QUANTUM CHEMICAL STUDIES OF METAL-DNA INTERACTIONS

Konstantinos Gkionis

A thesis submitted to
Cardiff University
in accordance with the requirements for the degree of

Doctor of Philosophy

School of Chemistry
Cardiff University

September 2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed .....(Candidate) Date ....18 Sep 2010

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed .....(Candidate) Date ....18 Sep 2010

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed .....(Candidate) Date ....18 Sep 2010

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .....(Candidate) Date ....18 Sep 2010

ACKNOWLEDGEMENTS

Firstly, I would like to thank my family for their continuous support throughout all years of my studies that made this work possible. Also, Dr George Sfyris for our endless discussions over the years, the rest of my long-term devoted friends who have believed in me and supported me in their own way and Dr Chris Kefalidis and Dr Anna Lourantou for our interesting discussions and for sharing their thoughts on their projects. Also, I would like to specially thank Professor C.A. Tsipis, by whom I had the great luck to be introduced to Quantum Chemistry and who inspired me to pursue academic research in this field.

At Cardiff University, I would like to thank all teaching staff of the Theoretical Chemistry Group for their constructive and insightful remarks, Dr Ian Merrick for his valuable assistance with technical issues and past and present members of lab 1.95 for creating a pleasant working environment. In particular, Dr James Landon for having provided assistance with writing scripts and especially Dr Grant Hill for his assistance with technical issues and for being willing to discuss and advice on aspects of this work and science in general. Also, I am grateful to the Engineering and Physical Sciences Research Council for funding of this work and to the UK National Service for Computational Chemistry for use of its facilities.

Finally and most of all, I would like to thank my supervisor, Dr James A. Platts for giving me the opportunity to work on this project, for his outstanding support and guidance and for always being interested and willing to discuss my questions and provide me with new ideas and knowledge.

ABSTRACT

A series of density functional theory (DFT) and quantum mechanics/molecular mechanics (QM/MM) calculations are used to investigate the binding of platinum and ruthenium anticancer drugs to DNA. The qualitative and quantitative features of Becke's half-and-half (BHandH) functional for calculating geometries, binding energies and harmonic frequencies of non-covalently bound systems are tested and the intermolecular interactions are characterised and quantified using the QTAIM electron densities. Application of this DFT-QTAIM approach to complexes of the type $[(\eta^6\text{-arene})Ru(en)(nucleobase)]^{2+}$ shows a clear preference for binding at guanine over any other base both in gas phase and in aqueous solution, a trend explained on the basis of QTAIM and molecular orbital data. Key parameters of the QM/MM methodology within the ONIOM scheme and efficient geometry optimisation strategies are examined for applications involving DNA oligonucleotides. Calculations on cis-$[Pt(NH_3)_2]^{2+}$ (cisplatin) bound to d(CpCpTpGpGpTpCpC).(GpGpApCpCpApGpG) reveal that proper consideration of the electrostatics between the QM an MM regions can lead to acceptable geometries, especially when explicit solvent molecules are present. This approach is used to explore the effects of methyl substitution on the binding of a series of $[Pt(en)]^{2+}$ (en: ethylenediamine) complexes to dinucleotides. Among the examined methyl derivatives, significant differences are observed for the variants whose en nitrogen atoms are multiply methylated. Binding energies are found to be in excellent correlation with *in vitro* cytotoxicity data expressed as $-\log(IC_{50})$. The above mentioned cisplatin-oligonucleotide complex is compared against three clinically approved platinum drugs (carboplatin, heptaplatin and lobaplatin). Calculations on truncated models show a stronger binding for cisplatin among the four complexes and numerous intermolecular interactions are located via QTAIM analysis in the lobaplatin and heptaplatin complexes. Additionally, subtle differences in key geometrical parameters are observed among the complexes around the sites of platination, with the exception of unusually short interplanar base - base distances in the complexes of loba- and heptaplatin. Finally, the same QM/MM methodology is applied to oligonucleotide sequences of five base pairs that contain difluorotoluene or mismatched base pairs, which are shown to be too flexible to be optimised at reliable structures at the chosen level of truncation. Comparisons among obtained structures using different input parameters further validate the followed QM/MM approach.

Contents

1. Introduction

This thesis presents a summary of efforts to use quantum mechanics and hybrid quantum mechanics/molecular mechanics (QM/MM) methods to model the interaction of cisplatin and related transition metal complexes with DNA. The present literature review therefore encompasses the chemistry and biochemistry of cisplatin and previous attempts to use QM and QM/MM methods to model key aspects of this. Also included is an overview of the structure and properties of DNA, believed to be the primary biological target of such drugs, and the nature of its interactions with potential or approved drugs, thus necessitating discussion of the suitability and limitations of current methodologies in describing such interactions.

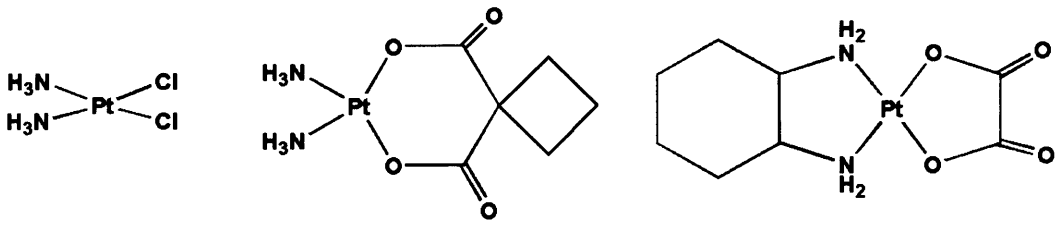
1.1 Introduction to cisplatin chemistry & biochemistry

1.1.1 Chemistry and biochemistry of cisplatin and related molecules

Cisplatin (*cis*-[Pt Cl_2 $(NH_3)_2$]) is one of the best-selling anti-cancer drugs of the last four decades. First synthesised in the 19th century,[1] interest was sparked in the 1960's following Rosenberg's serendipitous discovery of cytotoxicity,[2,3] ultimately leading to approval by the US Food & Drug Administration in 1978. Marketed as Platinol, this deceptively simple complex is now widely used as an effective first line treatment for many cancers.

Despite the success of cisplatin as an anticancer drug (it is active against testicular cancer, ovarian cancer, cervical cancer, colorectal cancer and relapsed lymphoma), its toxicity in tumour cells is coupled with several drawbacks[4] that have stimulated interest in the development of improved platinum drugs, and consequently in understanding the molecular mechanism that explains the biological activity of platinum compounds.[5] This lead to the discovery of carboplatin (azanide cyclobutane-1,1-dicarboxylic acid platinum) and oxaliplatin (cyclohexane-1,2-diamine oxalate platinum), which are both widely in use for the treatment of cancers, as well

as other complexes (*e.g.* nedaplatin, lobaplatin, heptaplatin) that have been approved for use in some parts of the world but not globally.[6]



cis-DDP, Cisplatin Carboplatin Oxaliplatin

Figure 1.1.1: Structures of the approved platinum drugs cisplatin, carboplatin and oxaliplatin.

Carboplatin is similar to cisplatin in its proposed mechanism of action, and is better tolerated by the body but less efficient than cisplatin due to a chelate effect. Carboplatin is used in the treatment of ovarian cancer, cervical cancer, head and neck cancer, non-small cell lung cancer and relapsed lymphoma. Oxaliplatin is a third generation platinum anticancer drug with a diaminocyclohexane (DACH) entity, predominantly used in the treatment of colorectal cancer. All of these complexes show antitumour activity due to the formation of cytotoxic lesions on DNA with platinum adducts, preventing replication and eventually causing cell death.

The method by which platinum complexes enter cells is a matter of debate, with both passive diffusion across membranes and active transport by copper transporter proteins suggested. However, once into the cytoplasm, cisplatin becomes hydrated with two chloride ligands being replaced by two water ligands to form a positively charged species. Products of its hydrolysis have the ability to interact with nucleophilic molecules within the cell; such molecules include DNA, RNA and proteins. The currently accepted thesis is that cisplatin induces its cytotoxic properties through binding to nuclear DNA. Platinum drugs favour binding to N7 atoms of the imidazole rings of guanosine(G) and adenosine(A) bases of DNA, potentially resulting in monoadducts, interstrand and intrastrand crosslinks, the latter being the most likely binding mode. Among all the possible intrastrand crosslinks, there are more 1,2-d(GpG) crosslinks than

any other, resulting in a significant DNA distortion. A view of intrastrand 1,2-d(GpG) cisplatin-DNA complex is shown in Figure 1.1.2.

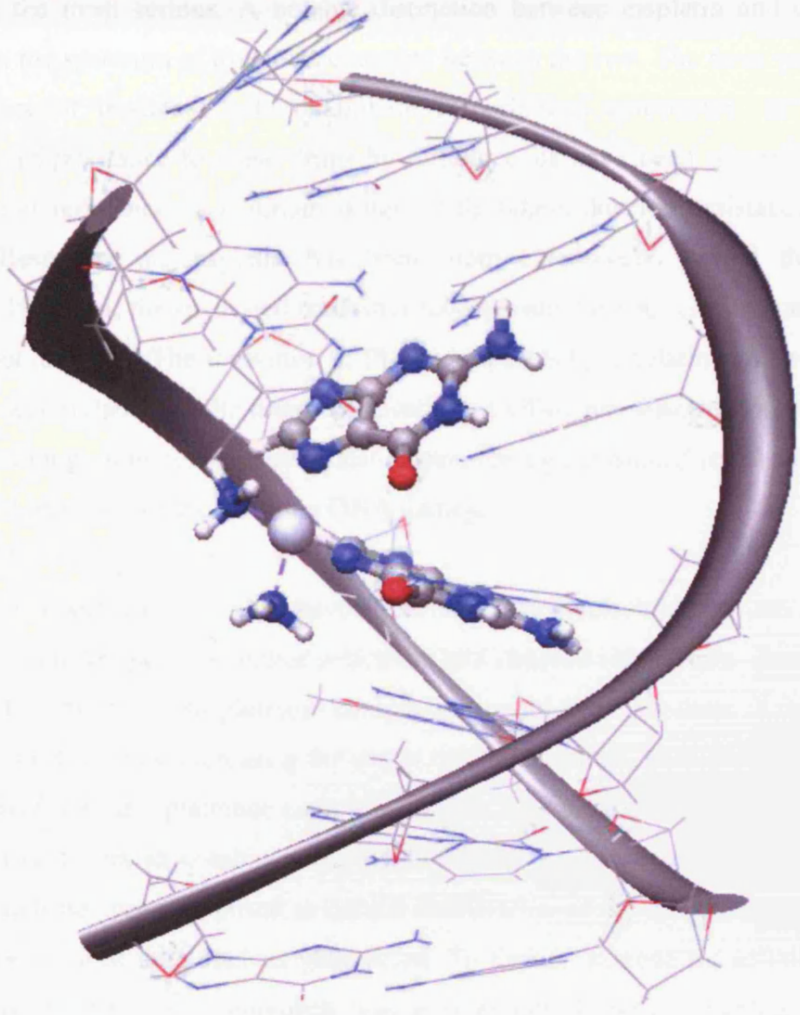


Figure 1.1.2: Intrastrand 1,2-d(GpG)-cisplatin complex.

Oxaliplatin forms fewer crosslinks than cisplatin at equimolar concentrations as its adducts are bulkier and more hydrophobic than those formed from cisplatin or carboplatin, leading to different effects in the cell. The general consensus is that 1,2-intrastrand crosslinks are responsible for the observed cytotoxicity, as comparative studies with inactive *trans*-diamine-dichloroplatinum (transplatin) show that this is unable to form 1,2-intrastrand crosslinks, but does form 1,3-intra and interstrand crosslinks.

The major limitations associated with use of these complexes are the toxicity of and resistance to the drugs. Toxicities range from mild to severe nephrotoxicity to peripheral neurotoxicity, the latter being the most serious. A notable distinction between cisplatin and carboplatin is the difference in the spectrum of toxicities observed between the two. The most common toxicity as a consequence of treatment with oxaliplatin is peripheral neuropathy. In addition, several mechanisms of resistance to these drugs by tumour cells have been observed. Some tumours have a natural resistance to platinum drugs while others develop resistance after the initial treatment. Resistance to cisplatin has been more extensively studied than resistance to carboplatin. However, the suggested resistance mechanisms for both cisplatin and carboplatin are similar, if not identical. The formation of Pt-DNA adducts by cisplatin can be limited (resisted) by reduced accumulation of the drug, enhanced drug efflux and inactivation by coordination to sulphur containing proteins. Resistance also occurs through enhanced repair of Pt-DNA adducts and increased tolerance of the resulting DNA damage.

These resistance and toxicity issues have led to intensive research efforts into finding new drug candidates with better potency and/or selectivity and reduced side effects. Promising avenues of research include multinuclear platinum complexes, in which two or three Pt centres are coupled together *via* linker groups, increasing the extent of DNA damage. Biologically active ligands can be incorporated into the platinum complexes, either to increase the concentration of platinum into the cell or to act in combination with DNA lesions. Increased bulk and rigidity of the nitrogen ligands has been proposed to reduce deactivation of the drug by slowing reaction with sulfur containing molecules such as glutathione. To further improve the affinity of cisplatin for nucleic acids, an alternative approach was also proposed, which involves the tethering of different DNA-binding ligands, such as oligonucleotides, intercalators and DNA-groove binders, to a cisplatin-derivative.[7] *Trans*-platinum complexes have also come back into vogue recently, with NH_3 ligands replaced by heterocyclic ligands such as pyridine, imidazole, thiazole *etc.*[8,9]

Metals other than platinum have also been a focus for discovery of novel drugs: promising candidates include titanium, ruthenium, rhenium and gold. Most are believed to operate in a similar fashion to cisplatin, with cationic and nucleophilic metal centres forming adducts with DNA, although specific details may differ.

Quantum chemical methods can play a significant role in gaining a deeper insight into the mechanism of action of these drugs, and hence suggest novel avenues for research into new and improved treatments. Specific areas of interest include the geometrical and electronic structure of both the drugs themselves and their DNA adducts, the mechanisms and potential energy surfaces of activation and subsequent reaction with DNA, and means by which cells distribute through the body and enter cells.

This chapter will discuss some selected work in these areas, including some by our own group as well as the many groups working worldwide on these topics. We focus on studies based on *ab initio* and/or density functional theory (DFT) methods, and leave to one side the large body of literature using molecular mechanics methods.

1.1.2 Calculation of cisplatin structure, activation & DNA interactions

One of the first efforts to apply quantum chemistry to gain a better understanding of cisplatin chemistry was reported in 1985 by Basch *et al.*,[10] who used Hartree-Fock (HF) methods together with a double-ζ Gaussian basis set and pseudopotential on Pt to compare *cis*- and *trans*- isomers of $PtCl_2(NH_3)_2$ and hydrated complexes. In this way, they found that the *trans*- isomer is *ca.* 19 kcal mol^{-1} more stable than the *cis*- one, a difference attributed to lesser repulsion between chlorides. Geometry optimisation gave rather longer Pt—N bonds in cisplatin compared to transplatin, interpreted as evidence for a substantial trans effect in the former, and also geometrical evidence for N—H···Cl hydrogen bonding in the *cis*-isomer.

Recognising the need for electron correlation to obtain accurate results, Carloni *et al.* used gradient corrected exchange correlation DFT functionals and periodic plane-wave basis sets to examine the isomers.[11] This reproduced Pt—N and Pt—Cl bond lengths to within the estimated error in experimental data (obtained from X-ray diffraction) and confirmed the stability of the *trans*- over the *cis*- isomer, albeit by 8 kcal mol^{-1} at this level. These authors also found evidence for N—H···Cl hydrogen bonds, not least in non-zero values of Mayer bond orders between donor and acceptor atoms. Vibrational spectra were also calculated for cisplatin and transplatin in the

range 100 to 3500 cm^{-1}, resulting in very good agreement with the available experimental data, especially in the Pt—N and Pt—Cl stretching frequencies, as well as some fine details in the N—H region that are highly sensitive to the symmetry and environment. The bonding in these complexes was examined *via* difference densities, obtained by subtraction of fragment electron density from that of the entire complex.

Following the work of Carloni *et al.*, Pavankumar *et al.* carried out a systematic study of the structure, bonding, charge density and vibrational frequencies of cisplatin, with particular focus on the dependence of these properties with respect to theoretical methods and basis sets.[12] Electron correlation was included using Møller-Plesset (MP) perturbation theory. 14 different basis sets on Cl, N, and H were employed, and combined with two pseudopotential schemes, namely those of Stevens, Basch, and Krauss (SBK) and of Hay and Wadt (LanL2DZ). Finally, three different possible conformations were explored, *i.e.*, two with C_{2V} and one with C_S symmetry, all three confirmed as minima by harmonic frequency calculation. Agreement with experimental geometry improved with use of larger basis sets, and with higher orders of the MP series, and was also noticeably better with SBK rather than LanL2DZ pseudopotentials. Vibrational frequency analysis showed considerable variation in calculated frequencies with method and basis set, and the authors proposed MP2/6-311++G(2d,2p) as the best choice, giving low overall errors across the range of values observed. Bonding within cisplatin was analysed using the calculated molecular orbitals, electrostatic potential, and electron density analysis.

1.1.3 Modelling of cisplatin activation

While the structure of cisplatin itself is of undoubted interest, the fact that it must be activated by hydration has led to significant interest in the interactions and reactions of cisplatin with water. In this context, 'hydration' can have two quite different meanings: one refers to the solvation of an intact cisplatin molecule, with water in the second solvation shell. The other refers to the reaction of cisplatin with one or more water molecules, leading to substitution of one or both chloride ligands. Both are important in the chemistry and biochemistry of cisplatin, and have been examined using theoretical methods.

Kozelka and co-workers investigated potential energy surfaces (PES) for interaction of a single water molecule with cisplatin in various orientations.[13] From this, favoured cisplatin-water conformations and distance dependence were observed. Significant differences between these parameters from HF, MP2 and DFT levels of theory indicated the importance of dispersion energy in determining the details of interaction with water. Intriguingly, these studies demonstrated the existence of 'inverse' hydration,[14] in which Pt···H—O interactions dominate, as well as the more expected Pt···O orientations. Similar PES curves were recently established for a larger number of possible orientations of cisplatin-water complexes by Lopes *et al.*,[15] who also extracted Lennard-Jones parameters for faster/larger-scale simulation of cisplatin hydration from these results. Robertazzi *et al.*[16] used DFT methods to examine the optimum geometry of 1:1 cisplatin-water complexes, reporting just three stable minima and characterising the interactions present in each on the basis of electron density. Coordination to platinum induces substantial polarisation of chloride and ammine ligands, causing the N—H groups to act as strong hydrogen bond donors and Cl as strong hydrogen bond acceptors. Extending this treatment gave an estimate of the first solvation shell of cisplatin is a 10:1 complex that contained N—H···O, O—H···Cl and O—H···O hydrogen bonds. In addition, the analysis of the electron density revealed that the hydrogen bonds between cisplatin and water become slightly weaker when explicit solvation is included, whereas water···water interactions are strongly enhanced by their proximity to cisplatin.

Characterisation of the hydrolysis of cisplatin has been of much interest, including the barriers to successive replacement of chlorides, the mechanisms of this process and the overall energy changes accompanying such reactions. Zhang *et al.*[17] reported DFT studies of the hydrolysis reaction using a range of popular exchange-correlation functionals, pseudopotentials on Pt, and a self-consistent reaction field (SCRF) estimation of aqueous solvation. They found the reaction to proceed *via* a 5-coordinate, trigonal bipyramidal transition state, described as belonging to the S_N2 class of substitution reaction. This mechanism gives rise to a barrier of *ca.* 23 kcal mol^{-1}, slightly in excess of an experimental value of 20 kcal mol^{-1}. Comparison of gas-phase with solvated data indicated that solvation has a major effect on calculated barriers, while barely changing reaction thermodynamics. Since this work, similar studies have tested variations on level of theory, basis set and treatment of solvation, but have in general found little difference in

calculated barriers to reaction and predicted rate constants for hydrolysis. For instance, Robertazzi followed the progress of hydrolysis instigated by one of the water molecules in their 10:1 water-cisplatin complex, *i.e.* including explicit solvation rather than implicit SCRF effects, while evident differences were found in the mechanism of chloride substitution, the energetics of the reaction was hardly affected by the inclusion of explicit solvation.

Deubel and co-workers investigated all the hydrolysis steps of cisplatin employing state-of-the-art DFT calculations.[18] The computational scheme was carefully chosen in order to tackle problems arising from the estimation of solvation free energies and entropic effects. Such an approach was corroborated by the close agreement of experimental and calculated pK_a values of cisplatin and its hydrolysis products. Interestingly, calculations predicted very similar activation barriers for the three hydrolysis steps, ranging between 25 and 27 kcal mol^{-1}, with reaction free energies of 0-2 kcal mol^{-1}. In addition, a critical comparison of these results with previous studies,[19,20] indicated that the evident disagreement of the second and third hydrolysis barriers, previously predicted to be strongly endergonic, may be due to factors such as the choice of the reference state to calculate the reaction energy and barrier as well as the estimation of free solvation energy and entropic effects. Notably, these calculations supported, for the first time, the experimental hypothesis of the diaqua form of cisplatin being the actually active species.

1.1.4 Modelling of cisplatin - DNA interactions

Once activated *via* hydrolysis, the interaction of solvated cisplatin with DNA is the generally accepted mechanism for the observed cytotoxicity of the drug. The details of this interaction have therefore been of great interest to theoretical researchers. Amongst the first such studies was by Basch *et al.*, using HF methods to probe binding of $Pt(NH_3)^{2+}$ to the individual bases guanine, adenine, cytosine, and thymine.[21] This showed preferential binding to the N7 position of guanine (see Figure 1.1.3 for numbering), with reasonably strong binding to N and O sites across most other bases. The order of strength of binding was proposed to be G(N7) > C(N3) > C(O2) > G(O6) > A(N3), A(N1) > A(N7) > G(N3) > T(O4) > T(O2), with approximately 19 kcal mol^{-1} difference between G(N7) and the next most stable site. Bifunctional, or chelating, binding

to both N7 and O6 of a single guanine was not found to be a stable orientation, supporting the suggestion that 1,2 and 1,3- intrastrand binding might be the major source of DNA damage by cisplatin. The effects of binding to N7 and O6 sites was monitored *via* Mulliken population analysis, indicating substantial polarisation of the electron density of guanine by the Pt fragment.

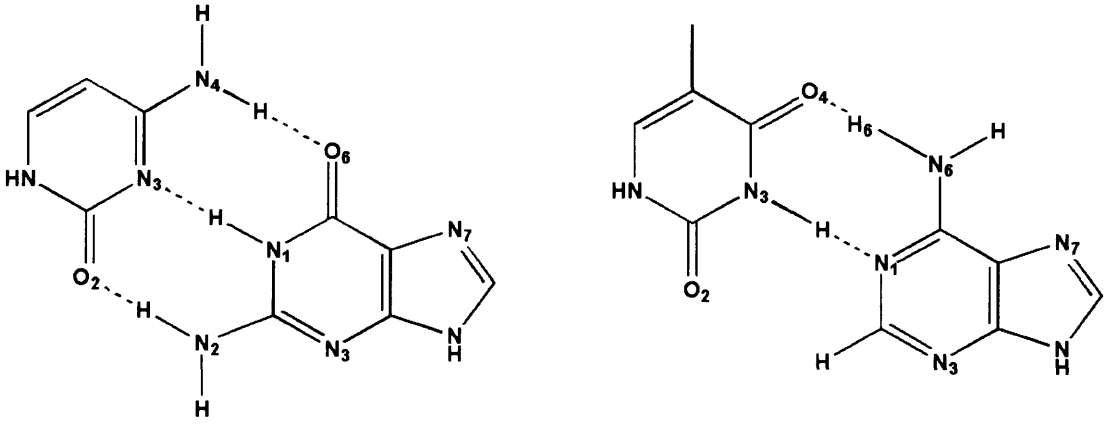


Figure 1.1.3: Numbering of Watson-Crick paired cytosine-guanine (CG, left) and thymine-adenine (TA, right) pairs.

Baik *et al.* used DFT methods, pseudopotentials and SCRF solvation to address the key question of why cisplatin binds preferentially to guanine over adenine.[22] The N7 sites of these bases are electronically similar, and are also exposed in the major groove of standard B-DNA. This detailed study confirmed previous findings that cisplatin shows a thermodynamic preference for binding at guanine over adenine, with gas phase SCF values differing by around 15 kcal mol^{-1}. Hydrogen bonds from coordinated N—H groups of cisplatin to O6 of guanine and N6 of adenine were clearly observed, the latter involving significant pyramidalisation of the NH_2 group. The strengths of these hydrogen bonds were estimated at *ca.* 7 and 5 kcal mol^{-1}, respectively, and are hence insufficient to account for the difference in stability. Solvation reduces binding energies substantially and also closes the difference between guanine and adenine binding, due to the greater solvation energy of guanine. Zero point energy (ZPE) and entropy effects, calculated from harmonic vibrational frequency data, reduce the difference in binding further still, giving a $\Delta\Delta G(Sol)$ of *ca.* 7 kcal mol^{-1}. These authors also examined the kinetics of guanine and adenine

adduct formation from platinum chloroaqua complexes, consistently finding lower barriers for reaction with guanine, quoting values of 24.6 kcal mol^{-1} for guanine and 30.2 kcal mol^{-1} for adenine. These differences in barriers were further probed by an energy decomposition scheme and in terms of the frontier molecular orbitals of the nucleic acid bases.

Carloni *et al.* used Car-Parrinello molecular dynamics (CPMD) to monitor the reaction of activated cisplatin with DNA fragments,[23] particularly guanine-phosphate-guanine (*d*GpG), for which X-ray crystallographic data are available. Simulation of the $Pt(NH_3)_2^{2+}$ adduct of this in water was stable, and gave average Pt—N bond lengths of 2.05(7) and 2.03(5) Å (estimated standard deviation in parentheses) for ammonia and guanine ligands, and angles fluctuating by *ca.* 3° from the X-ray values. The angle formed between coordinated guanines was notably smaller than observed in the solid state, an effect ascribed to the effect of solvation. Hydrogen bonding patterns to solvent water and to the 5′-phosphate group were monitored across 2 ps of dynamical simulation.

Burda and Leszczynski used the popular B3LYP DFT method, supplemented by MP2 data where necessary, to monitor Pt-DNA interactions, and in particular the structure and energetics of bridged structures with two purine bases.[24] Hydrogen bonds and trans effects were observed to play a role in determining the relative energies of PtA_2, PtG_2 and PtAG adducts, surprisingly showing that the mixed complex PtAG is relatively stable, but that PtG_2 is more so. Molecular orbitals and natural bond orbitals (NBO) were used to rationalise these observations in terms of donation from and back-donation to ligands. Robertazzi *et al.* tackled a similar problem, but this time used topological analysis of the electron density as the principal tool for analysis.[25] Stationary (or critical) points in the electron density unambiguously identify bonds, and properties evaluated at these points can be related to the strength and other properties of the bonds. In this way, the ubiquity of hydrogen bonds in cisplatin adducts of DNA becomes clear: these consist not only of hydrogen bonds from the coordinated N—H ammonia ligands, but also between nucleic acid bases. Estimating the strength of these hydrogen bonds on the basis of electron density properties allowed the remaining covalent binding energy to be computed, quantifying the increased Pt affinity of G(N7) over other potential binding sites. The effect of platination on the standard Watson-Crick base pairing of guanine with cytosine was examined in

similar fashion: in general, the overall pairing energy of G with C is little affected by the presence of Pt, but the strengths of individual hydrogen bonds (and hence the relative orientation of the duplex DNA) are altered markedly. A clear increase in the Pt—N bond strength in Pt-G≡C over Pt-G was also noted.

Mantri *et al.* used DFT to address why cisplatin forms AG links but not GA ones, where the difference lies in the directionality of the nucleic acid chains in DNA, taking *d*(pApGpA) and smaller segments of this as models of DNA.[26] This work showed little thermodynamic difference between AG and GA adducts, whereas kinetic differences for formation of the two motifs are large. The transition state leading to AG formation is stabilised by hydrogen bonding to a backbone phosphate group, whereas the right-handed nature of the DNA helix prevents this interaction in the GA orientation. The geometrical constraints of binding a single Pt centre to two adjacent nucleic acid bases also induces large changes in the backbone dihedrals of the DNA chain, especially in the puckering of the sugar rings, allowing better relaxation and greater tilt angle between purine bases.

1.1.5 QM/MM models of cisplatin-DNA complexes

The studies discussed above use small segments of DNA as models of larger scale behaviour. While there are clearly many lessons that can be drawn from this approach, most experimental data on cisplatin-DNA binding are obtained for larger DNA oligomers, typically duplex DNA structures of between 6 and 16 bases in length. Full quantum mechanical simulation of such large structures is likely to be computationally prohibitive for many years to come so an alternative approach is required. Hybrid QM/MM methods, in which a small region of a larger structure is treated with an accurate quantum mechanical method while the remainder is described by faster molecular mechanics (MM) methods, shows much promise in this regard. However, only a few examples of this approach have been published to date.

Spiegel *et al.* coupled CPMD methods to both GROMOS and AMBER MM forcefields in order to model cisplatin binding to larger DNA oligomers such as d(CCTCTG*G*TCTCC)-

d(GGAGACCAGAGG), where * indicates the location of cisplatin binding.[27] This was prompted by observation of significant differences between X-ray and NMR data. Their QM/MM partitioning scheme included $[Pt(NH_3)_2]^{2+}$, two guanines and, in one case, a single phosphate group in the QM region, linked to the remaining MM region by hydrogen link atoms. Single point energies were used to check this partition by extending the QM region to comprise more bases, including the paired cytosines and flanking thymines in the same strand as the Pt-bound guanine. Simulation of this system over 5 to 7 ps allowed the structure to relax and geometrical features used to monitor DNA structures such as rise, roll, axis bend, buckle and propeller twist were checked. In this way, the simulated structure was found to be more similar to the NMR than to the X-ray structure, perhaps unsurprisingly since the former is obtained from solution phase. Hydrogen bonds between ammonia ligands and phosphate groups were lost over the course of the simulation, but those to O6 of guanine persisted throughout. "Docking" of $[Pt(NH_3)_2]^{2+}$ to the regular structure of B-DNA resulted in a highly distorted structure, but this relaxed to an essentially identical structure to that obtained from the experimental starting point after just a few picoseconds of simulation.

Robertazzi and Platts used the ONIOM procedure for QM/MM to couple DFT methods with AMBER, starting with relatively small models of a single strand of DNA.[28] In order to account for possible dispersion interactions between nucleobases, the BHandH functional was employed in combination with medium/large-sized basis sets. The partitioning employed here included $[Pt(NH_3)_2]^{2+}$ and two - four nucleic acid bases in the QM region, with their associated sugar-phosphate backbones in the MM region. Compared with isolated bases, the effect of the backbone was not found to alter the main trends in binding energies, such as the preference for G over A. However, subtle differences in energies and especially optimised geometries were evident, particularly when the Watson-Crick paired cytosine (for G) or thymine (for A) and their respective sugar-phosphate backbones were included. A similar partitioning was then employed to study cisplatin binding to a larger DNA oligomer, d(CCTG*G*TCC)·d(GGACCAGG), solvated by around 400 water molecules, whose structure is known from NMR experiment. This DFT/AMBER approach satisfactorily reproduced the experimental structure, and again revealed subtle differences in structure and binding than would be obtained from a purely QM approach.

Matsui *et al.* examined whether coordination of cisplatin and transplatin to DNA can induce proton transfer reactions between the nucleobases, using two- and four-base pair models.[29] In this study the ONIOM method was employed for geometry optimisation of the four base pair model, using the mPW1PW91 DFT functional and the Universal Force Field (UFF), while the proton transfer reactions were modelled fully at the DFT level using the SDD ECP and basis set on the platinum atom and the 6-31G(d,p) basis set on the rest of the atoms. The inclusion of large-scale DNA structure through the ONIOM method did little to change the coordination geometry around Pt. Comparison with available experimental structural data indicated generally good agreement for the platinum coordination site, but slightly larger differences in the lengths of the many hydrogen bonds present within these complexes. Single proton transfer reactions were found to be feasible within the platinated GC pair, but multiple proton transfers were not.

1.2 Platinum-based alternatives

The long history in drug development, small size and well-defined coordination geometry of cisplatin make it an ideal candidate for study using quantum chemical methods. Accordingly, fewer publications on the structure and properties of non-cisplatin drugs such as carboplatin have appeared in the literature. Tornaghi performed one of the first comparisons of cisplatin with carboplatin using DFT,[30] reporting good general agreement with the experimental X-ray structure. Discrepancies of 3% in bond lengths and 4% in angles were found and explained as errors in DFT/basis set as well as environmental effects in the crystal. Calculation of harmonic frequencies allowed bands in the infra-red spectrum to be assigned, with distinct Pt—N bands but Pt—O modes that are strongly coupled to motions of the 6-membered ring. Molecular orbital energies showed significant differences to cisplatin, including a rather larger HOMO-LUMO gap due to coupling with ligand-based orbitals. In both cases the LUMO is almost solely the d_{x2-y2} orbital, but the HOMO differs, with strong mixing of Pt with O orbitals in cisplatin.

Wysokiński and Michalska compared the performance of several DFT methods in calculating the structure and vibrational frequencies of cisplatin and carboplatin.[31] Findings for cisplatin were not very different from those discussed above. Carboplatin was found to have essentially planar

coordination around Pt with the 6-membered ring in a boat conformation and a slightly puckered cyclobutane fragment. The exchange-correlation functional mPW1PW91 was found to give the best agreement with the X-ray crystallographic structure, although differences due to theoretical methods were small. Vibrational spectra were also well calculated by mPW1PW91, in terms of both the normal mode frequencies and the intensities of infra-red and/or Raman bands. Recently, Wysokiński *et al.* revisited their predictions of the Raman spectrum of carboplatin, including new measurements of this spectrum, confirming the accuracy of their chosen mPW1PW91 functional.[32] Highly detailed analysis of the vibrational modes within carboplatin was presented, allowing confident assignment of the entire experimental spectrum. NBO analysis was used to study the charge distribution and bonding in carboplatin, indicating a charge of +0.81 on Pt and –1.07 on N, and evidence for a trans-influence from N to O.

Pavelka *et al.* have monitored the hydrolysis reaction of carboplatin, which is of interest since this drug is designed to undergo slower *in vivo* activation than cisplatin due to the chelating nature of the leaving group.[33] Using B3LYP with the Stuttgart-Dresden pseudopotentials, along with SCRF aqueous solvation, they found a barrier of *ca.* 31 kcal mol^{-1} for initial disruption of the chelated carboplatin structure, the reaction proceeding *via* a 5-coordinate transition state similar to that now well-established for cisplatin. This first reaction is endothermic by around 15 kcal mol^{-1}. The second hydrolysis step, which leads to loss of the malonato ligand, has a lower barrier of *ca.* 22 kcal mol^{-1}, such that the first step appears to be rate limiting. The effects of pH were monitored by comparing hydroxide (HO^-) with aqua (H_2O) ligands, acidification being known experimentally to speed up this reaction. Indeed, with protonated ligands, the rate-limiting barrier was reduced to around 23 kcal mol^{-1}. It was proposed that changes to the ligand structure on protonation leads to formation of stronger hydrogen bonds, which then act to stabilise the relevant transition state.

Sarmah and Deka have used DFT and SCRF solvation to compare solvation and reactivity indices for cisplatin, carboplatin and oxaliplatin.[34] These indices include global and local hardness and softness, electrophilicity and frontier molecular orbital (Fukui) functions. Optimised geometries for cisplatin and carboplatin were similar to those discussed above, while that for oxaliplatin was similar to the X-ray crystallographic structure with a chair-like geometry of the cyclohexane ring and NH_2 groups in the equatorial position. The calculated reactivity

indices reproduced the experimental trend in reactivity of cisplatin, carboplatin, and oxaliplatin. These values were also used to construct a QSAR equation for the experimental cytotoxicity (IC_{50}) of seven platinum complexes, in which a single electrophilicity parameter explained 90% of the variance in the experimental data.

JM118, shown in figure 1.2.1, is a candidate platinum drug including a hydrophobic cyclohexane ring, designed to slow hydrolysis and improve cell uptake. Zhu *et al.* used standard B3LYP, LanL2DZ and SCRF methods to examine the hydrolysis of this complex.[35] Similar mechanisms to other Pt(II) complexes were found, with trigonal bipyramidal transition states. The details of the energy barrier were found to be sensitive to the model used, and inclusion of one or more explicit water molecules was found to be necessary as they act to solvate the anionic leaving group. This work confirmed that JM118 should undergo slower hydrolysis than cisplatin, hence allowing time for the complex to reach its cellular target before being deactivated, and also hopefully reducing some of the toxic side effects associated with cisplatin.

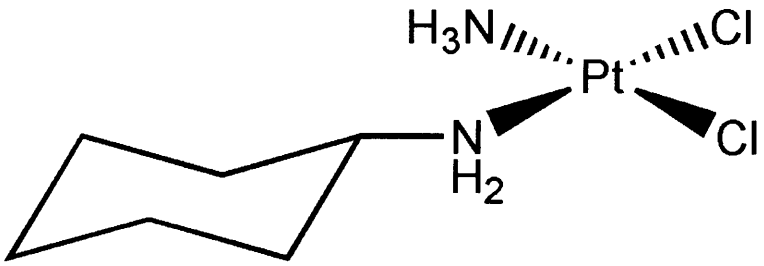


Figure 1.2.1: Structure of JM118 (*cis*-amminedichlorocyclohexylamineplatinum(II))

Other ligand architectures studied using theoretical methods include a complex with orotic acid (vitamin B_{13}), which is known to act as a biological carrier for metal ions such as Mg^{2+} and Ca^{2+}.[36] As with previous work on carboplatin, the Raman spectrum of this complex was used as the principle source of experimental data, and again mPW1PW91 performed well in reproducing this. NBO analysis was used to examine the bonding and potential hydrogen bonding within this complex. Dos Santos *et al.* studied complexes of $Pt(Cl)_2$ with tetracycline, a potent broad-spectrum anti-microbial compound.[37] The range of possible coordination sites of the platinum moiety were explored with HF, MP2 and B3LYP methods, leading to 14 separate coordinated

forms, in all of which the tetracycline acts as a chelating ligand, the most stable being coordinated through one N and one O centre, both in gas phase and SCRF water. Changes in predicted ^{13}C NMR chemical shift data from the values for free tetracycline were used to compare with experiment. Hydrolysis of the tetracycline-$Pt(Cl)_2$ complex was also examined, and found to be broadly similar to previous studies of other Pt(II) complexes.

Synthesis of complexes containing more than one Pt centre is an increasingly popular strategy in discovery of new drugs, it being proposed that they give 1,2-intrastrand adducts more selectively than single Pt complexes. Deubel has studied dinuclear complexes bridged by hydroxo groups, including their formation from single Pt complexes, activation by hydrolysis and reactions with guanine.[38] Barriers to reaction were found to be higher than for mononuclear compounds. No evidence for direct Pt—Pt bonding interactions was found on the basis of calculated electron density, instead a ring critical point was located between Pt nuclei. Magistrato *et al.* studied dinuclear Pt complexes bridged by azoles, including those linked by adjacent (1,2) N and non-adjacent (1,3) atoms in the azole using QM/MM methods to follow complexation to a decamer of duplex DNA.[39] Geometries of the complexes themselves were close (3% relative error) to the X-ray crystallographic values, with the typical square-planar coordination geometry expected of Pt(II). DNA complexes were close to NMR structures, with RMS deviation of *ca.* 0.6 Å; local structural parameters such as "rise", "roll", and "tilt" were similar to NMR, but major differences to the local structure around Pt from equivalent cisplatin structures were evident. It was suggested that the lack of major changes in the DNA duplex structure may help the Pt adduct escape recognition by repair enzymes, and hence lead to amelioration of the problem of resistance to treatment observed with cisplatin drug therapy.

The above results for the diplatinum azole-bidged complexes were well reproduced by Spiegel *et al.* using force field parameters derived from *ab initio* forces generated by QM/MM calculations following the "force matching" methodology.[40] The parameters were used to perform 10 ns MD simulations, in which the (1,2)-azole derivative – DNA adduct structural characteristics were well reproduced, while larger discrepancies were observed for the (1,3) derivative. This was attributed to the fact that the distortions induced by the (1,3) complex are larger than those of the (1,2). The MD simulations based on *ab initio* forces, apart from reproducing the QM/MM

results, have the advantage of implicitly incorporating solvent and temperature effects into the simulation.

QM/MM MD simulations for modelling drug DNA interactions were reviewed in 2006 by Spiegel and Magistrato.[41] In this work, emphasis was given to cisplatin- and dinuclear azole-bridged - DNA studies as well as on heterocyclic antitumour antibiotics which bind covalently to DNA. Despite the drawback of limited time scale of the methodology that is pointed out by the authors, the usefulness of QM/MM MD simulations in providing information on the induced structural distortions on the DNA helix upon drug complexation and on the underlying chemistry of the relevant processes is highlighted. As a result, it is concluded that this approach can contribute to the understanding of recognition processes and is a promising method for future studies of non-DNA drug targets.

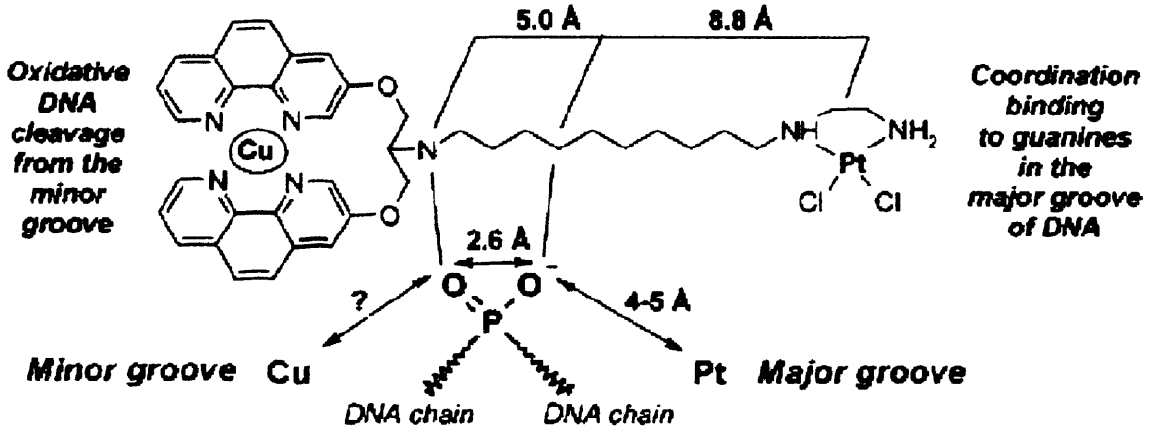


Figure 1.2.2: Schematic view of Cu(3-Clip-Phen) and its interaction with DNA (reproduced from ref. 42)

In an effort to improve upon cisplatin activity, Reedijk and co-workers recently synthesised a multifunctional drug (Figure 1.2.2) combining a cisplatin-derivative and a copper-based artificial DNA-cleaving agent, Cu(3-Clip-Phen).[42] This was found to bind to DNA from the major groove *via* the cisplatin subunit and the minor groove *via* the copper subunit. This combination resulted in the typical effects of cisplatin, *e.g.* DNA bending, together with double-strand cleavage promoted by the copper centre, leading to promising biological results.[43] Two theoretical studies

focus on the copper phenanthroline subunit: Robertazzi *et al.*[44] employed B3LYP and BLYP DFT functionals to characterise the structural and electronic properties of the parent compound copper-1,10-phenantrholine complex, as well as the more effective derivatives Cu(2-Clip-Phen) and Cu(3-Clip-Phen) (Phen = phenanthroline), with the two aromatic rings linked by a serinol bridge (Clip). The interaction with a DNA fragment was then explored by docking calculations, indicating pseudo-planarity of Cu(3-Clip-Phen) to be one of the key factors of activity. In addition, preliminary calculations[45] on the entire complex showed that the geometries of cisplatin and copper subunits hardly change when these are combined together. This suggested, in line with experiments, that DNA binding of ditopic cisplatin-copper complex is similar to that of the single components. Further theoretical studies are required to shed light on the structural and electronic properties of such an intriguing example of a potential multifunctional drug.

1.3 Non-platinum alternatives

Platinum is by no means the only metal that is able to form DNA adducts and hence show potential anti-cancer properties. The long history and precedent for Pt complexes, along with the highly regular stereochemistry associated with the common oxidation states, mean that structure-activity rules for Pt complexes are well-established, but in recent years complexes of other transition metals, most notably titanium, ruthenium, and rhenium, have become prominent in the literature. Robertazzi and Platts examined the entire *d*-block of transition metals for their interactions with guanine, and their effect on guanine-cytosine pairing, using DFT and QTAIM methods, keeping the oxidation state, overall charge, and ligands as similar as possible to those found in cisplatin.[46] Most metals show a clear preference for the N7 coordination site, as is found in Pt complexes, but the "early" transition metals such as titanium and vanadium are thermodynamically more stable when complexed to the O6 site, with a crossover region in group 6 (chromium). The manganese group were found to have the weakest binding to guanine, whereas the nickel group were found to have the strongest binding.

The effect of metal complexation on base pairing to cytosine was strikingly different depending on whether the N7 or O6 site is preferred, since O6 is also involved as a hydrogen bond acceptor

in this pairing. As with cisplatin, the overall energy of pairing is barely affected by N7 coordination, although individual hydrogen bonds are affected. In contrast, almost all O6 complexes exhibit much weaker pairing and distorted geometry of base pairs, due to the almost complete loss of N_4—H_4···O_6 hydrogen bond to cytosine. Molecular electrostatic potentials of O6 and N7 complexes and the differences from free guanine were used to rationalise the observed changes.

Ruthenium complexes show much promise as new, non-platinum drugs. One of the most popular is ImH[*trans*-Ru(III)Cl_4(DMSO)(Im)] (Im = imidazole), which has been termed NAMI-A for short. This complex, a strong antimetastatic agent, recently completed Phase I clinical trials. The mechanism of action of this potential drug remains unknown, but similarities in the activation of the complex by substitution of chlorides with water have been noted, this being required for biological activity. Recently, three studies have been reported on the hydrolysis of this complex. Chen *et al.*[47] used DFT with SCRF solvation to establish the pseudo-octahedral nature of transition states and a barrier of 23 kcal mol^{-1}, in reasonable agreement with experiment. Bešker presented a similar work[48] using larger basis sets and explicit water molecules, obtaining better agreement with experiment. Significantly, they showed that the first hydrolysis step is faster than the second one.

Vargiu *et al.*[49] studied the same process along with the analogous one in (ImH)[trans-$RuCl_4(Im)_2$] (ICR), with a view to understanding the difference in biological activity between these complexes, *e.g.*, ICR is active against primary tumours and NAMI-A is an antimetastatic agent. Unlike previous studies, this work considered both Ru(II) and Ru(III) states, all the possible hydrolysis routes (Dimethylsulfoxide and Imidazole hydrolysis were also studied) and the reduction potentials for the most relevant metabolites. Reduction is indeed believed to play a key role in the biological activity of these complexes, *i.e.*, the kinetically inert Ru(III) may be converted into the more labile Ru(II) complex. Similar reaction profiles to previous work for Ru(III) were found, but their data suggested that Ru(II) di-aqua complexes should be more abundant and possibly play a more important role in the biological activity of NAMI-A. Reduction of ICR was found to be more difficult than of NAMI-A, and had less effect on the overall hydrolysis path.

Chiorescu *et al.*[50] used DFT and SCRF methods to study the Ru(III)/Ru(II) process in more detail, reporting "unprecedented accuracy" in their predictions of experimental data for 61 ruthenium complexes in four solvents, including NAMI-A. The effects of basis set, and especially details of the SCRF models, were systematically tested, with radii for atoms that make up the solute cavity a particular focus.

Ruthenium features in another class of promising compounds first developed and tested by Sadler,[51] in which arene and ethylenediamine ligands are coordinated to Ru, with a single chloride ion. As with other drugs, this chloride is believed to be lost by hydrolysis, leaving a single site for DNA adduct formation. Peacock *et al.* studied the hydrolysis and nucleobase binding of some ruthenium complexes, including O,O- and N,N-chelating ligands, and compared these to the behaviour of equivalent osmium compounds, using DFT and SCRF methods.[52] This work showed that hydrolysis of the O,O-complexes was significantly faster than that of N,N-complexes.

Bešker *et al.* used DFT to study the binding of some simple ruthenium complexes such as $[Ru(NH_3)_5]^{n+}$ to nucleobases, and also to some common fragments of amino acids, as well as hydrolysis of chloride complexes.[53] The complexes conform to the expected octahedral geometry and contain multiple hydrogen bonds (typically to O6 of guanine or N6 of adenine). Bond dissociation energies of Ru(II) and Ru(III) complexes were calculated, using SCRF solvation models, leading to similar trends in binding to Pt complexes and stronger binding for Ru(III) than Ru(II). Some trends between bond strengths and proton affinity and/or pK_a of the corresponding bases were observed, suggesting that the former might be a useful guide to biological activity.

Gossens *et al.*[54] used standard DFT and MP2 methods as well as CPMD to study the binding of these complexes to different DNA bases, finding a similar order to the preference for binding site observed for cisplatin and related species, G(N7) >> C(O2) ≈ C(N3) > A(N7) > G(O6) > OH2. Very low barriers for the rotation of the arene were found, indicating that the ligand can re-orient itself to maximise or minimise interactions with its environment. Strong hydrogen bonding was

observed between coordinated N—H groups of the ethylene-diamine ligand and O6 of guanine, as was flexibility of the backbone N—C—C—N dihedral of the ethylenediamine ligand. The complex to N7 of adenine was found to be significantly less stable than the guanine complex, in accord with experimental findings that greater preference for guanine is found with these Ru complexes than with cisplatin-type complexes. Evidence for a weak hydrogen bond to N6 of adenine was found, which requires a change of hybridisation (and hence destabilisation) of the amino group.

Dorcier *et al.*[55] proposed a combined experimental and theoretical study of organometallic ruthenium(II) and osmium(II) anticancer complexes binding to an oligonucleotide. They firstly characterized such complexes by means of NMR, mass spectroscopy and crystal X-ray diffraction. Secondly, they explored the biological activity by investigating the interaction between these complexes with a DNA fragment, employing electrospray ionization mass spectroscopy. Unlike Sadler's compounds, loss of arene groups was observed in certain cases, which may represent a new binding mechanism for a metal-DNA adduct. DFT calculations were then used to rationalize these findings. These calculations suggested that several factors, including change of the metal centre (ruthenium *vs.* osmium), methylation/protonation of the ligand and nature of the arene group (*p*-cymene vs. benzene) affected the metal-arene binding energies.

Rhodium is another transition metal that shows promise for treatment of cancers. Deubel recently reported DFT studies of the formation of dirhodium-DNA adducts.[56] Unlike most other metal complexes, this preferentially forms adducts to adenine rather than guanine, and in doing so appears to stabilise an unusual tautomeric form of the nucleobase which can act as a bridging ligand between the two rhodium centres. DFT methods identified the binding energy of various tautomers of adenine to rhodium in both monodentate and bidentate forms, supporting the experimental observation of preference for adenine over guanine. Transition states for formation of bidentate complexes, and for the necessary tautomerisation, were identified and associated with barriers of between 20 and 30 kcal mol-1. Modification of the DNA interactions of such complexes, for instance by increasing the lability of carboxylate leaving groups, were also examined, leading to concrete proposals for synthesis of new drug targets.

A similar problem was examined by Burda and Gu,[57] who used B3LYP, NBO analysis and electrostatic potentials to examine the structures, stabilities and properties of dirhodium adducts to adenine and guanine. Head-to-head and head-to-tail orientations of nucleobases were considered and again the stability of adenine adducts was noted. However, since formation of these adducts requires proton transfer (tautomerisation), the energetics of this also affects the final thermodynamic stability of complexes. This is easier in guanine than in adenine, counteracting the apparent stability of adenine adducts.

Šponer *et al.*[58] used DFT to examine titanocene, $[Ti(Cp)_2Cl_2]$, Cp=cyclopentadiene. Despite superficial similarities to cisplatin, the mode of action of this complex is believed to be rather more complex, and may involve loss of Cp as well as chloride ligands. A range of alkylammonium-substituted titanocenes, designed to increase aqueous solubility without reducing biological activity, were examined and the effects of the substituents on electronic properties and Cp binding energies determined. This work suggested that proton-induced loss of Cp might play an important role in the biochemistry of these complexes, and the energy of protonation shows some correlation with *in vivo* anti-tumour activity. This supports the hypothesis that both Cl and Cp are lost during the action of this drug.

1.4 Non-covalent interactions

Most of the systems described in the previous sections are characterised by the presence of non-covalent interactions. This term refers in general to interactions between atoms of the same (intramolecular) or different molecules (intermolecular) which are not linked directly by a covalent bond. All non-covalent interactions are weak in comparison to covalent binding, however their interplay and additivity effects render them as an important topic in biological systems, materials science and molecular recognition processes. The DNA molecule is perhaps the most prominent example of the importance of these interactions, as discussed in the following two sections.

1.4.1 Structure of DNA

During the last decades the DNA molecule has met increasing publicity, due to its fundamental biological importance and its relation to heredity and storage of information. The well known double helix consists of two interwound nucleic acid polymer chains and the details of its structure can be found in most common textbooks, such as those of references 59,60,61.

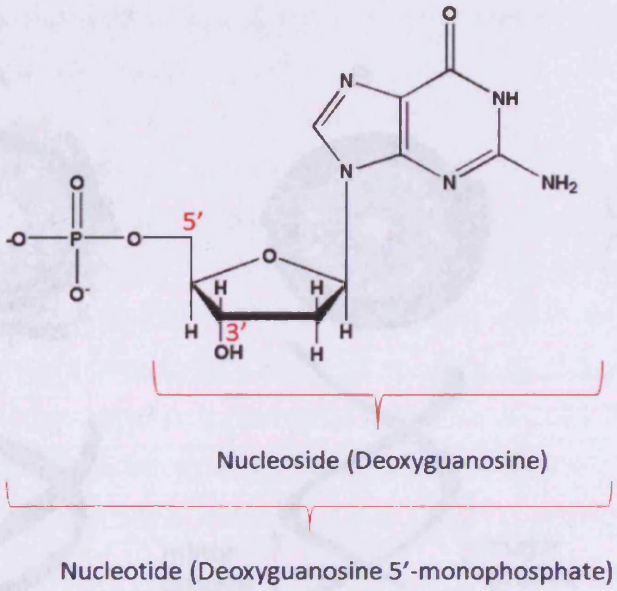


Figure 1.4.1: Building block of DNA.

The DNA molecule is characterised by its primary, secondary and tertiary structure. The building blocks of each nucleic acid chain are termed as nucleosides and the way they are linked together forms the primary structure. Each nucleoside is a combination of a deoxyribose sugar and one of four bases adenine, thymine, guanine and cytosine (A, T, G and C). Two nucleosides are linked through phosphate groups, which form esteric bonds with the 5′- and 3′- oxygen atoms of the ribose sugar. A nucleoside with one, two or three phosphate groups attached to it is referred to as a nucleotide (figure 1.4.1).

The secondary structure is the double helical arrangement of the two nucleotide chains that was shown by Watson and Crick. The two helices run in antiparallel directions with the complementary base pairs (A-T and G-C) positioned in the interior of the double helix. In this arrangement, each base is hydrogen bonded to its pair, while consecutive base pairs are

positioned in a stacked manner. The sugar-phosphate units lie at the exterior of the helix and serve as the backbone of the structure. Another feature that is observed in the secondary structure is the presence of the major and minor grooves (shown in figure 1.4.2). This feature stems from the fact that the bonds of the paired bases to their corresponding sugar ring are not directly opposite in space.

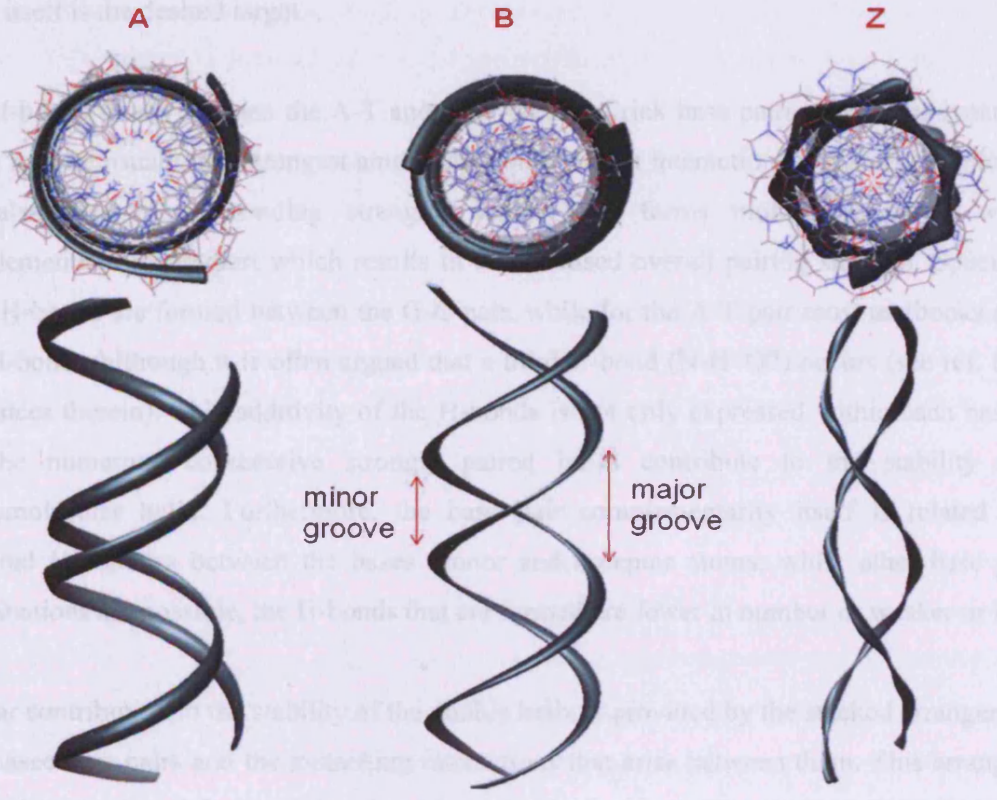


Figure 1.4.2: Axial (top) and side (bottom) view of A-, B- and Z-DNA forms.

Finally, the double helix can occur in various different forms, the best known of which are the A, B and Z. These types of helices differ from each other in the orientation (right- or left-handed) and the conditions under which they are synthesized (e.g. humidity and salt concentration), as well as in the spatial characteristics of the major and minor grooves and the number of bases per turn. The most common DNA form *in vivo* is B, since it occurs in high humidity conditions such as those found in the interior of living cells.

1.4.2 Non-covalent interactions in DNA

Evidently, the DNA molecule is characterised mainly by the presence of H-bonding and π-stacking among the spectrum of non-covalent interactions. A better consideration of its structure, especially at a macromolecular level, reveals the importance of the presence of these interactions for the stability of the helix, as well as their role in molecular recognition processes when the DNA itself is the desired target.

The H-bonds found between the A-T and G-C Watson-Crick base pairs are thermodynamically stable and are usually the strongest among the non-covalent interactions, yet stacking interaction can also resemble H-bonding strengths. Each base forms multiple H-bonds with its complementary counterpart which results in an increased overall pairing strength. Specifically, three H-bonds are formed between the G-C pair, while for the A-T pair most textbooks refer to two H-bonds, although it is often argued that a third H-bond (N-H···O2) occurs (see ref. 62 And references therein). This additivity of the H-bonds is not only expressed within each base pair, but the numerous consecutive strongly paired bases contribute to the stability of the macromolecular helix. Furthermore, the base pair complementarity itself is related to the favoured H-bonding between the bases' donor and acceptor atoms; while other base pairing combinations are possible, the H-bonds that are formed are fewer in number or weaker or both.

Similar contribution to the stability of the double helix is provided by the stacked arrangement of the consecutive pairs and the π-stacking interactions that arise between them. This arrangement is on the one hand favoured by the relatively rigid sugar rings. Furthermore, in aqueous environment entropic effects arise and are manifested as hydrophobic interactions between the nucleobases and water, thus further assisting the stacking of the base pairs. This kind of stabilising interaction is also additive across the extent of the DNA structure and enhances its chemical stability.

Another type of non-bonded interaction encountered in DNA systems is the purely electrostatic interaction between the negatively charged phosphate groups of the helix backbone and positively charged ions, such as Na^+ and Mg^{2+}. The presence of such positive ions contributes to

the overall stability by balancing the development of strongly repulsive forces between the negative phosphate groups.

Non-covalent interactions can also have a key role when DNA interacts with other molecules, as for example, in anticancer chemotherapy when DNA is the drug target. One class of drugs is the intercalators, which are inserted between consecutive nucleobase base pairs. Such drugs contain planar aromatic groups and thus stacking interactions arise between the drug's aromatic rings and the DNA bases. Another class of drugs are alkylating agents, which form a covalent bond with appropriate sites on the DNA, such as in the case of metals that bind to the nucleophilic N7 nitrogen atoms of guanine and adenine. Apart from the N7 sites, exocyclic groups of the nucleobases can participate in secondary interactions with appropriate carrier ligands of the metals, as is the case with the Ru(II) complexes that were mentioned in section 1.3. The carbonyl oxygen of guanine can serve as an H-bond acceptor and the amino group of adenine as an H-bond donor. Such secondary interactions may increase both the drug target specificity and the binding strength.

1.4.3 Modelling of non-covalent interactions

Due to the importance of non-covalent interactions, as exemplified through the overview of the DNA's structure, a proper description of such interactions is an essential feature of any suitable method for modelling of drug-DNA interactions. As a result, the development of appropriate modelling methods that are able to accurately describe similar systems is an active research field. H-bonding is mainly electrostatic in nature and modelling of H-bonded systems does not exhibit significant difficulties. For example, standard methods such as Hartree-Fock and DFT with small to medium sized basis sets reproduce all key features of hydrogen bonded complexes, including geometries and binding energies. This is not the case, however, with π-stacking interactions. The latter arise as a result of London dispersion forces, which in turn are theoretically associated to electron correlation. Details of theoretical methods are given in Chapter 2.

A first expected consequence is that methods that do not account for electron correlation fail to properly describe dispersion interactions. This is the case for the Hartree-Fock (HF) theory, whose formalism lacks of electron correlation and thus cannot accurately describe stacked systems, not only quantitatively but qualitatively as well.

Post-HF methods account for electron correlation and therefore provide better results when dealing with dispersion interactions. At present, the most accurate method is considered to be the coupled cluster theory with non-iterative triple excitations (CCSD(T)). For best results, at least an aug-cc-pVTZ basis set[63] is required, a fact that is accompanied by a significant computational cost. Additionally, the necessity to calculate counterpoise (CP) corrections[64] to the BSSE further increases the computational cost and limits the applicability of CCSD(T) to small and medium sized systems. A compromise method for larger systems involves extrapolation of MP2 correlation energy (see below) to the basis set limit, followed by correction for the shortcomings of MP2 by adding the difference ($E_{corr}^{CCSD(T)} - E_{corr}^{MP2}$), which is relatively basis-set independent and so can be evaluated with smaller basis sets such as cc-pVDZ

The first *ab initio* alternative that incorporates electron correlation with lower computational effort than CCSD(T) is the second order Moller-Plesset perturbation theory (MP2). However, it is well established that stacking interactions are overestimated at this level of theory, even after extrapolation to the complete basis set (CBS) limit. Both CCSD(T) and MP2 methods are still used to test the performance of other methods. In 2005 Hobza *et al.* gathered binding energies of a large number of non-covalently bound complexes evaluated by their group[65]. The resulting dataset, named JSCH-2005, consisted of the binding energies of several H-bonded, stacked and interstrand nucleobase pairs, as well as amino acid pairs, calculated at the CCSD(T) and MP2 levels and extrapolated to the complete basis set limit. A smaller subset of JSCH-2005, named S22 dataset, was proposed by the authors as a useful benchmark set for comparison of newly developed methods and consisted of 7 H-bonded, 8 dispersion dominated and 7 mixed complexes.

The disadvantages of the CCSD(T) and MP2 methods have directed modern research into the path of both increasing the accuracy of the calculations and lowering the computational cost. For example, local correlation (L) MP2 methods use localised orbitals to restrict the excitations to

sets of virtual orbitals (domains) that are spatially close to the occupied orbitals, and in doing so reduce computational cost[66,67,68,69,70]. Additionally, they are by construction free of BSSE[66,67,68] when used with moderately large basis sets and also scale linearly with size of the systems[69,70]. The use of the density fitting approximation (DF) (or resolution of the identity) replaces expensive 4-index 2-electron integrals with a combination of 2- and 3-index ones, thus reducing the time required for the MP2 calculation by around an order of magnitude[71,72,73]. The combination of DF and LMP2 methods, referred to as DF-LMP2[74], reproduces traditional MP2 potential energy surfaces for the benzene dimer[75].

Despite this advance in achieving faster calculations and less disk storage requirements, DF-LMP2 did not improve on the MP2's overestimation of stacking interactions. An alternative modification of MP2 scales the same-spin and opposite-spin electron pair contributions to the correlation energy by different factors[76]. This method is termed as spin component scaling MP2 (SCS-MP2) and has been applied successfully to pyridine dimers and trimers[77] and azulene dimers[78]. The same method combined with DF (DF-SCS-MP2) was applied to the benzene dimer[75] but failed to accurately describe stacked nucleic acid complexes[79]. Different scaling parameters optimised for non-covalent interactions have been published, such as SCSN[79] that is optimised for stacking interactions in nucleic acids and SCS(MI)[80] that is optimised against the more diverse S22 set. The application of DF-SCSN-LMP2 was tested against Hobza's CCSD(T) data of the S22 set with a mean error of 0.27 kcal mol^{-1}.

Another strand of research that shows promise is the use of density functional theory (DFT). Standard methods of DFT perform well for hydrogen bonding, but fail completely in their description of π-stacking[81]. This failure is generally ascribed to the lack of dispersion in Kohn-Sham DFT. Several functionals have been proposed for better description of non-covalent interactions, such as the time-dependent DFT (TD-DFT)[82], van der Waals density functional (vdW-DF)[83,84], mPW2-PLYP[85], Truhlar's M05-2X[86], M06[87] and PWB6K[88] functionals, as well as KT1[89], BR and XX[90].

Other promising methods include dispersion-like terms in DFT, via Lennard-Jones or similar parametric forms (DFT-D)[91,92,93]. Another approach varies the description of exchange within the exchange-correlation functional to mimic the effects of dispersion[87,94,95]. Within the various

methods to take this approach, previous work done in our group showed that Becke's half-and-half functional, BHandH[96], performs surprisingly well in describing π-stacking interactions, performing better than MP2 on average across a wide range of stacked complexes[97]. Additionally, BHandH was employed successfully in previous work of our group for the study of DNA oligonucleotides in the gas phase[98] and the binding of cisplatin to oligonucleotides[28]. However, BHandH consistently overestimates the strength of hydrogen bonds, leading to an imbalance in description of structures where both stacking and H-bonding are present and previous work has shown the importance of environment (water, counterions etc.) in reproducing experimental data for the interactions within and between DNA strands[99,100].

1.5 Absorption/distribution/metabolism/excretion (ADME) aspects

It is increasingly acknowledged that favourable physico-chemical properties of molecules should be taken into account alongside the search for high *in vivo* or *in vitro* activity in the search for new drugs. The factors that lead to a molecule having appropriate chemical properties are generally grouped under the heading of absorption, distribution, metabolism, and excretion (ADME) properties. Absorption and distribution factors include aqueous solubility, lipophilicity, bioavailability, transport across barriers found in cell walls, intestinal walls, or between blood and brain. Metabolism is vital for many metal complexes, since the active species is not usually the one administered to patients. The hydrolysis of cisplatin is one example of this, but is so fundamental to the chemistry and biochemistry of this drug that it is not often described as an ADME problem.

An important factor in the design of new organic drugs is the lipophilicity of the molecule, usually taken as the logarithm of the partition coefficient of the species between *n*-octanol and water, log*P*. Methods to estimate the log*P* of typical organic species are common, but applications to metal complexes are scarce, and have been limited by a lack of experimental data. Several years ago members of our group set out a method to estimate log*P* for platinum complexes, based on statistical correlations between exposed surface areas of polar, halogen and metal atoms.[101] Geometries of 24 compounds, mostly Pt(II) but also some Pt(IV) were obtained

by HF optimisation, and surface areas calculated from these geometries. Acceptable correlations were found and predicted log*P* was shown to have exponential relation with the uptake of platinum compounds into cells. Subsequently, our group collaborated with colleagues who measured log*P* data for a further 24 Pt(II) complexes by RP-HPLC.[102] Geometries of each compound were optimised using B3LYP and properties related to lipophilicity such as dipole moment, polarisability and electrostatic potentials were calculated at these structures. Rather better statistical relations were found than in our initial study, which seemed to stem from the fact that all measurements were carried out using the same protocol in the same laboratory, rather than being gathered from many diverse literature sources.

More recently, we collated a literature dataset consisting of 43 Pt(II) complexes and a further 21 Pt(IV) complexes, each of which has log*P* reported.[103] DFT optimisation and property calculations were carried out as before, but the methods used previously did not work as well as might be hoped. The lack of reproducibility across different methods is certainly one source of error, but the limited variety of complexes used in previous studies also seems to be a limitation. Acceptable models of log*P* were constructed by use of a genetic algorithm to select the calculated descriptors that best model the literature data,[104] but the neatness of previous work is lost in this fashion. The model was then used to calculate log*P* for some metabolites of cisplatin, which cannot be easily measured, and used to rationalise observations on the distribution and uptake of these metabolites.

Metabolism is an important factor in metal drugs in two ways: *in vivo* activation of complexes to their active forms, and deactivation by proteins and other species. A recent alternative to variations on the theme of cisplatin has been the use of Pt(IV) complexes, which typically have octahedral coordination geometry. These are kinetically more inert than Pt(II) analogues and hence can be orally administered with fewer side effects. The Pt(IV) complex is seen as a "pro-drug", which must be altered by reduction to the Pt(II) form before it can form DNA adducts and exert its biological effect. Reduction potentials for the Pt(II)/Pt(IV) couple are therefore an important aspect of design: if reduction is too difficult the concentration of active species will be reduced, while if reduction is too easy the pro-drug reaching its target intact will be prevented.

Reduction potentials for common ligands and metals can be predicted from tabulated constants, computational attempts to calculate reduction potentials directly are scarce though.[105]

Deactivation of the active species is widely thought to occur through complexation by sulphur ligands, especially the side chains of cysteine and methionine amino acids, and by glutathione. Increased levels of glutathione have been associated with the onset of resistance to cisplatin therapy, such that this aspect has received a great deal of attention. Zimmermann *et al.*[106] used B3LYP to model the interactions of cisplatin and its metabolites with some model thiol ligands. Monodentate and chelate complexes were found to be stable and strong Pt—S bonds were found, in accordance with the Hard-Soft Acid-Base (HSAB) principle. Reaction of cisplatin with cysteine is exothermic and thermodynamically competitive with complexation with guanine, whereas methionine is a weaker ligand to Pt. Da Silva came to similar conclusions from their DFT and *ab initio* studies.[107] Predicted rate constants for substitution with sulphur ligands were in good agreement with experimental data.

Deubel and co-workers investigated the potential implications of ammine loss from cisplatin, linking this phenomenon to inactivation, storage and resistance of cisplatin.[108] In particular, they employed the B3LYP functional and implicit solvation to study several cisplatin complexes with various biomolecules (including nitrogen heterocycles, neutral and anionic sulphur ligands) to evaluate the trans influence and trans effect of the ligands. Upon binding of cisplatin hydrolysis products with these ligands, loss of ammine was predicted as the predominant reaction. It was found that the charge of Pt(II) centre has a little effect for ammine displacement, while anionic and neutral sulphur ligands exert the strongest trans influence/effect.

Chapter 2: Theory

The following paragraphs outline the basic theoretical background of the methods and techniques that were used for the present work. Most of the content comes from standard textbooks [109,110,111,112,113,114,115,116,117,118], in which further information on each topic can be found, unless otherwise cited.

2.1 Hartree-Fock Theory

2.1.1 Introduction – Schrödinger Equation

The non-relativistic, time-independent Schrödinger equation has the shorthand form:

$$H\Psi = E\Psi \tag{2.1.1}$$

and is the basis for the development of the Quantum Mechanics theory. Expansion of this equation would show that it is a 2^{nd}-order partial differential equation, where H is a differential operator (Hamiltonian operator), Ψ is a wavefunction that describes the system and E is a scalar number (energy). It is therefore concluded that equation 2.1.1 is an eigenvalue equation and thus can be solved by determining the function Ψ and the eigenvalues E.

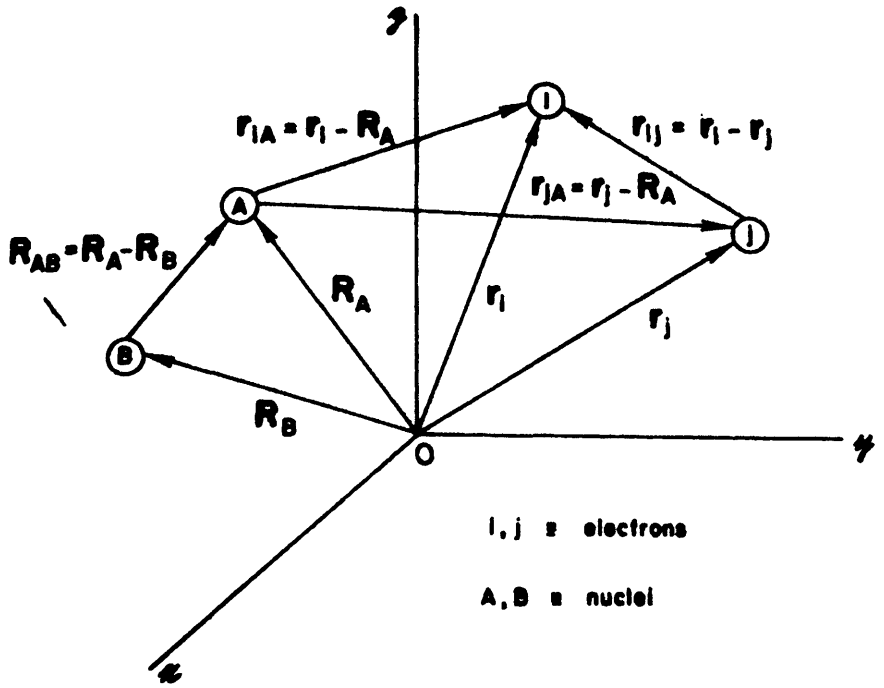


Figure 2.1.1 Molecular coordinates (Fig. taken from ref.110)

The Hamiltonian operator is easily constructed for a system of particles. Figure 2.1.1 displays a molecular coordinate system, in which $r_{iA} = |r_i - R_A|$ is the distance between the i-th electron and A-th nucleus, $r_{ij} = |r_i - r_j|$ is the distance between the i-th and j-th electrons and $R_{AB} = |R_A - R_B|$ is the distance between the A-th and B-th nuclei. Generalisation of this system for M nuclei and N electrons, using atomic units, gives the Hamiltonian the general form:

$$H = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \qquad (2.1.2)$$

where Z_A is the nuclear atomic number, M_A is the ratio of the mass of nucleus A to the mass of an electron and ∇^2 is the Laplacian operator:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \qquad (2.1.3)$$

Furthermore, in the expression 2.1.2 of the Hamiltonian the first two terms represent the kinetic energy operators for the electrons and nuclei, respectively. The remaining three terms represent the Coulombic interactions between the particles: electron-nucleus attraction, electron-electron and nucleus-nucleus repulsion, respectively.

The solution of the Schrödinger equation for three or more particles is impossible, due to the correlated motions of the particles, which is a fundamental problem, termed as the 3-body problem. The Born-Oppenheimer (B.O.) approximation is one of the fundamental approximations used in quantum chemistry in order to finally tackle the problem of solving the Schrödinger equation for many particle systems.

2.1.2 Born-Oppenheimer Approximation

The B.O. approximation introduces a number of sophisticated approximations. The nuclear kinetic energy is neglected and the nuclear-nuclear repulsion is treated as a constant. In this way, the motions of the particles are decoupled. This approximation does not introduce large errors and is easily justified by the large mass of the nuclei compared to the electronic mass which results in the nuclear motion being much slower than the electronic. As a consequence, the electronic Hamiltonian takes the form:

$$H_{elec} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \qquad (2.1.4)$$

Additionally, the wavefunction $\Psi_{tot}(r,R)$ of the system is separated into a product of an electronic $\Psi_{elec}(r,R)$ and a nuclear $\Psi_{nuc}(R)$ part :

$$\Psi_{tot}(r,R) = \Psi_{elec}(r,R)\Psi_{nuc}(r,R) \qquad (2.1.5)$$

$\Psi_{elec}(r,R)$ is the solution of the Schrödinger equation for the electronic Hamiltonian:

$$H_{el}\Psi_{el}(r;R) = E_{el}\Psi_{el}(r;R) \qquad (2.1.6)$$

and depends explicitly on the electronic coordinates and only parametrically on the nuclear coordinates. The same coordinate dependence holds for the total energy, which consists of the electronic energy and the constant internuclear repulsion.

The total energy derived in the way described above serves as a potential under which the nuclei move, i.e. the nuclei move in an average field of electrons, resulting in a potential energy surface (PES). The nuclear Hamiltonian is therefore:

$$H_{nuc} = -\sum_{A=1}^{M} \frac{1}{2M_A} + \varepsilon_{tot}(\{R_A\}) \tag{2.1.7}$$

Solution of the nuclear Schrödinger equation can describe the vibration, rotation and translation of a molecule.

2.1.3 Interpretation of the wavefunction

The wavefunction that obeys the above equations and describes the system has no apparent physical meaning. The universally accepted interpretation of the wavefunction is known as the "Born interpretation", which regards the quantity

$$|\Psi_i|^2 = \Psi_i^* \Psi_i \tag{2.1.9}$$

as a probability density. When $|\Psi_i|^2$ is multiplied by the volume element $d\tau$ it provides the probability of finding an electron in this certain volume. A further consequence is that integration of the probability over all space should equal the total number of electrons that are contained in the system:

$$\int \Psi_i^* \Psi_i d\tau = N \tag{2.1.10}$$

Additionally, if Ψ is an eigenfunction of an operator $\hat{A}$ which describes a physical quantity (observable), its value can be calculated as:

$$\langle A\rangle = \frac{\int \Psi_i^* \hat{A} \Psi_i d\tau}{\int \Psi_i^* \Psi_i d\tau} \qquad (2.1.11)$$

The denominator in the last expression equals 1, if the wavefunction is normalised.

2.1.4 Description of the wavefunction - Molecular Orbital Approximation

The exact solution of the Schrödinger equation for many electron systems is not possible, so further approximations must be made. The first of these is the molecular orbital (MO) approximation, in which the total wavefunction is made up of a product of one electron wavefunctions, the molecular orbitals. For the proper description of electrons, apart from their spatial coordinates spin functions are also essential in order to describe the two distinct electronic spin orientations (up and down). Therefore, each electron depends on four coordinates, three spatial and one spin. Single-particle wavefunctions expressed as products of a spatial ($\psi(r)$) and a spin ($\alpha(\omega)$, $\beta(\omega)$) function are called spin orbitals $\chi(x)$.

$$\chi(x) = \begin{cases} \psi(r)\alpha(\omega) \\ \quad or \\ \psi(r)\beta(\omega) \end{cases} \qquad (2.1.12)$$

Many electron systems can be simply represented as a product of spin orbitals. Such products are called Hartree products Ψ^{HP}.

$$\Psi^{HP}(x_1, x_2, ..., x_N) = \chi_i(x_1)\chi_j(x_2)....\chi_k(x_N) \qquad (2.1.13)$$

However, interchange of the four coordinates x between any two electrons must be followed by change on the sign of the wavefunction, in accordance with the antisymmetry principle. While Hartree products do not satisfy this requirement, linear combinations can be written which do and wavefunctions expressed as such linear combinations can equivalently be written in determinantal form (Slater determinants).

$$\Psi(x_1, x_2, ..., x_N) = (N!)^{-1/2} \begin{vmatrix} \chi_i(x_1) & \chi_j(x_1) & \cdots & \chi_k(x_1) \\ \chi_i(x_2) & \chi_j(x_2) & \cdots & \chi_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(x_N) & \chi_j(x_N) & \cdots & \chi_k(x_N) \end{vmatrix} \quad (2.1.14)$$

2.1.5 Hartree-Fock Approximation

Since the three-body problem cannot be solved, approximate solutions of the Schrödinger equation need to be sought. The Hartree-Fock method introduces the one-electron Fock operator (eq. 2.1.15) in order to tackle this problem and reduce the many-electron problem into one-electron problem.

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + U^{HF}(i) \quad (2.1.15)$$

In this way, the electron-electron repulsion is expressed in an average way by the average potential $U^{HF}(i)$ that the i-th electron experiences due to the presence of all the rest of the electrons. The term $U^{HF}(i)$ includes the Coulomb (J) and exchange (K) interactions:

$$U^{HF}(i) = \sum_u \{J_u(i) - K_u(i)\} \quad (2.1.16)$$

The former is the classical electrostatic repulsion between the charged electrons and the latter is related to the spin correlations and has no classical equivalent.

The equation to be solved after incorporating the Fock operator is the Hartree-Fock equation:

$$f(i)\chi(x_i) = \varepsilon\chi(x_i) \quad (2.1.17)$$

One implication that stems from the form of the Fock operator is that the solution for one electron affects the solution for the rest of the electrons through the average potential term. Thus, eq. 2.1.17 has to be solved iteratively. This iterative procedure is called the Self-consistent Field (SCF) method.

2.1.6 Roothaan-Hall equations – Orbital energies

Within the MO approximation, a common approximation is for each MO to be written as linear combination of atomic orbitals:

$$\phi_i = \sum_{\alpha}^{M_{basis}} c_{ai}\chi_\alpha \tag{2.1.18}$$

and the Hartree-Fock equations (2.2.17) take the form:

$$f(i)\sum_{a}^{M_{basis}} c_{ai}\chi_a = \varepsilon_i \sum_{a}^{M_{basis}} c_{ai}\chi_\alpha \tag{2.1.19}$$

Multiplication by the left of the equations (2.2.19) with a specific basis and integration leads to the Roothaan-Hall equations, which can be written in matrix form as:

$$\boldsymbol{FC=SC\varepsilon} \tag{2.1.20}$$

In the last expression, F is the Fock matrix, S the overlap matrix, C the matrix of the coefficients c_{ai} and ε is a matrix with the orbital energies. Each term ε_i represents the energy of an electron in a spin-orbital i, i.e. it is the sum of the core interaction and the Coulomb and exchange interactions with the other electrons in the system. The total energy of the system is therefore related to the orbital energies through the expression:

$$E = \sum_{i=1}^{N} \varepsilon_i - \sum_{i=1}^{N_2}\sum_{j=1}^{N_2}(2J_{ij} - K_{ij}) \tag{2.1.21}$$

2.2 Basis Sets

2.2.1 General Overview

Equation 2.1.18 is a mathematical expression of the Linear Combination of Atomic Orbitals (LCAO) approximation, the most widely used method for obtaining MO coefficients and energies. The most common types of basis set functions are the Slater-type orbitals (STO) (eq. 2.2.1) and the Gaussian-type orbitals (GTO) (eq. 2.2.2)

$$\text{STO: } \chi_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{n-1}e^{-\zeta r} \tag{2.2.1}$$

$$\text{GTO: } \chi_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{2n-2-l}e^{-\zeta r^2} \tag{2.2.2}$$

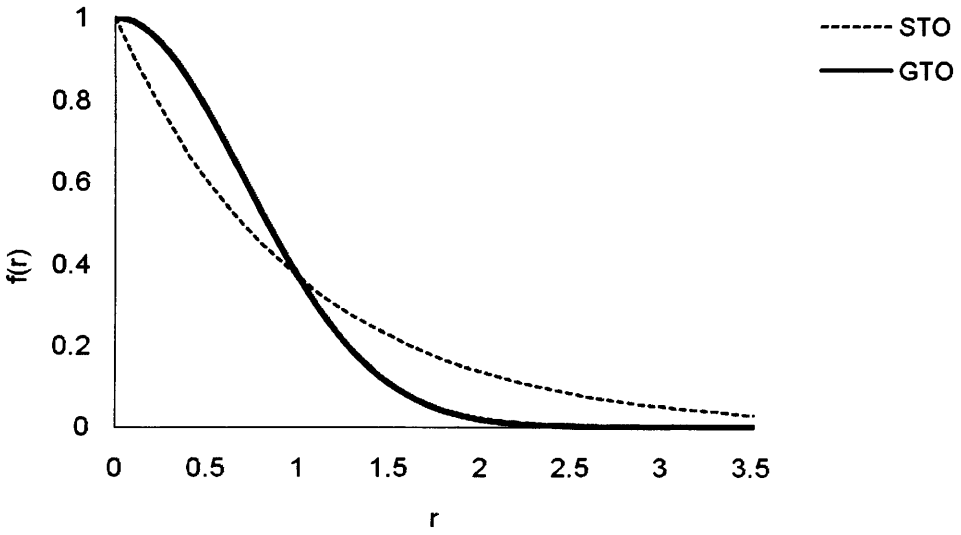


Figure 2.2.1: Comparison of STO and GTO functions.

The STOs provide a better behaviour of the systems at r=0 and at large r than the GTOs. For example, the former can describe the cusp at the nucleus by having a finite slope at r=0, while the latter have zero slope at r=0. For large r, GTOs show a more rapid decay than STOs (figure

2.2.1). However, while solving the HF equations, three- and four-electron integrals occur and the evaluation of such kinds of integrals using STOs that are located on different nuclei is extremely time consuming, while with the employment of GTOs this computational problem is avoided.

The qualitative shortcomings of the Gaussian functions can be countered by expressing the basis functions as linear combinations of primitive Gaussian functions. Such orbitals are called contracted Gaussian functions (CGF):

$$\text{CGF: } \phi_{\mu}^{CGF} = \sum_{i=1}^{L} d_{i\mu}\phi_i(\alpha_{i\mu}) \qquad (2.2.3)$$

where $d_{i\mu}$ are expansion coefficients and L is the number of functions in the expansion (or length of the contraction).

2.2.2 Minimal Basis Set - Example of Basis Set

A minimal basis set is the one that uses the least number of functions that are required to describe the occupied orbitals of each atom. For example, for the description of H or He atoms the minimal basis set would consist of a single s-type function. The STO-3G is a minimal basis set in which each orbital is represented by three Gaussian functions. Such basis sets are often not accurate enough for the proper description of a system and a further improvement can be achieved by doubling all the basis functions (double-zeta, DZ basis sets). Another approach is to use split-valence basis sets. In this case the number of functions is increased only for the valence electrons, which are the most crucial upon chemical transformations, while the inner electrons are described less accurately. Further improvements, according to the problem of interest, can be attained by addition of functions of higher angular momentum (polarisation functions) or diffuse functions.

One example of a basis set is the 6-31+G** of the Pople series of basis sets and has been used extensively in this work. It comprises of a contraction of six primitive Gaussian functions for the core orbitals, three primitives for the inner part of the valence electrons and one for the outer part of valence electrons. Additionally, it has one set (+) of s- and p- diffuse functions for the heavy

atoms, as well as a d- polarisation function for the heavy atoms and a p- polarisation function for the hydrogen atoms.

2.2.3 Basis Set Superposition Error

The description of the wavefunction as a linear combination of basis functions would be accurate if an infinite number of the latter is used. In practice, though, it is possible to use only a finite number of basis functions and this incompleteness in the expansion introduces an error to the calculated energy. One implication of the basis set incompleteness is that the electronic density of a nucleus can be described by additional functions located at another nucleus, a fact that artificially lowers the energy. This kind of error is known as Basis Set Superposition Error (BSSE) and is of high importance especially when weak intermolecular interactions are present.

Since employment of very large basis sets or the complete basis set (CBS) limit is not practical, other alternatives have been proposed to account for the BSSE, among which the most popular is the Counterpoise (CP) method[64]. According to the CP method, when estimating the interaction energy of a complex AB, the energy of each monomer (A and B) is calculated at the geometry it has in the complex using additionally the basis functions (a or b) of the other monomer, although without its nuclei present (ghost functions):

$$\Delta E = E^{a\cup b}(AB)_{AB} - E^{a\cup b}(A)_{AB} - E^{a\cup b}(B)_{AB} \qquad (2.2.4)$$

The superscripts in the above expression denote the basis set used and the subscripts the geometry in which each fragment considered.

2.2.4 Effective Core Potentials (ECP)

The quantum chemical treatment of metals is a challenging task, due to the increased number of electrons that are present. This at first implies that a large number of basis functions are required to properly describe the metal-containing system. The use of ECPs, or pseudopotentials, lies in

the philosophy of describing the core electrons by a potential, which is designed as an analytical function that accounts for the electrostatic effects.

Apart from the above computational consideration, heavy atoms such as metals, exhibit significant relativistic effects. The inner electrons that lie closer to the nucleus interact intensely with the increased nuclear charge and as a result their velocities increase to a non-negligible fraction of the speed of light. This effect must also be taken into account in order to have a more accurate description of a system. As a result, apart from the electrostatic effects, ECPs are often designed to account for relativistic corrections as well.

2.2.5 Design of ECPs

ECPs are designed by replacing the valence orbitals of a HF, Dirac-HF or DFT wavefunction by a set of nodeless pseudo-orbitals. The core electrons are replaced by a potential, which in turn is described by analytical functions of the distance between nuclei and electrons and can also include the above mentioned relativistic effects. Finally, the parameters of the analytical functions are fitted appropriately in order to reproduce the all-electron valence orbitals.
In this work, the Stuttgart-Dresden (SDD)[119] ECP has been employed in the calculations that involved metals.

2.3 Post-HF methods and Density Functional Theory (DFT)

In the previous discussion on the HF theory it was mentioned that the electron-electron interactions are described in an average way. As a consequence, the HF wavefunction is not ideal and the resulting energy is not the lowest possible (exact). This energetic difference between the HF energy and the exact energy is called the correlation energy (E_C):

$$E_C = E^{exact} - E^{HF} \qquad (2.3.1)$$

and it is often of high importance in the description of chemical systems. Using the HF theory as a starting point, more accurate but computationally expensive methods have been developed.

These methods are usually referred to as post-HF methods, such as Configuration Interaction (CI), Many-Body Perturbation Theory (MBPT) and Coupled Cluster (CC), as well as Density Functional Theory (DFT) which accounts explicitly for the correlation energy.

2.3.1 Configuration Interaction (CI)

This method accounts for the electron correlation by taking into account excited states of the system. In other words, the wavefunction is a linear combination of the ground state wavefunction and wavefunctions that correspond to excited states:

$$\Psi_{CI} = a_0\Phi_{HF} + \sum_S a_S\Phi_S + \sum_D a_D\Phi_D + \sum_T a_T\Phi_T + \ldots = \sum_{i=0} a_i\Phi_i \qquad (2.3.2)$$

where Φ_{HF} is a HF single determinant wavefunction and the rest of the Φ_i determinants represent excited states (singly S, doubly D, triply T etc.).

Inclusion of all the possible excitations leads to Full Configuration Interaction (FCI), which is the most accurate method. However, considering that each Φ_i is also written as a linear combination of basis functions it is evident that both CI and FCI are practical only for small systems, due to the large number of integrals that occur. Also, CI has the disadvantage of not being size-consistent, i.e. the energy of a system at large separation does not equal the sum of the energies of each isolated monomer.

There are modifications of the CI method aiming at reducing the number of computed integrals, by reducing the excitation level. Examples are the CI with singles (CIS), with doubles (CID), which include only singly and doubly excited determinants, respectively and CISD, which includes both types of excitation.

2.3.2 Many-Body Perturbation Theory (MBPT)

Another alternative for including electron correlation in quantum chemical calculations is the Many-body perturbation theory (MBPT). In contrast to the CI method, MBPT is size-consistent,

although not variational. In MBPT the Hamiltonian H of a system is written as a sum of a zeroth-order Hamiltonian H_0 and a perturbation V:

$$H = H_0 + \lambda V \tag{2.3.3}$$

where λ is a parameter such that $0 \le \lambda \le 1$.

The zeroth-order Hamiltonian refers to an unperturbed system with known eigenfunctions $\Psi_i^{(0)}$ and eigenvalues $E_i^{(0)}$. The eigenfunctions Ψ_i and eigenvalues E_i of the Hamiltonian of the real system are expanded as Taylor series in λ :

$$\Psi_i = \Psi_i^{(0)} + \lambda\Psi_i^{(1)} + \lambda^2\Psi_i^{(2)} + \ldots = \sum_{n=0} \lambda^n \Psi_i^{(n)} \tag{2.3.4}$$

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \ldots = \sum_{n=0} \lambda^n E_i^{(n)} \tag{2.3.5}$$

where n denotes the order of the corrections to the wavefunction and the energy. In chemical problems the zeroth-order Hamiltonian is chosen to be the Hartree-Fock Hamiltonian, as first applied by C. Møller and M.S. Plesset (MPPT).

Once the eigenfunction $\Psi_i^{(n)}$ of the desired order is is found, the energies can be calculated as in the following examples :

$$E_i^{(0)} = \int \Psi_i^{(0)} H_0 \Psi_i^{(0)} d\tau \tag{2.3.6}$$

$$E_i^{(1)} = \int \Psi_i^{(0)} V \Psi_i^{(0)} d\tau \tag{2.3.7}$$

$$E_i^{(2)} = \int \Psi_i^{(0)} V \Psi_i^{(1)} d\tau \tag{2.3.8}$$

$$E_i^{(3)} = \int \Psi_i^{(0)} V \Psi_i^{(2)} d\tau \tag{2.3.9}$$

From equations 2.3.6-2.3.9 it is apparent that at least second order MPPT must be applied so as to yield an improvement over HF theory. MPPT of this level is simply termed as MP2.

As mentioned above, MPPT is not a variational method, but it has been proved to be size-consistent for all orders. Also, in spite of being the *ab initio* method with the lowest computational cost, it is not simple enough to allow application on large scale systems. As a result, variations of the method have been developed that aim at reducing the calculations. For example, the use of localised orbitals reduces the computational cost by restricting the excitations to sets of virtual orbitals. These sets are termed as "domains" and consist only of those virtual orbitals that are spatially close to the occupied orbitals. Such methods are known as local correlation methods (e.g. LMP2)[66-70] and have the additional advantage of being free of basis sets superposition errors when a moderately large basis set is used.

A different approach makes use of the Density Fitting (DF) (or resolution-of-identity, RI) procedure, according to which expensive 4-index-2-electron integrals are replaced by a combination of 2- and 3-electron integrals and the computational effort of MP2 calculations can be reduced by an order of magnitude[71-73]. Finally, DF and local electron correlation (LMP2) have been effectively combined resulting in the method known as DF-LMP2[74].

One variation of this last method is the Spin-component-scaling MP2 (SCS-MP2)[76]. Here, the contributions of same- and opposite-spin electron pair contributions to the correlation energy are scaled by different scaling factors. On the same theoretical basis, the use of scaling parameters specifically optimised for nucleic acids comes under the term of Spin-component-scaling for nucleobases (SCSN)[79]. All the above modifications of MP2 have been mentioned in section 1.4.3.

2.3.3 Coupled Cluster (CC)

In the coupled cluster method the excited states are generated by applying an excitation operator T on a HF reference wavefunction Φ_0. The operator T has the form:

$$T = T_1 + T_2 + \ldots T_{N_{elec}} = \sum T_i \qquad (2.3.10)$$

where i denotes the excitation level. The coupled cluster wavefunction Ψ_{cc} is given by the expression:

$$\Psi_{CC} = e^{T}\Phi_0 \tag{2.3.11}$$

where

$$e^{T} = 1 + T + \frac{1}{2}T^2 + \frac{1}{6}T^3 + \ldots = \sum_{k=0}^{\infty}\frac{1}{k!}T^k \tag{2.3.12}$$

and the Schrödinger equation is:

$$He^{T}\Phi_0 = Ee^{T}\Phi_0 \tag{2.3.13}$$

When all the possible excitations are taken into account by applying the corresponding T operators, the coupled cluster wavefunction is equivalent to that of the FCI method. This implies a high computational cost and in practice truncated coupled cluster schemes are used instead. A few examples of such schemes are the CCSD (where $e^{T} = 1 + T + \frac{1}{2}T^2$), CCD (where $e^{T} = 1 + \frac{1}{2}T^2$), CCSDT (where $e^{T} = 1 + T + \frac{1}{2}T^2 + \frac{1}{6}T^3$).

2.3.4 Density Functional Theory (DFT)

The fundamental difference between DFT and the *ab initio* methods described so far, is that it is the total electronic density distribution which is calculated, instead of the wavefunction, and subsequently the total electronic energy of the system.

The origins of DFT date back to the 1920's, when the Thomas-Fermi model was developed in an attempt to derive the kinetic energy of the uniform electron as system. Remarkable advance in the development of DFT were the two Hohenberg-Kohn theorems. The first theorem demonstrated that the ground-state energy and other electronic properties of a system are fully

determined by its electron density in a unique manner. In other words, the energy E of the system is a functional of its density p(r). This dependence can be expressed as:

$$E[p(r)] = \int V_{ext}(r)p(r)dr + F[p(r)] \tag{2.3.14}$$

In the above expression the first term represents the interaction between the electrons and the external potential $V_{ext}(r)$ that the nuclei exert on the electrons. The term F[p(r)] contains both the kinetic energy of the electrons and all the inter-electronic interactions. The second Hohenberg-Kohn theorem proved that the calculated DFT energy of a system is an upper bound to its true energy, i.e. the variational principle is obeyed analogously to the HF theory.

A major difficulty in the application of DFT is that the exact form of the term F[p(r)] is not known and thus, through equation 2.3.14 the dependence of E on p(r) is also not known exactly. The Kohn-Sham formulation offered a way to tackle this setback by approximating the problematic term as a sum of terms that represent different contributions to the total functional:

$$F[p(r)] = E_{K.E.}[p(r)] + E_H[p(r)] + E_{XC}[p(r)] \tag{2.3.15}$$

The first term, $E_{K.E.}[p(r)]$, represents the electronic kinetic energy, $E_H[p(r)]$ is the electron-electron Coulomb repulsion and $E_{XC}[p(r)]$ expresses the exchange and correlation contributions, as well as a correction to the kinetic energy term, since the latter refers to non-interacting electrons.

Expansion of equation 2.3.14 to include the terms of equation 2.3.15 in their analytical form leads to the full expression of the energy dependence:

$$E[p(r)] = -\sum_{A=1}^{M}\int \frac{Z_A}{|r-R_A|}p(r)dr + \sum_{i=1}^{N}\int \psi_i(r)(-\frac{\nabla^2}{2})\psi_i(r)dr$$
$$+\frac{1}{2}\int\int \frac{p(r_1)p(r_2)}{|r_1-r_2|}dr_1dr_2 + E_{XC}[p(r)] \tag{2.3.16}$$

Furthermore, Kohn and Sham constructed one-electron orthonormal orbitals $\Psi_i(r)$, in the form of linear combinations of atomic orbitals, which served as a basis for expressing the electron density:

$$p(r) = \sum_{i=1}^{N} |\psi_i(r)|^2 \tag{2.3.17}$$

Use of the above formulation leads to the one-electron Kohn-Sham equations:

$$\left\{ -\frac{\nabla_1^2}{2} - \left(\sum_{A=1}^{M} \frac{Z_A}{r_{1A}} \right) + \int \frac{p(r_2)}{r_{12}} dr_2 + V_{XC}[r_1] \right\} \psi_i(r_1) = \varepsilon_i \psi_i(r_1) \tag{2.3.18}$$

Here, ε_i are orbital energies and $V_{XC}[r]$ is the exchange-correlation functional and the above equations can be solved in a self-consistent manner.

Most DFT methods that are used in practice to deal with chemical systems are based on the Kohn-Sham formalism and differ among them in the way they approximate the unknown exchange-correlation term. One class of such approximations is the Local Density Approximation (LDA), which is based on the uniform electron gas model. In this model, the electron density is considered as homogeneous and constant throughout all the system's space. Using the exchange-correlation energy per electron $\varepsilon_{xc}[p(r)]$ of the system, the $E_{XC}[p(r)]$ term estimated as:

$$E_{XC}[p(r)] = \int p(r) \varepsilon_{XC}[p(r)] dr \tag{2.3.19}$$

LDA has been generalised to distinguish between α and β densities in the exchange energy. This more general approach is known as Local Spin Density Approximation (LSDA). The idealistic model of a uniform electron gas is an obvious source of errors for the LSDA methods. As a result, effort has been taken to consider a non-uniform gas model as a starting point, by setting the exchange-correlation term to be dependent not only on the electron density, but on its

variation as well, i.e. on the derivatives of p(r). Such methods are known as gradient-corrected or Generalised Gradient Approximation (GGA) methods. Well known examples of exchange and correlation functional within the GGA context are, respectively, Becke's (B or B88)[120] and Lee-Yang-Parr's (LYP)[121] functional. Methods have also been developed to set the afore-mentioned dependence on the Laplacian of the density ($\nabla^2 p(r)$). Such methods are extensions of the GGA and are known as meta-GGA methods.

Finally, various DFT methods combine the exact exchange energy with the LSDA exchange energy and/or gradient correction terms. These methods are known as hybrid. The exact exchange can be calculated by a Slater determinant that consists of Kohn-Sham orbitals using the HF method. An example of a hybrid functional that has been extensively used in the present work is Becke's half-and-half (BHandH)[96], which uses half the exact and half the LSDA exchange energy:

$$E_{XC}^{HandH} = \frac{1}{2}E_X^{exact} + \frac{1}{2}E_X^{LSDA} + \frac{1}{2}E_C^{LSDA} \qquad (2.3.20)$$

Another well-known example is Becke's three-parameter functional which is used in the highly popular B3LYP[122,123] method. In B3LYP, the exchange energy is a combination of the exact, LSDA and Becke's correction to the LSDA exchange energies, while the correlation energy is a combination of the LSDA and LYP energies. All these terms are weighted using three parameters that are determined by fitting to experimental data.

It is apparent that regardless of the chosen DFT method, the electron correlation problem is accounted for. This feature combined with the computational efficiency of DFT calculations offer this method an advantage when compared to correlated *ab initio* methods, especially for applications on polyatomic systems. On the other hand, several DFT functionals exhibit limitations in the presence of interactions due to van der Waals forces, such as dispersion forces. Dispersion forces originate from electron correlation and the inexact description of the correlation energy terms often results in DFT interaction energies not having the proper R^{-6} behaviour at long range. There is however ongoing research for the derivation of functionals that

can account for dispersion interactions. Amongst these, Truhlar has proposed the M05[86] and M06[87] families of functionals that have been optimised specifically for use with non-covalent interactions. Also, Grimme and others have proposed adding an empirical R^{-6} dispersion term to DFT energies in order to model dispersion interactions[91]. Both of the above examples of functionals have been mentioned in section 1.4.3.

2.4 Potential Energy Surfaces

2.4.1 Minima and saddle points

The potential energy of a chemical system is dependent on the relative arrangement of its constituent atoms, as mentioned on section 2.1.2. A potential energy surface (PES) (fig. 2.4.1) is a hypersurface of the potential energy over all the possible arrangements of the atoms. There are 3N-6 coordinate dimensions that can describe the relative arrangement of the atoms of a system; therefore it is not possible to visualize the PES of polyatomic systems. Instead, the PES is explored in a mathematical sense, by locating its stationary points, i.e. the points were the first derivatives are zero. Among the possible stationary points, minima and saddle points are usually of chemical interest.

The lowest minimum is the global minimum and corresponds to the most stable configuration of a molecule. It is also possible for a molecule or chemical system in general to have more than one global minimum of equal energy that corresponds to structures related by symmetry. Local minima are characterised by low energy compared to the nearest neighbourhoods and do not imply the most stable configuration of the system.

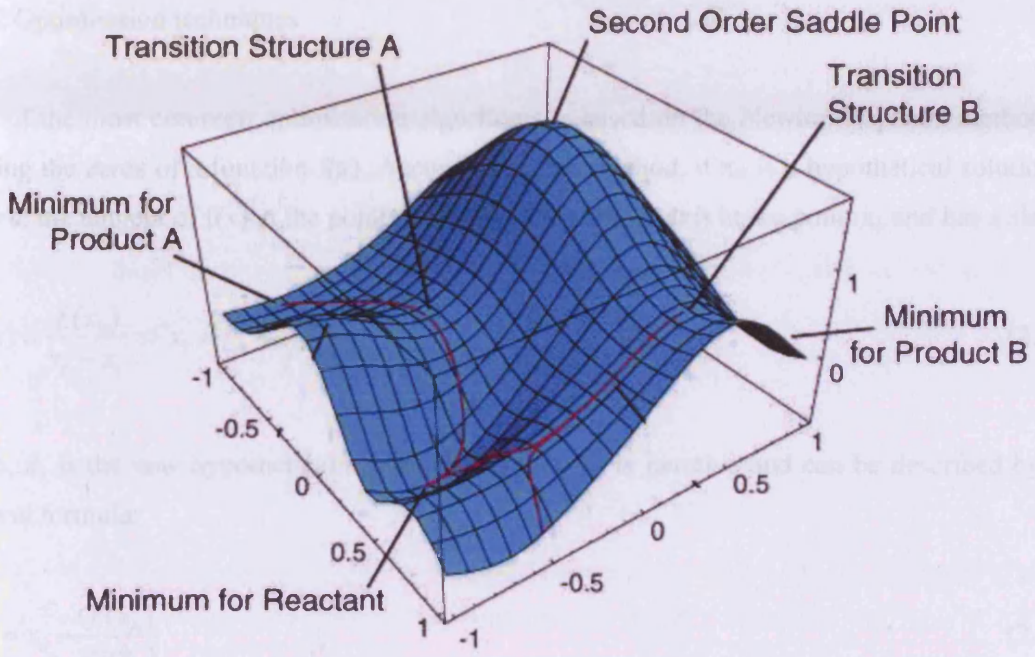


Figure 2.4.1: Example of a PES.

Whether the system is found at its global or at a local minimum, it also maintains zero-point energy which, if the surrounding barriers are not too high, can "move" the system from one minimum to another. The same effect can take place when enough energy is provided to the system leading it to overcome the surrounding barriers through the saddle points (or transition states). All stationary points are characterised by the first derivatives of the energy being zero. The classification as minima or saddle points is verified by calculating the second derivatives, i.e. the Hessian matrix; all positive eigenvalues correspond to a minimum and n negative eigenvalues (or imaginary frequencies) to a saddle point of n-th order.

Knowledge of the PES provides useful information for a system and its possible structures, as well as for the pathways it follows during isomerisation processes. Location of the minima and transition states of a molecule is achieved with the appropriate optimisation algorithms that are found in most computational chemistry packages.

2.4.2 Optimisation techniques

One of the most common optimisation algorithms is based on the Newton-Raphson method for finding the zeros of a function f(x). According to this method, if x_0 is a hypothetical solution of f(x)=0, the tangent of f(x) at the point $(x_0,f(x_0))$ crosses the x-axis at the point x_1 and has a slope:

$$f'(x) = \frac{f(x_0)}{x_0 - x_1} \Rightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \tag{2.4.1}$$

Then, x_1 is the new hypothetical solution. This process is iterative and can be described by the general formula:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{2.4.2}$$

The Newton-Raphson method has been combined with Pulay's extrapolation procedure of Direct Inversion in the Iterative subspace (DIIS)[124]. The latter is a method that aids the SCF convergence and its combination with the Newton-Raphson method is known as Geometry DIIS (GDIIS)[125], which is efficient for optimizing the geometry of large molecules.

2.5 QM/MM Methods: ONIOM approach

2.5.1 Introduction

The electronic structure methods that have been described so far are adequate for studying a plethora of chemical systems and transformations, such as prediction of geometries and modelling reactions that involve breaking or formation of bonds between atoms. Their impracticability for large systems can often be balanced by truncating the system to a model of appropriate size that allows for application of the desired method. However, there are cases in which a truncated model cannot suitably represent the whole system, such as enzyme reactions,

the DNA molecule or when the presence of explicit solvent molecules is necessary. Although electronic structure methods on their own fail to offer a solution to such problems, their combination with methods of comparatively negligible computational cost, such as molecular mechanics (MM), has proven to be a useful alternative. Such hybrid methods are termed as quantum mechanical/molecular mechanical (QM/MM) methods. The idea of QM/MM methods was firstly introduced in a study of lysozyme by A.Warshel and M. Levitt in 1976[126]. In the following paragraphs a description of MM methods precedes a brief overview of the QM/MM schemes.

2.5.2 Molecular Mechanics (MM)

In MM or force field (FF) methods the molecular energy is calculated as a function of the nuclear positions and is based on a "ball and spring" model, while the electronic motion is completely neglected. As a result, painstaking quantum mechanical calculations are avoided and the dynamics or equilibriums of molecular systems can be approximated by Newtonian dynamics and Hooke's law, respectively.

The energy expression of a typical FF consists of five interatomic potential terms that contribute to the total energy (eq.2.5.1). Contributions from pairs of bonded atoms and interatomic angles are represented by harmonic potential terms and from torsional angles by a periodic term. For non-bonded and electrostatic interactions the Lennard-Jones and standard Coulomb potentials are used respectively and are applied both to atoms of different molecules and atoms of the same molecule that lie at least three bonds apart from each other.

$$E=\frac{1}{2}\sum_{bonds}k_i(l_i-l_{i,0})^2+\frac{1}{2}\sum_{angles}k_i(\theta-\theta_{i,0})^2+\frac{1}{2}\sum V_n(1+\cos(n\omega-\gamma))$$
$$+\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]+\frac{q_iq_j}{4\pi\varepsilon_0 r_{ij}}\right) \qquad (2.5.1)$$

2.5.3 Overview of QM/MM methods

In a typical QM/MM scheme, the studied system is appropriately divided into two parts (layers or levels): the one layer encloses the atoms that necessitate the use of a high-level method (QM, semi-empirical or DFT) and the second layer the atoms that are described by the low-level method (MM) and serve as a "backbone" of the system (fig. 2.5.1).

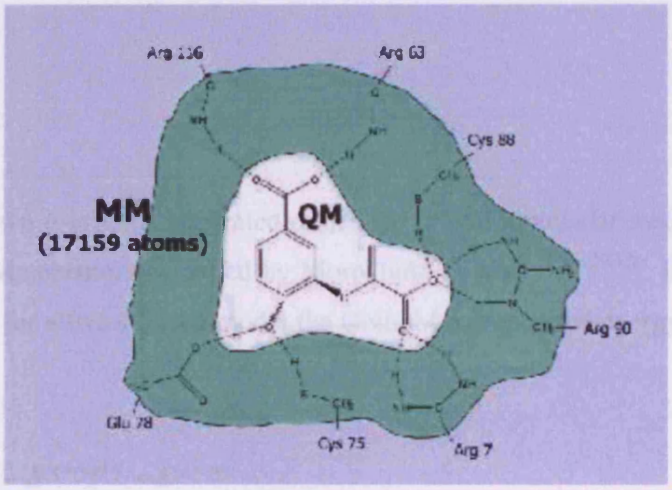


Figure 2.5.1: Partitioning into QM and MM regions.

The junction between the two layers is straightforward when there is no bond between atoms that are treated with different levels of theory. In the opposite scenario when the two layers are connected by a covalent bond, its breaking would leave the QM part with an unpaired electron. This can be balanced by addition of an appropriate "link" atom that maintains the polarity of the original bond. The link atom is invisible to the MM part and usually a hydrogen atom is used for this cause.

The Hamiltonian and energy of a QM/MM partitioned system are:

$$H_{total} = H_{QM} + H_{MM} + H_{QM/MM} \tag{2.5.2}$$

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM} \tag{2.5.3}$$

The first two terms of the energy are calculated with the standard QM and MM procedure. The third term deals with the important aspect of how the interaction between the different layers is handled in QM/MM schemes. The simplest approach is mechanical embedding, in which only the forces exerted on the QM atoms by the MM atoms are considered and vice versa and interaction between electronic parts is neglected. Alternatively, electronic embedding allows the QM atoms to be polarised by the presence of the MM atoms. Finally, there is the possibility to use polarisable FF that enables the polarisation of the MM region to be taken into account as well.

2.5.4 ONIOM

The ONIOM (our own n-layered integrated molecular orbital molecular mechanics) method is a generalised QM/MM scheme, developed by Morokuma *et al.*.[127,128,129,130]. It can combine more than two layers and for a two-layered model the ONIOM extrapolated energy has the subtractive form:

$$E_{ONIOM}^{high,real} = E^{high,model} - E^{low,model} + E^{low,real} \qquad (2.5.4)$$

The indices "high" and "low" refer to the level of the method that is used, while "model" and "real" refer to size of the system.

2.5.5 Practical considerations

For all the QM/MM calculations performed in this work the ONIOM method was used as implemented in the Gaussian03 software package[131] and the MM calculations were performed with the AMBER[132] force field. Assignment of atomic charges was made by Mulliken population analysis, charge equilibration method (QEq)[133] or by fitting to electrostatic potential (ESP).

2.6 Quantum Theory of Atoms-In-Molecules (QTAIM)

2.6.1 Introduction – QTAIM concepts

As seen from the previous discussion, all the necessary information that characterises a system, such as a molecule, is contained in the state function $\Psi(r)$ and properties of interest can be extracted by applying the appropriate operator. However, it is often desirable to be able to partition a given system into subsystems, i.e. define atoms within a molecule and derive atomic properties such as, for example, partial atomic charges. The quantum theory of atoms in molecules (QTAIM) is one of the various population analysis schemes that are used to tackle this problem and was developed by R.W. F. Bader.

The key concept within the context of QTAIM is the electron density of the system, which is defined as:

$$\rho(r;X) = N\int d\tau\psi^{*}(\chi;\mathrm{X})\psi(\chi;\mathrm{X}) \qquad (2.6.1)$$

where the integration is over all electronic spin coordinates and all but one Cartesian coordinates. The atoms within the molecule are defined by analysing the topology of the density; in other words, locating points in space where the first derivatives of the density vanish. Such points are termed as critical points $\rho(r_c)$ and are characterised as minima, maxima or saddle points, according to the curvature of the density on each point. The curvature is found by diagonalising the 3x3 Hessian matrix of the density and the resulting eigenvalues λ_i show the curvature in each direction. As a result, another crucial concept is the Laplacian of the density and it is equal to the trace of the Hessian matrix:

$$\nabla^2\rho(r) = \lambda_1 + \lambda_2 + \lambda_3 \qquad (2.6.2)$$

or

$$\nabla^2\rho(r) = \nabla\cdot\nabla\rho(r) = \partial^2\rho(r)/\partial x^2 + \partial^2\rho(r)/\partial y^2 + \partial^2\rho(r)/\partial z^2 \qquad (2.6.3)$$

2.6.2 Topology of $\rho(r)$ – Types of critical points

The critical points can be classified according to their rank, which is equal to the number of the non-zero eigenvalues of the Hessian matrix and their signature, which is simply the algebraic sum of the signs of the eigenvalues. In the case where all three eigenvalues are non-zero (rank=3), there are four possibilities of signature:

(3, -3): all eigenvalues are negative and the density shows a local maximum at r_c. These points indicate the nuclear positions.

(3, -1): two of the eigenvalues are negative. The density has a maximum in the plane defined by the two axes that correspond to the negative eigenvalues and a minimum along the third axis and the points are termed as Bond critical points (BCPs).

(3, +1): two eigenvalues are positive. The density has a minimum in the plane defined by the two axes that correspond to the positive eigenvalues and a maximum along the third axis. Such points are called Ring critical points (RCPs).

(3, +3): all eigenvalues are positive and the density has a local minimum at r_c, which is characterised as Cage critical point (CCP).

The following figure displays three different views of the density of the diborane molecule.

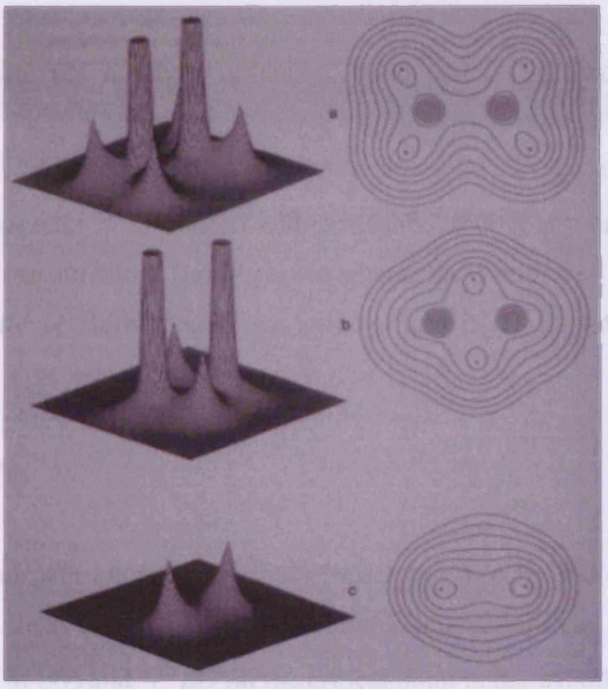


Figure 2.6.1: Different views of the diborane molecule's density as 3D relief (left) and Contour map (right). (Figure taken from ref. 115)

2.6.3 Gradient of the charge density – Bond paths

The use of the gradient allows for the representation of the charge density as a vector field and the tracing of its trajectory (or its gradient path). Two of the basic properties of such trajectories are : i) they are perpendicular to lines of constant density, because the gradient vector always points to the direction of greatest increase in the density, and ii) each trajectory originates or terminates at a point where $\nabla\rho(r)$ vanishes, i.e. at a critical point. For (3, -1) BCPs, the positive (λ_3) eigenvalue points along the bond direction.

The gradient paths between (3,-1) critical points and the (3,-3) nuclear attractors are called atomic interaction lines or bond paths. Depiction of all bond paths of a molecule at a given nuclear configuration defines its molecular graph. Moreover, the numbers of bond paths and of critical points that are present in a finite system are related by the Poincaré-Hopf relationship:

$$n-b+r-c=1 \qquad (2.6.4)$$

where n, b, r and c are the numbers of nuclei, bond paths, ring and cage critical points, respectively.

Finally, each nuclear attractor is associated with a neighbourhood, any originating trajectory of which is terminated at the attractor. The largest neighbourhood with this characteristic is called the basin of the attractor or atomic basin. An attractor and its basin define an atom in space, either bound or bound.

2.6.4 Properties of BCPs

The presence of a bond path between two nuclei does not directly indicate the presence of a covalent bond between those two nuclei. The interatomic interactions can be classified by taking into account the values of $\rho(r)$ and $\nabla^2\rho(r)$ at the critical point under examination.

When $\nabla^2\rho(r)<0$ and having a large value, and $\rho(r_c)$ is also large the interaction is characterised as covalent, polar or shared. When $\nabla^2\rho(r)>0$ and $\rho(r_c)$ is relatively low the interaction is a closed-shell one.

For example, according to Koch and Popelier's[134] criteria on hydrogen bonds, the electron density at the critical point of an H-bond is about an order of magnitude smaller than that of a covalent bond. The value of the density can vary from 0.0016 to 0.0205 a.u. depending on the nature of the H-bond.

2.6.5 Other properties

Other properties of interest within the context of QTAIM that have been used throughout this work are the quantities of kinetic energy densities, K(r) and G(r), defined as:

$$K(r)=-(\frac{\hbar^2}{4m})N\int d\tau'\{\Psi^*\nabla^2\Psi+\Psi\nabla^2\Psi^*\} \qquad (2.6.5)$$

$$G(r) = (\frac{\hbar^2}{4m})N\int d\tau' \nabla\Psi^* \cdot \nabla\Psi \qquad (2.6.6)$$

as well as the function of the Laplacian, L(r):

$$L(r) = (\frac{\hbar^2}{4m})\nabla^2\rho(r) \qquad (2.6.7)$$

Equations 2.6.4 – 2.6.6 are related through the relationship:

$$K(r)=G(r)+ L(r) \qquad (2.6.8)$$

Moreover, the electronic energy density is:

$$E(r)=G(r)+V(r)=-K(r) \qquad (2.6.9)$$

Where V(r) is called virial density and is equal to:

$$V(r) = -r \cdot \nabla \cdot \overleftrightarrow{\sigma} + \nabla(r \cdot \overleftrightarrow{\sigma}) \qquad (2.6.10)$$

2.6.6 Practical considerations

For the evaluation of the critical points and bond paths the AIM2000[115,135,136] software has been used, while the rest of the properties described above were evaluated using the Extreme2000 and AIMPAC[137] packages.

2.7 Polarised Continuum Model (PCM)

For a more accurate description of a system, it is usually desirable to take into account the effects of the environment of the system, as for example the effects of a solvent. Solvent molecules can

either be treated explicitly and added in the system as in the QM/MM methods or the solvent can be described as a continuous medium.

In models of the latter approach, the solvent is treated as a uniform polarisable medium of certain dielectric constant ε with a cavity in which the solute is placed.

The free energy of solvation is:

$$\Delta G_{solv}=\Delta G_{cavity}+\Delta G_{disp}+\Delta G_{elec} \quad (2.7.1)$$

where ΔG_{cavity} is the energetic cost for creating the cavity, ΔG_{disp} describes the solvent-solute van der Waals attraction/repulsion and ΔG_{elec} is the electrostatic stabilisation between the solvent and solute.

The interaction between the solute and the medium induces an electric field within the formed cavity, which can be taken into account as a perturbation of the Hamiltonian:

$$H_{total}=H_0+H_{RF} \quad (2.7.2)$$

where:

$$H_{RF}=-\hat{\mu}^{T}\frac{2(\varepsilon-1)}{(2\varepsilon+1)\alpha^{3}}\langle\Psi|\hat{\mu}|\Psi\rangle \quad (2.7.3)$$

H_0 is the unperturbed Hamiltonian of the system, μ is the dipole moment, α the radius of the cavity and ε the dielectric constant. This method is generally referred to as Self-Consistent Reaction Field (SCRF) method.

Within this general scheme lies the Polarisable Continuum Model (PCM)[138,139,140], which uses the atomic van der Waals radii of the solvent for the formation of the cavity.

2.8 Statistical Analysis

Individual sets of data and correlations between different sets of data have been analyzed and explored using simple techniques such as linear regression and multiple linear regression. Such techniques are accompanied by the appropriate statistical tools that aim at testing the accuracy and predictive ability of the derived models.

2.8.1 Single-descriptor linear regression

In the present work, regression refers to the specific case of fitting a line to a set of N data points (x_i,y_i) based on the standard least-squares fitting method. The term linear implies that the line of interest is a straight line of the form y=ax+b , i.e. the line of interest has linear dependence on the unknown parameters of the fit[141]. The parameters a, b are the regression coefficients, or, in this case, the slope and the intercept of the line of interest.

For a simple overview of the method, it is useful to define the quantity of the sum of squares:

$$SS_{xx} = \sum_{i=1}^{N}(x_i - \bar{x})^2 , SS_{yy} = \sum_{i=1}^{N}(y_i - \bar{y})^2 , SS_{xy} = \sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) \qquad (2.8.1)$$

The method proceeds by requiring that the quantity $D = \sum_{i=1}^{N}[y_i - y]^2$ be a minimum, i.e. satisfying the condition $\frac{\partial D}{\partial a} = 0$ and $\frac{\partial D}{\partial b} = 0$ for i=1,....N. The necessary algebraic manipulations lead to set of simultaneous equations for a and b. Using the equations (2.8.1) the solutions have the form:

$$b = \frac{SS_{xy}}{SS_{xx}}, a = \bar{y} - b\bar{x} \qquad (2.8.2)$$

2.8.2 Statistical tests: r^2 – s.d. – rms – t-ratio

Once the parameters a,b are determined and denoting the new y values that are calculated by the line equation as $y_{calc,i}$, the quality of the regression model is tested by the square correlation coefficient r^2:

$$r^2 = \frac{ESS}{SS_{yy}} = \frac{SS_{yy} - RSS}{SS_{yy}} = 1 - \frac{RSS}{SS_{yy}}, \quad 0 \le r^2 \le 1 \tag{2.8.3}$$

Where ESS, RSS are the explained and residual sums of squares, respectively:

$$ESS = \sum_{i=1}^{N} (y_{calc,i} - \bar{y})^2 \, , RSS = \sum_{i=1}^{N} (y_i - y_{calc,i})^2 \tag{2.8.4}$$

The correlation coefficient r^2 yields a fraction of the total variance that is explained by the model. Good fits have an r^2 close to 1, while r^2 close to zero is indicative of poor fit. However, because r^2 increases upon increase of the number of variables, the quality of fit is better tested in combination with the Standard Deviation (sd) and the Root Mean Square Error (RMSE):

$$sd = \sqrt{\frac{\sum_{i=1}^{N} (y_i - y_{calc,i})^2}{N-1}} \, , RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - y_{calc,i})^2}{N}} \tag{2.8.5}$$

It is apparent from equations 2.8.5 that for large datasets sd and RMSE are equivalent. Additionally, the sd divided by $\sqrt{N}$ yields the standard error.

Finally, the t-statistic provides information on the significance of the variable that is used and is equal to the coefficient of that variable divided by the standard error. If the used variable is significant the resulting t value is larger than the standard t value that can be found in tables.

2.8.3 Multiple Linear Regression

In multiple linear regression a surface $y=a_0+a_1x_1+a_2x_2+...$ is fitted to the data points analogously to linear regression. Again, a first idea of the quality of the fit is provided by the multiple correlation coefficient R^2. Alternatively, the adjusted R^2 (R^2_{adj}) can be used, which is equal to:

$$R^2_{adj} = 1 - \frac{SS_{err}}{SS_{tot}} \cdot \frac{df_{tot}}{df_{err}} \qquad (2.8.6)$$

where df_{tot} are the total degrees of freedom (number of data points) and df_{err} are the degrees of freedom minus the number of parameters used in the model. Models of good quality have similar values of R^2 and R^2_{adj}, while for poor models R^2_{adj} is less than R^2 or even negative. The significance of a model can be examined using Fischer's F-statistic (F) and as in the case of t-statistic the obtained value is compared to tabulated values for a certain level of confidence. Finally, another useful technique for testing the quality of a model is the cross-validated R^2 (R^2_{cv} or Q^2):

$$R^2_{CV} = 1 - \frac{\sum_{i=1}^{N}(y_{pred,i} - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \qquad (2.8.7)$$

Where ($y_{pred,i}$) refers to data not used in the derivation of the model.

Application of Statistical tools and techniques of this work was carried out using the JMP[142] software.

2.9 Intermolecular Forces

2.9.1 Introduction

Apart from the typical chemical bonds, the non-bonded forces that occur between atoms or molecules are crucial for the geometry of molecules themselves or for determining the properties

of matter in general. Such interactions act through space and can be either attractive or repulsive. Since they depend on the distance between the species they act on, one way to systematically study them is the classification as long- and short range interactions. However, the ultimate origin of most is of electrostatic nature.

2.9.2 Long range interactions

Long range interactions have an inverse power dependence ($\sim r^{-n}$) on the distance. The main representatives of this class are the electrostatic, induction and dispersion interactions. The first obey the classical coulomb law between charged particles, the sign of which determines an attractive or repulsive behaviour. Induction interactions on the other hand are always attractive and arise by the mutual distortion of the molecular charge distribution that atoms or molecules exert on each other. Finally, dispersion interactions have their origin in the correlated motion of electrons, as mentioned in section 2.3.4, i.e. their origin is quantum mechanical.

2.9.3 Short range interactions

In this category falls the exchange interaction the occurrence of which is obvious within the Hartree-Fock formalism. Usually exchange interaction is termed as exchange-repulsion to include the strong repulsion that occurs at very short distances. Finally, another type is donor-acceptor interactions, also known as charge transfer and which are of attractive nature. All short range interactions exhibit an exponential decrease ($\sim e^{-ar}$) with increasing distance.

2.10 Structural Analysis of DNA molecules (Curves)

The analysis and description of DNA helices can be made using rotational and translational parameters determined by the relative positions of the nucleic acid bases and the choice the helical axis. A workshop on the DNA curvature and bending was held in 1998 with the aim of recommending definitions and nomenclature of these parameters that can be widely adopted by research groups[143].

The definitions of the translations and rotations are displayed in figures 2.10.1 and 2.10.2, respectively. Depending on whether a global or local axis is used, parameters that are related to successive base pairs are also termed as global or local, respectively.

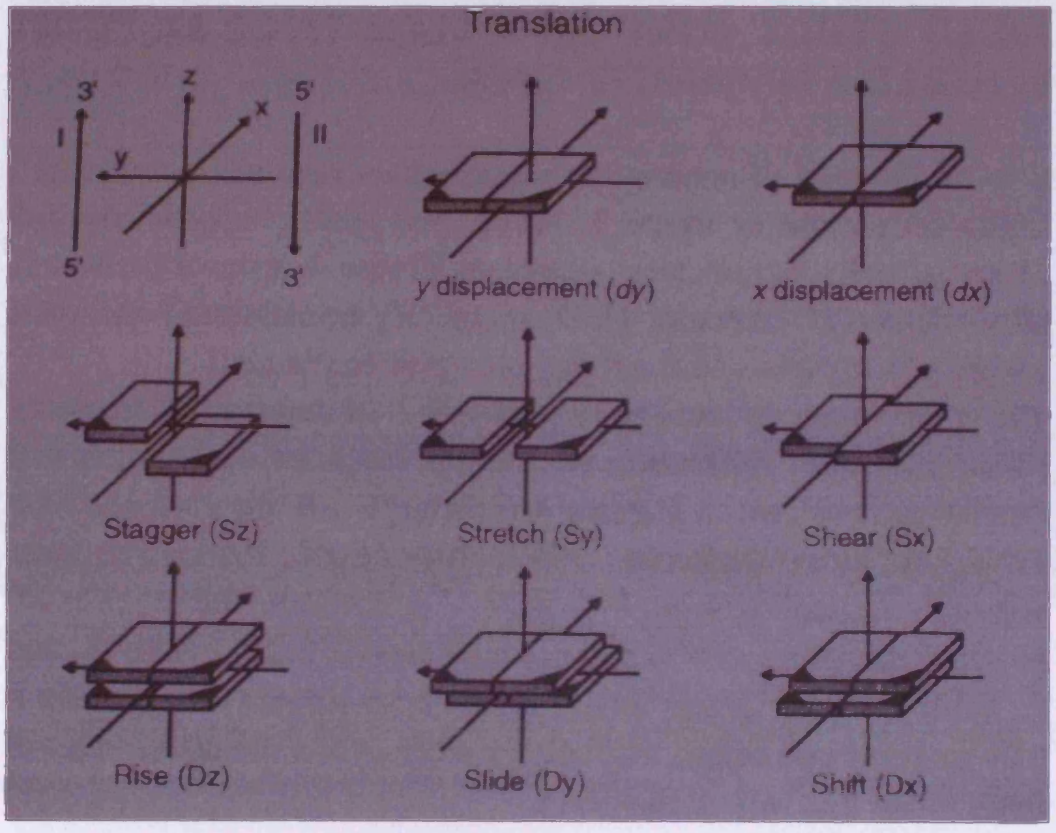


Figure 2.10.1: Translational parameters of DNA bases (Fig. taken from ref. 118).

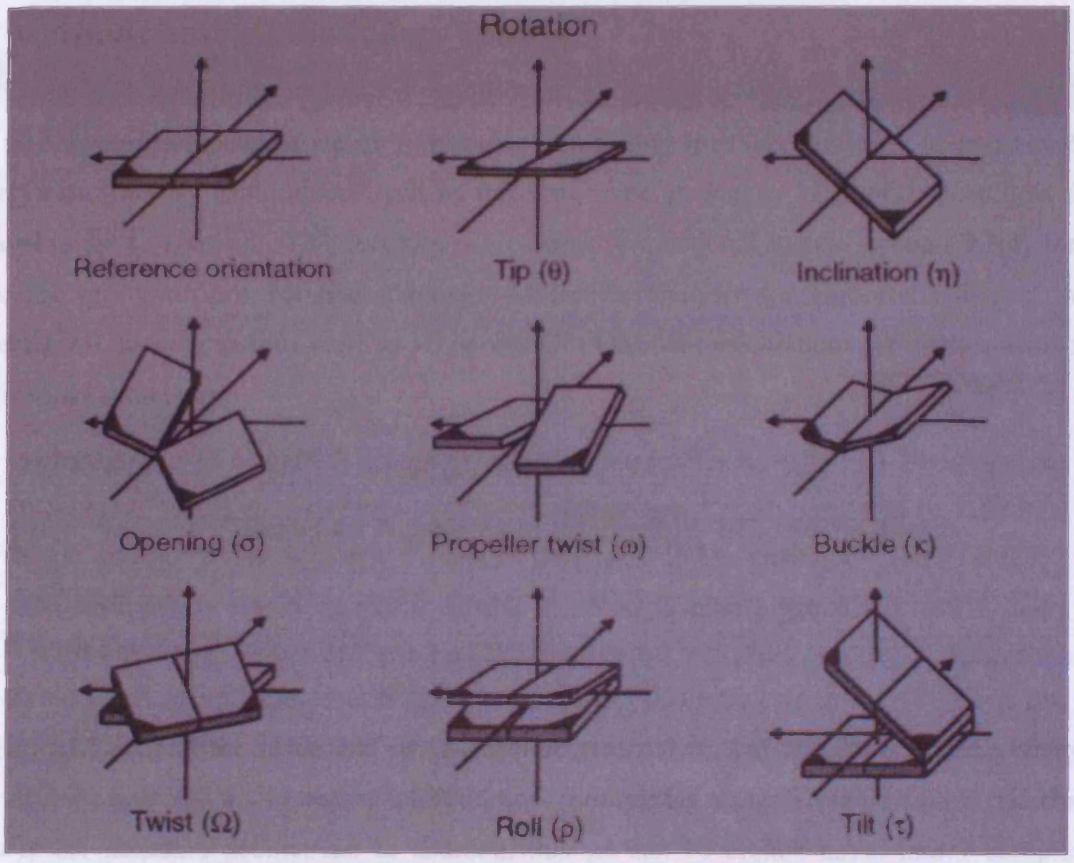


Figure 2.10.2: Rotational parameters of DNA bases (Fig. taken from ref. 118).

The analysis of structures of this work is based on the above definitions and the values of the parameters have been evaluated using the Curves5.1 software[144,145].

3. Methods and methodology testing

In this chapter, we concentrate on developing and testing methods for study of non-covalent interactions, mainly using model systems that have experimental or high-level theoretical data available for comparison. This includes use of density functional theory (section 2.3.4) for π-stacking and hydrogen bonding, Atoms-in-Molecules analysis for characterisation of such interactions, and the various ways to set up ONIOM QM/MM calculations for nucleic acids and their metal complexes.

3.1 Performance of BHandH for Non-covalent Interactions: S22 and JSCH Sets

3.1.1 Basis Set

Table 3.1.1 reports the mean and mean absolute errors (ME and MAE) of binding energies across the whole S22 set[65] (section 1.4.3) of model complexes, using BHandH with various basis sets at the literature geometries. In addition, this set can be broken down into 7 complexes dominated by hydrogen bonding (HB), 8 bound largely by dispersion forces (Disp), and 7 where both types of interaction are important (Mix), based on Symmetry Adapted Perturbation Theory (SAPT) calculations. These data show that the average absolute error across this varied set of complexes varies only slightly with basis set, in a range of 2 to 2.5 kcal mol^{-1}. Inclusion of diffuse functions improves predictions for HB and Disp complexes, but not for Mix, while *f*-type basis functions give slightly improved predictions for all three classes. The average error is uniformly positive, indicating that BHandH tends to over-bind the complexes.

Errors in the binding energies of hydrogen-bonded complexes are large for BHandH, on average around 5 kcal mol^{-1}, but are much smaller for dispersion bound and mixed complexes. To put these values in context, MP2 binding energies extrapolated to the basis set limit (taken from ref. 65) gives MAE = 0.80 (all), 0.15 (HB), 1.51 (Disp), and 0.64 (Mix). It should be noted that the MP2/CBS estimates were recently updated by Marchetti and Werner[146], using explicitly

correlated methods and accounting for errors related to the uracil dimer that were present in the original paper (see ref. 146 for further discussion). Thus, BHandH provides binding energies of much poorer quality than MP2 for H-bonded complexes, but performs rather better than MP2 for dispersion bound complexes, where MP2 is known to over-bind[75,147], while the two methods give comparable performance for mixed complexes. In general, the smallest basis set 6-31G(d) gives significantly larger errors than all others, in accord with the first paper of our group on BHandH, in which at least one diffuse function on heavy atoms was required for reasonable accuracy.[97] Mean counterpoise correction are small, varying from 1.94 kcal mol^{-1} for 6-31G(d) to 0.49 kcal mol^{-1} for 6-311++G(2dfd,2p).

Table 3.1.1: Mean and mean absolute errors (ME and MAE) for S22 binding energies at literature geometry, using BHandH with various basis sets (kcal mol^{-1}).

	ME	MAE	MAE HB	MAE Disp	MAE Mix	Mean BSSE
6-31G(d)	+1.73	2.40	5.22	1.08	1.10	1.94
6-31+G(d,p)	+1.90	2.12	4.89	0.60	1.10	0.59
6-311++G(d,p)	+1.93	2.10	4.81	0.54	1.16	0.70
6-311++G(2df,2p)	+1.81	1.99	4.58	0.55	1.05	0.49
aug-cc-pVDZ	+1.79	2.00	4.65	0.56	1.00	0.73

3.1.2 Functional

In our group's initial study of BHandH[97] it was concluded that the good performance for π-stacking stems from the use of local exchange, since the purely local density approximation (LDA) also gave reasonable results for the benzene dimer. This suggests that the performance of BHandH-like hybrid functionals may be tuned by altering the amounts of exact and Slater exchange employed. Table 3.1.2 reports errors for various combinations of exact and Slater exchange in the functional, keeping the LYP correlation functional throughout. This analysis reveals that increasing the amount of LDA exchange leads to worse predictions for all three

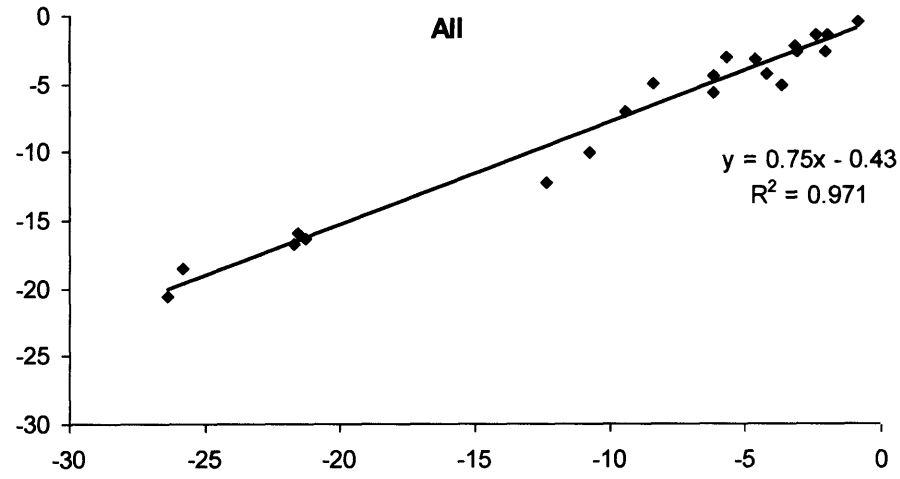
classes of interaction. Increasing the amount of HF exchange improves performance for H-bonded and mixed complexes, but degrades that for dispersion bound ones. Indeed, Table 3.1.2 makes it clear that BHandH as originally defined includes close to the optimal amount of HF and LDA exchange for the S22 set. The overall error is reduced slightly with increased fractions of HF exchange, but the gains are small.

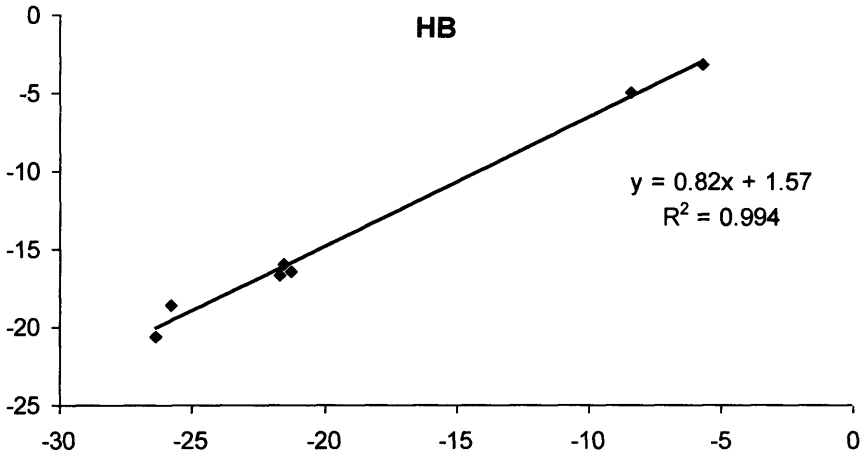
Table 3.1.2: Effect of variation of the amount of HF exchange contained in the functional for S22 binding energies, with 6-31+G(d,p) (kcal mol^{-1}).

%HF, LDA	ME	MAE	MAE HB[a]	MAE Disp[b]	MAE Mix[c]
30, 70	+3.06	3.06	6.30	1.31	1.83
40, 60	+2.47	2.52	5.57	0.79	1.46
50, 50	+1.90	2.12	4.89	0.60	1.10
60, 40	+1.36	1.93	4.25	0.89	0.80
70, 30	+0.84	1.88	3.65	1.36	0.72
80, 20	+0.35	1.86	3.09	1.86	0.64
90, 10	-0.11	1.89	2.56	2.40	0.63

[a]H-bonded, [b]Dispersion-dominated and [c]Mixed complexes.

The fact that BHandH consistently over-binds non-covalently bound complexes in general (just 2 out of 22 complexes are predicted to be under-bound) suggests that prediction of binding energy made using this functional can be simply scaled by a multiplicative factor to improve performance. As shown in Figure 3.1.1, this is indeed the case: the line of best fit is some way from the ideal $y = x$ line, but fits the data well, with $R^2 = 0.971$ and MAE = 0.83 kcal mol^{-1}. Thus, this scaling process reduces the overall error by more than half. It is also possible to individually scale the different classes of complex. The best fit for dispersion bound complexes is very close to $y = x$, while for H-bonded and mixed complexes this is closer to that shown in Figure 3.1.1. These fits give mean absolute errors of 0.60, 0.43, and 0.36 kcal mol^{-1} for dispersion, H-bonded, and mixed complexes, respectively.


All
y = 0.75x - 0.43
R2 = 0.971
0
-5
-10
-15
-20
-25
-30
-30
-25
-20
-15
-10
-5
0


HB
y = 0.82x + 1.57
R2 = 0.994
0
-5
-10
-15
-20
-25
-30
-25
-20
-15
-10
-5
0

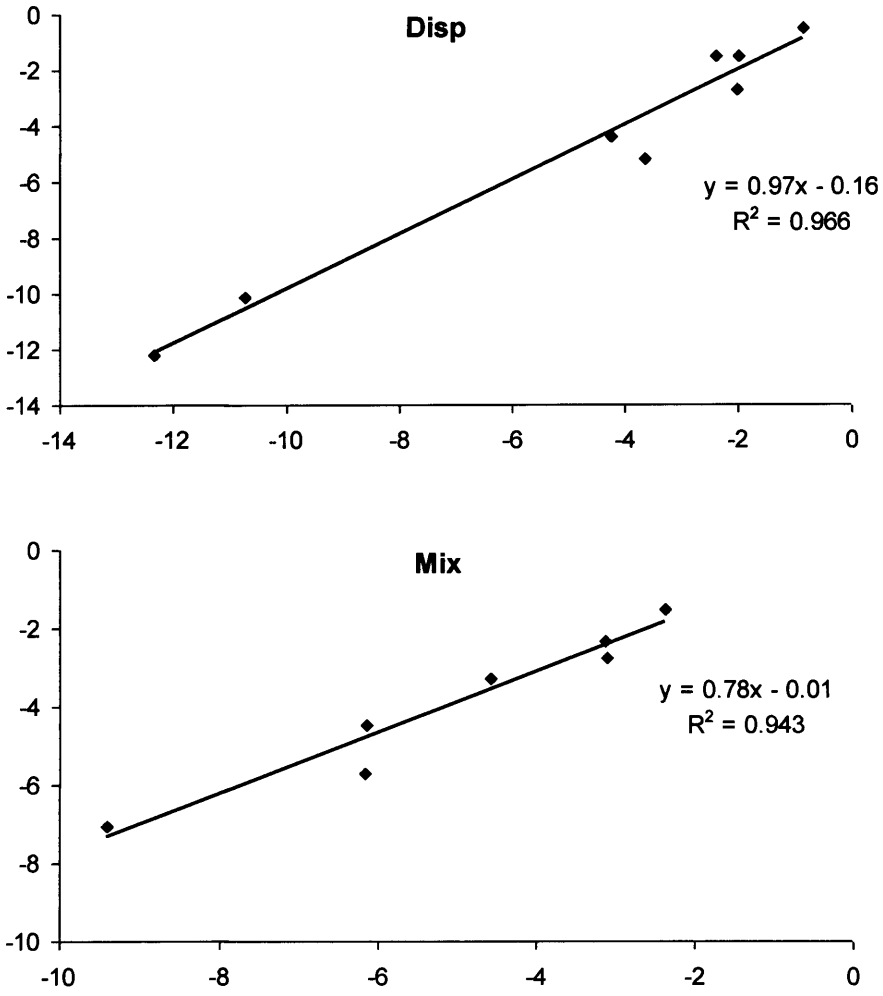


Figure 3.1.1 Plot of BHandH/6-31+G(d,p) (abscissa) *vs.* literature CBS(T) (ordinate) binding energies (kcal mol^{-1}) for the S22 set of model complexes.

3.1.3 QTAIM analysis

In addition to direct calculation of interaction *via* the supermolecular approach, the previous paper on BHandH[97] set out a surprisingly accurate correlation between interaction energy and the sum of the electron density at all BCP's found in the electron density of stacked complexes. This relation is summarised in equation (3.1.1)

$$\Delta E_{\pi} = -173.18 * \sum p_{\pi} - 0.02 \qquad (3.1.1)$$

where ΔE_{π} is the interaction energy, in kcal mol^{-1}, and $\sum p_{\pi}$ is the sum of electron density at all bond critical points located between stacked molecules, in atomic units. This relation gave an r^2 value of 0.950 and a root mean square error of just 0.48 kcal mol^{-1}. Similar relations have long been established for hydrogen bonding, with the electron density at the H-bond BCP and/or the change in density at the donor BCP particularly widely used[148],[149].

However, the direct analogue of equation (3.1.1) would not be particularly useful, given the errors noted for H-bonds in Table 3.1.1. Instead, we have taken an alternative approach, whereby BHandH electron density data is plotted against MP2 interaction energy. This interaction energy was evaluated for 26 small and medium sized H-bonded complexes, using counterpoise corrected MP2/aug-cc-pVTZ methods. This gives rise to the equation shown as equation (3.1.2):

$$\Delta E_{HB} = -162.55 * \rho_{\mathrm{HB}} + 0.44 \qquad (3.1.2)$$

where units are the same as for equation (3.1.1). This correlation has r^2 = 0.98 and RMSE = 1.07 kcal mol^{-1}. In this way, we aim to reproduce MP2 quality hydrogen bond energies from BHandH calculated electron densities, by identifying all intermolecular BCP's and applying the relevant linear correlation.

Data for application of these relations is reported in Table 3.1.3: at the literature geometry, agreement for dispersion bound complexes is reasonable, but less good for H-bonded and mixed complexes. Performance seems particularly poor for H-bonded complexes containing multiple H-bonds (*i.e.* all except ammonia and water dimers), which can be in error by almost 10 kcal mol^{-1}. Equation (3.1.2) was trained on simple complexes containing just one H-bond, and clearly struggles to describe these multiple H-bonded complexes. Table 3.1.3 also contains QTAIM analysis at the BHandH fully optimised geometry (see below for details), for which the performance is rather better, with MAE = 1.80 kcal mol^{-1}, comparable to direct evaluation of interaction energy. This error is rather constant across different types of complex (1.63, 2.07 and 1.74 kcal mol^{-1}).

Table 3.1.3: Estimated ΔE (kcal mol^{-1}) using equations (3.1.1) and (3.1.2) for the S22 set.

	CBS(T)	BHandH geometry	Hobza geometry
Ammonia dimer	-3.17	-2.32	-1.38
Water dimer	-5.02	-5.14	-3.42
Formic acid dimer	-18.61	-19.37	-14.26
Formamide dimer	-15.96	-12.46	-9.50
Uracil dimer	-20.65	-15.05	-11.14
Pyridoxine-aminopyridine	-16.71	-14.13	-10.14
Adenine-thymine WC	-16.37	-15.29	-10.28
Methane dimer	-0.53	-0.95	-0.48
Ethene dimer	-1.51	-2.60	-1.06
Benzene-methane	-1.50	-3.86	-2.42
Benzene dimer	-2.73	-2.44	-2.30
Pyrazine dimer	-4.42	-3.87	-3.40
Uracil dimer	-10.12	-7.13	-7.55
Indole-benzene	-5.22	-2.80	-4.58
Adenine-thymine	-12.23	-9.32	-10.66
Ethene-ethyne	-1.53	-1.11	-0.65
Benzene-water	-3.28	-1.91	-0.85
Benzene-ammonia	-2.35	-1.07	-0.61
Benzene-HCN	-4.46	-1.19	-0.83
Benzene dimer	-2.74	-1.75	-1.38
Indole-benzene T	-5.73	-1.79	-1.25
Phenol dimer	-7.05	-7.93	-4.86

The similarity between the slopes of equations (3.1.1) and (3.1.2) is striking: indeed, within statistical error they are effectively identical. This suggests that the electron density properties of non-covalent interactions are similar, no matter the type of interaction involved. To the best of our knowledge, this possibility has not been noted before. Such a property would be highly desirable, since separation of critical points into H-bond and stacking classes is time consuming and can be ambiguous. We therefore attempted to combine models of this form, taking the S22 literature binding energy as a reference, and testing correlations similar to equations (3.1.1), (3.1.2) for these. Taking BHandH/6-31+G(d,p) calculated electron densities, a fit with $R^2 = 0.926$ and RMSE = 1.78 kcal mol^{-1} was obtained, while the larger 6-311++G(d,p) basis gave marginal improvement to $R^2 = 0.931$ and RMSE = 1.72 kcal mol^{-1}. The errors involved in fitting to the entire S22 dataset are therefore rather worse than for the individual sets used to train equations (3.1.1) and (3.1.2), but are comparable to those found for direct evaluation of interaction energy. As a further test, MP2/6-311++G(d,p) BCP properties were calculated, giving $R^2 = 0.925$ and RMSE = 1.78 kcal mol^{-1}. Thus, it seems clear that neither method nor basis is crucial to calculated electron density properties, and that BHandH/6-31+G(d,p) is sufficient for such analysis.

To further investigate these correlations, we employed the H-bonded nucleic acid pairs of the rest of the JSCH-2005 database, which includes both experimental and optimised geometries. A trend was evident for this data, although the fit quality is not as good as for the S22 data (R^2 = 0.806, MAE = 2.84 kcal mol^{-1}). Including only optimised geometries improved results, lowering the mean unsigned error by 0.4 kcal mol^{-1}. Closer examination revealed two distinct trends: one for nucleic acid bases paired by two H-bonds, and another for base pairs with three H-bonds. Separating these classes resulted in better correlation for the doubly H-bonded pairs (R^2 = 0.901, MAE = 0.86 kcal mol^{-1}), although this was not the case for the base pairs with three H-bonds for which a rather poor fit was observed. However, there is an indication that donor (D) –acceptor (A) patterns within the complexes with three H-bonds play a significant role in the resulting trends. Distinguishing the complexes in those following a DDA pattern and those having an ADA pattern resulted in separate fits of similar quality, but with significantly different slopes. From these results it seems that synergy in H-bond effects might play a role in the resulting

binding energies, and hence that correlations between BCP properties and binding energies are family dependent.

3.1.4 Geometry Optimisations

Two optimisation strategies have been employed: firstly, a rigid monomer approach, in which both monomers are fixed at the geometry given by Hobza *et al.*, and only the 6 intermolecular degrees of freedom allowed to vary. Subsequently, a fully flexible optimisation including all intra- and intermolecular parameters was performed. The first set of data, reported in Table 3.1.4, indicates that in 14 of 22 cases, the optimised structure has an RMSD within 0.1 Å of the literature geometry, and in a further 7 within 0.2 Å. Despite the clear overestimation of the strength of hydrogen bonding, it appears that BHandH gives a reasonable estimation of the geometry of H-bonded complexes, with an average RMSD of 0.06 Å for these 6 complexes. Only in one case, the phenol dimer, is a rather large change from literature data observed. This stems from a change in relative orientation of the rings, as shown in Figure 3.1.2, in which the interplanar angle changes from 60.5° to 78.2°.

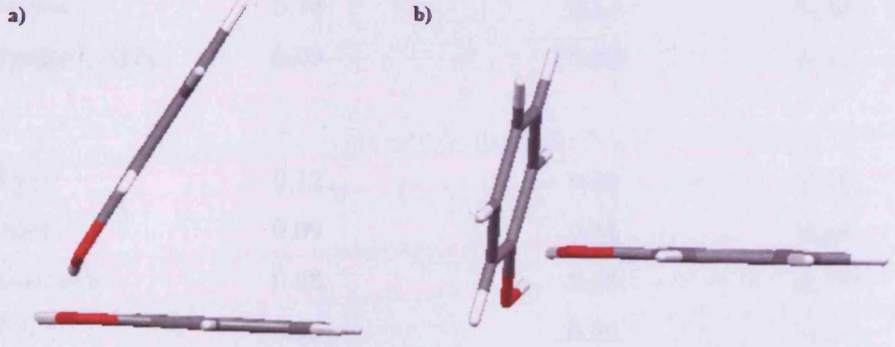


Figure 3.1.2: Phenol dimer geometry from a) CP-corrected MP2/cc-pVTZ (taken from ref. 65), and b) BHandH/6-311++G(d,p) optimisation.

Table 3.1.4: Root mean square deviation (RMSD) from literature geometry and energy difference (|ΔE|) between literature and BHandH optimised geometries for the S22 set.

	Rigid monomer RMSD (Å)	Full opt RMSD (Å)	|ΔE| (kcal mol^{-1})
Ammonia dimer	0.11	0.11	0.58
Water dimer	0.08	0.07	0.46
Formic acid dimer	0.05	0.06	0.97
Formamide dimer	0.05	0.06	0.36
Uracil dimer	0.05	0.07	0.07
Pyridoxine-aminopyridine	0.05	0.20	0.76
Adenine-thymine WC	0.07	0.11	0.15
Methane dimer	0.20	0.20	0.35
Ethene dimer	0.13	0.13	0.43
Benzene-methane	0.07	0.07	0.10
Benzene dimer	0.08	0.04	0.07
Pyrazine dimer	0.07	0.05	0.07
Uracil dimer	0.11	0.13	0.03
Indole-benzene	0.14	0.23	0.26
Adenine-thymine stack	0.09	0.09	0.31
Ethene-ethyne	0.12	0.10	0.26
Benzene-water	0.09	0.26	0.54
Benzene-ammonia	0.08	0.10	0.24
Benzene-HCN	0.06	0.06	0.22
Benzene dimer	0.04	0.06	0.09
Indole-benzene T	0.05	0.09	0.23
Phenol dimer	0.67	0.59	0.84
Average	0.11	0.13	0.34

RMSD data for the full optimisation of all geometrical parameters are also included in Table 3.1.4. In most cases, these are identical to or slightly larger than the rigid monomer values. One case stands out from this trend, namely benzene-water, in which the orientation of the water molecule alters to give two O—H···π H-bonds rather than one, as shown in Figure 3.1.3. As in the frozen monomer case, the phenol dimer, shows a rather larger change of almost 0.6Å, which represents a change in the relative orientation of the aromatic rings and formation of a C—H···π interaction that is not present in the literature geometry.

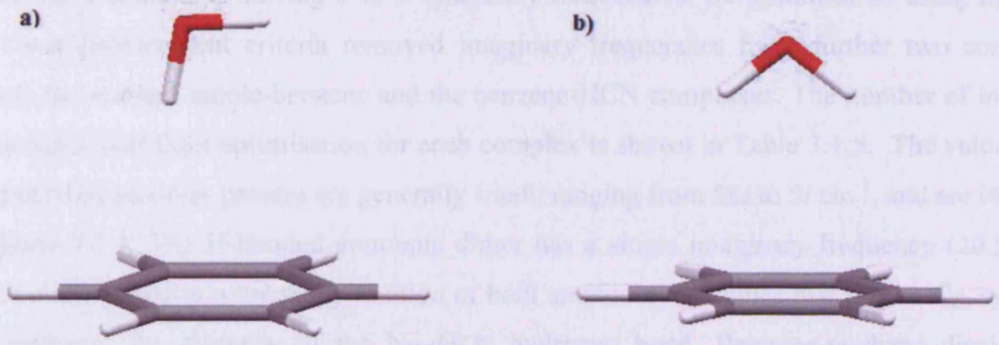


Figure 3.1.3: Benzene-water geometry from a) literature, and b) BHandH/6-311++G(d,p) optimisation.

To check the energetic consequences of these geometry changes, DF-LMP2 calculations of binding energy were carried out at the literature and fully optimised geometries. Non-covalent interactions are generally associated with rather flat potential energy surfaces, such that relatively large changes in geometry can result in very small energy changes. Table 3.1.4 confirms that this is indeed the case, with an average difference in binding energy of 0.34 kcal mol^{-1} between literature and optimised geometries. In general, energy changes are larger for H-bonded complexes than for dispersion bound or mixed complexes, despite the smaller changes in geometry observed. The largest changes are seen for the dimers of formic acid, which is strongly

H-bonded, and phenol, which undergoes the largest change in geometry during full optimisation. Even these changes are less than 1 kcal mol^{-1}, and on average the change in energy is 0.48 kcal mol^{-1} for all 6 H-bonded complexes, compared with 0.20 and 0.35 kcal mol^{-1} for the dispersion bound and mixed complexes, respectively.

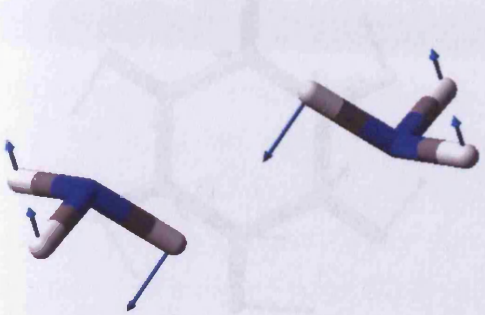
3.1.5 Frequencies

Initial frequency calculations were performed on BHandH/6-311++G(d,p) optimised geometries using the same method and basis set as for the optimisation, with the "ultrafine" grid option for numerical integration selected. These showed that 14 complexes are true minima at this level, the remaining 8 complexes having 1 or 2 imaginary frequencies. Re-optimisation using tight SCF and force/displacement criteria removed imaginary frequencies for a further two complexes, namely the stacked indole-benzene and the benzene-HCN complexes. The number of imaginary frequencies after tight optimisation for each complex is shown in Table 3.1.5. The values of the imaginary frequencies present are generally small, ranging from $58i$ to $5i$ cm^{-1}, and are illustrated in Figure 3.1.4. The H-bonded ammonia dimer has a single imaginary frequency ($20.5i$ cm^{-1}), which corresponds to a "rocking" motion of both ammonia molecules that breaks C_{2h} symmetry and increases the linearity of the N—H···N hydrogen bond. Benzene-methane displays two degenerate imaginary frequencies at $26.4i$ cm^{-1}, both of which correspond to rotation of methane relative to benzene, moving C—H off the C_6 axis of benzene to remove the C_3 symmetry of the complex. The stacked uracil dimer has a single imaginary frequency ($15.9i$ cm^{-1}), consisting of a "rocking" motion of each uracil to break the C_2 symmetry. Benzene-water and benzene-ammonia both have one imaginary frequency ($11.3i$ cm^{-1} and $57.8i$ cm^{-1}, respectively). In the former, this is a rotation of H_2O away from C_s symmetry, while in the latter is similar to benzene-methane, moving N—H off the C_6 axis of benzene. The T-shaped benzene dimer displays two imaginary frequencies ($28.6i$ and $5.1i$ cm^{-1}), both of which are rotations of one benzene relative to the other, moving the donor C—H off the C_2 axis of the C_{2v} complex.

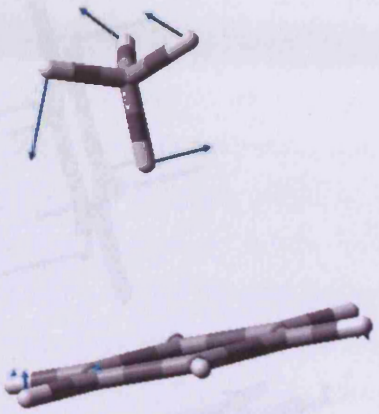
Table 3.1.5: Harmonic frequency analysis for the S22 set of model complexes.

	#Imag Freq	ΔZPVE [a] (kcal mol^{-1})
Ammonia dimer	1	1.77 (1.70)
Water dimer	0	2.51 (2.37)
Formic acid dimer	0	1.67
Formamide dimer	0	2.27
Uracil dimer	0	1.06
2-Pyridoxine...2-aminopyridine	0	0.89
Adenine-thymine WC	0	0.92
Methane dimer	0	0.61 (1.16)
Ethene dimer	0	0.98
Benzene-methane	2	0.54
Benzene dimer stack	0	0.18
Pyrazine dimer	0	0.52
Uracil dimer stack	1	0.97
Indole-benzene stack	0	0.33
Adenine-thymine stack	0	1.12
Ethene-ethyne	0	0.60 (0.71)
Benzene-water	1	0.74 (1.06)
Benzene-ammonia	1	0.70
Benzene-HCN	0	0.60
Benzene dimer T	2	0.34
Indole-benzene T	0	0.43
Phenol dimer	0	1.54

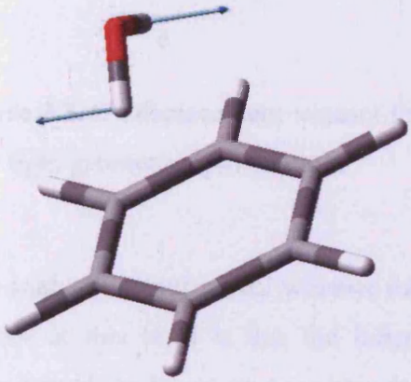
[a]BHandH/6-311++G(d,p) values at minima following re-optimisation, DF-LMP2/aug-cc-pVTZ values in parenthesis where available.



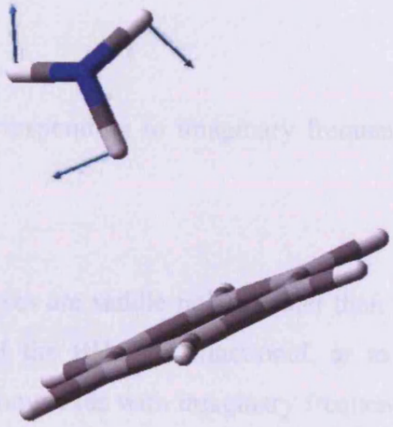
a



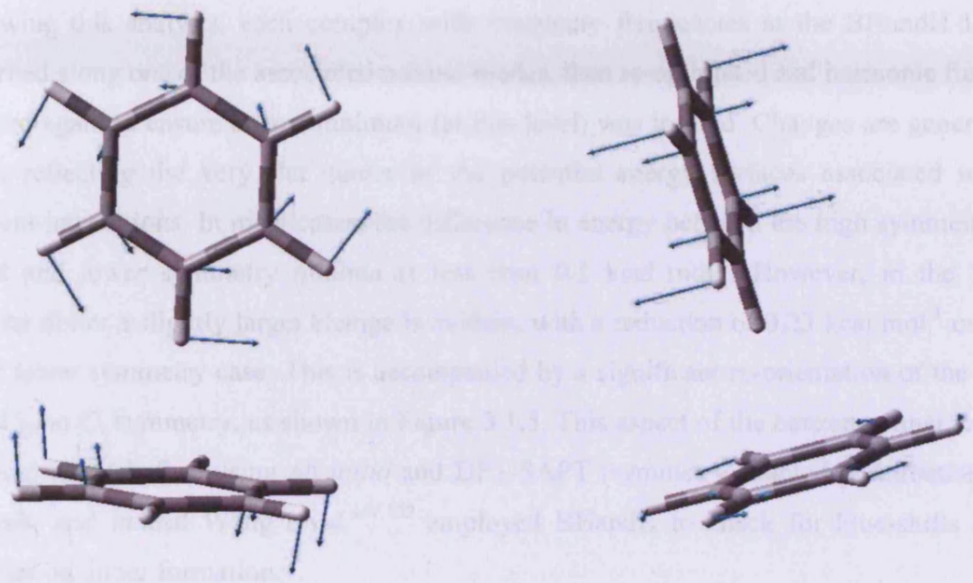
b



c



d



e f

Figure 3.1.4: Displacement vectors for normal modes corresponding to imaginary frequencies after tight geometry optimisation.

This analysis cannot reveal whether the fact that 6 complexes are saddle points rather than true minima at this level is due the inherent shortcomings of the BHandH functional, or to the symmetry of the literature geometry. It is notable that all complexes with imaginary frequencies do have some symmetry, and that in ref. 65 this is assumed with no further comment or test. Harmonic frequency calculation with suitably high-level *ab initio* methods is a difficult task, but the speed of DF-LMP2 makes this feasible, although the lack of analytical second derivatives means that numerical differentiation is required, limiting this to smaller complexes at this stage[150]. In the case of the ammonia dimer, DF-LMP2/aug-cc-pVTZ optimisation and frequency confirms that the literature C_{2h} geometry is appropriate and hence that BHandH is in error, presumably due to its tendency to overestimate the strength of H-bonding. However, benzene-water is a first order saddle point with DF-LMP2 as well as with BHandH in the literature C_s geometry, with a similar normal mode of imaginary frequency corresponding to rotation of water off the symmetry plane of benzene.

Following this analysis, each complex with imaginary frequencies at the BHandH level was perturbed along one of the associated normal modes, then re-optimised and harmonic frequencies checked again to ensure a true minimum (at this level) was located. Changes are generally very small, reflecting the very flat nature of the potential energy surfaces associated with non-covalent interactions. In most cases, the difference in energy between the high symmetry saddle points and lower symmetry minima is less than 0.1 kcal mol^{-1}. However, in the T-shaped benzene dimer a slightly larger change is evident, with a reduction of 0.23 kcal mol^{-1} on moving to the lower symmetry case. This is accompanied by a significant re-orientation of the complex from C_{2v} to C_s symmetry, as shown in Figure 3.1.5. This aspect of the benzene dimer's structure has been noted before using *ab initio* and DFT-SAPT (symmetry adapted perturbation theory) methods, and indeed Wang *et al.*[151,152] employed BHandH to check for blue-shifts in C—H stretches on dimer formation.

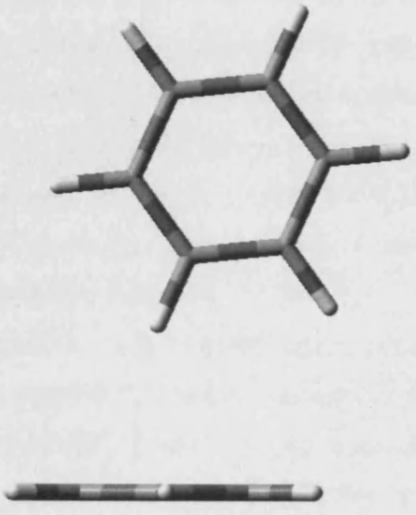


Figure 3.1.5: Re-optimised geometry of "T-shaped" benzene dimer.

Table 3.1.5 also reports the calculated change in zero-point vibrational energy (ΔZPVE). For those compounds initially with imaginary frequencies, those data were calculated at the re-optimised geometry. Those data show that ΔZPVE is small for the dispersion bound complexes, values rising above 1 kcal mol^{-1} only for stacked adenine-thymine, whereas ΔZPVE values are rather larger for the hydrogen bonded complexes. This is to be expected, since the force constants for displacement of hydrogen bonds are generally larger than for displacement of stacking interactions. Within the mixed complexes values are small except for the phenol dimer, which contains a relatively strong O—H···O H-bond. For selected complexes, the zero-point vibrational energy corrections were also calculated using the DF-LMP2 method. These are reported in parenthesis in Table 3.1.5, and where comparison is possible the agreement is generally excellent.

3.1.6 Toluene dimer

As an independent test of the conclusions drawn on the basis of Hobza's S22 dataset, we checked the performance of BHandH and related methods for the toluene dimer. Structures and binding energies for this complex in three distinct orientations, termed parallel, antiparallel, and cross, were reported by Wright[153], using CCSD(T) methods. At the geometries reported in ref. 153, BHandH overestimates binding, with an average absolute error of 1.45 kcal mol^{-1}. Scaling these binding energies following Figure 3.1.1 leads to rather better predictions (MAE = 0.73 kcal mol^{-1}). Importantly, BHandH reproduces the relative energy of the three forms, indicating that "antiparallel" is the most stable form, following by the "cross" form, with the "parallel" orientation the least stable. Optimisation using BHandH again leads to small changes in structure (RMSD = 0.13, 0.21 and 0.14Å, respectively) and in energy (0.21, 0.12, and 0.19 kcal mol^{-1}, respectively using DF-LMP2). QTAIM results are less accurate, underestimating binding throughout and, at the literature geometry, ordering the stability of complexes incorrectly. The above results are shown in the following table (table 3.1.6).

Table 3.1.6: Binding energy for three orientations of toluene dimer (kcal mol^{-1})

	CCSD(T)	BHandH unscaled	BHandH scaled	AIM[a,b]	AIM[a,c]
Parallel	-2.71	-3.79	-3.27	-2.18	-1.27
Antiparallel	-3.47	-5.40	-4.48	-2.69	-1.68
Cross	-3.24	-4.58	-3.87	-2.91	-1.25

[a]AIM refers to binding energies estimated using the linear relationships 3.1.1 and 3.1.2 [b]at literature CCSD(T) geometry; [c]at BHandH optimised geometry.

Additionally to the above mentioned tests, six more DFT functionals were tested on the same literature coordinates. All functionals predict correct ordering of stability for the three complexes in comparison to the CCSD(T) order of stability. The calculated interaction energies and signed errors with respect to the CCSD(T) values are shown in table 3.1.7. Apart from BHandH, all the rest of the tested functionals underestimated the interaction and the best performance varies according to which complex is considered. Among the BHandH energies, the errors are slightly increased when BHandH optimised geometries are considered although the literature and BHandH geometries are in excellent agreement (figure 3.1.6).

Table 3.1.7: CP-corrected interaction energies using the 6-31++G** basis set.

	antiparallel		cross		parallel	
Method	kcal mol^{-1}	Signed error	kcal mol^{-1}	Signed error	kcal mol^{-1}	Signed error
PWB6K	-2.64	-0.83	-2.08	-1.16	-1.56	-1.15
PW6B95	-2.64	-0.83	-2.08	-1.16	-1.56	-1.15
MPWB1K	-2.64	-0.83	-2.08	-1.16	-1.56	-1.15
M05	-2.26	-1.21	-1.74	-1.50	-1.26	-1.46
M05-2X	-2.26	-1.21	-1.74	-1.50	-1.26	-1.46
MPW1B95	-2.26	-1.21	-1.74	-1.50	-1.26	-1.46
BHandH	-5.40	1.93	-4.58	1.34	-3.79	1.08
BHandH(opt)	-5.93	2.46	-5.26	2.02	-4.39	1.68
CCSD(T)	-3.47	-	-3.24	-	-2.71	-

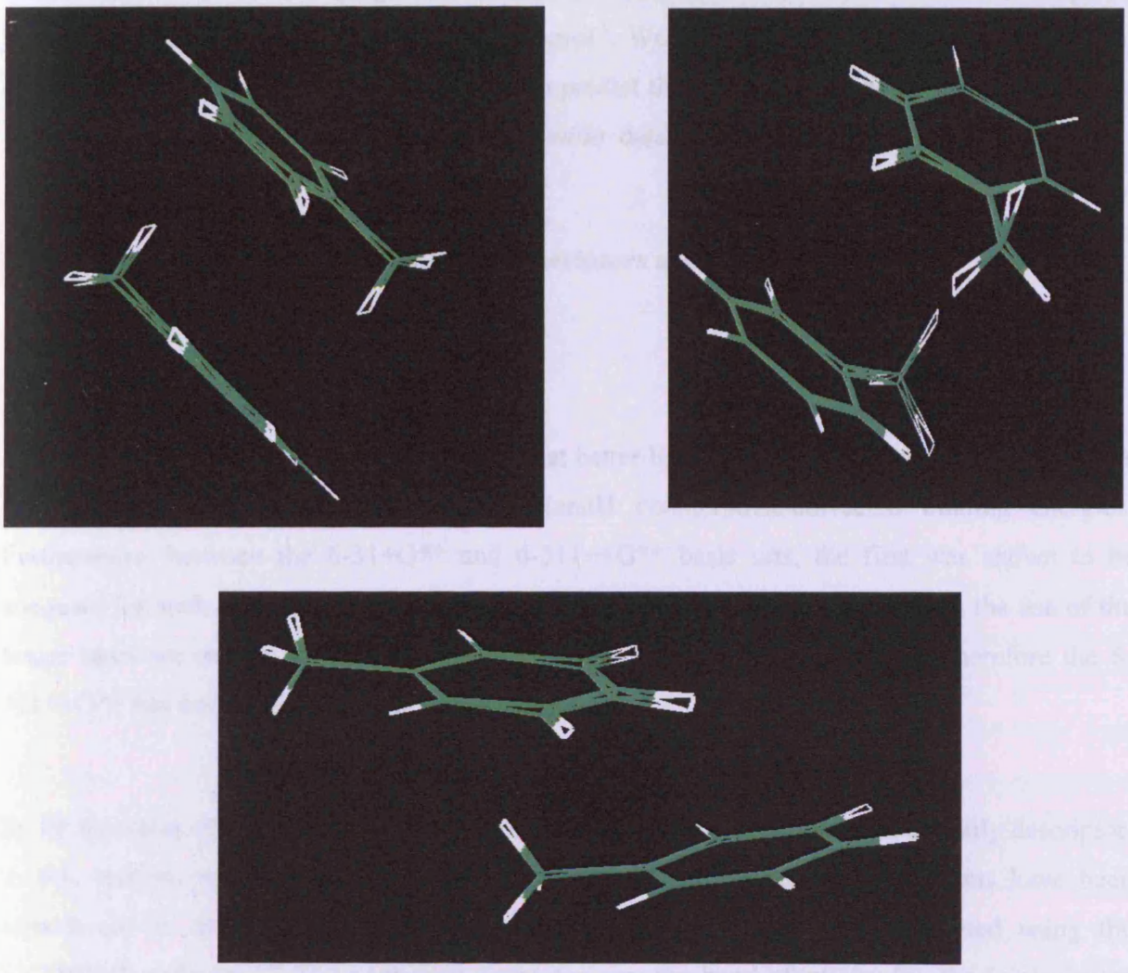


Figure 3.1.6: Superposition of literature and BHandH optimised geometries for the three toluene dimers.

3.1.7 Alkanes

Finally, recent work has shown that most common DFT functionals fail to recover the relative stability of linear and branched alkanes, a shortcoming that was ascribed to the poor description of weak interactions, most notably dispersion, in such functionals.[154] As a simple test of the performance of BHandH in this regard, geometries of *n*-octane and 2,2,3,3-tetramethylbutane were optimised using BHandH with both 6-31+G(d,p) and 6-311++G(d,p) basis sets. At these

minima, possessing C_{2h} and D_{3d} symmetry, respectively, the branched isomer is 2.06 (with the smaller basis set) and 2.93 (larger basis) kcal mol^{-1} more stable than the linear one, compared with an experimental value of 1.9 ± 0.5 kcal mol^{-1}. We note that this is in stark contrast to all other functionals considered in ref. 154, which predict the linear form to be the more stable one, with errors compared to experiment and/or *ab initio* data of between 5 and 12 kcal mol^{-1}.

3.2 Correlations between QTAIM-based descriptors and binding energy

3.2.1 Introduction

In the previous discussion it was mentioned that better linear fits are obtained between the BCPs of BHandH optimised geometries and BHandH counterpoise-corrected binding energies. Furthermore, between the 6-31+G** and 6-311++G** basis sets, the first was shown to be adequate for such analysis. However, because in some cases (stacked complexes) the use of the larger basis set reveals BCPs not present when the smaller basis set is used, therefore the 6-311++G** has been used in this further analysis.

So far the value of the density (ρ(r)) at the intermolecular BCPs was used as the only descriptor. In this section, apart from the ρ, more properties of the density at such points have been considered, as defined in the context of the theory of QTAIM and generated using the EXTREME software.[137] The additional properties are: the bond ellipticity (ε), the Laplacian of the density ($\nabla^2\rho(r)$), the kinetic energy densities (G(r) and K(r)), the function L(r) of $\nabla^2\rho(r)$, the nuclear-electronic potential energy density ($V_{nuc}(r)$) and finally the electronic potential energy density (V(r)). All these properties have been defined in paragraph 2.6.5. The analysis is again performed on the subgroups of the S22 database (H-bonded, dispersion-dominated and mixed), as well as on the whole S22 dataset.

3.2.2 H-bonded complexes

Considering the H-bonded subgroup alone (7 complexes), the linear fits that are obtained are summarised in table 3.2.1.

Table 3.2.1: Summary of linear fits for the H-bonded complexes.

	slope	intercept	R^2	RMSE
ρ(r)	-227.599	-1.980	0.952	2.159
$\nabla^2\rho(r)$	-96.171	1.792	0.964	1.875
G(r)	-272.197	-1.873	0.951	2.196
K(r)	-742.072	-12.595	0.732	5.121
L(r)	384.683	1.792	0.964	1.875
Vnuc(r)	-0.287	-10.473	0.504	6.967
V(r)	205.403	-4.275	0.920	2.797
E(r)	742.072	-12.595	0.732	5.121

As seen from table 3.2.1, with the exception of K(r) and Vnuc(r), all properties yield high correlations with the binding energies, with R^2 above 0.900. The most satisfactory fits are obtained in the cases of $\nabla^2\rho(r)$ and L(r) which show the highest R^2 (0.964) and the lowest RMSE (1.875 kcal mol^{-1}). By definition of the property L(r), which is simply a function proportional to the Laplacian of the density, it is not a surprise that the linear fits for these two properties are of almost identical quality, yet exhibiting substantially different slopes. ρ(r) and G(r) give similar fits, both with RMSE above 2 kcal mol^{-1}, followed by V(r) which is of poorer quality. In most cases, the largest errors are introduced by the uracil, formamide and formic acid dimers, however at least at this stage the dataset is rather limited so as to consider improvement of the models by removing possible outliers.

3.2.3 Dispersion complexes

As noted in the previous analysis for the cases of the ethene dimer and the benzene-methane complex, the number of intermolecular BCPs (4 and 6 respectively) yields a sum of ρ(r) that leads to large errors for these two systems. Therefore the average value of the BCPs was used, which is also the case for all the properties in the present analysis.

Table 3.2.2: Summary of linear fits for the dispersion dominated complexes.

	slope	intercept	R^2	RMSE
ρ(r)	-231.040	-0.487	0.974	0.819
$\nabla^2\rho(r)$	-68.873	-0.683	0.979	0.739
G(r)	-313.420	-0.786	0.980	0.711
K(r)	2259.953	0.024	0.959	1.020
L(r)	275.504	-0.683	0.978	0.741
Vnuc(r)	-0.081	-1.025	0.949	1.136
V(r)	363.294	-0.923	0.982	0.679
E(r)	-2260.054	0.024	0.959	1.018

The overall picture for the subset of the dispersion complexes is quite different, with all the considered properties yielding high quality linear fits. Indicative of this fact is that the poorest fit which is again obtained against Vnuc(r) as in the case of the H-bonded complexes, shows a high R^2 equal to 0.949 and RMSE close to 1.14 kcal mol^{-1}. The best model is obtained for V(r) with $R^2 = 0.982$ and RMSE 0.679 kcal mol^{-1} (data shown in table 3.2.2).

3.2.4 Mixed Complexes

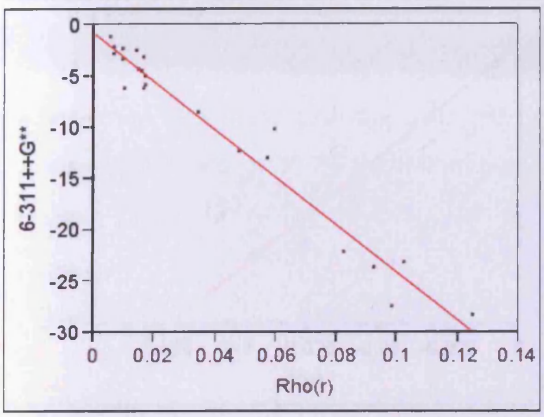
The case of the mixed complexes is more challenging to describe in terms of a linear fit, as the intermolecular interactions are more complex and this can be observed in the statistical measures of the following table (table 3.2.3).

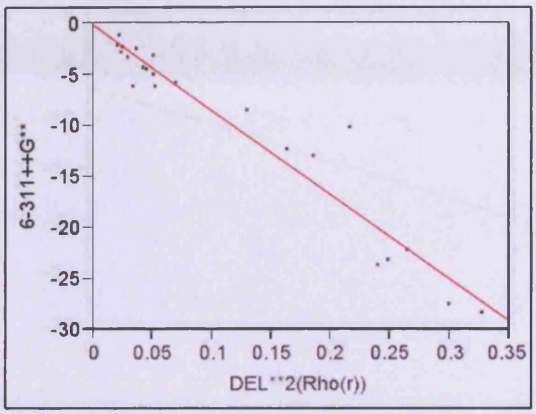
Table 3.2.3: Summary of linear fits for the dispersion mixed complexes.

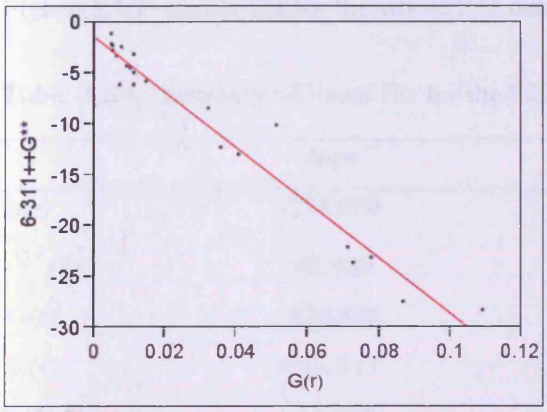
	slope	intercept	R^2	RMSE
ρ(r)	-133.631	-2.363	0.712	1.557
$\nabla^2\rho(r)$	-36.090	-2.659	0.735	1.495
G(r)	-147.617	-2.873	0.726	1.520
K(r)	4278.354	2.814	0.748	1.458
L(r)	144.359	-2.659	0.735	1.495
Vnuc(r)	-0.096	-2.178	0.716	1.546
V(r)	150.889	-3.100	0.716	1.549
E(r)	-4278.368	2.814	0.748	1.458

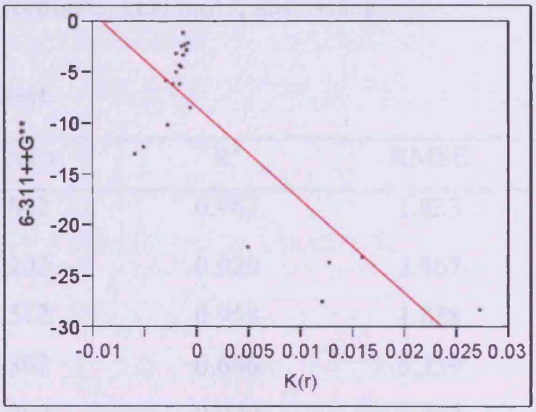
3.2.5 S22 dataset

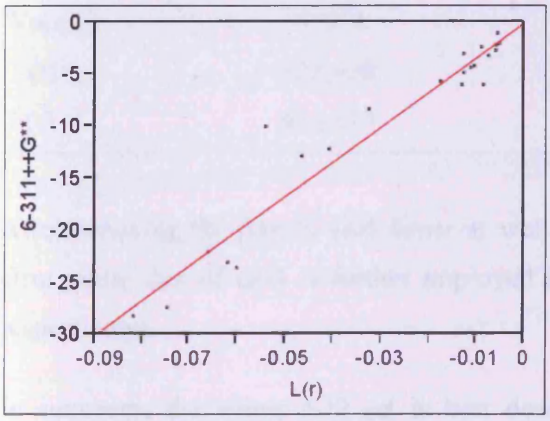
The main sources of error on the obtained linear fits for the whole of the S22 set are the molecules of the phenol dimer and the formic acid dimer. The fits are shown in figure 3.2.1 and table 3.2.4. In the case of the electron density, which is the best fit, along with G(r), the phenol dimer is an apparent outlier, as seen from the residuals plot. The molecule is characterized by the presence of 4 intermolecular BCPs, of which 2 are common H-bonds and 2 not. Removing the values of the unusual BCPs from the considered sum, decreases the quality of the fit, while completely removing this data point results in $R^2 = 0.973$ and RMSE = 1.54 kcal mol^{-1}, in the case of ρ(r). The change is more evident for G(r), for which the RMSE drops slightly lower to 1.5 kcal mol^{-1} with $R^2 = 0.975$. Further improvement is observed also for L(r), V(r) and $\nabla^2\rho(r)$, yet G(r) remains the best describing property.

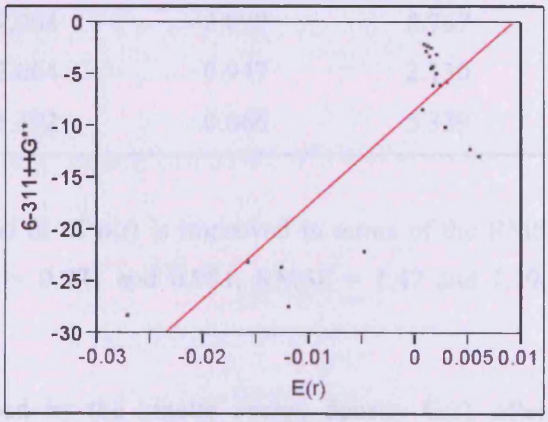

6-311++G**
Rho(r)


6-311++G**
DEL**2(Rho(r))


6-311++G**
G(r)


6-311++G**
K(r)


6-311++G**
L(r)


6-311++G**
E(r)

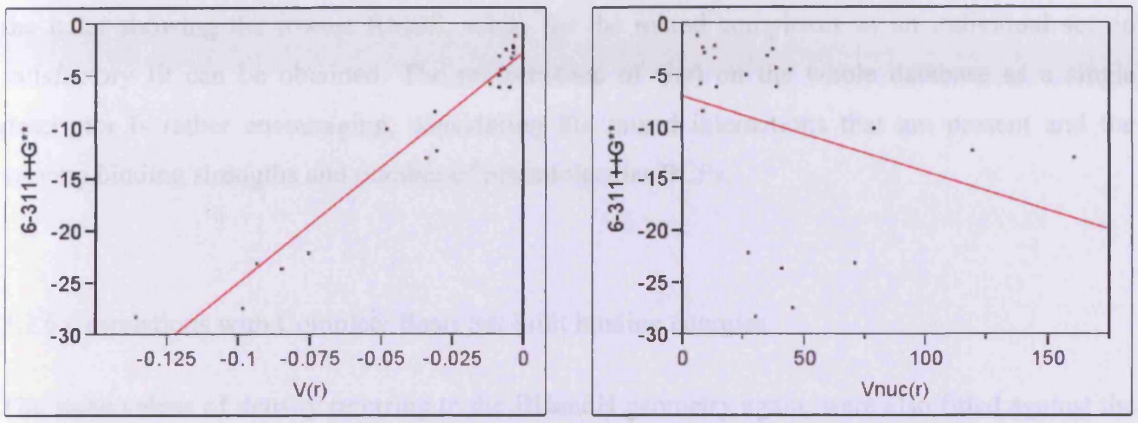


Figure 3.2.1: Linear fits for the whole S22 dataset (ordinate: kcal mol^{-1}, abscissa: a.u.).

Table 3.2.4: Summary of linear fits for the S22 dataset.

	slope	intercept	R^2	RMSE
ρ(r)	-234.069	-0.802	0.961	1.833
$\nabla^2\rho(r)$	-82.989	-0.202	0.929	2.467
G(r)	-271.676	-1.512	0.958	1.888
K(r)	-912.517	-8.392	0.666	5.339
L(r)	331.958	-0.202	0.929	2.467
Vnuc(r)	-0.074	-7.004	0.099	8.767
V(r)	222.430	-2.664	0.947	2.130
E(r)	912.517	-8.392	0.666	5.339

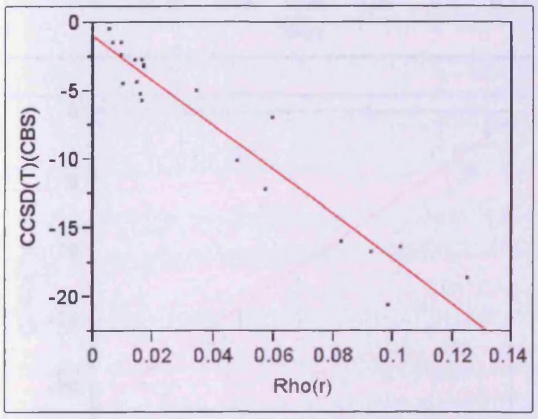
After removing the formic acid dimer as well, the fit of ρ(r) is improved in terms of the RMS error while that of G(r) is further improved (R^2 = 0.971 and 0.981, RMSE = 1.47 and 1.19, respectively).

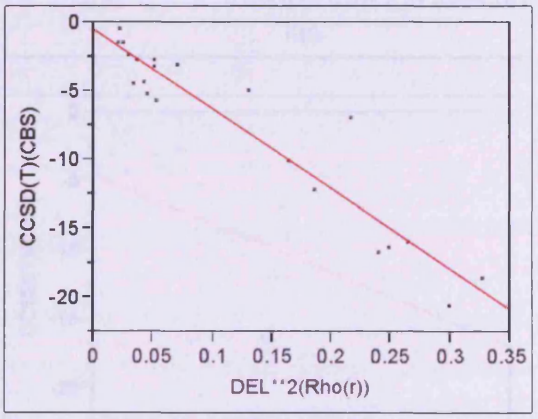
In summary, the whole S22 set is best described by the kinetic energy density G(r), after removing the formic acid and phenol dimers as outliers, while a satisfactory fit is also obtained in terms of the density ρ(r). For the individual subsets, the H-bonded complexes are best described by the Laplacian of ρ(r) and L(r). For the dispersion complexes all the considered properties yield linear fits of high quality, with G(r) and V(r) giving the highest correlation and

the latter showing the lowest RMSE, while for the mixed complexes as an individual set no satisfactory fit can be obtained. The performance of G(r) on the whole database as a single descriptor is rather encouraging, considering the mixed interactions that are present and the varying binding strengths and number of intermolecular BCPs.

3.2.6 Correlations with Complete Basis Set limit binding energies

The same values of density referring to the BHandH geometry again, were also fitted against the CCSD(T)/CBS and MP2/CBS interaction energies as estimated by Hobza, as well as the respective CBS estimates of Werner and are presented collectively for the whole S22 dataset. The following graphs (figure 3.2.2) summarise the fits against the CCSD(T)/CBS binding energy estimates, without the removal of outliers. The overall performance of the fitting procedure for all the considered properties is similar to that against the BHandH binding energies. Table 3.2.5 summarises the fitting parameters against the CCSD(T) energies before and after the removal of outliers.





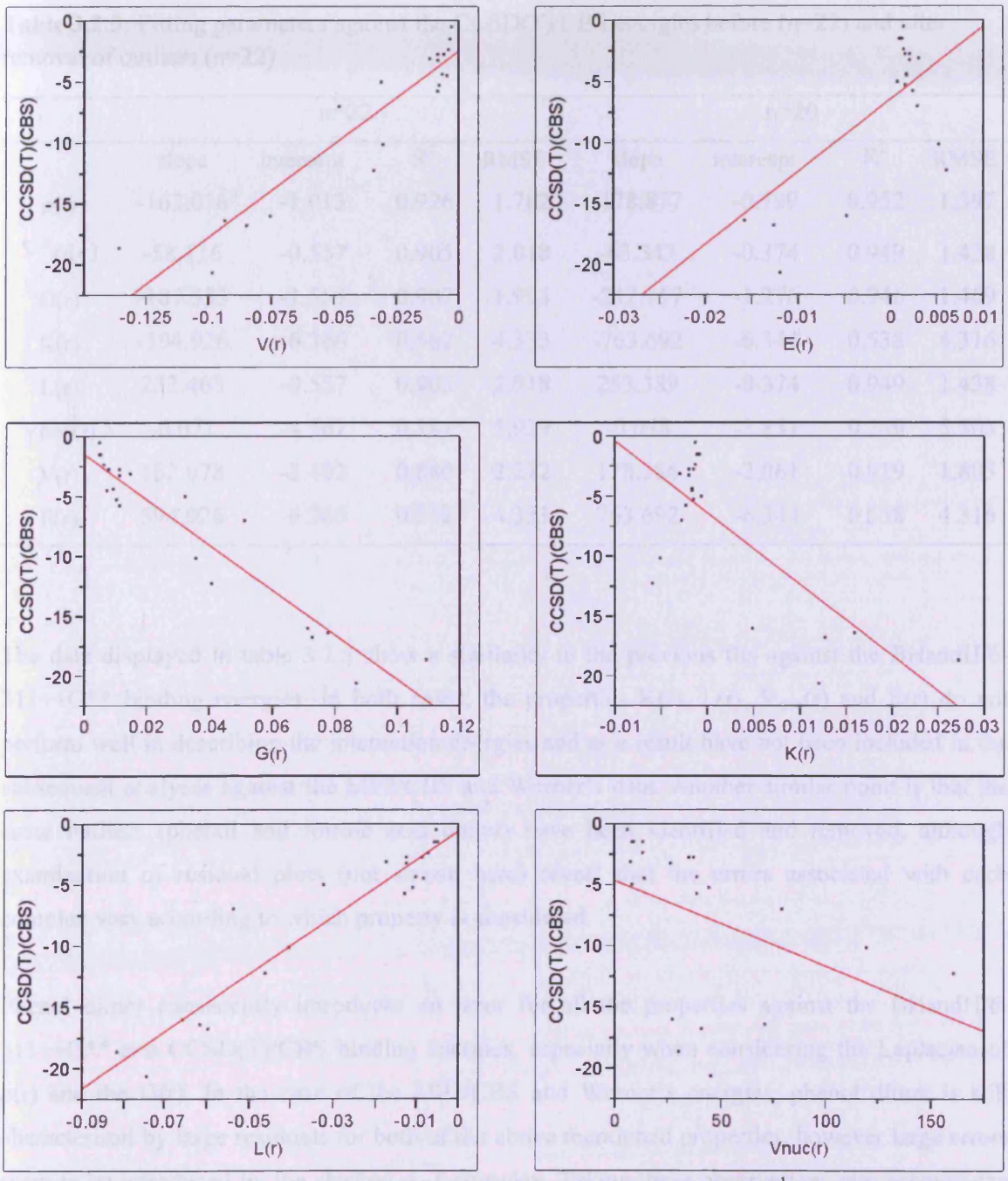


Figure 3.2.2: Linear fits against CCSD(T)/CBS B.E.'s (ordinate: kcal mol^{-1}, abscissa: a.u.).

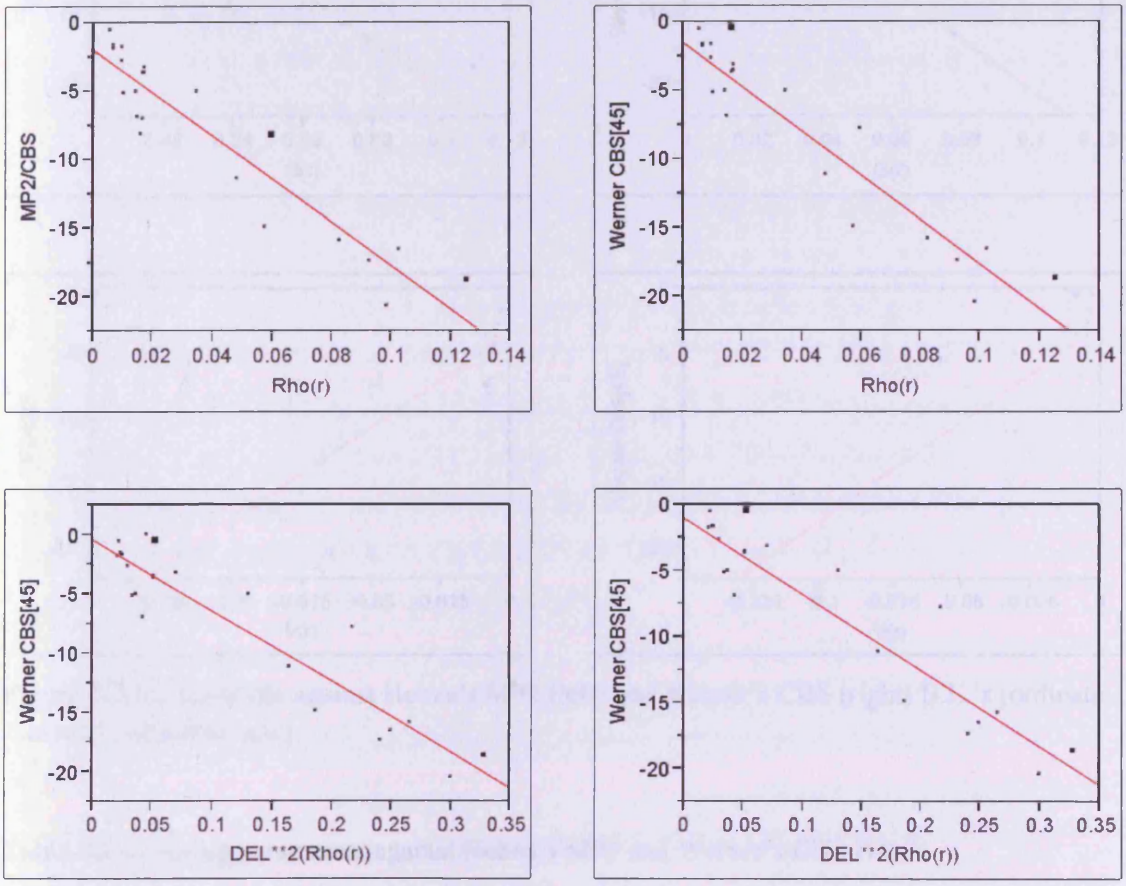
Table 3.2.5: Fitting parameters against the CCSD(T)/CBS energies before (n=22) and after removal of outliers (n=22).

	n=22				n=20			
	slope	intercept	R^2	RMSE	slope	intercept	R^2	RMSE
ρ(r)	-163.016	-1.013	0.926	1.782	-178.877	-0.799	0.952	1.397
$\nabla^2\rho(r)$	-58.116	-0.557	0.905	2.018	-63.347	-0.374	0.949	1.438
G(r)	-187.533	-1.559	0.907	1.993	-212.757	-1.276	0.946	1.469
K(r)	-594.926	-6.366	0.562	4.333	-763.692	-6.344	0.538	4.316
L(r)	232.463	-0.557	0.905	2.018	253.389	-0.374	0.949	1.438
Vnuc(r)	-0.071	-4.567	0.181	5.929	-0.078	-3.831	0.249	5.503
V(r)	152.078	-2.402	0.880	2.272	178.386	-2.061	0.919	1.805
E(r)	594.926	-6.366	0.562	4.333	763.692	-6.344	0.538	4.316

The data displayed in table 3.2.5 show a similarity to the previous fits against the BHandH/6-311++G** binding energies. In both cases, the properties K(r), L(r), $V_{nuc}(r)$ and E(r) do not perform well in describing the interaction energies and as a result have not been included in the subsequent analyses against the MP2/CBS and Werner's data. Another similar point is that the same outliers (phenol and formic acid dimer) have been identified and removed, although examination of residual plots (not shown here) reveal that the errors associated with each complex vary according to which property is considered.

Phenol dimer consistently introduces an error for all the properties against the BHandH/6-311++G** and CCSD(T)/CBS binding energies, especially when considering the Laplacian of ρ(r) and the G(r). In the case of the MP2/CBS and Werner's energies, phenol dimer is still characterised by large residuals for both of the above mentioned properties, however large errors seem to be introduced by the stacked A-T complex. Taking these observations into account can help justify the removal of the phenol dimer as an outlier for the analysis, but the removal of the formic acid dimer is still questionable.

Fitting of ρ(r), $\nabla^2\rho(r)$, G(r) and V(r) against the Hobza MP2/CBS and Werner's energies for the whole S22 dataset is shown below (fig. 3.2.3) and their parameters are shown in table 3.2.6.



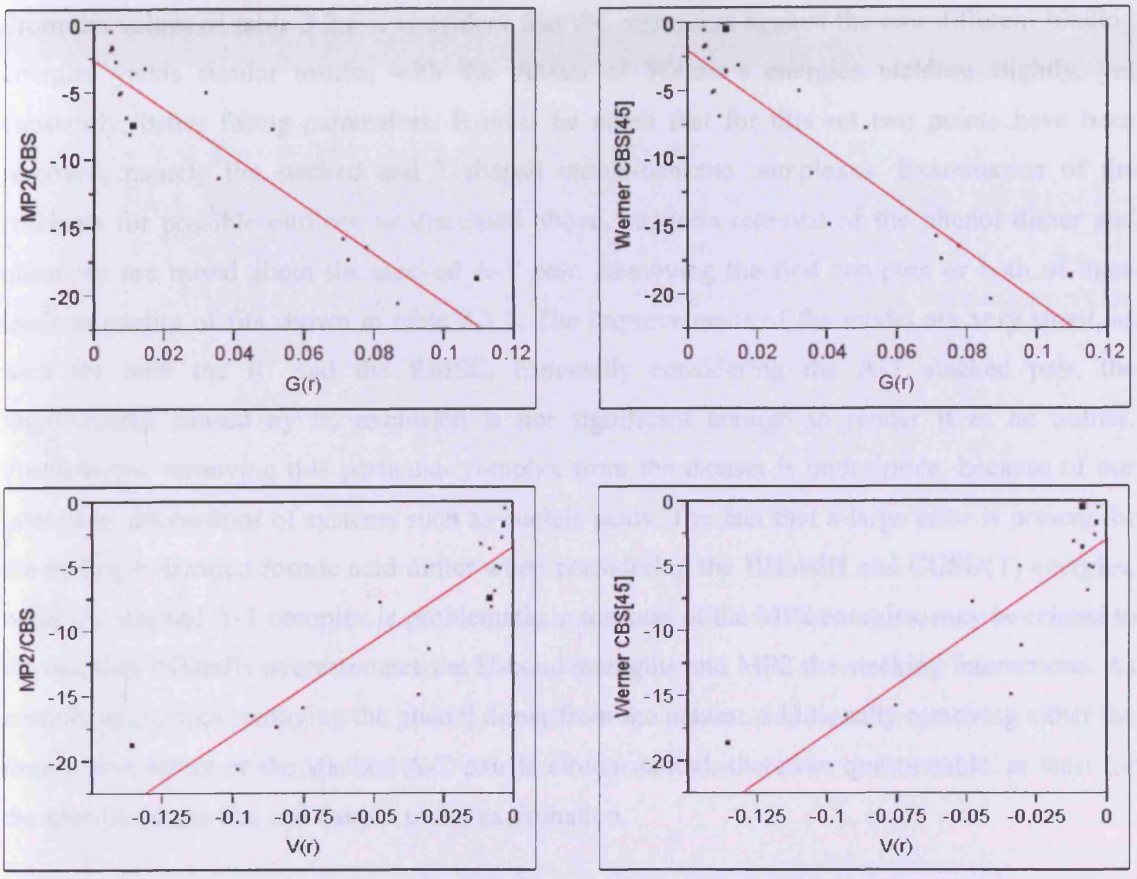


Figure 3.2.3: Linear fits against Hobza's MP2 (left) and Werner's CBS (right) B.E.'s (ordinate: kcal mol^{-1}, abscissa: a.u.).

Table 3.2.6: Fitting parameters against Hobza's MP2 and Werner's CBS B.E.'s.

	MP2/CBS				Werner CBS[45]			
	slope	intercept	R^2	RMSE	slope	intercept	R^2	RMSE
$\rho(r)$	-157.080	-2.025	0.873	2.315	-161.580	-1.482	0.899	2.162
$\nabla^2\rho(r)$	-60.151	-0.360	0.871	2.487	-57.905	-0.968	0.882	2.339
G(r)	-179.133	-2.600	0.841	2.593	-184.967	-2.038	0.869	2.468
V(r)	144.444	-3.432	0.806	2.864	148.777	-2.930	0.831	2.801

From the values of table 3.2.6 it is evident that the regression against the two different binding energies yields similar results, with the dataset of Werner's energies yielding slightly, yet constantly, better fitting parameters. It must be noted that for this set two points have been removed, namely the stacked and T-shaped indole-benzene complexes. Examination of the residuals for possible outliers, as discussed above, suggests removal of the phenol dimer and questions are raised about the stacked A-T pair. Removing the first complex or both of them leads to quality of fits shown in table 3.2.7. The improvements of the model are very small, as seen for both the R^2 and the RMSE. Especially considering the A-T stacked pair, the improvement caused by its exclusion is not significant enough to render it as an outlier. Furthermore, removing this particular complex from the dataset is undesirable, because of our interest in interactions of systems such as nucleic acids. The fact that a large error is present for the hydrogen-bonded formic acid dimer when considering the BHandH and CCSD(T) energies, while the stacked A-T complex is problematic in the case of the MP2 energies, may be related to the fact that BHandH overestimates the H-bond strengths and MP2 the stacking interactions. As a result, apart from removing the phenol dimer from the dataset, additionally removing either the formic acid dimer or the stacked A-T pair is circumstantial, therefore questionable, at least for the specific properties and dataset under examination.

Table 3.2.7: Fitting parameters after removal of possible outliers from the dataset.

	MP2/CBS				Werner CBS[45]			
	phenol out n=21		A-T(stacked) out n=20		phenol out n=19		A-T stack out n=18	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
ρ(r)	0.890	2.214	0.902	2.085	0.914	2.058	0.930	1.858
$\nabla^2\rho(r)$	0.898	2.133	0.900	2.108	0.925	1.923	0.930	1.858
G(r)	0.861	2.487	0.884	2.268	0.887	2.360	0.915	2.045
V(r)	0.815	2.871	0.857	2.513	0.838	2.821	0.888	2.347

3.3 QTAIM metal-arene

QTAIM analysis has been previously employed in a work of Palusiak *et al.*[155] which involved η^5-C_5H_5 - metal binding. The authors reported that the resulting topology did not correspond to η^5-binding and only 2 or 3 metal – arene bond critical points were found instead of 5. On the one hand, this effect was ascribed to the complex topology around the metal atom resulting from the use of ECP, while emphasising the fact that the Poincaré-Hopf relationship (section 2.6.4) is not satisfied. On the other hand, observed variations of both the atomic charges of the cyclopentadienyl carbon atoms as well as of the metal – π bond lengths led the authors to doubt the equivalence of all arene atoms in the π-bond.

In this work the topology of the density of the metal-arene binding was explored using the iron(II) complex $[(ben)Fe(en)Cl]^+$ (ben=benzene, en=ethylenediamine). Our interest was in the binding between ruthenium-arene complexes and nucleobases and for this initial test the Cl^- was used instead of nucleobase in order to decrease the computational cost. Also, Fe(II) was chosen over Ru(II) in order to perform wavefunction calculations both with and without an ECP.

3.3.1 Structure of the complex $[(ben)Fe(en)Cl]^+$

The complex was firstly optimised using B3LYP and BHandH at the 6-31+G**/SDD level. The two resulting structures did not have significant differences and are shown superimposed in figure 3.3.1. The arene-metal distance using BHandH is shorter than using B3LYP by an average of 0.046 Angstroms. The BHandH structure was chosen for the subsequent wavefunction calculations.

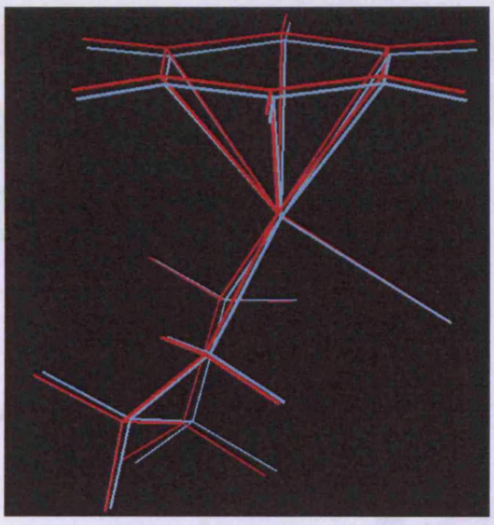


Figure 3.3.1: Optimised geometries of the $[(\eta^6\text{-ben})Fe(en)Cl]^+$ complex (blue: B3LYP, red: BHandH).

3.3.2 Functional

The wavefunction of the BHandH structure was evaluated at the same level of theory as the optimisation (6-31+G**/SDD) using tight SCF criteria. The analysis of the density revealed only three arene-metal BCPs, instead of six that would be expected for η^6-binding, in accordance with the observations of Palusiak *et al.* Two of these BCPs are found between the C5 and C6 arene atoms the metal, while the third seems to be between the metal and C2-C3 BCP. Furthermore, the topology of the density around the metal atom is complex, which is a result of describing the metal with the ECP. Figure 3.3.2 displays the molecular graph for this wavefunction.

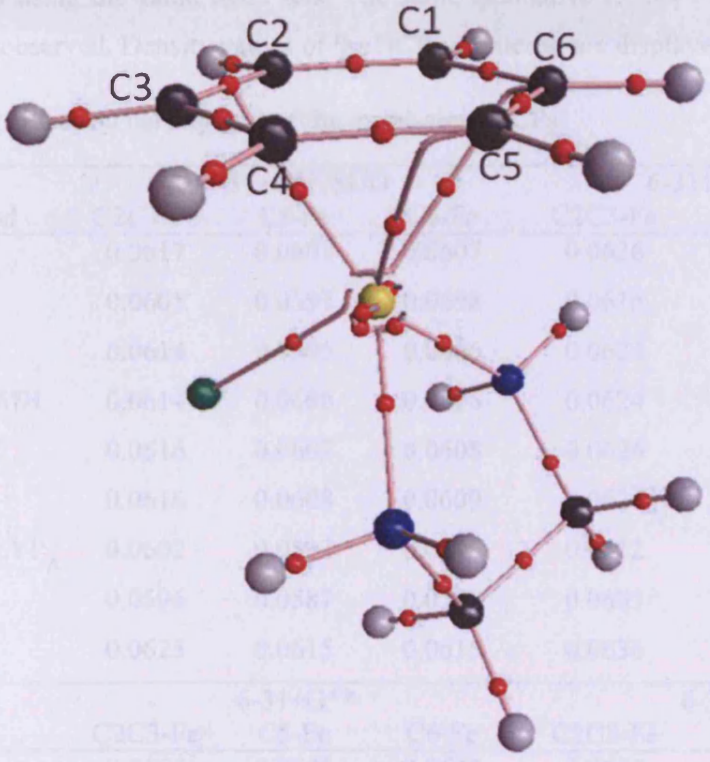


Figure 3.3.2: Molecular graph at the BHandH/6-31+G**/SDD level.

Further calculations were performed on the same structure using other DFT functionals (B3LYP, B3PW91, MPW1PW91, PBE1PBE, O3LYP, BHandHLYP), as well as MP2 and HF theory. In all cases three iron-benzene BCPs were predicted at the same positions, indicating that this is not a result based on the functional that is employed.

3.3.3 Basis Set

The dependence of the density topology was also examined with respect to the used basis set and ECP. Apart from 6-31+G**, that was mentioned in the previous paragraph, the larger 6-311++G** basis set was combined with the SDD ECP, as well as two sets of all-electron

calculations using the same basis sets. The same qualitative results of the previous paragraph were again observed. Density values of the BCPs of interest are displayed in table 3.3.1.

Table 3.3.1: Electron density (au) of the metal-arene BCPs.

	6-31+G**/SDD			6-311++G**/SDD		
method	C2C3-Fe	C5-Fe	C6-Fe	C2C3-Fe	C5-Fe	C6-Fe
BHandH	0.0617	0.0607	0.0607	0.0626	0.0615	0.0615
B3LYP	0.0605	0.0597	0.0598	0.0616	0.0605	0.0606
B3PW91	0.0614	0.0605	0.0606	0.0624	0.0613	0.0614
MPW1PW91	0.0614	0.0606	0.0606	0.0624	0.0614	0.0614
PBE1PBE	0.0616	0.0607	0.0608	0.0626	0.0616	0.0616
O3LYP	0.0616	0.0608	0.0609	0.0626	0.0616	0.0616
BHandHLYP	0.0602	0.0593	0.0593	0.0612	0.0601	0.0601
HF	0.0596	0.0587	0.0588	0.0605	0.0594	0.0595
MP2	0.0625	0.0615	0.0615	0.0636	0.0624	0.0624
	6-31+G**			6-311++G**		
	C2C3-Fe	C5-Fe	C6-Fe	C2C3-Fe	C5-Fe	C6-Fe
BHandH	0.0655	0.0645	0.0645	0.0647	0.0636	0.0636
B3LYP	0.0647	0.0637	0.0638	0.0640	0.0630	0.0630
B3PW91	0.0654	0.0645	0.0645	0.0648	0.0637	0.0637
MPW1PW91	0.0655	0.0645	0.0645	0.0648	0.0637	0.0637
PBE1PBE	0.0657	0.0647	0.0647	0.0649	0.0639	0.0639
O3LYP	0.0654	0.0646	0.0646	0.0649	0.0639	0.0639
BHandHLYP	0.0644	0.0634	0.0634	0.0635	0.0624	0.0624
HF	0.0638	0.0628	0.0628	0.0623	0.0854	0.0613
MP2	0.0662	0.0652	0.0652	0.0650	0.0639	0.0639

One point to mention is that the density of the BCP that lies between the metal and the C2-C3 BCP is slightly larger than that of the other two BCPs, irrespective of the method and basis set used. The numerical differences observed for different basis sets are not found to be larger than 0.0042 a.u.

3.3.4 Scan

Additionally to the above calculations a relaxed scan was performed on the Cl-Fe-X-C dihedral angle, where C is a carbon atom of the benzene ring and X is the centroid of the ring. The scan was performed at steps of 5 degrees and verified that a small barrier to rotation exists at 0^{o} of this dihedral angle and a minimum for 30^{o}. The barrier comes up to 0.72 kcal mol^{-1} and the obtained curve is shown in figure 3.3.3.

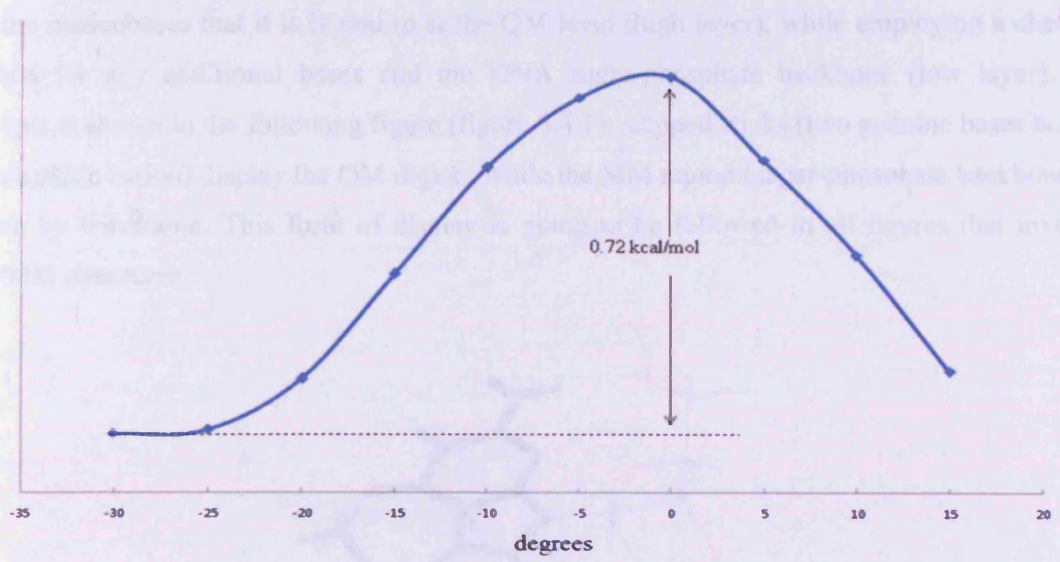


Figure 3.3.3: Energy barrier to rotation of the benzene ring in the $[(\eta^6\text{-ben})Fe(en)Cl]^+$ complex.

QTAIM analysis of the structure at the maximum and at a random point (dihedrals of 0^0 and -10^0 respectively) was performed, using at the BHandH/6-31+G**/SDD level. Again only three arene – metal BCPs were found. A notable difference is that for these non-equilibrium structures the BCPs of interest are associated with alternating carbon atoms of the benzene ring, rather than consecutive atoms, as previously mentioned.

From all the above, it can be concluded that the observed topology of the density between the metal and the arene is not an artefact of the method or the ECP.

3.4 ONIOM calculations

3.4.1 Introduction – Input requirements

The intention to study larger systems using the ONIOM method requires the partitioning of the systems into high and low layers, as was described in section 2.5.3. In the present study where binding between metal drugs and DNA is involved, an obvious choice is to treat the metallodrug and the nucleobases that it is bound to at the QM level (high layer), while employing a cheaper method for any additional bases and the DNA sugar-phosphate backbone (low layer). An example is shown in the following figure (figure 3.4.1): capped sticks (two guanine bases bound to a cisplatin cation) display the QM region, while the MM region (sugar-phosphate backbone) is shown by wireframe. This form of display is going to be followed in all figures that involve QM/MM structures.

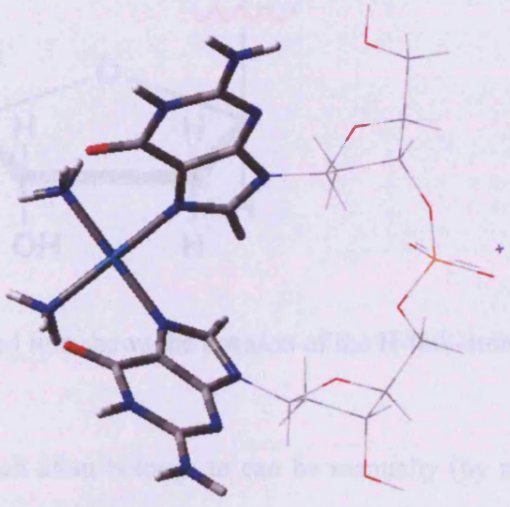


Figure 3.4.1: Display of QM (stick) and MM (wireframe) regions. Purple dot depicts a Na^+ counter-ion.

Following the previous discussion, the BHandH DFT method with the 6-31+G**/SDD basis set and ECP has been chosen for the high layer. The MM calculation on the low layer is performed using the AMBER force field which has been parameterised for nucleic acids.

The two layers are separated by link atoms, where the nucleobase – sugar covalent bond is present. Figure 3.4.2 shows an example of the position of a link atom when breaking the N-C bond between a guanine and its sugar backbone. H-atoms were used for the substitution of the carbon atom, while maintaining the polarity of the original N-C bond. In cases where more bases are surrounding the bases of the high level, either as Watson-Crick H-bonded pairs or as stacked bases, a non-bonded approach was followed.

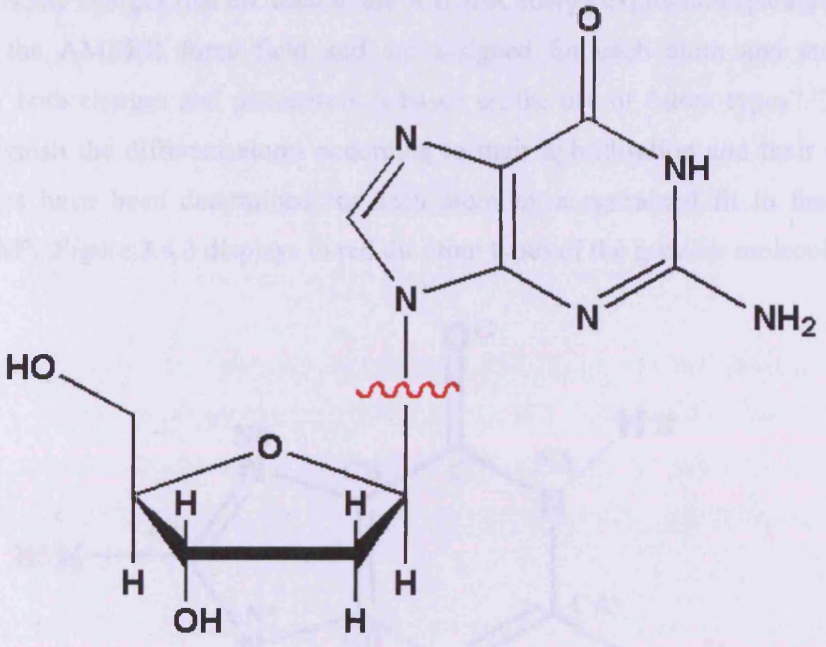


Figure 3.4.2: The red curved line shows the position of the H-link-atom.

The layer (high or low) each atom belongs to can be manually (by mouse selection) specified using the GaussView package which further automatically determines the position of the link atoms. Apart from the link atom position, use of the AMBER force field requires that "atom-types" and charges be specified for each of all the atoms in the input, along with their coordinates. The required input format of the calculations is automatically generated by GaussView (including atom types for the whole molecule), but due to technical problems atomic charges are not assigned for all of the atoms. The following paragraph deals with the

specification of charges at the low layer. Alternative charge schemes for the high layer are tested in paragraph 3.4.4 using large complexes.

3.4.2 Atom types and atomic charges

The parameters and charges that are used in the AMBER energy expression (paragraph 2.5.2) are contained in the AMBER force field and are assigned for each atom and atom-pair. The assignment of both charges and parameters is based on the use of "atom-types". The latter are used to distinguish the different atoms according to their hybridisation and their neighbouring atoms. Charges have been determined for each atom by a restrained fit to the electrostatic potential (RESP). Figure 3.4.3 displays in red the atom types of the guanine molecule.

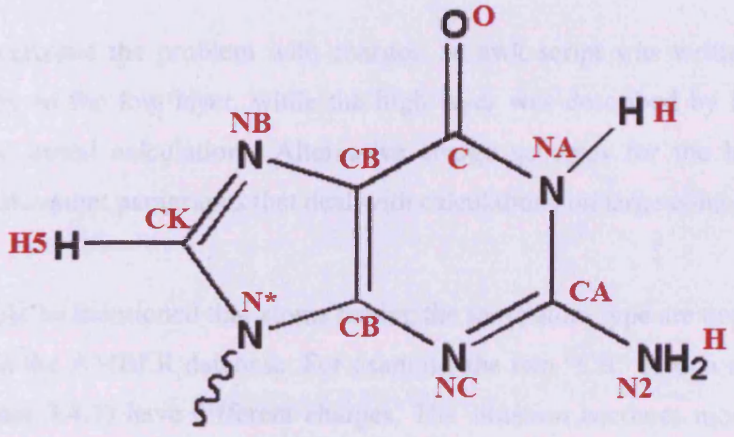


Figure 3.4.3: Atom types on guanine.

As an example, the carbon and oxygen atoms of the carbonyl group of fig. 3.4.3 have charges of 0.4918 and -0.5699e, respectively. The force constant (k) of the double bond between these two atoms is 570.00 (kcal mol^{-1})/$Å^2$ and the equilibrium bond length (r_0) is 1.229Å. However, such parameters and charges are not available for all the atoms of the periodic table, including the metals that are involved in the present work. As a result, charges for the high-layer atoms are not

available and are not automatically generated as described in the previous paragraph. Furthermore, in most (if not all) cases, charges are not assigned for several low-layer atoms as well, presumably due to the presence of the metal atoms.

Following previous work of our group[28], the lack of parameters for the metals has been dealt with by using the so-called "soft" parameters, i.e. approximate equilibrium bond lengths and angles associated with a small force constant. In this way a flattened parabolic potential is defined for the metals. Specifically in the present work, metal-involving bonds and angles were assigned equilibrium values of 1.08Å and 120.00 degrees, respectively and a force constant of 5.0 (kcal mol^{-1})/$Å^2$. The effect of using soft parameters does not introduce significant errors if they are used exclusively in the QM region, since ONIOM is a subtractive scheme. This effect is better explored in section 3.4.4.

In order to overcome the problem with charges, an awk script was written in order to assign atomic charges on the low layer, while the high layer was described by Merz-Kollman (MK) charges in the initial calculations. Alternative charge schemes for the high layer atoms are explored in subsequent paragraphs that deal with calculations on large complexes (section 3.4.4).

Firstly, it should be mentioned that atoms having the same atom type are not accompanied by the same charge in the AMBER database. For example, the two "CB" carbon atoms on the guanine molecule (figure 3.4.3) have different charges. The situation becomes more complicated when considering all the different bases (A, T, G, C, U) and whether the molecule is RNA or DNA. Initially, the database files that contain the AMBER charges for all nucleic acids bases (both RNA and DNA) were downloaded from the AMBER website[156]. An awk script was written (dna_types.awk) in order to collect all the charges encountered for any given atom type and estimate the average, minimum and maximum charge for each atom type. The purpose of this script was to examine if the variations of the charge for each atom type is small and allows for an average charge to be assigned for each atom type. However, in many cases the variations were larger than 0.2 charge units (a value above which we considered the variation non-negligible) and even worse in some cases for a given atom type charges with opposite signs were observed.

The above led us to consider distinguishing the charges for each atom type, each base and the sugar-phosphate backbone. Average charges were used only in cases that were difficult to be distinguished, yet the average value was close to the values from which it was estimated. Below follows an example of how the script works, based on one of the "complicated" cases of atom types:

In figure 3.4.4 the atom types that are present in the A and G molecules are shown. We consider the case of the CB atom types, which are present twice on each molecule, i.e. we have to assign 4 different charges to one atom type. The distinction is made by defining a "subtype" in the script using distance criteria. Therefore the four CB types are distinguished as CBNB and CBNN (if they are close to a NB or N* atom, respectively). The two subtypes are further distinguished by whether there is a carbonyl group in the molecule (guanine) or not (adenine). In all the cases that were also not straightforward to distinguish, the distinction was achieved in a similar manner.

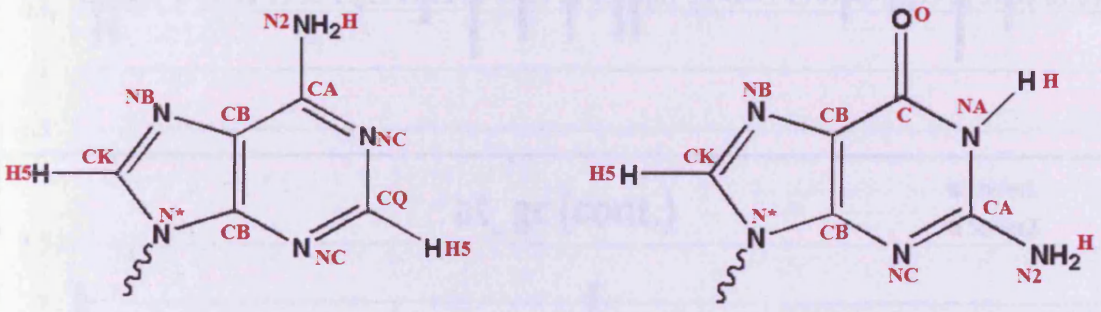


Figure 3.4.4: Atom types on A (left) and G (right) molecules.

After the script was written to include all possible atom types that might be encountered, it was tested against nucleic acid fragments generated by the Hyperchem database. The tests were performed on three "test sets" A, B and C. Set A consisted of any single base with the deoxyribonucleic (d-) or ribonucleic (r-) backbone (d-A, d-G, d-T, d-C, r-A, r-G, r-U, r-C), as well as the Watson-Crick A-T and G-C H-bonded pairs with d- backbone (AT, GC). Set B contained combinations of two stacked bases connected with d-sugar-phosphate backbone (AT,

AC, AG, AA, GT, GC, GG, TC, TT, CC) and set C included double-stranded DNA and RNA pairs (dAT-dAT, dAT-dGC, dGC-dGC, rAU-rAU, rGC-rGC, rAU-rGC). Finally, each of the A, B and C test sets was prepared for three different conformations (A, B and Z forms) of the nucleic acids. The results of the application of the script on all of the above structures are not reported here. Instead only one example follows, that of the B-form of the dAT-dGC structure (figure 3.4.5).

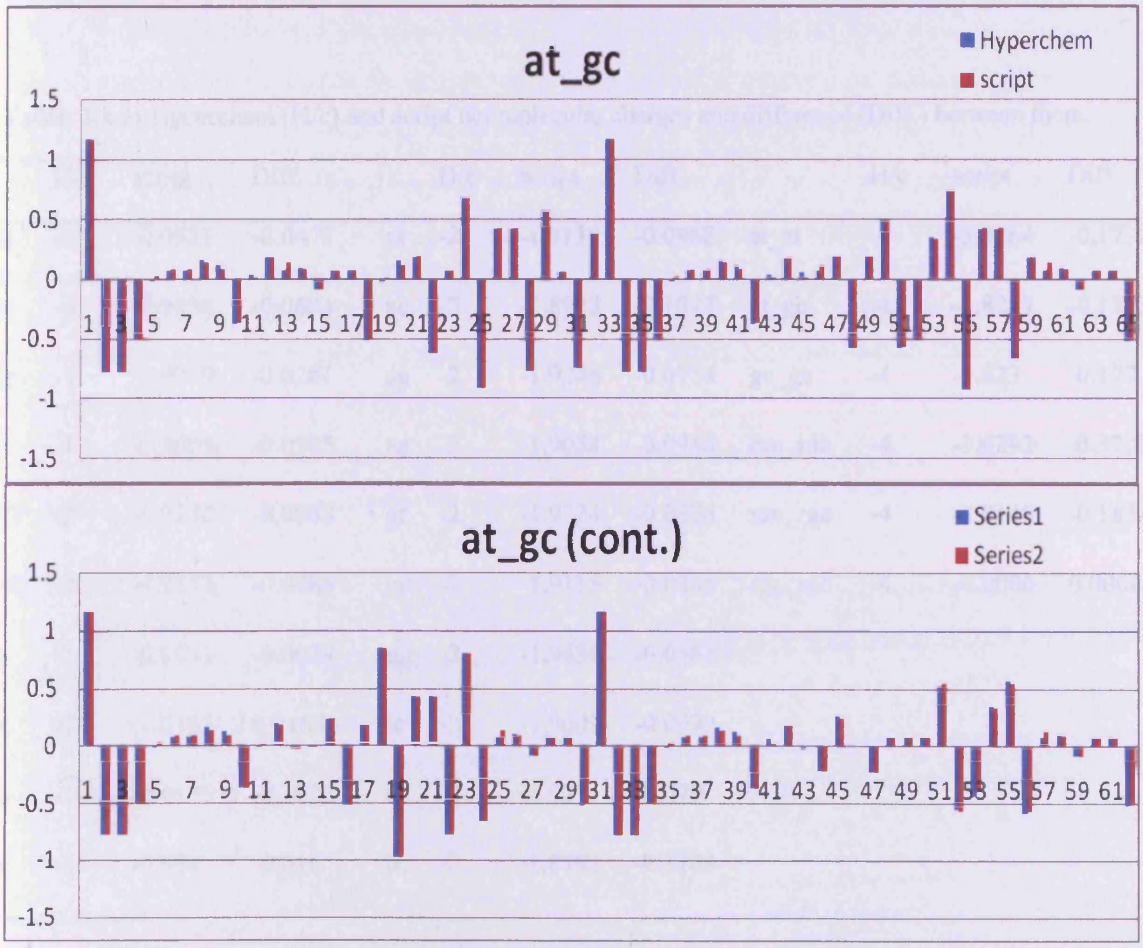


Figure 3.4.5: Comparison of script and Hyperchem charges for the dAT-dGC molecule.

Another point worth mentioning is related to the total sum of the charges for each the above described tests. In Table 3.4.1 the sum of the script assigned charges are compared to the

Hyperchem sums for the B form of all the three test sets (A,B and C). The observed differences are in most cases less than 0.1 charge units and are considered a good approximation of the real total charge. However, error accumulates as more bases are included in the molecule and for the double stranded DNA fragments (at-at, at-gc adn gc-gc) the absolute difference from the actual charge sums up to 1.7 and 1.8 units. Larger differences are observed (and were expected) for r-structures, since the script was mainly modified to describe DNA structures. However, the source of the error is known and a single change of an average value in the script is required before applying to RNA structures.

Table 3.4.1: Hyperchem (H/c) and script net molecular charges and difference (Diff.) between them.

	H/c	script	Diff.		H/c	script	Diff.		H/c	script	Diff.
da	-1	-0.9527	-0.0473	at	-2	-1.9132	-0.0868	at_at	-4	-3.8264	-0.1736
dc	-1	-0.9396	-0.0604	ac	-2	-1.8923	-0.1077	at_gc	-4	-3.8247	-0.1753
dg	-1	-0.9719	-0.0281	ag	-2	-1.9246	-0.0754	gc_gc	-4	-3.823	-0.177
dt	-1	-0.9605	-0.0395	aa	-2	-1.9054	-0.0946	rau_rau	-4	-3.6292	-0.3708
AT	-2	-1.9132	-0.0868	gt	-2	-1.9324	-0.0676	rau_rgc	-4	-3.8149	-0.1851
GC	-2	-1.9115	-0.0885	gc	-2	-1.9115	-0.0885	rgc_rgc	-4	-4.0006	0.0006
ra	-1	-0.9971	-0.0029	gg	-2	-1.9438	-0.0562				
rg	-1	-1.0163	0.0163	tc	-2	-1.9001	-0.0999				
ru	-1	-0.8175	-0.1825	tt	-2	-1.921	-0.079				
rc	-1	-0.984	-0.016	cc	-2	-1.8792	-0.1208				

3.4.3 Geometry optimisation of small complexes

The input strategy developed so far was tested on an initial optimisation on cisplatin bound to a single-stranded GpG dinucleotide. This system was chosen since there is a crystal structure available in the literature[157] and allows for comparison of the resulting geometry. Furthermore, Scheeff and co-workers have derived force field parameters for cisplatin and oxaliplatin[158], which enables further comparisons to be made with the "soft" parameters for the metal.

The soft metal parameters are expressed in the input as "wildcards". As a result, during the various tests for setting up the input and exploring keyword combinations, the wrong atom types that were assigned by GaussView were not detected immediately, but were described by the same parameters as the metal. This lead into repeating almost all of the calculations that will be further mentioned and details are given only for the complex of cisplatin with GpG. For this complex three geometries (A,B and C) have been obtained (figure 3.4.6):

A: soft parameters for metal, contains atom type-errors on the bases (covered by the soft parameters) and wrong Na position.
B: literature parameters for metal, correct atom types for the rest, correct Na
C: soft for metal, correct atom types for the rest, correct Na.

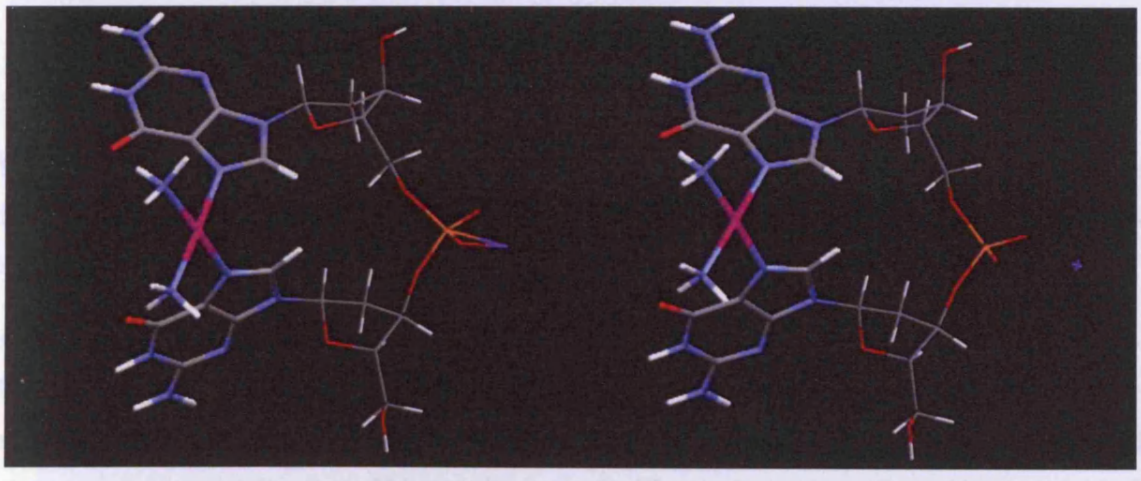


Figure 3.4.6: Structures of A (left) and B(right). Structure of C is visually identical to B.

The extrapolated energy (in hartree) and its components for A,B and C are displayed in table 3.4.2 and the energy differences among them in table 3.4.3:

Table 3.4.2: ONIOM calculated energies of A, B, C (hartree).

	A	**B**	**C**
low model	-1.102633346	-1.044531630	-1.053528404
high model	-1309.951235877	-1309.911817290	-1309.911921
low real	0.187581753	-0.108392267	-0.119015608
extrapolated	-1308.661021	-1308.975678	-1308.977408

Table 3.4.3: Energetic differences (kcal mol^{-1}) between the A-B, A-C and B-C pairs of complexes.

	A-B	A-C	B-C
low model	-36.46	-30.81	5.65
high model	-24.73	-24.67	0.06
low real	185.72	192.39	6.67
extrapolated	197.45	198.53	1.09

Most of the extrapolated energy difference between A and B, C comes from a difference in the MM calculation on the real system (around 186 and 192 kcal mol^{-1} respectively). This indicates that the error originates from the incorrect Na – P and Na – O description in the A complex.

In B, C complexes, the Na – P and Na – O interaction is correctly described only by the non-bonded and coulombic terms of the force field, while in complex A the "soft" parameters were mistakenly applied to describe these Na interactions as bonded (due to the "wildcards" and absence of connectivity list). This results in a significant shortening of the Na-O and Na-P distances, which in turn increase the non-bonded and coulomb interaction. However, further inspection (by hand) of the AMBER energy terms indicates the most of the error comes from the

quadratic bond and angle expressions. This also explains the A-B and A-C non-negligible differences in the "low model" energies, since the same soft parameters were used in all the cases where wrong atom types were detected on the bases.

Also, between the B and C complexes, the MM energy differences on the model and the real system are much lower (around 5.6 and 6.7 kcal mol^{-1}) and are solely because of the different description of the Pt interaction with its neighbouring atoms. This is indicated by the similar magnitude of the "low model" and "low real" energy differences in B-C. This further leads to an acceptable difference in the extrapolated energy, due to the subtractive scheme that is followed in the ONIOM strategy.

For the "high" level, the energy difference between B and C is negligible and this is in agreement with the RMSD for all the atoms between the two structures (table 3.4.4). The high-level differences between the A–B and A–C complexes are likely to originate again from the incorrect description of the Na interactions in the A complex. Since electrostatic embedding is turned on, such effects are incorporated in the Hamiltonian and thus may have affected the QM calculation.

Table 3.4.4: RMSD between the structures of A, B and C (Å).

RMSD	A-B	A-C	B-C
model	0.065	0.064	0.002
real	0.182	0.175	0.010
Na	0.783	0.775	0.023

Overall, the observed differences between B and C are encouraging for the use of soft parameters for the metal, for the complex structure can be reasonably described and not affected by the MM calculation and the energy deficiency of this strategy is accounted for by the

subtractive energy scheme that results in cancellation of most of the error. The errors associated with the complex A suggest that soft parameters be applied with caution, so as to describe only the desired interactions. Incorrect descriptions in the MM part of the calculation may result in non-negligible errors in the QM part as well, especially if charge embedding is applied.

Finally, the following table (table 3.4.5) displays a comparison of the geometry (bonds and angles) around the Pt atom (numbering of atoms shown in figure 3.4.7) between complexes B,C and the crystal structure of cis$Pt(NH_3)_2$-N7GpGN7p. The agreement with the crystal structure is satisfactory.

Table 3.4.5: Comparison of structures B and C to the crystal structure.

	C	B	crystal
Pt-N1	2.031	2.031	2.050(0.036)
Pt-N2	2.033	2.033	2.055(0.045)
Pt-N7A	2.027	2.027	1.968(0.055)
Pt-N7B	2.024	2.025	2.015(0.063)
N7A-Pt-N1	88.52	88.54	89.6(1.3)
N7A-Pt-N2	174.85	174.81	176.8(2.5)
N7A-Pt-N7B	90.85	90.87	88.3(2.2)
N1-Pt-N2	92.78	92.76	91.7(1.01)
N1-Pt-N7B	177.1	177.15	179.9(2.5)
N2-Pt-N7B	87.6	87.58	90.3(1.8)
guan/guan	86.54	86.5	81.2(4.3)

Additionally to the above optimizations, several attempts were made to include microiterations, that are reported to increase the speed of optimisation calculations[159]. The use of microiterations was tested with a series of calculations that involved trying different keywords and a detailed exploration of the ONIOM route of the calculation in order to use non-standard routes in the G03. Details are not shown here, as all of these tests failed. The conclusion is that optimisation with microiterations does not work when using an ECP.

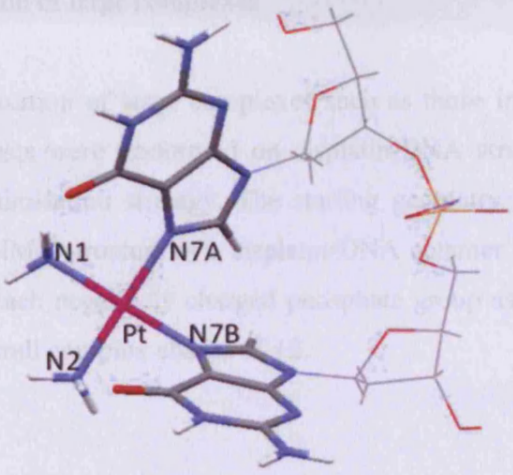


Figure 3.4.7: Atom numbering around the Pt central atom.

Another test for the usefulness of microiterations was optimising a stacked and an H-bonded A-T pair with and without microiterations. The time for each calculation is shown in table 3.4.6, in which it is shown that for the H-bonded pairs no significant gain is observed. On the contrary, employing microiterations for geometry optimization of the stacked pair significantly reduces (almost by a factor of 2) the time length of the calculation. The advantages of using microiterations are more profound in the case of the stacked pair and this fact can be ascribed to the greater flexibility of this complex than that of the H-bonded.

Table 3.4.6: Duration of geometry optimizations with (micro on) and without (micro off) use of microiterations.

	Micro off	Micro on
A-T (H-bonded)	1h,0min,58.4sec	57min,55.6sec
A-T (stacked)	3h,4min,58.9sec	1h,35min,0sec

3.4.4 Geometry Optimisation of large complexes

Full, unconstrained optimisation of large complexes such as those involved in this work is an ambitious task. Several tests were performed on cisplatin-DNA structures in order to decide upon a more efficient optimisation strategy. The starting geometry for the optimisations was taken from Yang *et al.*'s NMR structure of a cisplatin-DNA octamer[160] (pdb entry: 1AU5) after adjusting sodium ions to each negatively charged phosphate group as described in the previous section, resulting in an overall complex charge of +2.

The first testing approach was to freeze parts of the initial octamer during the optimization. In this way the number of degrees of freedom that are allowed to relax is reduced, which in turn can lead to a faster optimization. In the following discussion the terms "dimer", "tetramer", "hexamer" and "octamer" are used to indicate the part of the structure that is optimised each time. For instance, calculations that refer to the "dimer" imply optimisation of the two guanines of the QM level, their corresponding WC pairs and the O3'-O5' part of the backbone that encloses these four bases, with all the rest of the atoms frozen. Similarly the "tetramer" and "hexamer" are defined, while the "octamer" refers to a full unconstrained optimization of the whole complex.

Among these four calculations, only the dimer was fully optimized within a reasonable number of steps and time limit and the rest of the calculations were terminated as inefficient. Table 3.4.7 summarises the number of steps and the initial and final forces for each of the four optimization strategies and figures 3.4.8 - 3.4.11 indicate the more rapid convergence of the dimer optimization compared to rest. Except for the dimer calculation, as final forces we considered the ones of the last step before the termination of the calculation. Therefore, these are not necessarily the lowest forces encountered until the time of termination, keeping in mind the oscillations that occur during an optimization procedure. However the values well indicate that the forces are still far from the desired thresholds.

Table 3.4.7: Number of steps and forces of the optimizations.

	Number of Steps	Maximum Force initial	Maximum Force last	RMS force initial	RMS force last
Dimer	60	0.627936	0.00007	0.032431	0.000013
Tetramer	79	0.626068	0.092286	0.027803	0.010866
Hexamer	62	0.630684	0.015432	0.031304	0.001661
Octamer(full)	329	0.679613	0.022945	0.040829	0.002121

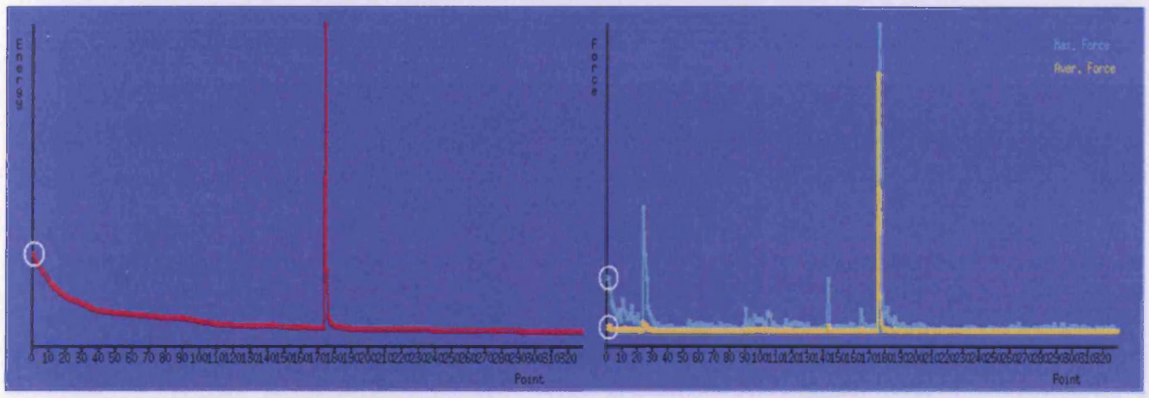


Figure 3.4.8: Octamer energy (left) and forces (right).

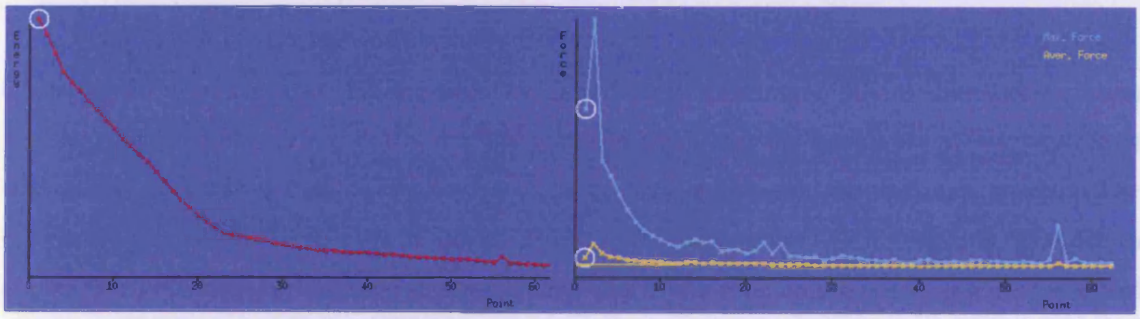


Figure 3.4.9: Hexamer energy (left) and forces (right).

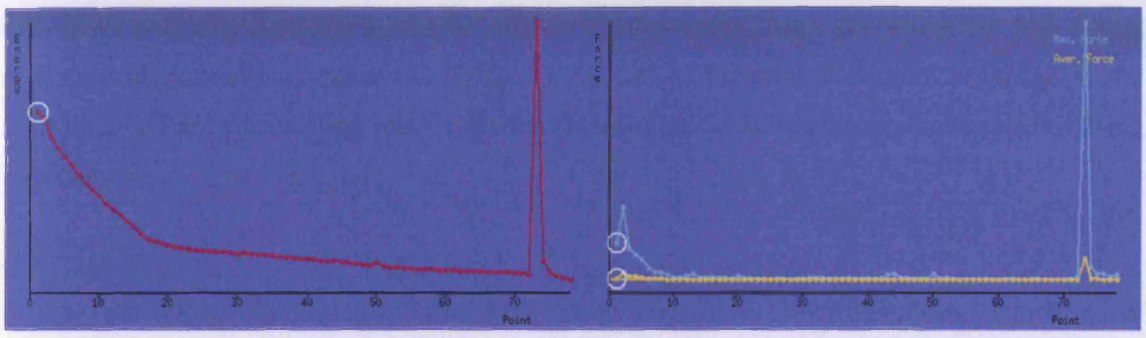


Figure 3.4.10: Tetramer energy (left) and forces (right).

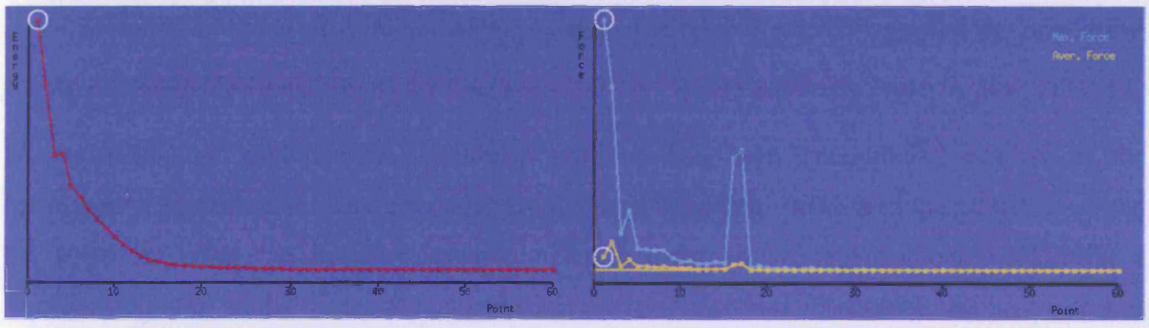


Figure 3.4.11: Dimer energy (left) and forces (right).

Choosing the geometry that resulted from the dimer's optimization, an intermediate optimization step was performed, during which the QM part, its pairs and the backbone atoms were kept frozen (i.e. the atoms that were allowed to relax previously were in this step frozen). The remaining part of the complex was subjected to optimization using microiterations, in order to adjust to the geometry of the central part (dimer). Due to the problems associated with the use of microiterations while using an ECP, the platinum atom was intended to be replaced by a metal atom, which can be described by the all-electron basis set thus avoiding the use of ECP. This replacement, although unrealistic, is expected to lead not to large errors, since the Pt atom is both kept frozen and relatively "buried". The main source of error of such a replacement of atoms is that it induces changes in the atomic charges of the QM region and thus affects the AMBER calculation. A reasonable choice was considered to be the nickel (Ni) atom, however the absence of an atomic radius for Ni in the MK scheme leads to failure of the calculation. Another

alternative that was considered was the choice of magnesium (Mg), for which the MK fitting scheme runs normally and also the charge of +2 can be maintained without a change in the multiplicity. The choice of Mg was finally the one adopted for subsequent calculations involving microiterations.

Having performed this intermediate optimization step using microiterations (which is of negligible computational cost considering the size of the system studied) the complex was subjected to full optimization in order to reach the final optimized structure, using the GDIIS method that is recommended for large systems[161]. The initially calculated forces of this final step were considerably lower than the initial forces shown in table 3.4.7. This implies the usefulness of the intermediate calculation where the microiterations were turned on.

During this final optimization, a further problem has been encountered related to the optimization algorithm. At least two steps are required to be performed until the GDIIS is turned on, so as to define the iterative subspace. During these initial steps without the GDIIS, a significant "jump" of the forces occurs and several steps are required in order to achieve forces at least as low as the initial, as shown in figure 3.4.12. This is leads to significant increase of the required time for the full optimization.

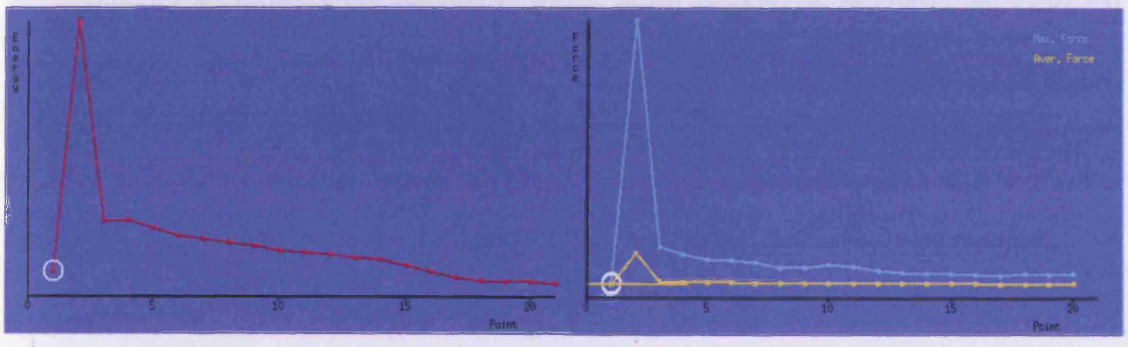


Figure 3.4.12: Jump of energies (left) and forces (right graph).

The second strategy followed, related to the just mentioned problem of the GDIIS method, was to check the performance of different optimization algorithms starting from the geometry

obtained after the microiterations. The complex was initially optimized using different algorithms, so as to check if a large drop of the forces can be achieved at reasonably lower cost and subsequently use these steps to form the iterative subspace for the GDIIS method. The alternatives that were considered were the steepest descent algorithm, the fast solving methods as well as performing the optimization in the Cartesian system, rather than using the internal coordinates. Graphs similar to the previously shown for the three last mentioned options are shown in the following figures (3.4.13 - 3.4.15).

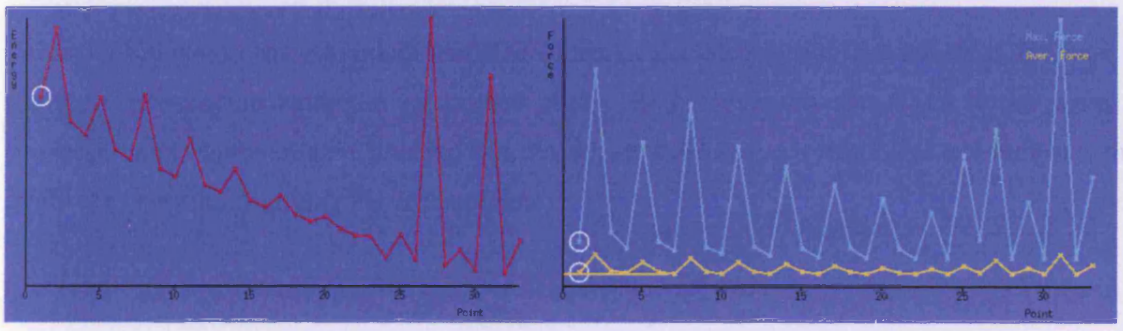


Figure 3.4.13: Changes in energy (left) and forces (right) using the Steepest Descent method.

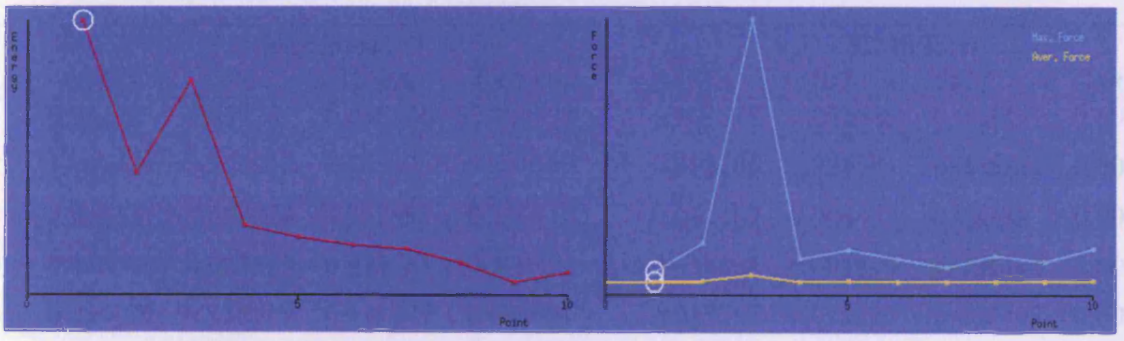


Figure 3.4.14: Changes in energy (left) and forces (right) using the Fast solving methods.

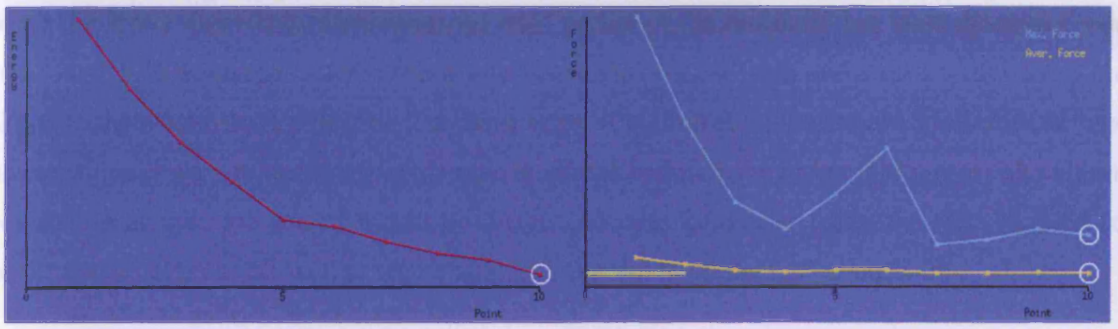


Figure 3.4.15: Changes in energy (left) and forces (right) using the Cartesian coordinates.

Table 3.4.8 displays the maximum and RMS forces calculated in the first ten steps for each of the four optimization strategies mentioned above. Both maximum and RMS forces seem to converge more rapidly when Cartesian coordinates are used and as a result this test indicates that this is the best choice among the four options.

Table 3.4.8: Generated forces for each optimization algorithm (SD:Steepest descent, FSM: Fast solving methods).

	max force				RMS force			
step	GDIIS	SD	FSM	Cartesian	GDIIS	SD	FSM	Cartesian
1	0.03428	0.03428	0.038436	0.035794	0.002649	0.002649	0.00339	0.002645
2	4.162129	0.214912	0.14533	0.021167	0.492708	0.021591	0.006201	0.001712
3	0.59135	0.044424	0.967997	0.010315	0.047107	0.004457	0.02785	0.000888
4	0.475509	0.027559	0.086568	0.006532	0.038806	0.002419	0.004498	0.000604
5	0.390022	0.138269	0.119276	0.011521	0.035267	0.013733	0.007793	0.000801
6	0.3752	0.034371	0.085734	0.017712	0.032146	0.003474	0.005693	0.000798
7	0.343075	0.025016	0.056697	0.00442	0.02824	0.001871	0.004014	0.000448
8	0.274837	0.177964	0.096569	0.005127	0.026012	0.017699	0.004436	0.000444
9	0.267486	0.028307	0.076254	0.006498	0.024118	0.003185	0.00386	0.000642
10	0.319678	0.022791	0.123808	0.005708	0.020949	0.001754	0.006876	0.00052

In all the above described calculations the MM region of the molecule has been described using the standard AMBER charges, while the atoms of the QM region were assigned charges using the MK fitting to the electrostatic potential. The three steps of optimization (constrained, microiterations, final unconstrained) were repeated using alternative charge assignments to the QM atoms. The alternatives considered include the use of Mulliken charges, charges generated using the QEq method (charge equilibration scheme), as well as zero charge (no charge) to all the QM atoms.

Table 3.4.9: Binding energies for the constrained optimized structures (dimers).

	Pt (hartree)	dna (hartree)	Complex (hartree)	kcal mol^{-1}
MK	-230.46654	-1083.468515	-1314.715574083	-489.77
QEq	-230.466855	-1083.460115	-1314.653676212	-456.00
no charge	-230.466543	-1083.468485	-1314.715574380	-489.79
Mulliken	-230.466766	-1083.497928	-1314.729075461	-479.65
no embed	-230.471758	-1083.394233	-1314.277394514	-258.15

The components of the energy for the complex and the DNA are shown in the following table as well.

Table 3.4.10: DNA and complexed DNA ONIOM energies using different charge schemes (in hartree).

	DNA ONIOM energies			Complex ONIOM energies		
	low/model	high/model	low/system	low/model	high/model	low/system
MK	-26.4496	-1104.433688	-5.48445	-26.3764	-1335.68	-5.41171
Qeq	-26.9325	-1104.431098	-5.96156	-27.2226	-1335.68	-6.19632
no charge	-26.4485	-1104.432794	-5.48415	-26.3764	-1335.68	-5.41169
Mulliken	-26.876	-1104.433592	-5.94038	-27.1328	-1335.68	-6.18236
no embed	0.036258	-1077.816455	-5.54152	0.089011	-1308.69	-5.49524

Table 3.4.11: Charges on selected atoms of the QM region.

	MK	MK noembed	mulliken
Pt	0.171244	0.160857	0.284491
H3N (1)	-0.732821	-0.7515	-0.922465
H3N (2)	-0.889663	-0.7515	-0.94539
NB (1)	-0.139075	-0.10434	-0.158574
NB (2)	0.011618	-0.46324	-0.051772
O (1)	-0.634948	-0.51838	-0.52176
O (2)	-0.60246	-0.51174	-0.562731
N* (1)	-0.367278	-0.38494	-0.337882
N* (2)	-0.433121	-0.46324	-0.344192

Table 3.4.12: Base-base interplanar angles (degrees) and Pt-N bond lengths (Å) for the "dimers".

	G1/G2	C1/C2	G1/C1	G2/C2	Pt-N (G1)	Pt-N(G2)
MK	49.58	22.07	7.06	25.56	2.019	2.017
Qeq	50.63	22.05	6.98	25.28	2.016	2.02
no charge	49.57	21.81	7.52	25.53	2.018	2.017
Mulliken	49.94	22.3	7.83	25.1	2.015	2.019
no embed	60.57	13.85	19.2	52.18	2.013	2.021
1AU5	56.85	27.34	16.91	39.69	1.963	1.984

After the initial optimization of the central part (dimer), optimization using microiterations and a Mg atom instead of Pt was performed and subsequently, a full optimization. The full optimization was performed combining the GDIIS method and Cartesian coordinates. This proved to be inefficient as well and the calculations were stopped. For the case of the QEq charges, the final optimization step was repeated using only Cartesian coordinates, which resulted in significantly quicker optimizations.

From the above charge schemes that were examined no clear advantage in either speed or accuracy was evident over use of MK charges. As a result the MK scheme was used, which is as mentioned previously more compatible with the charges of the AMBER force field.

Another set of tests involved full optimizations both with and without electrostatic embedding. For these final tests the MK charges and cartesian coordinates were used. The starting structures for these optimizations were again the ones obtained after the intermediate steps of microiterations. The use of Cartesian coordinates for the optimization leads to rapid convergence of the RMS forces below the default threshold value of the software. The same is not observed for the maximum forces, therefore full optimization is not reached strictly, in terms of the default cutoffs. The energies gradually drop for several steps and result in a trend to reach a plateau. At this point the energies start to oscillate as a consequence of oscillation of the maximum forces, as seen in figure 3.4.16. The oscillation of the energy values is in a small scale and taking into account the size of the systems and the long-converged RMS force the obtained geometry is not expected to be altered further if the optimization is continued.

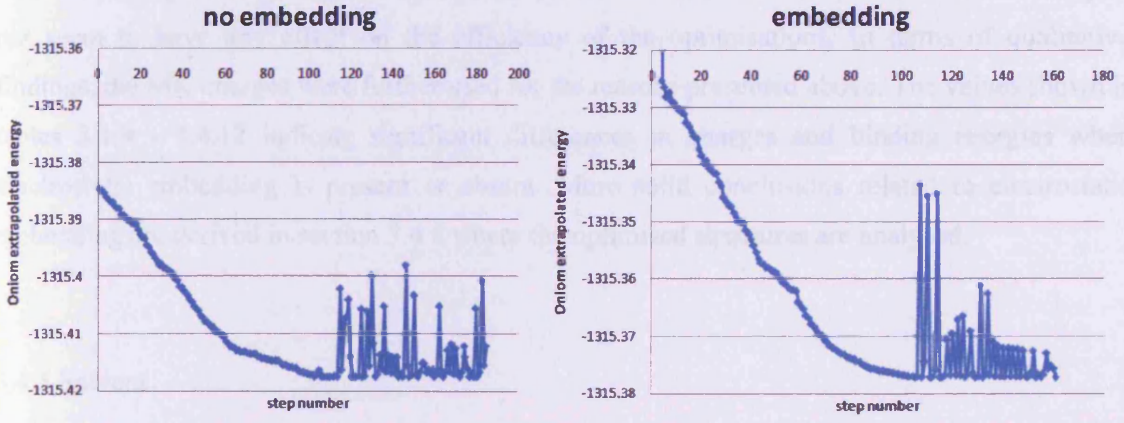


Figure 3.4.16: Observed oscillations during optimizations using Cartesian coordinates.

Another final approach was to reach a fully optimised structure by gradually building the octamer. In other words, the central dimer as defined earlier in this section was fully optimised in the absence of any additional base pairs. Accordingly, pairs of bases were added manually using the Avogadro software. The newly added base pairs were initially allowed to relax and adjust to the dimer using microiterations before another full optimisation was carried out. The process was repeated until the desired number of base pairs was reached. Apparently, such an approach is not efficient in terms of time and was soon abandoned.

Summarising the conclusions of the present section, the use of Cartesian coordinates could be the best choice among the options that were tested, although full convergence is not met. However, use of Cartesian coordinates is not compatible with atom freezing and the problems associated with the GDIIS method can be mediated by controlling the step size and without updating the trust radius at each step. Such optimisations typically proceeded reasonably smoothly although the large displacements at the early stages of the optimisation are not eliminated. Additionally, access to more powerful computing resources at this stage aided the performance of such large scale optimisations. As a result, further calculations were performed using the GDIIS method.

The different QM charge alternatives and the presence or absence of electrostatic embedding do not seem to have any effect on the efficiency of the optimisations. In terms of qualitative findings, the MK charges were further used for the reasons presented above. The values shown is tables 3.4.9 - 3.4.12 indicate significant differences in charges and binding energies when electrostatic embedding is present or absent. More solid conclusions related to electrostatic embedding are derived in section 3.4.6 where the optimised structures are analysed.

3.4.5 Solvent

A more realistic description of the systems necessitates the use of explicit solvent molecules. The original structure that has been used so far (1AU5) was loaded into Hyperchem in order to build a solvated model using a periodic box. Water was selected as a solvent and the dimensions of the box were set at a large value (x=y=z=56.104Å), which ensured full solvation of the input

structure by thousands of water molecules. Even though the solvent would be described at the MM level, such a large number is impractical for the calculation. A script was written to truncate the solvated model to a smaller size by using a cutoff solute-solvent distance, above which solvent molecules would be ignored. Smaller scale solvation can as well be achieved by selecting a smaller and non-cubical periodic box in advance. This is less convenient since the size of the solute in three directions has to be taken into account each time, while using the script the size of solvent is irrelevant and the truncation is achieved uniformly around the solute.

Two different solvated structures were prepared by this procedure; for the first a cutoff distance of 2.7Å was used resulting in 50 water molecules (solv50) and for the second the cutoff and number of solvent molecules were increased to 3Å and 90, respectively (solv90). For both solvated structures the water molecules and sodium atoms were initially allowed to relax using microiterations prior to a full optimisation, by freezing the DNA-drug part and replacing the Pt atom by Mg. Finally, full geometry optimisation was carried out for both models following the procedure that was finally determined in the previous paragraph. An image of a solvated structure (solv90) is shown in figure 3.4.17 and the results of the optimisations are discussed in the following section.

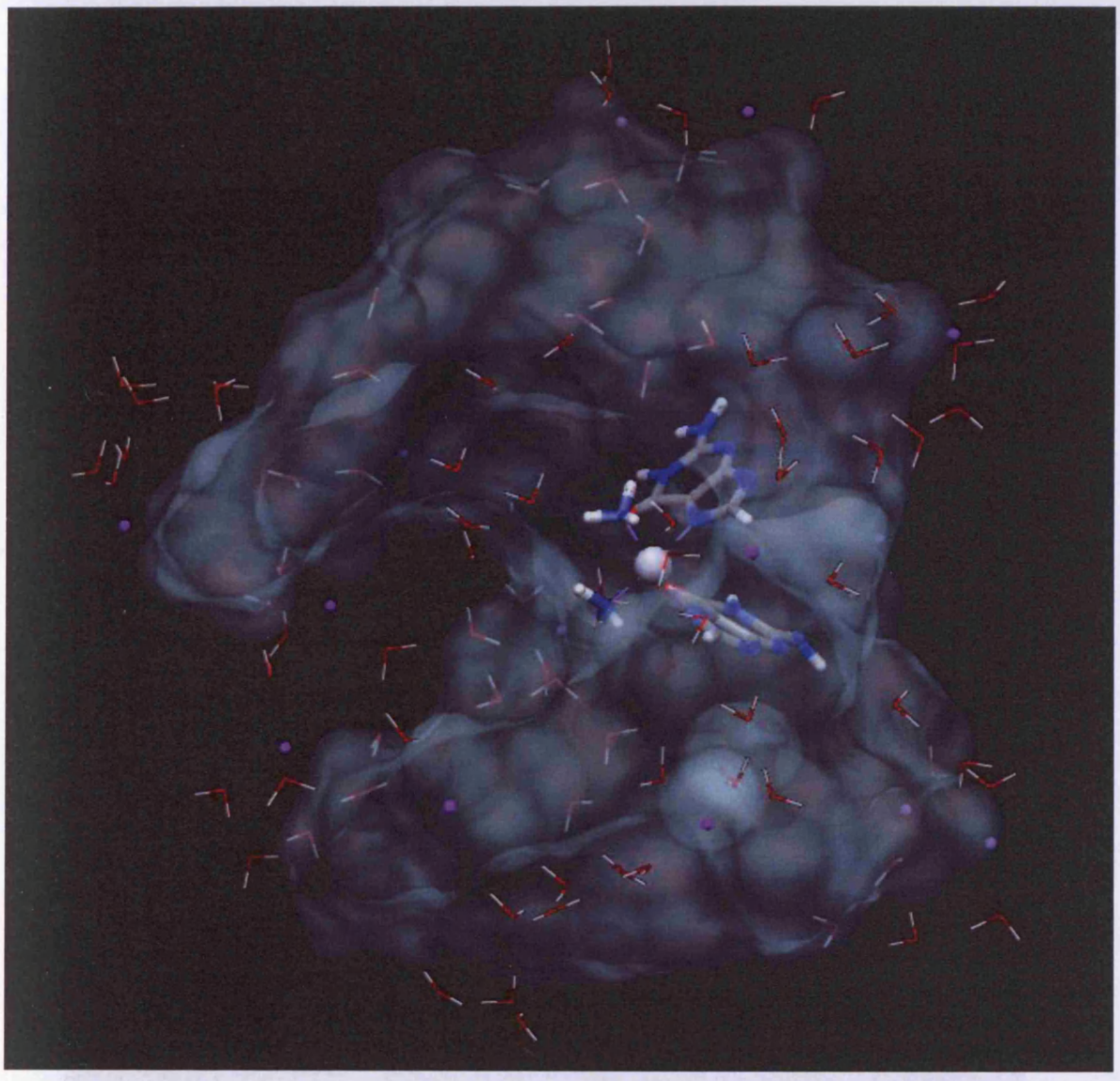


Figure 3.4.17: Optimised structure of the solvated cisplatin-DNA structure (solv90). The Pt atom is shown as white sphere and the rest of the QM region as capped sticks; Na^+ ions shown as purple spheres and solvent molecules as wireframe. The MM part has been replaced by a solid surface of 60% transparency for clarity.[162,163]

3.4.6 Results – Curves analysis

The primary test of the suitability of the optimised geometries obtained with various methods was their agreement with the solution phase NMR structure, as reported in the PDB entry 1AU5. To allow proper comparison, Curves analysis was performed on the NMR structure and on each optimised one. Figure 3.4.18 and tables 3.4.13, 3.4.14 display the NMR structure and its Curves data, respectively. The following collections of graphs (figures 3.4.19 and 3.4.20) display comparatively the differences from the NMR structure of the above mentioned parameters. Our interest lies particularly in the QM region, *i.e.* G4 and G5.

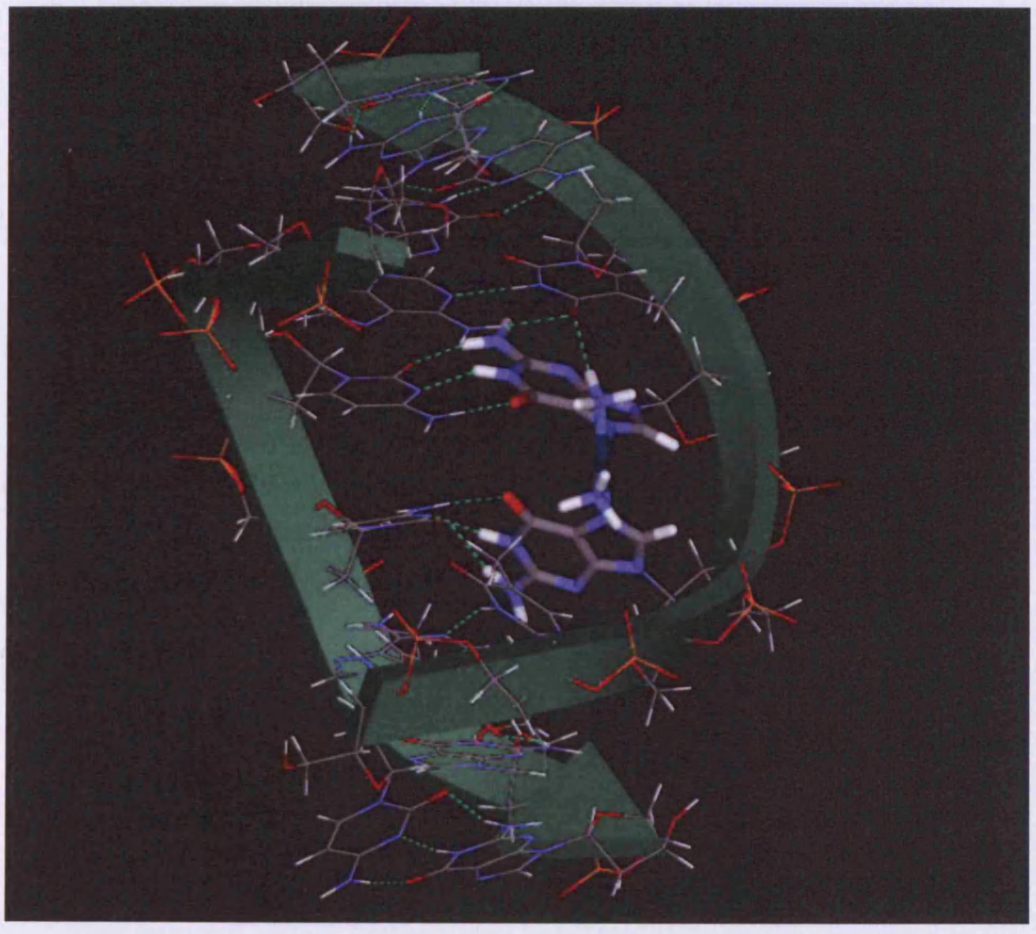


Figure 3.4.18: NMR structure of 1AU5.

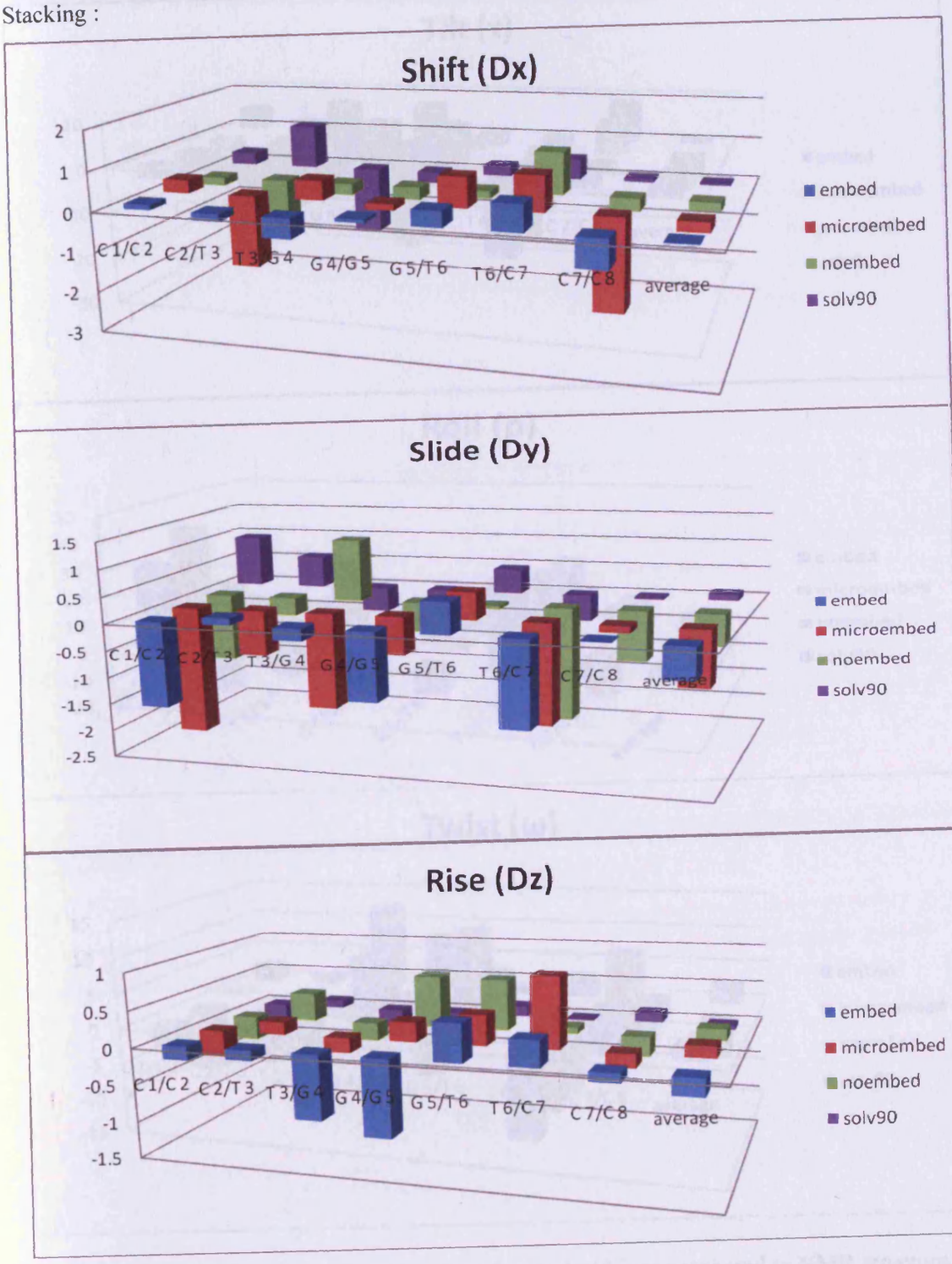
Table 3.4.13: Local inter-base pair parameters for 1AU5 (Dx,Dy,Dz (Å), τ,ρ,ω(degrees)).

Local Inter-Base pair parameters						
1st strand	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (τ)	Roll (ρ)	Twist (ω)
C1/C2	0.7	0.04	4.32	7.09	-56.78	44.64
C2/T3	1.35	-1.91	3.94	1	1.6	43.69
T3/G4	-0.03	0.06	2.51	22.9	-6.94	25.73
G4/G5	1.2	-1.55	5.5	-13.18	58.72	23.24
G5/T6	0.86	-1.53	2.78	11.18	-5.36	42.22
T6/C7	0.35	-0.32	2.57	15.53	-24.76	34.08
C7/C8	0.55	-2.12	4.17	-11.89	7.45	33.6
2nd strand	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (τ)	Roll (ρ)	Twist (ω)
G8/G7	-1.08	-0.89	3.49	-0.66	-21.93	37.68
G7/A6	-0.13	-1.98	3.78	8.71	0.86	29.7
A6/C5	-2.02	1.09	3.73	-6.77	14.05	33.69
C5/C4	0.57	-0.03	3.73	-12.66	20.71	18.36
C4/A3	-0.34	-1.1	2.53	-10.25	14.61	17.27
A3/G2	-0.55	-0.69	3.6	5.07	-18.9	45.51
G2/G1	0.23	-2.04	3.55	0.45	-20.39	36.48

Table 3.4.14: Global base-base pair parameters for 1AU5 (Sx,Sy,Sz(Å),κ,ω,σ (degrees)).

	Shear (Sx)	Stretch (Sy)	Stagger (Sz)	Buckle (κ)	Propel (ω)	Opening (σ)
C1-G8	-0.16	0.22	0.13	12.57	9.29	-1.09
C2-G7	0.87	-0.34	0.28	8.94	-36.47	-3.64
T3-A6	-0.27	0.14	0.82	-22	-29.03	9.53
G4-C5	0.02	-0.28	-1.05	-12.37	-41.51	5.64
G5-C4	-0.94	-0.14	0.31	-17.2	1.4	-3.98
T6-A3	0.11	0.2	0.6	2.31	-10.04	19.99
C7-G2	0.34	-0.13	-0.55	4.31	-18.25	10.13
C8-G1	0.33	0.31	0.26	-10.42	11.64	6.57
Average:	0.04	0	0.1	-4.23	-14.12	5.39

Stacking :


Shift (Dx)
2
1
0
-1
-2
-3
C1/C2
C2/T3
T3/G4
G4/G5
G5/T6
T6/C7
C7/C8
average
embed
microembed
noembed
solv90
Slide (Dy)
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
C1/C2
T3/G4
G5/T6
T6/C7
C7/C8
average
embed
microembed
noembed
solv90
Rise (Dz)
1
0.5
0
-0.5
-1
-1.5
C1/C2
C2/T3
T3/G4
G4/G5
G5/T6
T6/C7
C7/C8
average
embed
microembed
noembed
solv90

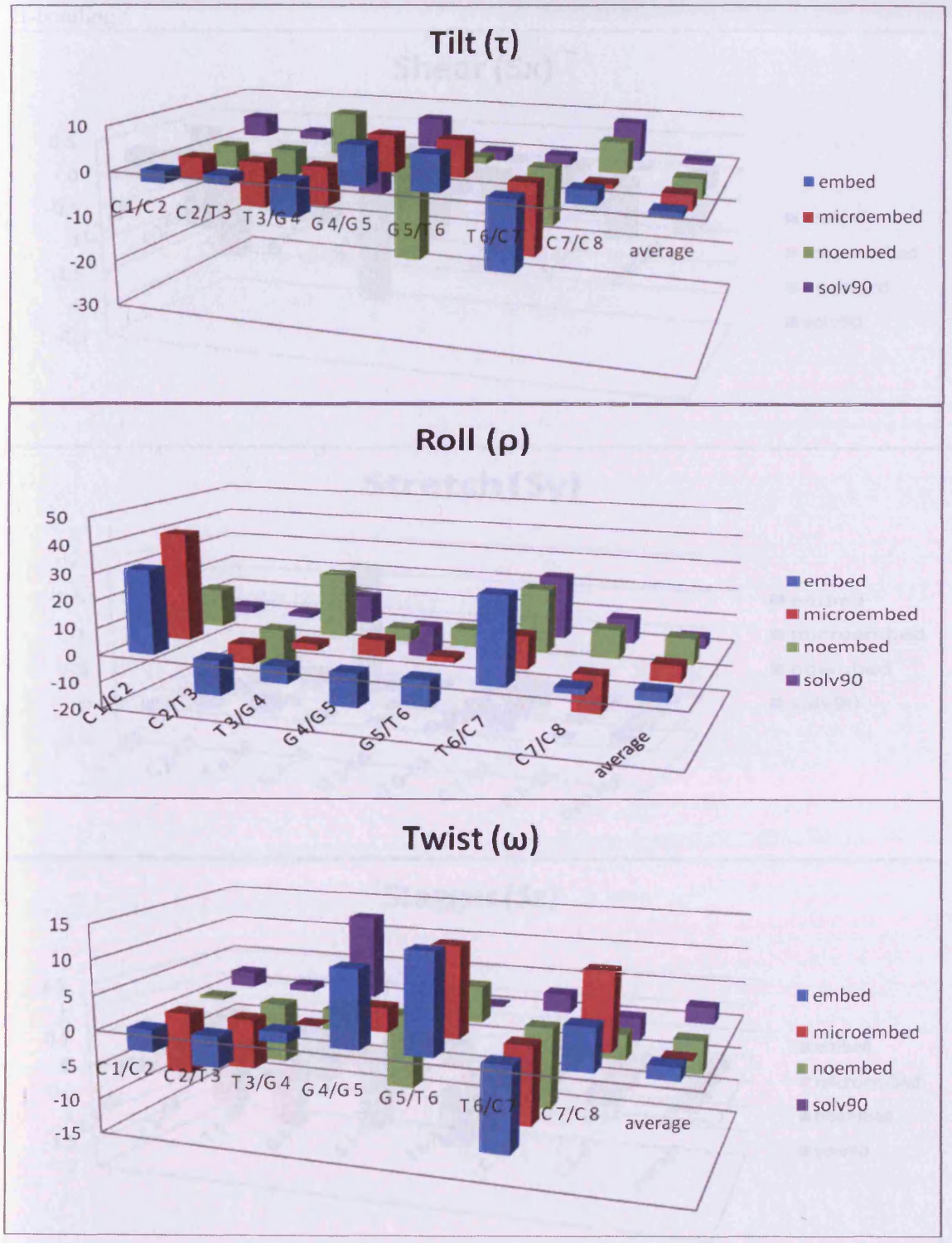
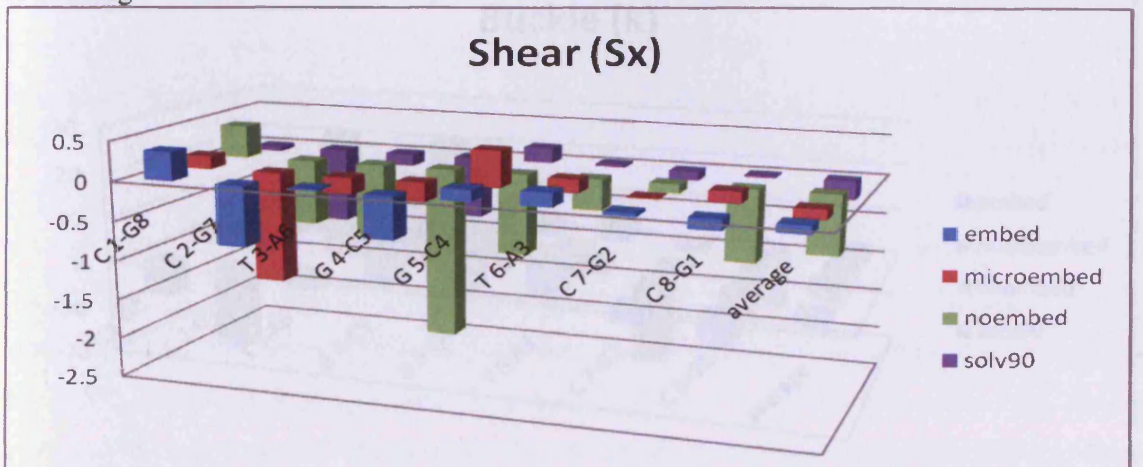
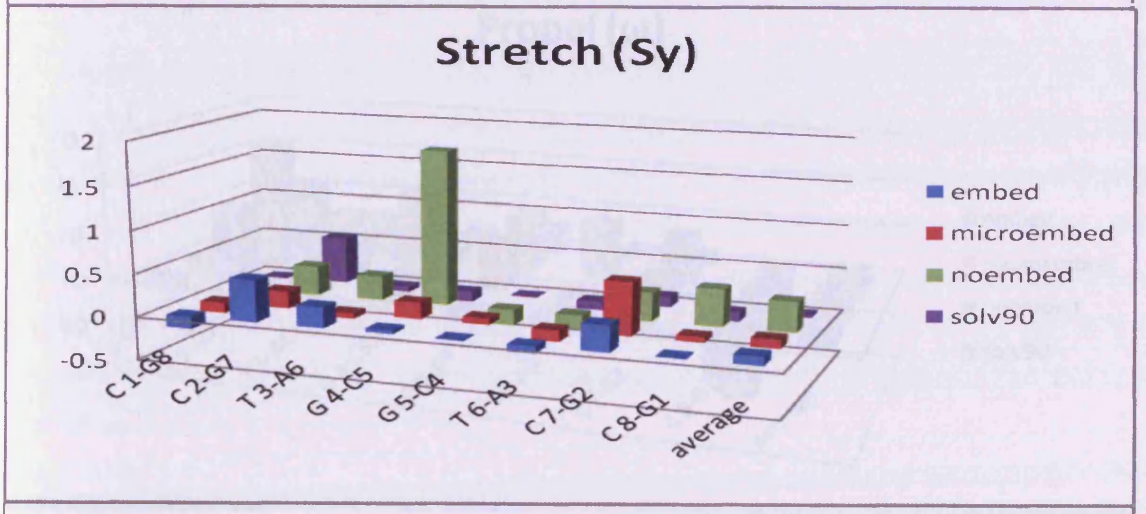
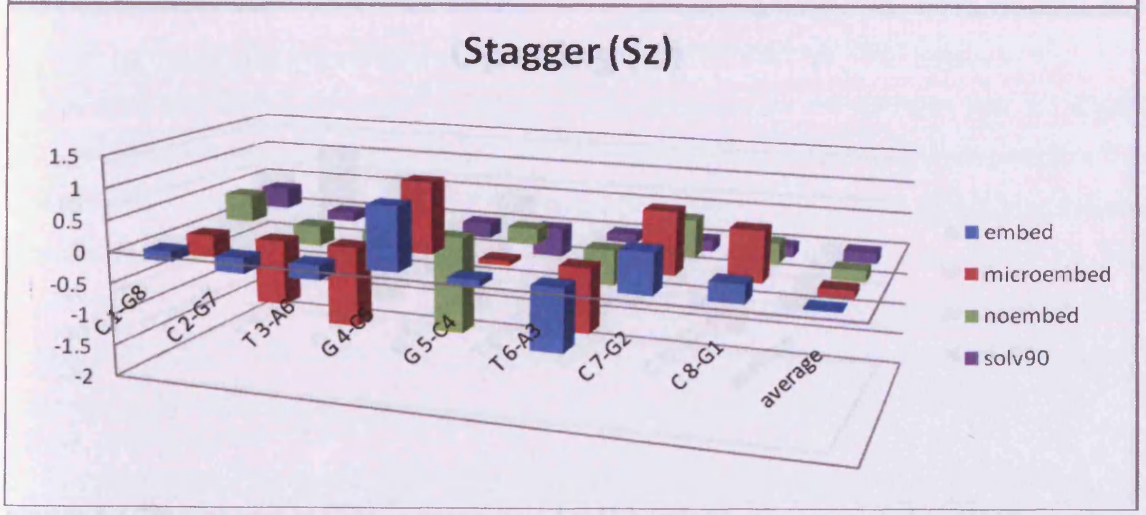


Figure 3.4.19: Changes in Curves parameters for stacked bases compared to NMR structure.

H-bonding :


Shear (Sx)
0.5
0
-0.5
-1
-1.5
-2
-2.5
C1-G8
C2-G7
T3-A6
G4-C5
G5-C4
T6-A3
C7-G2
C8-G1
average
embed
microembed
noembed
solv90


Stretch (Sy)
2
1.5
1
0.5
0
-0.5
C1-G8
C2-G7
T3-A6
G4-C5
G5-C4
T6-A3
C7-G2
C8-G1
average
embed
microembed
noembed
solv90


Stagger (Sz)
1.5
1
0.5
0
-0.5
-1
-1.5
-2
C1-G8
C2-G7
T3-A6
G4-C5
G5-C4
T6-A3
C7-G2
C8-G1
average
embed
microembed
noembed
solv90

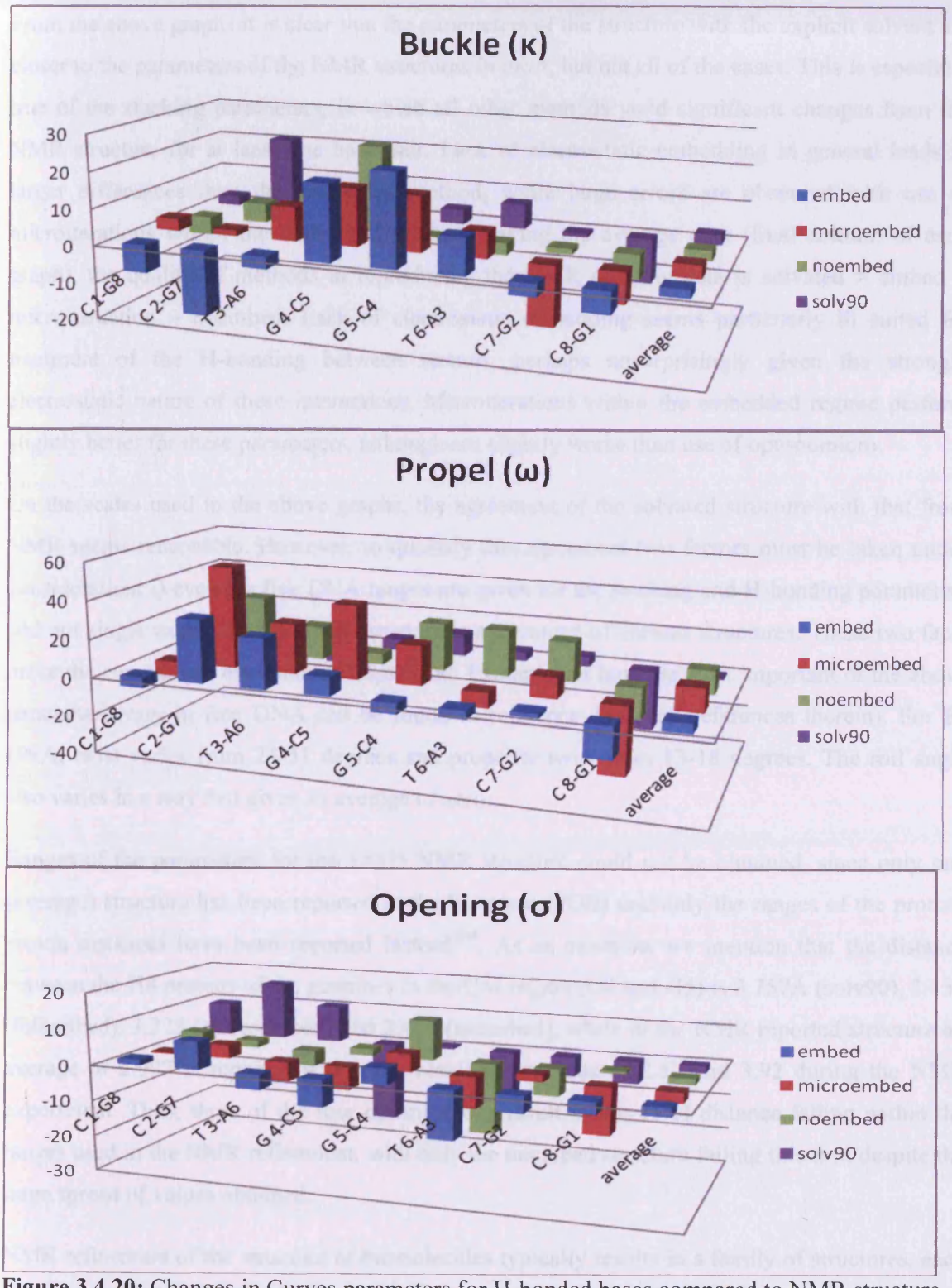


Figure 3.4.20: Changes in Curves parameters for H-bonded bases compared to NMR structure.

From the above graphs it is clear that the parameters of the structure with the explicit solvent are closer to the parameters of the NMR structures in most, but not all of the cases. This is especially true of the stacking parameters, in which all other methods yield significant changes from the NMR structure for at least one base pair. Lack of electrostatic embedding in general leads to larger differences than the embedded method, while large errors are observed with use of microiterations within the embedded scheme. Taking the average data (final column of each graph), the quality of methods in reproducing the NMR stacking data is solvated > embed > microiteraction ~ noembed. Lack of electrostatic embedding seems particularly ill suited for treatment of the H-bonding between strands, perhaps unsurprisingly given the strongly electrostatic nature of these interactions. Microiterations within the embedded regime perform slightly better for these parameters, although are slightly worse than use of opt=nomicro.

On the scales used in the above graphs, the agreement of the solvated structure with that from NMR seems reasonable. However, to quantify this agreement two factors must be taken under consideration: i) even for free DNA ranges are given for the stacking and H-bonding parameters and not single values, ii) the NMR structure is an average of various structures. These two facts make the comparison even more complicated. Examples of how the most important of the above parameters vary in free DNA can be found in reference 118 (and references therein). For B-DNA, twist varies from 24-51 degrees and propeller twist from 13-18 degrees. The roll angle also varies in a way that gives an average of zero.

Ranges of the parameters for the 1AU5 NMR structure could not be obtained, since only one (average) structure has been reported in the literature (PDB) and only the ranges of the proton-proton distances have been reported instead[164]. As an example, we mention that the distance between the H8 protons of the guanines in the QM region (G4 and G5) is 3.757Å (solv90), 3.134 (fullembed), 3.228 (microembed) and 2.474 (noembed), while in the NMR reported structure an average of 2.907 is reported within the obtained extremes of 2.52 and 3.92 during the NMR experiment. Thus, three of the four optimisations result in this H···H distance falling within the ranges used in the NMR refinement, with only the noembed structure failing this test, despite the large spread of values obtained.

NMR refinement of the structure of biomolecules typically results in a family of structures, each of which satisfies the proton-proton distance ranges found in the NMR experiment. In the

original paper reporting the 1AU5 structure, an unspecified number of such structures were obtained but only one (presumably that deposited in the PDB) shown. In order to gain further insight into the structural variations that can be observed in an NMR experiment, another structure from the literature (pdb entry : 2NQ0) was selected.[165] In this case, 15 different structures have been added in the PDB database that correspond to an intra-strand binding of cisplatin to a DNA dodecamer, each of which has been analysed using Curves5.1 It should be noted that this PDB entry corresponds to both a different DNA sequence and different experimental conditions, especially temperature, to those used in entry 1AU5. However, this analysis should at least provide us with a context in which to place our optimisation results. Tables 3.4.15 and 3.4.16 display the minimum and maximum values encountered in the analysis of the previously mentioned parameters, thus providing an idea of expected ranges in these values.

Concentrating on the platinated guanine residues (G6 and G7 in this structure), this analysis shows that ranges of 0.5 to 1.0 Å in shift, slide and rise, and up to 10 ° in tilt, roll and twist are found within the family of NMR structures (table 3.4.15). Ranges are smaller within the H-bonding parameters, with changes of 0.2 to 0.5 Å and 2 to 8 ° (table 3.4.16). Similar results are obtained for the rest of the bases.

Table 3.4.15: Minimum and maximum values of local inter-base pair parameters for the 2NQ0 structure.

	shift		slide		rise		tilt		roll		twist	
	min	max	min	max	min	max	min	max	min	max	min	max
G6/G 7	0.67	1.12	-2.06	-0.94	3.9	4.61	-3.07	-1.39	24.44	34.45	19.75	28.2

Table 3.4.16: Minimum and maximum values of Global base-base pair parameters for the 2NQ0 structure.

	shear		stretch		stagger		buckle		propel		opening	
	min	max	min	max	min	max	min	max	min	max	min	max
G 6-C 7	-0.56	-0.22	-0.12	0.08	-0.11	0.05	8.51	18.14	-14.12	-8.71	2.41	4.1
G 7-C 6	-0.58	-0.11	0.01	0.38	0	0.11	-5.28	-0.17	-2.84	6.05	6.42	9.39

3.5 QTAIM analysis on large metal-DNA systems

This final section of the present chapter is related to the performance of QTAIM analysis of truncated models from larger complexes. In order to perform QTAIM analysis a wavefunction is necessary to be available for the whole system; therefore the computational cost is prohibitively large for large scale DNA systems. The question that arises is whether the truncation of a system to a smaller model has a significant effect on the BCP's of interest.

For this reason four truncated models of the 1AU5 crystal structure were prepared. The first (structure I) is shown in figure 3.5.1 and consists of the two WC-GC pairs to which cisplatin is bound. For discussion purposes the WC pair and the cisplatin ligand that lie at the top of figure 3.5.1 are denoted as "A" and the WC pair and cisplatin ligand that are at the bottom are denoted as "B". Structures II and III consist of the model with an additional WC pair (TA) adjacent to A and B, respectively; therefore each of II and III consists of three WC pairs. Finally, structure IV consists of four WC pairs, i.e. it is a combination of II and III. In all structures the sugar-phosphate backbone has been removed and the relevant bonds have been saturated by H atoms. Simple sketches of structures I-IV are displayed in figure 3.5.2. Single point calculations on these structures were performed at the BHandH/6-31+G**/SDD level.

Figure 3.5.1: Truncated model from the 1AU5 structure.

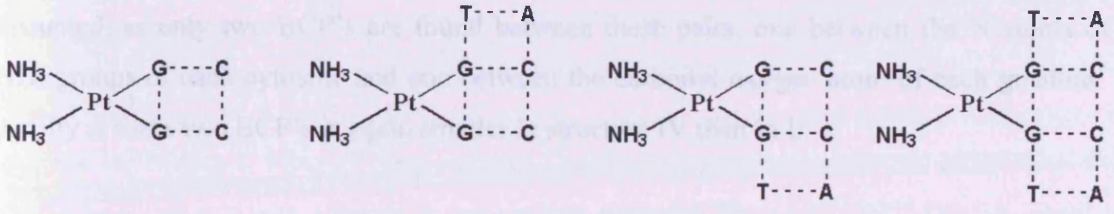


Figure 3.5.2: Schematic view of calculated structures I - IV.

Interactions involving cisplatin which are present in all four structures I-IV are the two Pt-N covalent bonds, as well as two BCP's between the Pt atom and the exocyclic O atoms of the two guanines Pt s bound to. The density of these two types of BCP's for each structure is shown in table 3.5.1.

Table 3.5.1: Electron density at selected BCPs.

	I	II	III	IV
Pt-N7 (A)	0.1291	0.1294	0.1286	0.1289
Pt-N7 (B)	0.1365	0.1366	0.1363	0.1363
Pt-O (A)	0.0084	0.0083	0.0084	0.0084
Pt-O (B)	0.0103	0.0103	0.0101	0.0101

From the above table it is evident that between structures I and IV there is a slight drop of all densities when four WC pairs are included. In the case of three WC pairs (II and III) no definite observation can be made. When adding a pair next to "B" (structure III) again a small drop on the values of Pt-N7 (B) and Pt-O (B) is observed, but this is not the case for structure II, in which a similar drop of the value Pt-N7 (A) is not observed. However, the structures are not symmetrical and subtle geometrical effects may have an effect on the resulting values. Overall, the differences observed are of rather small magnitude and at this level there is essentially no change. As a result, the smaller model (I) is adequate for such analysis. It is also worth mentioning that in all four structures the stacking between the "A" and "B" GC pairs seems disrupted, as only two BCP's are found between these pairs: one between the N atoms of the NH_2 groups of each cytosine and one between the carbonyl oxygen atom of each guanine. The density at these two BCP's is again smaller in structure IV than in I.

On the other hand, apart from the numerical effects, there is a potential risk of missing interactions when the model is truncated. In the present example the addition of each base pair to structure I reveals one more kind of interaction. This interaction involves an H atom of the cisplatin carrier ligands and its spatially neighbouring additional base. For example, in structure IV both NH_3 groups of cisplatin are involved in an interaction with the external base pairs: ligand A participates in a N-H···C interaction with thymine and ligand B in a N-H···O interaction with thymine. The density at the corresponding BCP's is 0.0057 and 0.0262 a.u. respectively. The additive effect of such interactions may not be negligible, especially if ligands bulkier than those of cisplatin are present.

4. Applications

In this chapter the methods and methodologies that were explored previously are employed to investigate systems of chemical interest. Such systems mostly involve metal-DNA complexes of varying sizes (sections 4.1 – 4.4) as well as non-metallic systems, such as the distortions induced on DNA by modified bases (section 4.5).

4.1 Arene Complexes of Ruthenium(II) and Osmium(II): Role of Hydrogen Bonding and π-stacking.

Recently developed ruthenium(II) complexes of the type $[(\eta^6\text{-arene})Ru^{II}(en)\text{-}Cl][PF_6]$ (en = ethylenediamine) show promising anticancer activity,[51] comparable to or better than platinum anticancer complexes, as mentioned in the introduction. Chemical and spectroscopic studies[166,167] reveal a stronger preference for guanine than was observed with platinum-based drugs, with experimental NMR data suggesting that hydrogen bonding to guanine could explain the observed preferential binding.

The size of the arene ligand strongly affects the activity;[51] arenes containing two or more rings, such as, for example, biphenyl (bip) or tetra- or dihydroanthracene (tha and dha), have crystal structures[166] that reveal π-stacking interactions between the arene and base, in addition to hydrogen bonds. Hence, the ruthenium(II) arene complexes can be considered as potential intercalators.[166,168] Experiments on DNA duplexes with ruthenium(II) arene complexes showed that the tha complex has a cancer cell cytotoxicity approximately 20 times higher than that of a ruthenium cymene complex.[169] It is important to note, however, that while DNA binding is believed to be the mechanism of action, other factors such as cell uptake will affect the cytotoxicity and could be strongly dependent on the nature of the arene. Studies on DNA duplexes reveal the importance of the arene ligand, which distorts the DNA duplex through intercalation in the biphenyl complex (although partially saturated arenes such as tha and dha may not be "true" intercalators) and through steric interactions in the cymene complex.[170] The

same study showed that the distortion caused by the cymene complex was extended to more nucleic acid pairs than the distortion caused by the tha complex.

These complexes were studied by a combined DFT and QTAIM analysis in a manner as described in chapter 3. In all cases the SDD basis set was used for the Ru atoms and the 6-31+G** on all other atoms. The aqueous solution of complexes and their separated moieties was estimated using the polarizable continuum model (PCM) approach that was described in chapter 2, employing a dielectric constant of 78.39 and atomic radii from the UFF force field.[171] In this way, the binding energy corrected for hydration effects was estimated. Finally, DF-LMP2 and SCSN calculations have been performed using Dunning's aug-cc-pVTZ basis set.[63]

4.1.1 Structures of $[Ru(en)XY]^{2+}$ complexes

Figure 4.1.1 shows the nucleobases and arenes considered, along with atom numbering. The N9-methyl purines and N1-methyl pyrimidines are referred to simply as G, I, A and C, T respectively. The A rings of bip, dha, and tha are those η^6-coordinated to the Ru atom. In addition to these arenes, which have been studied experimentally,[166,167] we have also examined the parent anthracene molecule, which lacks the flexibility of dha and tha, and is therefore not expected to bind as tightly or undergo π-stacking to the same extent.

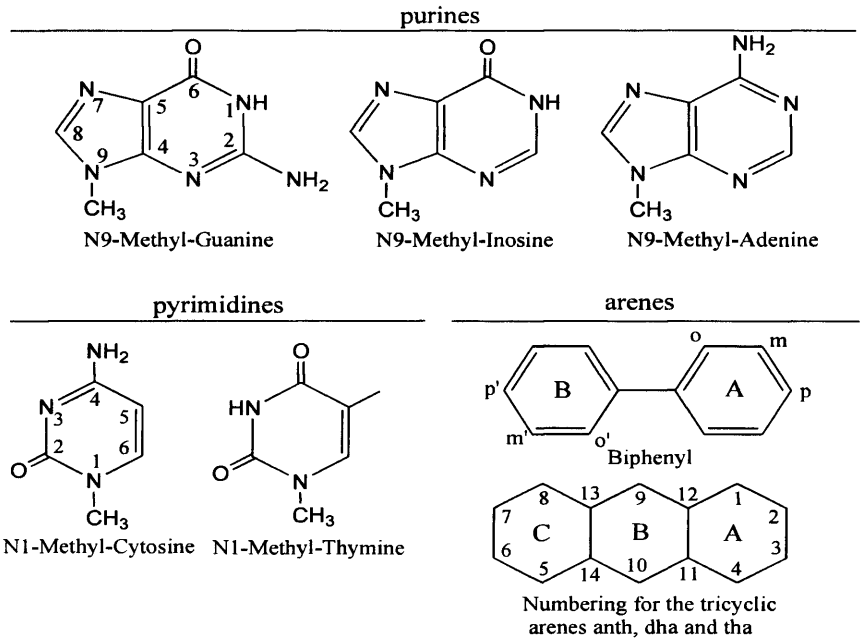


Figure 4.1.1: Arenes and bases studied, including numbering scheme.

Optimized structures of selected complexes are displayed in Figure 4.1.2. These geometries show both covalent binding to Ru and N—H...X hydrogen bonds between the en ligand (donor) and O or N atoms of the bases (acceptors). These H-bonds are ubiquitous throughout all complexes studied, with the O6 of G and I, N6 of A, O2 of C and O2 and O4 of T acting as acceptors to the en N—H donor. Given the previous success of BHandH in describing π-stacking,[28,97,98] it is encouraging to note the arene ligands bip, dha, tha are positioned over the nucleobases, in a similar manner to the reported crystal structures.[166] Table 4.1.1 summarizes some key geometrical features of the optimised structures, and where possible compares these with the crystal structures reported in reference 166. In general the agreement is good, with Ru—N_{base} and Ru—N_{en} distances within 0.02Å of experimental values and Ru-arene distances approximately slightly longer than experimental data. The tendency of BHandH to over-estimate H-bond strengths (and hence underestimate lengths by *ca.* 0.1Å) is apparent from these comparisons, but this is a systematic error and so should have less effect on the trend in binding energies. The mutual orientation of arene and guanine rings is well reproduced, both in terms of the separation between mean planes (R) and the angular orientation of the arenes over the

nucleobases. The latter is described by the hinge angle (θ) on the C9-C10 atoms for the anth, dha and tha ligands and the arene - base interplanar angle (φ).

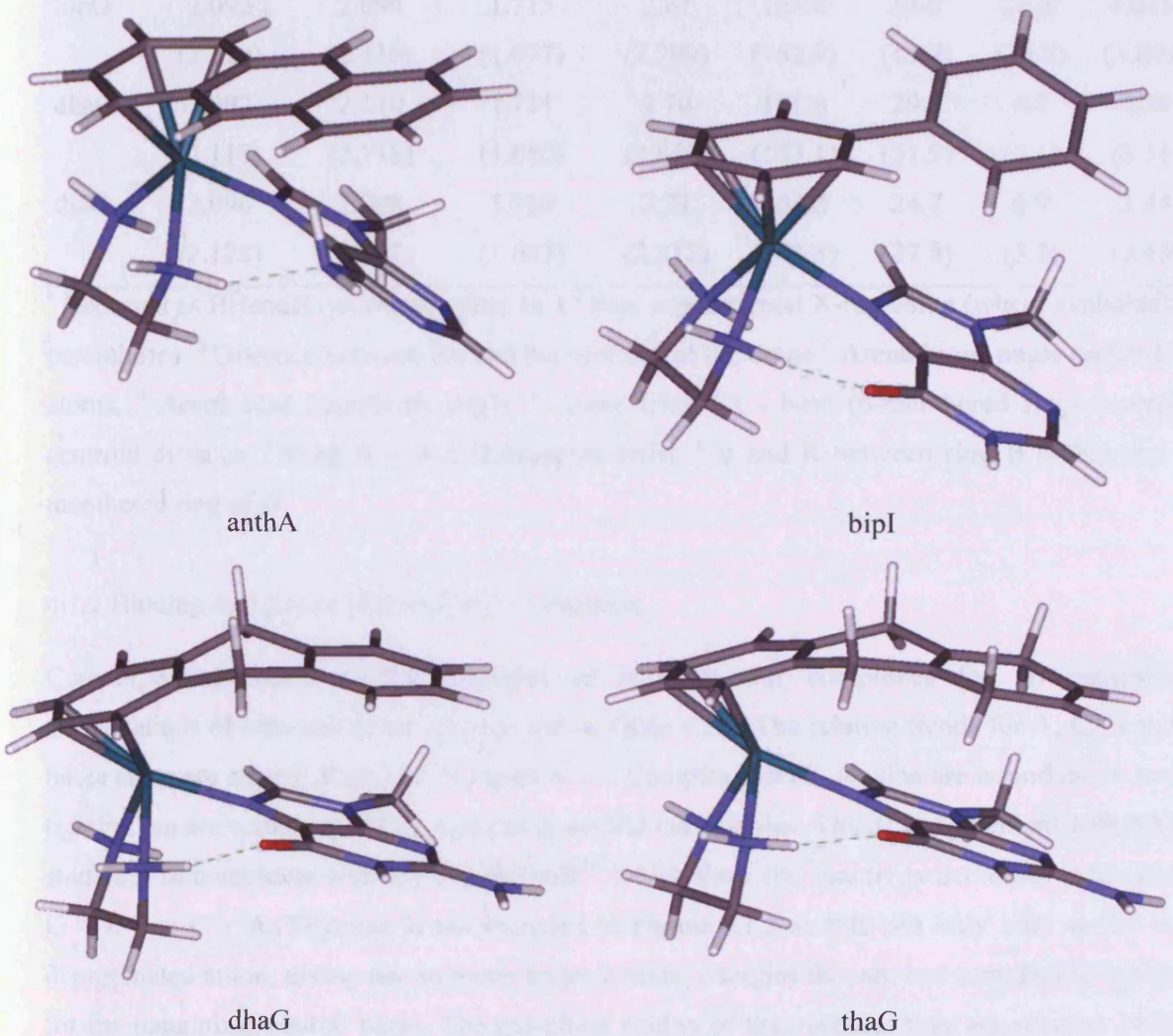


Figure 4.1.2: Optimized geometries of selected compounds.

Table 4.1.1: Selected optimized geometrical parameters and comparison with experiment[a].

	Ru—N_{base} (Å)	Ru—N_{en} (Å)	Ru—arene[b] (Å)	N...X (Å)	N-H...X (deg)	θ[c] (deg)	φ[d] (deg)	R[e] (Å)
benG	2.094	2.097	1.707	2.68	162.8	-	29.6	-
anthG	2.095	2.083	1.750	2.70	162.8	0.0	28.9	5.095
bipG	2.092	2.096	1.715	2.67	164.8	29.6[f]	24.8[g]	4.011[g]
	(2.120)	(2.116)	(1.677)	(2.799)	(162.9)	(13.7)	(23.8)	(3.801)
dhaG	2.092	2.110	1.711	2.70	161.8	29.5	4.7	3.39
	(2.117)	(2.118)	(1.680)	(2.840)	(163.1)	(31.9)	(3.1)	(3.31)
thaG	2.096	2.098	1.710	2.71	162.2	24.7	6.9	3.34
	(2.128)	(2.127)	(1.683)	(2.812)	(163.3)	(27.8)	(3.3)	(3.45)

[a] Reported as BHandH optimised value on 1st line, experimental X-ray value (where available) in parentheses. [b] Distance between Ru and the centroid of the arene [c] Arene hinge angle on C9-C10 atoms. [d] Arene-base interplanar angle. [e] Arene (ring C) - base (6-membered ring) centroid-centroid distance. [f] Ring A – ring B propeller twist. [g] φ and R between ring B of bip and 5-membered ring of G.

4.1.2 Binding energies of $[Ru(en)XY]^{2+}$ complexes

Counterpoise-corrected binding energies of bases to Ru complexes for all considered combinations of base and arene are reported in Table 4.1.2. The relative trends for A, C, G and I bases are more clearly displayed in Figure 4.1.3. Complexes with guanine are bound much more tightly than are complexes of the remaining neutral nucleobases. This is in agreement with NMR studies[167] of complexes with $[(\eta^6\text{-bip})Ru(en)]^{2+}$, which show that reactivity decreases in the order G > I, T > C > A. Thymine is not included in Figure 4.1.3 as this can only bind as the N3-deprotonated anion, giving rise to much larger binding energies that are not comparable to those for the remaining, neutral bases. The gas-phase acidity of thymine has been reported as 347.03 kcal mol^{-1},[172] such that in the absence of solvent, binding of thymine to Ru-arenes is predicted to be energetically unfavourable. Solvent will clearly have a larger effect on the relative energies of charged metal and ligand moieties than on the remaining complexes with neutral bases, rendering comparisons based on gas-phase data useless. We therefore have not considered

thymine complexes in any subsequent analysis and the problem of including solvent in the manner as described in section 3.4.5 was not possible due to lack of time.

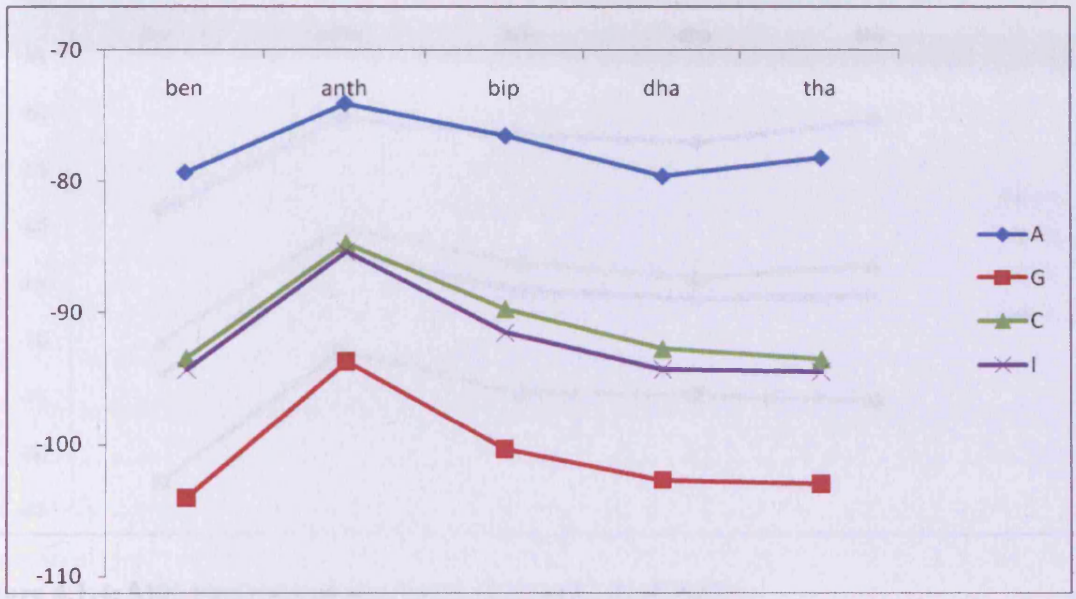
Table 4.1.2: Counterpoise corrected binding energies of $[(\eta^6\text{-arene})Ru(en)(base)]^{2+}$ complexes at the BHandH/6-31+G**(SDD) level in gas-phase (gas) and PCM aqueous solution (aq.) (kcal mol^{-1}).

Base\arene	ben	anth	bip	dha	tha
A (gas)	-79.31	-74.02	-76.49	-79.59	-78.19
(aq.)	-34.07	-34.56	-42.05	-45.22	-44.04
G (gas)	-104.00	-93.64	-100.32	-102.64	-102.94
(aq.)	-38.98	-39.14	-46.33	-50.40	-50.69
C (gas)	-93.44	-84.76	-89.74	-92.76	-93.54
(aq.)	-37.18	-36.64	-42.43	-45.50	-47.79
I (gas)	-94.35	-85.34	-91.54	-94.30	-94.49
(aq.)	-37.69	-38.09	-43.64	-47.88	-48.32
T (gas)	-266.54	-252.41	-257.68	-261.55	-262.57
GC (aq.)	-116.16	-	-	-	-115.32

Table 4.1.3: Binding energies of $[(\eta^6\text{-arene})Ru(en)(base)]^{2+}$ complexes (kcal mol^{-1}) at the M05/6-31+G*/SDD level.

Base\arene	ben	anth	bip	dha	tha
A	-58.67	-50.49	-51.81	-52.49	-50.57
G	-82.51	-70.89	-74.54	-74.81	-75.27
C	-70.06	-60.11	-63.16	-64.45	-63.27
I	-73.00	-62.29	-65.56	-66.28	-66.05

a.



b.

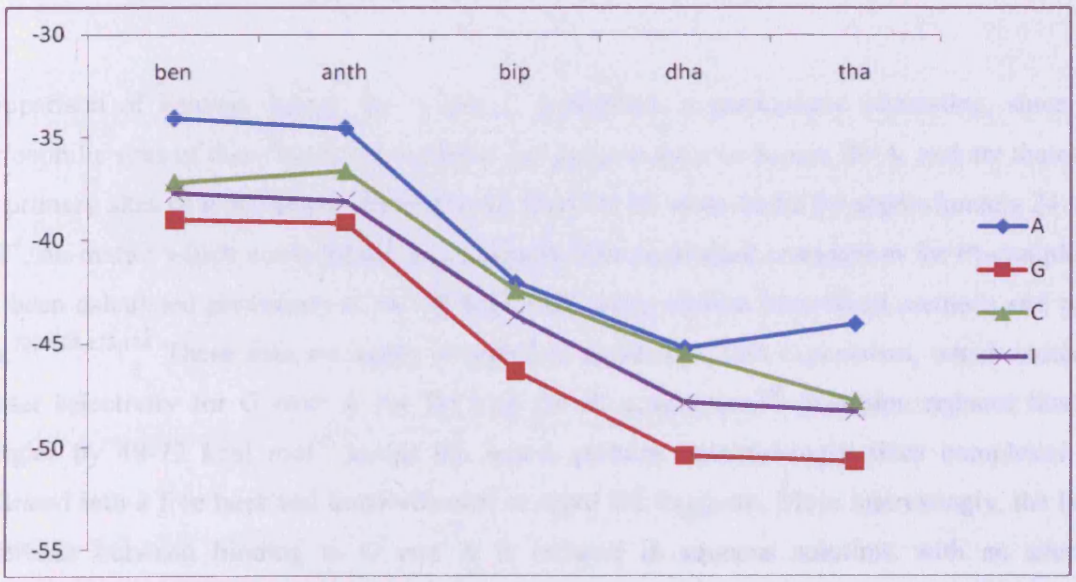


Figure 4.1.3: Counterpoise corrected binding energies for A, G, C and I in (a) gas phase; and (b) PCM aqueous solution.

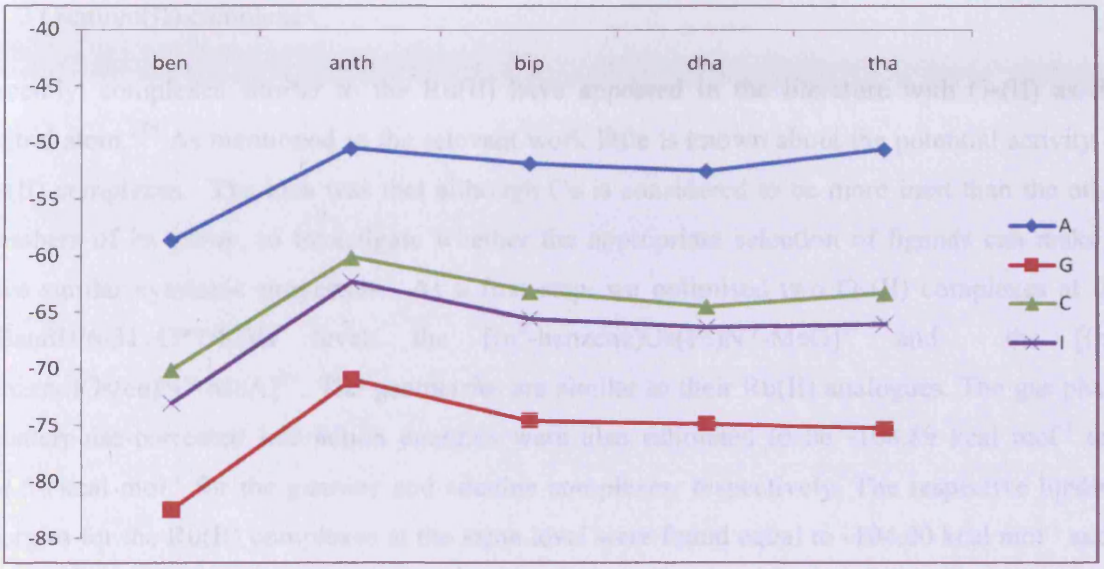


Figure 4.1.4: M05 binding energies for A, G, C and I (kcal mol^{-1}).

Comparison of binding energy for A and G complexes is particularly interesting, since the nucleophilic sites of these bases are available for complexation in duplex DNA, and are therefore the primary sites of metallation. The latter are found to be more stable by approximately 24 kcal mol^{-1}, no matter which arene ligand is considered. The equivalent comparison for Pt-complexes has been calculated previously at *ca.* 15 kcal mol^{-1} using various theoretical methods and basis sets.[22,25,28,173,174] These data are again in complete agreement with experiment, which indicates greater selectivity for G over A for Ru than for Pt complexes.[166] Solvation reduces binding energies by 48-72 kcal mol^{-1} across the board, perhaps unsurprisingly since complexes are separated into a free base and uncoordinated charged Ru fragment. More interestingly, the large difference between binding to G and A is reduced in aqueous solution, with an average difference of 5.14 kcal mol^{-1}. Baik *et al.* reported similar trends for cisplatin, and assigned this to preferential solvation of G and the accessibility to a solvent of polar groups in A complexes, resulting in an energy difference of 4.59 kcal mol^{-1}.[22] Thus, the differential binding of G and A of these Ru-arene complexes is suggested to be closer to that observed for cisplatin in aqueous solution than in the gas phase.

4.1.3 Osmium(II) complexes

Recently, complexes similar to the Ru(II) have appeared in the literature with Os(II) as the central atom.[175] As mentioned in the relevant work little is known about the potential activity of Os(II) complexes. The idea was that although Os is considered to be more inert than the other members of its group, to investigate whether the appropriate selection of ligands can make it have similar cytotoxic properties. As a first step, we optimised two Os(II) complexes at the BHandH/6-31+G**/SDD level: the $[(\eta^6$-benzene)Os(en)N7-MeG$]^{2+}$ and the $[(\eta^6$-benzene)Os(en)N7-MeA$]^{2+}$. The geometries are similar to their Ru(II) analogues. The gas phase counterpoise-corrected interaction energies were also estimated to be -108.89 kcal mol^{-1} and -84.54 kcal mol^{-1} for the guanine and adenine complexes, respectively. The respective binding energies for the Ru(II) complexes at the same level were found equal to -104.00 kcal mol^{-1} and -79.31 kcal mol^{-1}. In other words the above initial calculations suggest that changing the central atom from Ru(II) to Os(II) results in a tighter binding to the nucleobases of the order of 5 kcal mol^{-1}, while the preference for guanine over adenine is practically the same for the two metals (-24.35 kcal mol^{-1} for Os and -24.69 kcal mol^{-1} for Ru). The tighter binding for the Os(II) over the Ru(II) complexes corresponds well with the presumed inertness of the Os atom, which in turn may result in reduced side-effects in the case of Os(II), as more of the intact complexes can reach the cell without reacting with other biomolecules.

4.1.4 H-bonding and π-stacking

It has been suggested that the discrimination of G and A is due to the strong N—H...O H-bond in complexes with G, and to repulsion between amino groups of en and A. However, our optimised geometries indicate H-bonds are ubiquitous, with amino groups of A and C adopting a non-planar geometry that allows N to act as an H-bond acceptor, in accordance with experimental and theoretical findings.[174,176,177,178,179,180] The sum of angles around the amino – NH_2 atom for the A and C complexes was used in previous work to quantify this non-planarity. In A complexes, these angles sum to around 335°, indicative of relatively strong H-bonding, whereas in C complexes these angles sum to *ca.* 355°, suggesting less re-orientation due to H-bonding.

The ubiquity of H-bonds between amino groups of the ethylenediamine ligand and bases, and the intermediate stability of the inosine complex, suggest that the observed selectivity for guanine is not driven solely by N-H...O6 H-bonding. It is evident from Figure 4.1.2, and from experimental X-ray and NMR studies,[166,167] that the complexes adopt conformations in which arenes are oriented over the bases, giving rise to additional stabilising ring-ring stacking interactions. Table 4.1.2 and Figure 4.1.3 show that the binding energy follows the order anth > tha > dha > bip, *i.e.* increases with the size of the arene and with more parallel, face-to-face orientation of arene and base. This broadly agrees with the observation that cytotoxic activity of $[\eta^6$-arene)Ru(en)$]^{2+}$ complexes is increased by increasing the size of the arene,[51] which also involves the ability to intercalate between DNA base-pairs.[166,168] Thus, for these ligands at least, the importance of π-stacking seems clear.

Anthracene has not been studied experimentally, and was included in our theoretical study to examine the effect of arene rigidity. All anthracene complexes studied show weaker binding to a given base than any other arene considered. This seems most likely to be attributed to the lack of flexibility of the anth ligand, which prevents it from adopting a favourable positioning over the bases, that would allow for further stacking interactions. The reduced binding energy in the anth complexes (between 2.5 and 10 kcal mol^{-1}) is of the order expected for stacking interactions between DNA bases.[181,182,183,184,185,186] On the other hand, the relative twist of the two benzene rings in bip, and the sp^3 character of the C9, C10 atoms of dha and tha offer the required flexibility for developing stacking interactions between arenes and bases.

A clear exception to this trend is observed for complexes with benzene as the arene. Surprisingly, the binding energy of benzene complexes is comparable to that of dha and tha for all the studied nucleobases. In the gas phase, benzene complexes have comparable binding energies to the tha complexes. Since stacking interactions should not be present in these complexes, the origin of this stability is not clear, and will be explored in more detail below. However, in aqueous solution this apparently anomalous behaviour is not present, and benzene shows similar behaviour to anthracene. The lack of intercalative ability[166,168] of monocyclic arenes has been studied experimentally[169] with p-cymene (p-isopropyl-toluene) as a ligand, in which distortion and thermal destabilisation of DNA was found.

To probe the origin of these binding energies in more detail, further calculations were carried out on fragments of the whole complexes. Two sets of fragments were identified: i) with the Ru(en) moiety removed to leave just the arene and base; and ii) with the base removed to leave a $[\eta^6$-arene)Ru(en)$]^{2+}$ fragment. Both sets of fragments were fixed at the overall complex geometries. The former should shed light on stacking and other direct interactions between arene and base, while the latter can be used to monitor the electronic effect of arene on the metal centre and its potential interaction with Ru. Table 4.1.4 reports BHandH counterpoise-corrected arene-base binding energies, which confirm that interactions with benzene as the arene are small, and in most cases slightly destabilising. Interactions with anthracene are also generally small but stabilising, although rather large stabilisation is observed for adenine. Interestingly, this anthracene-adenine combination is also relatively stable from the data reported in Table 4.1.2. The more flexible arenes bip, dha, and tha show increased arene-base stabilisation, increasing with the same trend as seen for the overall binding energy for a given base. It seems clear, therefore, that the observed trend in binding energy against arene can be explained by these non-covalent interactions, with the obvious exception of benzene.

Table 4.1.4: Arene-base BHandH binding energies (kcal mol^{-1}).

	ben	anth	bip	dha	tha
A	+0.85	-3.59	-2.78	-4.07	-4.04
G	+0.50	-0.80	-3.79	-6.22	-6.56
I	+0.59	-1.24	-4.01	-6.40	-6.63
C	-1.56	-1.79	-4.92	-7.33	-8.65

In order to test the performance of BHandH, additional calculations on the same arene-base fragments were performed at the DF-LMP2 and SCSN levels, using the aug-cc-pVTZ basis set. The binding energies obtained are shown in table 4.1.5 and a comparison with BHandH for the tha complexes is illustrated in figure 4.1.5. All three methods show the same trend in binding energy, but as is well known DF-LMP2 significantly overestimates stacking energies. Excellent agreement between BHandH and SCSN results is observed, with an RMS error between these methods of just 0.76 kcal mol^{-1}. For the purines A, G and I, BHandH values are slightly less

negative (~0.6 kcal mol^{-1}) than SCSN values, whereas with cytosine as the base, the BHandH energy drops below the SCSN value. The tendency of BHandH to overestimate H-bond energy may be the cause of this change; however, the same drop in energy is observed in the benzene series, for which no H-bonds should be present.

Table 4.1.5: Arene-base SCSN and DF-LMP2 binding energies (kcal mol^{-1})

	ben		anth		bip		dha		tha	
	DF-LMP2	SCSN	DF-LMP2	SCSN	DF-LMP2	SCSN	DF-LMP2	SCSN	DF-LMP2	SCSN
A	+0.36	1.25	-6.25	-4.50	-6.12	-3.92	-10.42	-6.43	-8.32	-4.56
G	-0.21	+0.72	-5.12	-3.12	-7.28	-5.02	-12.55	-8.80	-10.62	-7.19
I	-0.78	+0.40	-5.69	-3.58	-7.68	-5.42	-12.66	-8.99	-10.43	-7.23
C	-1.92	-0.74	-4.83	-3.00	-7.45	-5.43	-11.13	-8.04	-10.43	-7.50

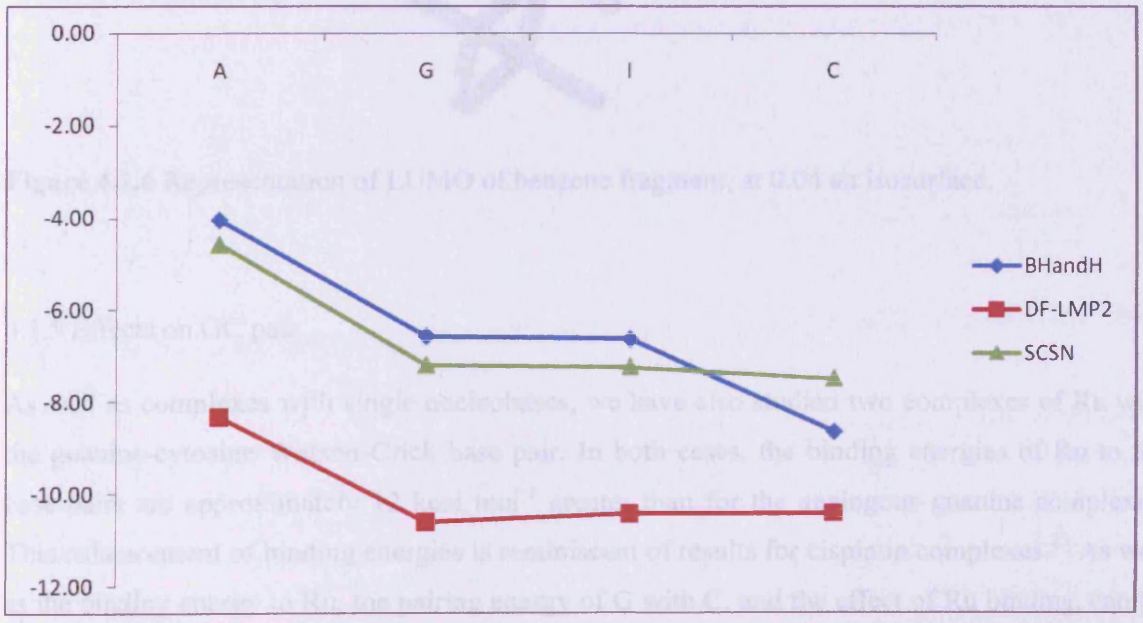


Figure 4.1.5: Binding energies of tha with nucleobases (kcal mol^{-1}).

Consideration of the second set of fragments goes some way to explaining the apparently anomalous behaviour of benzene complexes. Orbital energies of each $[\eta^6$-arene)Ru(en)$]^{2+}$

complex were calculated at both HF and DFT levels. An image of the LUMO of the benzene complex is shown in Figure 4.1.6: similar plots are observed for all other complexes. The LUMO of the benzene-containing fragment is markedly lower in energy (-0.424 au) than any of the remaining fragments (-0.376 to –0.396 au). On this basis, one would expect better overlap of benzene complexes with the HOMO of the incoming base, and hence a stronger covalent bond, than with other arenes. This will be investigated in more depth, using QTAIM methods, below.

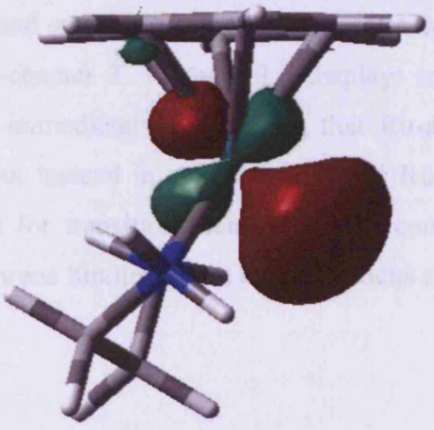


Figure 4.1.6 Representation of LUMO of benzene fragment, at 0.04 au isosurface.

4.1.5 Effects on GC pair

As well as complexes with single nucleobases, we have also studied two complexes of Ru with the guanine-cytosine Watson-Crick base pair. In both cases, the binding energies of Ru to the base pairs are approximately 12 kcal mol^{-1} greater than for the analogous guanine complexes. This enhancement of binding energies is reminiscent of results for cisplatin complexes.[25] As well as the binding energy to Ru, the pairing energy of G with C, and the effect of Ru binding, can be calculated from these results. The pairing energy of free GC is 20.10 kcal mol^{-1}, a value that is little changed in the benGC complex (20.80) or the thaGC complex (19.98). A more detailed

analysis of the observed effects on the GC pair, again using QTAIM analysis, will be discussed below.

4.1.6 QTAIM analysis

From these results, it is apparent that the high selectivity of the studied Ru(II) complexes towards guanine cannot be fully understood by simple structural criteria. The fact that all nucleobases act as H-bond acceptors suggests that the interplay of all interactions gives rise to the observed selectivity. In order to decompose the interactions present and hence to explore this interplay of covalent binding, H-bonding and π-stacking, we turn to QTAIM analysis, employed in the manner that was described in chapter 3. Figure 4.1.7 displays molecular graphs for selected complexes. One point that is immediately apparent is that Ru-arene bonding is not present between all six arene atoms, but instead in most cases 3 or 4 Ru-arene BCPs are found. This bonding pattern is not unusual for transition metal sandwich complexes, as was discussed in section 3.3 and in any case Ru-arene binding is not the main focus of this investigation.

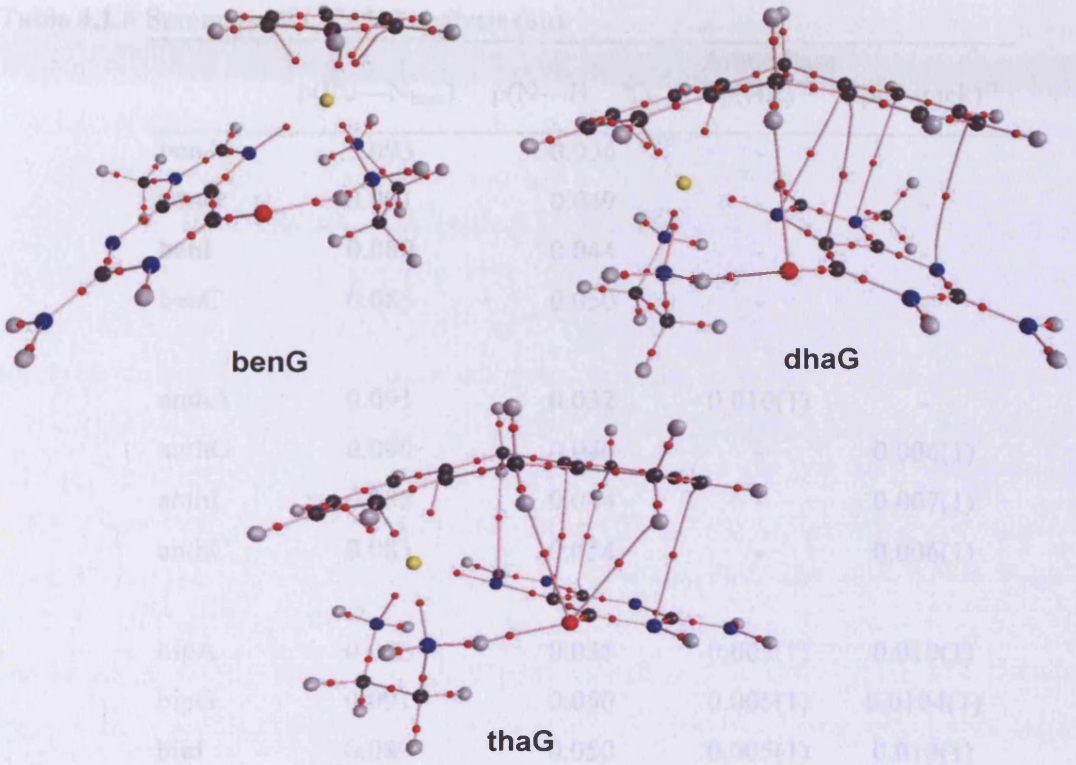


Figure 4.1.7: Molecular graphs of selected guanine complexes.

QTAIM analysis results in no arene-base interactions in any benzene complex, and progressively more in anth, bip, dha and tha complexes. This analysis reveals both stacking and H-bonding arene base interactions, the strength of which can be estimated from properties of the relevant BCPs. In this way, the interactions can be decomposed into covalent binding of the bases to Ru, H-bonding between bases and en, and H-bonding and π-stacking between bases and arenes. Table 4.1.6 summarises this QTAIM analysis for all complexes considered and lists the number of the arene-base H-bond and stacking BCPs observed in each complex. More details of individual interactions, such as between the en ligand and the purine and pyrimidine, bases can be found in the appendix (tables 7.1-7.4).

Table 4.1.6 Summary of QTAIM analysis (au).

	ρ(Ru—N_{base})	ρ(N—H…X)	Arene-base $\Sigma\rho$(HB) [a]	$\Sigma\rho(\pi$-stack) [a]
benA	0.093	0.036	-	-
benG	0.091	0.049	-	-
benI	0.089	0.044	-	-
benC	0.085	0.050	-	-
anthA	0.091	0.032	0.010(1)	-
anthG	0.090	0.046	-	0.006(1)
anthI	0.088	0.044	-	0.007(1)
anthC	0.083	0.054	-	0.006(1)
bipA	0.096	0.035	0.003(1)	0.010(1)
bipG	0.091	0.050	0.005(1)	0.0104(1)
bipI	0.089	0.050	0.005(1)	0.010(1)
bipC	0.084	0.054	-	0.009(1)
dhaA	0.092	0.039	0.017(2)	0.032(4)
dhaG	0.091	0.046	0.008(1)	0.031(4)
dhaI	0.089	0.046	0.007(1)	0.025(3)
dhaC	0.085	0.057	0.022(3)	0.028(4)
thaA	0.092	0.037	0.031(3)	0.013(2)
thaG	0.090	0.049	0.026(3)	0.017(2)
thaI	0.089	0.048	0.026(3)	0.016(2)
thaC	0.084	0.055	0.041(5)	0.018(2)
benGC	0.095	0.052	-	-
thaGC	0.093	0.054	0.040(5)	0.033(4)

[a] Values are the sum of the electron density at all BCPs located, with number of BCPs in parenthesis.

Table 4.1.7 also quantifies the strengths of the various non-covalent interactions, including both arene-base and en-base. This indicates that the hydrogen bonds from en-NH_2 ligands to O6 of guanine are the most stabilizing, closely followed by those to inosine. These are in the range 6.9–7.9 kcal mol^{-1} for G and 6.7–7.6 kcal mol^{-1} for I, with only a small effect coming from the nature of the arene. The strongest such hydrogen bond is found in bipG, approximately of equal strength to that in bipI. The strength of this hydrogen bond is slightly increased when guanine is paired with C, but only by ca. 0.5 kcal mol^{-1}, which is insufficient to account for the increase in the overall binding energy of ruthenium found in these complexes. These hydrogen bonds are 1.2–2.4 kcal mol^{-1} stronger than the N-H···N6 bonds to adenine, which are observed in crystal structures of adenine.[180] In the pyrimidine complexes, N-H···O2 of C are of strength comparable to those seen for G and I. In this case, the second -NH_2 of the en ligand is also involved in rather weak (~1.9 kcal mol^{-1}) hydrogen bonds to C. An unexpected type of BCP was found in all C complexes, in which the N-H of the -NH_2 group is the donor and the Ru atom is the acceptor, as shown in Figure 4.1.8. Such hydrogen bonds to metals are not without precedent, for example, in the "inverse hydration" studied by Kozelka *et al.*[14,187] Using the linear relationship employed for all other hydrogen bonds, the energies of these interactions are estimated to be in the range 2.1–2.4 kcal mol^{-1}, which is of magnitude similar to that observed by Kozelka *et al.* However, this figure should be treated with caution because this interaction seems quite different from those used to train such relationships.

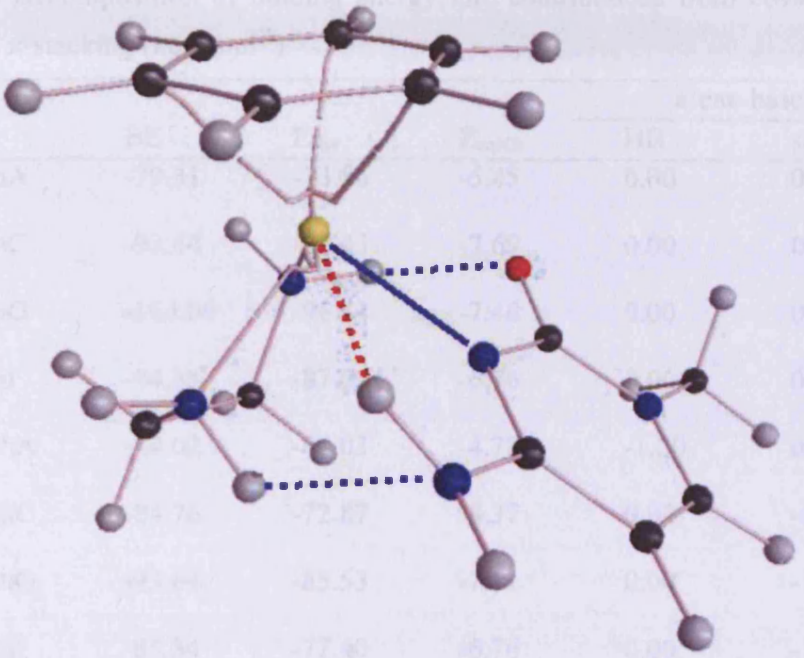


Figure 4.1.8: Molecular graph of the benC complex, with Ru—N bond shown as a continuous blue line, H-bonds shown as dashed blue lines, and Ru···H-N interaction as dashed red line. Other BCPs are omitted for clarity.

Between zero and four BCPs corresponding to arene-base stacking are also found, depending on the nature of the arene and base. These data are also summarized in Table 4.1.6. The largest number of stacking BCPs are found for dha, while in tha complexes, a number of these convert into C-H···π hydrogen bonds. In contrast, anth and bip complexes show fewer such BCPs, typically just one and, in one case (anthA), none at all. These data make it clear that the interplay of hydrogen-bonding and π-stacking effects is complex and depends on the details of both the arene and base.

Table 4.1.7 also summarizes a decomposition of the overall binding energy into contributions from covalent bonding, hydrogen bonding, and π stacking, based on QTAIM analysis and previously established relations between the electron density and energy (section 3.1.3). Covalent energies are estimated by subtracting the hydrogen-bonding and stacking contributions from the total binding energy, assuming that the remaining energy is due to covalent bonding.

Table 4.1.7: Decomposition of binding energy into contributions from covalent bonding, H-bonding, and π-stacking (kcal mol^{-1}) [a]

				arene-base	
	BE	E_{cov}	E_{enHB}	HB	stack
benA	-79.31	-73.86	-5.45	0.00	0.00
benC	-93.44	-83.43	-7.69	0.00	0.00
benG	-104.00	-96.54	-7.46	0.00	0.00
benI	-94.35	-87.58	-6.76	0.00	0.00
anthA	-74.02	-68.03	-4.78	-1.20	0.00
anthC	-84.76	-72.87	-8.37	0.00	-1.01
anthG	-93.64	-85.53	-7.11	0.00	-1.01
anthI	-85.34	-77.40	-6.76	0.00	-1.17
anthA	-74.02	-68.03	-4.78	-1.20	0.00
bipA	-76.49	-69.47	-5.29	0.00	-1.73
bipC	-89.74	-77.56	-8.34	0.00	-1.48
bipG	-100.32	-88.97	-7.81	-1.71	-1.84
bipI	-91.54	-82.06	-7.69	0.00	-1.79
dhaA	-79.59	-65.88	-5.86	-2.38	-5.48
dhaC	-92.76	-73.54	-8.80	-3.19	-4.81
dhaG	-102.64	-89.31	-7.07	-0.90	-5.36
dhaI	-94.30	-82.16	-7.02	-0.75	-4.37
thaA	-78.19	-65.74	-5.58	-4.54	-2.33
thaC	-93.54	-73.23	-8.55	-6.15	-3.17
thaG	-102.94	-87.37	-7.63	-5.04	-2.90
thaI	-94.49	-80.53	-7.28	-3.81	-2.86

benGC	-116.16	-108.19	-7.97	0.00	0.00
thaGC	-115.32	-95.18	-8.37	-6.02	-5.75

[a] Data for thymine complexes omitted due to different scale

These data make it clear that the bulk of the energetic preference for guanine over other bases comes from this covalent bonding. These contributions are, on average, around 8.1, 13.4, and 21.5 kcal mol^{-1} more favorable for guanine than for inosine, cytosine, and adenine, respectively. In contrast, non-covalent interactions are relatively constant across bases for a given arene, although the increased binding of dha and tha over anth and bip is clearly related to these non-covalent interactions.

Interestingly, we find correlations between E_{LUMO} and E_{cov} values for a given base. For all five G complexes, this yields $R^2 = 0.99$, for I complexes $R^2 = 0.98$, for C complexes $R^2 = 0.88$, and for A complexes $R^2 = 0.60$. This supports our argument that the apparently anomalous behavior of the benzene complexes is due to the electronic nature of this arene ligand, manifesting itself in a lower LUMO energy on ruthenium and hence stronger binding to a given base. Some relation is also seen between E_{cov} and the electron density in the Ru-N_{base} bond, although the quality of linear correlations is rather poor because of the limited range of data available. More detail and plots can be found in the appendix (figures 7.1, 7.2).

As mentioned above, binding to the GC pair is considerably more favorable than that to G alone. Table 4.1.7 shows that much of this extra stabilization comes from enhanced covalent bonding in the GC case, with a small increase in the strength of the N-H···O hydrogen bond. In the thaGC complex, further stabilization stems from stacking interactions between the arene and cytosine as well as with guanine. This is shown schematically in Figure 4.1.9. In thaG structure, the arene is positioned over guanine, whereas in thaGC, it moves slightly to be positioned over both G and C, leading to the formation of BCPs between tha and C as well as between tha and G.

QTAIM analysis can also be used to monitor the individual hydrogen bonds within the GC pairing, as shown in Figure 4.1.10. Compared to free GC, in which N4-H4···O6 is the strongest

of the three hydrogen bonds present, both benGC and thaGC show substantial weakening of this hydrogen bond. In contrast, the two hydrogen bonds in which G acts as a proton donor are strengthened, such that the overall pairing energy of GC is approximately equal in all cases. This is again similar to the situation seen in complexes of cisplatin,[25] as well as other late transition metals.[46] In previous work[46] using a different set of ligands, values of 5.14, 8.29, and 8.65 kcal mol^{-1} were estimated for the hydrogen bonds between a Ru-G_{N7} complex and C, values only slightly different from those seen in Figure 4.1.10. From this, we suggest that ruthenium complexation to G sites in DNA should have a relatively small effect on GC pairing within the duplex.

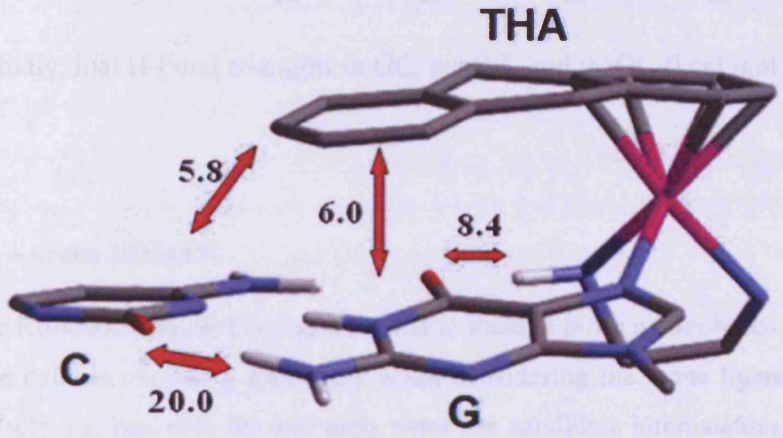


Figure 4.1.9: Intermolecular energy contributions in the thaGC complex [kcal mol^{-1}] (H-atoms other than those of the G-C, G-en H-bonds are not shown for simplicity).

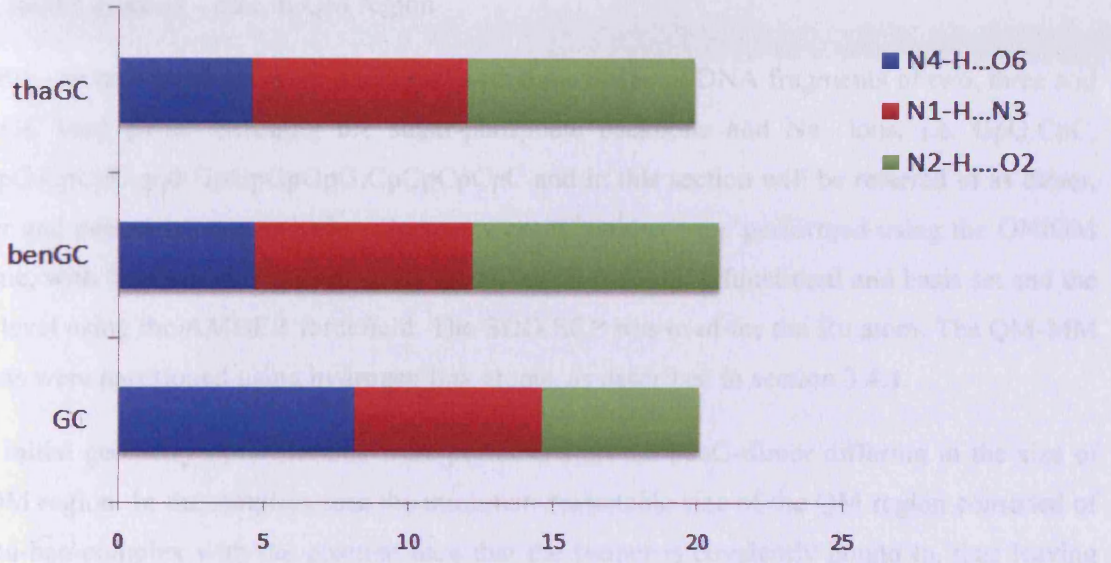


Figure 4.1.10: Individual H-bond strengths in GC, benGC, and thaGC (kcal mol^{-1})

4.2 Ruthenium – arene ONIOM

The study of the Ru-arene complexes was intended to include more nucleobases. The addition of extra nucleobase pairs is necessary especially when considering the arene ligands that can have intercalating effects, i.e. bip, dha, tha and anth, since the candidate intercalators will most likely interact with both the nucleic acid pairs between which they are inserted. The complexes presented in section 4.1 are already of computationally considerable size and therefore QM/MM was the suitable choice for any increase of the present complexes' size. Initially, efficient optimization strategies were sought in collaboration with F. Marozzelli, who at that time was an MSc. student and who further extended the number of the studied complexes and also explored variations of Ru-arene-DNA adducts and their energetics.[188] The work described in the following two sections was carried out by F. Marozzelli within the context of his MSc. project, the supervision of which to a large extent was done by myself.

4.2.1 Model systems – size of QM region

The Ru – arene complexes were combined with three different DNA fragments of two, three and five GC base pairs, including the sugar-phosphate backbone and Na^+ ions, i.e. GpG.CpC, GpGpG.CpCpC and GpGpGpGpG.CpCpCpCpC and in this section will be referred to as dimer, trimer and pentamer, respectively. Geometry optimizations were performed using the ONIOM scheme, with the QM level described by the BHandH/6-31+G** functional and basis set and the MM level using the AMBER forcefield. The SDD ECP was used for the Ru atom. The QM-MM regions were partitioned using hydrogen link atoms, as described in section 3.4.1.

Two initial geometry optimisations were performed on the benG-dimer differing in the size of the QM region. In the simplest case the minimum reasonable size of the QM region consisted of the Ru-ben complex with the guanine base that the former is covalently bound to, thus leaving the respective cytosine and the additional GC pair, as well as the sugar-phosphate backbone in the MM region. In the second approach, the QM region was extended to include the above mentioned cytosine base and therefore including the Ru-ben and one GC pair. The lack of differences in the resulting optimised structures indicated that calculations could be performed at a lower computational cost using the small QM region.

4.2.2 Geometry optimisations

Increasing either the size of the Ru-ligands or of the DNA fragments proved to be a technically challenging task. In several cases the resulting structures were characterized by significant distortions of the DNA part that were considered unrealistic. Such problems were countered by running preliminary optimizations that imposed restraints on atoms or groups of atoms. This strategy was followed for all the ligands, apart from benzene, in the cases of the DNA pentamers.

Examples of optimised geometries are shown below: figure 4.2.1 displays optimised geometries of the pentamer that contains benzene and figure 4.2.2 displays two different binding modes of the Ru-dha to a trimer. As seen in figure 4.2.2 (left), in the first of the examined binding modes the cyclic ligand that is located on top of the Ru atom points towards the oligonucleotide (t) and in the second case (figure 4.2.2-right) the same ligand points away from the oligonucleotide (a).

The optimization of a complex twice with different binding mode of the metal-drug was not limited to the case of the Ru-dha trimer, but was followed for all trimers whose ligands contain more than one ring. The same approach was not followed for the pentamer structure due to time limitations of the project.

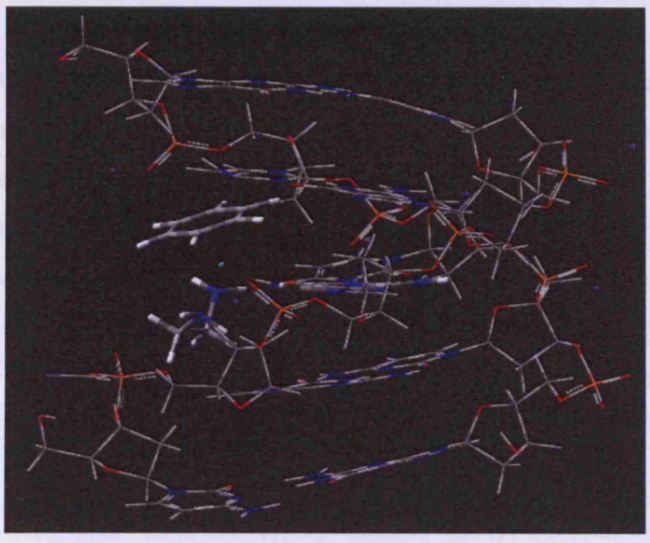


Figure 4.2.1: Ru-ben pentamer (picture taken from ref.188).

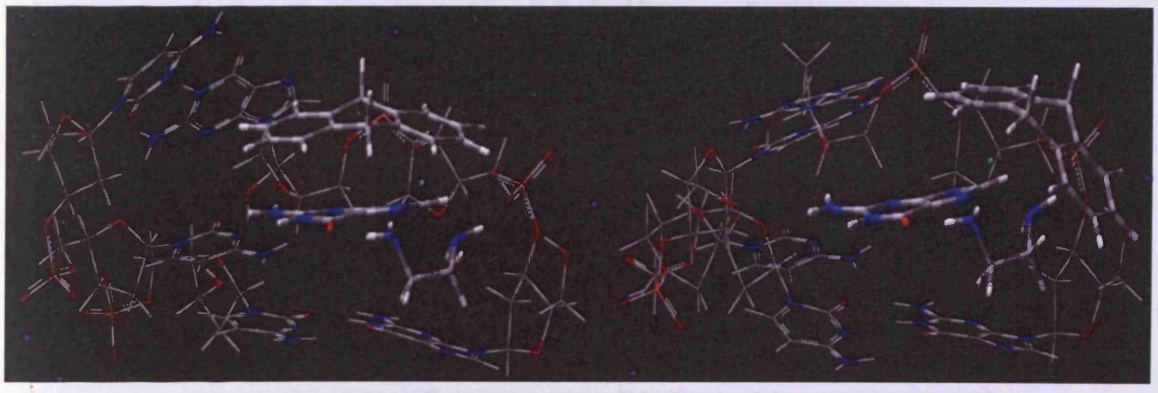


Figure 4.2.2: Ru-dha (t) (left) and Ru-dha (a) (right) bound to DNA trimers (picture taken from ref.188).

4.2.3 Binding energies of the complexes

Apart from the distortions caused on the oligonucleotides, we were also interested in the binding energies of the complexes, as well as on the energetic differences between the two binding modes of the same complex. Interestingly, in the case of the Ru-bip and Ru-dha trimers, the structures where the arene ligand points away from the DNA fragment are more stable in terms of the QM/MM-derived energy (table 4.2.1). This fact was suprising, since when the arenes are in closer proximity to the nucleobases more stabilizing interactions were expected. However, counterpoise-corrected binding energies performed fully on the DFT level and at truncated systems of the optimised complexes that do not involve the MM part showed that the binding is favoured as was expected; i.e. between the same complexes, the one with the ligands pointing towards the helix is favoured, a fact that is attributed to the relative energies of the monomers (Table 4.2.2). More detailed discussion on the systems summarized here can be found in reference 188.

Table 4.2.1: Comparison of energies between the trimers of Ru-bip and Ru-dha (table taken from ref. 188).

		towards	away	ΔE_{tow-aw}
		hartree	hartree	kcal mol^{-1}
Ru-bip	DFT/QM	-1,291.82	-1,291.87	28.34
	ONIOM	-1,284.45	-1,284.52	48.63
	AMBER/MM	-2.08	-2.19	69.03
Ru-dha	DFT/QM	-1,368.72	-1,368.73	6.28
	ONIOM	-1,361.37	-1,361.38	6.28
	AMBER/MM	-2.22	-2.21	-6.28

Table 4.2.2: CP-corrected binding energies.

Ligands	Binding Energy CP corrected (kcal mol^{-1})
ben	-342.59
bip towards	-333.22
bip away	-327.80
tha towards	-327.54
dha towards	-326.47
dha away	-324.56
ant towards	-322.98

4.3 Effect of methyl substitution on Pt(en) complexes

Since the discovery of the anticancer activity of cisplatin, cis-$[Pt(NH_3)_2Cl_2]$ (1), in the 1960s[2,3] , much interest has been directed towards the relation between structure and activity of platinum(II) complexes. The classic rules of Cleare and Hoeschele show that the ideal requirements for high activity are square-planar Pt(II) complexes with two halide or carboxylate ligands to act as leaving groups, along with two amines as ''carrier groups'', in a cis conformation[189,190]. Moreover, the presence of at least one N–H on each amine enhances activity, an effect that has been ascribed to the formation of hydrogen bonds to acceptor atoms on the DNA target. Cisplatin and related platinum complexes, cis-$[PtA_2X_2]$, are believed to enter cells intact via passive diffusion across membranes, and are then activated by hydrolysis to the reactive cis-$[PtA_2(OH_2)]^{2+}$ complex, which then undergoes nucleophilic substitution by electron-rich groups in DNA, most notably the N7 sites on guanine[173,191,192,193,194].

Within these structure–activity rules, the scope for modification of both carrier and leaving groups is immense, and such modifications have been employed to generate new drugs with improved therapeutic properties. Alteration of the leaving groups, X, leaves the final DNA adduct unchanged, and hence such changes do not alter the range of cancers treated by a drug, but can alter the physical and pharmacokinetic properties of the drug. Carboplatin, the second

platinum drug to be licensed for treatment of cancers, is one example of such an approach, in which the replacement of chloride ligands for chelating dicarboxylate slows metabolism to the active species and hence reduces systemic toxicity [5,195].

Modification of amine carrier groups, A, on the other hand, can alter both pharmacokinetics and the nature of the final DNA adduct, and hence offers the prospect of treating a wider range of cancers than are amenable to cisplatin therapy. Oxaliplatin is perhaps the best known of such drugs, in which the two ammine groups of cisplatin are replaced by a chelating 1,2-diaminocyclohexane ligand[196]. This not only alters the electronic and steric properties of the platinum centre, but is also believed to increase the lipophilicity of the drug and hence increase uptake in cells. Another common modification, albeit one that has not yet been included into any licensed drug, is the use of ethylenediamine as a chelating ligand, as exemplified in the parent complex cis-[Pt(en)Cl_2] (**2**; where en is ethylene-diamine)[197]. The carbon backbone of ethylenediamine offers further possibility for structural modification, for example by attachment of lipophilic groups or pharmacophores to increase uptake into cells [198,199].

The following sections (4.3.1-4.3.5) are concerned with the effects of such modifications on the binding of platinum complexes to DNA, based around a set of variations on the theme of 2, in which C–H and/or N–H groups are replaced by up to four methyl groups. Activity data, in the form of IC_{50} values for A2780 human ovarian cancer cells, have been reported for this set of molecules, along with molecular mechanics (MM; AMBER) calculations on their adducts with an octamer of double-stranded DNA[200]. Following this work, hybrid quantum mechanics QM/MM are employed here to explore structural and energetic aspects of binding to DNA. In this approach, the platinum centre and the groups most closely bound to it are treated at the DFT level, whereas the remainder of the DNA structure is treated with a much faster MM method. DFT methods are widely used to study cisplatin and related molecules and their interactions with DNA and other biomolecules although QM/MM methods are less well studied in this field, as mentioned in the introduction. Our hypothesis is that this approach should be able to better reflect the subtle changes in electronic and steric properties that stem from substitutions in carrier ligands than the purely MM approach employed in ref. 200. Figure 4.3.1 displays the parent chloro-Pt(II) compounds cisplatin (1), dichloroethylenediamine platinum(II) (2) and the latter's methyl derivatives (3–13).

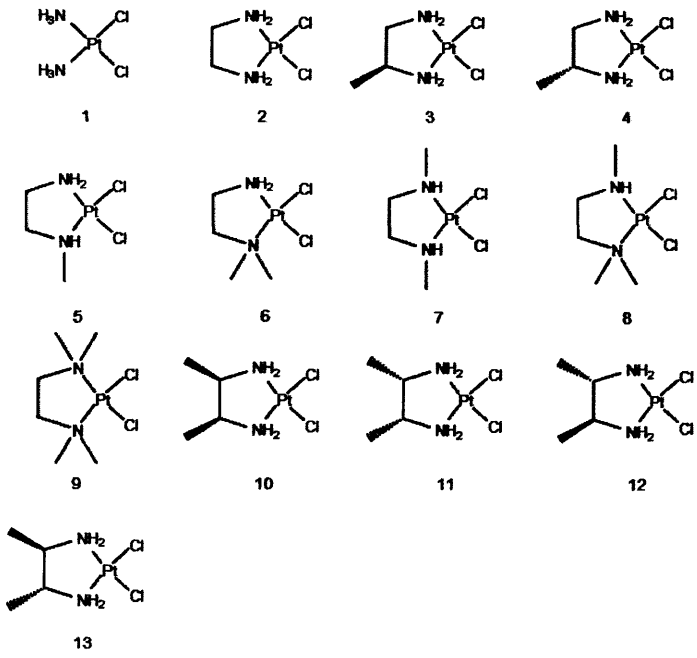


Figure 4.3.1: Structures of compounds 1 – 13.

4.3.1 Geometries of complexes

Figure 4.3.2 displays the optimised geometries for ss and ds cisplatin-DNA complexes, also showing graphically the choice of QM and MM region, with the QM region in bold and MM region as wireframe. Key geometrical parameters, including bond lengths and angles around Pt as well as guanine-guanine interplanar angles, for all complexes are reported in Tables 4.3.1 and 4.3.2, for ss and ds complexes, respectively. Geometrical parameters for **1** are close to available experimental data[201] and previous calculations for cisplatin-GpG, maintaining the distorted square-planar geometry seen in many cisplatin complexes. The differences between single and double-strand DNA adducts are generally small, with slightly shorter Pt—N7 distances in the latter. However, one substantial change between ss and ds complexes is in the relative orientation of the guanines, as measured by the angle between the mean planes of each base. In the ss-complexes, most values are close to 90°, with the exception of the tri- and tetra- N-methylated complexes **8** and **9**. In the ds-complexes, the two guanines are much closer to the idealised parallel orientation, at 25 - 30°, due to the restraining effect of the complementary CpC fragment,

as can be observed in Figure 4.3.2. Since the latter is a more realistic representation of cisplatin-DNA interactions, this significant change in geometry demonstrates the potential of the QM/MM approach to describe larger-scale fragments of DNA than could be easily tackled using QM methods alone.

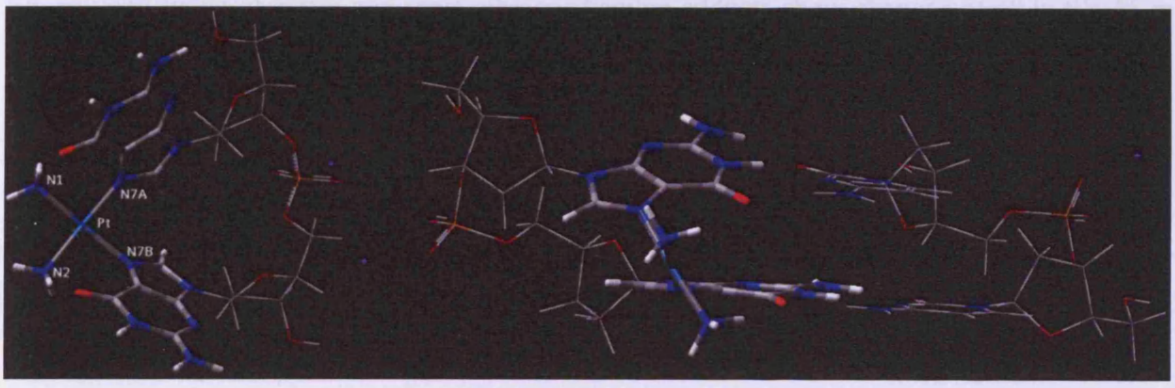


Figure 4.3.2 Optimised geometry of cisplatin adduct with ss- and ds-DNA.

Modification from cisplatin to the parent ethylenediamine adduct (entry **2**) causes significant decrease in the N1-Pt-N2 angle, due to the constraints of the 5-membered metallocycle, with corresponding increases in the remaining angles around N. However, the key geometrical features of the adduct of $[Pt(en)]^{2+}$ with GpG are essentially unchanged. Introduction of a single methyl group on the C or N backbone of en (**3**, **4** and **5**) makes little difference to the geometrical parameters in Tables 4.3.1 and 4.3.2, nor does the presence of two methyl groups on backbone carbons (**10**, **11**, **12** and **13**) or on separate nitrogens (**7**). In contrast, double methylation on a single nitrogen (**6**, **8**) or on both nitrogens (**9**) causes significant changes in geometrical parameters. The Pt—N distance for the methylated nitrogen is increased, while the interplanar guanine-guanine angle is significantly reduced.

As well as the local geometrical parameters around Pt and the coordinated guanines, Table 4.3.3 reports base-pair geometrical parameters extracted from Curves analysis of the optimised geometries. It is instructive to compare these to the equivalent data from ds-GpG.CpC optimised at the same QM/MM level. It is evident that, in most cases, platination causes significant disruption to the base-pair structure. Shift and slide parameters for cisplatin and most other

adducts are altered by between 0.2 and 0.3Å from the free DNA value, but the change induced in rise (the vertical separation of base pairs) is rather smaller. The tri- and tetra-methylated complexes **8** and **9** buck this trend, with shift and slide values much closer to free DNA and rise values much smaller than in free DNA. This pattern is also reflected in the angular parameters tilt, roll and twist: cisplatin and most ethylenediamine adducts show changes of 4° in tilt, 8° in roll and 4° in twist, whereas complexes **8** and **9** show larger changes of 5°, 13° and 7°. Thus it is evident that even with the minimal model of DNA employed in this work, the disruption caused by platination and the effect of the number and position of methyl substituents can be estimated using the QM/MM approach. Table 4.3.3 also reports geometrical data for GC pairing, averaged over both pairs present. Once again, changes across most complexes are relatively small except for complexes **8** and **9**, which show significant deviations from all other complexes, most notably in the angular terms buckle and opening.

Table 4.3.1: Selected geometrical data for single-stranded complexes.

	Pt-N1	Pt-N2	Pt-N7A	Pt-N7B	N7A-Pt-N7B	N1-Pt-N2	guan/guan [a]
1	2.031	2.033	2.027	2.024	90.8	92.8	86.5
expt [b]	2.054	2.071	1.972	2.043	88.2	92.1	84.3
(sd)	(0.040)	(0.016)	(0.046)	(0.062)	(2.1)	(1.3)	(3.7)
2	2.028	2.027	2.037	2.027	91.0	83.2	88.9
3	2.027	2.022	2.039	2.027	90.9	83.0	89.2
4	2.027	2.027	2.037	2.027	90.7	83.1	89.4
5	2.026	2.036	2.040	2.028	90.5	83.7	89.8
6	2.053	2.029	2.018	2.024	86.9	84.2	70.0
7	2.030	2.035	2.040	2.032	90.6	84.3	89.4
8	2.044	2.053	2.014	2.027	83.4	84.1	43.1
9	2.057	2.053	2.016	2.030	83.2	84.9	43.8
10	2.026	2.021	2.041	2.029	91.0	82.7	87.7
11	2.022	2.026	2.037	2.029	90.8	82.8	89.7
12	2.02	2.02	2.039	2.03	90.9	82.6	89.3
13	2.026	2.027	2.039	2.027	90.7	82.9	89.8

[a] Interplanar angle between coordinated guanines. [b] Experimental data from ref. 201 averaged over four independent molecules in asymmetric unit, with sd calculated from variation across each molecule, rather than from the reported experimental esd.

Table 4.3.2: Selected geometrical data for double-stranded complexes.

	Pt-N1	Pt-N2	Pt-N7A	Pt-N7B	N7A-Pt-N7B	N1-Pt-N2	guan/guan [a]
1	2.045	2.035	2.019	2.016	84.9	94.2	26.4
2	2.039	2.027	2.024	2.014	84.4	83.5	25.6
3	2.037	2.023	2.026	2.017	84.3	83.2	26.2
4	2.037	2.027	2.025	2.017	84.3	83.4	26.5
5	2.036	2.032	2.029	2.108	84.2	83.9	26.4
6	2.048	2.024	2.026	2.024	84.2	84.8	25.7
7	2.039	2.031	2.031	2.021	84.3	84.7	26.0
8	2.037	2.065	2.019	2.030	82.8	84.9	29.6
9	2.048	2.064	2.022	2.031	82.7	85.6	28.6
10	2.035	2.024	2.025	2.020	84.2	82.8	26.5
11	2.032	2.026	2.026	2.018	84.2	83.1	26.4
12	2.030	2.020	2.027	2.019	84.2	82.8	26.3
13	2.036	2.028	2.024	2.020	84.2	83.2	26.7

Table 4.3.3: Base-pair geometries for platinated GC pairs (Å and °) [a]

	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (τ)	Roll (ρ)	Twist (ω)	Shear (Sx)	Stretch (Sy)	Stagger (Sz)	Buckle (κ)	Propel (ω)	Opening (σ)
1	1.11	-2.07	3.90	-1.31	6.91	31.92	-0.17	-0.13	-0.20	8.19	-4.42	7.35
2	1.22	-1.83	3.67	-1.27	6.06	30.82	-0.03	-0.15	-0.24	0.6	-4.41	3.07
3	1.09	-2.10	3.92	-1.68	6.79	32.66	-0.21	-0.08	-0.16	10.16	-5.33	8.42
4	1.09	-2.10	3.91	-1.72	6.84	32.64	-0.21	-0.08	-0.15	10.2	-5.29	8.45
5	1.08	-2.11	3.92	-1.79	6.69	32.73	-0.21	-0.08	-0.15	9.95	-5.35	8.31
6	1.09	-2.01	3.85	-1.41	7.34	31.94	-0.18	-0.10	-0.17	10.28	-4.58	7.81
7	1.10	-2.08	3.95	-1.31	6.45	33.16	-0.21	-0.08	-0.16	8.58	-6.71	7.67
8	0.77	-1.88	3.56	-3.70	13.00	28.84	-0.48	-0.06	0.01	29.36	0.83	14.51
9	0.81	-1.86	3.60	-2.51	12.1	29.50	-0.33	-0.05	-0.02	24.56	-1.36	13.35
10	1.07	-2.04	3.85	-1.46	7.04	32.19	-0.22	-0.09	-0.14	9.95	-5.49	7.79
11	1.09	-2.11	3.91	-1.64	6.76	32.68	-0.21	-0.08	-0.15	10.16	-5.38	8.4
12	1.10	-2.11	3.92	-1.71	6.67	32.72	-0.21	-0.08	-0.16	10.05	-5.28	8.4
13	1.06	-2.06	3.85	-1.79	7.31	32.10	-0.25	-0.08	-0.13	11.4	-5.06	8.53
GpG_relax [b]	0.87	-1.73	3.83	+2.27	-1.02	36.16	-0.02	0.14	0.02	14.25	-7.79	4.41

[a] Shift, slide, rise, tilt, roll and twist refer to the orientation between GC pairs, while shear, stretch, stagger, buckle, propel and opening refer to orientations within GC pairs, averaged over both; [b]Obtained from QM/MM optimisation of GpG.CpC using the same methods.

4.3.2 Binding Energies

After obtaining the optimal geometries of Pt–DNA adducts that were described in the previous section, several methods for calculation of the binding energy were tested. The simplest approach, denoted ΔE^{direct} in this work, simply calculates the energy of the platinum-carrier fragment and the DNA fragment independently, with no geometrical relaxation of either fragment from the complex. Allowing relaxation of each fragment to its optimum geometry results in data denoted ΔE^{relax}. In both cases, the platinum-containing fragment was calculated with DFT, and the DNA fragment with DFT/AMBER, using the same link between QM and MM regions as for the complex. A shortcoming of both approaches is that they include the energy of the bare $[PtA_2]^{2+}$ complex, which is unstable and chemically unrealistic. We therefore also calculated ΔE^{reac}, i.e. the change in energy for the model reaction shown in Eq. 4.3.1:

$$\text{cis - } [PtA_2(H_2O)_2]^{2+} + \text{ss/ds-GpG} \rightarrow \text{cis - } [Pt(NH_3)(\text{ss/ds-GpG})]^{2+} + 2H_2O \qquad (4.3.1)$$

in which nucleotide-containing species are optimised using ONIOM, as described above, and the remaining species are optimised with DFT. ΔE^{direct} and ΔE^{relax} were calculated according to Eq. 4.3.2, with the former holding DNA and platinum fragments at their geometries in the Pt–DNA complex, and the latter allowing each to relax to its nearest local energy minimum:

$$\text{cis - } [Pt(NH_3)]^{2+} + \text{ss/ds-GpG} \rightarrow \text{cis - } [Pt(NH_3)(\text{ss/ds-GpG})]^{2+} \qquad (4.3.2)$$

Furthermore, counterpoise-corrections for the BSSE problem have been computed purely at the DFT level for truncated models by replacing the sugar-phosphate backbone by methyl groups. Counterpoise corrections estimated in this manner have been added to the ΔE^{direct}. The same truncated DFT model was also used to estimate the relative aqueous solvation energy of the complex and fragments using the polarisable continuum model (PCM), yielding a quantity termed here as ΔE^{solv}.

Table 4.3.4 summarizes the binding energies (kcal mol^{-1}) for ss and ds complexes, calculated directly or *via* reaction with two water molecules; these data are shown graphically in figure 4.3.3. Direct evaluation of binding energy shows that binding to ds-DNA is more stable than to ss-DNA by between 3 and 14 kcal mol^{-1}, presumably due to the greater attraction of the Pt^{2+} centre to the negatively charged phosphates, which are in close proximity despite not being

formally bound. Cisplatin is the most tightly bound of all adducts considered in this work, forming an adduct that is approximately 12 kcal mol^{-1} more stable than $[Pt(en)]^{2+}$. Successive methylation further reduces binding energy, and the position of methylation again strongly affects the calculated results. Introduction of a methyl group on the carbon of en (**3**, **4**) reduces binding by 4 - 5 kcal mol^{-1}, while two methyls (**10**, **11**, **12**, **13**) reduce binding by a further 3 - 5 kcal mol^{-1}. A single methyl on nitrogen (**5**) has a much larger effect of 8 - 10 kcal mol^{-1}, while multiple substitution on N has still larger effects of 30 - 40 kcal mol^{-1}. Binding energies calculated *via* the reaction energy of $[PtA_2(H_2O)_2]^{2+}$ with ss- and ds-DNA show similar trends, with cisplatin more strongly bound than any other complex. However, the differences between single and double methylation, and between substitution on C or N are much less pronounced in these data. Particularly in the ds-DNA adducts, the only significant difference in binding energy between all variations on the en backbone occur for 3 or 4 methyls on nitrogen (**8**, **9**), both of which are much less strongly bound than the remaining complexes.

Table 4.3.4 Binding energies to ss- and ds-DNA (kcal mol^{-1})

	ΔE^{direct}		ΔE^{react}		ΔE^{solv}	
	ss	ds	ss	ds	ss	ds
1	-279.57	-285.54	-112.40	-137.97	-73.90	-56.73
2	-267.63	-272.02	-109.42	-132.03	-64.13	-46.84
3	-263.03	-266.53	-106.95	-130.91	-63.83	-46.14
4	-263.79	-267.44	-98.11	-131.45	-83.54	-46.43
5	-259.06	-262.29	-106.66	-130.95	-66.89	-49.21
6	-247.23	-256.63	-99.58	-129.93	-69.99	-53.3
7	-252.14	-255.36	-104.65	-128.57	-68.29	-50.51
8	-230.02	-242.56	-82.57	-115.73	-71.56	-58.12
9	-224.76	-237.54	-88.87	-115.59	-74.73	-60.29
10	-259.79	-264.04	-106.29	-131.35	-65.41	-48.41
11	-259.29	-262.87	-95.96	-129.25	-82.87	-45.26
12	-258.66	-261.90	-104.82	-128.76	-63.49	-45.56
13	-260.47	-264.98	-97.43	-131.82	-84.54	-48.77

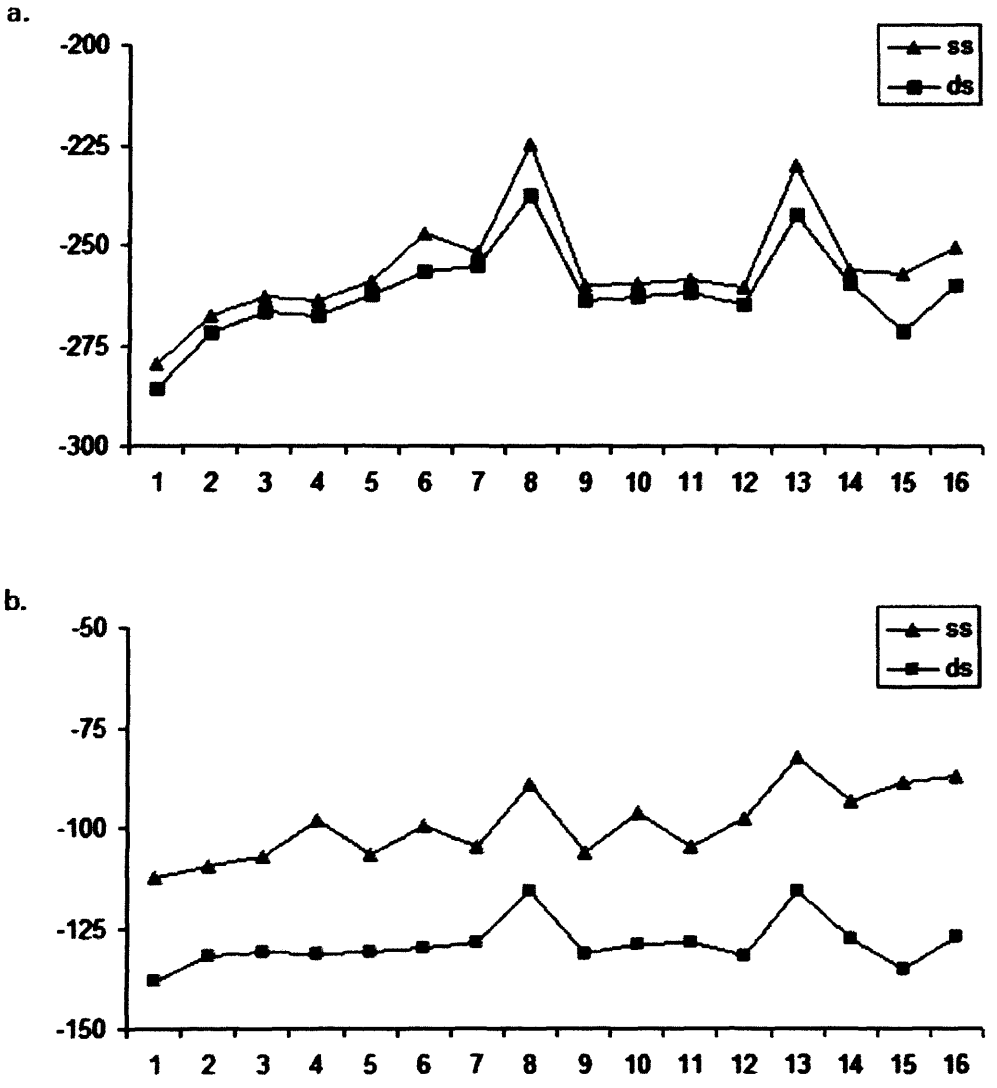


Figure 4.3.3 Binding energies to ss- and ds-DNA (kcal mol^{-1}) (a) ΔE^{direct} and (b) ΔE^{react}.

Binding energies that allow geometrical relaxation of the unbound Pt and DNA species, ΔE^{relax}, are smaller than ΔE^{direct} by approximately 20 kcal mol^{-1}. A scatter plot of ΔE^{direct} *vs.* ΔE^{relax} gives an r^2 value of 0.99, indicating that the effect of relaxation on binding energy is essentially constant, such that ΔE^{relax} carries no significant information that is not present in ΔE^{direct}. These values are therefore not reported in Table 4.3.4. Corrections for the aqueous solvation of each

species, using the polarizable continuum model (PCM) were applied to generate ΔE^{solv}, also reported in Table 4.3.4. Interestingly, these data do not reflect the trend observed in all other methods of calculating binding energy, showing the most strongly bound adducts to be those of the tri- or tetra- N-methyl complexes **8** and **9**. Closer examination of the results leading to this data indicates that this behaviour stems from relative solvation of the dissociated platinum moieties. Methyl substitution of the polar N—H groups drastically reduces the solvation energy of these species, thus increasing binding energy relative to better solvated fragments.

4.3.3 QTAIM analysis

Insight into the origins of the trends in binding can be obtained from Atoms in Molecules (QTAIM) analysis, which locates covalent and non-covalent interactions between DNA and the Pt fragments and allows estimation of the strengths of individual interactions. The necessary wavefunction calculations were performed in the truncated models described in the beginning of section 4.3.2. This analysis, reported in Table 4.3.5 and illustrated in Figure 4.3.4, reveals that in the ss-DNA adducts of cisplatin, $[Pt(en)]^{2+}$, and all carbon-substituted variations of this, two N—H···O hydrogen bonds between ligand N—H and O6 of guanine are formed. On the basis of the electron density at the H-bond critical point, these H-bonds are estimated to contribute 12 kcal mol^{-1} to the overall stability of the adduct. A single N-methyl substitution (**5**, **7**) preserves this pattern of H-bonding, but replacement of both N—H's on a single centre leads to C—H···O H-bonding in place of N—H...O. Such contacts contribute much less, approximately 1.5 kcal mol^{-1}, to the overall stability of the complex and are examined in more detail in the following section. In complexes **8** and **9**, only C—H···O contacts are observed. However, the changes in H-bond energy are relatively small compared to those reported in overall binding energy in Table 4.3.4. Subtracting the H-bond contribution from the overall ΔE^{direct} gives an estimate of the intrinsic energy of the covalent Pt—N bonds, ΔE_{cov}. There is no significant correlation between ΔE_{cov} and the electron density in the Pt—N bonds ($r^2 = 0.32$), but close correlation between ΔE_{cov} and the LUMO energy of the Pt fragment is observed ($r^2 = 0.96$). The most weakly bound complexes (**8** and **9**) have, by some margin, the highest energy LUMO values, and are less able to interact with the HOMO of guanine. It is apparent, therefore, that methylation on nitrogen significantly

diminishes both covalent and non-covalent interactions between the Pt-fragment and DNA, both effects contributing to the relative weakness of binding of such complexes.

Table 4.3.5: QTAIM Energy decomposition for ss complexes.

	N-H···O	C-H···O	Sum of hydrogen bonds (kcal mol^{-1})	ΔE_{cov}(kcal mol^{-1})	Sum of (Pt-N) (au)	E_{LUMO} (eV)
1	2	0	-12.90	-266.66	0.2342	-0.503
2	2	0	-12.74	-254.89	0.2302	-0.474
3	2	0	-12.42	-250.61	0.2297	-0.464
4	2	0	-12.35	-251.44	0.2304	-0.465
5	2	0	-12.44	-246.62	0.2291	-0.452
6	1	1	-8.75	-238.49	0.2375	-0.439
7	2	0	-12.21	-239.94	0.2281	-0.437
8	0	3	-4.41	-225.61	0.2390	-0.423
9	0	3	-4.50	-220.26	0.2383	-0.414
10	2	2	-13.82	-245.97	0.2284	-0.456
11	2	0	-12.12	-247.17	0.2296	-0.456
12	2	0	-12.14	-246.52	0.2286	-0.454
13	2	2	-13.60	-246.87	0.2293	-0.454

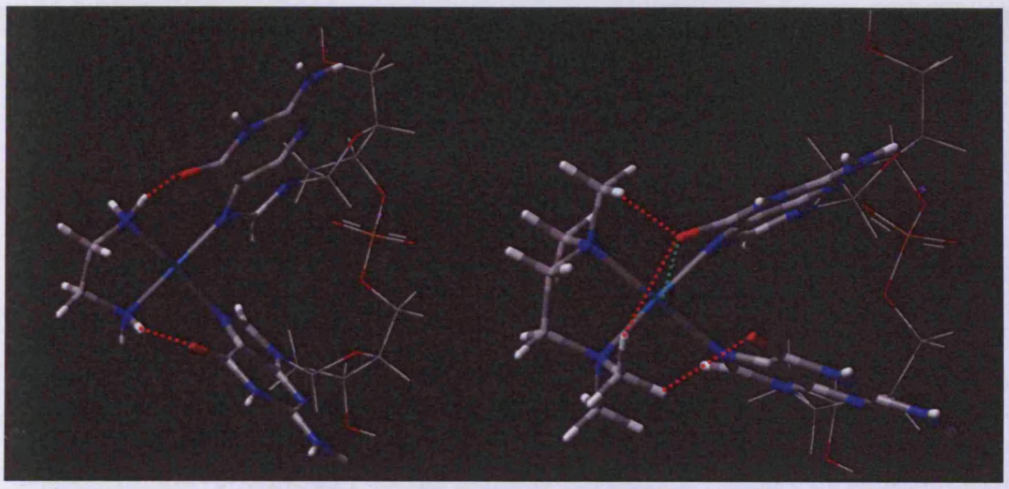


Figure 4.3.4 Non-covalent interactions located by QTAIM analysis for i) **2**-ss, showing two N—H···O contacts; and ii) **9**-ss, showing three C-H…O (red) and one weak Pt-O_6 contacts (green).

4.3.4 Discussion of the C-H…O contacts

The contacts observed in 9-single stranded DNA are displayed again in Figure 4.3.5 and are marked with the letters a,b and c for discussion purposes. Table 4.3.6 shows C..O and C-H…O distances and the C-H…O angles observed for each of the a,b, and c. Distances of H…O contacts range from 2.07 to 2.76Å and hence are within the limits set out by Desiraju and Steiner[202].

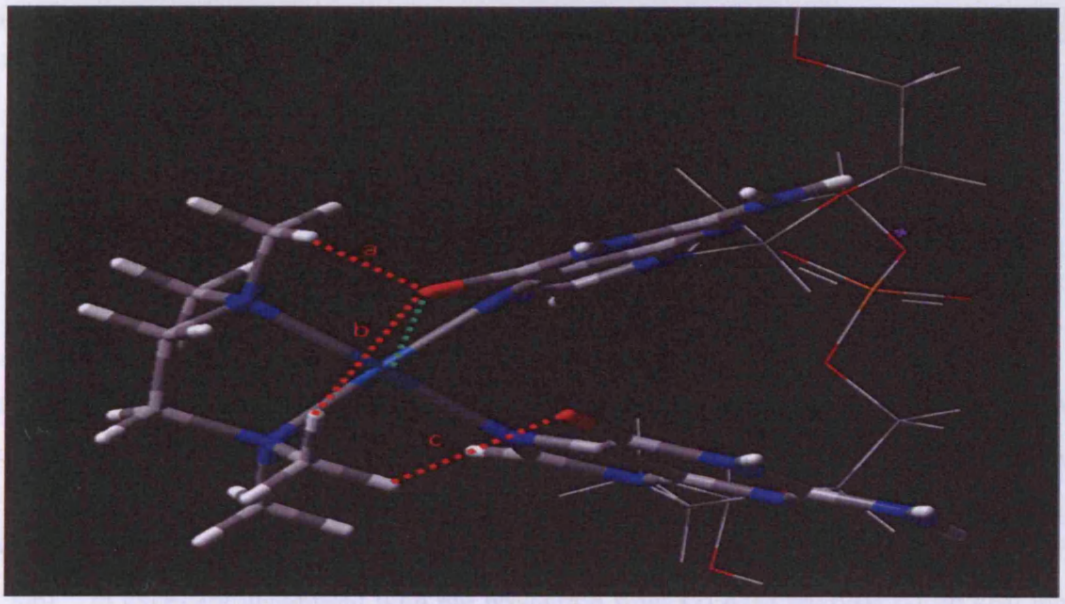


Figure 4.3.5: Observed interactions in complex 9.

Table 4.3.6: Geometric features of observed C-H…O contacts in complex 9.

	A	b	c
C-H···O (Å)	2.068	2.761	2.594
C···O (Å)	2.992	3.67	3.283
C-H···O(deg)	140.99	140.67	120.53

Detailed analysis using Koch and Popelier's criteria[134] based on atomic properties was not possible, owing to the complex topology of the density in the 9–single-stranded DNA complex, preventing accurate integration of atomic properties. However, electron density and its Laplacian at the H…O bond critical point are within the limits set out by Koch and Popelier, as are changes in natural bond orbital charges on hydrogen (+0.276 in 9–single-stranded DNA, +0.283 in the platinum fragment) (table 4.3.7).

Table 4.3.7: Charges, electron densities and Laplacian of the complex 9-ss and the model systems.

	Mulliken	nbo	ρ(r)	L
9-ss	0.218	0.276	0.008381	0.0299
$Pt(en)(CH_4)(H_2O)_2$	0.226	0.283	-	-
CH_4G	0.171	0.257	0.008395	0.0296
CH_4	0.174	0.255	-	-

A model of the interaction c of figure 4.3.5 was constructed, consisting of CH_4 interacting with O6 of guanine: a scan of C…O distance (keeping angular and dihedral parameters fixed as in 9–single-stranded DNA) indicated maximum stability for a C…O distance of approximately 3.3Å.(figure 4.3.6) At this distance, the binding energy of the CH_4…guanine complex is -1.57 kcal mol^{-1}, in excellent agreement with the prediction from QTAIM analysis (-1.46 kcal mol^{-1}). Electron density and Laplacian properties of this model were in close agreement with those from 9–single-stranded DNA (shown in table 4.3.7).

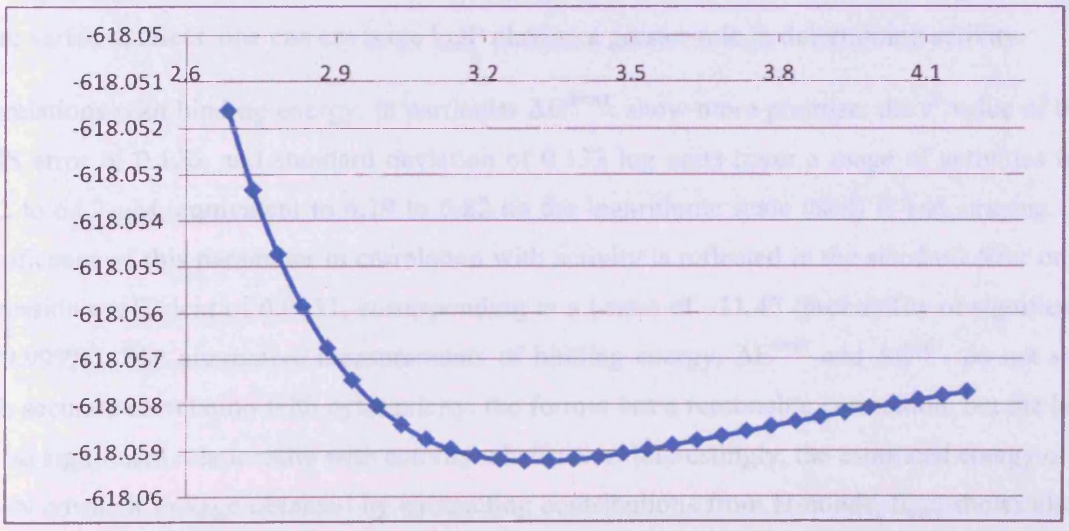


Figure 4.3.6: Scan of the C…O distance (ordinate in hartree, abscissa in Å)

4.3.5 Correlations with cytotoxicity

Cytotoxic activities of compounds **1 - 13** were reported by Milanesio *et al.* in ref. 200, where it was shown that the number and position of methyl substituents has a major effect on *in vitro* biological activity for human cultured ovarian carcinoma cells (A2780). We employed calculated geometrical and binding energy data as potential descriptors of cytotoxicity, to ascertain whether such calculations might be used to understand and predict biological activity. In addition, the octanol-water partition coefficient, logP, for each compound was estimated using the ALOGPS program, recently modified to incorporate predictions for Pt^{II} compounds[103]. Table 4.3.7 reports the results of linear regression of each selected calculated property against cytotoxicity, transformed to a free energy in the standard way as $-\log(IC_{50})$. This shows some relation between logP and activity, but one that cannot be used for prediction with any reliability. It is perhaps slightly surprising that logP shows no such correlation, since lipophilicity is widely believed to play a role in the distribution and uptake of organic as well as inorganic drugs[104, 203,204]. However, within this limited set of molecules the *position* of methyl groups, rather than simply the number of such groups, is the major factor in determining equilibrium structure and binding energy, whereas the number of methyl groups dominates the calculated logP. In larger, more varied datasets, one can envisage logP playing a greater role in determining activity.

Correlations with binding energy, in particular ΔE^{direct}, show more promise: the r^2 value of 0.94, RMS error of 0.126, and standard deviation of 0.133 log units (over a range of activities from 1.52 to 64.7 μM, equivalent to 4.19 to 5.82 on the logarithmic scale used) is encouraging. The significance of this parameter in correlation with activity is reflected in the standard error on the regression coefficient of 0.0031, corresponding to a t-ratio of −11.47 (probability of significance > 99.999%). The alternative measurements of binding energy, ΔE^{react} and ΔE^{solv}, do not show such accurate correlation with cytotoxicity: the former has a reasonable correlation, but the latter has no significant relationship with activity whatsoever. Interestingly, the estimated energy of the Pt—N covalent linkage obtained by subtracting contributions from H-bonds, E_{cov}, shows almost as accurate a correlation with activity as does ΔE^{direct}, perhaps indicating that the strength of the covalent bond formed plays a greater role in determining activity than do the non-covalent interactions present in a given complex. Figure 4.3.7 shows a plot of ΔE^{direct} *vs.* activity.

Attempts to combine descriptors into multiple linear correlations showed no improvement over the simple linear fit with ΔE^{direct}.

Table 4.3.7 Correlation between calculated properties and cytotoxicity.

Property	r^2	r^2_{adj}	RMS error
logP	0.505	0.460	0.327
ΔE^{direct}	0.943	0.935	0.126
ΔE^{react}	0.779	0.758	0.218
ΔE^{solv}	0.247	0.178	0.403
ΔE_{cov}	0.933	0.925	0.136
shift	0.532	0.489	0.318
Slide	0.218	0.147	0.411
Rise	0.298	0.234	0.389
tilt	0.397	0.342	0.361
Roll	0.507	0.462	0.326
twist	0.275	0.209	0.396

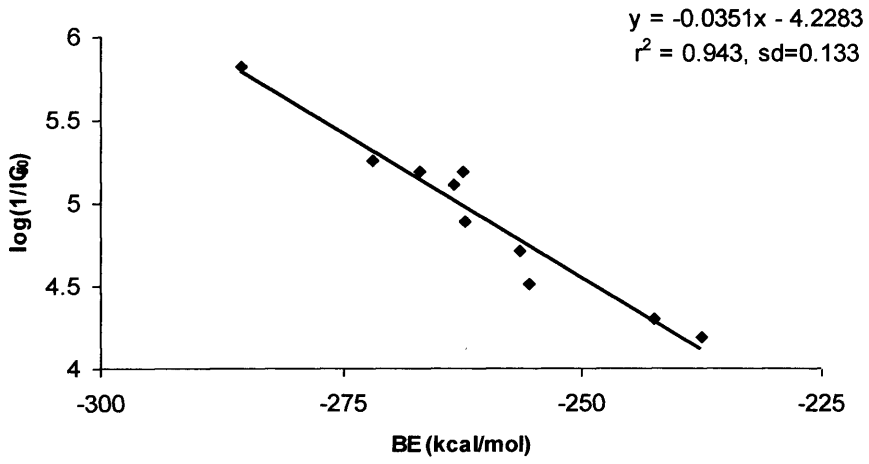


Figure 4.3.7 Plot of ΔE^{direct} *vs.* activity.

4.4 Comparison of DNA binding of cisplatin, oxaliplatin, heptaplatin, lobaplatin

Apart from cisplatin, carboplatin and oxaliplatin that were mentioned in section 4.3, three more platinum-based drugs have found approval for clinical use, namely nedaplatin, heptaplatin and lobaplatin[6]. All six drugs are collectively shown in figure 4.4.1. Among these, nedaplatin and carboplatin are characterized by a modification of the leaving group relative to cisplatin, while oxaliplatin, heptaplatin and lobaplatin have both carrier ligands and leaving groups altered. Therefore, only the last three drugs are expected to show variations in binding to DNA when compared to cisplatin and are presently discussed.

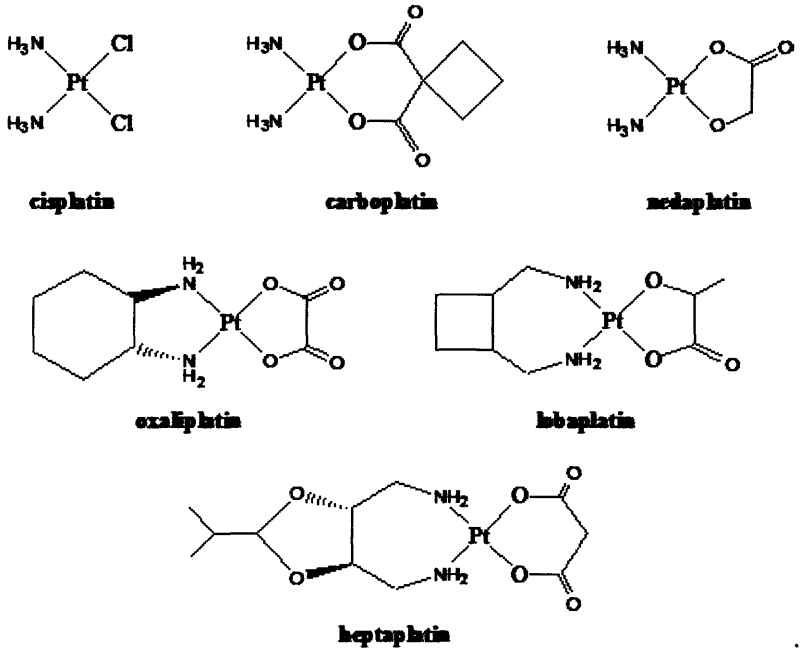


Figure 4.4.1: Approved platinum drugs.

In the following sections oxaliplatin (oxaliPt), heptaplatin (heptaPt) and lobaplatin (lobaPt) are initially compared to cisplatin (cisPt) in an identical manner as the complexes of section 4.3. Optimised geometries and energies have all been computed in an identical manner as in the previous section.

4.4.1 Complexes with single- and double-stranded dinucleotides

Optimised geometries of the oxaliPt, heptaPt and lobaPt are shown in figure 4.4.2. Due to the lack of significant variations in the ss complexes, only geometries of the ds complexes are shown. Key geometrical parameters for both the ss and ds complexes are displayed in table 4.4.1. The Pt-N7 bond lengths of oxaliPt are slightly shorter in the ds complex than in the ss, as observed with complexes of section 4.3. In the cases of heptaPt and lobaPt this picture is somehow altered, with the one of the two Pt-N7 bonds of the ds complexes being slightly shorter and the other being slightly longer than their respective bonds in the ss complexes. All three complexes also exhibit the same change in the guanine interplanar angles when moving from the ss to the ds complexes, with the latter approaching an almost parallel orientation of the two guanines.

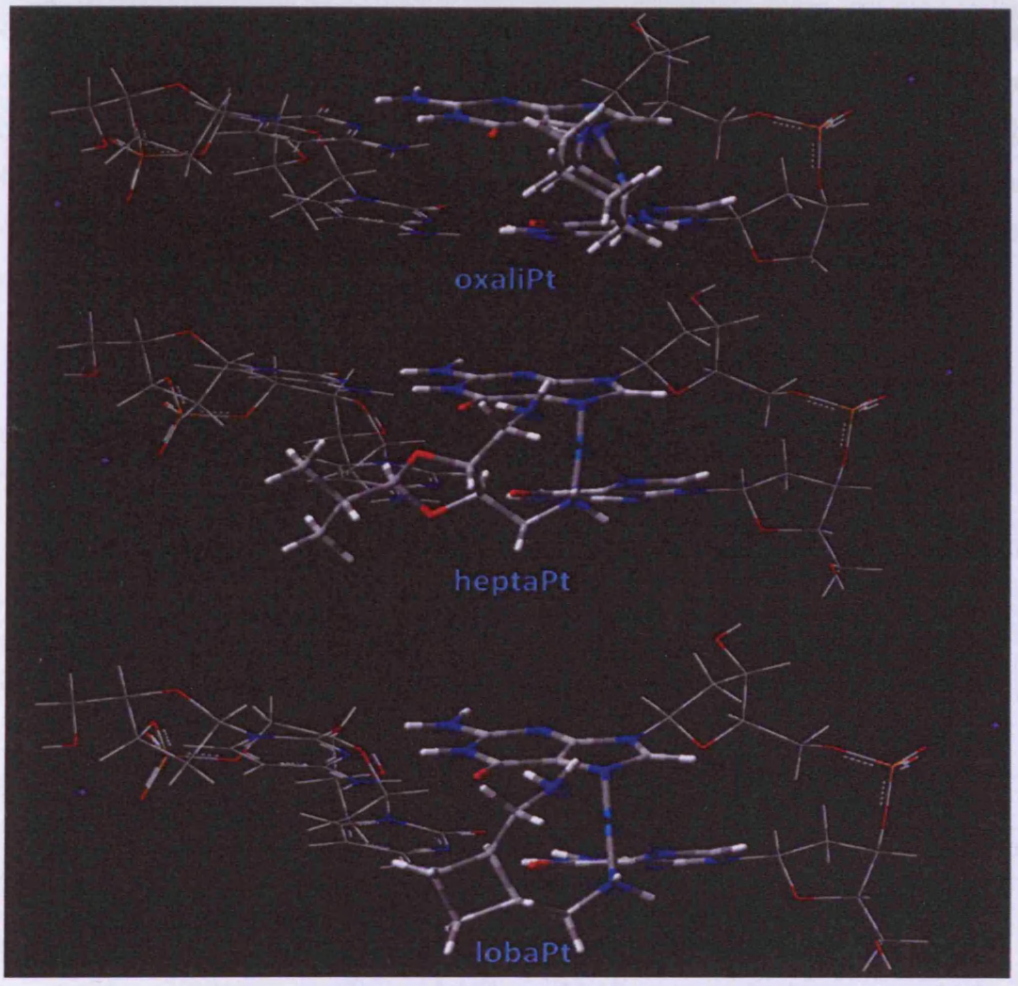


Figure 4.4.2: Optimised geometries of the three platinum complexes with ds-GpG.CpC.

Curves analysis was performed again in order to compare the induced distortions of the complexation with respect to the unbound GpG.CpC DNA fragment. The changes in the "shift" and "slide" base-pair parameters are similar to that of cisPt for all three complexes, ranging from 0.2-0.3Å. A different behaviour of the "rise" parameter is observed for heptaPt and lobaPt: cisPt and oxaliPt show a slight increase for this parameter in comparison to the free DNA, while a slight decrease is observed for heptaPt and lobaPt. Similar differences between the four complexes are observed for the angular parameters (tilt, roll, twist), as well as for the average GC pairing parameters. In general, induced distortions on the GpG.CpC show similar trends

between cisPt and oxaliPt, but different patterns are observed for heptaPt and lobaPt. It should however be noted, that in contrast to the ethylenediamine derivatives explored in the previous section, the present four complexes do not form a "family" in terms of their carrier ligands and such differences in the observed trends could be expected. Furthermore, the ligands of hetpaPt and lobaPt are significantly bulkier than those of the rest of the complexes and their effects on the CpC.GpG are likely to be more complicated.

Table 4.4.1: Selected geometrical data for single- and double-stranded complexes.

				Single-stranded			
	Pt-N1	Pt-N2	Pt-N7A	Pt-N7B	N7A-Pt-N7B	N1-Pt-N2	guan/guan
cisPt	2.031	2.033	2.027	2.024	90.8	92.8	86.5
oxaliPt	2.024	2.027	2.029	2.033	90.77	83.28	88.3
heptaPt	2.062	2.053	2.032	2.023	89.55	101.44	78.9
lobaPt	2.047	2.048	2.031	2.025	89.16	99.05	77.9
				Double-stranded			
	Pt-N1	Pt-N2	Pt-N7A	Pt-N7B	N7A-Pt-N7B	N1-Pt-N2	guan/guan
cisPt	2.045	2.035	2.019	2.016	84.9	94.2	26.4
oxaliPt	2.035	2.025	2.026	2.023	84.17	83.43	27.58
heptaPt	2.048	2.046	2.022	2.030	84.67	94.26	25.24
lobaPt	2.042	2.046	2.021	2.036	84.40	97.86	26.14

This fact is more clearly observed by performing QTAIM analysis on the ds complexes, appropriately truncated not to include the sugar-phosphate backbone. While in the cases of cisPt and oxaliPt only one BCP is found between the drug's ligands and the DNA fragment, for lobaPt the number BCPs comes up to four and for heptaPt it is further increased to nine. Among these, one and four BCPs are found between Pt's ligands and cytosine atoms for lobaPt and hetpaPt, respectively, indicating that the latter two complexes are characterised by more subtle interactions. Figure 4.4.3 attempts to display these interactions for the case of the heptaPt complex although such molecular graphs are rather complex to visualise.

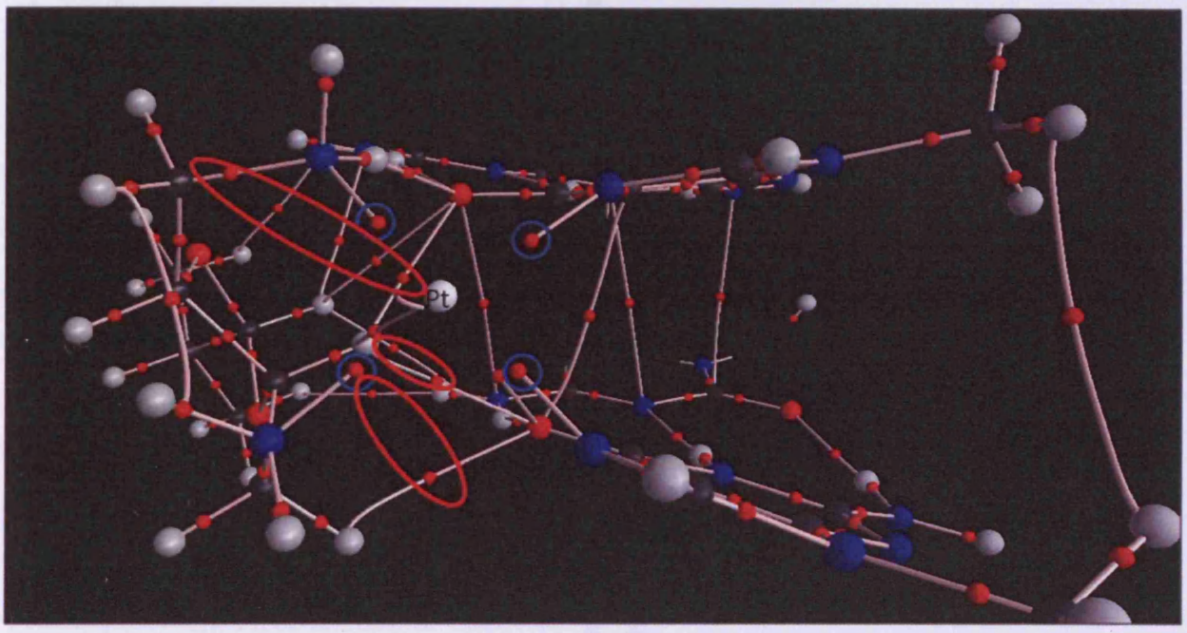


Figure 4.4.3: Molecular graph of the heptaPt-GpG.CpC complex. For clarity, the Pt atom is marked with its symbol and around it are the four Pt-N BCPs circled in blue. Most of the right-hand side of the image is covered by the GC pairs while on the left is the heptaPt drug. The nine intermolecular BCPs between the drug's ligands and the nucleobases have been circled in red.

Table 4.4.2: Base-pair geometries for platinated GC pairs (Å and °) [a]

	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (τ)	Roll (ρ)	Twist (ω)	Shear (Sx)	Stretch (Sy)	Stagger (Sz)	Buckle (κ)	Propel (ω)	Opening (σ)
cisPt	1.11	-2.07	3.90	-1.31	6.91	31.92	-0.17	-0.13	-0.20	8.19	-4.42	7.35
oxaliPt	1.11	-2.16	3.98	-1.72	6.97	32.42	-0.2	-0.11	-0.19	9.87	-4.2	8.5
heptaPt	1.16	-2.03	3.70	-2.31	4.43	30.75	-0.32	-0.26	-0.23	15.36	5.3	9.07
lobaPt	1.05	-2.05	3.76	-2.06	8.70	29.89	-0.26	-0.19	-0.21	14.64	0.9	9.51
GpG_relax [b]	0.87	-1.73	3.83	+2.27	-1.02	36.16	-0.02	0.14	0.02	14.25	-7.79	4.41

[a] Shift, slide, rise, tilt, roll and twist refer to the orientation between GC pairs, while shear, stretch, stagger, buckle, propel and opening refer to orientations within GC pairs, averaged over both; [b]Obtained from QM/MM optimisation of GpG.CpC using the same methods.

Quantification of the strength of these interactions based on the BCPs as described in section 3.1.3 yields a sum of -7.32, -6.90 and -9.19 kcal mol^{-1} for cisPt, oxaliPt and lobaPt, respectively, while for heptaPt the strength is nearly twice that of cisPt, summing up to -14.55 kcal mol^{-1}. However, these values are only indicative, since the nature of the BCPs found are quite different from those used to develop equations 3.1.1 and 3.1.2, indicating that a more detailed analysis is required.

Table 4.4.3 reports binding energies of the ss and ds complexes. As shown for the ethylenediamine derivatives, ds complexes are characterised by increased stability relative to the ss complexes. This fact could be expected at least for the lobaPt and heptaPt, since on the basis of the QTAIM analysis that was discussed in the above paragraph, the cytosine pairs are involved in interactions with the Pt's ligands. This is most probably the reason for the larger differences between the ss and ds binding energies observed for the lobaPt and heptaPt compared to those of cisPt and oxaliPt, a fact that was also observed in the analysis of the complexes of section 4.3.2. The ds complex of cisPt is found more stable than the ss by 5.79 kcal mol^{-1} in terms of the ΔE^{direct} measure and by 25.57 kcal mol^{-1} in terms of the ΔE^{react}. The respective values for heptaPt, which is characterised by the most additional interactions, are found to be 14.30 and 46.52 kcal mol^{-1}. The trend is reversed when considering the ΔE^{solv} and the ss complexes are shown to be more stable which most possibly stems from the relative solvation of the individual fragments.

Table 4.4.3: Binding energies to ss- and ds-DNA (kcal mol^{-1}).

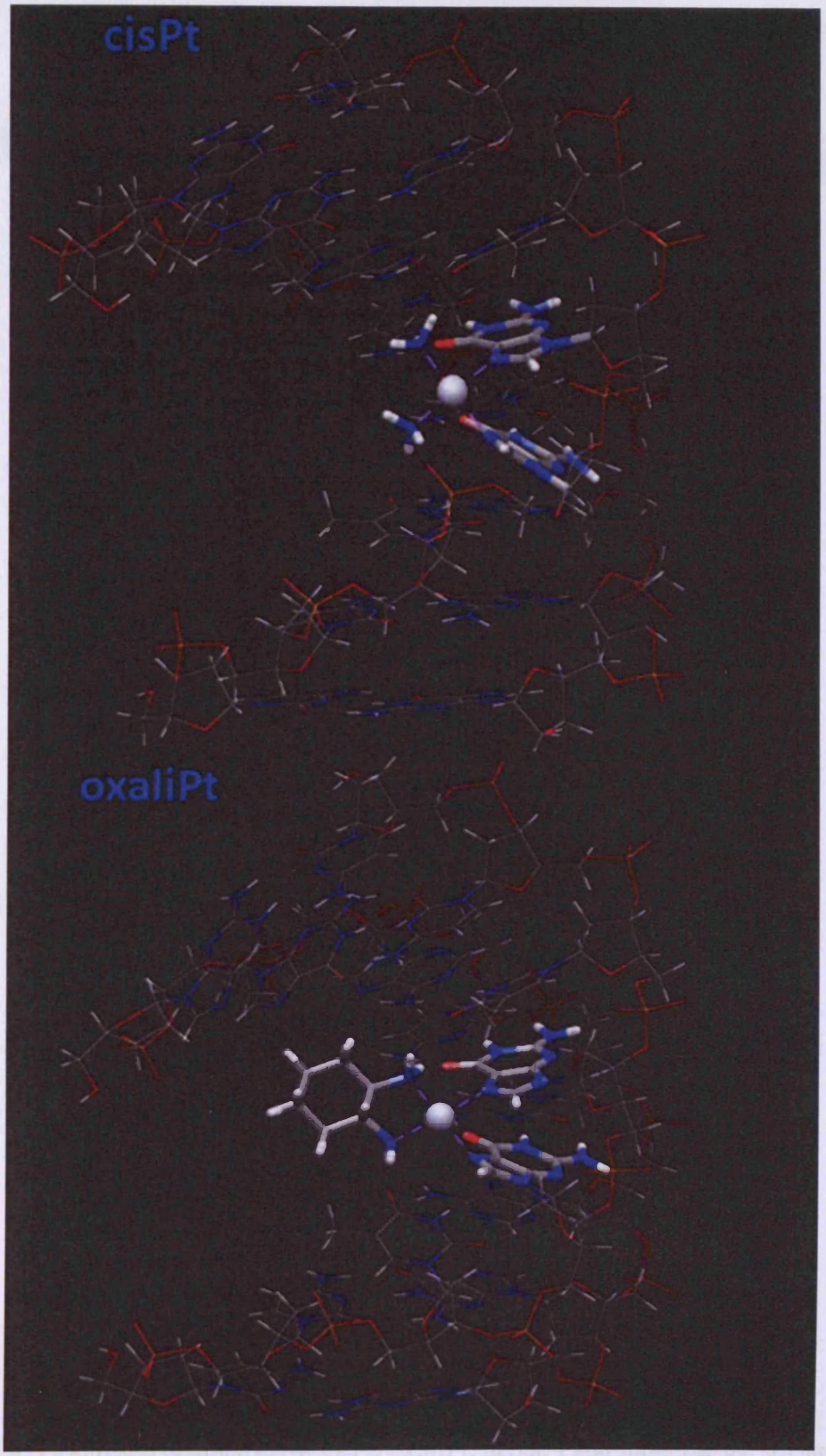
	ΔE^{direct}		ΔE^{react}		ΔE^{solv}	
	ss	ds	ss	ds	ss	ds
cisPt	-279.57	-285.54	-112.40	-137.97	-73.90	-56.73
oxaliPt	-255.96	-259.71	-93.47	-127.65	-83.77	-46.46
heptaPt	-257.09	-271.39	-88.64	-135.16	-69.95	-54.67
lobaPt	-250.38	-260.23	-87.01	-127.34	-66.41	-52.16

4.4.2 Oxaliplatin, heptaplatin and lobaplatin complexes with DNA octamers.

Additionally to the analysis of the ss and ds complexes of oxaliPt, lobaPt and heptaPt geometry optimizations were performed for complexes with double-stranded octamers of the B-DNA form, aiming at gaining insights of the induced structural distortions at a larger scale. The sequence of the DNA octamer used here is the d(CpCpTpG*pG*pTpCpC).(GpGpApCpCpApGpG), with "*" indicating the sites of platination. This sequence was chosen to match the one used for the results of section 3.4.6, thus allowing for comparisons with the respective cisPt complex that was analysed there. The optimizations were performed using the ONIOM method using electrostatic embedding, without use of solvent due to time limitations. Therefore, the structures of oxaliPt, lobaPt and heptaPt complexes with the DNA octamers are presently compared to the cisPt "embed" of section 3.4.6, which in turn has already been compared to the respective crystal structure. All the calculation details are identical to the ones used for deriving the cisPt "embed" structure.

The optimised geometries of the four complexes are displayed in figure 4.4.4. Due to the size of the systems default convergence criteria could not be met, especially for the complex heptaPt. However, the maximum and RMS forces were sufficiently low with respect to the size of the complexes, while at the same time no significant geometry perturbations occurred around the QM region, which is the area of main interest. Thus, the geometries obtained and shown on figure 4.4.4 were considered suitable for evaluation and comparison of the complexes' helical parameters.

From figure 4.4.4 it can also be readily seen that in all four complexes platination induces a bending of the helix towards the major groove, with a tendency to enclose the drug. This bending of the helix brings bulky Pt ligands spatially close to nucleobases other than those of the site of platination (G4, G5), therefore possibly giving rise to additional non-bonded stabilizing interactions between atoms of the QM and MM regions. For this reason, no energetic or QTAIM analyses were performed at this stage and the present discussion is limited to the Curves parameters around the QM region.


cisPt
oxaliPt

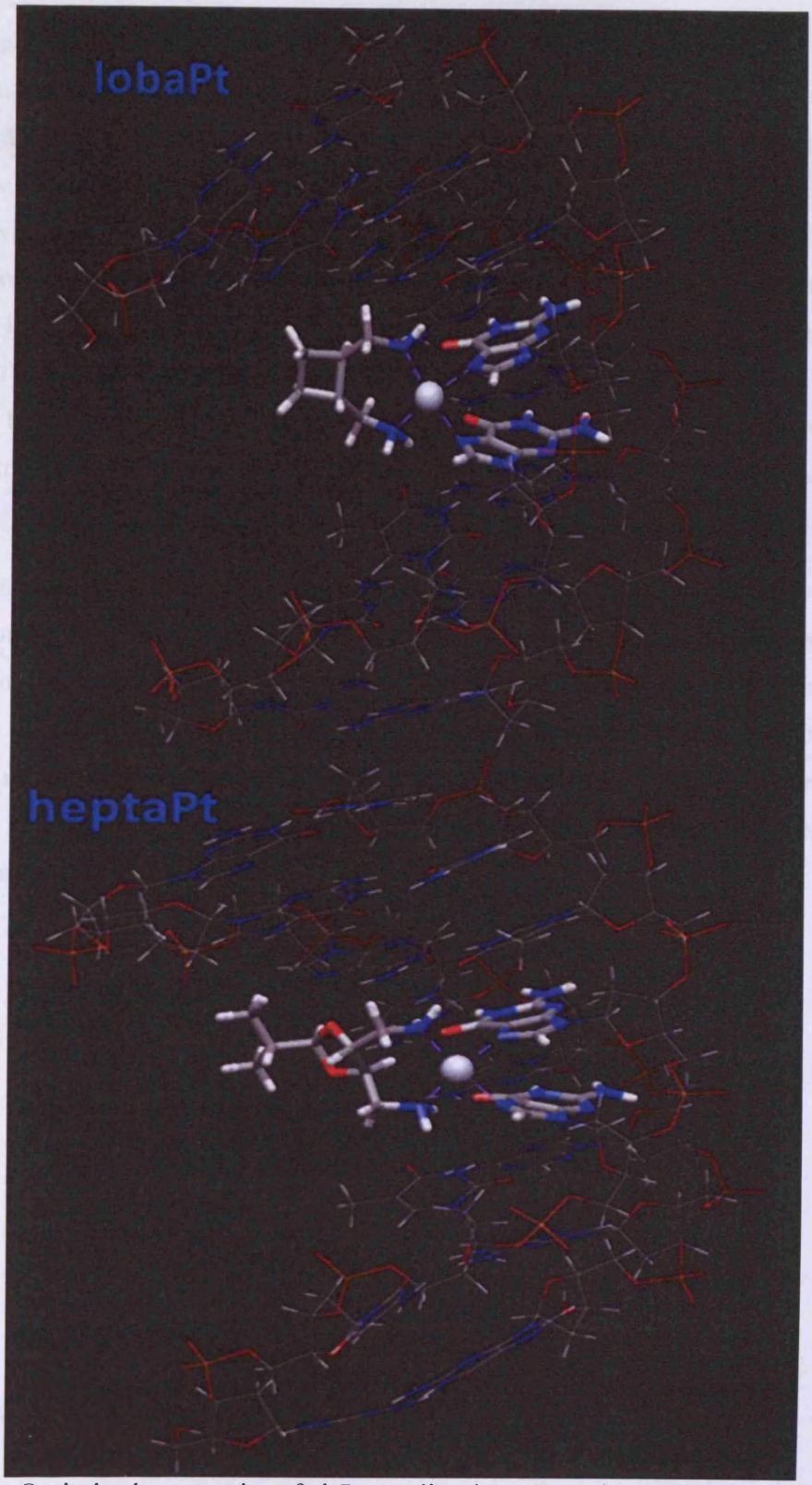


Figure 4.4.4: Optimised geometries of cisPt, oxaliPt, heptaPt and lobaPt at the BHandH/6-31+G**/SDD level.

The local parameters between the stacked G4-C5 and G5-C4 base pairs are graphically displayed in figure 4.4.5. A first look at the graphs reveals that the four helices are similarly affected by the complexation with platinum and subtle quantitative differences are observed. The positive values of "roll" indicate an opening towards the minor groove or equivalently the bending towards the major groove that was mentioned in the previous paragraph. This opening is more pronounced in the cisPt complex with a value of 28.2° and drops to 20.6° and 15.8° for oxaliPt and lobaPt, respectively, while heptaPt has a "roll" angle only slightly greater than that of lobaPt (17.3°). Differences less than 2° among the complexes are observed for the "tilt" angles. The latter have the same (negative) sign in all four cases indicating an opening towards the same strand, which is not surprising since the binding site is the same among all four adducts. Insignificant differences (less than 4°) are also observed for the "twist" angles.

A similar picture is observed for the translational parameters of the stacked pair. Translation along the y-axis ("slide") occurs towards the same direction for all the complexes with small variations among them, while the respective dislocation along the x-axis ("shift") is practically the same in all cases. Finally, the vertical separation of the two base pairs indicated by the "rise" parameter is almost the same between cisPt and oxaliPt (3.55Å and 3.6Å, respectively). LobaPt and heptaPt are characterized by slightly shorter inter-base pair distances, 3.27Å for the former and 2.99 Å for the latter.

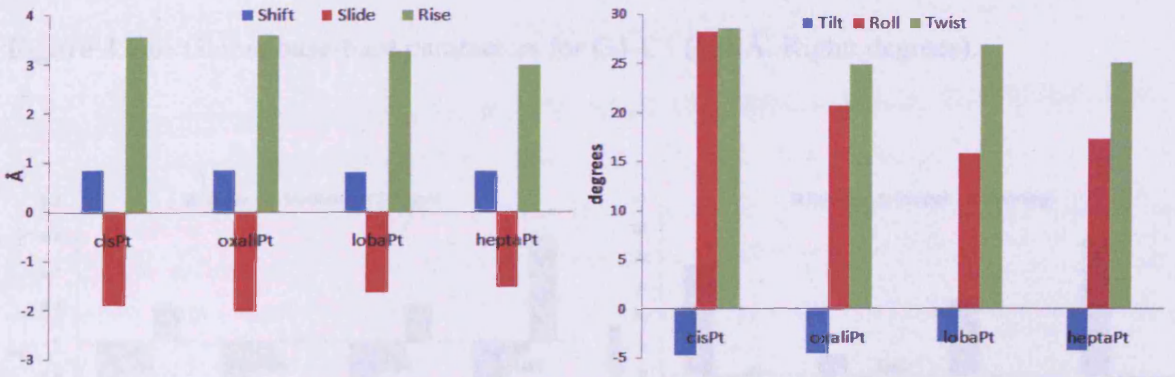


Figure 4.4.5: Local inter-base pair parameters for G4/G5.

The consequences of platination on the two GC pairs (G4C5 and G5C4) are also graphically displayed in figures 4.4.6 and 4.4.7. It can be seen that for each GC pair the qualitative

features are the same among the complexes with only two exceptions, both related to the heptaPt complex. The first of the exceptions is observed for the "stagger" parameter of the G4C5 pair (figure 4.4.6) and the second is related to the "opening" angle of the G5C4 pair (figure 4.4.7). These two qualitative differences most likely stem from the fact that among the four complexes the heptaPt complex is characterized by the maximum number of interactions between the Pt's ligand and the neighbouring nucleobases.

Quantitatively, a significant difference is observed in the relative propeller twist ("propel") of the G4C5 pairs. The maximum difference is between the cisPt and oxaliPt reaching almost 15° and the values of the "propel" parameter follow the order oxaliPt < lobaPt < heptaPt < cisPt and is to some extent related to the roll angle of figure 4.4.7. A full detailed quantitative analysis of the Curves parameters is beyond the scope of the present section and the details for each complex can be found in the appendix (tables 7.5 - 7.12).

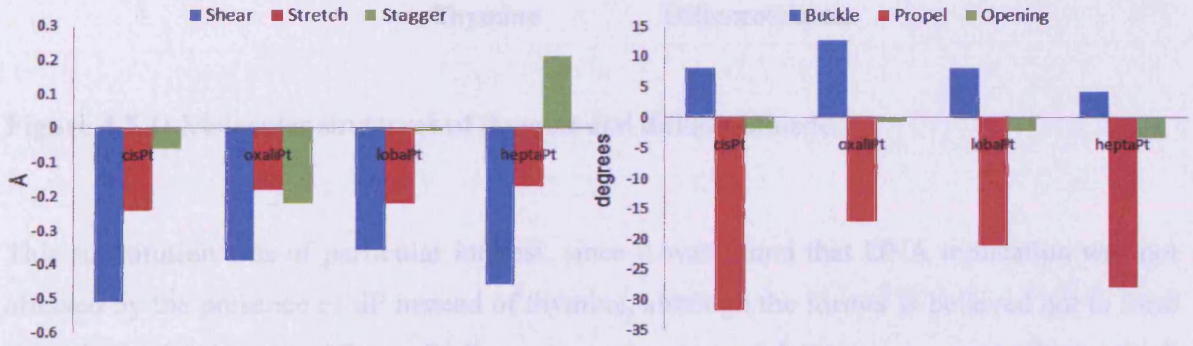


Figure 4.4.6: Global base-base parameters for G4-C5 (left:Å. Right: degrees).

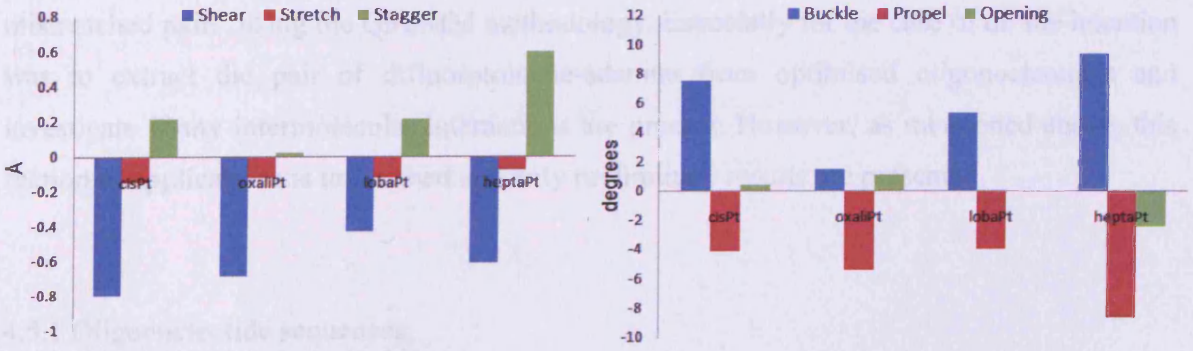
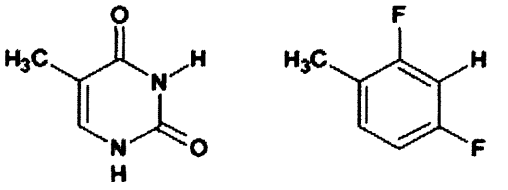


Figure 4.4.7: Global base-base for G5-C4.

4.5 Modified nucleic acid bases

In 1994, Schweitzer and Kool intended to study the importance of the polarity of nucleic acid bases in DNA-related processes by replacing a DNA base with molecules of similar size and less polar[205]. The resulting nucleosides were termed as "nonpolar nucleoside isosteres". One example of such an isostere was the 2,4-difluorotoluene (dF), which closely matches the size of thymine (figure 4.5.1), while the replacement of the nitrogen and carbonyl oxygen atoms of thymine by carbon and fluorine atoms, respectively, alters the polarity of the molecule.



Thymine **Difluorotoluene**

Figure 4.5.1: Molecular structures of thymine and difluorotoluene.

This substitution was of particular interest, since it was found that DNA replication was not affected by the presence of dF instead of thymine, although the former is believed not to form Watson-Crick H-bonds. These findings gave rise to a debate among researchers which included the question whether dF is H-bonded to adenine or not[206].

The present section, although largely unfinished, as a first step deals with modeling the structural deformation of the DNA helix, as a result of the presence of dF or various mismatched pairs, using the QM/MM methodology. Especially for the case of dF the intention was to extract the pair of difluorotoluene-adenine from optimised oligonucleotides and investigate if any intermolecular interactions are present. However, as mentioned above, this section of applications is unfinished and only preliminary results are presented.

4.5.1 Oligonucleotide sequences

In ref. 206 (and relevant references therein) it is reported that a series of oligonucleotide structures involving dF and other isosteres, as well as mismatched nucleobase pairs were

studied, aiming at determining their relative thermodynamic stability. Our first step was to carry out geometry optimisations on oligonucleotide fragments of the same sequence as the one reported in ref. 206 and is shown in table 4.5.1.

Table 4.5.1: Oligonucleotide sequences.

– CTTTTC(X)TTCTT – GAAAAG(Y)AAGAA	
XY	**XY**
TA	AT
TG	GT
TC	CT
TT	TT
FA	AF
FG	GF
FC	CF
FT	TF

Initially, only the sequence of the left column of table 4.5.1 was considered and the DNA duplexes were truncated to pentamers with the XY pair as the central fragment. Thus, the studied sequence was 5' -TCXTT, 3' –AGYAA. Taking into account the XY combinations leads to 8 possible oligonucleotides, 4 with the F molecule as X and each of the A,G,C,T paired to it and 4 with T as X paired to each of the four common nucleobases. The combinations of XY and the numbering used to identify the nucleotides are shown in table 4.5.2.

Table 4.5.2: X-Y pairs and numbering of complexes.

X	T	T	T	T	F	F	F	F
Y	A	G	C	T	A	G	C	T
#	1	3	5	7	2	4	6	8

4.5.2 Fluorine parameters and QM region

Initial calculations were performed on nucleotide 2, which necessitates the definition of MM parameters related to the fluorine atom. The carbon atom of the difluorotoluene ring that binds

to the backbone was assigned the atom type C*, while the carbon atoms that bear the F atom were described as CJ. Both the C* and CJ atom types, as well as the F atom, were described by the same "soft" forcefield parameters that were used in our calculations that involved Ru or Pt. Parameters for the difluorotoluene molecule have been derived by Cubero *et al.*[207] and have also been used in our calculations. Additionally, the absence of metal atoms in the present systems allows for the use of microiterations during the QM/MM optimisations. As a result, six different approaches for structure 2 (2a – 2f) were performed, differing in the parameters that were used, the definition of the QM region and the use of microiterations. The details of the six optimisations are summarised in Table 4.5.3. In all the calculations the GDIIS algorithm was used for geometry optimisations at the BHandH/6-31+G** level and Na counter-ions have been used to neutralise the molecules, while time-limitations again prevented us from including solvent molecules.

Table 4.5.3: Summary of calculations on complex 2.

molecule	parameters	micro	QM region
2a	soft	on	AF H-bond
2b	soft	off	AF H-bond
2c	lit.[a]	on	AF H-bond
2d	soft	on	FC stack
2e	lit.[a]	on	FC stack
2f	soft	on	FT stack

[a] Literature parameters

Figure 4.5.2 displays the geometry of 2a optimised using microiterations, soft parameters and the F-A pair at the QM level. Turning the microiterations off resulted in 2b structure (Fig. 4.5.3) which shows a severe distortion of the system, compared to a regular DNA nucleotide, with the neighbouring bases to the QM region having a tendency to adopt a perpendicular orientation towards the central pair. The structure shown in figure 4.5.3 is not fully optimised, since the observed distortion combined with the length of the optimisation (50 and 505 optimisation steps for 2a and 2b, respectively) were considered unreasonable and thus the optimisation was stopped. As a result, all of the following calculations were performed using microiterations.

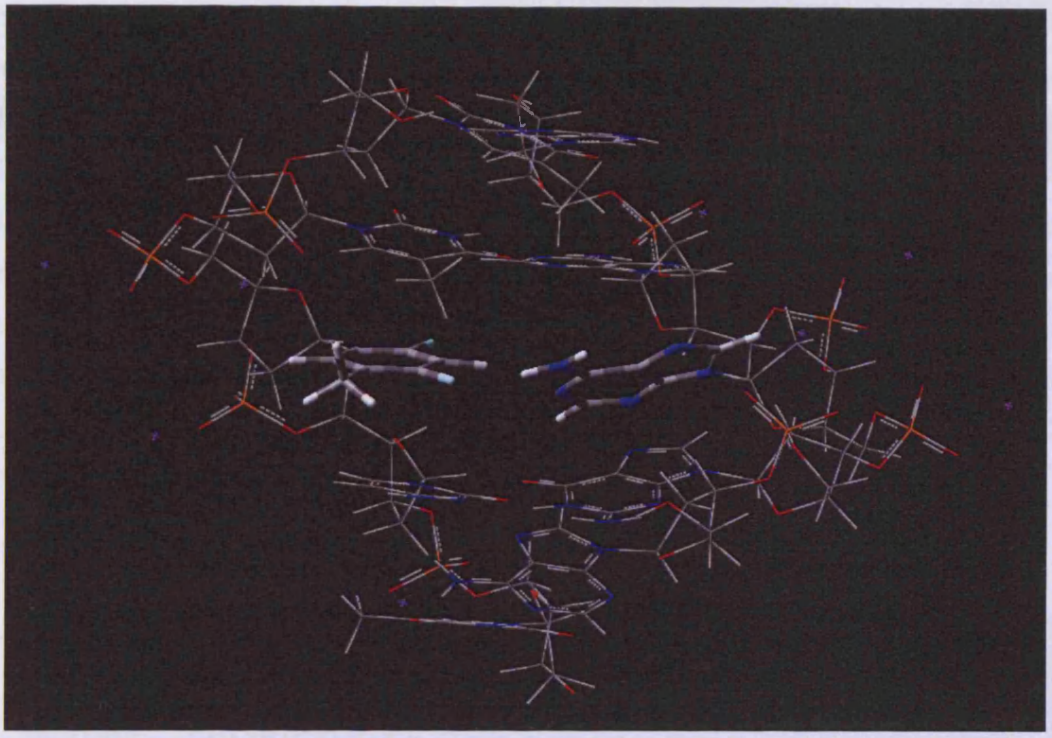


Figure 4.5.2: Optimised geometry of complex 2a at the BHandH/6-31+G** level, using microiterations and soft parameters.

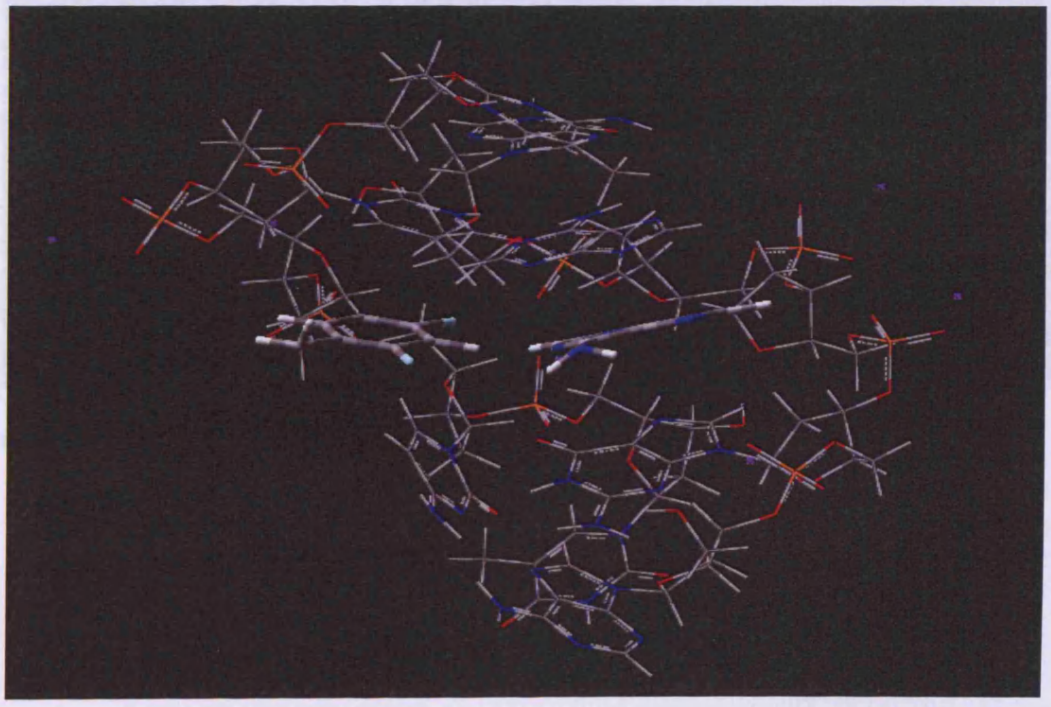


Figure 4.5.3: Distorted structure during optimization of 2b.

Repeating the calculation with microiterations and using the literature parameters (structure 2c) had practically no effect in the resulting geometry or the computational cost (43 steps) and the fully optimised nucleotides 2a and 2c are almost indistinguishable if aligned (Figure 4.5.4). This fact again supports the soft parameters approach that was also discussed in section 3.4.2 Changing the QM region to be the F-C stacked pair, again with soft and literature parameters (2d and 2e, respectively) leads to more evident discrepancies between the two approaches (fig. 4.5.5), especially when moving towards the external bases. (It must be noted, that the C* atom type used in 2a and 2b was replaced by CA in all the rest of the calculations.) The same is observed when comparing 2d to 2f, which differ in the QM region, as well as 2a to 2d and 2a to 2f.

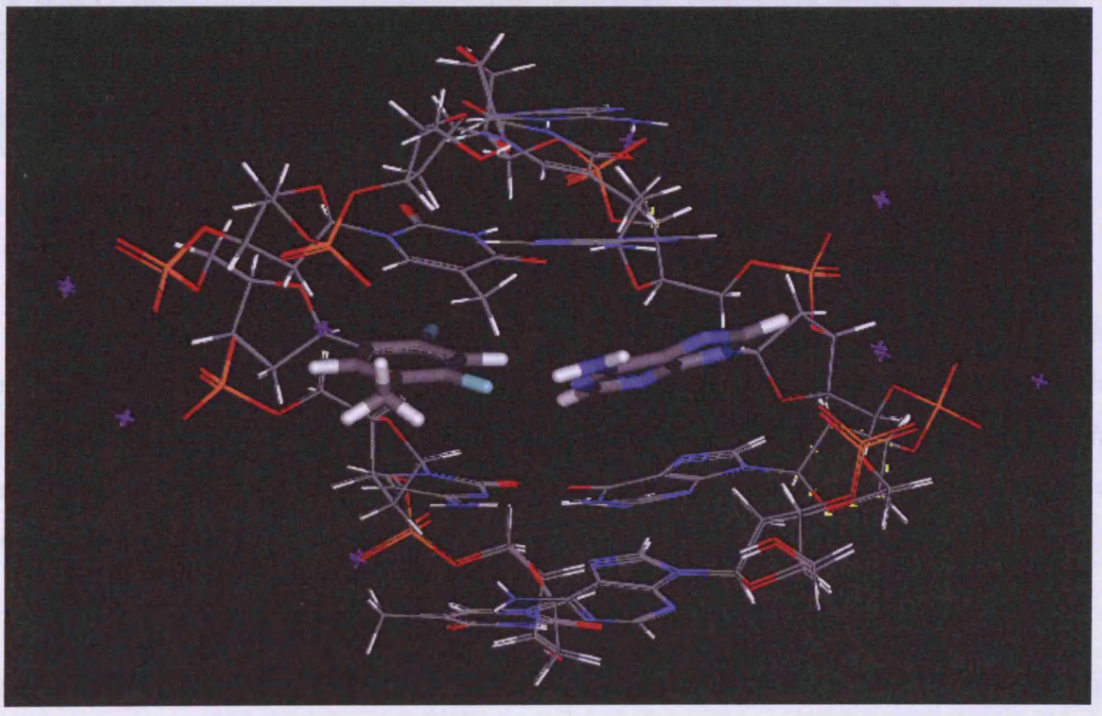


Figure 4.5.4: Aligned structures for 2a and 2c.

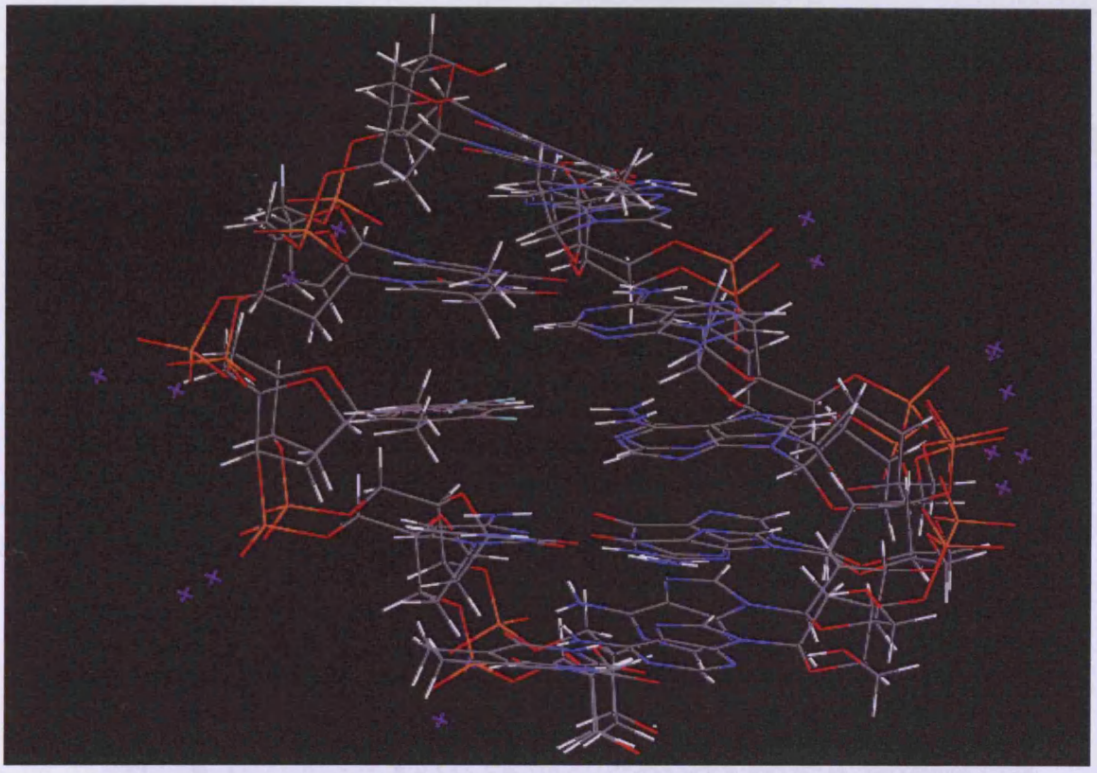


Figure 4.5.5: Aligned structures for 2d and 2e.

The fact that the dissimilarities become more evident for the external bases, as opposed to the inner ones, as well as initial observations when trying to optimize a trimer, implies that the flexibility of the systems plays a crucial role in the resulting differences. Furthermore, the overall agreement between the structures is satisfactory and can serve as a basis to consider the obtained geometries realistic.

Finally, an alternative optimization was performed on structure 2 using Cartesian coordinates and soft parameters without microiterations. As seen in the case of the cisPt calculations, this approach shows slow convergence and is time consuming. A graph of the energy and forces for this optimization is shown below (figure 4.5.6). Moreover, the resulting structure is similar to the one obtained using the GDIIS method and microiterations, which implies again that there is not significant gain by the use of Cartesian coordinates, at least for the systems considered here.

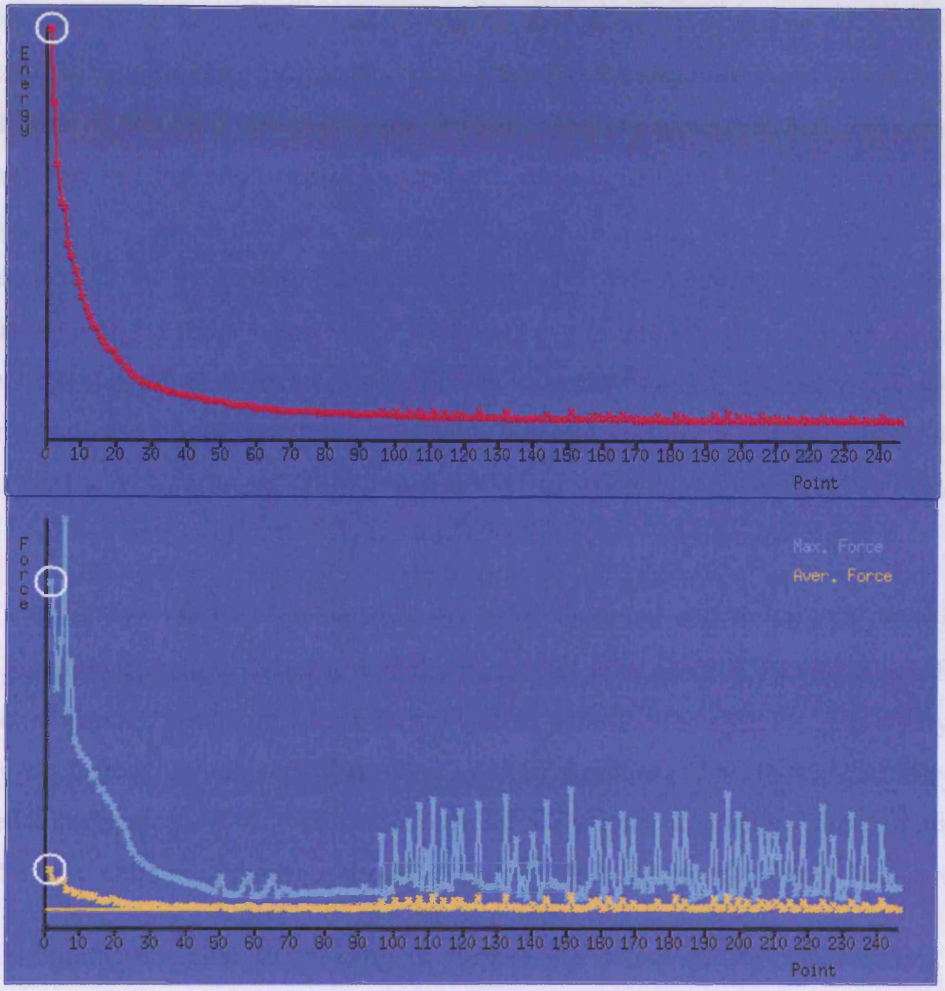


Figure 4.5.6: Energy (top) and forces (bottom) during optimisation of complex 2.

4.5.3 Geometry optimizations

Following the calculations described in the previous section, the use of soft parameters was selected for carrying out optimizations on systems that contain atoms (atom-types) not already parameterized in AMBER (in this case molecules 4,6 and 8), in order to maintain consistency between the calculations. Specification of new atom types was not necessary for molecules 1,3,5 and 7, which contain standard nucleic acid bases. However, these systems, apart from 1, contain nucleobase mismatches which can destabilize the structures and thus are challenging, at least at the level of a pentamer.

Indeed, comparison of structures 1 and 2 shows that when A is paired to T, the helical structure tends to unwind (fig. 4.5.7), a case not observed for structure 2a in which A is paired to F. The effects on the helix are even more dramatic when G is mismatched T (3) and C to T (5), both resulting in a structure that is too complex to visualize and the helical arrangement is not preserved at all. This is not case when pairing G with F (4), which results in a structure more reminiscent of the unwinding that occurs for 1, while when pairing C to F (6), the helical distortion is severe, yet mostly due to the external base pairs. The final pair of molecules, 7 and 8 also exhibit similar behavior. In particular, mismatching of T to itself (7) leads to a structure in which the 5'- strand shows a helical character and the 3'-strand tends to have the phosphate backbones in a straight line. Again, the distortion is more evident in the external bases. The structure of complex 8 is again similar in behavior to molecules 1 and 4.

The above results indicate that the systems are rather complex and flexible to optimize. It is not certain whether the same distortion of the helix will take place if moving away from the pentamer to add more base pairs which would further stabilize the systems. Especially for the case of pairing difluorotoluene to adenine, a solution structure of a DNA dodecamer shows that the helical arrangement is well-preserved and significant distortions are extended only locally to the base pairs neighbouring the dF-A pair.[208] This structure is shown in figure 4.5.8.

It would be interesting also to investigate further the effect of the mismatching on the structures, especially in cases where the distortion occurs in half of the molecule. (This may imply that the mismatching has crucial effect not only in the H-bond pairing, but also in the stacking between the bases.) Also, the specific sequence as well as the size and type of the QM region are parameters that need to be taken into account in detail.

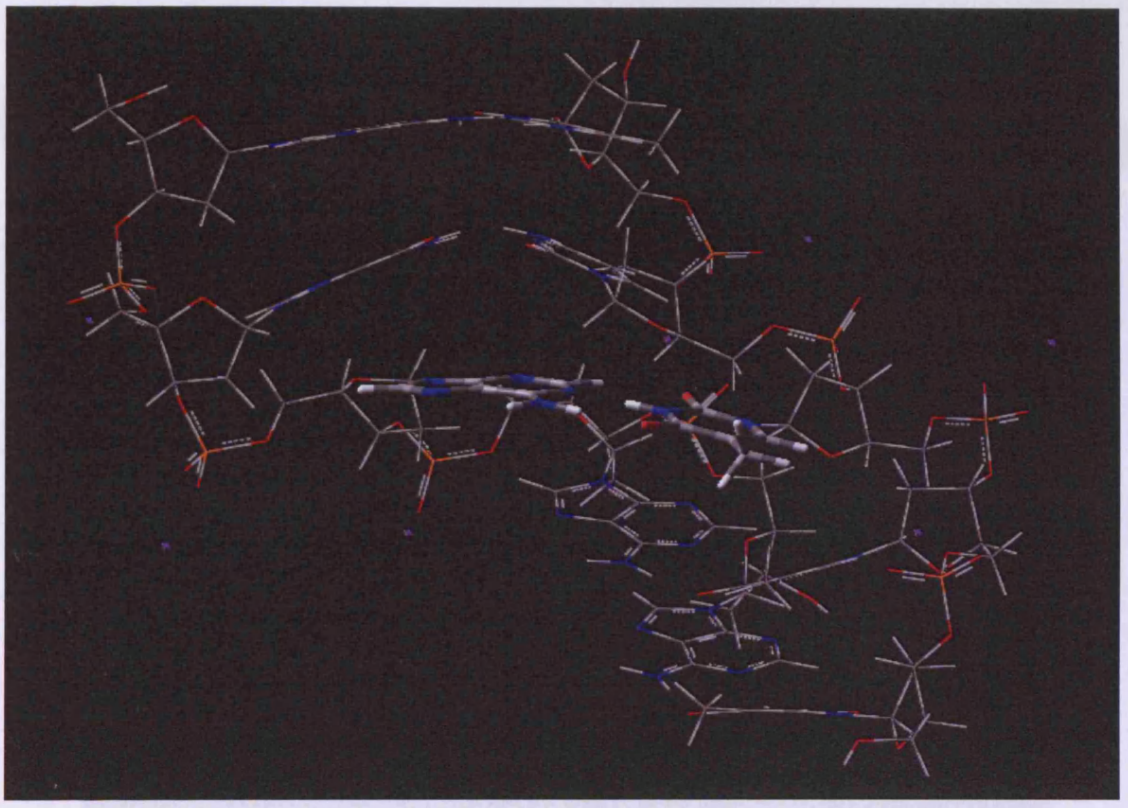


Figure 4.5.7: Optimised geometry of structure 1 at the BHandH/6-31+G** level.

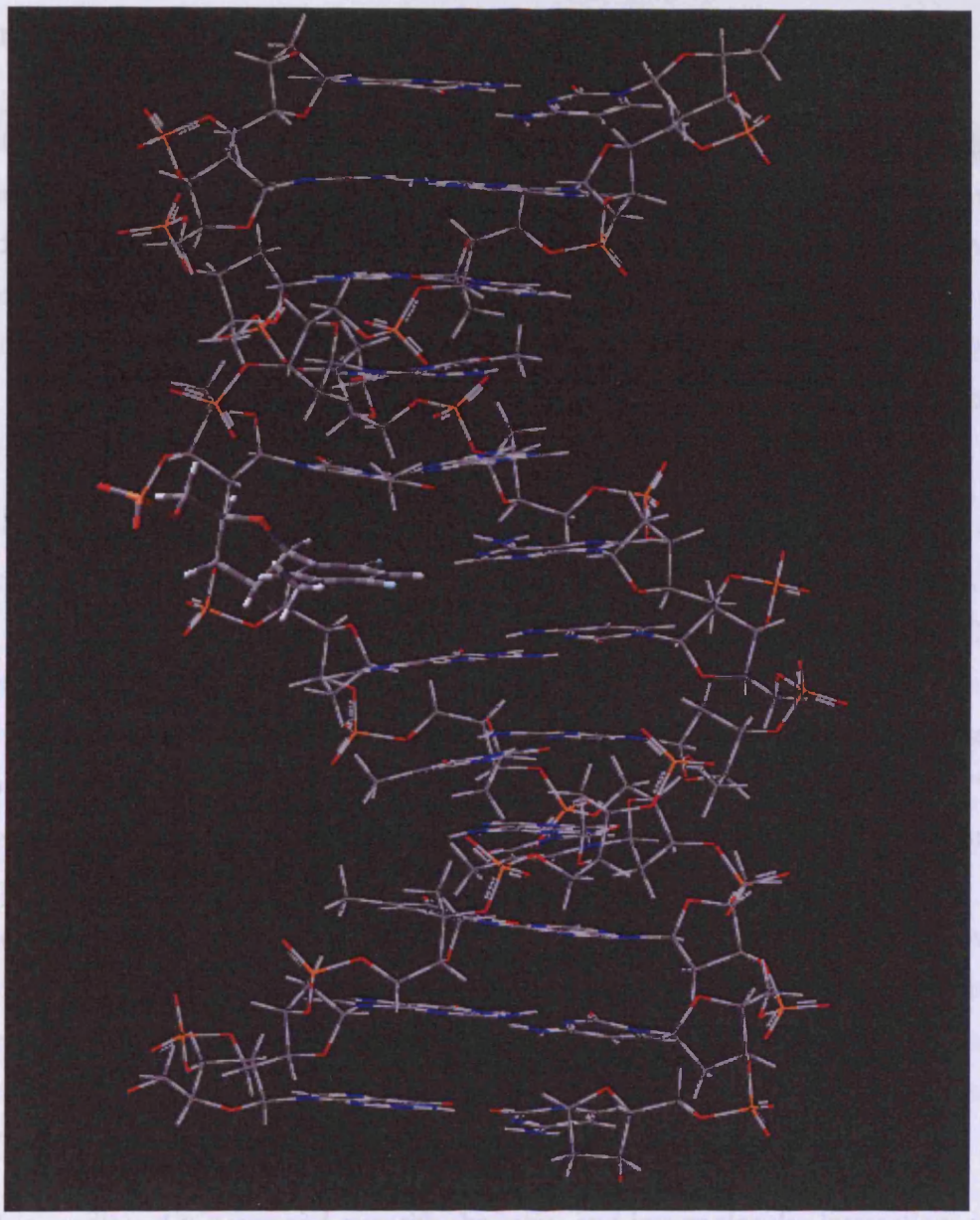


Figure 4.5.8: Structure of DNA dodecamer containing the dF-A pair (pdb entry: 1BW7).

5. Conclusions

A series of DFT and QM/MM calculations has been carried out, aiming at the investigation of the interactions between metal-containing anticancer drugs and the DNA molecule which is believed to be the biological target of such drugs. The data are supported by experiment and ab initio benchmark calculations were appropriate, while additional insights into the specific interactions that occur in the studied systems are gained by the application of the Quantum Theory of Atoms in Molecules (QTAIM), frontier molecular orbital theory and statistical fitting.

The proper description of non-covalent interactions and subsequently the study of systems characterised by such interactions is still a challenging research task, one that requires careful method selection. Based on previous observations on the successful application of BHandH for non-covalently bound systems, a systematic investigation of its performance and limitations against a dataset of 22 dimers verified that it consistently overestimates the strengths of hydrogen bonds, but performs rather better for dispersion bound and "mixed" complexes. Altering the percentage of the exchange-correlation contribution of the functional does not improve overall performance, but shortcomings can be ameliorated by scaling calculated interaction energies. In all cases reasonable geometries are obtained with an average root mean square deviation from literature geometries of 0.11 Å. In six out of 22 cases, the optimized geometries that result were found not to be true minima if the point group symmetry of literature structures is conserved. Additionally, BHandH predicts correctly the order of stability among three toluene dimer isomers and two isomers of n-octane.

QTAIM is employed to trace and quantify the intra- and intermolecular interactions with the aid of previously observed linear relationships between electron densities and interaction energies of H-bonded and π-stacked complexes. Despite the similarities of the relationships that suggested that the electron density properties of non-covalent interactions are similar, an improved linear model based on the S22 data, able to account for both types of interaction, is not achieved. However, this analysis revealed that the method or basis set is not crucial for such relationships. Deeper investigation on the S22 set shows similar correlations for more QTAIM properties, while examination on the whole JSCH database indicates that donor-acceptor patterns in H-bonded complexes have significant role in the observed trends.

Additionally, it is clarified that the disagreement in the observed topology of the density between metals and arenes and the η^6-binding is not an artefact of method or basis set.

The promising performance of BHandH and QTAIM makes it a promising method for study of biological systems such as nucleic acids and their interactions with drugs. These methods were applied to complexes of the type $[(\eta^6\text{-arene})Ru(en)(nucleobase)]^{2+}$, yielding geometries in good agreement with available crystal structures and the arenes bip, dha and tha are positioned in a stacked manner over the nucleobases. Both BHandH and M05 agree on the binding energies as a function of the arene and base considered and a clear preference of the Ru complex for guanine over any other base is demonstrated. Specifically, the preference for binding at guanine over adenine is approximately 24 kcal mol^{-1} and is reduced to around 5 kcal mol^{-1} in aqueous solution. The origins of this preference were explained on the basis of QTAIM and frontier molecular orbital data, and the effect of Ru complexation on GC pairing quantified.

To study larger fragments of DNA than single bases or base pairs, it becomes unfeasible to use DFT alone, so QM/MM (ONIOM) methods were employed. Many tests of the input parameters required for such calculations were performed, including atomic charges and types and MM parameters for the metal centres. Optimisations of double-stranded dinucleotide fragments fully converge using the GDIIS algorithm, while for larger DNA fragments more efficient geometry optimisation strategies were sought for by employing microiterations, changing the optimisation algorithms and freezing parts of the complexes during the calculations.

The combined DFT/MM approach using the ONIOM scheme was further employed for a series of geometry optimisations of cisplatin bound to the d(CpCpTpGpGpTpCpC).(GpGpApCpCpApGpG) DNA octamer. Comparison with the respective NMR structure shows that electrostatic embedding is crucial for the proper description of the systems, while inclusion of explicit solvent molecules further improves the performance of the approach, and optimised structural parameters are within acceptable ranges compared to NMR data.

The above approach was further applied to investigate the effects of methyl substitution on the binding of cis-$[Pt(en)]^{2+}$ complexes to DNA fragments, yielding structural and energetic information as consequences of the number and position of methylation. It was shown that

triple and quadruple methylation on nitrogen atoms of ethylenediamine leads to significant changes compared to other methylated variants, a fact that is partially due to the replacement of strong N-H···O hydrogen bonds with weaker C-H···O contacts, as suggested by QTAIM analysis, and also to effects on the electronic structure of Pt, as revealed by frontier molecular orbitals. Furthermore, lipophilicity measures and geometrical parameters of this series of complexes showed rather poor correlations with *in vitro* cytotoxicity data, while in contrast the latter are shown to be in excellent correlation with calculated binding energies.

Additionally, comparison of the DNA binding of the four approved platinum-based drugs cisplatin, carboplatin, oxaliplatin and heptaplatin based on the same ONIOM approach showed subtle differences occurring in key geometrical parameters at the platination sites, using both small and larger DNA fragments. Among the four complexes cisplatin is shown to be energetically more favoured in terms of calculated binding energies. QTAIM analysis reveals numerous subtle interactions between the DNA bases and the bulky ligands of lobaplatin and heptaplatin, the complexes of which are additionally characterised by surprisingly short nucleobase interplanar distances.

Finally, preliminary ONIOM results on small oligonucleotides containing difluorotoluene in place of thymine or mismatched base pairs indicate that these systems are sensitive to the choice of the QM region due to their flexibility, thus suggesting that larger nucleotide sequences be used in order to attain reliable geometries. However, comparisons among selected structures of these non-metallic systems verify the performed tests of the ONIOM input parameters for metallic systems, thus adding to the validity of our applied QM/MM approach.

6. References

[1] Peyrone, M. *Ann. Chemie Pharm.* **1845**, 51, 129

[2] Rosenberg, B. ; Van Camp, L. V. ; Krigas, T. *Nature* (London), **1965**, 205, 698-699

[3] Rosenberg, B. ; Van Camp, L. V.; Trosko, J. E. ; Mansour, V. H. *Nature* (London), **1969**, 222, 385-386

[4] Wong, E. ; Giandomenico, C. M. *Chem. Rev.* **1999**, 99, 2451

[5] Reedijk, J. *Chem. Commun* **1996**, 801

[6] Galanski, M. ; Jakupec, M. A. ; Keppler, B. K. *Curr. Med. Chem.* **2005**, 12, 2075

[7] Reedijk, J. *Coord. Chem. Rev.* **2005**, 249, 2845-2853

[8] Van Beusichem, M. ; Farrell, N. *Inorg. Chem.* **1992**, 31, 634-639

[9] Farrell, N. ; Kelland, L. R. ; Roberts, J. D. ; Van Beusichem, M. *Cancer Res.* **1992**, 52, 5065-5072

[10] Basch, H. ; Krauss, M. ; Stevens, W. J. ; Cohen, D. *Inorg. Chem.* **1985,** 24, 3313-3317

[11] Carloni, P. ; Andreoni, W. ; Hutter, J. ; Curioni, A. ; Giannozzi, P. ; Parinello, M. *Chem. Phys. Lett.* **1995**, 234, 50-56

[12] Pavankumar, P. N. V. ; Seetharamulu, P. ; Yao, S. ; Saxe, J. D. ; Reddy, D. G. ; Hausheer, F. H. *J. Comp. Chem.* **1999,** 20, 365-382

[13] Kozelka, J. ; Bergès, J. ; Attias, R. ; Fraitag, J. *Angew. Chem. Int. Ed.* **2000**, 39, 198-201

[14] Bergès, J. ; Caillet, J. ; Langlet, J. ; Kozelka, J. *Chem. Phys. Lett.* **2001**, 344, 573-577

[15] Lopes, J. F. ; Rocha, W. R. ; Dos Santos, H. F. ; De Almeida, W. B. *J. Chem. Phys.* **2008**, 128, 165103

[16] Robertazzi, A. ; Platts, J. A. *J. Comp. Chem.* **2004**, 25, 1060-1067

[17] Zhang, Y. ; Guo, Z. ; You, X.-Z. *J. Am. Chem. Soc.* **2001**, 123, 9378-9387

[18] Lau, J. K.-C. ; Deubel, D. V. *J. Chem. Theory Comput.* **2006**, 2, 103-106

[19] Raber, J. ; Zhu, C. ; Eriksson, L. A. *Mol. Phys.* **2004**, 102, 2537-2544

[20] Burda, J. V. ; Zeizinger, M. ; Leszczynski, *J. Comput. Chem.* **2005**, 26, 907-914

[21] Basch, H. ; Krauss, M. ; Stevens, W. J. ; Cohen, D. *Inorg. Chem.* **1986**, 25, 684-688

[22] Baik, M.-H. ; Friesner, R. A. ; Lippard, S. J. *J. Am. Chem. Soc.* **2003**, 125, 14082-14092

[23] Carloni, P. ; Sprik, M. ; Andreoni, W. *J. Phys. Chem. B* **2000**, 104, 823-835

[24] Burda, J. V. ; Leszczynski, J. *Inorg. Chem.* **2003**, 42, 7162-7172

[25] Robertazzi, A. ; Platts, J. A. *Inorg. Chem.* **2005**, 44, 267-274

[26] Mantri, Y. ; Lippard, S. J. ; Baik, M.-H. *J. Am. Chem. Soc.* **2007**, 129, 5023-5030

[27] Spiegel, K. ; Rothlisberger, U. ; Carloni, P. *J. Phys. Chem. B* **2004**, 108, 2699-2707

[28] Robertazzi, A. ; Platts, J. A. *Chem. Eur. J.* **2006**, 12, 5747-5756

[29] Matsui, T. ; Shigeta, Y. ; Hirao, K. *J. Phys. Chem. B* **2007**, 111, 1176-1181

[30] Tornaghi, E. ; Andreoni, W. ; Carloni, P. ; Hutter, J. ; Parinello, M. *Chem. Phys. Lett.* **1995**, 246, 469-474

[31] Wysokiński, R. ; Michalska, D. *J. Comput. Chem.* **2001**, 22, 901-912

[32] Wysokiński, R. ; Kuduk-Jaworska, J. ; Michalska, D. *J. Mol. Struct. Theochem* **2006**, 758, 169-179

[33] Pavelka, M. ; Lucas, M. F. A. ; Russo, N. *Chem. Eur. J.* **2007**, 13, 10108-10116

[34] Sarmah, P. ; Deka, R. C. *Int. J. Quant. Chem.* **2008**, 108, 1400-1409

[35] Zhu, C. ; Raber, J. ; Eriksson, L. A. *J. Phys. Chem. B* **2005**, 109, 12195-12205

[36] Wysokiński, R. ; Hernik, K. ; Szostak, R. ; Michalska, D. *Chem. Phys.* **2007**, 333, 37-48

[37] Dos Santos, H. F. ; Marcial, B. L. ; De Miranda, C. F. ; Costa, L. A. S. ; De Almeida, W. B. *J. Inorg. Biochem.* **2006**, 100, 1594-1605

[38] Deubel, D. V. *J. Am. Chem. Soc.* **2006**, 128, 1654-1663

[39] Magistrato, A. ; Ruggerone, P. ; Spiegel, K. ; Carloni, P. ; Reedijk, J. *J. Phys. Chem. B.* **2006**, 110, 3604-3613

[40] Spiegel, K. ; Magistrato, A. ; Maurer, P. ; Ruggerone, P. ; Rothlisberger, U. ; Carloni, P. ; Reedijk, J. ; Klein, M. L. *J. Comput. Chem.* **2008**, 29, 38-49

[41] Spiegel, K. ; Magistrato, A. *Org. Biomol. Chem.* **2006**, 4, 2507-2517

[42] De Hoog, P. ; Boldron, C. ; Gamez, P. ; Sliedregt-Bol, K. ; Roland, I. ; Pitié, M. ;Kiss, R. ; Meunier, B. ; Reedijk, J. *J. Med. Chem.* **2007**, 50, 3148-3152

[43] De Hoog, P. ; Louwerse, M. J. ; Gamez, P. ; Pitié, M. ; Baerends, E. J. ; Meunier, B. ; Reedijk, J. *Eur. J. Inorg. Chem.* **2008**, 4, 612-619

[44] Robertazzi, A. ; Magistrato, A. ; De Hoog, P. ; Carloni, P. ; Reedijk, J. *Inorg. Chem.* **2007**, 46, 5873-5881

[45] Robertazzi, A. (unpublished data)

[46] Robertazzi, A. ; Platts, J. A. *J. Biol. Inorg. Chem.* **2005**, 10, 854-866

[47] Chen, J. ; Chen. L. ; Liao, S. ; Zheng, K. ; Ji, L. *J. Phys. Chem. B* **2007**, 111, 7862-7869

[48] Bešker, N. ; Coletti, C. ; Marrone, A. ; Re, N. *J. Phys. Chem. B Lett.* **2008**, 112, 3871-3875

[49] Vargiu, A.V. ; Robertazzi, A. ; Magistrato, A. ; Ruggerone, P. ; Carloni, P. *J. Phys. Chem. B* **2008**, 112, 4401-4409

[50] Chiorescu, I. ; Deubel, D. V. ; Arion, V. B. ; Keppler, B. K. *J. Chem. Theory Comput.* **2008**, 4, 499-506

[51] Morris, R. E. ; Aird, R. E. ; Murdoch, P. D. ; Chen, H. M. ; Cummings, J. ; Hughes, N. D. ; Parsons, S. ; Parkin, A. ; Boyd, G. ; Jodrell, D. I. ; Sadler, P. J. *J. Med. Chem.* **2001**, *44*, 3616–3621

[52] Peacock, A. F. A. ; Melchart, M. ; Deeth, R. J. ; Habtemariam, A. ; Parsons, S. ; Sadler, P. J. *Chem. Eur. J.* **2007**, 13, 2601-2613

[53] Bešker, N. ; Coletti, C. ; Marrone, A. ; Re, N. *J. Phys. Chem. B* **2007**, 111, 9955-9964

[54] Gossens, C. ; Tavernelli, I. ; Rothlisberger, U. *J. Chem. Theory Comput.* **2007**, 3, 1212-1222

[55] Dorcier, A. ; Dyson, P. J. ; Gossens, C. ; Rothlisberger, U. ; Scopelliti, R. ; Tavernelli, I. *Organometallics* **2005**, 24, 2114-2123

[56] Deubel, D. V. *J. Am. Chem. Soc.* **2008**, 130, 665-675

[57] Burda, J. V. ; Gu, J. *J. Inorg. Biochem.* **2008**, 102, 53-62

[58] Šponer, J. E. ; Leszczynski, J. ; Šponer, J. *J. Phys. Chem. B* **2006**, 110, 19632-19636

[59] Jeremy M. Berg, John L. Tymoczko, Lubert Stryer - "Biochemistry", 6th edition, W.H. Freeman and Company (2007)

[60] Lizabeth A. Allison - "Fundamental Molecular Biology", Blackwell Publishing (2007)

[61] Graham L. Patrick - "An Introduction to Medicinal Chemistry", 3rd edition, Oxford (2005)

[62] Matta, C. F. ; Castillo, N. ; Boyd, R. J. *J. Phys. Chem. B* **2006**, 110, 563-578

[63] Dunning, T. H. J. *Chem. Phys.* **1989**, 90, 1007

[64] Boys, S. F. ; Bernardi, F. *Mol. Phys.* **1970**, 19, 553

[65] Jurečka, P. ; Šponer, J. ; Černý, J. ; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, 8, 1985-1993

[66] Pulay, P. *Chem. Phys. Lett.* **1983**, 100,151

[67] Saebø, S. ; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, 44, 213

[68] Hampel, C. ; Werner, H. J. *J. Chem. Phys.* **1996**, 104, 6286

[69] Schütz, M. ; Hetzer, G. ; Werner, H. J. *J. Chem. Phys.* **1999**, 111, 5691

[70] Hetzer, G. ; Schütz, M. ; Stoll, H. ; Werner, H. J. *J. Chem. Phys.* **2000**, 113, 9443

[71] Vahtras, O. ; Almlöf, J. ; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, 213, 514

[72] Hättig, C. ; Weigend, F. *J. Chem. Phys.* **2000**, 113, 5154

[73] Schütz, M. ; Manby, F. R. *Phys. Chem. Chem. Phys.* 2003, 5, 3349

[74] Werner, H. J. ; Manby, F. R. ; Knowles, P. J. *J. Chem. Phys.* **2003**, 118, 8149-8160

[75] Hill, J. G. ; Platts, J. A. ; Werner, H. J. *Phys. Chem. Chem. Phys.* **2006**, 8,4072

[76] Grimme, S. *J. Chem. Phys.* **2003**, 118, 9095

[77] Piacenza, M. ; Grimme, S. *Chem. Phys. Chem.*, **2005**, 6, 1554

[78] Piacenza, M. ; Grimme, S. *J. Am. Chem. Soc.* **2005**, 127, 14841

[79] Hill, J. G. ; Platts, J. A. *J. Chem. Theory Comput.* 2007, 3, 80

[80] Distasio JR, R. A. ; Head-Gordon, M. *Mol. Phys.* **2007**, 105, 1073 - 1083.

[81] Černý, J. ; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, 7, 1624-1626

[82] Hesselman, A. ; Jansen, G. *Chem. Phys. Lett.* **2003**, 367, 778-784

[83] Dion, M. ; Rydberg, H. ; Schröder, E. ; Langreth, D. C. ; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, 92, 246401

[84] Thonhauser, T. ; Cooper, V. R. ; Li, S. ; Puzder, A. ; Hyldgaard, P. ; Langreth, D. C. *Phys. Rev. B*, **2007**, 76, 125112

[85] Schwabe, T. ; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, 8, 4398

[86] Zhao, Y. ; Schultz, N. E. ; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, 2, 364-382

[87] Zhao, Y. ; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, 120, 215

[88] Zhao, Y. ; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, 7, 2701

[89] Keal, T. W. ; Tozer, D. J. *J. Chem. Phys.* **2004**, 121, 5654-5660

[90] Becke, A. D. ; Johnson, E. R. *J. Chem. Phys.* **2005**, 123, 154101

[91] Grimme, S. *J. Comput. Chem.* **2004**, 25, 1463-1473

[92] Schwabe, T. ; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, 9, 3397-3406

[93] Jurečka, P. ; Černý, J. ; Hobza, P. ; Salahub, D. R. *J. Comput. Chem.* **2007**, 28, 555

[94] Zheng J. J. ; Zhao Y. ; Truhlar D. G. *J. Chem. Theory. Comput.* **2007**, 3, 569

[95] Swart M. ; van der Wijst, T. ; Fonseca Guerra C. ; Bickelhaupt, F. M. *J. Mol. Model.* **2007**, 13, 1245

[96] Becke A. D. *J. Chem. Phys.* **1993**, 98, 1372

[97] Waller M. P. ; Robertazzi, A. ; Platts, J. A. ; Hibbs, D. E. ; Williams, P. A. *J. Comput. Chem.* **2006**, 27,491

[98] Robertazzi, A. ; Platts, J. A. *J. Phys. Chem. A* **2006**, *110*, 3992–4000

[99] Fonseca Guerra C. ; Bickelhaupt, F. M. ; Snijders, J. G. ; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, 122, 4117

[100] van der Wijst T. ; Fonseca Guerra C. ; Swart, M. ; Bickelhaupt, F. M. *Chem. Phys. Lett.* 426,415

[101] Platts, J. A. ; Hibbs, D. E.; Hambley, T. W. ; Hall, M. D. *J. Med. Chem.* **2001**, 44, 472-474

[102] Platts, J. A. ; Oldfiled, S. P. ; Reif, M. M. ; Palmucci, A. ; Gabano, E. ; Osella, D. *J. Inorg. Biochem.* **2006**, 100, 1199-1207

[103] Tetko, I. V. ; Jaroszewicz, I. ; Platts, J. A. ; Kuduk-Jaworska, J. *J. Inorg. Biochem.* **2008**, 102, 1424-1437

[104] Oldfield, S. P. ; Hall, M. D. ; Platts, J. A. *J. Med. Chem.* **2007**, 50, 5227-5237

[105] Dobrogorskaia-Méreau, I. A. I. ; Nemukhin, A.V. *J. Comput. Chem.* **2005**, 26, 865-870

[106] Zimmermann, T. ; Zeizinger, M. ; Burda, J. V. *J. Inorg. Biochem.* **2005**, 99, 2184-2196

[107] Da Silva, V. J. ; Costa, L. A. S. ; Dos Santos, H. F. *Int. J. Quant. Chem.* **2008**, 108, 401-414

[108] Lau, J. K.-L. ; Deubel, D. V. *Chem. Eur. J.* **2005**, 11, 2849-2855

[109] F. Jensen, "Introduction to Computational Chemistry", 2nd edition, Wiley (2007)

[110] A. Szabo and N. S. Ostlund, "Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory", Dover (1996)

[111] A. R. Leach, "Molecular Modelling: Principles and Applications", 2nd edition, Pearson Prentice Hall (2001)

[112] P. W. Atkins and R. S. Friedman, "Molecular Quantum Mechanics", 3rd edition, Oxford University Press (1997)

[113] C. J. Cramer, "Essentials of Computational Chemistry: Theories and Models", 2nd edition, Wiley (2005)

[114] A. J. Stone, "The Theory of Intermolecular Forces, Oxford University Press (2002)

[115] R. F. Bader, "Atoms In Molecules: A Quantum Theory", Oxford University Press (2003)

[116] P. L. Popelier, Atoms in Molecules: An Introduction, Prentice Hall, 2000

[117] E. Steiner, "The Chemistry Maths Book", Oxford Science Publications (2004)

[118] V. A. Bloomfield, D. M. Crothers, I. Tinocco, JR., "Nucleic Acids: Structures, Properties, and Functions", University Science Books (2000)

[119] Andrae, A. ; Hausserman, U. ; Dolg, M. ; Stoll, H. ; Preuss, H. *Theor.Chim. Acta* **1990**, *77*, 123–141

[120] Becke, A. D. *Phys. Rev. A* **1988**, 38, 3098-

[121] Lee, C. ; Yang, W. ; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789

[122] Becke A. D. *J.Chem.Phys.* **1993**, 98, 5648-

[123] Stephens, P. J. ; Devlin, F. J. ; Chabalowski, C. F. ; Frisch, M. J. *J.Phys.Chem.* **1994**, 98, 11623

[124] Pulay, P. J. *Comp. Chem.* **1982**, 3, 556-

[125] Csaszar, P. ; Pulay, P. *J. Mol. Struct.* **1984**, 114, 31-

[126] Warshel, A. ; Levitt, M. *J.Mol.Biol.* **1976**, 103, 227-249

[127] Maseras, F. ; Morokuma, K. *J. Comput. Chem.* **1995**, 16, 1170-1179

[128] Svensson, M. ; Humbel, S. ; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. J. *Phys. Chem.* **1996**, 100, 19357-19363

[129] Dapprich, S. ; Komáromi, I. ; Byun, K. S. ; Morokuma, K. ; Frisch, M. J. *J. Mol. Struct.* (THEOCHEM) **1999**, 461, 1-21

[130] Vreven, T. ; Byun, K. S. ; Komáromi, I. ; Dapprich, S. ; Montgomery, J. A. Jr.; Morokuma, K. ; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, 2, 815-826

[131] Frisch, M. J. *et al. Gaussian03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004

[132] Cornell, W. D. *et al. J. Am. Chem. Soc.* **1995**, 117, 5179-5197

[133] Rappé, A. K. ; Goddard III, W. A. *J. Phys. Chem.* **1991**, 95, 3358-63

[134] Koch, U. ; Popelier, P. L. A. *J. Phys. Chem.* **1995**, 99, 9747-9754

[135] Bader, R. W. F. *Pure Appl. Chem.* **1988**, *60*, 145–155.

[136] Biegler-König, J. ; Schönbohm, J. *J. Comput. Chem.* **2002**, *23*, 1489–1494

[137] Bieger-Konig, F. W. ; Bader, R. F. ; Tang, T.-H. *J. Comp. Chem.* **3** (1982), p. 317

[138] Miertus, S. ; Scrocc, E. ; Tomasi, J. *Chem. Phys.* **1981**, 55, 117

[139] Miertus, S. ; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239–245

[140] Cossi, M. ; Barone, V. ; Cammi, R. ; Tomasi, J. *Chem. Phys. Lett.* **1996**,*255*, 327–335

[141] http://mathworld.wolfram.com/LeastSquaresFitting.html (last accessed: 04/09/09)

[142] JMP, Revision 5, SAS Institute, Inc., Cary, NC (USA), 2002.

[143] Dickerson, R. E. *Nuc. Ac. Res.* **1989**, 17, 1797-1803

[144] Lavery, R. and Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, 6, 63-91

[145] Lavery, R. and Sklenar, H., *J. Biomol. Struct. Dyn.* **1989**, 6, 655-67

[146] Marchetti, O. ; Werner, H.-J. *Phys. Chem. Chem. Phys.* **2008**, 10, 3400-3409

[147] Šponer, J. ; Riley, K. E. ; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, 10, 2595.

[148] Boyd, R. J. ; Choi, S. J. *Chem. Phys. Lett.* **1985**, 120, 80.

[149] Carroll, M. T.; Bader, R. F. W. *Mol. Phys.* **1988**, 65, 695.

[150] DF-LMP2/aug-cc-pVTZ optimisation and frequency calculations, commencing from the literature geometry, were performed for smaller complexes. Analogous calculations for the remaining complexes are in progress, and will reported: Hill, J.G.; Platts, J.A. *in preparation.*

[151] Podeszwa, R. ; Bukowski, R. ; Szalewicz, K. *J. Phys. Chem. A* **2006**, 110, 10345.

[152] Wang, W. ; Pitoňák, M. ; Hobza, P. *Chem. Phys. Chem.* **2007**, 8, 2107.

[153] Rogers, D. M. ; Hirst, J. D. ; Lee, E. P. F. ; Wright, T. G. *Chem. Phys. Lett.* **2006**, 427, 410.

[154] Grimme, S. *Angew. Chem. Int. Ed. Engl.* **2006**, 45, 4460.

[155] Palusiak, M. ; Rudolf, B. ; Zakrzewski, J. ; Pfitzner, A. ; Zabel, M. ; Grabowski, S. J. *J. Organometallic Chem.* **2006**, 691, 3232-3238

[156] http://amber.scripps.edu/

[157] Orbell, J. D. ; Marzilli, L. G. ; Kistenmacher, T. J. *J. Am. Chem. Soc.* **1981**, 103, 5126-5133

[158] Scheeff, E. D. ; Briggs, J. M. ; Howell, S. B. *Mol. Pharm.* **1999**, 56, 633-643

[159] Vreven, T. ; Morokuma, K. ; Farkas, Ö. ; Schlegel, H. B. ; Frisch, M. J. *J. Comp. Chem.* **2003**, 24, 760-769

[160] Yang, D. ; van Boom, S. S. G. E. ; Reedijk, J. ; van Boom, J. H. ; Wang, A. H. J. *Biochem.* **1995**, 34, 12912-12920

[161] Gaussian03 manual pages: http://www.gaussian.com/g_dl/g03_man.zip

[162] Molecular graphics image was produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).

[163] Pettersen, E. F. ; Goddard, T. D. ; Huang, C. C. ; Couch, G. S. ; Greenblatt, D. M. ; Meng, E. C. ; Ferrin, T. E. *J Comput Chem.* **2004**, 13, 1605-1612.

[164] (a) Doreleijers, J. F. ; Mading, S. ; Maziuk, D. ; Sojourner, K. ; Yin, L. ; Zhu, J. ; Markley, J. L. ; Ulrich, E. L. *J. Biomol. NMR* **2003,** 26, 139-146. (b) Doreleijers, J. F. ; Nederveen, A. J. ; Vranken, W. ; Lin, J. ; Bonvin, A. M. ; Kaptein, R. ; Markley, J. L. ; Ulrich, E. L. BioMagResBank databases DOCR and FRED with converted and filtered sets of experimental NMR restraints and coordinates from over 500 protein PDB structures. *J. Biomol. NMR* **2005,** 32, 1-12. (c) Nederveen, A. J. ; Doreleijers, J. F. ; Vranken, W. ; Miller, Z. ; Spronk, C. A. ; Nabuurs, S. B. ; Guntert, P. ; Livny, M. ; Markley, J. L. ; Nilges, M. ; Ulrich, E. L. ; Kaptein, R. ; Bonvin, A. M. (2005). RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **2005**, 59, 662-672. (d) Nabuurs, S. B. ; Nederveen, A. J. ; Vranken, W. ; Doreleijers, J. F. ; Bonvin, A. M. ; Vuister, G. W. ; Vriend, G. ; Spronk, C. A. DRESS: a database of REfined solution NMR structures. *Proteins* **2004**, 55, 483-486.

[165] Wu, Y. ; Bhattacharyya, D. ; King, C. L. ; Baskerville-Abraham, I. ; Huh, S.-H. ; Boysen, G. ; Swenberg, J. A. ; Temple, B. ; Campbell, S. L. ; Chaney, S. G. *Biochem.* **2007**, 46, 6477-6487

[166] Chen, H. ; Parkinson, J. A. ; Parsons, S. ; Coxall, R. A. ; Gould, R. O. ; Sadler, P. J. *J. Am. Chem. Soc.* **2002**, *124*, 3064–3082.

[167] Chen, H. ; Parkinson, J. A. ; Morris, R. E. ; Sadler, P. J. *J. Am. Chem. Soc.* **2003**, *125*, 173–186.

[168] Novakova, O. ; Chen, H. ; Vrana, O. ; Rodger, A. ; Sadler, P. J. ; Brabec, V. *Biochemistry* **2003**, *42*, 11544–11554.

[169] Novakova, O. ; Kasparkova, J. ; Bursova, V. ; Hofr, C. ; Vojtiskova, M. ; Chen, H. ; Sadler, P. J. ; Brabec, V. *Chem. Biol.* **2005**, *12*, 121–129.

[170] Liu, H.-K. ; Wang, F. ; Parkinson, J. A. ; Bella, J. ; Sadler, P. J. *Chem.sEur. J.* **2006**, *12*, 6151–6165.

[171] Rappé, A. K. ; Casewit, C. J. ; Colwell, K. S. ; Goddard, W. A. III ; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.

[172] Chen, E. C. M. ; Herder, C. ; Chen, E. S. *J. Mol. Struct.* **2006**, *798*, 126–133.

[173] Sherman, S. E. ; Lippard, S. J. *Chem. Re*V. **1987**, *87*, 1153–1181.

[174] Vlieghe, D. J. ; van Meervelt, L. *Biochemistry* **1999**, *38*, 16443–16451.

[175] Peacock, A. F. A. ; Parsons, S. ; Sadler, P. J. *J.Am.Chem. Soc.* , **2007**, 129, 3348-3357

[176] Hovorun, D. M. ; Gorb, L. ; Leszczynski, J. *Int. J. Quant. Chem.* **1999**, *75*, 245–253.

[177] Šponer, J. ; Hobza, P. *Int. J. Quant. Chem.* **1996**, *57*, 959–970.

[178] Šponer, J. ; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161–3164.

[179] Bludský, O. ; Šponer, J. ; Leszczynski, J. ; Špirko, V. ; Hobza, P. *J. Chem. Phys.* **1996**, *105*, 11042–11050.

[180] Luisi, B. ; Orozco, M. J. ; Luque, F. J. ; Shakked, Z. *J. Mol. Biol.* **1998**, *279*, 1123–1136.

[181] Rěha, D. ; Kabeláč, M. ; Ryjáček, F. ; Šponer, J. ; Šponer, J. E. ; Elstner, M. ; Suhai, S. ; Hobza, P. *J. Am. Chem. Soc.* **2002**, *124*, 3366–3766.

[182] Hobza, P. ; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802–11808.

[183] Jurecka, P. ; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.

[184] Šponer, J. ; Gabb, H. A. ; Leszczynski, J. ; Hobza, P. *Biophys. J.* **1997**, *73*, 76–87.

[185] Leininger, M. L. ; Nielsen, I. M. B. ; Colvin, M. E. ; Jannsen, C. L. *J. Phys. Chem. A* **2002**, *106*, 3850–3854.

[186] Yanson, I. K. ; Teplitsky, A. B. ; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149–1170.

[187] Barratta, W. ; Mealli, C. ; Herdtweck, E. ; Ienco, A. ; Mason, S. A. ; Rigo, P. *J. Am. Chem. Soc.* **2004**, *126*, 5549–5562, and references cited therein

[188] F. Marozzelli, MSc dissertation, Cardiff University, Cardiff, UK, **2009**.

[189] Cleare, M. J. ; Hoeschele, J. D. *Bioinorg. Chem.* **1973**, 2, 187–210

[190] Cleare, M. J. ; Hoeschele, J. D. *Platinum Met. Rev.* **1973**, 17, 2–13

[191] Basolo, F. ; Gray, H. B. ; Pearson, R. G. *J. Am. Chem. Soc.* **1960**, 82,4200– 4203

[192] Gray, H. B. ; Olcott, R. J. *Inorg. Chem.* **1962**, 1, 481–485

[193] Ushay, H. M. ; Tullius, T. D. ; Lippard, S. J. *Biochemistry* **1981**, 20, 3744–3748

[194] Johnson, N. P. ; Hoeschele, J. D. ; Rahn, R. O. *Chem. Biol. Interact.* **1980**, 30, 151–169

[195] Eastman, A. ; Barry, M. A. *Biochemistry* **1987**, 26, 3303–3307

[196] Raymond, E. ; Faivre, S. ; Woynarowski, J. M. ; Chaney, S.G. *Semin. Oncol.* **1998**, 25:4–12

[197] Monti, E. ; Gariboldi, M. ; Maiocchi, A. ; Marengo, E. ; Cassino, C. ; Gabano, E. ; Osella, D. *J. Med. Chem.* **2005**, 48, 857–866

[198] Alink, M. ; Nakahara, H. ; Hirano, T. ; Inagaki, K. ; Nakanishi, M. ; Kidani, Y. ; Reedijk, J. *Inorg. Chem.* **1991**, 30, 1236–1241

[199] Inagaki, K.; Nakahara, H. ; Alink, M. ; Reedijk, J. *J. Chem. Soc. Dalton Trans.* **1991**, 1337–1341

[200] Milanesio, M. ; Monti, E. ; Gariboldi, M. B. ; Gabano, E. ; Ravera, M. ; Osella, D. *Inorg. Chim. Acta* **2008**, 361, 2803–2814

[201] Sherman, S. E.; Gibson, D. ; Wang, A. H. J. ; Lippard, S. J. *Science* **1985**, 230, 412–417

[202] "The Weak Hydrogen Bond in Structural Chemistry and Biology" - Gautam R. Desiraju, Thomas Steiner, Oxford University Press (1999)

[203] Bouchard, G. ; Carrupt, P.-A. ; Testa, B. ; Gobry, V. ; Girault, H. H. *Chem. Eur. J.* **2002**, 8,3478–3484

[204] Testa, B. ; Crivori, P. ; Reist, M. ; Carrupt, P.-A. *Persp. Drug Disc Des.* **2000**, 19, 179–211

[205] Schweitzer, B. A. ; Kool E. T. *J. Org. Chem.* **1994**, 59, 7238-7242

[206] Kool, E. T. ; Sintim, H. O., *Chem. Commun.* **2006**, 3665-3675

[207] Cubero, E. ; Sherer, E. C.; Luque, F. J. ; Orozco, M. ; Laughton, A. C. *J. Am. Chem. Soc.*, **1999**, *121*, 8653–8654

[208] Guckian, K. M. ; Krugh, T. R. ; Kool, E. T. *Nat. Struct. & Mol. Biol.* **1998**, 5, 954 - 959

7. Appendix

Table 7.1: Number of BCP's found in AIM analysis.

	stack	H-bond	total		stack	H-bond	total
anthA	0	1	1	dhaA	4	2	6
anthC	1	0	1	dhaG	4	1	5
anthG	1	0	1	dhaC	4	3	7
anthI	1	0	1	dhaT	3	3	6
anthT	1	1	2	dhaI	3	1	4
bipA	1	1	2	thaA	2	3	5
bipC	1	0	1	thaG	2	3	5
bipG	1	1	2	thaC	2	5	7
bipT	1	0	1	thaT	3	6	9
bipI	1	1	2	thaI	2	3	5

Table 7.2: Full details of base-en hydrogen bonding.

	acceptor	rho (BCP) au	kcal mol^{-1}	D-A dist	D-H..A angle
anthA	N6(A)	0.0321	-4.78	2.855	152.3
anthG	O6(G)	0.0464	-7.11	2.7	162.8
anthI	O6(I)	0.0443	-6.76	2.707	161.3
benA	N6(6NH2,A)	0.0361	-5.43	2.845	154.1
benG	O6(G)	0.0481	-7.38	2.682	162.8
benI	O6(ino)	0.0443	-6.76	2.685	162.7
bipA	N6(6NH2,A)	0.0352	-5.29	2.846	152.9
bipG	O6(G)	0.0501	-7.70	2.667	164.8
bipI	O6(I)	0.0500	-7.69	2.67	164.5
dhaA	N6(6NH2,A)	0.0387	-5.86	2.824	156.5
dhaG	O6(G)	0.0459	-7.02	2.697	161.8
dhaI	O6I)	0.0459	-7.02	2.696	161.5
thaA	N6(6NH2,A)	0.0370	-5.58	2.83	153.7
thaG	O6(G)	0.0491	-7.54	2.706	162.2
thaI	O6(I)	0.0475	-7.28	2.685	162.4
benGC	O6(G)	0.0517	-7.97	2.668	164.5
thaGC	O6(G)	0.0542	-8.37	2.657	166.9
cymG	O6(G)	0.0472	-7.24	2.69	163.08
anthC	O2(C)	0.0397	-6.02	2.652	142.2
	N4 (4NH2,C)	0.0145	-1.92	3.002	21.5
anthT	O2(T)	0.0406	-6.16	2.671	144.4
	O4(T)	0.0488	-7.50	2.647	152
benC	N4(4NH2,C)	0.0153	-2.05	2.995	122.7
	O2(C)	0.0394	-5.97	2.649	140.8
benT	O4(T)	0.0454	-6.94	2.653	148.5
	O2(T)	0.0402	-6.10	2.676	144.5
bipC	N4(4NH2,C)	0.0154	-2.07	2.987	123
	O2(C)	0.0387	-5.86	2.654	140.9
bipT	O4(T)	0.0418	-6.36	2.672	145.8
	O2(T)	0.0411	-6.25	2.67	144.9
	N3(T)	0.0126	-1.61	3.061	108
dhaC	O2(V)	0.0412	-6.26	2.641	142.6
	N4(4NH2,C)	0.0156	-2.10	2.972	122
dhaT	O2(T)	0.0428	-6.52	2.655	145.3
	O4(T)	0.0421	-6.41	2.671	146.4
	N3(T)	0.0126	-1.61	3.058	107.8
thaC	O2(C)	0.0399	-6.05	2.649	141.7
	N4(4NH2,C)	0.0154	-2.07	2.983	122.5
thaT	O2(T)	0.0413	-6.28	2.666	144.9
	O4(T)	0.0424	-6.46	2.668	146.5
	N3(T)	0.0126	-1.61	3.062	108

Table 7.3: Full details of base-arene hydrogen bonding

	donor	acceptor	rho (BCP) au	kcal mol^{-1}	D-A dist	D-H..A angle
anthA	N6(A)	C13	0.0101	-1.20	3.393	148.4
anthG	-	-	-	-	-	-
anthI	-	-	-	-	-	-
benA	-	-	-	-	-	-
benG	-	-	-	-	-	-
benI	-	-	-	-	-	-
bipA	C(9CH3,A)	C m' (B)bip	0.0028	-0.02	4.018	126.3
bipG	C9(9CH3,G)	Cm'(B)bip	0.0051	-0.39	3.505	124.7
	C(9CH3,I)	C m' (B)bip	0.0051	-0.39	3.72	
bipI	N2(2NH2,A)	H9(dha)	0.0087	-0.98	3.141	170
dhaA	N7(A)	C8(dha)	0.0086	-0.96	3.304	117
	C9(dha)	O6(G)	0.0082	-0.90	3.341	123.4
dhaG	C9(dha)	O6(I)	0.0073	-0.75	3.367	122.5
dhaI	N6(NH2,A)	H9(tha)	0.0122	-1.55	2.944	160.7
thaA	C8(tha)	N6(6NH2,A)	0.0096	-1.12	3.407	125.6
	C5(tha)	N9(A)	0.0088	-0.99	3.558	141.3
	C9(tha)	O6(G)	0.0082	-0.89	3.268	116.1
thaG	C8(tha)	O6(G)	0.0101	-1.20	3.297	125.3
	C5(tha)	C4(G)	0.0076	-0.80	3.433	121.2
	C9(tha)	O6(I)	0.0088	-0.99	3.201	115.2
	C8(tha)	O6(I)	0.0098	-1.16	3.276	124.1
thaI	C5(tha)	C4(I)	0.0074	-0.77	3.481	126.4
	C2(ino)	O6(G)	0.0068	-0.66	3.635	144.98
	C3(ino)	C5(G)	0.0095	-1.12	3.48	132.86
cymG	C2(ino)	O6(G)	0.0068	-0.66	3.635	144.98
	C3(ino)	C5(G)	0.0095	-1.12	3.48	132.86
anthC	-	-	-	-	-	-
anthT	C(5CH3,T)	C5(anth)	0.0057	-0.49	3.843	152.2
benC	-	-	-	-	-	-
benT	-	-	-	-	-	-
bipC	-	-	-	-	-	-
bipT	-	-	-	-	-	-
dhaC	C9(dha)	O2(C)	0.0083	-0.91	3.159	108.5
	C1(1CH3,C)	C8(dha)	0.0074	-0.77	3.534	130.1
	C10(dha)	N4(4NH2,C)	0.0065	-0.62	3.488	116.1
dhaT	C9(dha)	O2(T)	0.0111	-1.37	3.087	113.9
	C1(1CH3,T)	C8(dha)	0.0083	-0.91	3.446	127.6
	C10(dha)	O4(C)	0.0082	-0.90	3.418	129.7
thaC	C9(tha)	O2(C)	0.0099	-1.17	3.062	107.2
	C8(tha)	O2(C)	0.0074	-0.77	3.317	115.5
	C1(1CH3,C)	C7(tha)	0.0097	-1.14	3.461	141.6
	C10(tha)	N4(4NH2,A)	0.0074	-0.77	3.55	128.2
	C5(tha)	C5(C)	0.0061	-0.55	3.777	135.5
thaT	C9(tha)	O2(T)	0.0136	-1.78	3.047	119.3
	C8(tha)	O2(T)	0.0051	-0.39	3.602	121.1
	C8(tha)	C1(1CH3,T)	0.0080	-0.87	3.438	125

C1(1CH3,T)	C7(tha)	0.0078	-0.83	3.481	130.1
C5(tha)	C5(T)	0.0090	-1.03	3.448	127
C10(tha)	O4(T)	0.0094	-1.09	3.379	132.4
C1(1CH3,T)	O2(T)	0.0216	-3.08	2.694	105.2

Table 7.4: Full details of base-arene stacking.

	atom	atom	rho (BCP) au	kcal mol^{-1}	D-A dist
anthA	-	-	-	-	-
anthG	O6(G)	C13(anth)	0.0057	1.00	3.27
anthI	O6(I)	C13(anth)	0.0066	1.16	3.2
benA	-	-	-	-	-
benG	-	-	-	-	-
benI	-	-	-	-	-
bipA	C8(A)	C o' (B)bip	0.0099	1.73	3.102
bipG	C o'(B)bip	C8(G)	0.0104	1.83	3.061
bipI	C8(I)	C o' (B)bip	0.0103	1.80	3.095
dhaA	C8(dha)	C6(A)	0.0077	1.35	3.232
	N7(A)	C14(dha)	0.0098	1.71	3.104
	C4(A)	C5(dha)	0.0069	1.21	3.378
	N9(A)	C5(dha)	0.0071	1.25	3.305
dhaG	C5(G)	C13(dha)	0.0090	1.58	3.225
	N3(G)	C7(dha)	0.0062	1.09	3.444
	N9(G)	C5(dha)	0.0075	1.31	3.281
	N7(G)	C14(dha)	0.0087	1.53	3.186
dhaI	C13(dha)	C5(I)	0.0089	1.56	3.221
	C14(dha)	N7(I)	0.0088	1.54	3.183
	C5(dha)	N9(I)	0.0074	1.30	3.274
thaA	C7(tha)	N1(A)	0.0049	0.87	3.522
	C14(tha)	N7(A)	0.0085	1.49	3.202
thaG	C7(tha)	N1(G)	0.0079	1.38	3.27
	C14(tha)	C5(G)	0.0090	1.58	3.269
thaI	C7(tha)	N1(I)	0.0073	1.28	3.312
	C14(tha)	C5(I)	0.0091	1.59	3.24
cymG	-	-	-	-	-
anthC	O2(cyt)	C10(anth)	0.0057	1.01	3.377
anthT	O2(T)	C14(anth)	0.0076	1.33	3.183
benC	-	-	-	-	-
benT	-	-	-	-	-
bipC	N3(C)	C o' (B)bip	0.0085	1.48	3.232
bipT	C o'(B)bip	C2(T)	0.0095	1.67	3.029
dhaC	N1(C)	C8(dha)	0.0058	1.02	3.45
	C2(C)	C13(dha)	0.0080	1.40	3.49
	N3(C)	C14(dha)	0.0086	1.50	3.432
	C5(G)	C5(dha)	0.0053	0.93	3.577
dhaT	C2(T)	C13(dha)	0.0083	1.45	3.091
	N3(T)	C14(dha)	0.0086	1.51	3.228

	C5(T)	C5(dha)	0.0069	1.21	3.43
thaC	C13(tha)	O2(C)	0.0094	1.64	3.048
	C14(tha)	N3(C)	0.0089	1.55	3.247
thaT	C7(tha)	N1(T)	0.0052	0.92	3.513
	C6(tha)	C6(T)	0.0054	0.96	3.448
	C14(tha)	N3(T)	0.0092	1.61	3.256

Figure 7.1: Plots of Ecov *vs.* ρ(Ru—N)

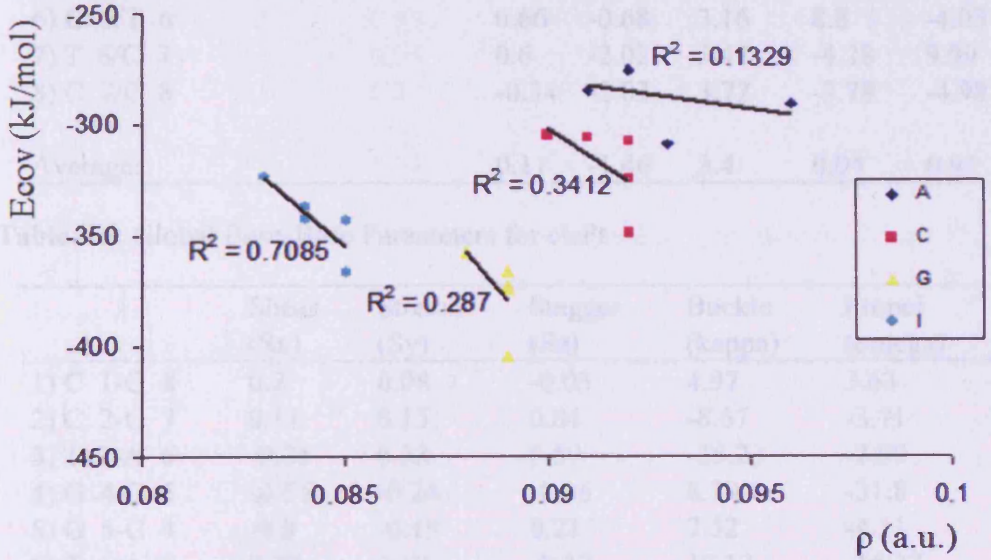


Figure 7.2: Plots of Ecov *vs.* E_{LUMO}

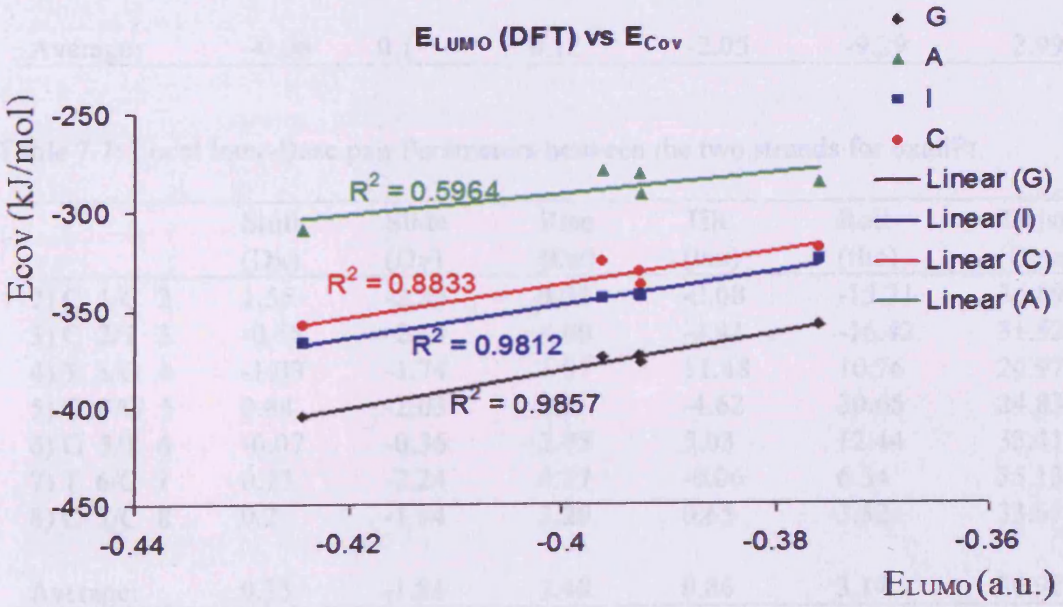


Table 7.5: Local Inter-Base pair Parameters between the two strands for cisPt

	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (tau)	Roll (rho)	Twist (Omega)
2) C 1/C 2	-0.06	-2.11	3.76	-0.5	-8.93	35.43
3) C 2/T 3	0.34	-1.78	3.77	3.16	-11.99	33.7
4) T 3/G 4	-1.24	0.35	2.41	1.66	-1.73	33.66
5) G 4/G 5	0.83	-1.91	3.55	-4.8	28.21	28.57
6) G 5/T 6	0.66	-0.68	3.16	8.8	-4.03	44.15
7) T 6/C 7	0.6	-2.02	3.47	-4.28	9.99	28.04
8) C 7/C 8	-0.34	-2.03	3.72	-3.78	-4.98	40.37
Average:	0.11	-1.46	3.4	0.04	0.93	34.85

Table 7.6: Global Base-Base Parameters for cisPt

	Shear (Sx)	Stretch (Sy)	Stagger (Sz)	Buckle (kappa)	Propel (omega)	Opening (sigma)
1) C 1-G 8	0.2	0.08	-0.03	4.97	3.63	0.41
2) C 2-G 7	0.11	0.15	0.04	-8.67	-3.71	4.12
3) T 3-A 6	-0.25	0.38	0.59	-25.23	-2.99	5.35
4) G 4-C 5	-0.51	-0.24	-0.06	8.19	-31.8	-1.86
5) G 5-C 4	-0.8	-0.15	0.21	7.52	-4.11	0.39
6) T 6-A 3	0.28	0.12	-0.37	12.13	-14.37	6.7
7) C 7-G 2	0.3	0.17	0.07	0.89	-16.21	5.54
8) C 8-G 1	0.2	0.32	0.54	-16.2	-4.76	3.3
Average:	-0.06	0.1	0.12	-2.05	-9.29	2.99

Table 7.7: Local Inter-Base pair Parameters between the two strands for oxaliPt.

	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (tau)	Roll (rho)	Twist (Omega)
2) C 1/C 2	1.55	-2.35	4.51	-2.08	-15.31	36.89
3) C 2/T 3	-0.58	-2.19	4.00	-4.41	-16.42	31.52
4) T 3/G 4	-1.03	-1.74	1.97	11.48	10.76	20.97
5) G 4/G 5	0.84	-2.03	3.6	-4.62	20.65	24.83
6) G 5/T 6	-0.07	-0.36	2.75	5.03	12.44	33.41
7) T 6/C 7	0.73	-2.24	4.27	-0.06	6.34	35.13
8) C 7/C 8	0.2	-1.94	3.29	0.65	3.52	33.67
Average:	0.23	-1.84	3.48	0.86	3.14	30.92

Table 7.8: Global Base-Base Parameters for oxaliPt.

	Shear (Sx)	Stretch (Sy)	Stagger (Sz)	Buckle (kappa)	Propel (omega)	Opening (sigma)
1) C 1-G 8	0.22	0.21	-0.16	23.31	-18.61	2.52
2) C 2-G 7	0.11	0.23	-0.02	-8.11	1.22	4.76
3) T 3-A 6	-0.19	0.81	1.31	-24.36	-8.35	12.66
4) G 4-C 5	-0.39	-0.18	-0.22	12.77	-16.96	-0.89
5) G 5-C 4	-0.69	-0.12	0.02	1.5	-5.47	1.06
6) T 6-A 3	0.42	0.35	-0.65	20.87	-17.99	6.45
7) C 7-G 2	0.42	0.25	-0.11	-18.04	-13.07	5.46
8) C 8-G 1	0.1	0.4	0.89	-25.05	-12.85	5.37
Average:	0	0.24	0.13	-2.14	-11.51	4.67

Table 7.9: Local Inter-Base pair Parameters between the two strands for lobaPt.

	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (tau)	Roll (rho)	Twist (Omega)
2) C 1/C 2	1.53	-2.4	4.43	-2.32	-13.89	35.85
3) C 2/T 3	-0.86	-2.27	4.06	-4.94	-16.61	31.78
4) T 3/G 4	-0.97	-1.72	2.09	10.47	6.62	24.43
5) G 4/G 5	0.79	-1.65	3.27	-3.47	15.78	26.87
6) G 5/T 6	0.26	-0.69	2.98	5.54	5.63	42.27
7) T 6/C 7	1.05	-2.19	3.7	-1.76	8.92	29.55
8) C 7/C 8	-0.8	-1.96	3.67	-2.25	-2.04	37.62
Average:	0.14	-1.84	3.46	0.18	0.63	32.62

Table 7.10: Global Base-Base Parameters for lobaPt.

	Shear (Sx)	Stretch (Sy)	Stagger (Sz)	Buckle (kappa)	Propel (omega)	Opening (sigma)
1) C 1-G 8	0.24	0.17	-0.19	23.57	-16.36	2.43
2) C 2-G 7	0.11	0.2	-0.04	-6.24	-0.28	5.45
3) T 3-A 6	-0.26	0.69	1.31	-24.48	-11.47	9.79
4) G 4-C 5	-0.36	-0.22	-0.01	8.14	-21.07	-2.09
5) G 5-C 4	-0.43	-0.12	0.21	5.32	-3.95	0
6) T 6-A 3	0.25	0.11	-0.52	13.7	-10.36	2.4
7) C 7-G 2	0.19	0.28	0.13	-6.5	-19.62	6.27
8) C 8-G 1	0.23	0.53	0.71	-23.26	-9.46	7.41
Average:	0	0.2	0.2	-1.22	-11.57	3.96

Table 7.11: Local Inter-Base pair Parameters between the two strands for heptaPt.

	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (tau)	Roll (rho)	Twist (Omega)
2) C 1/C 2	0.58	-3.36	4.81	-6.31	-12.06	27.58
3) C 2/T 3	-1.31	-2.24	3.83	-0.49	-29.1	34.54
4) T 3/G 4	-1.19	-1.29	2.22	10.96	0.91	30.19
5) G 4/G 5	0.81	-1.54	2.99	-4.29	17.34	25.12
6) G 5/T 6	0.35	-0.44	2.76	3.31	2.19	29.37
7) T 6/C 7	1.05	-1.07	3.66	2.94	-11.6	42.72
8) C 7/C 8	-0.55	-1.78	4.32	-7.27	-4.69	43.76
Average:	-0.04	-1.67	3.51	-0.16	-5.29	33.32

Table 7.12: Global Base-Base Parameters for heptaPt.

	Shear (Sx)	Stretch (Sy)	Stagger (Sz)	Buckle (kappa)	Propel (omega)	Opening (sigma)
1) C 1-G 8	0.52	-0.19	-0.51	18.09	2.56	4.21
2) C 2-G 7	-0.06	0.36	0.2	-22.68	19.12	11.76
3) T 3-A 6	-0.09	0.77	1.43	-23.42	-12	15.07
4) G 4-C 5	-0.46	-0.17	0.21	4.19	-27.93	-3.27
5) G 5-C 4	-0.61	-0.08	0.6	9.33	-8.62	-2.41
6) T 6-A 3	-0.33	0.41	0.32	27.44	-24.99	9.41
7) C 7-G 2	0.13	-0.03	-0.35	15.22	-23.07	3.65
8) C 8-G 1	-0.07	0.45	0.72	-21.76	-13.88	5.75
Average:	-0.12	0.19	0.33	0.8	-11.1	5.52