may be worthwile to mention that if a certain conformation is energetically favourable, it is by the *Boltzmann statistics* also more likely to occur. In that sense, the calculation of energies can provide an explanation as to why certain structures are preferred over others.

There are two main ways to do this, and those are either by classical or quantum mechanics.

Classical Mechanics At the center of molecular energetics calculations by classical mechanics lies the *ergodic postulate*, which simply states that the ensemble average equals the time average at the limit where time goes towards infinity. Hence,

$$[A] = \lim_{t\to\infty} \frac{1}{t} \int_0^t dt'\, A(t') \quad , \tag{2.1}$$

where by the Boltzmann statistics

$$[A] = \frac{\sum_i A_i \exp(-\beta E_i)}{\sum_i \exp(-\beta E_i)} \quad \text{or} \quad [A] = \frac{\int d\Gamma_{q,p}\, A(q,p) \exp(-\beta H(q,p))}{\int d\Gamma_{q,p} \exp(-\beta H(q,p))}\,, \tag{2.2}$$

where $d\Gamma_{q,p}$ is a volume element in phase space. Generally speaking, evaluation of the right hand side of Eq. (2.1) leads to *molecular dynamics* and the left side to *Monte Carlo simulations*.

In molecular dynamics, the time evolution of the ith atom in the system is given by Newton's second law

$$\ddot{\mathbf{x}}_i = -\frac{1}{m_i}\nabla\Phi$$

where $\mathbf{x}_i$ is the position of the atom, m its mass and Φ is the force field potential.

Monte Carlo simulations, however, try to find the most probable conformations by evaluating the *canonical distribution*, i.e. $\exp(-\beta H(q,p))/Z$, where Z is the denominator of the continuous ensemble average in Eq. (2.2).

Quantum Mechanics In quantum mechanical treatments, on the other hand, the time evolution of the system is given by the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}|\,\alpha,\,t\,\rangle = H|\,\alpha,\,t\,\rangle \quad ,$$

where the Hamiltonian operator H is the sum of the potential and kinetic energy of the state, and the *ket*, represented by $|\,\rangle$, is the generalised representation of its solutions. As in classical mechanics, the time evolution is generated by the Hamiltonian, and for a stationary state, which is necessarily prepared in eigenstates to the Hamiltonian, the time dependence of the state can be factored out.

$$|\,a',\,t_o \to t\,\rangle = \exp\left(-\frac{i}{\hbar}E_{a'}(t-t_o)\right)|\,a'\,\rangle.$$

Inserting this into the Schrödinger equation gives the time independent formulation

$$H|\,a'\,\rangle = E_{a'}|\,a'\,\rangle \quad ,$$

provided that the Hamiltonian is time independent.

The obvious drawback of the classical approach is that molecules are not classical objects[3] and that important quantum mechanical effects will be neglected in such treatments.

This is partly remedied by approximating such effects by terms in the potential Φ, either based on experimental results or quantum mechanical calculations. However, there can never be any certainty as to whether a potential derived specifically for one system is applicable to another, unless this has been shown either experimentally or by quantum mechanical *ab initio* calculations.

On the other hand, since *ab initio* calculations are much more computationally demanding than the classical treatments, they have been mostly confined to smaller single molecular systems until very recently.

[3]Whereas this statement is easy to motivate for molecules, it is harder to justify for macromolecular borderline classical – quantum mechanical cases, such as proteins. For an enlightening measurement of the transition from quantum to classical mechanics in fullerenes, by the renowned physicist Anton Zeilinger, please consult reference [27], and for a theoretical treatment of decoherence and open systems, the impressive Ph. D. thesis of Johan Åberg is recommended [1].

There is an abundance of studies on the base stacking interactions using molecular mechanics, where the work of Cristopher Hunter may be of special interest[4] [29, 30].

Whereas semi-empirical treatment of base-pair hydrogen bonding can be traced back to the late sixties [37], *ab initio* calculations on base stacking has been extensively studied fairly recently by Jiří Šponer, in the group of Pavel Hobza [57, 34, 58], but others have also begun to realize the potential in this field [38, 55], with new quantum mechanically related experimental data on DNA being published with increasing frequency.

2.3 The Standard Reference Frame

There is a unified reference frame for internal helical coordinates, referred to in the literature as *the Cambridge Accord* [11]. These coordinates are subdivided into *local base-pair parameters*, *local step parameters* and *global helical parameters* respectively.

Local base-pair parameters describe deviations from the ideal Watson-Crick base-pairing geometry between complementary bases, local step parameters define the transformations between adjacent base-pairs when going from one base-pair to another, and global helical parameters characterize base-pair orientation with respect to a globally defined coordinate system.

In the base-pair coordinate frame, the y-axis is taken to be parallel to either the line connecting the C1′ of the first strand to the corresponding C1′ on the complementary strand, or the line between the C6 atom of the pyrimidine with the C8 on the purine, as shown in figure 2.2. Whatever the choice, the y-axis needs to intersect the C6 atom of the pyrimidine.

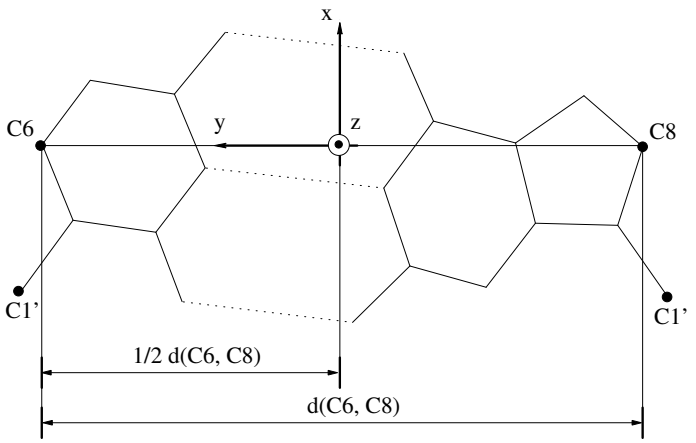


Figure 2.2: The base-pair oriented coordinate frame as given by the Cambridge Accord.

The x-axis must also lie in the plane of the base-pair, perpendicular to the y-axis, so that it intersects the bisector of C6-C8. Consequently, the point of intersection will lie at half the length between C6 and C8. Finally, the z-axis

[4] Actually, the latter citation has inspired much of the work presented here.

completes the coordinate system and will, by the right hand rule, point outwards and be orthogonal to the plane of the paper.

With such a choice of coordinate frame, two consecutive base-pairs in ideal B-DNA conformation are displayed in figure 2.3.

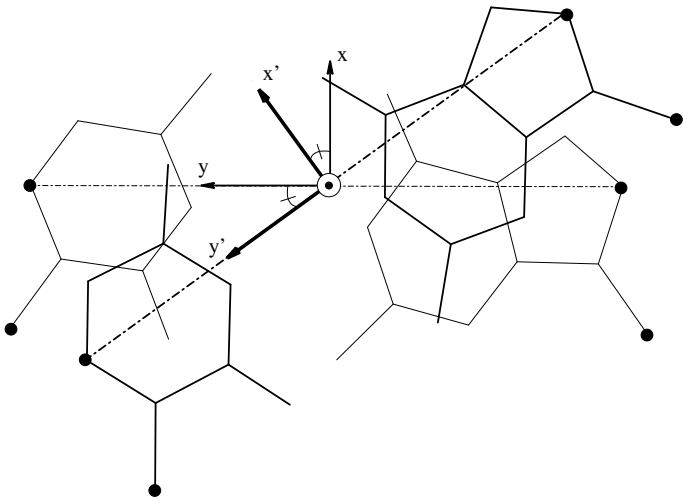


FIGURE 2.3: A graphical representation of two consecutive base-pairs in ideal B-DNA, together with the local base-pair coordinate systems given by the Cambridge Accord. The coordinate system of the first base-pair is given by the xy-axes, the second base-pair by the $x'y'$-axes and the two z-axes are parallell. By this definition, the base-pairs in B-DNA are rotated with respect to the z-axis by 36^o.

Indeed, this choice of coordinate system provides, to a good approximation, a description of the helical turn in B-DNA, where translations in the xy-directions can be considered negligible. Actually, the rotation shown above is the primary rotational degree of freedom in the Cambridge Accord, termed *helical twist* (ω).

By representing the base-pairs with slabs, as shown in figure 2.4, the six local base-pair parameters are schematically depicted in figure 2.5 and the local step parameters are shown in figure 2.6.

In 2001, the joint committee of IUPAC-IUBMB approved the parameters in the Cambridge Accord, and the resulting work is termed *the Standard Reference Frame for the Description of Nucleic Acid Base-pair Geometry* and also contained cartesian coordinates of the atoms in the bases given in this reference frame [41]. These coordinates will be explicitly used in this work and are presented in table A.1, located in the appendix section A.1 on page 71.

2.4 The Mid-Step Triad

Having defined the helical coordinates, we require that the coordinates in this frame should be invariant under the two-fold symmetry of the double helix,

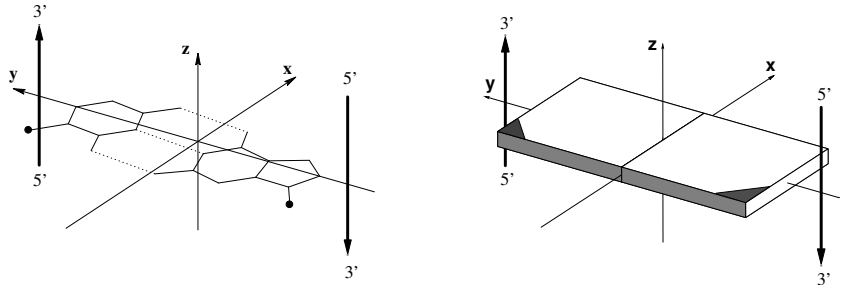
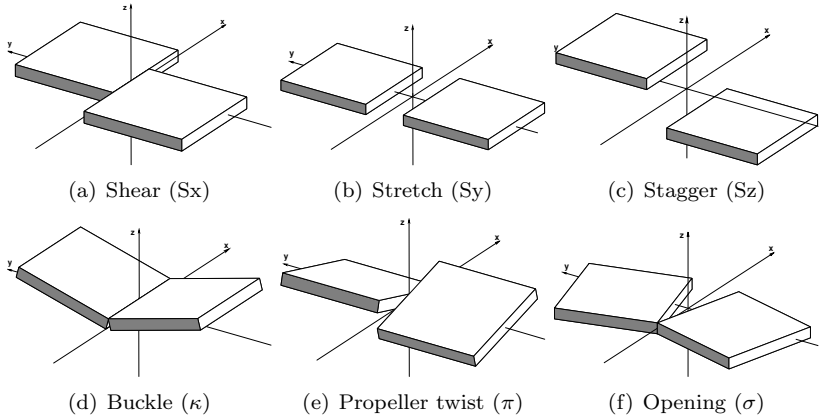


FIGURE 2.4: A full atomic representation of the bases, to the left, can be simplified using slabs to denote the planar bases, on the right.



(a) Shear (Sx) (b) Stretch (Sy) (c) Stagger (Sz)

(d) Buckle (κ) (e) Propeller twist (π) (f) Opening (σ)

FIGURE 2.5: Local base-pair parameters as given by the Cambridge accord. Each column represents translations and rotations about the x-, y-, and z-axis respectively. (Constructed and visualized with `3DNA` [36] and `XFig`.)

or at least up to a sign change. In other words, the helical coordinates in an arbitrary sequence

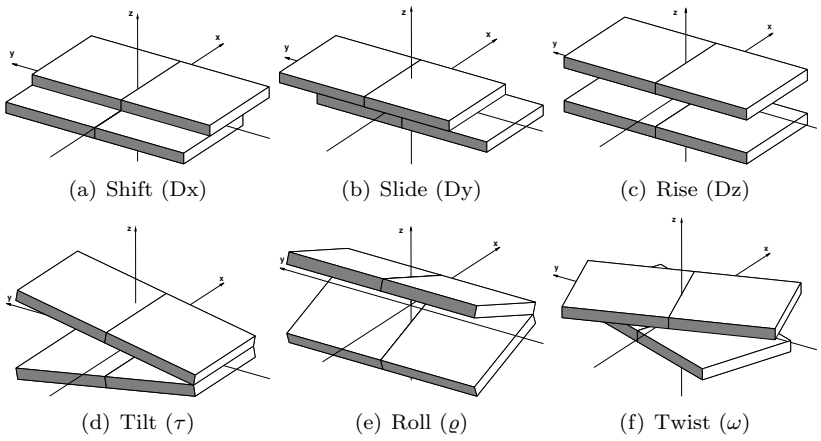
$$\begin{array}{c} 5'\text{--}\mathtt{ACGT}\text{--}3' \\ \quad\;\; |\,|\,|\,| \\ 3'\text{--}\mathtt{TGCA}\text{--}5' \end{array}$$

should be identical when calculated from the left to right as from the right to left in the corresponding helix.

In order for this to hold, the rotation-translation operator taking the first base to the second should commute with the operator that brings it back. It can be shown[5] that this holds in the plane if the translation is taken to be along the bisector of the rotational angle.

In practice, a *virtual intermediate step* is defined to lie between the base-pairs at half the translation vector lengths from base-pair i to $i+1$. As long as

[5]This will not be shown here, since it is slightly beyond the scope of this background. However, an outline of a proof using rotation-translation commutators can be given upon request.



(a) Shift (Dx) (b) Slide (Dy) (c) Rise (Dz)

(d) Tilt (τ) (e) Roll (ϱ) (f) Twist (ω)

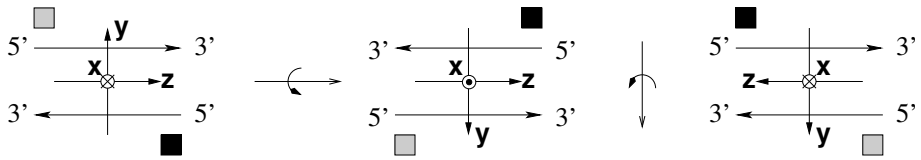
FIGURE 2.6: Base-pair step parameters as given by the Cambridge accord. Each column represents translations and rotations about the x-, y-, and z-axis respectively. (Constructed and visualized with `3DNA` [36] and `Xfig`.)

the base-step parameters are evaluated with respect to that theoretical intermediate coordinate frame, the rotation-commutation operators will commute.

This concept of defining helical parameters is refered to as *the mid-step triad* (MST) procedure in the literature, originating from [65], and is sketched in 2.7.

Two-fold Symmetry Invariance

However, the prefered invariance under the two-fold symmetry operation is not altogether retained in the MST procedure, which can be seen in the following schematical transformations.



From this representation we see that whereas the y- and z-axes follow the symmetry operation, the x-axis remains unchanged. Thus, every helical parameter that is symmetric with respect to the x-axis, namely tilt τ and shift Dx, will retain their magnitude but suffer a sign-change.

Finally, moving beyond the planar description, we are forced to accept the general non-commutativity of rotation matrices and resign to either stating the precise order by which the operations have been performed or making small angle approximations. Hence, to get further insight into helix geometry, we need to turn to the specifics of the calculation scheme at hand, which for the `3DNA` software [36] used in this report can be found in [35, 28].

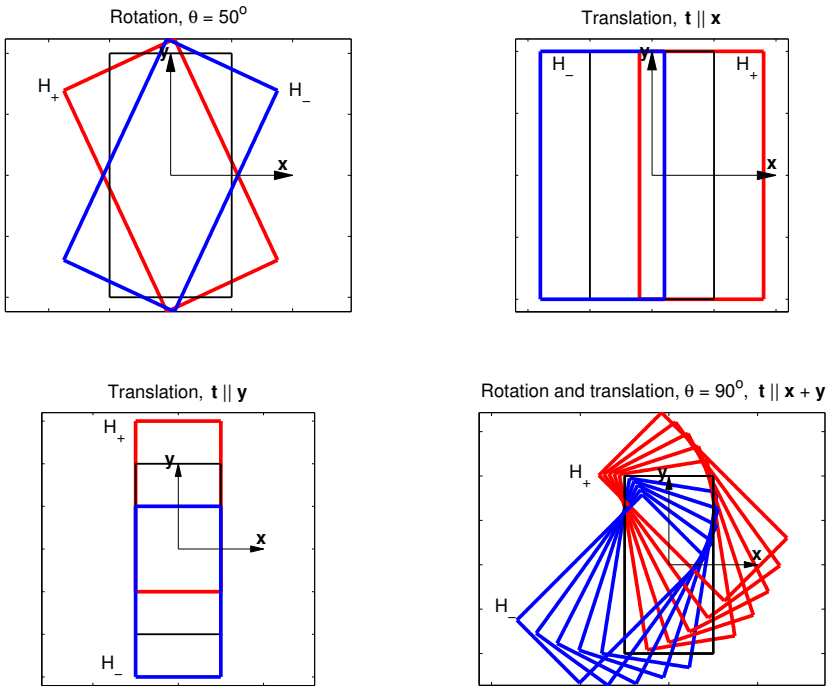


FIGURE 2.7: A graphical representation of the commuting rotation-translation operators in the mid-step triad choice of reference frame. Here, basepair i and $i+1$ are represented by a blue and red rectangle, respectively, whereas the mid-step triad intermediate is given with thin black lines. Notice the equivalency with the helical step parameter representations in figure 2.6. (Created in `Matlab`)

Also note that under the two-fold symmetry there are ten unique base-pair steps, and by considering the following array

	A	G	C	T
T	**TA**	TG	TC	TT
C	**CA**	**CG**	CC	CT
G	**GA**	**GG**	**GC**	GT
A	**AA**	**AG**	**AC**	**AT**

we realize that the sequences found below the diagonal are symmetrical with respect to those above the diagonal. Whether we choose the upper diagonal or the lower diagonal sequences as our reference is arbitrary, but any parameter evaluated from a lower diagonal sequence will be equivalent by definition with one from the upper diagonal, save for a sign change.

3 Theory

> It is by logic that we prove,
> but by intuition that we discover.
>
> Henri Poincaré

In this section, the theoretical background of the most important quantum chemical methods used in this work will be outlined, namely *the Self Consistent Field method*, *Møller-Plesset perturbation* and *the Counterpoise correction*, along with a brief description of *gaussian basis sets* and the *basis set superposition error*.

To this end, the original articles of Roothan [47], Boys [7] and Møller [39] have been reviewed and rewritten in terms of the Dirac formalism[1], along with a few notes and comments.

For a brief and non-technical survey, please consult the summary on page 29.

3.1 Hartree-Fock Theory

The Hartree-Fock hamiltonian operator

In the unrestricted treatment, where each electron is associated with its own spinorbital $|\, i\uparrow\rangle$, the state of an atom or a molecule consisting of N electrons is given by the *Slater determinant*

$$|\,\Psi\,\rangle = (N!)^{-\frac{1}{2}}\,\epsilon_{ijk\ldots l}\,|\,i\uparrow\rangle \otimes |\,j\downarrow\rangle \otimes |\,k\uparrow\rangle \otimes \ldots \otimes |\,l\uparrow\rangle \quad , \tag{3.1}$$

which guarantees that the overall state is antisymmetric upon permutation of any electron, in agreement with the Fermi-Dirac statistics of fermions.

For ground states, the restricted treatment, where two electrons of paired spins occupy the same orbital, is a common simplification, and is given by the similar Slater determinant[2]

$$|\,\Psi\,\rangle = (N!)^{-\frac{1}{2}}\,\epsilon_{ijk\ldots l}\,|\,i\uparrow\downarrow\rangle \otimes |\,j\uparrow\downarrow\rangle \otimes |\,k\uparrow\downarrow\rangle \otimes \ldots \otimes |\,l\uparrow\downarrow\rangle \quad . \tag{3.2}$$

The electronic hamiltonian operator in the adiabatic approximation is given by

$$\hat{H} = \sum_i \hat{h}_i + \frac{1}{2}e^2 \sum_{i\neq j} \frac{1}{|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j|} \tag{3.3}$$

[1] For an introduction to the Dirac formalism, the excellent book by J.J. Sakurai is warmly recommended [49].

[2] Where it is understood that $|\,i\uparrow\downarrow\rangle = |\,i\uparrow\rangle \otimes |\,i\downarrow\rangle$, such that $|\,i\uparrow\rangle \otimes |\,i\downarrow\rangle = -|\,i\downarrow\rangle \otimes |\,i\uparrow\rangle$, and that the summation is also carried out over the spin states.

where $\hat{h}_i$ is the hamiltonian operator for the ith electron moving in the field of the naked nuclei and $\hat{\mathbf{x}}_i$ is its position. In terms of the antisymmetrized product in the restricted treatment of Eq. (3.2), it can be shown that the expectation value of the total Hamiltonian simplifies to

$$E = \langle \Psi | \hat{H} | \Psi \rangle = 2 \sum_i h_i + \sum_{ij} (2J_{ij} - K_{ij}) \quad , \tag{3.4}$$

where J_{ij} and K_{ij} are the matrix elements of the *Coulomb* and the *exchange operator* respectively. Defining the two-electron product basis $\{|\, i \,\rangle \otimes |\, j \,\rangle\}$, of orbital i and j, these elements may be succintly written

$$J_{ij} = e^2 \langle\, ij \,| \frac{1}{|\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2|} |\, ij \,\rangle \quad \text{and} \quad K_{ij} = e^2 \langle\, ij \,| \frac{1}{|\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2|} |\, ji \,\rangle \quad . \tag{3.5}$$

We may also define the corresponding one-electron operators $\hat{J}_i$ and $\hat{K}_i$ through the properties

$$J_{ij} = \langle\, i \,| \hat{J}_j |\, i \,\rangle = \langle\, j \,| \hat{J}_i |\, j \,\rangle \quad \text{and} \quad K_{ij} = \langle\, i \,| \hat{K}_j |\, i \,\rangle = \langle\, j \,| \hat{K}_i |\, j \,\rangle \quad ,$$

by which the closed shell *Fock operator* $\hat{F}$ is written

$$\hat{F} \equiv \hat{h} + \sum_i (2\hat{J}_i - \hat{K}_i) \quad . \tag{3.6}$$

From these representation independent definitions, we can expand the coulomb and exchange operators in terms of the position eigenbasis, and convince ourselves that the Coulomb operator represents the repulsion energy between the electrons in the orbitals $|\, i \,\rangle$ and $|\, j \,\rangle$, whereas the exchange operator has no obvious classical interpretation.

The Independent Particle Model. Furthermore, by summing over the j occupied states in the position eigenbasis expansion of the Coulomb and exchange operator, the interaction of the electrons in the ith orbital is seen to be given with respect to the mean-field of all the other orbitals. This approximation is known as *the independent particle model*.

The Exchange Interaction. Finally, by associating the positions with spins, we can also see that the Coulumb operator accounts for the electrostatic interaction between all orbitals, whereas the exchange operator removes charge repulsion energy due to parallell spin, whose coexistence in the same position is forbidden by the Pauli principle.

The Roothaan Equations

In order to find the orbitals and their associated energies in the independent particle model, we may begin by assuming that they can be formed from the set of normalized atomic orbitals of the constituent atoms. Given the set of normalized atomic orbitals $\{|\, a_j \,\rangle\}$, the corresponding transform of the ith molecular orbital by such a *linear combination of atomic orbitals* (LCAO) is given by

$$|\, i \,\rangle \sim \sum_k |\, a_k \,\rangle \langle\, a_k \,|\, i \,\rangle = \sum_k c_k |\, a_k \,\rangle \quad . \tag{3.7}$$

Equipped with the solutions of each atom, our only remaining concern lies in finding the set of coefficients $\{c_k\}$ for which the Schrödinger equation is solved.

The Matrix Representation. To simplify the notation, we may associate the expanded molecular orbital with the column vector $\mathbf{c}_i$, where

$$|\, i \,\rangle \doteq \mathbf{c}_i = \begin{pmatrix} c_1 & c_2 & \dots & c_k \end{pmatrix}^T \quad .$$

In this manner, an integral involving two molecular orbitals with respect to an arbitrary one-electron operator $\hat{A}$ may then be associated with the matrix product

$$\langle\, i \,|\hat{A}|\, j \,\rangle \sim \sum_{kl} \underbrace{\langle\, i \,|\, a_k \,\rangle}_{\mathbf{c}_i^\dagger} \underbrace{\langle\, a_k \,|\hat{A}|\, a_l \,\rangle}_{A} \underbrace{\langle\, a_l \,|\, j \,\rangle}_{\mathbf{c}_j} \doteq \mathbf{c}_i^\dagger A\, \mathbf{c}_j \quad ,$$

where A is the matrix whose elements are given by the integrals $\langle\, a_k \,|\hat{A}|\, a_l \,\rangle$.

For instance, the Coulomb operator J_{ij} in Eq. (3.5) can be written as the matrix product

$$J_{ij} = \begin{cases} \langle\, i \,|\hat{J}_j|\, i \,\rangle \doteq \mathbf{c}_i^\dagger J_j \mathbf{c}_i \\ \langle\, j \,|\hat{J}_i|\, j \,\rangle \doteq \mathbf{c}_j^\dagger J_i \mathbf{c}_j \end{cases} \quad ,$$

and by the general non-orthogonality of $\{|\, a_k \,\rangle\}$

$$\langle\, i \,|\, j \,\rangle \sim \sum_{kl} \underbrace{\langle\, i \,|\, a_k \,\rangle}_{\mathbf{c}_i^\dagger} \underbrace{\langle\, a_k \,|\, a_l \,\rangle}_{S} \underbrace{\langle\, a_l \,|\, j \,\rangle}_{\mathbf{c}_j} \doteq \mathbf{c}_i^\dagger S\, \mathbf{c}_j = \delta_{ij} \quad ,$$

where S is appropriately termed *the overlap matrix*, since its elements are given by the overlap between the kth and lth atomic orbital. We see that the overlap matrix serves as a connection between the non-orthogonality of the atomic orbitals and the orthonormality of the molecular orbitals.

The Variational Procedure. In order to find these coefficients we can use the positive semidefinitness of the expectation value of the Hamiltonian to convince ourselves that there can be no arbitrary choice of state which has a lower energy than the true eigenstate.

Hence, by the variational principle, a necessary, but not sufficient, requirement for our best guess of the ground state ket is

$$\delta\, E = \delta\, \langle\, \Psi \,|\hat{H}|\, \Psi \,\rangle = \langle\, \delta\, \Psi \,|\hat{H}|\, \Psi \,\rangle + \langle\, \Psi \,|\hat{H}|\, \delta\, \Psi \,\rangle = 0 \quad ,$$

where, by Eq. (3.4),

$$\begin{aligned} \delta E \;&=\; 2\sum_i \delta h_i + \sum_{ij} (2\delta J_{ij} - \delta K_{ij}) \doteq 2\sum_i (\delta \mathbf{c}_i^\dagger)(h - \sum_j (2J_j - K_j))\, \mathbf{c}_i \\ &+\; 2\sum_i (\delta \mathbf{c}_i^T)(h^* - \sum_j (2J_j^* - K_j^*))\, \mathbf{c}_i^* = 0 \quad . \end{aligned} \tag{3.8}$$

We also require that the molecular orbitals remain orthonormal during the variation,

$$\delta\, \langle\, i \,|\, j \,\rangle = \langle\, \delta\, i \,|\, j \,\rangle + \langle\, i \,|\, \delta\, j \,\rangle \doteq (\delta \mathbf{c}_i^\dagger) S \mathbf{c}_j + \mathbf{c}_i^\dagger S (\delta \mathbf{c}_j) = 0 \quad . \tag{3.9}$$

where by the method of lagrangian multipliers we may choose the set of multipliers $-2\varepsilon_{ij}$, and combine Eq. (3.8) with Eq. (3.9) to obtain

$$\delta E' = 2\sum_i (\delta \mathbf{c}_i^\dagger)(F\mathbf{c}_i - \sum_j S\mathbf{c}_j\varepsilon_{ji}) + 2\sum_i (\delta \mathbf{c}_i^T)(F^*\mathbf{c}_i^* - \sum_j S^*\mathbf{c}_j^*\varepsilon_{ji}) = 0 \quad , \tag{3.10}$$

where the Fock matrix[3] F is given by the sum of matrices

$$F = h - \sum_j (2J_j - K_j) \quad ,$$

whose entries are given by the integrals

$$F_{kl} = \langle\, a_k \,|\hat{h}|\, a_l \,\rangle + \sum_j \sum_{mn} c_{jm}c_{jn} \left(2\langle\, a_m\, a_k \,|\frac{e^2}{\hat{x}_1 - \hat{x}_2}|\, a_n\, a_l \,\rangle - \langle\, a_m\, a_k \,|\frac{e^2}{\hat{x}_1 - \hat{x}_2}|\, a_l\, a_n \,\rangle \right) . \tag{3.11}$$

Finally, the conditions in Eq. (3.10) are fulfilled when

$$\begin{cases} F\mathbf{c}_i = \sum_j S\mathbf{c}_j\varepsilon_{ji} \\ F^*\mathbf{c}_i^* = \sum_j S^*\mathbf{c}_j^*\varepsilon_{ij} \end{cases} ,$$

where ε_{ij} can actually be interpreted as being elements in the hermitian matrix ε. Hence, these two equations become equivalent with the single matrix equation

$$FC = SC\varepsilon \quad , \tag{3.12}$$

where the matrix C is given by the coefficient column vectors

$$C = \begin{pmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_k \\ | & | & & | \end{pmatrix} \quad .$$

The set of equations in (3.12) is known as the Roothaan equations [47], and will provide us with the coefficients we seek. In addition, there exists a unitary tranformation so that the matrix ε_{ij} is diagonalized, and then the diagonal entries equals the molecular orbital energies.

The Self-Consistent Field Method

Unfortunately, to find a solution to these equations it is required that we are able to form the Fock matrix, whose elements are in turn dependent on the coefficients that we seek (cf. Eq.s (3.11) and (3.12)). In other words, the Roothaan equations are inherently recursive.

Now, the only consistent feature we can require from a solution to a recursive relation is that it converges. In practice, one usually begins by guessing an initial set of coefficients, then form the Fock matrix, by Eq. (3.11), from which new coefficients are calculated, through Eq. (3.12), and are subsequently inserted back into Eq. (3.11), and so on.

The solution is given by a convergence criterion for the total energy

$$\lim_{i\to k} E_{i-1} - E_i \leqslant \delta \quad , \tag{3.13}$$

[3] cf. the Fock operator in Eq. (3.6) on page 20.

where δ is a small positive value[4] and E_i is the ith Hartree-Fock energy, given by Eq. (3.4) on page 20. The procedure is repeated until δ has been reached, within a certain k number of steps. This procedure is called the *self consistent field* (SCF) method, and is the starting point for many orbital calculations.

3.2 Møller-Plesset Perturbation Theory

So far we have found a means of obtaining orbital energies and expansion coefficients in terms of atomic orbitals for a ground state molecule. Whereas Hartree-Fock theory incorporates the Pauli principle we realize that we are still somewhat limited by the lack of pair interactions between the electrons by the independent particle model. This neglected interaction is termed *electron correlation*.

However, through the SCF procedure we have a complete description of the static states which the electrons may occupy, and that completeness also blesses us with the total conformal space that is available to them.

Assuming that the inclusion of the electron correlation is a small correction to the HF treatment, the correlation energy can be found by a perturbative expansion of the exact hamiltonian, which is known as *Møller-Plesset perturbation theory*, introduced in 1934 [39].

At this point, it should be pointed out that there are also many other quantum chemical methods that take electron correlation into account[5], and the motivation behind the choice of Møller-Plesset perturbation theory for this work is that it is relatively cheap and known to give a satisfactory description of non-bonding interactions [51].

Now, assume that the exact hamiltonian H in Eq. (3.3) harbors such solutions that are not captured by the Hartree-Fock hamiltonian H_{HF}, and define the difference Γ to be

$$\Gamma \equiv H - H_{HF} = \frac{1}{2}\sum_{i\neq k}^{n} V_{ik} - \sum_{i}^{n}\left(J_i^s - K_i^s\right) + \frac{1}{2}Tr\left(\rho^s\left(J_i^s - K_i^s\right)\right) \quad ,$$

with respect to the density operator, where a superscript s denotes the stationary operators. If Γ is sufficiently small, we can regard it as a correction to the Hartree-Fock hamiltonian

$$(H_{HF} + \lambda\Gamma)|\,\Psi\,\rangle = E|\,\Psi\,\rangle \tag{3.14}$$

and express the exact solution $|\,\Psi\,\rangle$ in orders of λ through the perturbative expansion

$$|\,\Psi\,\rangle = |\,\Psi^{(0)}\,\rangle + \lambda|\,\Psi^{(1)}\,\rangle + \lambda^2|\,\Psi^{(2)}\,\rangle + \ldots + \lambda^n|\,\Psi^{(n)}\,\rangle \tag{3.15}$$

[4]Hypothetically, it is concievable that the limit $\lim_{i\to\infty} E_{i-1} - E_i = 0$ exists, since the variational procedure guarantees at least one local minimum somewhere in the conformation space of the coefficients, provided that the basis set is complete. However, in reality this is rarely, if ever, seen.

[5]Usually, *coupled cluster theory* is the method of choice.

with the associated energies

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \ldots + \lambda^n E^{(n)} \quad . \tag{3.16}$$

Substituting the expressions in Eq.s (3.15) and (3.16) into (3.14), and comparing the result to the order of λ then gives the following equations for the perturbed energies and states.

$$\begin{array}{lrcl} \mathcal{O}(\lambda^0): & H_{HF}|\,\Psi^{(0)}\,\rangle & = & E^{(0)}|\,\Psi^{(0)}\,\rangle \\ \mathcal{O}(\lambda^1): & \left(H_{HF} - E^{(0)}\right)|\,\Psi^{(1)}\,\rangle & = & \left(E^{(1)} - \Gamma\right)|\,\Psi^{(0)}\,\rangle \\ \mathcal{O}(\lambda^2): & \left(H_{HF} - E^{(0)}\right)|\,\Psi^{(2)}\,\rangle & = & E^{(2)}|\,\Psi^{(0)}\,\rangle\left(E^{(1)} - \Gamma\right)|\,\Psi^{(1)}\,\rangle \\ \vdots & \vdots & & \vdots \end{array} \tag{3.17}$$

On the other hand, let $\{|\,\psi_i\,\rangle\}$ be a basis of antisymmetrized product eigenstates to the Hartree-Fock hamiltonian, such as in Eq. (3.1). By the completeness of this basis, we are free to expand $|\,\Psi\,\rangle$ in terms of such determinants,

$$|\,\Psi\,\rangle = \sum_i |\,\psi_i\,\rangle\langle\,\psi_i\,|\,\Psi\,\rangle = \sum_i c_i|\,\psi_i\,\rangle \quad . \tag{3.18}$$

The perturbed states could then also be expanded to the nth order in λ.

$$|\,\Psi^{(1)}\,\rangle = \sum_i a_i^{(1)}|\,\psi_i\,\rangle \tag{3.19}$$

$$|\,\Psi^{(2)}\,\rangle = \sum_i a_i^{(2)}|\,\psi_i\,\rangle \tag{3.20}$$

$$\vdots$$

$$|\,\Psi^{(n)}\,\rangle = \sum_i a_i^{(n)}|\,\psi_i\,\rangle \tag{3.21}$$

Inserting the expansions from Eq.s (3.19) and (3.20) into (3.17) gives the first and second order states and energies. It will not be shown here, but by expanding the resulting integrals the first order energy is

$$E^{(1)} = \frac{1}{2}Tr(\rho^s(J^s - K^s)) - Tr(\rho(J^s - K^s)) + \frac{1}{2}Tr(\rho^s(J^s - K^s)) = 0 \quad ,$$

which means that the Hartree-Fock solution is correct to the first order in the perturbative expansion.

In the second order correction it is interesing to note that triple and higher order substitutions vanish[6], along with the single substitutions, leaving the only contribution to the doubly substituted antisymmetrized products.

Finally, the expression for the second order energy is given by

$$E^{(2)} = \sum_i \frac{\langle\,\Psi^{(0)}\,|\Gamma|\,\psi_i\,\rangle\langle\,\psi_i\,|\Gamma|\,\Psi^{(0)}\,\rangle}{E^{(0)} - E_i} \quad ,$$

[6]Which can be seen by considering the condition on the pair interaction of the μth and νth electron, $\langle\cdots i^\mu\cdots j^\nu\cdots k\cdots\,|\,\Gamma_{\mu\nu}|\,\cdots k^\mu\cdots l^\nu\cdots m\cdots\rangle\delta_{km}$. Hence, orbital $|\,i\,\rangle$ and $|\,j\,\rangle$ can be substituted, but not $|\,k\,\rangle$, due to the ON of the MOs.

where we note that the largest contributions come from the substituted states closest to the ground state. Since the SCF procedure only produces a finite set of Rydberg states, this is very fortunate.

As an aside, we may notice that Møller and Plesset consciously make the choice of expressing their theory with respect to the Hartree-Fock density operator ρ^s, and start out by stating the equation of motion for ρ^s as given in the Schrödinger picture,

$$i\hbar\dot{\rho}^s = -[\rho^s, H^{HF}] \quad .$$

Whereas both the HF density and the exact density operators are stationary with respect to their hamiltonians, it may be noted as a matter of curiosity the time evolution of the HF ground state, which is a pure ensemble, with respect to the exact hamiltonian is

$$i\hbar\dot{\rho}^s = -[\rho^s, H] = E^{corr} \quad ,$$

with the possible interpretation that the exact solution corresponds to a mixed ensemble of all stationary HF states, hence

$$\rho = \sum_i w_i \,|\, \psi_i \,\rangle\langle\, \psi_i \,| \quad ,$$

such that the density has reached an equilibrium with respect to the exact hamiltonian. In other words, as $t \to \infty$.

If this would be physically true, which it need not be[7], it would be analogous to the ergodic postulate and classical ensemble averages in Eq.s (2.1) and (2.2), which were introduced on page 11, but here with respect to the motion of the electrons.

3.3 Basis sets

In the methods outlined above, the molecular orbitals are given in terms of expansion coefficients for the linear combination of atomic orbitals, as stated in Eq. (3.7).

Although it was assumed that this basis set is complete, and thus infinite, this cannot be achieved in practice where we are always restriced to finite treatments. The choice of this finite basis set is therefore always a significant constraint to the accuracy of the given method.

With this in mind, there are two different types of atomic orbital sets at hand, namely

i. Slater type orbitals (STO) and

ii. Gaussian type orbitals (GTO).

[7]Remember that we are still under the blessing of the completeness relation, and that the mathematical truth may not have a physical interpretation at all.

Slater Type Orbitals. Slater type orbitals are similar to the common solutions of the hydrogenic atom, as presented in most introductory courses in quantum mechanics, and are generally given by

$$\langle\, r\,\theta\,\phi \,|\, n\,l\,m \,\rangle = N_{nlm}\, Y_l^m(\theta,\phi)\,\zeta\, r^{n-1}\exp(-\zeta r) \quad ,$$

where N is a normalization constant, Y_l^m is a *spherical harmonic* and ζ is a coefficient which can be adjusted to accurately describe the radial behaviour or, loosely speaking, the size of the atomic orbital at hand.

Although the Slater type orbitals provide a relatively precise description of the atomic orbitals, their implementation in these methods results in integrals that are too computationally expensive to solve (cf. Eq.s (3.4) and (3.5)) since complicated integrands centered on different nuclei in the polyatomic molecules are involved. This is historically known as *the nightmare of the integrals* and was resolved by the introduction of the Gaussian-type orbitals by S.F. Boys in 1950 [7].

Gaussian Type Orbitals. By defining the atomic orbitals in terms of the cartesian Gaussian type orbitals centered at $\mathbf{r}'$,

$$\langle\, x\,y\,z\,\mathbf{r}' \,|\, \mu\,\nu\,\kappa;\,\zeta\,\rangle = N_{\mu\nu\kappa;\zeta}\, x^\mu y^\nu z^\kappa \exp\left(-\zeta[\mathbf{r}-\mathbf{r}']^2\right), \quad \mu,\nu,\kappa \in \mathbb{N}, \quad \zeta \in \mathbb{R}$$

the product of orbitals situated at different atomic centers in the integrands can be described by a gaussian centered at a single point[8].

Gaussian contractions

The drawback with the GTOs is that they do not describe the atomic orbitals as accurately as the STOs do, so it is customary to expand the atomic orbitals in terms of several GTOs. For an orbital belonging to the atom a, this approximate transform reads

$$|\, a;\, n\,l\,m\,\rangle \sim \sum_i^k \langle\, \mu\,\nu\,\kappa;\,\zeta_i \,|\, a;\, n\,l\,m\,\rangle |\, \mu\,\nu\,\kappa;\,\zeta_i\,\rangle = \sum_i^k c_i \,|\, \zeta_i\,\rangle \quad ,$$

where k is the number, c_i are the coefficients and ζ_i the exponents of the *primitive gaussians* $|\,\zeta_i\,\rangle$.

In the last equality, the $|\,\mu\,\nu\,\kappa\,\rangle$ part was dropped since there is a simple correspondence between the polynomial exponents $|\,\mu\,\nu\,\kappa\,\rangle$ and the atomic lm orbital quantum numbers that is given by symmetry[9].

It is important to note that the sets of coefficients $\{c_i\}$ and $\{\zeta_i\}$ in the general expansion given above are not affected by the calculation. By convention, such a sum of gaussian primitives is refered to as a *contraction*, and the

[8] Actually, this is rather easily shown for two gaussians $\langle \mathbf{r}' \,|\, \zeta' \rangle \langle \mathbf{r}'' \,|\, \zeta'' \rangle$ by completing the square in the exponential functions.

[9] To give a few examples, the combination $|\,2\,0\,0\rangle + |\,0\,2\,0\rangle + |\,0\,0\,2\rangle \sim x^2+y^2+z^2$ describes an s orbital since it is clearly spherically symmetric. By mere inspection, we also find that $|\,1\,0\,0\rangle \sim x \leftrightarrow p_x$, $|\,0\,1\,0\rangle \sim y \leftrightarrow p_y$ and $|\,0\,0\,1\rangle \sim z \leftrightarrow p_z$, whereas $|\,1\,1\,0\rangle \sim xy \leftrightarrow d_{xy}$, $|\,1\,0\,1\rangle \sim xz \leftrightarrow d_{xz}$, and so on, including higher order angular momenta with increasing $\mu\nu\kappa$.

conventional way of describing contractions in terms of the primitives is to put the number of primitives of a given type withing parentheses, in the above case (k), and the number of contractions within square brackets, hence [1].

Furthermore, several different contractions can indeed be combined to describe one and the same atomic orbital, but then they are treated as distinct orbitals in the calculation. To give an example, in an arbitrary basis set of GTOs the carbon $2s$-orbital could hypothetically be given by the following two contractions

$$|\, C;\, 2s\,\rangle \sim d\sum_{i}^{9} c_i |\, s;\zeta_i\,\rangle + d'\sum_{j}^{4} c'_j |\, s;\zeta'\,\rangle \quad \leftrightarrow \quad (13s)/[2s] \quad ,$$

where $\{c_1, c_2, \ldots\}$, $\{c'_1, c'_2, \ldots\}$, $\{\zeta_1, \zeta_2, \ldots\}$ and $\{\zeta'_1, \zeta'_2, \ldots\}$ are the contraction coefficients and exponents. Then the coefficients d and d' are available for variation in the calculation.

Two important conclusions can be drawn from this, namely that the accuracy of the calculation is both dependent on the number of primitives in a contraction and the number of contractions available to the method. Since the integrals are evaluated at the level of primitives and the variation at the level of contractions, the most computationally efficient basis set must consist of the right composition of contractions and primitives given for the system under consideration.

Fortunately, there are many standard basis sets to choose from and established rules of thumb as to how they relate to most model systems.

Standard Basis sets

The available standard basis sets can be subdivided into the following classes

i. Minimal basis sets

ii. Split-Valence and Multiple zeta (ζ) basis sets

iii. Polarized basis sets

iv. Diffuse function augmented basis sets

Minimal basis sets. In minimal basis sets, each atomic orbital is only represented by one contraction, and are typically only used as last resort in very large qualitative calculations.

Split-Valence and Multiple zeta basis sets. In these basis sets, the inner shell atomic orbitals are typically represented by one contraction, whereas the valence orbitals are given by two or more contractions. Consequently, if two contractions are used the basis set is termed double zeta, if three are used triple zeta, and so on.

Polarized basis set. In most basis sets it is possible to add basis functions of higher order angular momenta to accurately describe the polarization of orbitals, which are typically added uncontracted[10].

Diffuse function augmented basis sets. Diffuse functions are uncontracted gaussians where the exponents are taken to be significantly smaller than those already present in the basis set, allowing the molecular orbitals to allocate farther out from the atoms, as the name implies.

To give a few examples, the primitives and contractions for first row atoms[11], $Li - Ne$, using *Dunning's correlation consistent* basis sets, which will be extensively used in this work, are presented in the standard notation in table 3.1. Note that this is a multiple zeta basis set where d, f and g polarization functions are included in the definition, and that diffuse functions are added with the *aug-* prefix.

TABLE 3.1: The contraction schemes for Dunnings correlation consistent basis sets for first row atoms from double zeta (DZ) to quintet zeta (5Z), with and without diffuse functions.

X	cc-pVXZ	aug-cc-pVXZ
D	$(9s\,4p\,1d) \rightarrow [3s\,2p\,1d]$	$(10s\,5p\,2d) \rightarrow [4s\,3p\,2d]$
T	$(10s\,5p\,2d\,1f) \rightarrow [4s\,3p\,2d\,1f]$	$(11s\,6p\,3d\,2f) \rightarrow [5s\,4p\,3d\,2f]$
Q	$(12s\,6p\,3d\,2f\,1g) \rightarrow [5s\,4p\,3d\,2f\,1g]$	$(13s\,7p\,4d\,3f\,2g) \rightarrow [6s\,5p\,4d\,3f\,2g]$
5	$(14s\,8p\,4d\,3f\,2g1h) \rightarrow [6s\,5p\,4d\,3f\,2g\,1h]$	$(15s\,9p\,5d\,4f\,3g\,2h) \rightarrow [7s\,6p\,5d\,4f\,3g\,2h]$

Generally speaking, the more contractions and primitives that are present in the basis set, the better the resulting description. Hence, triple ζ basis sets are expected to produce better results than double ζ, but to a higher computational cost.

Polarization functions are important when describing correlation and chemical bonds in general, since they allow a slight reshaping of the molecular orbitals, whereas diffuse functions are used when long range interaction is modelled, for obvious reasons.

The only danger in including too many functions in the basis set, especially diffuse functions, is that the convergence, as defined in Eq. (3.13), can become harder to achieve.

3.4 The Basis Set Superpostion Error — BSSE

We have so far noted that the conformational freedom of the electrons in the practical realization of the theory is limited by the choice of basis sets. This limitation is a direct consequence of the incompleteness of the basis, which also causes other artefacts in the calculations.

[10] Just as with the multiple contractions described above, they will appear as distinct atomic orbitals in the output of the calculation.

[11] In quantum chemistry, the second row of the periodic system of elements is refered to as the first row.

One of these deficiencies is *the Basis Set Superposition Error* (BSSE), which arises when comparing otherwise size consistent calculations on a molecular complex with those of its fully separated constituents.

Since the basis sets are usually taken to be the same in the complex as in the monomers, the complex will experience a lowering in the energy simply because there are more basis functions — and hence a larger conformal space for the electrons — available in the complex than in the fully separated monomers. This difference in energy is the basis set superposition error[12].

The Counterpoise Correction

To remedy this, we realize that we have to find a way of evaluating the energy of fragment A when provided with the basis set centered on fragment B. If we let this energy be $E^{\mathbf{A}\{B\}}$, and the corresponding energy of fragment B be $E^{\{A\}\mathbf{B}}$, the BSSE energy is given by

$$E^{BSSE} = E^{\mathbf{A}\{B\}} + E^{\{A\}\mathbf{B}} - E^{\mathbf{A}} - E^{\mathbf{B}} \quad .$$

With this definition of the BSSE energy, *the counterpoise correction* to the energy of the complex is

$$E_{cp} = E^{\mathbf{AB}} - \left(E^{\mathbf{A}\{B\}} + E^{\{A\}\mathbf{B}} - E^{\mathbf{A}} - E^{\mathbf{B}} \right) \quad . \tag{3.22}$$

This may naturally be generalized to any number of fragments $A, B, C, \ldots$ by taking

$$E^{BSSE} = E^{\mathbf{A}\{BC\ldots\}} + E^{\{A\}\mathbf{B}\{C\ldots\}} + E^{\{AB\}\mathbf{C}\{\ldots\}} + \cdots - E^{\mathbf{A}} - E^{\mathbf{B}} - E^{\mathbf{C}} - \cdots \quad .$$

Intuitively we may also realize that the BSSE vanishes per definition as the inter-fragmental distance goes towards infinity or the basis set is expanded to completeness. Admittedly, at that limit other inconsistencies emerge which are too slight to make a difference here, so their description will be left for future theoretical treatments.

3.5 Summary

a. Hartree-Fock (HF) theory is the starting point of many molecular orbital calculations, including those given here.

b. In the Hartree-Fock hamiltonian the electronic interaction is approximated to occur in an average fashion, by the *Coulomb* and the *exchange operators*.

c. Whereas the Coulomb operator describes the electrostatic interaction, the purely quantum mechanical exchange operator accounts for the Pauli principle by removing the interaction energy from electrons which coincide in space with parallell spin.

[12] However, it should be noted that the BSSE is not an error *per se*, since the inclusion of more basis functions in the complex provides a better decription of its orbitals. The actual error arises in the comparison between the total energy of the complex and its constituents.

d. The orbitals and energies of an atom or molecule within the Hartree-Fock theory are formed by the *Self Consistent Field* method (SCF) which is a variational procedure that relies on recursion.

e. The average treatment of the electron interaction in HF theory is corrected by *Møller-Plesset Perturbation theory*, where the motion of the electrons is explicitly taken into account — an effect known as *electron correlation* — by using the solutions from the SCF method.

f. The molecular orbitals are formed by *linear combinations of atomic orbitals* (LCAO), but since the explicit usage of atomic orbitals, to the extent that they are known, is too computationally demanding, they are approximated with an expansion of Gaussian-Type Orbitals (GTO), in mathematically incomplete but numerically consistent basis sets.

g. The incompleteness of the basis sets introduces the *basis set superposition error* (BSSE) in treatments where interaction energies in complex of molecules are studied.

h. The BSSE can be corrected by the *Counterpoise procedure*.

4 The Base-Pair Coordinate Frame

They are joined together in pairs.

James Watson
& Francis Crick

There are two main strategies to perform coordinate transformations in terms of the helical parameters given in section 2.3. These are either by explicitly producing cartesian coordinates for the bases by some external software, which is compliant with the Cambridge Accord, and subsequently inputing the result into the computation software, or to directly write them so that the quantum mechanical software itself can perform the transformations.

In this study, the latter strategy has been implemented by rewriting the Cambridge Accord coordinate frame in a version that is compatible with the quantum mechanical `Gaussian` software.

In addition, the resulting `Gaussian` compliant coordinate frame has been implemented in two Perl scripts, named `ssDNA` and `dsDNA`, which take a DNA sequence as argument and outputs a ready-to-run input file of single or double stranded DNA, where the backbone has been omitted.

4.1 Z-Matrix Formulation of the Standard Reference Frame

The Z-Matrix

There are two ways of specifying the coordinates of the atoms in a molecule when performing a quantum mechanical calculation, namely in terms of *cartesian* or *internal* coordinates.

A cartesian coordinate input file simply consists of the atom symbol followed by the x, y and z coordinate. For water this would look something like

```
0 1
O  -0.464  0.177  0.0
H  -0.464  1.137  0.0
H   0.441 -0.143  0.0
```

Here the first line specifies the charge, 0, and the multiplicity, 1, of the molecule. In the subsequent lines, the atom type and its xyz-coordinate is given.

On the other hand, in an internal coordinate input file the coordinates are given in terms of bond lengths, bond angles and dihedral angles. One such construction is the *Z-matrix*, which for the same water molecule would read

```
0 1
O
H  1  1.0
H  1  1.0  2  104.5
```

Charge and multiplicity is the same, of course, but the second line contains only the oxygen atom type specification. In the third line, the hydrogen atom is given with respect to atom 1, which is the oxygen atom, with the bond length 1.0 Å. Finally, in the fourth line, a hydrogen atom is defined lying at a bond length of 1.0 Å from atom 1, the oxygen, making an angle 104.5^o with respect to atom 2, the other hydrogen.

We realize that the Z-matrix is the correct setting for the helical coordinate frame, since it is also given with respect to internal coordinates. Therefore, we would like to produce a Z-matrix in which the translations are given as bond lengths and the rotations as angles, in a way that is compliant with the Cambridge Accord and the CEHS scheme.

The Helical Scaffold

Since we are anticipating that the helical calculations will be demanding, we would prefer to write the helical Z-matrix in such a way that the bases can be treated separately and subsequently be merged into a helix.

Specifically, this would make it possible to do costly geometry optimizations in the appropriate theory on the bases, which is a small problem, then assemble those optimized structures into the helix and subsequently freeze all internal coordinates, save for the helical parameters, or perform a partial optimization.

The problem of writing Z-matrices for the DNA-bases is handled by the graphical user interface `GaussView` , which is an extension of the `Gaussian` package, and will not be described here. Henceforth, we may assume that we have Z-matrices for adenine, guanine, thymine and cytosine.

This reduces to finding a helical scaffold Z-matrix onto which we may attach the individual bases. Indeed, by the considerations outlined in section 2.3, and by the use of dummy atoms, the following choice of Z-matrix is tentatively compliant with shift, slide, rise and twist.

```
X
X, 1,  E1
X, 2, hDz, 1, A
X, 3,  Dx, 2, A, 1,   hOm, 0
X, 4,  Dy, 3, A, 2, -90.0, 0
X, 5, hDz, 4, A, 3, -90.0, 0
X, 6,  E2, 5, A, 4,   hOm, 0
```

In this Z-matrix $E1 = E2 = 1.0$ Å and $A = 90.0^o$, whereas the remaining parameters are the helical coordinates. The `X` atom type is a *dummy atom*, meaning that it does not take part in the calculation and is interpreted by

`Gaussian` simply as a geometrical point in the Z-matrix. Note that rise (Dz) is given by half of it's value, and that twist (ω) is given by

$$\texttt{hOm} = -90.0 + \frac{1}{2}\omega \quad .$$

A graphical representation of the scaffold is given in figure 4.1.

The attachment points in this scaffold are given by dummy atoms 2 and 6 in this Z-matrix, and is defined to lie in the origin of the local base pair reference frame. Consequently, the location of the base with respect to these dummy atoms can be calculated from the reference frame cartesian coordinates presented in table A.1 on page 71.

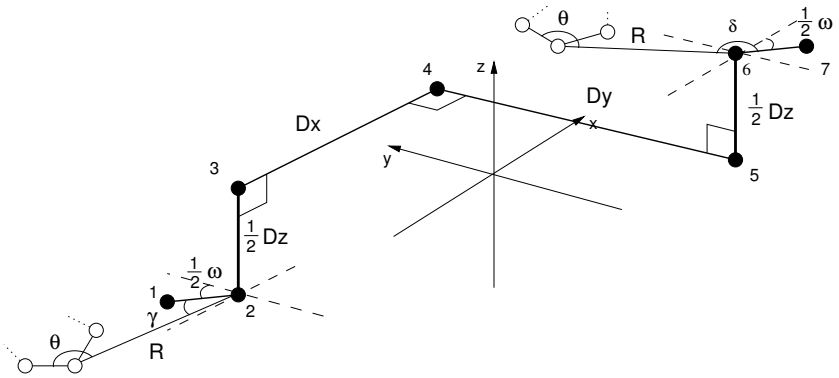


Figure 4.1: The Z-Matrix scaffold onto which the individual bases will be attached is given by the thick lines, and its dummy atoms are shown by filled black circles, along with their respective atomic numbering. The helical parameters are defined by the the bond lengths and angles between these dummy atoms. The attachment point for the base atoms, here shown in white circles, are dummy atom number 2 and 6, which are positioned at the origin of the bases' local reference frame, as described in section 2.3. The mid-step reference triad coordinate frame is shown by the xyz-axes.

The corresponding attachment points on the bases are the $N9$ atoms of purines and $N1$ atoms of pyrimidines. As implied in figure 4.1, we need to calculate one bond distance R and three angles θ, γ and δ given by

$$R = ||\mathbf{r}_a||, \quad \cos\theta = \frac{\mathbf{r}_a \cdot (\mathbf{r}_b - \mathbf{r}_a)}{||\mathbf{r}_a||\,||\mathbf{r}_b - \mathbf{r}_a||}, \quad \cos\delta = \frac{\mathbf{r}_a \cdot \hat{\mathbf{x}}}{||\mathbf{r}_a||} \quad \text{and} \quad \cos\gamma = \frac{\mathbf{r}_a \cdot \hat{\mathbf{y}}}{||\mathbf{r}_a||} \quad , \tag{4.1}$$

where $\mathbf{r}_a$ is the position vector of the $N9$ atom of purines and $N1$ atom of pyridines, and $\mathbf{r}_b$ is the position vector of the $C8$ atoms of purines and $C6$ atoms of pyrimidines. In addition, let us define the angle ε to be

$$\varepsilon = 180^o - \gamma \quad .$$

These values were calculated in `Matlab` and the results are given in table 4.1.

TABLE 4.1: The scaffold attachment parameters as given by equations 4.1 to the precision by which they were calculated. R is given in [Å], γ, θ and δ in [^{o}].

	R	γ	θ	δ
Adenine	4.6796	16.0143	90.8646	106.0143
Guanine	4.7300	15.8139	91.5779	105.8139
Thymine	4.6796	15.9252	97.9233	105.9252
Cytosine	4.7203	15.7970	96.8293	105.7970

There are four different scenarios by which attachments can be made.

i. Connection between scaffold and residue on 5′-end.

ii. Connection between scaffold and residue on 3′-end.

iii. Connection between scaffold and complementary residue on 5′-end.

iv. Connection between scaffold and complementary residue on 3′-end.

The generalized Z-matrix entries for these cases are given by

i. 5′-end

```
N     2     R     1     g     3    90.0   0
C     i    B1     2     t     3    90.0   0
H   i+1    B2     i    A1     2   180.0   0
```

ii. 3′-end

```
N     6     R     7     d     5   -90.0   0
C     i    B1     6     t     7     0.0   0
H   i+1    B2     i    A1     6   180.0   0
```

iii. Complementary 5′-end

```
N     2     R     1     e     3    90.0   0
C     i    B1     2     t     3   -90.0   0
H   i+1    B2     i    A1     2   180.0   0
```

iv. Complementary 3′-end

```
N     6     R     7     d     5    90.0   0
C     i    B1     6     t     7     0.0   0
H   i+1    B2     i    A1     6   180.0   0
```

where `g` $= \gamma$, `d` $= \delta$, `e` $= \varepsilon$, `t` $= \theta$ and i is the relative atomic number of the N atom in the first row.

The first row of these matrices is the first row in the base Z-matrix, so we need to add a reference to either dummy atom number 2 or 6 in the scaffold, a bond angle and a dihedral. We append a bond angle and dihedral to the second row, and only the dihedral angle 180^o to the third row. Hence, the `B1`, `B2` and `A1` are internal coordinates of the bases, and the fourth row or higher belongs to the base entirely and is not affected by the scaffold.

Finally, this can be extended by attaching another scaffold to dummy atom 6 with the following Z-matrix.

```
 X     6    E1     7    90.0    5    -90.0    0
 X     6   hDz     8    90.0    7     90.0    0
 X     9    Dx     6       A    8      hOm    0
 X    10    Dy     9       A    6    -90.0    0
 X    11   hDz    10       A    9    -90.0    0
 X    12    E2    11       A   10      hOm    0
```

Yet another base pair can then be attached to the 12^{th} dummy atom of this Z-matrix.

4.2 dsDNA

The considerations in the preceeding section could very well be generalized to include any number of scaffolds and base pairs. This was done in the scripting language Perl, chosen because of its extensive text processing capabilities through regular expressions, and the resulting scripts are termed `ssDNA` and `dsDNA`.

When supplemented with the `Gaussian` output files from a geometry optimization in the current directory, both `dsDNA` and `ssDNA` will write the scaffolds, record and connect the Z-matrices and optimized variables from ouput files and provide a `Gaussian` input file. Wheras `ssDNA` creates a single stranded DNA molecule devoid of backbone structure, `dsDNA` also makes the complementary strand.

To give an example, the command

```
>> dsDNA AT
```

produces a `Gaussian` input file which is named according to the submitted sequence, with the extension `.com`, and reads

```
%chk=AT.chk
...
#p MP2/aug-cc-pVDZ SCF=tight

AT Twist  Dx  Dy  Dz

0 1
 X
 X    1    E1_S1
 X    2    hDz_S1    1    A_S1
 X    3    Dx_S1     2    A_S1    1    hOm_S1    0
```

```
 X    4    Dy_S1    3    A_S1    2    -90.0    0
 X    5    hDz_S1    4    A_S1    3    -90.0    0
 X    6    E2_S1    5    A_S1    4    hOm_S1    0
 N    2    R_A1    1    g_A1    3    90.0    0
 C    8    B1_A1    2    t_A1    3    90.0    0
 H    9    B2_A1    8    A1_A1    2    180.0    0
 N    9    B3_A1    8    A2_A1    10    D1_A1    0
...
 H    8    B14_A1    21    A13_A1    20    D12_A1    0
 N    6    R_T2    7    d_T2    5    -90.0    0
 C    23    B1_T2    6    t_T2    7    0.0    0
 H    24    B2_T2    23    A1_T2    6    180.0    0
 C    24    B3_T2    23    A2_T2    25    D1_T2    0
...
 H    53    B14_A4    66    A13_A4    65    D12_A4    0

 B1_A1    1.374
...
 D12_A4    0.00256312
 R_A4    4.6796
 d_A4    106.0143
 t_A4    90.8646
 E1_S1    1.00000000
 hDz_S1    2.14496520
 Dx_S1    0.001
 Dy_S1    0.001
 E2_S1    1.00000000
 hOm_S1    -75.00000000
 A_S1    90.00000000
```

Note that parts of the link 0 section, the detailed Z-matrix and variables of adenine and thymine have been omitted for clarity.

In the output Z-matrix each variable will have a suffix consisting of the residue number and base type to avoid ambiguities. In this example, the variable `D12` in the fourth adenine residue has aquired the suffix `_A4`, and is now given as `D12_A4`.

The header is set in a special subroutine in the program named `header{}` and can be modified according to the current needs. Please also note that it is required that the structure optimization files are named `A.log`, `G.log`, `C.log` and `T.log` in order for `dsDNA` and `ssDNA` to recognize them.

Any structure optimization procedure in `Gaussian` is tolerated. The only thing that is absolutely required is that the optimized structure be given in terms of the original Z-matrix at the end of the output file.

In addition, both `dsDNA` and `ssDNA` takes the option `-c` which includes the fragment specification for a counterpoise correction in `Gaussian`.

In that case, the counterpoise keyword is added to the route section, whatever it may be, along with the number of fragments. The scripts will interpret each base residue as one fragment and add the appropriate fragment number

at the end of each Z-matrix row.

For instance, the command

```
>> dsDNA -c GCAATGCGTC
```

writes the input file `GCAATGCGTC.com`, where the route section now reads

```
#p MP2/aug-cc-pVDZ SCF=tight Counterpoise=20
```

and the Z-matrix row of the $N1$ atom from the last cytosine residue is

```
 N     54     R_C10     55     d_C10     53     -90.0     0 10
```

showing that it will indeed be interpreted as belonging to the tenth residue in the counterpoise procedure.

The `ssDNA` and `dsDNA` scripts are available to anyone upon request.

4.3 Evaluation of the Z-Matrix Formulation

The `dsDNA` script, and the scaffold formulation upon which it relies, was evaluated first on a base-pair level, and then on a dinucleotide level by the following procedures.

i. The standard reference base structures given in table A.1 on page 71 were rewritten in terms of Z-matrices.

ii. The Z-matrices were supplemented to the `dsDNA` procedure.

iii. The two ideal base-pairs `G:C` and `A:T` were created by the following commands

```
>> dsDNA G
>> dsDNA A     .
```

iv. The resulting `Gaussian` input files, `G.com` and `A.com`, were read by the `Molden` software and converted to a format supported[1] by `Molekel` [46, 21].

v. The two ideal base-pairs were also created with the Standard Reference Frame compliant `3DNA` software bundle [36].

vi. Six inter-base distances and angles were measured in the A:T structure, and five in the G:C structure, as given by `dsDNA` and `3DNA`.

vii. The deviation in the values from the structures of `dsDNA` and `3DNA` were calculated.

The results of these measurements and the calculated differences are given in table 4.2.

[1] In the current version, `Molekel` only seems to support reading of `Gaussian` *output* files.

TABLE 4.2: These are the results from the evaluation of `dsDNA` on the level of a single base pair. The atoms are labelled according to the IUPAC convention, shown on page 8, where the subscripts I and II denotes whether the atoms are found on the first or second strand, respectively.

C:G							
Distance [Å]	`3DNA`	`dsDNA`	$\|\Delta\|$	Angle [^{o}]	`3DNA`	`dsDNA`	$\|\Delta\|$
$N3_I$:$N1_{II}$	3.001	3.000	0.001	$N1_I$:$C6_I$:$C8_{II}$	67.636	67.639	0.003
$C4_I$:$C6_{II}$	4.348	4.347	0.001	$C2_I$:$O2_I$:$N2_{II}$	117.657	117.662	0.005
$C2_I$:$C2_{II}$	4.278	4.277	0.001	$N3_I$:$C2_I$:$C2_{II}$	60.097	60.099	0.002
$O2_I$:$N2_{II}$	2.866	2.865	0.001	$C4_I$:$N3_I$:$N1_{II}$	119.019	119.025	0.006
$C6_I$:$C8_{II}$	10.030	10.030	0.000	$N4_I$:$C4_I$:$C6_{II}$	56.436	56.435	0.001
T:A							
Distance [Å]	`3DNA`	`dsDNA`	$\|\Delta\|$	Angle [^{o}]	`3DNA`	`dsDNA`	$\|\Delta\|$
$O4_I$:$N6_{II}$	3.046	3.046	0.000	$C2_I$:$O2_I$:$N3_{II}$	112.770	112.764	0.006
$C4_I$:$C6_{II}$	4.340	4.341	0.001	$N3_I$:$C2_I$:$C2_{II}$	64.154	64.185	0.031
$N3_I$:$N1_{II}$	2.962	2.965	0.003	$C4_I$:$N3_I$:$N1_{II}$	117.945	117.944	0.001
$C2_I$:$C2_{II}$	4.182	4.185	0.003	$O4_I$:$C4_I$:$C6_{II}$	58.964	58.966	0.002
$C6_I$:$C8_{II}$	9.954	9.954	0.000	$C4_I$:$O4_I$:$N6_{II}$	123.040	123.039	0.001
$O2_I$:$N3_{II}$	4.904	4.907	0.003	$N1_I$:$C6_I$:$C8_{II}$	66.428	66.425	0.003

Furthermore, to assess whether `dsDNA` is capable of representing shift, slide and twist correctly, the following procedure was adopted.

1. The standard reference base structure Z-matrixes created above were supplemented to `dsDNA` and four structures were created with the commands

   ```
   >> dsDNA CC
   >> dsDNA TT        .
   ```

2. In the resulting `Gaussian` input files, the helical parameters were set to two different sets of values.

3. Four identical structures were created with `3DNA`.

4. The structures were *RMS* fitted in `Molekel`.

5. The root mean squared deviation was calculated along with the length of the vector representing the largest deviation.

Immediately it was realized that `Dy` in the `dsDNA` and `ssDNA` scaffold Z-matrix is actually $-Dy$, as given in the Standard Reference Frame. This can also be realized from figure 4.1. The impact of this unfortunate sign error on the calculations will be elaborated in the discussion. Correcting for this error, the values are given in table 4.3.

4.4 Dinucleotide Base-Pair Geometry

Now that we are equipped with a Z-matrix formulation which is consistent with the Standard Reference Frame parameters shift, slide, rise and twist, we would

Table 4.3: The results from the evaluation of `dsDNA` with respect to shift, slide, rise and twist, given by the two doublestranded dinucleotides 5′−CC−3′ and 5′−TT−3′. The translations and deviations are given in [Å] and twist in [^{o}].

	Dz	Dx	Dy	ω	$\max(\|\|\Delta_i\|\|)$	σ_{rms}
C:G→C:G	3	3	-4	30	0.011	0.005
	3	6	-3	70	0.003	0.001
T:A→T:A	3	3	-4	30	0.007	0.002
	3	6	-3	70	0.007	0.002

also need to know if some of these values could be fixed while we examine the effect of twist on the base orbital energies in single stranded dinucleotide B-DNA.

This was done by collecting a data set of helical parameters from 46 B-DNA crystal structures in the Nucleic Acid Database [5], with a resolution better than 2.0 Å, and subsequently performing a modest statistical analysis.

The NDB identification numbers of these structures are given in table A.2 in the appendix, section A.2 on page 72.

To begin with, the total distributions of the helical step parameters were plotted in `Matlab` and are shown in figure 4.2. The mean and standard deviations from these distributions are given in table 4.4.

Subdividing this data by the ten unique base pair steps given in section 2.4 on page 16, the calculated means, standard deviations and standard errors are given in table 4.5. Note that to compensate for the sign change in the important slide coordinate, the mean and standard deviation of the absolute value of shift is also given, where the two-fold symmetry argument is applicable.

The spread in this data, as given by the standard deviation is quite low for the translational parameters and seem the largest for twist. Nevertheless, there do seem to be statistically significant differences between certain dinucleotide sequences even with respect to this parameter.

Since GC will be of particular interest to us, we may also notice the difference in mean and spread for the twist parameter between the 5′−GC−3′ and the 5′−CG−3′ dinucleotide, being 38.0^o (4.4) in the former and 33.2^o (11.8) in the latter. This difference is not reflected in any of the other parameters for these two dinucleotides.

However, before drawing any conclusions based on the mean and standard deviation, the distributions themselves should be examined. Since the role of twist is the primary focus in this report, the distributions of this parameter for the ten unique sequences were calculated in `Matlab`, and are plotted in figure 4.3.

To qualitatively examine whether the distributions may be considered symmetrical, the median was calculated and compared to the mean, and the results are given in table 4.6.

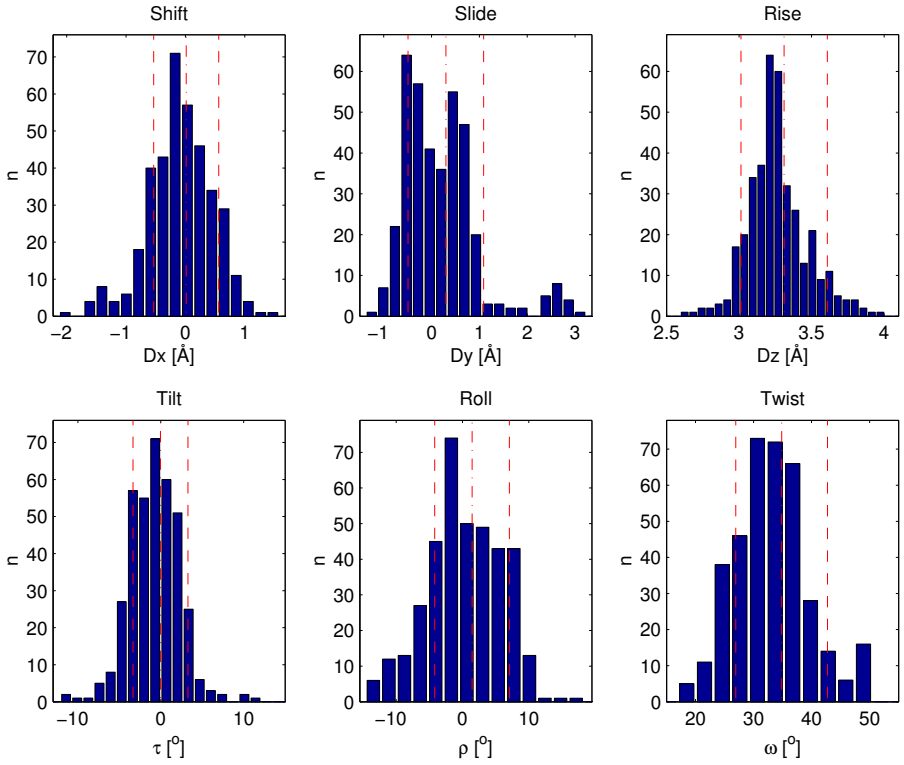


FIGURE 4.2: Total distributions of the helical step parameters in the B-DNA dataset as calculated and plotted in `Matlab`. The dash-dotted lines show the arithmetic mean and the dashed lines the standard deviations, whose values are given in table 4.4.

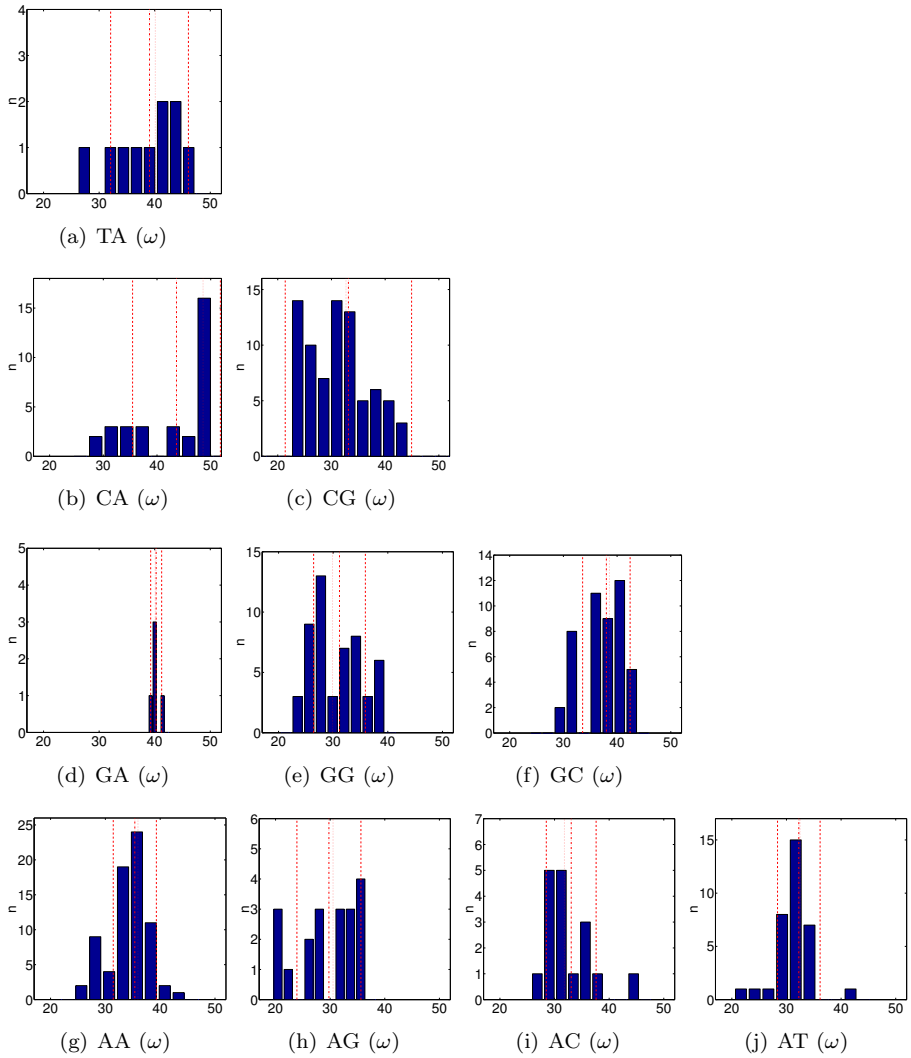
TABLE 4.4: The mean and standard deviations of the helical step parameters as calculated from the dataset. The translations are given in [Å] and the angles in [^{o}].

	Dx	Dy	Dz	τ	ρ	ω
$\bar{x}$	0.01	0.30	3.32	-0.04	1.43	34.84
σ	0.55	0.79	0.30	3.31	5.62	7.91

Table 4.5: Means and standard deviations for the helical step parameters in the data set considered here, along with the number of observations n. The values are presented with their respective two-fold symmetry equivalent sequence, for which also the absolute value of the sign sensitive shift parameter has been given. All translational parameters are given in [Å] and angles in [^{o}].

		Dx	$\|Dx\|$	Dy	Dz	τ	ρ	ω	n
AT	$\overline{x}$	0.03	—	-0.46	3.21	0.1	-0.6	32.3	35
	σ	0.50	—	0.27	0.17	2.2	3.0	3.9	
AC	$\overline{x}$	-0.21	0.45	-0.25	3.27	-0.6	-0.1	33.0	18
	σ	-0.20	0.44	0.56	0.16	2.8	4.1	4.6	
GC	$\overline{x}$	-0.13	—	0.21	3.36	-0.2	-3.6	38.0	48
	σ	0.83	—	0.48	0.25	2.8	5.1	4.4	
AG	$\overline{x}$	0.35	0.40	-0.02	3.28	-1.2	4.2	29.8	20
	σ	0.33	0.27	0.55	0.24	4.0	4.3	5.9	
GG	$\overline{x}$	0.06	0.52	0.39	3.29	0.4	4.9	31.2	54
	σ	0.60	0.31	0.48	0.22	4.0	4.0	4.7	
AA	$\overline{x}$	0.02	0.28	-0.12	3.26	-0.1	0.1	35.4	73
	σ	0.36	0.23	0.31	0.15	2.6	4.0	3.9	
CG	$\overline{x}$	0.04	—	0.59	3.38	0.2	5.0	33.2	81
	σ	0.58	—	0.56	0.51	4.4	5.4	11.8	
GA	$\overline{x}$	0.07	0.32	0.14	3.45	0.5	0.8	40.3	6
	σ	0.50	0.36	0.67	0.08	1.2	2.1	1.0	
CA	$\overline{x}$	0.05	0.24	1.75	3.30	0.2	-1.6	43.6	33
	σ	0.31	0.20	1.14	0.20	2.0	6.5	8.2	
TA	$\overline{x}$	-0.24	—	0.53	3.39	-0.9	2.6	39.1	11
	σ	0.28	—	0.68	0.25	1.9	6.1	7.0	

Avoiding the complication of assuming a normal distribution for the twist parameter, the above mentioned GC and CG distributions were tested for significant difference using *Wilcoxon's rank sum test*. Indeed, they are sampled from different distributions to a confidence degree of $\alpha < 0.001$.



(a) TA (ω)

(b) CA (ω) (c) CG (ω)

(d) GA (ω) (e) GG (ω) (f) GC (ω)

(g) AA (ω) (h) AG (ω) (i) AC (ω) (j) AT (ω)

FIGURE 4.3: Distribution plots of the twist parameter as calculated from the data set. The dashed lines show the standard deviation, the dash-dotted lines the mean and the dotted lines the median of the parameter with respect to the dinucleotide step, based on the values in table 4.5 and 4.6. All values are given in $[^o]$, and the ranges are fixed to $[20, 50]^o$ to simplify comparison.

TABLE 4.6: The median of the twist distributions compared with the mean values.

	AT	AC	GC	AG	GG	AA	CG	GA	CA	TA
$\bar{\omega}$	32.3	33.0	38.0	29.8	31.2	35.4	33.2	40.3	43.6	39.1
$\tilde{\omega}$	32.5	31.8	38.6	30.5	30.0	36.0	32.7	40.1	48.6	40.1
$\lvert\Delta\rvert$	0.2	2.2	0.6	0.7	1.2	0.6	0.5	0.2	5.0	1.0

5 Computations

In this section, all quantum mechanical calculations that have been performed during this project will be described, beginning with the correlation calculations on the Ne_2 model system, followed by the DNA single base and dimer calculations.

The `Gaussian 03` software bundle [25] has been used throughout this project and most of the minor computations were performed on a desktop 64-bit dual core computer system running under Red Hat Fedora Linux Core 4, specifically built and set up for this purpose[1].

The heavier nucleobase dimer correlation computations were performed at the `Ra` opteron cluster, which is — with its 280 64-bit processors and peak capacity of 1.34 Tflops — among the most powerful computer systems in Sweden[2].

5.1 Model Systems

The Neon Dimer

A potential energy scan (PES) was performed with respect to the inter-atomic distance, r, in the neon dimer, Ne_2, using Møller-Plesset perturbation theory, as outlined in section 3.2. From a previous undisclosed calculation[3], a minimum in the potential energy curve was obtained near $r = 3.25$ Å, which served as a starting point for this PES, consequently taken in the range of $r = [2.8, 6.4]$ Å.

This scan was performed to the fourth order in the perturbation theory in the frozen core approximation, including singles, doubles, triples and quarters substitutions in the Slater determinants, thus also automatically providing us with the second and third order corrections to the energies. The basis sets used were Dunning's correlation consistent double, triple, quadruple and quintuple zeta (ζ), both with and without diffuse functions, as described in table 3.1 on page 28.

To account for the basis set superposition error, counterpoise corrections were performed at every point. Although `Gaussian` will explicitly provide this correction for the highest order energy considered, the corrections need to be calculated by hand for remaining lower order energies. This was done using Eq. (3.22) on page 29.

[1] This had not been possible without the generous help from Martin Agback, at the Quantum Chemistry institution at Uppsala University (see the acknowledgements on page 3).

[2] See the `Ra` homepage at `http://www.uppmax.uu.se/systems/ra/ra-cluster`.

[3] Then using second order perturbation theory and Dunning's correlation consistent double zeta basis set, augmented with diffuse functions (aug-cc-pVDZ).

Since `Gaussian` fully supports C shell scripting, the entire procedure can be succintly stated by using the `Scan` keyword and two array loops in the following way.

```
#! /bin/csh
foreach basisset (cc-pVDZ cc-pVTZ cc-pVQZ cc-pV5Z
                  aug-cc-pVDZ aug-cc-pVTZ aug-cc-pVQZ aug-cc-pV5Z)
  echo "Calculating potential curve of Ne:Ne with basisset $basisset"
  mkdir $basisset
  set steps=(9 2 4)
  set point=("2.8" "3.8" "4.4")
    foreach delta ("0.1" "0.2" "0.5")
      g03 <<EOF> $basisset/Ne2_$point[1]_$basisset:r.log
%chk=$basisset/Ne2_$point[1]_$basisset.chk
%mem=600MB
%nproc=2
# MP4/$basisset gfoldprint SCF=tight pop=reg Scan Counterpoise=2

MP4 Ne2 potential with $basisset point $point[1]

0 1
Ne 0.0 0.0 0.0 1
Ne 1 B1 2

B1 $point[1] $steps[1] $delta

EOF
      shift steps
      shift point
    end
    echo "Done!"
end
```

Notice that the step length `delta` is set to be smaller in the vicinity of the minimum, and is progressively increased to save computing time. After execution, the results may conveniently be gathered by harvesting scripts, hence reducing the effort to a minimum.

Since the MP4^{FC}/aug-cc-pV5Z level of theory proved to be challenging even for this limited diatomic system, the calculation was truncated after reaching 3.4 Å, where it was deemed that enough information on the cusp had been obtained.

The resulting counterpoise corrected single point energies are plotted in figures 5.1 and 5.2, compared with experimental results from UV adsorption spectroscopy in [59] as interpreted by [48].

The calculated points have also been compared to the often used empirical *Lennard-Jones* pair potential, which is known to give a good description of

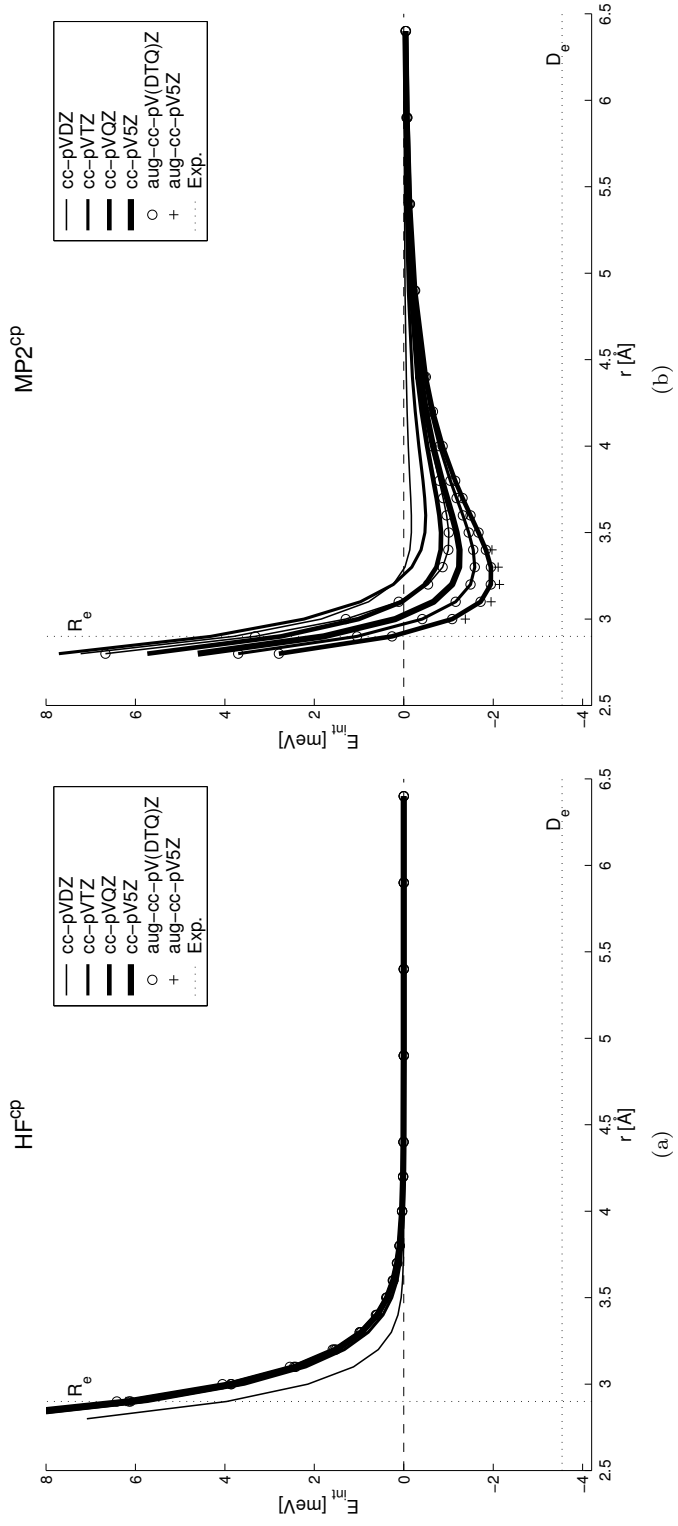


FIGURE 5.1: The results from the Ne_2 potential energy surface scan are plotted with the zeroth order energy in (a) and the second order in (b) for the basis sets (aug-)cc-pV(DTQ5)Z. In each diagram, the experimental values of the equilibrium distance R_e and well depth D_e are given by dotted lines, and the axes have been equally scaled between the plots for comparison. (See also figure 5.2.)

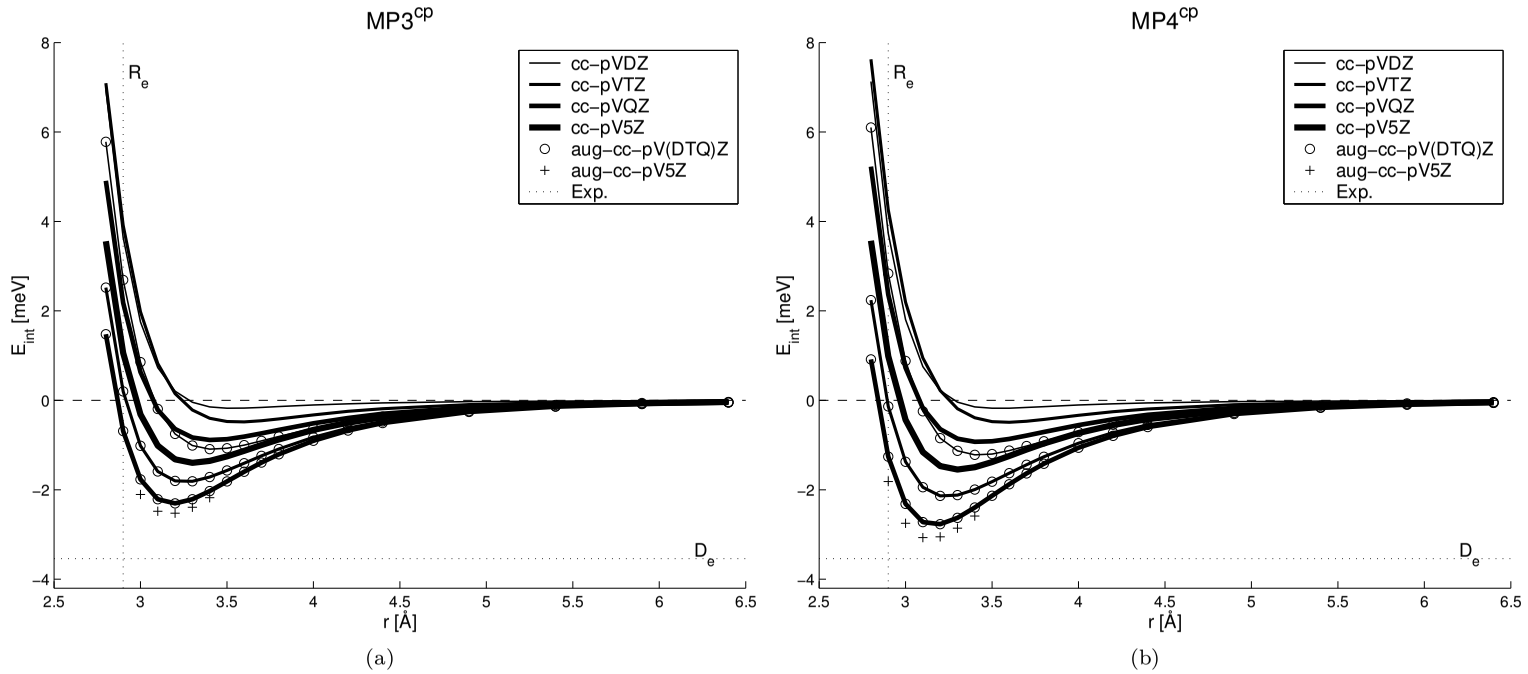


Figure 5.2: The results from the Ne_2 potential energy surface scan are plotted with the third order energy in (a) and the fourth order in (b), for the basis sets (aug-)cc-pV(DTQ5)Z. In each diagram, the experimental values of the equilibrium distance R_e and well depth D_e are given by dotted lines, and the axes have been equally scaled between the plots for comparison. (See also figure 5.1.)

binding energies in rare gas dimers and clusters [32], and is given by

$$U_{ij}^{LJ} = D_e \left[\left(\frac{r_e}{r_{ij}} \right)^{12} - 2 \left(\frac{r_e}{r_{ij}} \right)^{6} \right] \quad ,$$

where r_e is the equilibrium distance and D_e the well depth.

The chemical interpretation of the first term in this potential is the electrostatic repulsion energy excerted between the atoms, which steeply increases at small distances, whereas the second term is usually identified with the *van der Waals* interaction which is a comparatively long distanced interaction.

Adopting the well depth and equilibrium distance derived from the adsorption spectra mentioned above, i.e. $r_e = 2.9$ Å and $D_e = 3.55$ meV, the corresponding Lennard-Jones potential is drawn in figure 5.3.

Also plotted in this figure is the Lennard-Jones potential based on the presently calculated values of r_e and D_e to the highest order, along with the zeroth order SCF energy and the fourth order Møller-Plesset correction.

By henceforth defining the *electron correlation* energy to be the difference between the zeroth and n^{th} order energy as obtained by Møller-Plesset theory, it is seen to follow the *van der Waals* term in the corresponding Lennard-Jones potential.

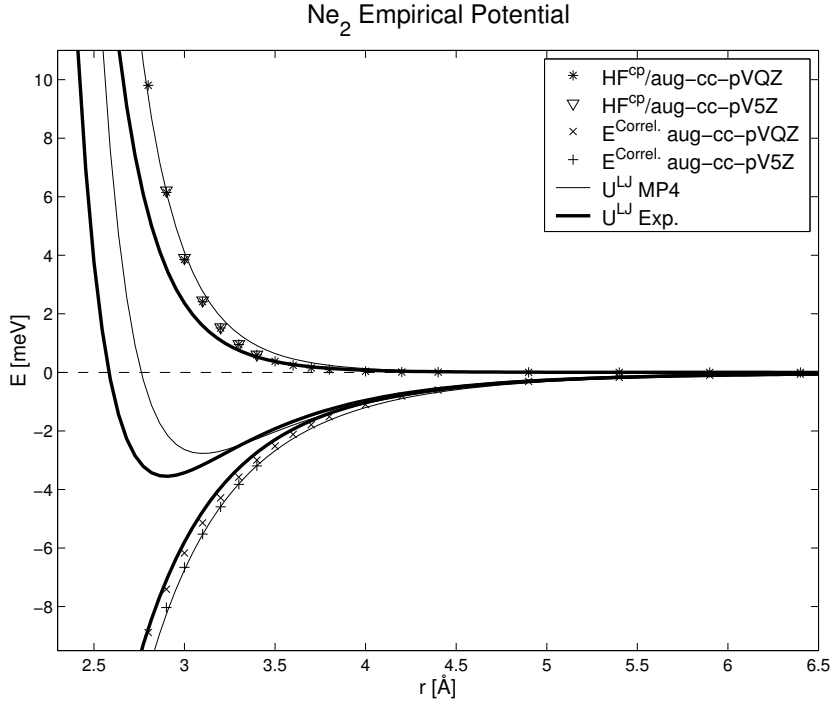


Figure 5.3: Plots of the Lennard-Jones potential with respect to the experimental equilibrium distances and well depth, and those estimated from the present MP4^{FC}/aug-cc-pV(Q,5)Z calculations.

5.2 The DNA Bases

Geometries

The single base geometries were optimized both with DFT, using the Becke three parameter Lee-Yang-Parr hybrid functional B3LYP, and Møller-Plesset second order perturbation theory, MP2.

For adenine, thymine and guanine, the DFT optimizations were performed with Pople's split valence 6-31G(d,p) basis set, with the contraction scheme

$$(10s, 4p, 1d \,/\, 4s, 1p) \rightarrow [3s, 4p, 1d \,/\, 2s, 1p] \quad .$$

To verify the effect of diffuse functions for the geometry calculations, the DFT optimization for cytosine was performed both with the 6-31G(d,p) and the heavier 6-31+G(3df,p) basis set, thus including the contractions

$$(11s, 5p, 3d, 1f \,/\, 5s, 1p) \rightarrow [4s, 5p, 3d, 1f \,/\, 3s, 1p] \quad .$$

The MP2 optimizations were performed with Dunning's correlation consistent double zeta basis set with diffuse functions (aug-cc-pVDZ) as described in table 3.1 on page 28.

The detailed results of the optimizations with respect to bondlengths and angles are given in the appendix, section A.3 on page 72. From those tables a condensed presentation of the root mean square deviations is shown here in table 5.1.

TABLE 5.1: Root mean square deviations of the bond lengths (B) in [mÅ] and angles (A) [^{o}] as compared between calculations and experimental results [41]. For comparison, the RMSD between the methods are also shown, which for cytosine implies the DFT calculation with the heavier basis set vis-à-vis MP2.

	$\sigma_{rms}^{Calc-Exp}$							
	B3LYP				MP2			
	6-31G(d,p)		6-31+G(3df,p)		aug-cc-pVDZ		$\sigma_{rms}^{DFT-MP2}$	
	B	A	B	A	B	A	B	A
Adenine	7.8	0.8	—	—	15.9	0.8	9.8	0.3
Thymine	14.1	1.2	—	—	17.0	1.2	7.3	0.3
Guanine	20.0	1.4	—	—	22.5	1.4	9.3	0.4
Cytosine	18.4	1.4	17.9	2.0	21.7	1.9	9.0	0.4

The methods show roughly equal accuracy, and the bond lengths all range within the precision of the experimental results, here being given by averaged geometries from high-resolution X-ray crystallographic structures as described in [41] and [9].

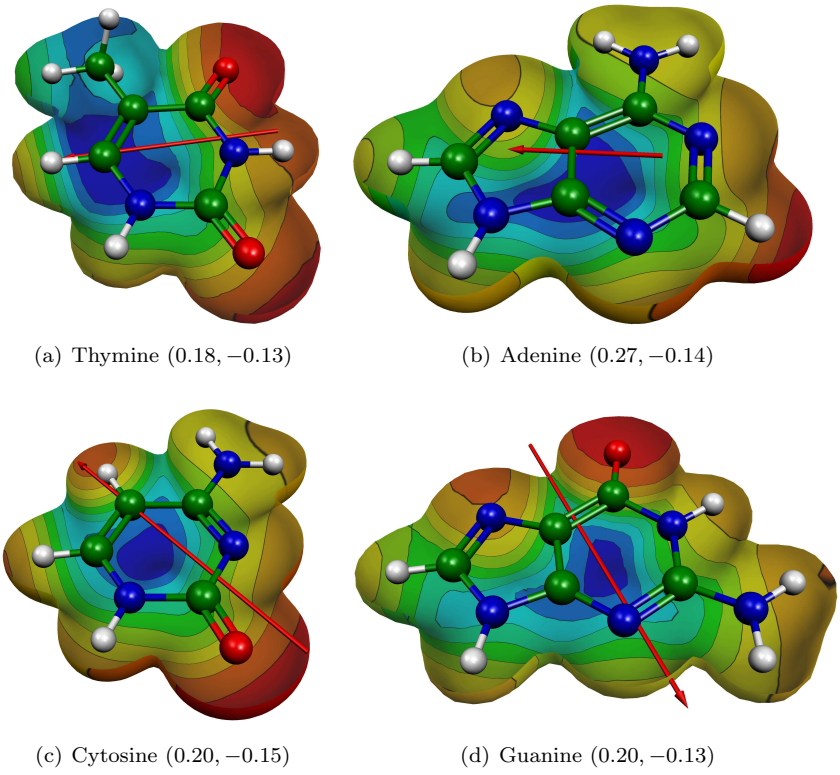
The largest dissemblance in bond lengths for MP2 are found in aromatic C–C and C–N bonds which are typically underestimated by 20-30 mÅ. The largest overall dissimilarity is in the guanine C6-N1 bond, which differs by 40-50 mÅ by both methods.

The bond angles, however, are not as well described, neither by DFT nor perturbation theory, and the deficiencies seem not to be remedied by the inclusion of diffuse functions. On the other hand, the methods show consistent results with low RMSDs when compared with each other.

Total Energies and Densities

The MP2 geometries obtained in the previous section were subsequently used to calculate the SCF orbital energies and the MP2 correction to the total energy, using the same aug-cc-pVDZ basis set.

Classical molecular electrostatic maps, derived from the Mulliken atomic charges, are mapped onto the MP2 density and presented along with scaled dipole vectors in figure 5.4.



(a) Thymine $(0.18, -0.13)$ (b) Adenine $(0.27, -0.14)$

(c) Cytosine $(0.20, -0.15)$ (d) Guanine $(0.20, -0.13)$

Figure 5.4: The molecular electrostatic potential as given by the Mulliken atomic charges mapped onto the total MP2 density isosurfaces for the four bases. The values beneath each picture are the charge ranges represented by the colors, red being negative and blue positive. Also shown in these pictures are the calculated dipole moment vectors, scaled by 0.5.

In addition, the total energies and dipole moments of the bases, from the SCF and MP2 methods, are given in table 5.2.

TABLE 5.2: The total energies at the SCF $E^{(0)}$ and MP2 $E^{(2)}$ level are shown together with the second order correction and the dipole moments of the bases, as given by the SCF and MP2 densities. Here, the energies are given in atomic units, or *Hartrees* [H], and the dipole moments in *Debye* [D].

		Adenine	Guanine	Thymine	Cytosine
Total Energies	$E^{(0)}$	-464.58	-539.47	-451.57	-392.67
	$E^{(2)}$	-466.10	-541.17	-452.97	-393.91
	$\Delta^{(2)}$	-1.51	-1.71	-1.41	-1.24
Dipole Moments	$\mu^{(0)}$	2.51	7.39	5.00	7.42
	$\mu^{(2)}$	2.56	6.63	4.34	6.55

Orbital Energies and Ionization Potentials

The SCF densities and orbital energies obtained in the previous section were subjected to a one body Green's function correction, termed *the Outer Valence Green's Function* method (OVGF), which corresponds to one electron — one hole rearrangments in the outer valence region during ionization [8]. The orbital energies are presented in table 5.3 along with the resulting ionization potentials and pole strengths.

Whereas one body corrections will accurately account for the relaxation effects in the outer valence region, many body Green's function methods are required to describe ionization processes from the inner valence shells. Since such methods are not bundled with the `Gaussian 03` package used here, the results from the three-body Green's function method ADC(3) as calculated by Trofimov *et al.*[60] are also included in table 5.3 for comparison.

Finally, having accounted for at least some of the relaxation effects, the ionization potentials arising from these SCF densities can be compared to experimental energies of adenine, cytosine and thymine, which have been obtained fairly recently by photoelectron spectroscopy [60].

These spectra are included here[4] and are plotted together with the ionization energies and pole strengths from the present OVGF calculation[5], in figure 5.5. In addition, the relaxation effects and concominant orbital energy changes are shown for all four bases in figures 5.6 and 5.7.

Interestingly, extensive relaxation effects can be seen in the nitrogen and oxygen lone pair orbitals of the bases, which are also plotted in figures 5.8 and 5.9.

5.3 Stacked Dimers

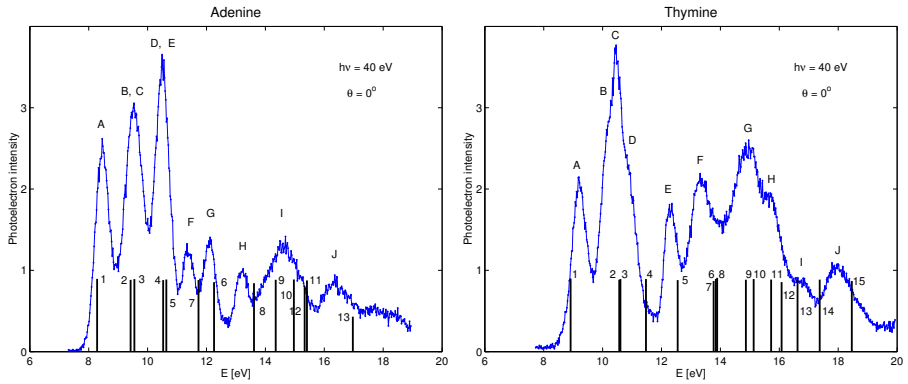
Continuing with the MP2 optimized structures from previously, stacked dimers corresponding to singly stranded 5′−CC−3′, 5′−GG−3′ and 5′−GC−3′ B-DNA were produced with the `ssDNA` script, as described previously.

[4] Courtesy of D.M.P. Holland.

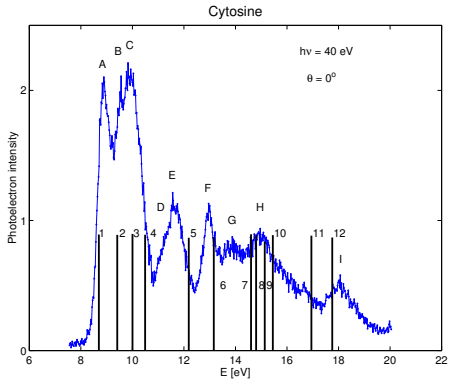
[5] However, since the spectrum of cytosine is dominated by another tautomer, no peak to orbital energy comparison is made in this case.

Table 5.3: SCF orbital energies and types presented with one body outer valence Green's functions calculations (OVGF), as compared with higher order three body ADC(3) calculations and photoelectron spectra from Trofimov *et al.* [60]. Here, all energies are given in $[eV]$.

			aug-cc-pVDZ				Trofimov *et al.*[60]		
			SCF	OVGF			ADC(3)		Exp.
	MO	Type	E	N^o	E	P	E	P	E
Adenine	25a	π_6	8.33	1	8.29	0.90	7.93	0.89	8.47
	24a	π_5	10.11	2	9.42	0.88	9.36	0.88	9.54
	23a	$\sigma_{N\,LP}$	11.17	3	9.55	0.89	9.20	0.89	9.45
	22a	π_4	11.50	4	10.53	0.88	10.26	0.87	10.45
	21a	$\sigma_{N\,LP}$	12.32	5	10.64	0.89	10.23	0.89	10.51
	20a	π_3	13.45	6	12.26	0.85	12.09	0.71	12.09
	19a	$\sigma_{N\,LP}$	13.55	7	11.73	0.88	11.20	0.87	11.35
	18a	π_2	15.10	8	13.61	0.84	13.54	0.65	13.20
	17a	σ	16.05	9	14.35	0.88	14.47	0.81	14.58
	13a	σ	18.36	10	16.97	0.43	16.80	0.64	16.35
Thymine	24a	π_6	9.60	1	8.91	0.90	8.59	0.90	9.19
	23a	π_5	11.79	2	10.56	0.88	10.00	0.87	10.45
	22a	$\sigma_{O\,LP}$	12.18	3	10.60	0.89	10.12	0.88	10.14
	21a	$\sigma_{O\,LP}$	13.03	4	11.47	0.89	10.95	0.88	10.89
	20a	π_4	13.86	5	12.55	0.88	12.31	0.84	12.27
	18a	π_3	15.15	6	13.77	0.87	13.31	0.55	13.31
	15a	σ	16.76	7	15.13	0.89	14.89	0.80	14.90
	14a	σ	17.42	8	15.72	0.89	15.16	0.47	15.75
	11a	σ	19.26	9	17.37	0.89	17.50	0.84	16.86
Cytosine	21a	π	9.33	1	8.70	0.89	8.17	0.88	8.89
	20a	π_5	10.43	2	9.41	0.89	8.93	0.89	9.89
	19a	$\sigma_{N\,LP}$	11.41	3	10.00	0.90	9.46	0.89	9.55
	18a	$\sigma_{O\,LP}$	12.07	4	10.49	0.89	9.91	0.88	11.20
	17a	π_4	13.27	5	12.19	0.87	11.90	0.78	11.64
	16a	π_3	14.58	6	13.16	0.87	12.92	0.80	12.93
	15a	σ	16.09	7	14.60	0.89	14.11	0.83	13.86
	14a	σ	16.57	8	14.80	0.89	14.81	0.81	14.94



(a) Adenine (b) Thymine



(c) Cytosine

FIGURE 5.5: The photoelectron spectra of adenine, thymine and cytosine as reported by Trofimov *et al.* [60] together with the energies and pole strengths of the present calculations, with the pole numbering as given in table 5.3.

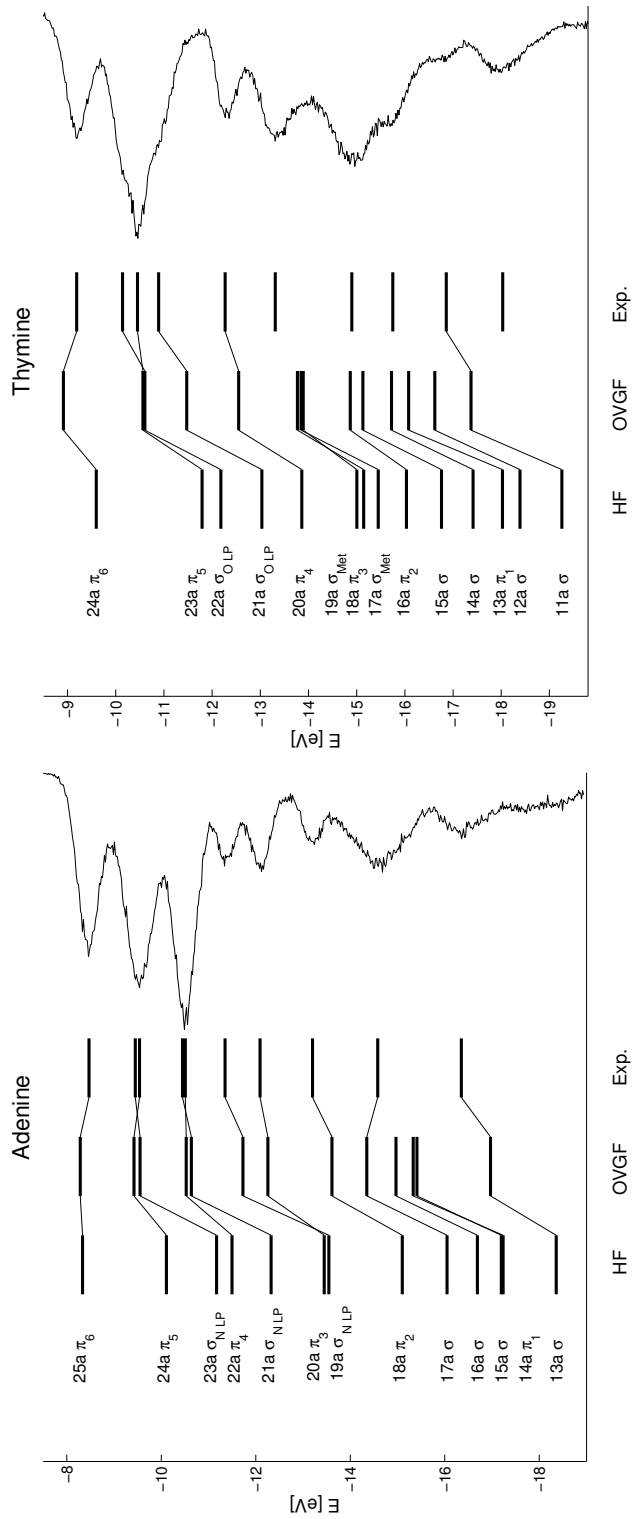
They were all subjected to a rigid potential energy scan (PES) at the counterpoise corrected MP2/aug-cc-pVDZ level, varying the twist parameter in the range $\omega = [0, 80]^o$, in steps[6] of 5^o. All other parameters of the bases, including bond lengths, angles and dihedrals, were kept fixed, under the assumption that the stacking energies are too slight to significantly affect them.

The settings of the frosen helical parameters were chosen to essentially comply with those given in table 4.5 on page 41, and are presented in table 5.4. The reasons behind the differences between these values and table 4.5, the most prominent being the sign change of slide, are elaborated in the discussions.

An attempt was made to do a relaxed[7] PES of the cytosine dimer, but it was eventually abandoned due to convergence problems.

[6]With the exception of the $\omega = 75^o$ point which was not calculated for the GG and CC dimers.

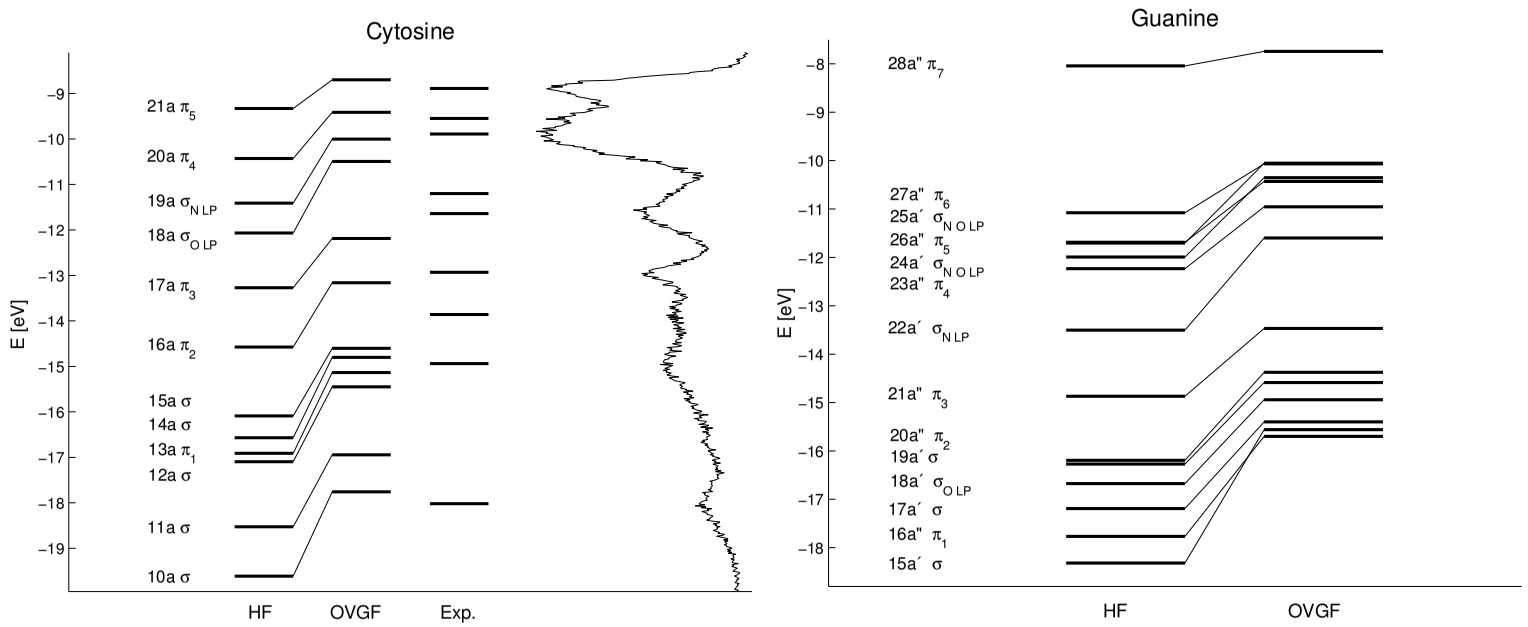
[7]Meaning that the structures of the bases are subjected to optimization for each value of the scanned parameter



(a) Adenine

(b) Thymine

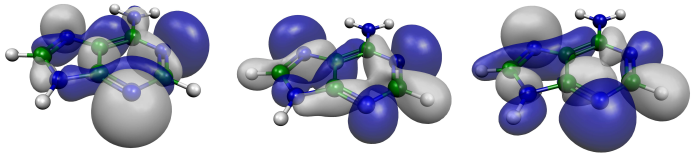
Figure 5.6: A graphical representation of the OVGF correction to the orbital energies for adenine and thymine, as calculated here, and the energies of the spectra given in figure 5.5. (See also figure 5.7.)



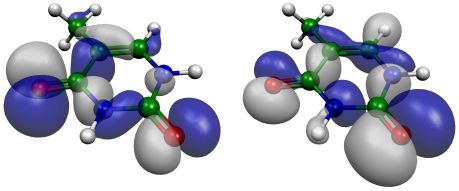
(a) Cytosine

(b) Guanine

Figure 5.7: A graphical representation of the OVGF correction to the orbital energies for cytosine and guanine, as calculated here, and the energies of the cytosine spectrum given in figure 5.5. (See also figure 5.6.)

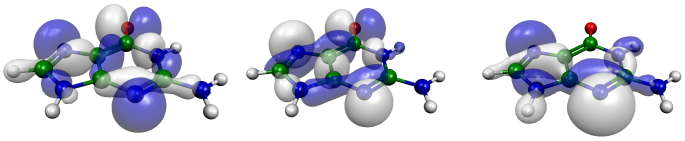


(a) A — 23a $\sigma_{LP\,N}$ (b) A — 21a $\sigma_{LP\,N}$ (c) A — 19a $\sigma_{LP\,N}$

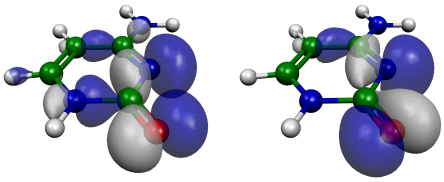


(d) T — 22a $\sigma_{LP\,O}$ (e) T — 21a $\sigma_{LP\,O}$

Figure 5.8: The nitrogen and oxygen lone pair orbitals of adenine, in (a), (b) and (c), along with those of thymine, in (d) and (e).



(a) G — 24a $\sigma_{LP\,O}$ (b) G — 25a $\sigma_{LP\,O}$ (c) G — 22a $\sigma_{LP\,N}$



(d) C — 19a $\sigma_{LP\,O}$ (e) C — 18a $\sigma_{LP\,O}$

Figure 5.9: The oxygen and nitrogen lone pair orbitals of guanine, in (a), (b) and (c), together with those of cytosine, in (d) and (e).

Table 5.4: Frozen helical parameters in the dimer calculations, in [^{o}] and [Å].

	Dx	Dy	Dz	τ	ρ
CC	0.06	-0.38	3.28	0	0
GG	0.06	-0.38	3.28	0	0
GC	0.00	-0.45	3.37	0	0

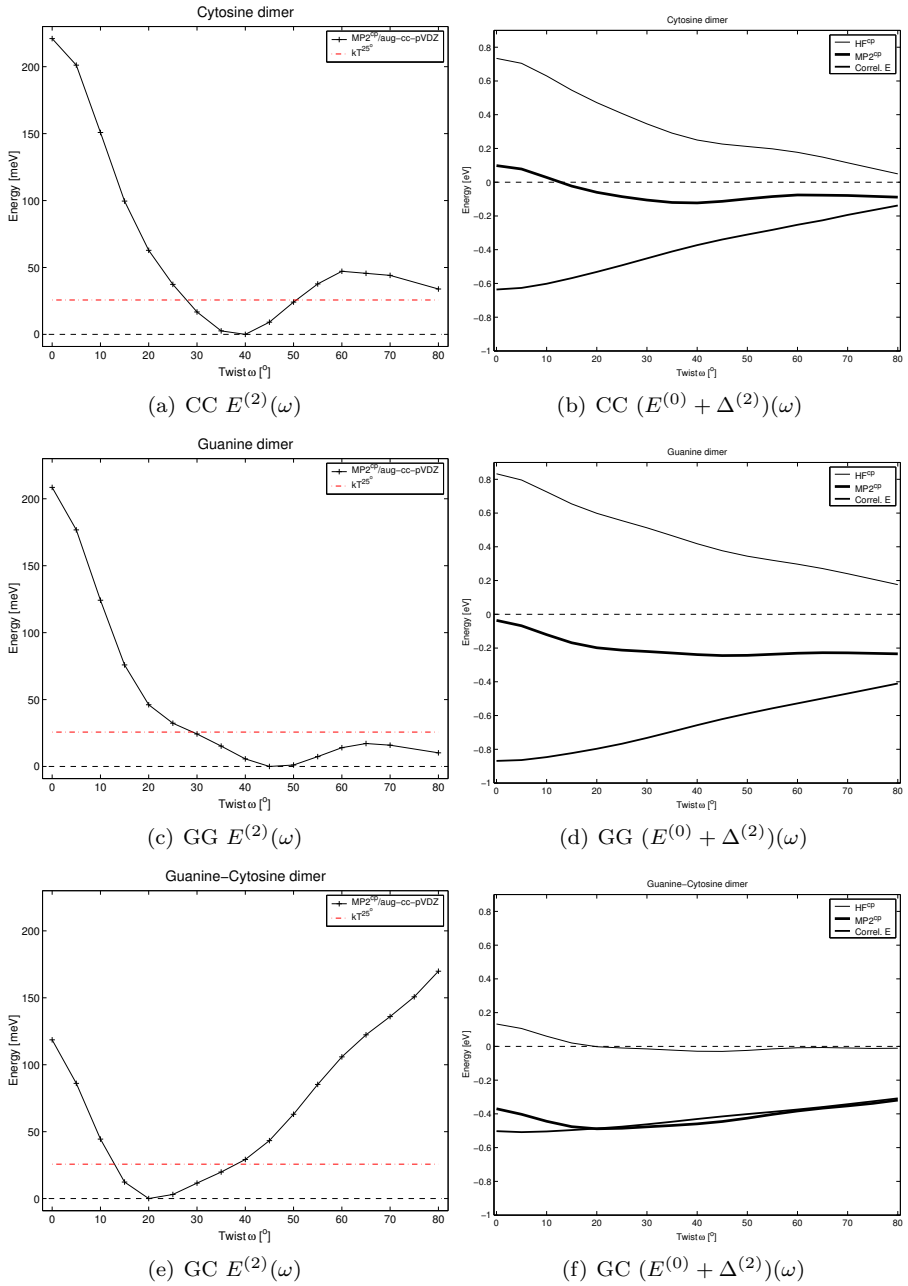
Dimer Energies

The single point energies from the potential scans are plotted with respect to the minimum energies in the left column of figure 5.10. Also, to obtain estimates of the relative contributions from the electron correlation energy given at this level of theory, the counterpoise corrected MP2^{cp} and SCF^{cp} monomer energies were subtracted from the corresponding dimer total energies. The resulting potential curves are plotted in the rightmost column of figure 5.10.

The total and relative MP2^{cp} and SCF^{cp} energies for $\omega = 0^o$ and the minima, i.e. where we estimate that $\partial_\omega E \to 0$, are also given in table 5.5.

TABLE 5.5: The total energies at the SCF^{cp} ($E^{(0)}$) and MP2^{cp} ($E^{(2)}$) levels of theory for the dimers, taken at $\omega = 0^o$, are presented with their respective approximate minima and corresponding $E^{(2)}$ total energy difference.

				Min			
	$E^{(0)}$ $[H]$	$E^{(2)}$ $[H]$	ω $[^o]$	$E^{(0)}$ $[H]$	$E^{(2)}$ $[H]$	ω $[^o]$	$\Delta_{0^o}^{min}$ [meV]
CC	-785.321	-787.818	0	-785.338	-787.826	40	-221
GG	-1078.906	-1082.352	0	-1078.923	-1082.360	45	-208
GC	-932.142	-935.100	0	-932.137	-935.104	20	-119



(a) CC $E^{(2)}(\omega)$ (b) CC $(E^{(0)} + \Delta^{(2)})(\omega)$

(c) GG $E^{(2)}(\omega)$ (d) GG $(E^{(0)} + \Delta^{(2)})(\omega)$

(e) GC $E^{(2)}(\omega)$ (f) GC $(E^{(0)} + \Delta^{(2)})(\omega)$

Figure 5.10: The calculated energies of the GG, CC, and GC dimers given with respect to the minimum value, along with the value of kT^{25^o}, in (a), (c) and (e), and the monomer energies in (b), (d) and (f). In the right hand side diagrams, the relative contributions from the SCF and the electron correlation to the second order are shown.

6 Discussion

Seek simplicity — and distrust it!

Alfred North Whitehead

In this chapter, the calculated potential curves will be discussed, interpreted and compared with the statistical data of double stranded DNA. The conclusions from these sections are summarized in section 6.3.

6.1 The Dimer Interactions

There are at least three different kinds of interaction that are expected to take place in the dimers, namely electrostatic interaction, electron correlation and direct orbital interaction.

To determine the effects of these interactions, and their ability to give an account for the observed potential curves, they will be briefly examined in this section.

Electrostatic interaction

To estimate the electrostatic multipole interaction energy, we may begin by noting from table 5.2 on page 50 that both guanine and cytosine have strong dipole moments.

Since higher order multipoles are generally not expected to make a significant contribution to the expansion of the charge density if the leading dipole term is non-vanishing, this estimation is reduced to merely evaluating the dipole-dipole interaction energy.

The dipole-dipole interaction energy can be explicitly calculated by the following simple formula[1],

$$U_{DD} = \frac{1}{r^3}\left[(\mathbf{p}_1 \cdot \mathbf{p}_2) - 3\,(\mathbf{p}_1 \cdot \hat{\mathbf{r}})\,(\mathbf{p}_2 \cdot \hat{\mathbf{r}}) \right] \quad , \tag{6.1}$$

in this case given that

$$\mathbf{r} = Dz\hat{\mathbf{z}} + R_{\hat{\mathbf{z}}}(\omega)\mathbf{t} - \mathbf{t} \quad , \tag{6.2}$$

where $\mathbf{t}$ is the vector from the origin of the Standard Reference Frame to the `Gaussian` standard orientation, in terms of which the dipole vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ are provided from the calculations.

[1] Conveniently given in atomic units, where $4\pi\varepsilon_o = 1$.

Furthermore, approximating shift and slide to zero, the dipole vectors will merely transform by a rotation about the rise vector. Thus, keeping $\mathbf{p}_1$ fixed,

$$\mathbf{p}_2' = R_{\hat{\mathbf{z}}}(\omega)\,\mathbf{p}_2 \quad ,$$

where R is an SO(3) rotation matrix.

In this approximation, the dipole-dipole interaction energy for GG and CC at $\omega = 0^o$ reduces to

$$E_{DD} = \frac{\mu^2}{r^3} \quad ,$$

due to symmetry, where μ is the dipole moment of guanine and cytosine respectively, and $r \equiv Dz$ for the corresponding dimer, whose values are given in table 5.4 on page 55. Evaluating this expression for the CC and GG dimer gives 0.76 eV and 0.78 eV, which compares to the calculated HF^{cp} energies of 0.73 eV and 0.83 eV .

Since these interaction energies are in fair agreement at $\omega = 0^o$, the higher order multipole contributions are expected to be small. In addition, the dipole-dipole potential is expected to remain similar throughout the transformation of the dipole vectors, if the electrostatic model is indeed valid.

However, evaluating the full expression of the dipole-dipole potential gives the potential curve plotted in figure 6.1 where it can be seen that this model is only valid in the vicinity of $\omega = 0^o$, for GG and CC, and gives a quite incorrect picture for the GC dimer.

Actually, whereas it is tempting to believe that the large deviation seen in the GC dimer dipole-dipole curve in 6.1(c) is due to a miscalculation, this result also follows from intuition. In figure 5.4 on page 49, the dipole vectors of guanine and cytosine are seen to run in almost opposite directions, which consequenlty also holds true at $\omega = 0^o$ where the bases are stacked in an approximately parallel fashion. As ω increases, the vectors are expected to reach a point where they are oppositely directed, where the negative attractive energy is at its maximum, and then diminish as the dipole interaction declines with increasing r. This is exactly what is seen in the dipole-dipole curve.

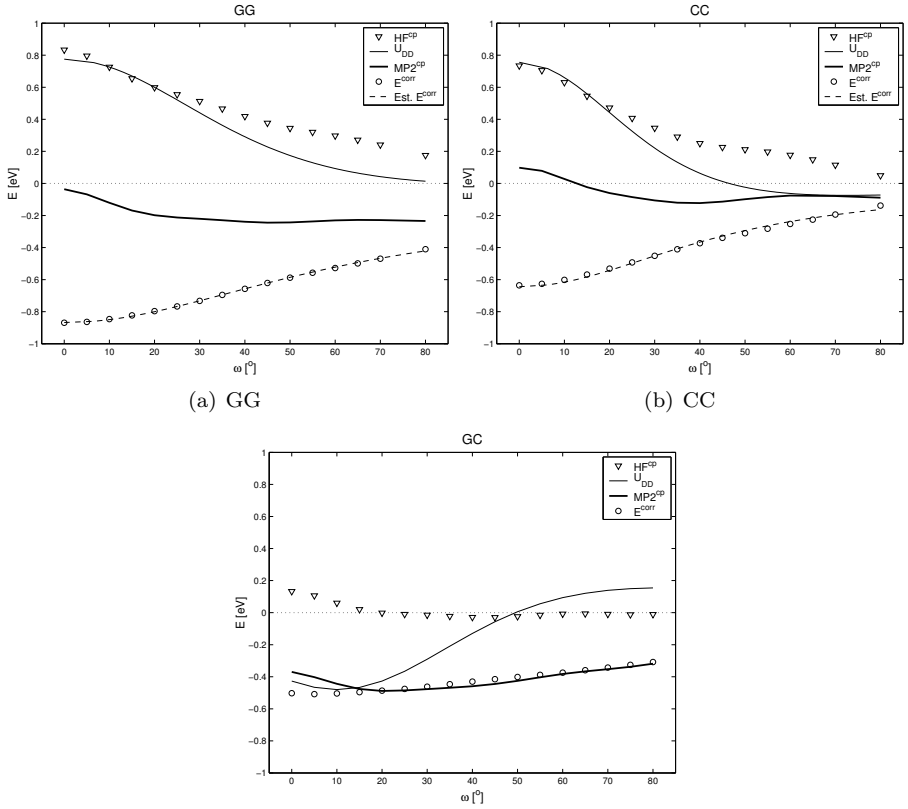
On the contrary, the HF^{cp} GC energy starts out at its maximum and then quickly declines towards zero, where it stays for the remainder of the transformation, leaving the electron correlation as the major contributor to the total potential energy at this level of theory.

From these considerations, we draw the conclusion that the *ab initio* calculated potential curves cannot be adequately explained by simple electrostatic dipole interaction, and that higher order multipoles are probably not able to account for the observed differences, although this remains to be shown explicitly.

Electron correlation

In the potential plots of the neon dimer, it seems that the electron correlation energy, as given by the Møller-Plesset perturbation theory, is fairly well decribed by the van der Waals term in the Lennard-Jones potential, as shown in figure 5.3 on page 47, at least at the fourth order correction with the augmented quintuple zeta basis set.

To substantiate this, we may perform a coordinate change by simply taking the natural logarithm of r and E^{corr}, which for some levels of theory are plotted



(a) GG (b) CC

(c) GC

Figure 6.1: The quantum mechanical potential curves, provided with the dipole-dipole potential energy as given by U_{DD} in equation 6.1, and the least-squares fitted logaritmic curves of the electron correlation in dashed lines, after the removal of the angular dependencies for GG and CC.

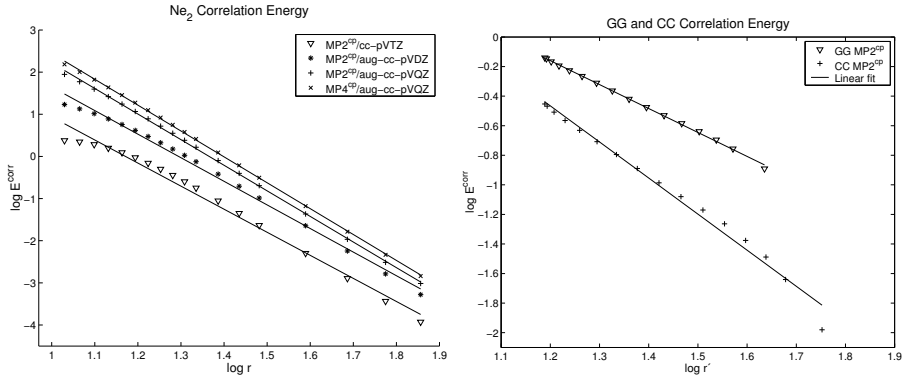
in figure 6.2(a). From this plot, it can be seen that the correlation energy is increasingly well described by the simple first order polynomial,

$$\log E^{corr} = \log(k) - a\log(r) \Longleftrightarrow E^{corr} = k\frac{1}{r^a}$$

where the corresponding exponents a can estimated by a linear least-squares fit.

The resulting exponents of the linear fit as shown in figure 6.2(a) are given in table 6.1, and are seen to converge towards r^{-6}, which is one of the characteristic properties of the van der Waals interaction.

The remaining differences at the high level MP4$^F C$/aug-cc-pV5Z calculation between the experimental and calculated spectroscopic constants of Ne_2, may be due to the omittance of core-core and core-valence correlation in the frozen core approximation. Judging from the core correlation effects on the spectroscopic constants of the $X\,^1\Sigma^+$ state in the CH^+ molecule as calculated by Taylor *et al.* [3], the differences are seen to be in the same order of magnitude with respect to D_e.



(a) Ne_2 (b) GG & CC

FIGURE 6.2: Logarithmic plots of the neon dimer correlation energies, in (a), and the base dimers, GG and CC in (b), after the removal of the angular dependencies .

TABLE 6.1: The exponents of the least-squares fit to the correlation energy in the neon dimer, as shown in figure 6.2(a).

	MP2			MP4
r^{-a}	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVQZ	aug-cc-pVQZ
a	5.47	5.59	6.07	6.14

Before attempting to interpret the decline in the electron correlation energy for the GG, CC and GC dimer, it would be beneficial to remove the angular dependencies in these curves, and instead try to relate them to some distance parameter[2] r'.

Actually, suitable coordinate transformations for the GG and CC dimers are easily found due to symmetry, for instance by choosing the distance between the centers of the aromatic rings. The corresponding coordinate transformation is given by

$$r' = \sqrt{4 \sin^2\left(\frac{\omega}{2}\right) ||\mathbf{v}||^2 + ||Dz\,\hat{\mathbf{z}}||^2} \quad ,$$

where $\mathbf{v}$ is a vector pointing from one monomer to the other. A linear least squares fit in the logaritmic plot of r' and E^{corr} is shown in figure 6.2(b).

Performing an inverse coordinate transformation on the least-squares fitted lines back to the angular representation gives the dashed curves shown in figures 6.1(a) and 6.1(b). However, no suitable choice of coordinates was found for the asymmetric GC dimer.

In addition, it can also be seen directly from figures 5.10(b), 5.10(d) and 5.10(f) on page 57 that the correlation energy is monotonously declining for all dimers, and that the specific features of the calculated *ab initio* potentials are

[2]It is expected that the correlation energy will depend on the spatial proximity of electron densities, even if the choice of parameter here is only approximative.

inherited from the HF energy rather than the correlation energy.

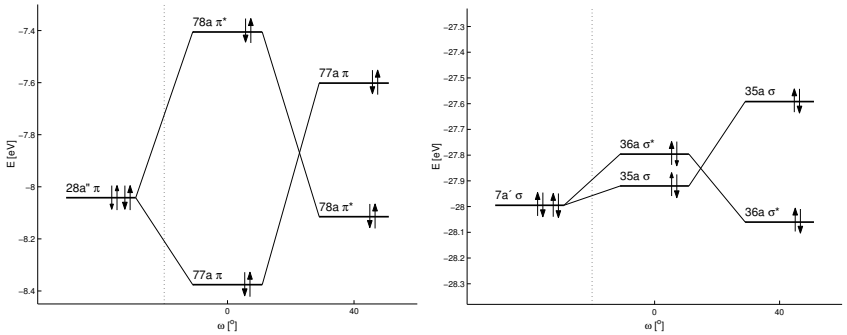
From these considerations, we draw the conclusion that whereas the correlation energy seems manifestly isotropic, the details of the potential curves are determined by the zeroth order Hartree-Fock energy.

Since we, by the previous paragraph, are unable to draw any conclusion about the HF energy by considering the dipole-dipole interaction, the orbitals themselves need to be examined.

Orbital interaction

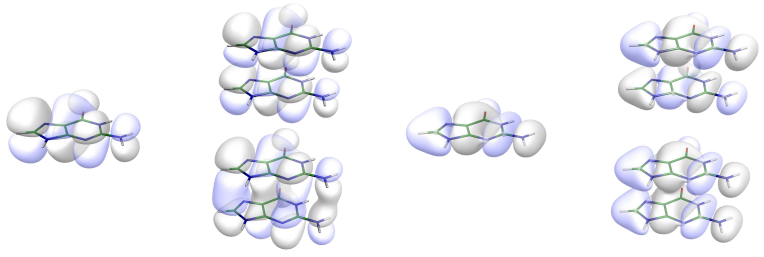
The dimer orbitals can in most cases be regarded as linear combinations of the molecular orbitals found in the monomers, and their corresponding orbital energies are subjected to energy splittings. As such, it is expected that the π orbital energies of the dimer experience a larger splitting than the σ orbitals, as a direct consequence of the π interaction.

To give an example of this, the energy splittings and orbital renderings of the HOMO π, HOMO-1 π, 35a σ and 36a σ orbitals of the GG dimer, which are made up of combinations from the 22a" π_7 HOMO and 7a' σ_7 orbitals of the guanine monomers, are shown in figure 6.3.



(a) π orbital splitting

(b) σ orbital splitting



(c) Dimer π orbitals

(d) Dimer σ orbitals

FIGURE 6.3: The π and σ orbital splitting as observed in the guanine dimer, in (a) and (b) respectively, together with renderings of the corresponding orbitals at $\omega = 0^o$ with isosurface probabilites $p = 0.01$, in (c) and (d).

In these diagrams, the scale is taken to be the same, and the considerable difference in the energy splitting between the π and σ orbitals at $\omega = 0^o$ is

apparent.

The GG dimer. Plotting the energies of all the valence orbitals in the guanine dimer, given the range of $\omega = [0, 80]^o$, reveals that the σ splitting consistently increases as the total energy is lowered, shown in figure 6.4.

At the same time, the two highest π orbitals undergo a decrease in energy splitting, which eventually leads to a rearrangement of the orbitals, estimated to occur somewhere between 20^o and 30^o. Judging from this plot, it seems likely that all of the π orbitals undergo a similar transformation.

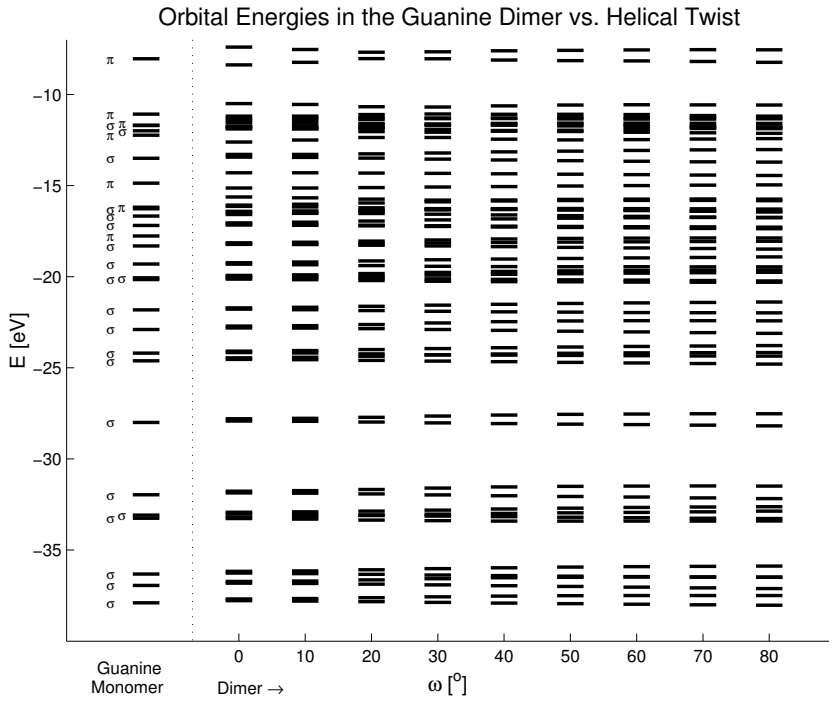


FIGURE 6.4: The orbital energies of the guanine monomer is given left of the dotted line, whereas the orbital energies of the guanine dimer in the range of $\omega = [0, 80]^o$ is shown to the right. In this plot, the whole of the valence region is encompassed.

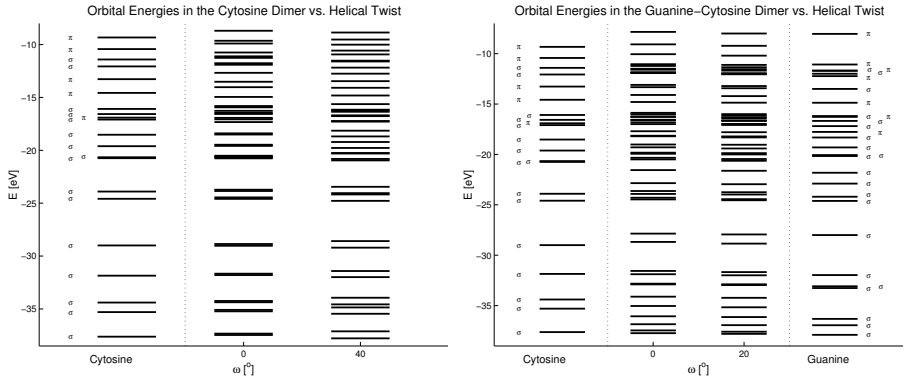
At this point it should be stressed that as the energies are split up and approach neighbouring single molecular orbital energies, they will in turn start to make significant contributions to the dimer orbitals and consequently often give rise to a blend which share little likeness with the 0^o state. In fact, some of the orbitals in the dense regions seen in figure 6.4 can hardly be considered to be made up of well determined single molecular orbitals at all, save for in the 0^o conformation.

With this in mind, the general behaviour of the orbitals in the guanine dimer during this transformation seems to be to move away from the near degeneracy of the σ orbitals at 0^o, which at the same time decreases the energy splitting of the π orbitals. Whereas the energy contributions from the π orbitals are

estimated to be very slightly increased, the σ orbital energies are significantly lowered, which, as such, is reminiscent of the *pseudo Jahn-Teller effect*.

The CC dimer. Adopting such a reasoning, we would expect that the picture is similar in the CC dimer, since it is also symmetric at 0^o. Indeed, by regarding the energies plotted for 0^o and the minimum point at 40^o in figure 6.5(a), this seems also to be true.

The σ orbitals are very slightly split, and hence nearly degenerate, in the 0^o confromation, whereas the splitting increases when moving away from the symmetrical state, which results in a significant overall energy lowering of the orbital energies.



(a) Cytosine dimer (b) Guanine-Cytosine dimer

Figure 6.5: The orbital energies of the cytosine and guanine-cytosine dimer in the valence region for $\omega = 0^o$ and the minima are shown along with the monomer energies separated by dotted lines.

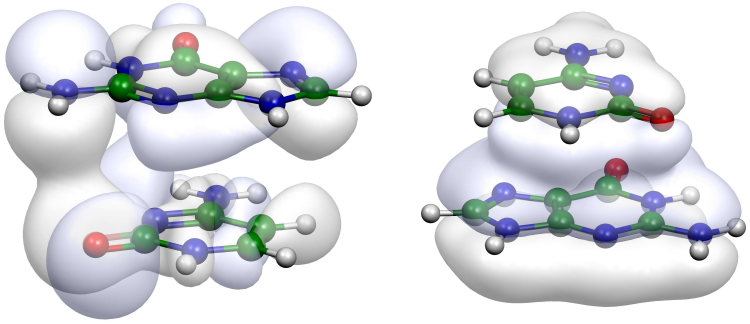
Finally, seeing that this effect is not present in the manifestly asymmetric GC-dimer, whose energies are shown in figure 6.5(b), reinforces the idea that symmetry and orbital degeneracies may play a role in determining the energetics of the vertical base stacking.

The GC dimer. In the GC dimer, where the total HF energy lowering is modest and the potential curve is mostly determined by the electron correlation energy, the orbitals are seen to have stronger monomer character, either derived from the guanine or the cytosine, which is natural when taking the energy differences between the different monomer orbitals into account.

Nonetheless, in the cases where guanine and cytosine orbitals are found in energetical proximity, splitting does occur, as is seen in figure 6.5(b).

Here, the energy lowering is given by factors which have not been elucidated here. However, among the orbitals that lower their energy the most, interesting features such as *lone pair–lone pair* interaction, as shown in figure 6.6(a), are observed.

Although lone pair–lone pair interaction would be an oxymoron in the valence bond description, it is frequently spotted also in the symmetrical dimers and actually appears in the π orbital example of figure 6.3(c).



(a) GC 59a (b) GC 46a

FIGURE 6.6: Interesting features of the dimer orbitals for the guanine-cytosine dimer, given at $\omega = 20^o$. (a) Lone pair-lone pair interaction is seen in orbital 59a. (b) An example of the observed direct π-interaction.

6.2 Comparison with double stranded DNA

Although we have only considered base dimers of single stranded DNA, it may be of some interest to compare the potential curves with the statistical distributions that were calculated in section 4.4. Remembering that the GG and CC dimers are complementary, they will both be compared with the dsGG twist distribution as given in figure 4.3(e) on page 42, whereas the GC dimer will be rendered with the dsGC distribution in figure 4.3(f).

Before doing so, there are a number of weaknesses in the dimer description that need to be discussed.

To begin with, the slide parameters in the dimers are about ~ 0.8 Å off from the intended mean value, due to the sign error in `dsDNA`. This will have an impact on the energy but it will probably be slight, as judged from previous undisclosed calculations on the cytosine dimer at $Dx = Dy = 0$ Å. Furthermore, the overall slide distribution for the dinucleotides is seen to be almost symmetrical about $Dy = 0$ in figure 4.2 on page 40, with a peak near -0.5 Å, which may also justify a qualitative comparison with the B-DNA data set.

Secondly, since we have suggested in the previous section that the energies in this system may very well rely on σ orbital conformation and energies, which are subsequently affected by the hydrogen bonds in double stranded DNA, it is important to remember that these hydrogen bonds have not been taken into account in the dimer description, thus possibly leaving out important additional quantum mechanical effects.

Thirdly, comparing with the neon dimer model system, e.g. figure 5.1(b) on page 45, the theoretical level to which the dimer potential curves have been calculated is too modest to warrant any claim of quantitative agreement.

Finally, since the sugar-phosphate backbone is also absent from the dimer description, there is no way of telling whether or not the ω angles that differ greatly from those in ideal B-DNA are permissible, or if the reported energy lowerings of $\sim 0.1 - 0.2$ eV per dimer are enough to make a difference when

compared to the corresponding energies of the complete system, i.e. where the backbone is included.

Nevertheless, the potential curves that are plotted along with the distributions in figure 6.7 show some degree of qualitative similarity, suggesting that the energies that have been calculated here may indeed have implications in the biologically active B-DNA double helix.

6.3 Conclusions

From the deliberations outlined above, the following conclusions can be drawn.

i. The nucleobases guanine and cytosine prefer not to be stacked symmetrically on top of each other.

ii. When given the opportunity to rotate about a vector corresponding to the B-DNA helix axis, they will do so and take on a conformation which is similar to that found in biologically active DNA.

iii. The cause behind this preferred conformation cannot be explained by means of classical electrostatics — at least not at the level of molecular dipole interaction — but is seen to be mainly quantum mechanical.

iv. Among the quantum mechanical effects, the electron correlation — which is shown to correspond to the chemical van der Waals interaction — is very important but is also seen to be chiefly isotropical.

v. The specific and anisotropic quantum mechanical effects are attributed to the zeroth order Hartree-Fock solutions.

vi. There is clear evidence of both π and σ orbital interaction that confers the energy lowering in the Hartree-Fock solutions of the stacked nucleobases.

In addition to this, the energy lowering in the Hartree-Fock solution is suggested to be caused by a *pseudo Jahn-Teller*-like destabilization of the σ orbitals in the symmetric dimers, whereas the cause remains somewhat elusive for the assymetric guanine-cytosine dimer. Presumably, the latter is due to specific orbital interactions, such as the observed lone pair-lone pair interactions.

6.4 Future Improvements

dsDNA

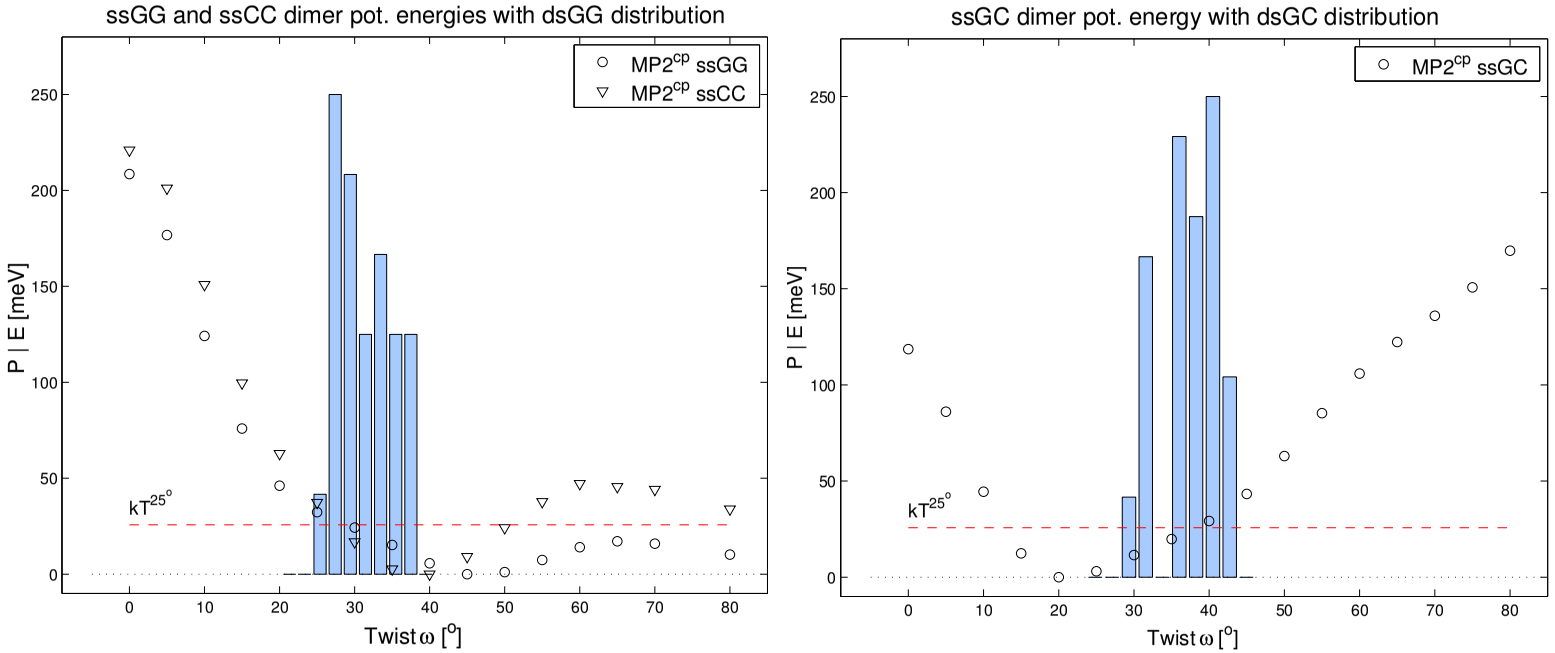
The most important improvements of `dsDNA` for future versions are

i. The inclusion of both signs for the translational helical parameters,

ii. a correct account of the important intra-base-pair *propeller twist* parameter and

iii. a description of the remaining local base step coordinates.

The first point stems from the fact that bond lengths in a Z-matrix cannot be negative, so sign-changes at this level must be given indirectly in terms of bond angles and dihedrals in the scaffold. It is concievable that future specifications will be by input files rather than by prompt arguments.



(a) ssGG & ssCC vs. dsGG

(b) ssGC vs. dsGC

Figure 6.7: The calculated MP2 potential curves plotted with the helical twist distributions of the double stranded B-DNA dinucleotide data set described in section 4.4.

Base interaction energies

The next step in solving the grand problem is to perform calculations on base-pairs, rather than base dimers, to the same level of theory or higher, using the same tranformations.

Perhaps the greatest weakness in the approach used in this study is that it is questionable whether the settings at $\omega = 0^o$ have any biological relevance at all. Thus, whereas we already know that such a conformation is not *favourable*, since we do not observe it, we would also need to verify whether it is at all *permissible*.

To remedy this, it would be interesting to verify whether there are experimental structures where base-pairs are found in the energetically unfavourable 0^o conformation and, if so, what the other helical parameters are in those cases.

In a study relying on such conformations, the removal of the sugar-phosphate backbone is easier to motivate since it has already had its say in determining the helical structure.

In fact, quickly skimming 1800 X-ray single crystal structures of double stranded DNA from the NDB, returns 65 base-pair steps, among a total of $\sim$ 24 000, having twist in the range $\omega = [1, 10]^o$.

One of these base-pair steps is found within a structure of the TATA box binding protein [33], when bound to the sequence

$$\begin{array}{c} 3'-\texttt{GCATATATATGC}-5' \\ \quad\;\; |||||||||||| \\ 5'-\texttt{CGTATATATACG}-3' \end{array}$$

where the underlined base-pair step has a twist of $\omega \sim 1^o$.

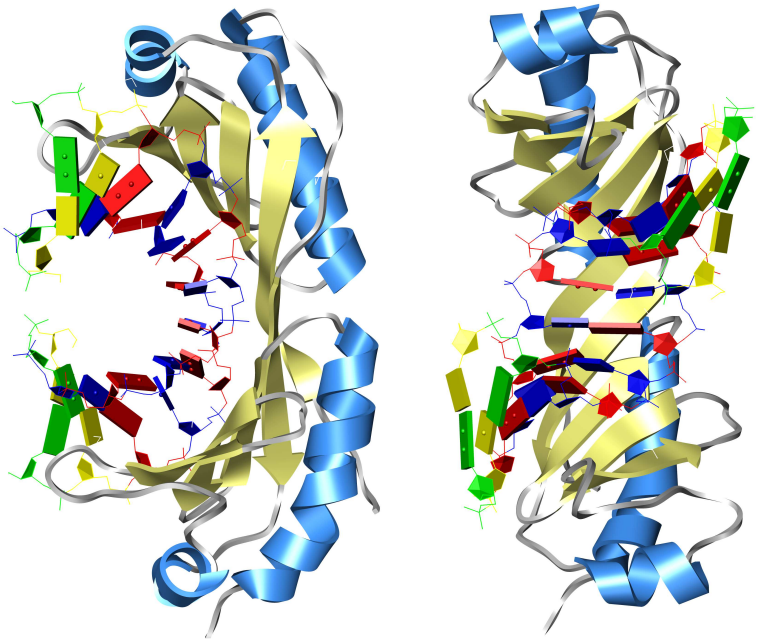
When the other parameters of that base-pair step are compared to the B-DNA averages from table 4.4 on page 40, they are seen to essentially comply, save for the roll parameter, as is shown in table 6.2. The bound protein-DNA complex is also shown in figure 6.8.

On the other hand, since this single observation could prove to be spurious, a future comprehensive study is suggested.

Table 6.2: Comparison between the low twist base-pair step in the bound protein TATA-box complex with the average values of free B-DNA as given in table 4.4 on page 40. The translations are given in [Å] and the angles in [^{o}].

		Dx	Dy	Dz	τ	ρ	ω
B-DNA	$\bar{x}$	0.01	0.30	3.32	-0.04	1.43	34.84
	σ	0.55	0.79	0.30	3.31	5.62	7.91
TATA-box	x	0.80	0.95	3.14	-0.16	21.26	1.46

On a quantum chemical level, partial optimizations for a few points along the potential curves would be valuable to ascertain that the energy is not constrained by the rigidity of the intra-base geometries. The convergence problems that were encountered here could presumably be solved by reducing the size



(a) Side view (b) Front view

FIGURE 6.8: Rendering from the X-ray crystal structure `1TGH`, as reported in [33], with the `Chimera` software [45]. The bases in the base-pair step seen in the middle bear a similar geometry to those computed at $\omega = 0^o$ in this work, save for a larger roll.

of the basis set until convergence is met, and then gradually increasing it until the geometry at the proper level is found.

On the other hand, this is easier said than done since the basis set used here, Dunning's aug-cc-pVDZ, is the lowest level correlation consistent basis set with associated diffuse functions. By considering the most important first row contractions, Pople's split valence 6-31+G(d) basis set is a likely candidate. Comparing $(11s\,5p\,1d) \rightarrow [4s\,5p\,1d]$, with that of aug-cc-pVDZ, $(10s\,5p\,2d) \rightarrow [4s\,3p\,2d]$, we see that Pople's basis set has four d functions less[3] per first row atom than Dunning's, but one 1 s-function more. Whether this is enough to make a difference remains to be seen.

[3]Here, it must be taken into account that Pople's basis set provide 6 cartesian d-functions, whereas Dunning's include 5 pure d-functions per contraction as stated above.

Appendix A

Appendix

A.1 The Standard Reference Frame

Here, the cartesian coordinates of the IUPAC-IUBMB joint commission document *A Standard Reference Frame of for the Description of Nucleic Acid Base-pair Geometry* [41] are presented in table A.1.

TABLE A.1: Cartesian coordinates, given in Å, of the base atoms in ideal Watson-Crick base-pairs as described by the Standard Reference Frame [41], using the nomenclature on page 8. Note that these coordinates are based on average geometries as observed in high-resolution crystal structures.

Adenine		x_0	y_0	z_0	Guanine		x_0	y_0	z_0
	C1′	-2.479	5.346	0.000		C1′	-2.477	5.399	0.000
	N9	-1.291	4.498	0.000		N9	-1.289	4.551	0.000
	C8	0.024	4.897	0.000		C8	0.023	4.962	0.000
	N7	0.877	3.902	0.000		N7	0.870	3.969	0.000
	C5	0.071	2.771	0.000		C5	0.071	2.833	0.000
	C6	0.369	1.398	0.000		C6	0.424	1.460	0.000
	N6	1.611	0.909	0.000		O6	1.554	0.955	0.000
	N1	-0.668	0.532	0.000		N1	-0.700	0.641	0.000
	C2	-1.912	1.023	0.000		C2	-1.999	1.087	0.000
	N3	-2.320	2.290	0.000		N2	-2.949	0.139	-0.001
	C4	-1.267	3.124	0.000		N3	-2.342	2.364	0.001
						C4	-1.265	3.177	0.000
Cytosine		x_0	y_0	z_0	Thymine		x_0	y_0	z_0
	C1′	-2.477	5.402	0.000		C1′	-2.481	5.354	0.000
	N1	-1.285	4.542	0.000		N1	-1.284	4.500	0.000
	C2	-1.472	3.158	0.000		C2	-1.462	3.135	0.000
	O2	-2.628	2.709	0.000		O2	-2.562	2.608	0.000
	N3	-0.391	2.344	0.000		N3	-0.298	2.407	0.000
	C4	0.837	2.868	0.000		C4	0.994	2.897	0.000
	N4	1.875	2.027	0.001		O4	1.944	2.119	0.000
	C5	1.056	4.275	0.000		C5	1.106	4.338	0.000
	C6	-0.023	5.068	0.000		$C5_M$	2.466	4.961	0.001
						C6	-0.024	5.057	0.000

A.2 Dinucleotide Base-Pair Geometry

TABLE A.2: The double stranded DNA X-ray crystal structures used in this thesis, given by the Nucleic Acid Database identification numbers.

UDJ049	UD0029	BDL084	BDL020	BDL005	BDL001
BDJ081	BDJ069	BDJ061	BDJ060	BDJ052	BDJ051
BDJ037	BDJ036	BDJ031	BDJ025	BDJ019	BDJ017
BD0087	BD0084	BD0082	BD0081	BD0080	BD0079
BD0077	BD0073	BD0070	BD0067	BD0066	BD0054
BD0051	BD0041	BD0040	BD0037	BD0036	BD0035
BD0034	BD0033	BD0032	BD0029	BD0023	BD0019
BD0018	BD0006	BD0005	BD0001		

A.3 Base Geometry Optimization

The detailed results from the geometry optimization described in section 5.2, are given for adenine in tables A.3 and A.4, thymine in tables A.5 and A.6, guanine in tables A.7 and A.8, and cytosine in tables A.9 and A.10.

TABLE A.3: Adenine bond lengths given by experimental results and geometry optimization using DFT and MP2, as described in section 5.2. All atoms are denoted by the IUPAC convention on page 8, the bond lengths are given in [Å], with the deviations and RMSDs in [mÅ].

	DFT	MP2	Exp.	Δ_{Exp}^{DFT}	Δ_{Exp}^{MP2}	Δ_{DFT}^{MP2}
N9:C8	1.381	1.379	1.374	-6.8	-4.8	-2.0
C8:H8	1.082	1.088	—	—	—	6.3
C8:N7	1.311	1.336	1.311	0.2	-25.0	25.1
N7:C5	1.385	1.387	1.389	3.8	2.4	14
C5:C6	1.412	1.418	1.405	-6.8	-12.6	5.8
C6:N6	1.353	1.361	1.335	-17.6	-25.5	8.0
N6:H6^a	1.006	1.011	—	—	—	4.1
N6:H6^b	1.006	1.010	—	—	—	4.5
C6:N1	1.345	1.348	1.351	6.0	2.8	3.2
N1:C2	1.344	1.360	1.337	-6.7	-23.3	16.7
C2:H2	1.088	1.093	—	—	—	4.8
C2:N3	1.337	1.346	1.331	-5.7	-14.6	8.9
N3:C4	1.339	1.349	1.343	4.0	-6.4	10.4
N9:H9	1.009	1.014	—	—	—	5.6
			σ_{rms}	7.8	15.9	9.8

Table A.4: Adenine angles given by experimental results and geometry optimization using DFT and MP2, as described in section 5.2. All atoms are denoted by the IUPAC convention on page 8, angles and deviations are given in [^{o}].

	DFT	MP2	Exp.	Δ_{Exp}^{DFT}	Δ_{Exp}^{MP2}	Δ_{DFT}^{MP2}
N9:C8:H8	121.3	121.7	—	—	—	0.4
N9:C8:N7	113.5	113.4	113.7	0.3	0.4	-0.1
C8:N7:C5	103.9	103.4	103.9	0.1	0.5	-0.5
N7:C5:C6	132.6	132.1	132.3	-0.4	0.2	-0.5
C5:C6:N6	122.2	121.9	123.7	1.5	1.8	-0.3
C6:N6:H6^a	119.0	119.2	—	—	—	0.2
C6:N6:H6^b	120.2	120.0	—	—	—	-0.2
C5:C6:N1	118.9	118.8	117.6	-1.3	-1.2	0.0
N1:C2:N3	118.3	118.4	118.6	0.3	0.2	0.1
N1:C2:H2	115.2	115.1	—	—	—	-0.1
N1:C2:N3	129.0	128.9	129.4	0.4	0.5	-0.1
C2:N3:C4	111.1	110.8	110.5	-0.6	-0.3	-0.3
C4:N9:H9	126.7	125.9	—	—	—	-0.8
			σ_{rms}	0.8	0.8	0.3

Table A.5: Thymine bond lengths given by experimental results and geometry optimization using DFT and MP2, as described in section 5.2. All atoms are denoted by the IUPAC convention on page 8, the bond lengths are given in [Å], with the deviations and RMSDs in [mÅ].

	DFT	MP2	Exp.	Δ_{Exp}^{DFT}	Δ_{Exp}^{MP2}	Δ_{DFT}^{MP2}
N1:C6	1.380	1.385	1.378	-2.1	-7.1	5.1
C6:H6	1.085	1.092	—	—	—	7.6
C6:C5	1.352	1.366	1.339	-13.0	-26.7	13.7
C5:C7	1.500	1.504	1.496	-4.3	-7.6	3.3
C7:H7^a	1.095	1.101	—	—	—	6.2
C7:H7^b	1.095	1.101	—	—	—	6.3
C7:H7^c	1.093	1.100	—	—	—	7.0
C5:C4	1.469	1.468	1.445	-23.5	-22.8	-0.7
C7:O4	1.222	1.234	1.228	6.3	-5.8	12.1
C4:N3	1.407	1.408	1.382	-25.2	-26.4	1.2
N3:H3	1.013	1.018	—	—	—	5.5
N3:C2	1.390	1.389	1.377	-12.6	-12.5	-1.0
N3:O2	1.218	1.229	1.220	2.3	-9.2	11.5
N1:H1	1.009	1.014	—	—	—	4.7
			σ_{rms}	14.1	17.0	7.3

TABLE A.6: Thymine angles given by experimental results and geometry optimization using DFT and MP2, as described in section 5.2. All atoms are denoted by the IUPAC convention on page 8, angles and deviations are given in [^{o}].

	DFT	MP2	Exp.	Δ_{Exp}^{DFT}	Δ_{Exp}^{MP2}	Δ_{DFT}^{MP2}
N1:C6:H6	115.1	115.4	—	—	—	0.3
N1:C6:C5	122.7	122.4	123.7	1.0	1.3	-0.2
C6:C5:C7	124.1	123.7	122.9	-1.2	-0.8	-0.4
C5:C7:H7^a	110.8	110.4	—	—	—	-0.4
C5:C7:H7^b	110.8	110.4	—	—	—	-0.4
C5:C7:H7^c	111.3	110.8	—	—	—	-0.6
C6:C5:C4	118.2	118.1	118.0	-0.2	-0.1	0.0
C5:C7:O4	125.0	124.9	124.9	-0.2	-0.1	-0.1
C5:C4:N3	114.6	114.7	115.2	0.7	0.6	0.1
C4:N3:H3	116.2	116.3	—	—	—	0.2
C4:N3:C2	123.9	123.8	121.3	-2.6	-2.6	-0.1
N3:C2:O2	123.2	123.4	123.0	-0.2	-0.3	0.2
C2:N1:H1	121.1	121.0	—	—	—	-0.1
			σ_{rms}	1.2	1.2	0.3

TABLE A.7: Guanine bond lengths given by experimental results and geometry optimization using DFT and MP2, as described in section 5.2. All atoms are denoted by the IUPAC convention on page 8, the bond lengths are given in [Å], with the deviations and RMSDs in [mÅ].

	DFT	MP2	Exp.	Δ_{Exp}^{DFT}	Δ_{Exp}^{MP2}	Δ_{DFT}^{MP2}
N9:C8	1.386	1.383	1.375	-10.8	-8.4	-2.3
C8:H8	1.081	1.088	—	—	—	6.3
C8:N7	1.307	1.332	1.305	-1.9	-27.4	25.5
N7:C5	1.383	1.385	1.389	6.4	4.0	2.3
C5:C6	1.439	1.447	1.418	-21.5	-29.4	8.0
C6:O6	1.219	1.229	1.238	19.5	9.1	10.4
C6:N1	1.441	1.435	1.391	-49.7	-44.4	-5.3
N1:H1	1.013	1.018	—	—	—	5.5
N1:C2	1.373	1.379	1.373	0.1	-5.7	5.8
C2:N2	1.362	1.371	1.342	-20.0	-29.0	9.0
N2:H2^a	1.004	1.008	—	—	—	3.7
N2:H2^b	1.006	1.010	—	—	—	3.7
C2:N3	1.315	1.319	1.322	7.2	2.6	4.7
N3:C4	1.357	1.370	1.349	-8.2	-20.9	12.7
N9:H9	1.009	1.014	—	—	—	5.5
			σ_{rms}	20.0	22.5	9.3

Table A.8: Guanine angles given by experimental results and geometry optimization using DFT and MP2, as described in section 5.2. All atoms are denoted by the IUPAC convention on page 8, angles and deviations are given in [^{o}].

	DFT	MP2	Exp.	Δ_{Exp}^{DFT}	Δ_{Exp}^{MP2}	Δ_{DFT}^{MP2}
N9:C8:H8	121.5	121.9	—	—	—	
N9:C8:N7	112.8	112.8	113.1	0.2	0.3	-0.1
C8:N7:C5	104.5	104.0	104.4	-0.1	0.5	-0.6
N7:C5:C6	130.3	130.1	130.5	0.2	0.4	-0.2
C5:C6:O6	131.5	131.1	128.5	-3.0	-2.6	-0.5
C5:C6:N1	109.4	109.3	111.7	2.2	2.4	-0.1
C6:N1:H1	112.9	113.3	—	—	—	0.3
C6:N1:C2	126.5	127.0	125.0	-1.6	-2.0	0.4
N1:C2:N2	117.4	116.8	116.1	-1.3	-0.7	-0.6
C2:N2:H2^a	123.0	122.6	—	—	—	-0.4
C2:N2:H2^b	117.5	117.7	—	—	—	0.3
N1:C2:N3	123.5	124.0	124.0	0.5	0.0	0.5
C2:N3:C4	112.4	111.5	112.0	-0.4	0.5	-0.9
C4:N9:H9	125.5	125.5	—	—	—	0.0
			σ_{rms}	1.5	1.4	0.4

Table A.9: Cytosine bond lengths given by experimental results and geometry optimization using DFT and MP2, as described in section 5.2. Here the B3LYP/6-31+G(3df, p) calculation is denoted by DFT$^+$. All atoms are denoted by the IUPAC convention on page 8, the bond lengths are given in [Å], with the deviations and RMSDs in [mÅ].

	DFT	DFT$^+$	MP2	Exp.	Δ_{Exp}^{DFT}	$\Delta_{Exp}^{DFT^+}$	Δ_{Exp}^{MP2}	Δ_{MP2}^{DFT+}
N1:C6	1.350	1.353	1.362	1.367	16.9	14.3	5.0	9.3
C6:H6	1.085	1.085	1.092	—	—	—	—	7.5
C6:C5	1.354	1.358	1.371	1.339	-15.0	-19.3	-31.6	12.3
C5:H5	1.085	1.082	1.090	—	—	—	—	8.5
C5:C4	1.452	1.439	1.444	1.424	-28.3	-14.8	-20.0	5.2
C4:N4	1.353	1.358	1.365	1.336	-17.1	-21.9	-29.2	7.3
N4:H4a	1.008	1.008	1.012	—	—	—	—	4.1
N4:H4b	1.005	1.005	1.009	—	—	—	—	4.1
C4:N3	1.315	1.319	1.328	1.335	19.6	16.0	6.8	9.2
N3:C2	1.356	1.369	1.383	1.353	-2.6	-15.7	-30.0	14.3
C2:O2	1.221	1.218	1.232	1.240	19.4	21.8	8.3	13.5
N1:H1	1.009	1.010	1.015	—	—	—	—	5.4
				σ_{rms}	18.4	17.9	21.7	9.0

TABLE A.10: Cytosine angles given by experimental results and geometry optimization using DFT and MP2, as described in section 5.2. Here the B3LYP/6-31+G(3df, p) calculation is denoted by DFT$^+$. All atoms are denoted by the IUPAC convention on page 8, angles and deviations are given in [^{o}].

	DFT	DFT$^+$	MP2	Exp.	Δ_{Exp}^{DFT}	$\Delta_{Exp}^{DFT^+}$	Δ_{Exp}^{MP2}	$\Delta_{MP2}^{DFT^+}$
N1:C6:H6	117.7	116.9	117.1	—	—	—	—	0.2
N1:C6:C5	119.8	120.0	119.6	121.1	1.2	1.0	1.4	-0.4
C6:C5:H5	119.5	121.5	121.3	—	—	—	—	-0.2
C6:C5:C4	119.9	116.0	116.0	117.5	-2.5	1.4	1.5	-0.1
C5:C4:N4	119.7	119.1	118.9	120.2	0.5	1.1	1.2	-0.2
C4:N4:H4^a	118.3	118.3	118.2	—	—	—	—	-0.1
C4:N4:H4^b	121.9	122.0	121.8	—	—	—	—	-0.2
C5:C4:N3	120.6	124.0	124.4	122.0	1.3	-2.0	-2.5	0.4
C4:N3:C2	120.8	120.4	119.8	119.9	-0.9	-0.5	0.1	-0.7
N3:C2:O2	120.4	125.6	124.9	121.8	1.4	-3.8	-3.1	-0.7
C2:N1:H1	120.2	121.5	121.2	—	—	—	—	0.2
				σ_{rms}	1.4	2.0	2.0	0.4

Bibliography

[1] Johan Åberg. *Open Quantum Systems — Effects in interferometry, quantum computation and adiabatic evolution.* PhD thesis, Uppsala University, 2005.

[2] S. Arai, T. Chatake, T. Ohhara, K. Kurihara, I. Tanaka, N. Suzuki, Z. Fujimoto, H. Mizuno, and N. Niimura. Complicated water orientations in the minor groove of the B-DNA decamer d(CCATTAATGG)$_2$ observed by neutron diffraction measurements. *Nucleic Acids Research*, 33(9):3017–3024, May 2005.

[3] C.W. Bauschlicher and S.R. Langhoff. Core-core and core-valence correlation. *J. Chem. Phys.*, 88(4):2540–2546, Feb. 1988.

[4] M.L.M. Beckers and L.M.C. Buydens. Multivariate analysis of a data matrix containing A-DNA and B-DNA dinucleotide monophosphate steps: multidimensional ramachandran plots for nucleic acids. *J. Comput. Chem.*, 19:695–715, 1998.

[5] H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider. The Nucleic Acid Database: A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, 63(3):751–759, Sep. 1992.

[6] H.M. Berman, J. Westbrook, G. Gilliland Z. Feng, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The protein data bank. *Nucleic Acids Res.*, 28(1):352–242, Jan. 2000.

[7] S.F. Boys. Electronic wave functions I. A general method of calculation for the stationary states of any molecular system. *Proc. Roy. Soc.*, A200:542–554, 1950.

[8] L.S. Cederbaum. One-body Green's function for atoms and molecules: theory and application. *J. Phys. B: Atom. Molec. Phys.*, 8(2):290–303, 1975.

[9] L. Clowney, S.C. jain, A.R. Srinivasan, J. Westbrook, W.K. Olson, and H.M. Berman. Geometric parameters in nucleic acids: Nitrongenous bases. *J. Am. Chem. Soc.*, 118(3):509–518, 1996.

[10] R.E. Dickerson, D.S. Goodsell, and S. Neidle. " ...the tyranny of the lattice ...". *Proc. Natl. Acad. Sci. U. S. A.*, 91(9):3579–3583, Apr. 1994.

[11] S. Diekmann. Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.*, 205(4):787–791, Feb. 1989.

[12] H. Drew and R.E. Dickerson. Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *J. Mol. Biol.*, 149(4):761–786, Jul. 1981.

[13] H. Drew and R.E. Dickerson. Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.*, 151(3):535–556, Sep. 1981.

[14] H. Drew, T. Takano, S. Tanaka, K. Itakura, and R.E. Dickerson. Crystal structure analysis of a complete turn of B-DNA. *Nature*, 287(5784):755–758, Oct. 1980.

[15] H. Drew, T. Takano, S. Tanaka, K. Itakura, and R.E. Dickerson. High-salt d(CpGpCpG), a left handed Z' DNA double helix. *Nature*, 286(5773):567–573, Aug. 1980.

[16] H.R. Drew, S. Samson, and R.E. Dickerson. Structure of a B-DNA dodecamer at 16 K. *Proc. Natl. Acad. Sci. USA*, 79:4040–4044, Jul. 1982.

[17] H.R. Drew, R.M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R.E. Dickerson. Structure of a B-DNA dodecamer: Conformation and dynamics. *Proc. Natl. Acad. Sci. USA*, 78(4):2179–2183, Apr. 1981.

[18] H.R. Drew, R.M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R.E. Dickerson. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. USA*, 78(4):2179–2183, Apr. 1981.

[19] K.M. Elsawy, M.K. Hodgson, and L.S.D. Caves. The physical determinants of the DNA conformational landscape: an analysis of the potential energy surface of single-strand dinucleotides in the conformational space of duplex DNA. *Nucleic Acids Research*, 33(18):5749–5762, Sep. 2005.

[20] *No authors listed.* IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Abbreviations and symbols for the description of conformations of polynucleotide chains. Recommendations 1982. *Eur. J. Biochem.*, 131(1):9–15, Mar. 1983.

[21] P. Flükiger, H.P. Lüthi, S. Portmann, and J. Weber. Molekel 4.3, 2000-2002. Swiss center for Scientific computing, Manno (Switzerland).

[22] R.E. Franklin and R.G. Gosling. Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature*, 172(4369):156–157, Jun. 1953.

[23] R.E. Franklin and R.G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171(4356):740–741, May 1953.

[24] A.V. Fratini, M.L. Kopka, H.R. Drew, and R.E. Dickerson. Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBr-CGCG. *J. Biol. Chem.*, 257(24):14686–14707, Dec. 1982.

[25] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, revision c.02. Gaussian, Inc., Wallingford CT.

[26] A.A. Gorin, V.B. Zhurkin, and W.K. Olson. B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.*, 247(1):34–48, Mar. 1995.

[27] L. Hackermuller, K. Hornberger, B. Brezger, A. Zeilinger, and M. Arndt. Decoherence of matter waves by thermal emission of radiation. *Nature*, 427(6976):711–714, Feb. 2004.

[28] M.A. El Hassan and C.R. Calladine. The assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme. *J. Mol. Biol.*, 251(5):648–664, Sep. 1995.

[29] C.A. Hunter. Sequence-dependent DNA structure. *J. Mol. Biol.*, 230(3):1025–1054, Apr. 1992.

[30] C.A. Hunter and X.J. Lu. DNA base-stacking interactions: a comparison of theoretical calculations with oligonucleotide X-ray crystal structures. *J. Mol. Biol.*, 265(5):603–619, Feb. 1997.

[31] E. Johansson, G. Parkinson, and S. Neidle. A new crystal form for the dodecamer C-G-C-G-A-A-T-T-C-G-C-G: symmetry effects on sequence-dependent DNA structure. *J. Mol. Biol.*, 300(3):551–561, Jul. 2000.

[32] Roy L. Johnston. *Atomic and Molecular Clusters*. Taylor & Francis, 2002. ISBN 0748409319.

[33] Z.S. Juo, T.K. Chiu, P.M. Leiberman, I. Baikalov, A.J. Berk, and R.E. Dickerson. How proteins recognize the TATA box. *J. Mol. Biol.*, 261(2):239–254, Aug. 1996.

[34] P. Jurecka, J. Sponer, J. Cerny, and P. Hobza. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, dna base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.*, 8(17):1985–1993, May 2006.

[35] 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(3):668–680, Oct. 1997.

[36] X.J. Lu and W.K. Olson. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, 31(17):5108–5121, Sep. 2003.

[37] S. Lunell and G. Sperber. Study of the hydrogen bonding in the adenine-thymine, adenine-cytosine, and guanine-thymine base pairs. *J. Chem. Phys.*, 46(6):2119–2124, Mar. 1967.

[38] P. Mignon, S. Loverix, J. Steyaert, and P. Geerlings. Influence of the π-π interaction on the hydrogen bonding capacity of stacked DNA/RNA bases. *Nucleic Acids Res.*, 33(6):1779–1789, Mar. 2005.

[39] C. Møller and M.S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46(7):618–622, 1934.

[40] G.E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8):114–117, Apr. 1965.

[41] 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.*, 1(313):229–237, Oct. 2001.

[42] M.J. Packer, M.P. Dauncey, and C.A. Hunter. Sequence-dependent DNA structure: dinucleotide conformational maps. *J. Mol. Biol.*, 295(1):71–83, Jan. 2000.

[43] M.J. Packer, M.P. Dauncey, and C.A. Hunter. Sequence-dependent DNA structure: tetranucleotide conformational maps. *J. Mol. Biol.*, 295(1):85–103, Jan. 2000.

[44] M.J. Packer and C.A. Hunter. Sequence-dependent DNA structure: the role of the sugar-phosphate backbone. *J. Mol. Biol.*, 280(3):407–420, Jul. 1998.

[45] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.*, 25(13):1605–1612, 2004.

[46] S. Portmann and H.P. Lüthi. Molekel: An interactive molecular graphics tool. *Chimia*, 54:766–770, 2000.

[47] C.C.J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23(2):69–89, Apr. 1951.

[48] R.J. Le Roy, M.L. Klein, and I.J. McGee. On the dissociation energy and interaction potential of ground-state Ne_2. *Mol. Phys.*, 28(2):587–591, 1974.

[49] J.J. Sakurai. *Modern Quantum Mechanics*. Addison-Wesley, 1994. ISBN 0201539292.

[50] B. Schneider, S. Neidle, and H.M. Berman. Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers*, 42(1):113 – 124, Dec. 1998.

[51] M. Shiotani, N. Isamoto, M. Hayashi, T. Fängström, and S. Lunell. Deuterium isotope effects on rotation of methyl hydrogens. a study of the dimethyl ether radical cation by ESR spectroscopy and ab initio and density functional theory. *J. Am. Chem. Soc.*, 122:12281–12288, 2000.

[52] X. Shui, G.G. Hu, L. McFail-Isom, and L.D. Williams. The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry*, 37(23):8341–55, Jun. 1998.

[53] X. Shui, C.C. Sines, L. McFail-Isom, D. VanDerveer, and L.D. Williams. Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry*, 37(48):16877–16887, Dec. 1998.

[54] G.E. Sims and S.H. Kim. Global mapping of nucleic acid conformational space: dinucleoside monophosphate conformations and transition pathways among conformational classes. *Nucleic Acids Research*, 31(19):5607–5616, Aug. 2003.

[55] A.L. Sobolewski and W. Domcke. The chemical physics of the photostability of life. *Europhys. news*, 37(4):20–23, 2006.

[56] M. Soler-Lopez, L. Malinina, V. Tereshko, V. Zarytova, and J.A. Subirana. Interaction of zinc ions with d(CGCAATTGCG) in a 2.9 å resolution X-ray structure. *J. Biol. Inorg. Chem.*, 7(4):533–538, Apr. 2002.

[57] J. Sponer, P. Jurecka, I. Marchan, F.J. Luque, M. Orozco, and P. Hobza. Nature of base stacking: reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chemistry.*, 12(10):2854–2865, Mar. 2006.

[58] J. Sponer, J. Leszczynski, and P. Hobza. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers.*, 61(1):3–31, 2002.

[59] Y. Tanaka and K. Yoshino. Adsorption spectra of Ne_2 and HeNe molecules in the vacuum-uv region. *J. Chem. Phys*, 57(7):2964–2976, Oct. 1972.

[60] A.B. Trofimov, J. Schirmer, V.B. Kobychev, A.W. Potts, D.M.P. Holland, and L. Karlsson. Photoelectron spectra of the nucleobases cytosine, thymine and adenine. *J. Phys. B: At. Mol. Opt. Phys*, 39:305–329, 2006.

[61] A.H. Wang, G.J. Quigley, F.J. Kolpak, J.L. Crawford, J.H. van Boom, G. van der Marel, and A. Rich. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, 282(5740):680–686, Dec. 1979.

[62] J.D. Watson and F.H.C. Crick. Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171(4361):964–967, May 1953.

[63] J.D. Watson and F.H.C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, May 1953.

[64] B. Wittig, T. Dorbic, and A. Rich. Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei. *Proc. Natl. Acad. Sci. USA*, 88(6):2259–2263, Mar. 1991.

[65] V.B. Zhurkin, Y.P. Lysov, and V.I. Ivanov. Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic Acids Res.*, 6(3):1081–1096, Mar. 1979.

Colophon

This document was typset using Leslie Lamport's LaTeX 2_ε, together with the memoir class developed by Peter Wilson. The chemical formulae in chapter 2 were typset using the XeTeX extension package by Shinsaku Fujita. All renderings and figures were produced with `Molekel` [46, 21], `3DNA` [36], `Chimera` [45], `Matlab` or `Xfig` and are the intellectual property of the author.