School of Chemistry

Cardiff University

Theoretical Studies of Interactions of Transition Metal Anticancer Complexes with DNA

Shaun T. Mutter

A thesis submitted to Cardiff University

in accordance with the requirements for the degree of

Doctor of Philosophy

April 2013

Acknowledgements

Firstly I would like to thank my Mum, Glenys, and Dad, Tim, for their constant support throughout all my years of study, for without them I wouldn't be where I am today. Also my other family members especially my brother Jason, for their support, belief, and general interest in what I do. My best and oldest friend Dan, not only for the interesting medical based conversations we have had but also for the welcome distractions and advice. My close friends from Cardiff University Alex, Tim, Flo, Cathy, and Emma, for all the good times we've had in the great outdoors and the long nights inside, and for continuous constructive conversations. A huge thank you to all the teaching staff of the Theoretical Chemistry Group at Cardiff University for their advice and guidance. All the past and present members of the lab, for all the enjoyable time I've spent working there and for all the bowling. Particularly Adam, Chris, Filippo, Nimah, Scott, and most importantly Kostas, as most of the work carried out here wouldn't have been possible without the frameworks he has set out. A mention also to the members of the inorganic department at Cardiff University where I made my 'second home' and spent many lunch times involved in interesting discussions and debates. Thanks should go to my collaborators overseas, Dr Nico Margiotta at Università degli Studi di Bari and Dr Paride Papadia at Università del Salento, whose experimental work has set the basis for much of the work presented here. An enormous thank you to Kate, who has been there always throughout the last seven years of my study and has constantly been willing to give me support and assistance, without her belief and endorsement none of this would have come to fruition. Finally and most importantly, I would like to thank my supervisor Dr Jamie Platts for giving me the opportunity to carry out this project, as well as for always giving expert guidance and support in all the areas involved, for without this none of the work here would be possible.

Abstract

Density functional theory (DFT) and combined quantum mechanics/molecular mechanics (QM/MM) calculations have been used to model inter- and intra-molecular non-covalent interactions of transition metal complexes and where applicable their interactions with DNA. Two DFT functionals, BHandH and B97-D, which have shown to be efficient in modelling systems containing non-covalent interactions, have been tested against high level *ab initio* calculations on test transition metal complexes, designed to represent the intermolecular interactions present in the benzene dimer and methane benzene systems. The DFT functionals above show good agreement with the benchmark calculations and have been used to study ruthenium arene 'piano stool' type complexes, of the general form $[\eta^6(\text{arene})\text{Ru(en)Cl}]^+$, which have shown potential as anticancer agents. The intramolecular interactions of these ruthenium complexes through coordination to guanine and adenine through the N7 nitrogen, has been explored using a selection of pure DFT, hybrid DFT, and post Hartree-Fock approaches against benchmark correlated wavefunction methods, where the best methods were found to be BHandH, B97-D2, and MP2(0.25). The B97-D2 functional was used to model these ruthenium complexes, with a selection of extended aromatic ligands with potential to act as intercalators, interacting with base pair steps. Calculated binding energies show a sensitivity to the nature of the arenes, where the more flexible ligands form more non-covalent interactions with DNA, as demonstrated by QTAIM analysis. Conformations and binding energies of a relatively new platinum anticancer drug, kiteplatin, with small single strand fragments of DNA, have been studied using B97-D and semi-empirical methods and compared to established drugs cisplatin and oxaliplatin. Isotropic shielding values and *J* coupling constants have also been calculated for these systems to relate these values to conformational data. Extended dual strand kiteplatin-DNA adducts have been studied using the QM/MM method ONIOM, combining BHandH with AMBER, to calculate binding energies and optimised structures. These results show that as the DNA adduct increases in size the values of the kiteplatin energies start to converge and comparison of base pair parameters show that around the site of coordination all fragments show comparable geometrical distortions.

Table of Contents

Chapter 1: Introduction

1.1 Overview

This thesis aims to present the work carried out, using a variety of quantum chemical and hybrid quantum mechanics / molecular mechanics methods, on modelling non-covalent interactions within transition metal complexes and modelling ruthenium and platinum anticancer drugs interacting with differing sizes of DNA adducts. The literature review presented in this chapter gives the background into non-covalent interactions, how they affect the structures of biological macromolecules, and the computational approaches that can be taken to model systems that contain these interactions. Also an overview of platinum and ruthenium anticancer drugs, with a focus on cisplatin and some of its newer analogues and ruthenium based 'piano stool' complexes is presented. This will not only look at the history of such molecules but how they have been modelled previously, as isolated molecules and within DNA adducts.

1.2 Non-Covalent Interactions

Much of the work presented within this thesis focuses on calculating properties of systems that contain non-covalent interactions, particularly at parts the energy of these interactions, which for some have shown to be a particular challenge for computational chemistry, due to their origin in electron correlation. These interactions can be defined as intermolecular or intramolecular interactions between atoms that are not characterised as covalent bonds; they are also sometimes referred to as weak interactions as their strength can be significantly less than that of covalent bonds. In spite of their relative weakness they have an additive nature that allows them to play an important role in many areas such as biological systems, material science, and molecular recognition.[1-9]

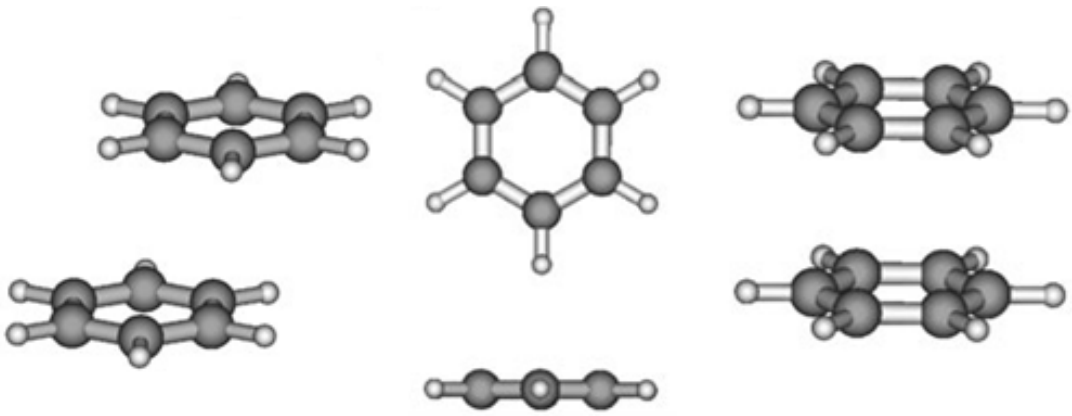
1.2.1 Origin of Non-Covalent Interactions

Non-covalent interactions are of great importance in many areas of science, particularly hydrogen bonding and van der Waals forces. However the ones of greatest interest within this work are those that arise due the correlated motion of electrons, this focuses on interactions such as π...π stacking and CH...π interactions. In contrast, hydrogen

bonding is a result of a charge-charge interaction of an electronegative atom with a hydrogen atom that is directly bonded to another electronegative atom, as well as orbital overlap effects and charge transfer which contribute to the overall strength of these interactions. As a result this makes hydrogen bonding mainly electrostatic in nature and they also can be significantly stronger than dispersion based interactions.[10] The π...π type of non-covalent interactions are a result of two induced dipoles, which occur instantaneously, they are also known as London dispersion forces.

The nature of π...π type interactions wasn't fully understood until Hunter *et al.* devised a model based on observations of porphyrin ring systems.[2] In this model the σ framework and π electrons are treated separately and postulates that the favourable π...π interaction is resultant of the π...σ attraction being greater than the repulsion of the π electron systems, giving a net attractive force.

A key feature in systems containing π...π interactions is how different relative orientations of species results in differing interaction strengths. This is especially prevalent in one of the key examples of this interaction, the benzene dimer, which exhibits three possible conformations, as shown in figure 1.2.1.1. The parallel displaced and T-shaped systems exhibit much more favourable energetic interactions than that of the sandwiched conformation.[11]



Parallel Displaced T-Shaped Sandwiched

Figure 1.2.1.1 Three conformations of the benzene dimer.

The Hunter model attempts to explain this result using three charges to represent the π systems and arrives at three rules. Rule 1, π...π repulsion dominates in a sandwiched π-stacked geometry; rule 2, π...σ attraction dominates in a T-shaped geometry; and rule 3, π...σ attraction dominates in an parallel displaced π-stacked geometry, as shown in figure 1.2.1.2.[2]

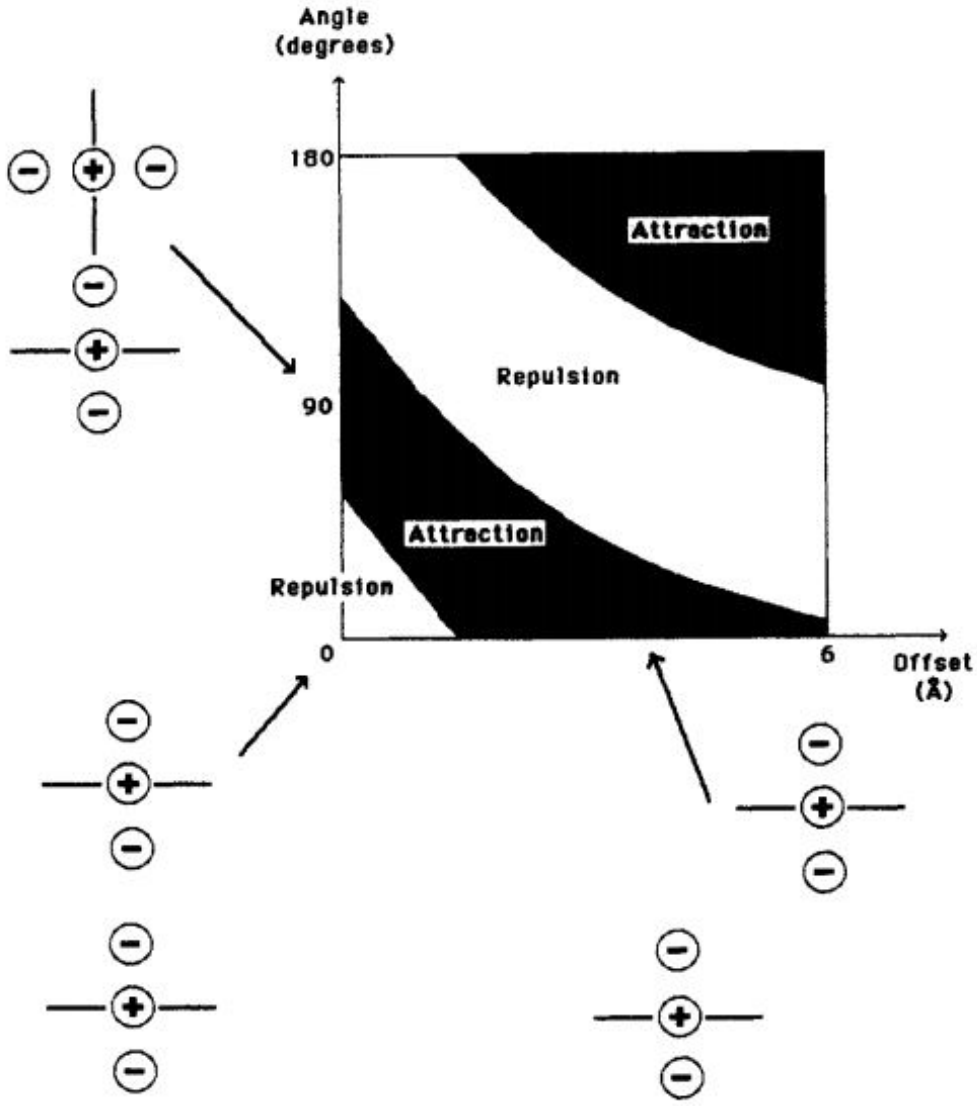


Figure 1.2.1.2 The interactions between two idealised π systems as a function of orientation and distance of parallel displacement.[2]

Although this model is generally considered as the standard explanation for π...π interactions it is not without flaws, the main noticeable one being that the model claims that two overlapping π systems, such as the sandwiched benzene dimer conformation, would result in an unfavourable interaction and therefore repulsion, whilst this is not the case. The benzene dimer preferentially exists in either the parallel displaced or T-shaped conformations but if the two benzenes are forced into the sandwich conformation the

interaction is still favourable, albeit of reduced stabilisation energy. Hill *et al.*[11] calculated these energies using CCSD(T) (*cf.* section 2.4.3) giving the parallel displaced dimer an interaction energy of -2.51 kcal/mol, the T-shaped dimer an energy of -2.49 kcal/mol, and the sandwiched complex an energy of -1.56 kcal/mol. This model does explain the geometrical preference observed in the benzene dimer, however to be able to calculate final binding energies dispersion is vital, which can be shown from errors that stem from computational approaches that do not account for electron correlation.

The CH...π interaction is another class of non-covalent interaction comprising of an energetically stable conformation arising when a C-H bond within a molecule is orientated towards a π system, such as one of the simplest examples the methane benzene system. This interaction can be validated by Hunter's rule 2 and the T-shaped benzene dimer can be thought of as somewhat analogous to a methane...benzene interaction where a CH points toward the centre of the benzene ring. However, the interaction energy of methane...benzene is less than the T-shaped benzene dimer, with CCSD(T) values of -1.45 kcal/mol[12] and -2.49 kcal/mol[11] respectively.

1.2.2 Non-Covalent Interactions in DNA

As mentioned above, non-covalent interactions are of great importance within biological systems, as they are crucial in defining the structures of many biological macromolecules, particularly DNA and proteins. DNA, or deoxyribonucleic acid, is a fundamentally important biological molecule with a double helix of nucleic acid polymer chains, as shown in figure 1.2.2.1.

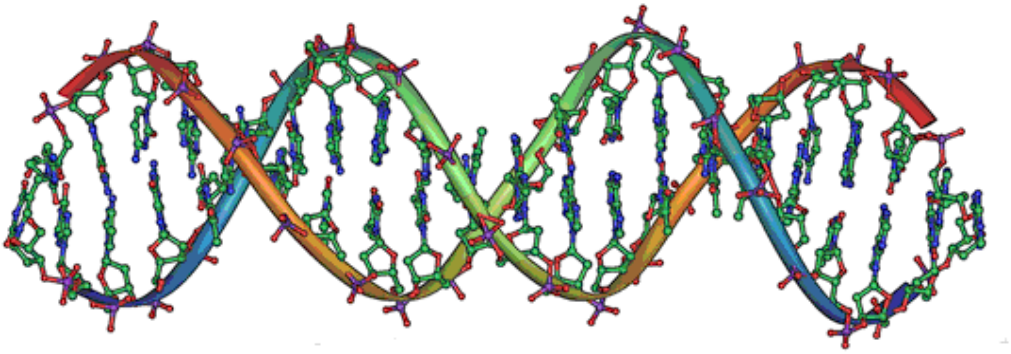
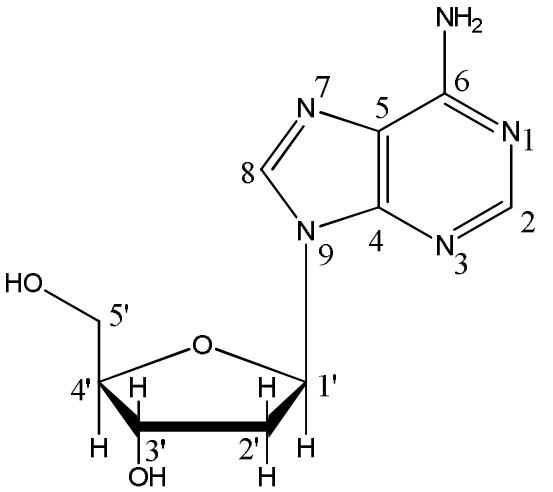
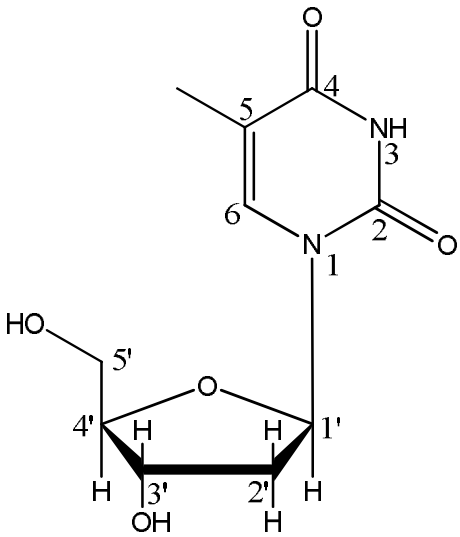


Figure 1.2.2.1 Double helical structure of DNA.

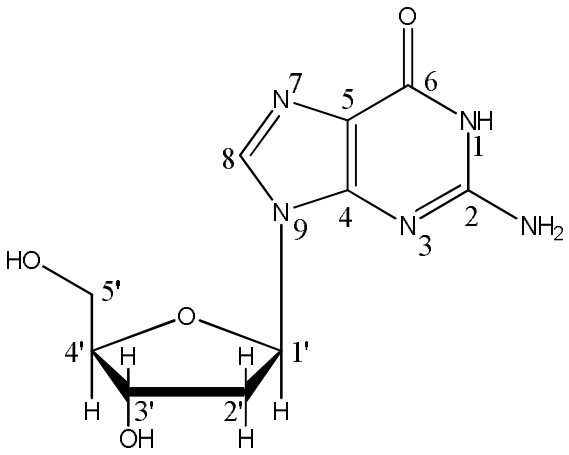
DNA can be characterised by its primary, secondary, and tertiary structure. The primary structure relates to the individual building blocks that make up the nucleic acid chain. These are formed from a deoxyribose sugar and one of four nucleobases, adenine, thymine, guanine, and cytosine (A, T, G, and C respectively) as shown in figure 1.2.2.2. Each of these units is termed a nucleoside and they are linked through phosphate groups, at the 3' and 5' oxygen atoms, to form nucleotides which make up the nucleic acid chains. In models of DNA the structures of the individual strands are usually terminated at the 3' and 5' oxygens, leading to definition of the term 3' and 5' end.



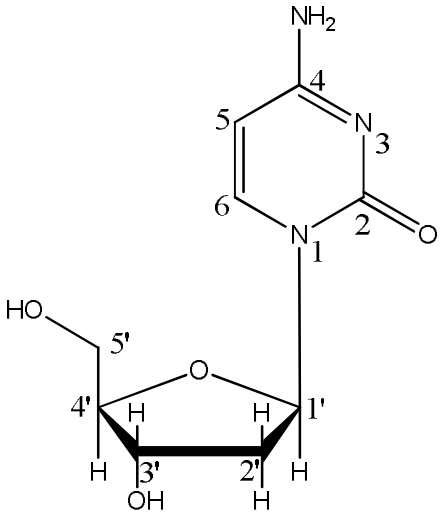
Deoxyadenosine



Deoxythymidine



Deoxyguanosine



Deoxycytidine

Figure 1.2.2.2 Structures of the four possible nucleosides of DNA, with the Crick and Watson numbering.

The secondary structure of DNA relates to the double helical structure that two nucleic acid chains form, as initially discovered by Crick and Watson.[13] The two chains run in anti-parallel directions and the pyrimidine nucleobases form stabilising hydrogen bonds with their complimentary purine bases, G to C and A to T. Nucleobases on the same strand also form additional stabilising interactions as a result of their stacked nature along that strand. As a result of the helical structure of DNA there are grooves present between the strands, adjacent to the nucleobases. These have the ability to act as a binding site, but due to the unsymmetrical nature of the two strands, as a result of the base pairs not being directly opposite, it is possible to form different size grooves, which are termed the minor and major grooves, as shown by figure 1.2.2.3.

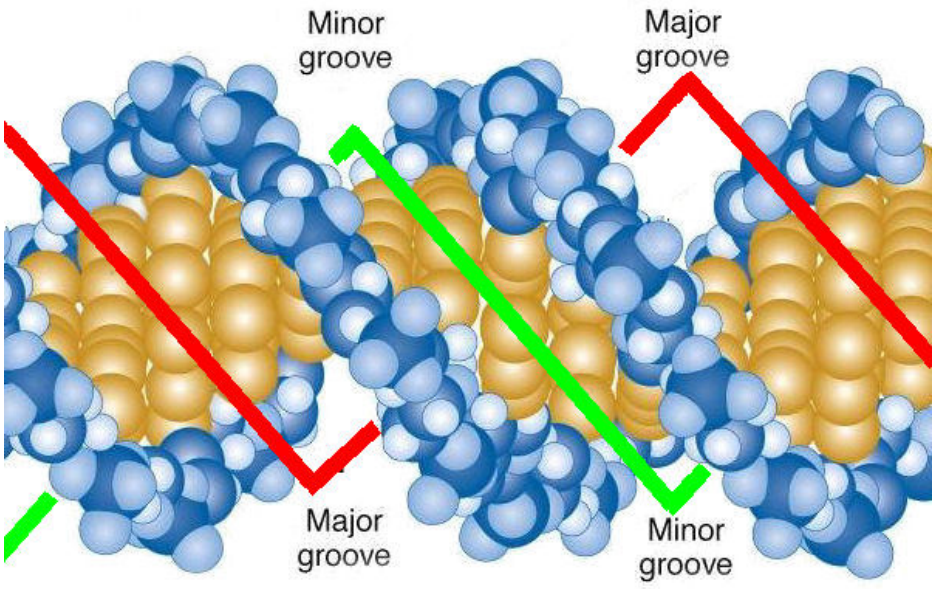
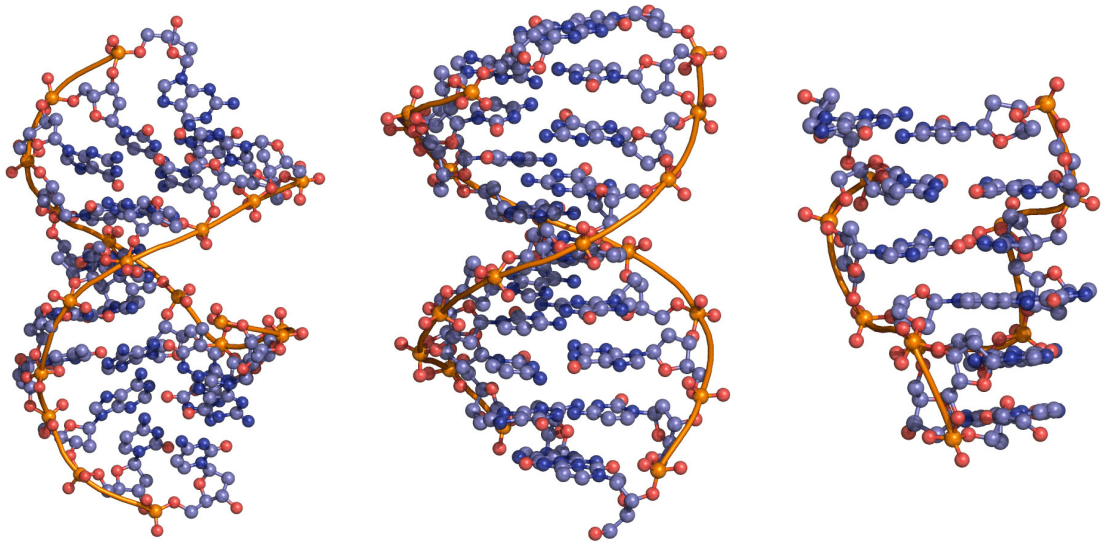


Figure 1.2.2.3 Minor and major groove of a DNA molecule.[14]

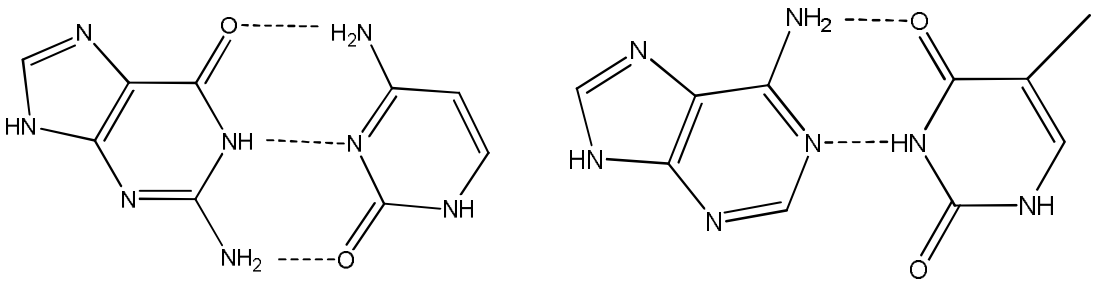
The tertiary structure of DNA relates to the form of the double helix, of which the most common are the A, B, and Z forms, as shown by figure 1.2.2.4. The B form of DNA is the one which is usually found within cells but the others can be observed depending on factors such as the level of hydration and through the binding of drugs.



A-DNA B-DNA Z-DNA

Figure 1.2.2.4 The three most common forms of DNA.[15]

The structure of DNA owes a lot to non-covalent interactions, especially the hydrogen bonds formed between base pairs and the π stacking occurring between adjacent nucleobases on a strand. Each base pair forms multiple hydrogen bonds, G-C forms three and A-T forms two, between the most favoured donor and acceptor atoms on the present bases, as shown by figure 1.2.2.5. This donor/acceptor pattern also allows for effective recognition of G for C and A for T. The multiple hydrogen bonds allow for extra stabilisation between the two nucleobases and this effect is greatly increased across all the consecutive base pairs in a DNA molecule.



GC AT

Figure 1.2.2.5 Hydrogen bonds formed between the GC and AT base pairs.

The hydrogens bonds are considered the strongest non-covalent interaction contributing to the structure of DNA but the π stacking occurring between adjacent nucleobases also gives significant stabilisation. There are also some electrostatic interactions occurring in the DNA molecule, the phosphate groups on the backbone are negatively charged and to balance this positive sodium and magnesium counter ions are present.

1.2.3 Modelling Non-Covalent Interactions

Accurately modelling non-covalent interactions presents an issue for many computational chemistry techniques. With their origin arising from the correlated motion of electrons any technique that approximates this incorrectly, will not result in a quantitatively accurate result and techniques that neglect it usually do not yield qualitatively accurate results. Ideally, to model systems containing these interactions high level correlated wavefunction methods are required, such as coupled cluster techniques. In addition, usually large basis sets of at least an augmented triple-ζ are also required, limiting applicability drastically.[16] At present the most accurate method that is applicable to more than a few atoms is termed CBS(T),[17] which combines an extrapolation of the second order Møller-Plesset correlation energy (MP2), to the basis set limit with a correction using coupled cluster, with single and double excitations and pertubative triples (CCSD(T)), at a smaller basis set, usually of double-ζ quality, as shown by equation 1.1. The extrapolation of the MP2 energy is usually carried out by the method of Helgaker *et al.*[18] where systematically improved basis sets are utilised, most commonly augmented double-ζ to augmented triple-ζ. This approach is believed to yield results within a small fraction of l kcal/mol of the true interaction energies.[19-22]

$$E(CBS(T)) = E(HF) + E^{corr}(MP2)_{CBS} + \Delta CCSD(T)_{VDZ} \qquad \text{(eq. 1.1)}$$

Systems containing hydrogen bonding do not give as much problem computationally as π...π and CH...π systems, as hydrogen bonding is a mainly electrostatic interaction. As a result of this it is possible to model these systems with Hartree-Fock (HF) and density functional theory (DFT) methods with relatively small sized basis sets and still calculate geometries and binding energies to a reasonable level of accuracy. However using these approaches on systems containing CH...π or π stacking often results in unfavourable interaction energies as a result of not properly describing the London dispersion forces,[23] which arise from electron correlation. The HF method is a prime example of

this, as it ignores dispersion by construction, *cf*. section 2.2, and therefore incorrectly models any system that contains stacking interactions.

As mentioned above the post Hartree-Fock methods, which account for electron correlation, are the ideal approach to treat non-covalent interactions, although some of the more inexpensive correlated wavefunction methods can still result in major errors. What could be considered the simplest of these is MP2 and it has been well established that this approach overestimates stacking energies, in some cases up to 100% overestimation can occur.[24,25] Riley *et al*.[26] attributed this failure of MP2, particularly in stacked complexes, to a loss in balance of the factors that compensate for error. Usually these compensating factors, that arise from overestimation in energy due to the failure to describe intramolecular correlation effects and the underestimation of total binding energy as a result of basis set deficiency, are balanced, yielding reasonable binding energies. However in the case of stacked complexes the balance between the two factors shifts resulting in less accurate values, which can explain the overestimations shown above. As a result of the errors within MP2, coupled cluster techniques, such as CCSD(T), and techniques that correct using coupled cluster, such as CBS(T), are widely regarded as those that will give the most accurate results for dispersion based systems.

The drawbacks to these high level methods mean they cannot be realistically used for many systems, as calculation time and therefore cost scales unfavourably with increasing system sizes, for example whilst Hartree-Fock and DFT generally scale at N^4, where N is the number of basis functions, MP2 scales at N^5 and CCSD(T) scales at N^7. As a result of this, an area which has garnered great interest is creating new methods which are computationally inexpensive but can still model non-covalent interactions, as well as modifying current approaches to make them more feasible without significantly losing accuracy.

To evaluate new approaches they are usually compared with benchmark datasets and one that is regularly used is that of Hobza's JSCH-2005 set.[27] This dataset uses CCSD(T) and MP2 to calculate the binding energies of a selection of non-covalently bound nucleobase and amino acid pairs, exhibiting a variety of H-bonding and stacking interactions. This set of over one hundred complexes was also reduced to a smaller set, termed the S22 set, which contains seven hydrogen bonding complexes, eight dispersion based complexes, and seven complexes containing mixed interactions. As this S22

dataset contains a variety of non-covalent interactions, it is often the set to which newly developed methods are benchmarked against. This set has however recently been superseded by the newer S66 dataset,[28,29] which uses CBS(T) to calculate interaction energies, equilibrium geometries, and dissociation curves of sixty six intermolecular complexes at nine distinct geometric configurations.

One direction in which research has been taken to accurately modelling dispersion based systems is to attempt to reduce the computational cost of the post Hartree-Fock methods. Not only does this aid in modelling non-covalent interactions but also allows these correlated wavefunction methods to be used on a wider range of systems, that previously would have been too computationally expensive. One such approach is the use of local correlation methods that exploit the short range nature of electron correlation, which decays as a function of r^{-6}.[30-32] This has been applied to MP2 to give the LMP2 method, *cf.* section 2.4.4. This approach uses localised orbitals to restrict excitations to a set of virtual orbitals, termed domains, which are spatially close to the occupied orbitals, resulting in an overall decrease in computational cost.[33,34] As an additional bonus, by construction these local correlation methods give results which are free of basis set superposition error when used with triple-ζ or larger basis sets, which further reduces cost as other methods such as counterpoise corrections don't need to be made. These methods can also be paired with the density fitting approximation (DF), also called resolution of identity (RI), which attempts to approximate the expensive 4-index-2-electron integrals with a cheaper combination of 2- and 3-index integrals. In some instances this has reduced the time of an MP2 calculation by an order of magnitude.[35-37] When these two approximations are paired with MP2, the resulting DF-LMP2 approach yields a method that in favourable cases has approached linear scaling with system size.[38,39]

DF-LMP2 has been shown to significantly reduce the computational effort when compared to a normal MP2 calculation, but this still does not alleviate the overestimation of stacking interactions. To improve on this it is possible to use an approach that separately scales the correlation energy of parallel and anti-parallel spin electron pairs, termed spin-component scaling (SCS).[40] In the initial form, proposed by Grimme, the scaling parameters were optimised to improve MP2 for a range of thermodynamic, kinetic, and geometrical properties. Different scaling parameters have also been published to account for specific interactions, such as SCSN,[41] SCS(MI),[42]

and SCSC.[43] SCSN has been optimised to improve binding energies of stacked nucleic acid base pairs, SCS(MI) has been optimised against the S22 dataset, and SCSC has been optimised to study catalysis and account for total energies as well as interaction energies. Both SCSN and SCS(MI) give errors of approximately 0.3 kcal/mol when compared to CBS(T) values over a set of representative intermolecular interactions.

Another approach in reducing the computational cost of MP2 calculations is to modify the basis set used. Combining MP2 with small basis sets such as 6-31G* or SVP, but using more diffuse polarization functions than normal yields results of similar accuracy to those carried out with larger basis sets.[44,45] Hobza *et al.*[46-50] used this approach in early electron correlation studies of nucleic acid base stacking, combining MP2 with the 6-31G* basis set but using diffuse d-functions with an exponent of 0.25 instead of the standard 0.8. This approach is sometimes termed MP2(0.25). Hobza and Ŝponer have however noted that this unbalanced basis set could not be used for other calculations, such as geometry optimisation, but this method does allow for a description of base stacking at a lower computational cost as a result of a decreased number of basis functions, compared to a basis set which would normally be deemed acceptable for calculating dispersion based interactions.[46]

Traditionally DFT methods do not accurately describe stacking interactions, but usually perform well for hydrogen bonding systems.[51] As with Hartree-Fock theory there is a lack of dispersion in the construction of standard Kohn-Sham DFT, *cf*. section 2.5. Although, as these methods are less computationally expensive than the correlated wavefunction methods, research has been directed into making DFT methods that can model systems containing dispersion based interactions. Some DFT approaches have been proposed that have been shown to model stacking interactions, these include time dependent DFT,[52] an inclusion of an empirical dispersion correction,[53] and modifying the exchange-correlation functional to model dispersion based interactions.[54,55]

The improvement of the exchange-correlation functional has led to a selection of functionals including KT1 and KT2 developed by Tozer *et. al*,[55] and the Minnesota family of functionals developed by Truhlar.[56-61] Truhlar's functionals include M05, M05-2X, M06, and M06-2X and have shown good performance in modelling non-covalent interactions,[62-66] which has been ascribed to an improved description of

correlation energy at the typical ranges of interacting molecules in van der Waals complexes.

The inclusion of an empirical dispersion correction to standard Kohn-Sham DFT is a method termed DFT-D.[53,67-72] This approach utilises the r^{-6} distance relationship of dispersion based interactions by correcting the Kohn-Sham DFT energy using an atom pairwise sum over r^{-6} potentials. The most common type of this correction takes the form known as D2 but has been refined for a broader range of applicability and higher accuracy in the form of D3, which includes r^{-8} terms as well as r^{-6}.[73] One of these methods that has been used extensively throughout this work is that of B97-D, which includes the D2 correction on the generalised gradient approximation method B97 which was developed by Grimme.[74]

Whilst many DFT approaches do not model stacking interactions, which as mentioned above is not unexpected due to a lack of dispersion in Kohn-Sham DFT, it is possible for methods to work surprisingly well due to what is usually attributed to cancellation of errors. One such approach, which has been used throughout this work, is that of Becke's half and half functional, BHandH.[75] Tests carried out on the benzene dimer show that BHandH performs better than MP2 at describing stacking interactions, whilst the similar hybrid DFT functional B3LYP does not give a favourable interaction at all.[76] This functional has been successfully applied to many systems by our group, including binding of platinum and ruthenium anticancer drugs to DNA and modelling stacking interactions within transition metal complexes.[76-83] However BHandH does consistently overestimate hydrogen bonding, which can lead to issues modelling systems containing a mix of non-covalent interactions.[80]

1.3 Anti-Cancer Drugs

It is possible to utilise the non-covalent interactions mentioned above in the design of anti-cancer drugs. Planar aromatic systems can be designed to act as intercalators, which effectively slot in between two nucleobases such that π...π stacking interactions stabilise this drug DNA adduct.[84-90] Molecules can also be designed to bind to the minor groove of DNA, whilst using non-covalent interactions to stabilise the overall complex.[90-96] Another class of anti-cancer drugs rely on formation of covalent bonds between nucleophilic sites on DNA bases and transition metal centres, as exemplified by

cisplatin and related drugs.[97-107] These are the class of drugs that are of focus here as it is generally believed that the distortion of the natural DNA structure induced by drug binding is significant enough to induce recognition by repair proteins, which leads to a cascade of biological responses ultimately ending in apoptosis, or programmed cell death.[108-110]

1.3.1 Cisplatin

Out of all the transition metal anticancer drugs the ones containing platinum are probably the most well known and out of these cisplatin or (cis-[$PtCl_2(NH_3)_2$]) is one of the best selling. Even though this complex was first synthesised in the 19th century,[111] the true possibility of using platinum to treat cancer wasn't fully realised until the 1960's, when Rosenberg discovered cytotoxicity, or that platinum metal inhibited cell division.[112,113] Although it is now widely used to treat many types of cancer, including testicular, ovarian, cervical, and colorectal cancer, it is not without drawbacks, which include nephrotoxicity and neurotoxicty when doses are exceeded.[114]

As a result of the drawbacks to cisplatin much work has been carried out into improving platinum based anti-cancer drugs as well as attempts to understand the mechanisms occurring when these complexes bind to DNA within tumour cells.[115] This research has led to the discovery of several platinum based drugs, two of which have seen global approval alongside cisplatin, namely carboplatin and oxaliplatin, whilst others such as heptaplatin, nedaplatin, and lobaplatin have been approved locally,[116] as shown in figure 1.3.1.1.

Carboplatin and nedaplatin have a similar proposed mode of action compared to cisplatin, which is attributed to the formation of the same active species when the anionic ligand is lost. Oxaliplatin, heptaplatin, and lobaplatin replace the NH_3 carrier ligands with bulkier chelating amines with an aim to increase cell uptake due to increased hydrophobicity. One of the main differences with these drugs is the observed toxicity, for example carboplatin exhibits less toxicity within cells when compared to cisplatin but has shown to be less efficient due to the chelating effect of the anionic ligand. Oxaliplatin, which is mainly used in treatment of colorectal cancer,[117] forms more hydrophobic adducts compared to both cisplatin and carboplatin and as a result different effects occur within cells.[118] It is also possible for toxicity issues to arise, such

that oxaliplatin treatment can result in severe peripheral neuropathy[114] and with any of these drugs there is also a possibility of tumour cells becoming resistant, through means such as reduced accumulation and repair of the drug DNA adducts.

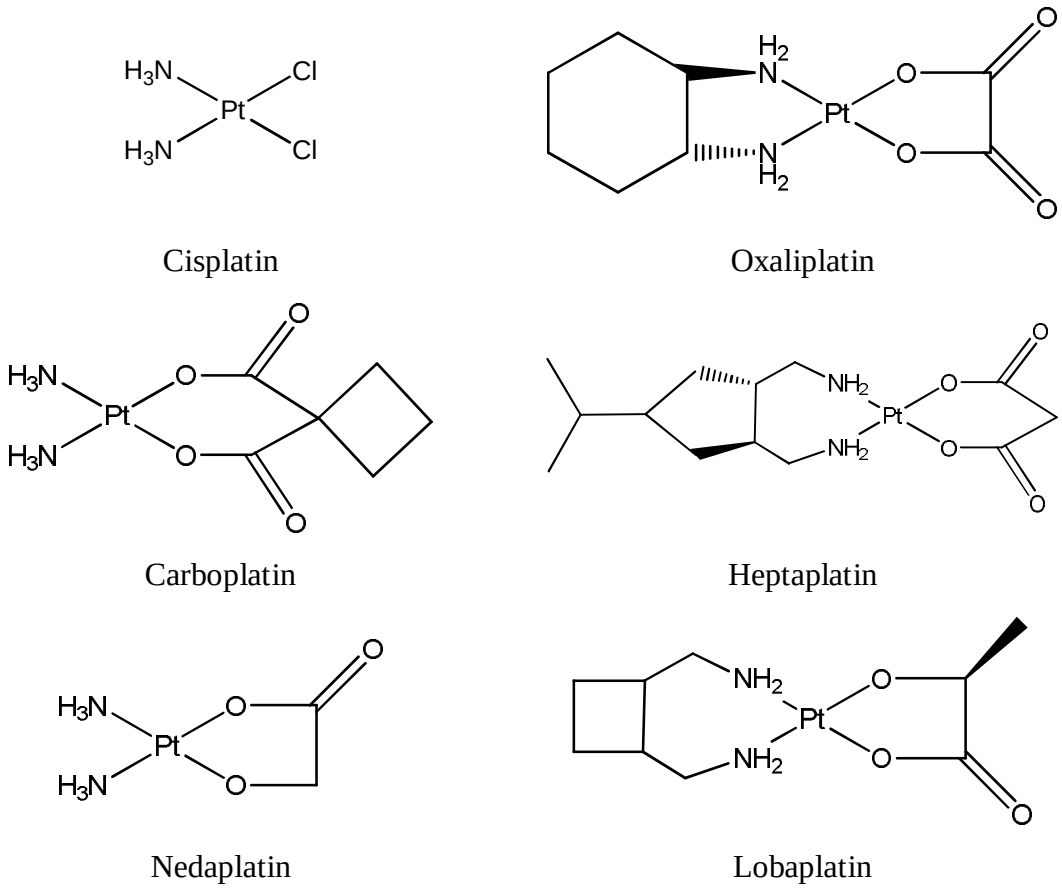


Figure 1.3.1.1 Structures of platinum based anti-cancer drugs

The reason for these complexes acting as anti-tumour agents is a result of formation of cytotoxic lesions on platinum DNA adducts, but there is still debate over how these drugs work. The currently accepted thesis for the mode of action of cisplatin is based around the drug entering cells intact, where there has also been debate between passive diffusion and active transport routes. Once in the cytoplasm, cisplatin is hydrolysed through displacement of the chloride ligands with water molecules, resulting in a positively charged complex. This can then interact with nucleophilic molecules within cells, with DNA a prime target. Cytotoxicity is induced through binding to N7 nitrogen sites on the purine nucleobases, guanine and adenine, which results in a combination of monoadducts and crosslinks causing distortion of the DNA helix. Whilst the 1,2-intrastrand crosslinks are usually associated with the cytotoxicity, it is also possible to form 1,3-intrastrand and interstrand crosslinks.

As there is no confirmed mode of action for these drugs, there is still great interest within this area of research. Computational studies can offer insight that can sometimes not be obtainable through experimental study. As a result of this cisplatin has been extensively studied using computational techniques to probe structure, activation, and the interactions which form DNA adducts. With the size of biomolecules, studying DNA drug adducts with molecular mechanics or semi-empirical methods would seem the ideal approach, as more accurate and computationally expensive approaches are impractical for large system sizes. However within this work there is a focus on using quantum chemical techniques to study DNA fragments where metal binding is occurring at a high level of accuracy, as such the literature here mainly represents studies using *ab initio* and density functional theory.

Some or the earliest attempts at calculations were carried out by Basch *et al.*,[119] to compare cisplatin with its geometrical isomer, transplatin, using Hartree-Fock and double-ζ Gaussian type basis sets, with pseudopotentials on platinum, to find that the trans form is approximately 19 kcal/mol more stable than the cis form, which was attributed to reduced repulsion from the chloride ligands. With the need to account for electron correlation, Carloni *et al.*[120] used gradient corrected exchange correlation DFT with plane wave basis sets and reproduced structural data for cis- and trans-platin, within an estimated error margin of crystal structure data. This DFT study also confirmed the extra stability of transplatin, but by a reduced value of 8 kcal/mol. Pavankumar *et al.*[121] continued this work on the structure and bonding of cisplatin using Møller-Plesset perturbation theory with a variety of basis sets and pseudopotentials, as well as exploring conformations of different symmetry. They found that calculated geometries improved, compared to experimental data, with an increase in number of basis functions and the order of the perturbation theory, up to fourth. Vibrational frequency analysis was used to confirm that found geometries were true minima, but as there was variation with calculated values between the methods and basis sets used, the authors proposed the most suitable approach as MP2/6-311++G(2d,2p).

Whilst extensive studies have shown that structural data and energetic properties can be reproduced within good accuracy, a lot of current interest has been on the activation of cisplatin through hydration. This has focused not only reactions between the complex and water to form the hydrolysed complex through loss of chloride ligands, but also the effect of water within a solvation shell around an intact cisplatin molecule.

Kozelka *et al.*[122] calculated potential energy surfaces for cisplatin with a single water molecule and found favoured conformations and distance dependence. There was also an indication that dispersion forces were in play, as results calculated at Hartree-Fock, DFT, and MP2 levels of theory exhibited significant differences. Lopes *et al.*[123] also calculated similar potential energy surface curves but with a larger number of possible orientations for cisplatin water complexes. Robertazzi *et al.*[124] used the DFT functional mPW1 to calculate optimum geometries of complexes containing cisplatin and a single water, reporting only three stable minima and then characterised the interactions at the minima by examining the electron density.

Not only is the effect of water as a solvent around cisplatin of interest but also the mechanism of hydrolysis to the aqua complex, as well as barriers to ligand substitution, and the energies associated with processes. Zhang *et al.*[125] used a range of pure and hybrid DFT methods with implicit solvent models to study the hydrolysis reaction of cisplatin. They found the reaction went through a 5-coordinate transition state, with a barrier of approximately 23 kcal/mol, with the reported experimental value at 20 kcal/mol. They also found that when comparing the gas phase and solvated data that inclusion of a solvent model greatly affected the barriers, by affecting the geometrical parameters of the transition state and concluded that a solvent model was essential.

Lau and Deubel used the B3LYP DFT functional with a continuum dielectric model to study all three hydrolysis reactions of cisplatin.[126] They calculated similar activation barriers of 25-27 kcal/mol and reaction free energies of 0-2 kcal/mol for the three reactions, in good agreement with experimental values. However the results for the second and third hydrolysis barriers are in strong disagreement with previous computational studies.[127,128] They attribute the differences to the other studies choice of reference states, along the reaction coordinate, to calculate barriers as well as estimating the free solvation and entropic effects. This study is also the first theoretical study to support the diaqua species being the active form of cisplatin, which is in agreement with previous experimental observations.

Mentioned above is that a hydrated cisplatin molecule will interact with DNA and preferentially bind to the nucleophilic oxygen and nitrogen sites on nucleobases, which is generally accepted as the reason for cisplatin's observed cytotoxicity. As a result of this the interaction of cisplatin with DNA, or individual nucleobases, is another area that

has garnered great interest. One of the first studies was carried out by Basch *et al.*,[129] using Hartree-Fock to model the binding of the $[Pt(NH_3)_3]^{2+}$ complex to the four nucleobases A, G, C and T. This showed the preference to bind to the N7 position on guanine (see figure 1.2.2.2), which was approximately 19 kcal/mol more stable than the next most stable site. There is still reasonably strong binding to the other nitrogen and oxygen sites, with the calculated binding strength series as G(N7) > C(N3) > C(O2) > G(O6) > A(N3) > A(N1) > A(N7) > G(N3) > T(O4) > T(O2).

Baik *et al.*[130] carried out studies using B3LYP and implicit solvent models, attempting to explain the reason for cisplatin preferentially binding to guanine over adenine, when these sites are electronically similar in DNA. Hydrogen bonds were observed from the NH moeities on cisplatin to the O6 on guanine and N6 on adenine, with calculated strengths of 7 kcal/mol and 5 kcal/mol respectively, significantly too small to account for the stability difference. Examination of the kinetics of formation of the platinum nucleobase adducts through vibrational frequency data also found that the reaction barrier was smaller for guanine than adenine, with calculated values of 24.6 kcal/mol and 30.2 kcal/mol respectively.

The research outlined above has utilised accurate computational techniques to model cisplatin and small fragments of DNA. There is however still interest in accurately modelling interactions of drugs, where there is more than just small fragments or single nucleobases. To be able to carry out studies of these interactions where the presence of larger biological molecules is desired hybrid quantum mechanics / molecular mechanics (QM/MM) methods are used. These schemes treat a small region of the system, such as the drug and coordinated nucleobase, with a high level method, such as MP2 or DFT, and the remainder of the complex with more approximate and faster molecular mechanics methods.

Robertazzi and Platts[77] used the ONIOM QM/MM approach to combine the DFT functional BHandH and AMBER forcefield to study cisplatin binding to systems of DNA, including small single strand adducts. In this scheme the platinum complex and nucleobases were treated with BHandH, chosen to account for any dispersion interactions present, and the sugar phosphate backbone was treated with AMBER. Inclusion of the backbone was found to not alter trends in binding energies, when compared to isolated nucleobases, but differences in energies and geometries were

observed. A similar approach was taken to model a larger cisplatin adduct, with the octamer of form d(CCTG*G*TCC). This BHandH/AMBER approach was found to satisfactorily reproduce the experimental structure for this, obtained from NMR studies.

Gkionis and Platts[83] carried out studies using QM/MM to test variants of the ONIOM method, by studying cisplatin adducts of a DNA dimer and octamer where BHandH was used for the QM region and AMBER forcefield for the MM region. By comparing optimised structures obtained by different strategies the authors concluded that electrostatic embedding is essential for proper descriptions of the complex and inclusion of an explicit solvent model in the form of water molecules further improves performance. This approach led to good agreement with experimentally obtained structures.

1.3.2 Carboplatin, Oxaliplatin, and Nedaplatin

Whilst there has been significant research into cisplatin, in part due to its small size, well defined geometry, and its significance in the area of anti-cancer agents, other analogous platinum based drugs have not garnered as much interest. In spite of this there is still literature, especially when focusing on comparisons between the newer drugs and cisplatin. One of the first such studies was carried out by Tornaghi *et al.*,[131] comparing carboplatin with cisplatin against experimental crystal structures. Using DFT methods there was relatively small discrepancies between bond lengths and angles, attributed to the inherent errors in the chosen methods.

Giese *et al.* used the B3LYP functional with the LANL2DZ basis set to calculate structures and vibrational spectra of cisplatin, carboplatin, and a variety of novel platinum complexes that exhibited anti-cancer activity.[132] The computational studies were to aid experimental data obtained by Raman spectroscopy and where available the calculations were useful in interpreting experimental results, but due to computational limitations at the time, the chosen methodology was not applicable to the larger complexes.

Wysokiński *et al.* evaluated the performance of several DFT functionals at calculating geometries and vibrational spectra of cisplatin and carboplatin.[133,134] The hybrid functional mPW1PW91 was found to best reproduce the experimental crystal structure

and good agreement with experimental Raman spectroscopy validated the accuracy of this method. Pavelka *et al.*[135] used B3LYP with implicit solvent models to study the hydrolysis reactions of carboplatin as the chelating anionic ligand means that it undergoes slower activation than cisplatin. The calculated initial barrier was found to be approximately 31 kcal/mol and found to proceed through a similar transition state and mechanism to that of cisplatin.

Sarmah and Deka used DFT methods to compare cisplatin and carboplatin with the third generation anti-cancer drug oxaliplatin.[136] Optimised geometries for carboplatin and cisplatin showed good agreement with results discussed above, whilst oxaliplatin was in good agreement with experimental crystal structure, exhibiting the cyclohexane ring in the chair conformation. Parameters, indicative of reactivity, were also calculated for all three drugs, including hardness and electrophilicity, which were in good agreement with the observed experimental trend in reactivities.

Gao[137] used several different DFT functionals with a variety of basis sets to calculate the molecular structure and vibrational spectra of carboplatin, with the aim of investigating performance and determining the most suitable methodology. The author suggests the most suitable approach is LSDA with either the SDD or LANL2DZ basis sets to calculate geometries, indicating this approach is better than the mPW1PW91 functional. For vibrational spectra calculations the PBE0/SDD approach was deemed most suitable.

Tyagi *et al.*[138] used the GGA functional PW91 to study structure and vibrational spectra of oxaliplatin and compare against experimental infra-red spectroscopy. With a focus on the conformation of the diaminocyclohexane ligand, the trans form was found to be more stable than the cis form. Out of the two equally stable trans enantiomers the trans-(1R,2R) form was found to be more biologically potent, which they attributed to chiral recognition by DNA. They also found good comparison between calculated gas-phase geometries and experimental structures to validate the chosen methodology.

The hydrolysis reactions of oxaliplatin have also been studied, by Russo *et al.*[139] using B3LYP with an implicit solvent model, to study the mechanisms in acidic and neutral conditions, such that a H_3O^+ and water molecule were included in the systems for the two conditions respectively. The acidic and neutral activation barriers were calculated at

23.60 kcal/mol and 27.95 kcal/mol respectively, in agreement with experimental data where neutral hydrolysis is slower than the acid counterpart. They also conclude that oxaliplatin exhibits a different behaviour from cisplatin, where in neutral conditions cisplatin will reach DNA in the mono-aqua form, oxaliplatin will be fully hydrolysed.

Whilst cisplatin, carboplatin, and oxaliplatin are the only platinum drugs that have gained global approval, there are a family of newer drugs that have seen local approval and as such there is less available literature based on computational studies of these complexes. As well as studying the hydrolysis of oxaliplatin above, Russo *et al.*[140] used similar methodologies to study the hydrolysis of nedaplatin in acidic and neutral conditions. The acidic and neutral activation barriers were calculated at 31.7 kcal/mol and 28.6 kcal/mol respectively. Through comparison with carboplatin and oxaliplatin they found similar trends and hypothesised that in neutral conditions nedaplatin will reach DNA as the fully hydrolysed complex.

Russo and co-workers have also carried out a comparative study between carboplatin, oxaliplatin, and nedaplatin based around formation of monoadducts with guanine and adenine nucleobases.[141] All complexes in question preferentially bound to guanine over adenine and that binding to guanine was a hydrogen bond controlled process, imposing both structural and kinetic control. Carboplatin was found to have the lowest activation barrier in neutral and acidic conditions for platination of guanine, which was attributed to hydrogen bond networks formed at the transition state geometry. The authors concluded that whilst binding to adenine was considerably slower than guanine, it could be observed in standard conditions, in contrast to previous studies on hydrolysed cisplatin.

1.3.3 Kiteplatin

Kiteplatin, or [(cis-1,4-DACH)$PtCl_2$], is a relatively new platinum anti-cancer drug with an isomeric form of the DACH ligand observed in oxaliplatin, as shown in figure 1.3.3.1. Early reports of the activities of this drug have shown that it exhibits greater potency against platinum resistant cell lines than cisplatin.[142] Further studies by Kohlar and co-workers have shown observed cytotoxicity in murine leukemia and human ovarian cancer cells.[143,144]

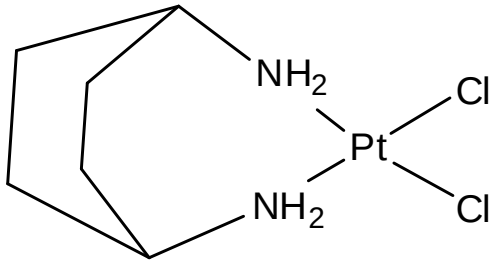


Figure 1.3.3.1 Structure of Kiteplatin.

Margiotta *et al.*[145] combined experimental and computational studies to examine the potential of kiteplatin at treating colorectal cancer. This type of cancer is one of the most common worldwide, with approximately one million new cases diagnosed each year,[146] and the only other drug effective at treating colorectal cancer is oxaliplatin. Comparisons of kiteplatin and oxaliplatin show that the former exhibits enhanced aqueous solubility. However, ONIOM calculations using BHandH/AMBER on kiteplatin and oxaliplatin DNA adducts failed to show significant differences between the two drugs but did show some evidence that kiteplatin induced DNA distortion closer resembles that of cisplatin than oxaliplatin. The authors conclude that kiteplatin could be used very effectively in the treatment of patients with oxaliplatin resistant colorectal cancer. As a result of the relatively recent discovery of the potential application of this drug, there have been very few other computational studies of this complex and as such this forms a large part of the work presented in chapter 5.

1.3.4 Ruthenium Anti-Cancer Drugs

The majority of research in the field of transition metal based anti-cancer drugs is focused on the development of platinum complexes as shown above, but this is not to say that other transition metals could not be used for this application. Some of them that have shown potential anti-tumour activity include titanium,[147] gold,[148,149] rhenium,[150] iridium,[151] copper,[152] cobalt,[153] and ruthenium.[154-156] Also Robertazzi and Platts[157] have studied all the transition metals in the d-block of elements, to examine interactions with guanine and guanine-cytosine base pairs. A preference for binding to the N7 site was observed for most metals, similar to platinum based drugs, but earlier transition metal drugs, such as titanium, favour the O6 site.

The focus here will be on the use of ruthenium transition metal complexes, such as those studied in chapters 3 and 4. Whilst there are several platinum drugs that have been approved and many more in trials there are only two notable ruthenium drugs in clinical trials including (trans-[tetrachlorobis(1H-indazole) ruthenate (III)]).[158] The other of these is (trans-[tetrachloro (dimethylsulfoxide) (imidazole) ruthenate (III)]), which has shown to be effective as an antimetastatic agent.[159] Although its mechanism of action of all steps is unknown, hydrolysis steps do occur with the chloride ligands similar to cisplatin. Several computational studies using B3LYP and implicit solvent models have been carried out on this complex to calculate barriers of the first hydrolysis step with values of 23.2 kcal/mol[160] and 24.8 kcal/mol,[161] with the discrepancies likely due to the use of different solvent models.

Another prominent class of ruthenium complexes with potential anti-cancer activity are those discovered by Sadler,[155,162-165] containing arene ligands. In these complexes the ruthenium is coordinated to an arene ligand, such as biphenyl or anthracene, through η^6 type coordination, with an ethylenediamine chelating ligand and a single chloride ligand, as shown in figure 1.3.4.1. Complexes of this type have shown to have a similar mode of action compared to cisplatin, through hydrolysis of the chloride ligand to give an activated aqua complex which can bind to nucleophilic sites on DNA via a ligand substitution reaction.[166-168]

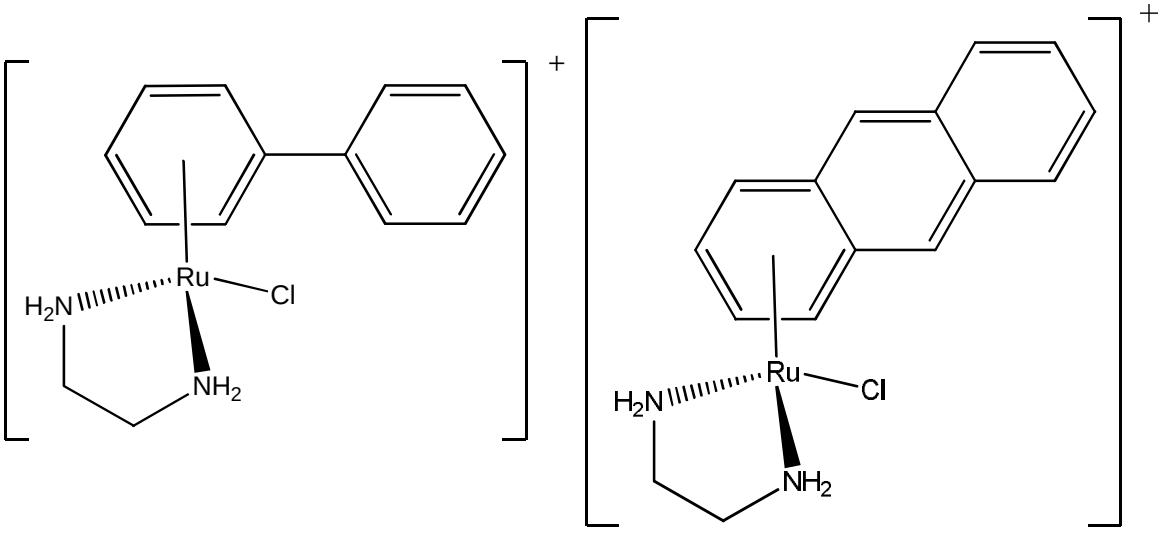


Figure 1.3.4.1 Sadler ruthenium complexes with biphenyl arene ligand (left) and anthracene arene ligand (right).

Sadler *et al.*[169,170] carried out studies on the interactions of these ruthenium complexes to compare the biphenyl and cymene (p-isopropyltoluene) ligands. It was found that the size of the arene has great effect on the activity of the drug, extended arene systems result in greater distortion of DNA, as a result of possible intercalation with the nucleobases. The smaller complexes are unable to form these interactions so a large part of the distortion is resultant of steric interactions.

Gkionis *et al.*[78] used BHandH to investigate the binding of these ruthenium arene complexes to isolated nucleobases with a variety of aromatic ligands. A clear preference to guanine over any other base was found and approximately a 23.9 kcal/mol difference in binding energy between guanine and adenine in the gas phase. Hydrogen bonding and π-stacking interactions were shown to play a significant role in stabilising the bound complexes. Hydrogen bonding occurred independently of the arene, between the nucleobases and the NH of the ethylenediamine ligand, whilst stacking interactions occurred with the larger aromatic systems.

Gossens *et al.*[171] also studied the binding of these ruthenium arene complexes, using benzene and cymene, to guanine, adenine, and cytosine using the DFT functional BP86 and MP2 methods. The calculated trend in strength of binding energies is comparative to that of cisplatin as, G(N7) >> C(O2) ~ C(N3) > A(N7) > G(O6). The same group have also examined the binding of the hydrolysed benzene ruthenium complex with BP86 calculations and docking studies.[166]

Deubel and Lau[172] used B3LYP with implicit solvent models to compare organometallic ruthenium complexes with cisplatin with regards to selectivity of biological targets. They conclude that both cisplatin and the ruthenium complex strongly bind to guanine and both drugs result in cell death, despite that the monofunctional ruthenium complexes are recognised and repaired in cell in a different manner to the bifunctional cisplatin adducts.

Gossens *et al.*[173] used classical and QM/MM molecular dynamics simulations to model a variety of ruthenium arene complexes binding to major grooves on DNA. The complexes studied with the ethylenediamine ligand, the ones of interest here, showed selectivity for guanine rich sequences and are driven to the binding site by electrostatic interactions. When close to binding distance the amino groups form hydrogen bonds

with the O6 atom, while ruthenium coordinates to the N7 atom, which ultimately results in distortion to the Watson-Crick base pair at the cymene side. The ruthenium DNA adduct was calculated to be approximately 11 kcal/mol more stable than the isolated hydrolysed ruthenium complex.

Futera *et al.*[174] used QM/MM techniques, combining B3LYP and AMBER, to study the ruthenium benzene complex interacting with DNA. A preference for binding to N7 of guanine is once again found, which is energetically stabilised by approximately 12 kcal/mol but the process was found to be kinetically inhibited with high barriers of approximately 28 kcal/mol. The possibility of ruthenium acting like cisplatin and forming intrastrand structures was also explored, while the structures were energetically preferred in the QM/MM model, significant geometrical rearrangement was required.

The computational studies carried out on these 'piano stool' ruthenium arene complexes have focused on hydrolysis reactions and interactions with nucleobases, while very few have examined a large range of potential extended aromatic systems. Sadler's studies have shown that the size of the arene can have a great effect on the activity and how much distortion is imbued onto the DNA, as such the work within chapter 3 and within reference 82 is based on a series of different ligands and how much distortion on relatively small base pair steps of DNA occurs.

1.4 References

1. Burley, S.K.; Petsko, G.A. *Science* **1985**, 229, 23.
2. Hunter, C.A.; Sanders, J.K.M. *J. Am. Chem. Soc.* **1990**, 112, 5525.
3. Hunter, C.A.; Singh, J.; Thornton, J.M. *J. Mol. Biol.* **1991**, 218, 837.
4. Gamez, P.; van Albada, G.A.; Mutikainen, I.; Turpeinen, U.; Reedijk, J. *Inorg. Chim. Acta.* **2005**, 358, 1975.
5. Aime, S.; Gianolio, E.; Uggeri, F.; Tagliapietra, S.; Barge, A.; Cravotto, G. *J. Inorg. Biochem.* **2006**, 100, 931.
6. Brammer, L.; Rivas, J.C.M.; Atencio, R.; Fang, S.Y.; Pigge, F.C. *Dalton Trans.* **2000**, 3855.
7. Meyer, E.A.; Castellano, R.K.; Diederich, F. *Angew. Chem-Int. Edit.* **2003**, 42, 1210.
8. Zimmerman, S.C.; Wu, W.M.; *J. Am. Chem. Soc.* **1989**, 111, 8054.
9. Hisamatsu, Y.; Takami, H.; Shirai, N.; Ikeda, S.I.; Odashima, K.; *Tetrahedron Lett.* **2007**, 48, 617-.
10. Hunter, C.A. *Angew. Chem-Int. Edit.* **2004**, 43, 5310.
11. Hill, J.G.; Platts, J.A.; Werner, H.-J.; *Phys. Chem. Chem. Phys.* **2006**, 8, 4072.
12. Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2000**, 122, 3746.
13. Watson, J.D.; Crick, F.H.C. *Cold Spring Harbor Symposia on Quantitative Biology* **1953**, 18, 123.
14. http://www.mun.ca/biology/scarr/MGA2_02-07.html
15. http://www.atdbio.com/content/5/Nucleic-acid-structure
16. Dunning, T.H. *J. Chem. Phys.* **1989**, 90, 1007.
17. Sponer, J.; Riley, K.E.; Hobza, P. *Phys .Chem. Chem. Phys.* **2008**, 10, 2595.
18. Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A.K. *Chem. Phys. Lett.* **1998**, 286, 243.
19. Morgado, C.A.; Jurečka, P.; Svozil, D.; Hobza, P.; Šponer, J. *J. Chem. Theory Comput.* **2009**, 5, 1524.
20. Zhao, Y.; Truhlar, D.G. *J. Phys. Chem. A* **2005**, 109, 6624.
21. Jurečka, P.; Hobza, P. *J. Am. Chem. Phys.* **2003**, 125, 15608.
22. Sinnokrot, M.O.; Sherrill, C.D. *J. Phys. Chem. A* **2006**, 110, 10656.
23. Johnson, E.R.; Wolkow, R.A.; Dilabio, G.A. *Chem. Phys. Lett.* **2004**, 394, 334.
24. Hobza, P.; Selzle, H.L.; Schlag, E.W. *J. Phys. Chem. B* **1996**, 100, 18790.
25. Jurečka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, 365, 89.
26. Riley, K.E.; Platts, J.A.; Řezáč, J.; Hobza, P.; Hill, J.G. *J. Phys. Chem. A* **2012**, 16, 4159-4169.
27. Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, 8, 1985-1993.
28. Řezáč, J.; Riley, K.E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, 7, 2427-2438.
29. Řezáč, J.; Riley, K.E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, 7, 3466-3470.
30. Pulay, P. *Chem. Phys. Lett.* **1983**, 100, 151.
31. Saebø, S.; Pulay. P. *Annu. Rev. Phys. Chem.* **1993**, 44, 213.
32. Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1996**, 104, 6286.
33. Schütz, M.; Hetzer, G.; Werner, H.-J. *J. Chem. Phys.* **1999**, 111, 5691.
34. Hetzer, G.; Schütz, M.; Stoll, H.; Werner, H.-J. *J. Chem. Phys.* **2000**, 113, 9443.
35. Vahtras, O.; Almlöf, J.; Feyereisen, M.W. *Chem. Phys. Lett.* **1993**, 213, 514
36. Hättig, C.; Weigend, F. *J. Chem. Phys.* **2000**, 113, 5154
37. Schütz, M.; Manby, F.R. *Phys. Chem. Chem. Phys.* **2003**, 5, 3349
38. Werner, H.-J.; Manby, F.R.; Knowles, P.J. *J. Chem. Phys.* **2003**, 118, 8149-8160

39. Polly. R.; Werner, H.-J.; Manby, F.R.; Knowles, P.J. *Mol. Phys.* **2004**, 102, 2311.
40. Grimme, S. *Chem. Eur. J.* **2004**, 10, 3423.
41. Hill, J.G.; Platts, J.A. *J. Chem. Theory Comput.* **2007**, 3, 80.
42. Distasio, R.A.; Head-Gordon, M. *Mol. Phys.* **2007**, 105, 1073.
43. Zhao, Y.; Truhlar, D.G. *J. Chem. Theory Comput.* **2009**, 5, 324.
44. Hobza, P.; Sponer, J.; Polasek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792-798.
45. Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590-5596.
46. Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, 124, 11802-11808.
47. Hobza, P.; Šponer, J. *J. Chem. Rev.* **1999**, 99, 3247-3276.
48. Sponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, 61, 3-31.
49. Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1997**, *101*, 9489-9495.
50. Sponer, J.; Leszczynski, J.; Hobza, P. *J. Comput. Chem.* **1996**, 17, 841-650.
51. Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, 7, 1624-1626.
52. Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2003**, 367, 778-784.
53. Grimme, S. *J. Comput. Chem.* **2004**, 25, 1463-1473.
54. Zhao, Y.; Truhlar, D.G. *Theor. Chem. Account* **2008**, 120, 215-241.
55. Keal, T.W.; Tozer, D.J. *J. Chem. Phys.* **2004**, 121, 5654-5660.
56. Zhao, Y.; Truhlar, D.G. *Chem. Phys. Lett.* **2011**, 502, 1-13.
57. Zhao, Y.; Truhlar, D.G. *J. Chem. Phys.* **2006**, 125, 194101.
58. Zhao, Y.; Schultz, N.E.; Truhlar, D.G. *J. Chem. Theory Comput.* **2006**, 2, 364.
59. Zhao, Y.; Schultz, N.E.; Truhlar, D.G. *J. Chem. Phys.* **2005**, 123, 161103.
60. Zhao, Y.; Truhlar, D.G. *J. Phys. Chem. A.* **2006**, 110, 13126.
61. Zhao, Y.; Truhlar, D.G. *J. Chem. Theory Comput.* **2008**, 4, 1849.
62. Zhao, Y.; Truhlar, D.G. *Acc. Chem. Res.* **2008**, 41, 157.
63. Zhao, Y.; Truhlar, D.G. *J. Am. Chem. Soc.* **2007**, 129, 8440.
64. Zhao, Y.; Truhlar, D.G. *Phys. Chem. Chem. Phys.* **2008**, 10, 2813.
65. Zhao, Y.; Truhlar, D.G. *J. Phys. Chem. C* **2008**, 112, 4061.
66. Zhao, Y.; Truhlar, D.G. *Org. Lett.* **2006**, 8, 5753.
67. Piacenza, M.; Grimme, S. *J. Am. Chem. Soc.* **2005**, *127*, 14841-14848.
68. Piacenza, M.; Grimme, S. *Chem.Phys.Chem.* **2005**, *6*, 1554-1558.
69. von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, -.
70. Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287-5293.
71. Peverati, R.; Baldridge, K. K. *J. Chem. Theory Comput.* **2008**, *4*, 2030-2048.
72. Pavone, M.; Rega, N.; Barone, V. *Chem. Phys. Lett.* **2008**, *452*, 333-339.
73. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H.; *J.Chem.Phys.* **2010**, 132, 154104.
74. Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787-1799.
75. Becke A.D. *J. Chem. Phys.* **1993**, 98, 1372.
76. Waller M.P.; Robertazzi, A.; Platts, J.A.; Hibbs, D.E.; Williams, P.A. *J. Comput. Chem.* **2006**, 27,491.
77. Robertazzi, A.; Platts, J.A. *Chem. A Eur. J.* **2006**, 12, 5747-5756.
78. Gkionis, K.; Platts, J.A.; Hill, J.G. *Inorg. Chem.* **2008**, 47, 3893-3902.
79. Gkionis, K.; Hill J.G.; Oldfield, S.P.; Platts, J.A. *J. Mol. Mod.* **2009**, 15, 1051-1060.
80. Gkionis, K.; Platts, J.A. *J. Bio. Inorg. Chem.* **2009**, 14, 1165-1174.
81. Mutter, S.T.; Platts, J.A. *Chem. A Eur. J.* **2010**, 16, 5391-5399.
82. Mutter, S.T.; Platts, J.A. *J. Phys. Chem. A* **2011**, 115, 11293-11302.
83. Gkionis, K.; Platts, J.A. *Comp. Theo. Chem.* **2012**, 993, 60-65.
84. Wall, R. K.; Shelton, A. H.; Bonaccorsi, L. C.; Bejune, S. A.; Dube, D.; McMillin, D. R. *J. Am. Chem. Soc.* **2001**, *123*, 11480-11481.
85. Reha, D.; Kabelac, M.; Ryjacek, F.; Sponer, J.; Sponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P. *J Am Chem Soc* **2002**, *124*, 3366-3376.

86. Cusumano, M.; Di Pietro, M. L.; Giannetto, A. *Inorg. Chem.* **1999**, *38*, 1754-1758.
87. Nakatani, K.; Matsuno, T.; Adachi, K.; Hagihara, S.; Saito, I. *J. Am. Chem. Soc.* **2001**, *123*, 5695-5702.
88. Hecht, C.; Friedrich, J.; Chang, T. C. *J. Phys. Chem. B* **2004**, *108*, 10241-10244.
89. Wheate, N. J.; Brodie, C. R.; Collins, J. G.; Kemp, S.; Aldrich-Wright, J. R. *Mini-Rev. Med.Chem.* **2007**, *7*, 627-648.
90. Strekowski, L.; Wilson, B. *Mutat. Res-Fund. Mol. M.* **2007**, *623*, 3-13.
91. Baraldi, P. G.; Tabrizi, M. A.; Preti, D.; Fruttarolo, F.; Avitabile, B.; Bovero, A.; Pavani, G.; Carretero, M. D. N.; Romagnoli, R. *Pure Appl. Chem.* **2003**, *75*, 187-194.
92. Baraldi, P. G.; Bovero, A.; Fruttarolo, F.; Preti, D.; Tabrizi, M. A.; Pavani, M. G.; Romagnoli, R. *Med. Res. Rev.* **2004**, *24*, 475-528.
93. Yamori, T.; Matsunaga, A.; Sato, S.; Yamazaki, K.; Komi, A.; Ishizu, K.; Mita, I.; Edatsugi, H.; Matsuba, Y.; Takezawa, K.; Nakanishi, O.; Kohno, H.; Nakajima, Y.; Komatsu, H.; Andoh, T.; Tsuruo, T. *Cancer Res.* **1999**, *59*, 4042-4049.
94. Rastogi, K.; Chang, J. Y.; Pan, W. Y.; Chen, C. H.; Chou, T. C.; Chen, L. T.; Su, T. L. *J. Med. Chem.* **2002**, *45*, 4485-4493.
95. Guddneppanavar, R.; Bierbach, U. *Anti-Cancer Agent Me.* **2007**, *7*, 125-138.
96. Baraldi, P. G.; Cacciari, B.; Guiotto, A.; Romagnoli, R.; Zaid, A. N.; Spalluto, G. *Farmaco* **1999**, *54*, 15-25.
97. Baker, E. S.; Manard, M. J.; Gidden, J.; Bowers, M. T. *J. Phys. Chem. B* **2005**, *109*, 4808-4810.
98. Clarke, M. J.; Zhu, F. C.; Frasca, D. R. *Chem. Rev.* **1999**, *99*, 2511-2533.
99. Calamai, P.; Guerri, A.; Messori, L.; Orioli, P.; Speroni, G. P. *Inorg. Chim. Acta.* **1999**, *285*, 309-312.
100. Berners-Price, S. J.; Mirabelli, C. K.; Johnson, R. K.; Mattern, M. R.; Mccabe, F. L.; Faucette, L. F.; Sung, C. M.; Mong, S. M.; Sadler, P. J.; Crooke, S. T. *Cancer Res.* **1986**, *46*, 5486-5493.
101. Collery, P.; Keppler, B.; Madoulet, C.; Desoize, B. *Crit. Rev. Oncol. Hemat.* **2002**, *42*, 283-296.
102. Rixe, O.; Ortuzar, W.; Alvarez, M.; Parker, R.; Reed, E.; Paull, K.; Fojo, T. *Biochem. Pharmacol.* **1996**, *52*, 1855-1865.
103. Bakhtiar, R.; Ochiai, E. I. *Gen. Pharmacol.* **1999**, *32*, 525-540.
104. Murray, J. H.; Harding, M. M. *J. Med. Chem.* **1994**, *37*, 1936-1941.
105. Sun, R. W. Y.; Ma, D. L.; Wong, E. L. M.; Che, C. M. *Dalton. T.* **2007**, 4884-4892.
106. Milacic, V.; Fregona, D.; Dou, Q. P. *Histol Histopathol* **2008**, *23*, 101-108.
107. Milacic, V.; Chen, D.; Giovagnini, L.; Diez, A.; Fregona, D.; Dou, Q. P. *Toxicol. Appl. Pharm.* **2008**, *231*, 24-33.
108. Sherman, S. E.; Lippard, S. J. *Chem. Rev.* **1987**, *87*, 1153-1181.
109. Gonzalez, V. M.; Fuertes, M. A.; Alonso, C.; Perez, J. M. *Mol. Pharmacol.* **2001**, *59*, 657-663.
110. Natile, G.; Marzilli, L. G. *Coordin. Chem. Rev.* **2006**, *250*, 1315-1331.
111. Peyrone, M. *Ann. Chemie Pharm.* **1845**, 51, 129.
112. Rosenberg, B.; Van Camp, L.V.; Krigas, T. *Nature* (London), **1965**, 205, 698.
113. Rosenberg, B.; Van Camp, L.V.; Trosko, J.E.; Mansour, V.H. *Nature* (London), **1969**, 222, 385.
114. Wong, E.; Giandomenico, C.M. *Chemical Rev*, **1999**, 99, 2451.
115. Reedijk, J. *Chem. Commun.* **1996**, 801.

116. Galanski, M.; Jakupec, M.A.; Keppler, B.K. *Curr. Med. Chem.* **2005**, 12, 2075.
117. Jakupec, M. A.; Galanski, M.; Keppler, B. K. *Rev. Physiol. Biochem. Pharmacol.*, **2003**, *146*, 1.
118. Woynarowski, J.M.; Faivre, S.; Herzig, M.C. *Mol. Pharmacol*. **2000**, 58, 920–7.
119. Basch, H.; Krauss, M.; Stevens, W. J.; Cohen, D. *Inorg. Chem.* **1985**, 24, 3313.
120. Carloni, P.; Andreoni, W.; Hutter, J.; Curioni, A.; Giannozzi, P.; Parrinello, M. *Chem. Phys. Lett.* **1995**, 234, 50.
121. Pavankumar, P.N.V.; Seetharamulu, P.; Yao, S.; Saxe, J.D.; Reddy, D.G.; Hausheer, F.H. *J. Comput. Chem.* **1999**, 20, 365.
122. Kozelka, J.; Berges, J.; Attias, R.; Fraitag, J. *Ang. Chem. Int. Ed.* **2000**, 39, 198.
123. Lopes, J.F.; Rocha, W.R.; Dos Santos, H.F.; De Almeida, W.B. *J. Chem. Phys.* **2008**, 128.
124. Robertazzi, A.; Platts, J.A. *J. Comput. Chem.* **2004**, 25, 1060.
125. Zhang, Y.; Guo, Z.; You, X.-Z. *J. Am. Chem. Soc.* **2001**, 123, 9378-9387.
126. Lau, J.K.-C.; Deubel, D.V. *J. Chem. Theory Comput.* **2006**, 2, 103-106.
127. Raber, J.; Zhu, C.; Eriksson, L. A. *Mol. Phys.* **2004**, 102, 2537-2544.
128. Burda, J. V.; Zeizinger, M.; Leszczynski, *J. Comput. Chem.* **2005**, 26, 907-914.
129. Basch, H.; Krauss, M.; Stevens, W.J. ;Cohen, D. *Inorg. Chem.* **1986**, 25, 684-688.
130. Baik, M-H.; Friesner, R.A.; Lippard, S.J. *J. Am. Chem. Soc.* **2003**, 125, 14082-14092.
131. Tornaghi, E.; Andreoni, W.; Carloni, P.; Hutter, J.; Parinello, M. *Chem. Phys. Lett.* **1995**, 246, 469-474.
132. Giese, B.; Deacon, G.B; Kuduk-Jaworska, J.; McNaughton D. *Biopolymers* **2002**, 67, 294-297.
133. Wysokiński, R.; Michalska, D. *J. Comput. Chem.* **2001**, 22, 901-912.
134. Wysokiński, R.; Kuduk-Jaworska, J.; Michalska, D *J. Mol. Struct. Theochem* **2006**, 758, 169-179
135. Pavelka, M.; Lucas, M.F.A.; Russo, N. *Chem. Eur. J.* **2007**, 13, 10108-10116.
136. Sarmah, P.; Deka, R.C. *Int. J. Quant. Chem.* **2008**, 108, 1400-1409.
137. Gao, H. *Spectrochimica Acta Part A* **2011**, 79, 687-693.
138. Tyagi, P.; Gahlot, P.; Kakkar, R. *Polyhedron* **2008**, 27, 3567-3574.
139. Lucas, M.F.A.; Pavelka, M.; Alberto, M.E.; Russo, N. *J. Phys. Chem. B* **2009**, 113, 831-838.
140. Alberto, M.E.; Lucas, M.F.A.; Pavelka, M.; Russo, N. *J. Phys. Chem. B* **2009**, 113, 14473-14479.
141. Alberto, M.E.; Butera, V.; Russo, N. *Inorg. Chem.* **2011**, 50, 6965-6971.
142. Hoeschele, J.D.; Hollis Showalter, H.D.; Kraker, A.J.; Elliott, W.L.; Roberts, B.J.; Kampf, J.W. *J. Med. Chem.* **1994**, 37, 2630-2636.
143. Shamsuddin, S.; Takahashi, I.; Siddik, Z.H.; Khokhar, A.R. *J. Inorg. Biochem.* **1996**, 61, 291−301.
144. Shamsuddin, S.; Santillan, C.C.; Stark, J. L.; Whitmire, K.H.; Siddik, Z. H.; Khokhar, A. R. *J. Inorg. Biochem.* **1998**, 71, 29−35.
145. Margiotta, N.; Marzano, C.; Gandin, V.; Osella, D.; Ravera, M.; Gabano, E.; Platts, J.A.; Petruzzella, E.; Hoeschele, J.D.; Natile, G. *J. Med. Chem.* **2012**, 55, 7182-7192.
146. Van Cutsem, E.; Peeters, M.; Siena, S.; Humblet, Y.; Hendlisz, A.; Neyns, B.; Canon, J.-L.; Van Laethem, J.-L.; Maurel, J.; Richardson, G.; Wolf, M.; Amado, R.G. *J. Clin. Oncol.* **2007**, 25, 1658-1664.
147. Ghosh, P.; D'Cruz, O.J.; Narla, R.K.; Uckun, F.M. *Clin. Cancer Res.* **2000**, 6, 1536-1545.

148. Hoke, G.D.; Macia, R.A.; Meunier, P.C.; Bugelski, P.J.; Mirabelli, C.K.; Rush, G.F.; Matthews, W.D. *Toxicol. Appl. Pharmacol.* **1989**, 100, 293-306.
149. Calamai, P.; Carotti, S.; Guerri, A.; Mazzei, T.; Messori, L.; Mini, E.; Orioli, P.; Speroni, G.P. *Anticancer Drug Des.* **1998**, 13, 67-80.
150. Katsaros, N.; Anagnostopoulou, A. *Crit. Rev. Oncol. Hematol.* **2002**, 42, 297-308.
151. Sava, G.; Giraldi, T.; Mestroni, G.; Zassinovich, G. *Chem. Biol. Interact* **1983**, 45, 1.
152. Yang, P.; Wang, H.F.; Gao, F.; Yang, B.S. *J. Inorg. Biochem.* **1996**, 62, 137-145
153. Jung, M.; Kerr, D.E.; Senter, P.D. *Arch. Pharm.* **1997**, 330, 173-176.
154. Sava, G.; Zorzet, S.; Giraldi, T.; Mestroni, G.; Zassinovich, G. *Eur. J. Cancer Clin. Oncol.* **1984**, 20, 841-847.
155. 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.
156. McNae, I.W.; Fishburne, K.; Habtemariam, A.; Hunter, T.M.; Melchart, M.; Wang, F.Y.; Walkinshaw, M.D.; Sadler, P.J. *Chem. Commun.* **2004**, 1786-1787.
157. Robertazzi, A.; Platts, J.A. *J. Biol. Inorg. Chem.* **2005**, 10, 854-866.
158. Hannon, M.A. *Pure Appl. Chem.* **2007**, 79, 2243-2261.
159. Levina, A.; Mitra, A.; Lay, P.A. *Metallomics* **2009**, 1, 458-470.
160. Chen, J.C.; Chen, L.M.; Liao, S.Y.; Zheng, K.C.; Ji, L.N. *J. Phys. Chem. B* **2007**, 111, 7862.
161. Besker, N.; Coletti, C.; Marrone, A.; Re, N. *J. Phys. Chem. B* **2008**, 112, 3871.
162. Chen, H. M.; Parkinson, J. A.; Parsons, S.; Coxall, R. A.; Gould, R. O.; Sadler, P. J. *J. Am. Chem. Soc.* **2002**, *124*, 3064-3082.
163. Chen, H. M.; Parkinson, J. A.; Morris, R. E.; Sadler, P. J. *J. Am. Chem. Soc.* **2003**, *125*, 173-186.
164. Habtemariam, A.; Melchart, M.; Fernandez, R.; Parsons, S.; Oswald, I. D. H.; Parkin, A.; Fabbiani, F. P. A.; Davidson, J. E.; Dawson, A.; Aird, R. E.; Jodrell, D. I.; Sadler, P. J. *J. Med. Chem.* **2006**, *49*, 6858-6868.
165. Peacock, A.F.A.; Sadler, P. J. *Chem-Asian J.* **2008**, *3*, 1890-1899.
166. Rothlisberger, U.; Gossens, C.; Tavernelli, I. *J. Phys. Chem. A* **2009**, *113*, 11888-11897.
167. Chval, Z.; Futera, Z.; Burda, J. V. *J. Chem. Phys.* **2011**, *134*, 024520.
168. Futera, Z.; Klenko, J.; Sponer, J. E.; Sponer, J.; Burda, J. V. *J. Comput. Chem.* **2009**, *30*, 1758-1770.
169. Novakova, O.; Kasparkova, J.; Bursova, V.; Hofr, C.; Vojtiskova, M.; Chen, H. M.; Sadler, P. J.; Brabec, V. *Chem. Biol.* **2005**, *12*, 121-129.
170. Liu, H. K.; Wang, F. Y.; Parkinson, J. A.; Bella, J.; Sadler, P. J. *Chem-Eur. J.* **2006**, *12*, 6151-6165.
171. Rothlisberger, U.; Gossens, C.; Tavernelli, I. *J. Chem. Theory Comput.* **2007**, *3*, 1212-1222.
172. Deubel, D. V.; Lau, J. K. C. *Chem. Commun.* **2006**, 2451-2453.
173. Rothlisberger, U.; Gossens, C.; Tavernelli, I. *J. Am. Chem. Soc.* **2008**, *130*, 10921-10928.
174. Futera, Z.; Platts, J.A.; Burda, J.V. *J. Comput. Chem.* **2012**, 33, 2092-2101.

Chapter 2: Theory

2.1 Introduction

To understand and analyse the results of calculations and to know how reliable a result is an understanding of the theoretical background is needed. This chapter provides the necessary background on the methods that underpin the computational approaches that have been carried out within this work. The basis of this comes from a selection of the standard computational chemistry textbooks.[1-6]

2.2 Hartree-Fock Theory

2.2.1 Schrödinger Equation

In modern quantum theory there is a fundamental postulate, that for any chemical system a wavefunction, Ψ, exists. This wavefunction contains all information about the system and through manipulation with operators it is possible to obtain the observable properties, with the one that is usually of greatest interest being the total energy of the system. In mathematical notation this equates to the shorthand form of the non-relativistic, time-independent Schrödinger equation.

$$H\Psi = E\Psi \qquad \text{(eq. 2.1)}$$

Expansion of equation 2.1, results in a second order partial differential equation, where H is the Hamiltonian operator and E is a scalar value, which represents the system energy. The operator is expressed in terms of the positions of nuclei, R, and electrons, r, and for a system of N electrons and M nuclei the Hamiltonian can be constructed in the form of equation 2.2.

$$H = -\frac{1}{2}\sum_{i=1}^{N}\nabla_A^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \qquad \text{(eq. 2.2)}$$

Where i and j are electrons, A and B are nuclei, Z is the nuclear atomic number, M_A is the ratio of the mass of nucleus A to an electron, and ∇^2 is the Laplacian operator, equation 2.3.

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad \text{(eq. 2.3)}$$

Equation 2.2 can be further split, where the first and second terms are the kinetic energy operators for the electrons and nuclei respectively and the third, fourth, and fifth terms are the Coulombic interactions representing electron-nucleus attraction, electron-electron repulsion, and nucleus-nucleus repulsion respectively.

2.2.2 Born-Oppenheimer Approximation

For any system containing three or more particles it is impossible to solve the Schrödinger equation exactly, this arises from the correlated motion of particles, and is a fundamental issue known as the three or many-body problem. Therefore to be able to carry out calculations on systems containing more than two particles, several approximations need to be made. One of the fundamental approximations used is the Born-Oppenheimer approximation and it allows the motion of particles to be decoupled by neglecting nuclear kinetic energy and treating the nuclear-nuclear repulsion as a constant. This is possible as there is a large ratio between the masses of the nuclei and the electrons and therefore change in position of electrons can be considered instantaneous in relation to the nuclei. As a result of this the full Hamiltonian can be split into the sum of the nuclear and electronic Hamiltonians.

$$H_{total} = H_{electronic} + H_{nuclear} \quad \text{(eq. 2.4)}$$

$$H_{electronic} = -\frac{1}{2}\sum_{i=1}^{N} \nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} \quad \text{(eq. 2.5)}$$

The total wavefunction of the system is also split into the product of the electronic and nuclear wavefunctions, which results in the simplified Schrödinger equation (eq. 2.7), where electrons move in a fixed field of nuclei.

$$\Psi_{total} = \Psi_{electronic}\Psi_{\text{nuclear}} \quad \text{(eq. 2.6)}$$

$$\mathrm{H}_{electronic}\Psi_{electronic} = E_{electronic}\Psi_{electronic} \quad \text{(eq. 2.7)}$$

The total energy of the system can then be calculated from the electronic energy and the nuclear-nuclear repulsion, which is constant for a given geometry.

2.2.3 Born Interpretation

It is important to be able to interpret the wavefunctions that are present in the above equations as alone they have no clear physical meaning. Born proposed an interpretation using the product of the wavefunction and its complex conjugate, Ψ^*.

$$|\Psi|^2 = \Psi^*\Psi \qquad \text{(eq. 2.8)}$$

The value $|\Psi|^2$, when multiplied by a volume element, $d\tau$, gives the probability of finding an electron within that volume of space. Integration of this probability density over all space should yield the total number of electrons in the system, N.

$$\int \Psi^*\Psi d\tau = N \qquad \text{(eq. 2.9)}$$

If the wavefunction is an eigenfunction of an operator, $\hat{A}$, it is possible to calculate an observable property, A, of that operator, (eq. 2.10). If the wavefunction is normalised the denominator of equation 2.10 equals 1.

$$A = \frac{\int \Psi^*\hat{A}\Psi d\tau}{\int \Psi^*\Psi d\tau} \qquad \text{(eq. 2.10)}$$

2.2.4 Molecular Orbital Approximation

The molecular orbital approximation is a further approximation that can be used in many electron systems, which cannot be solved exactly by the Schrödinger equation, in which the total wavefunction is split into the product of one electron wavefunctions, which are the molecular orbitals. As the electrons are in question here, it is important to not only take into account their spatial coordinates but also their spin orientation, which can be up or down. Electron spin can be described by either the functions $\alpha(\omega)$ or $\beta(\omega)$, where ω is the spin variable. Products of the spatial and spin functions for a single

electron give spin orbitals, χ, and many electron systems can be represented as a product of spin orbitals, which are called Hartree products, Ψ^{HP}.

$$\Psi^{HP}(1,2,\ldots N) = \chi_1(1)\chi_2(3)\ldots\chi_N(N) \qquad \text{(eq. 2.11)}$$

To obey the antisymmetry principle the interchange of the coordinates of any two electrons must also yield a change on the sign of the wavefunction and Hartree products do not satisfy this. However a linear combination of Hartree products does and wavefunctions expressed in this way can be expressed in the form of a Slater determinant.

$$\Psi = \frac{1}{\sqrt{N!}}\begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(N) & \chi_2(N) & \cdots & \chi_N(N) \end{vmatrix} \qquad \text{(eq. 2.12)}$$

This form does obey the antisymmetry principle, as interchanging the coordinates of two electrons corresponds to interchanging two rows of the Slater determinant, which results in a change in the sign of the determinant. As a result of this the Pauli exclusion principle is satisfied, such that two electrons cannot have the same set of quantum numbers, since this would correspond to two rows of the determinant being equal, equivalent to Ψ=0.

The Hartree-Fock method (HF) introduces this to the Schrödinger equation, by use of the one-electron Fock operator, f, which allows the electron-electron repulsion to be expressed in terms of an average potential, V^{HF}, that the i^{th} electron experiences from the remainder of the electrons in the system.

$$f = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + V^{HF} \qquad \text{(eq. 2.13)}$$

$$V^{HF} = \sum_j J_j - K_j \qquad \text{(eq. 2.14)}$$

The first two terms of the Fock operator are the one-electron Hamiltonian of an electron moving in space in relation to the nuclei. The potential, V^{HF}, contains Coulombic

interactions, J, which represent the electrostatic repulsion between the negatively charged electrons, and exchange interactions, K, which describe the spin correlation. Incorporation of the Fock operator gives the Hartree-Fock equation.

$$f\chi(1) = \varepsilon\chi(1) \quad \text{(eq. 2.15)}$$

Due to the form of the Fock operator, which depends on the coordinates of all electrons, a solution for one electron will directly affect the other electron, arising from the V^{HF} term, therefore the Hartree-Fock method must be solved iteratively in a procedure called the self-consistent field method.

2.2.5 Roothaan-Hall Equations

Solving the Hartree-Fock equations directly for molecules is not practical, so an alternate route is taken where each spin orbital is written as a linear combination of atomic orbitals.

$$\chi_i = \sum_{v}^{K} C_{vi}\phi_v \quad \text{(eq. 2.16)}$$

Using equation 2.16, the Hartree-Fock equations take the form of equation of 2.17. This can then be manipulated with a specific basis and through integration to give the Roothaan-Hall equations in matrix form, eq. 2.18.

$$f\sum_{v}^{K} C_{vi}\phi_v = \varepsilon_i \sum_{v}^{K} C_{vi}\phi_v \quad \text{(eq. 2.17)}$$

$$FC = SC\varepsilon \quad \text{(eq. 2.18)}$$

Where F is the Fock matrix, S is the overlap matrix, C is the matrix of coefficients, and ε is a diagonal matrix containing the orbital energies. These orbital energies represent the energy of an electron in a spin orbital and it is therefore possible to relate the total energy of the system to these values through equation 2.19.

$$E = \sum_{i=1}^{N} \varepsilon_i - \sum_{i=1}^{N/2} \sum_{j=1}^{N/2} (2J_{ij} - K_{ij}) \quad \text{(eq. 2.19)}$$

2.3 Basis Sets

The linear combination of atomic orbitals has already been expressed above, by equation 2.16, and is a widely used method for calculating molecular orbital energies and coefficients. The molecular orbitals are expressed as one-electron functions that are centred on the nuclei, also known as basis functions. A set of these functions results in a basis set, which are commonly constructed from one of two functions, Slater-type orbitals (STO), equation 2.20, or Gaussian type orbitals (GTO), equation 2.21.

$$\chi_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi) r^{n-1} e^{-\zeta r} \quad \text{(eq. 2.20)}$$

$$\chi_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi) r^{2n-1-l} e^{-\zeta r^2} \quad \text{(eq. 2.21)}$$

For instances where the electron and nucleus are at long distances or the distance between the two is zero the STOs provide a better description. STOs have a finite gradient at zero distance that correctly describes the cusp at the nucleus and have a more correct exponential decay, than the GTOs, at long range. With STOs, however, one cannot analytically calculate three or four centre two electron integrals that arise while solving the HF equations, which limits their use to high accuracy calculations on atomic systems or methods where the integrals are parameterized, such as semi-empirical methods.

The GTOs have clear shortcomings at small and large distances, as they show a more rapid decay than STOs and lack a description of the cusp when approaching the nucleus. As a result of this between the two functions the STOs are more accurate, but Gaussian functions can be used to express atomic orbitals as linear combinations of the primitive Gaussians to give an approximate STO, this is known as contraction of Gaussian functions. Even though there is an increase in number of basis functions using this approach, where the GTOs need approximately three times as many as the STOs, the integrals are easier to calculate which outweighs the increase in the number of functions.

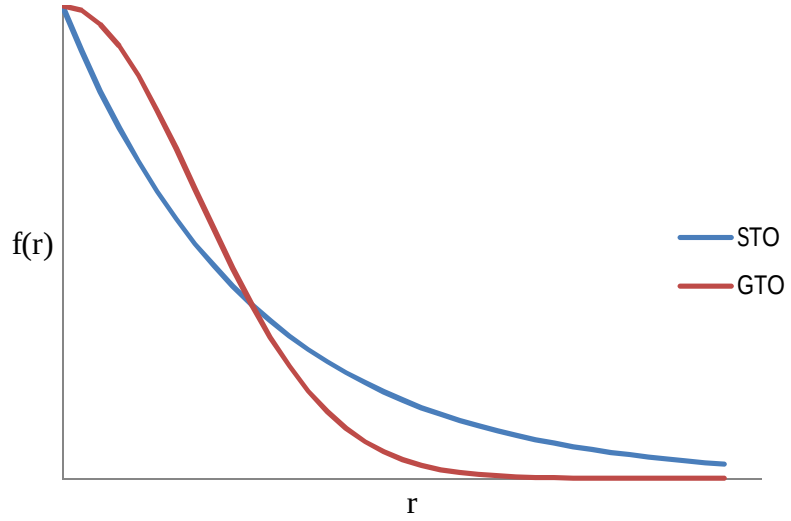


Figure 2.3.1 STO and GTO functions.

2.3.1 Minimal Basis Sets

The minimal basis set is the smallest that describes all electrons within the atoms making up the system in the ground state configuration, where each occupied orbital is described by one basis function. The STO-3G basis set[7] is an example of this where each orbital is represented by three primitive Gaussians. Problems arise with use of minimal basis sets as they are often not accurate enough to correctly describe the system so improvements need to be made. There are several approaches, which include doubling the number of basis functions, resulting in a basis set said to be double-ζ, as opposed to minimal basis sets which are single-ζ. In this case the number of primitive Gaussians does not necessarily change but the number of molecular orbital coefficients increases. The use of split-valence basis can also be applied, where only the valence orbitals have the number of basis functions increased and the core orbitals, which usually do not change during chemical transformations, are treated less accurately. An example of split valence basis sets are those of the Pople series, where one of the common ones is 6-31G, for this the core orbitals are represented by six primitive Gaussians and the valence orbitals are represented by two functions, made up separately of three and one primitive Gaussians.[8]

It is possible to further improve basis sets by including functions that describe higher angular momentum also called polarisation functions. Polarisation functions can include p-functions on hydrogen and d-functions on first row elements, although it is not always

necessary to include higher angular momentum functions on hydrogen, unless it is playing an active role in chemical transformations. Diffuse functions can also be included in a standard basis set that allow for expanded areas of electron density through smaller exponents, this can be useful in treatment of negatively charged species or lone pairs of electrons. Polarisation functions are usually denoted by a *, where the use of two includes both p and d polarisation functions whilst one just includes d polarisation functions. Diffuse functions are denoted by a +, where one applies just to heavy atoms and two include diffuse functions on all atoms. For example the Pople basis set, 6-31+G**, includes diffuse functions on heavy atoms, p polarisation functions for hydrogen, and d polarisation functions for heavy atoms.

2.3.2 Effective Core Potential

Calculations on systems that contain metals or the heavier main group elements, particularly the late transition metals, can cause issues to arise due to the increased amount of electrons. Not only is there an issue relating to large numbers of basis functions to accurately model the system but also relativistic effects can begin to play a part. As the core electrons interact with increasing nuclear charge, their velocities increase to fractions of the speed of light that can no longer be considered negligible. Effective core potentials (ECP), or pseudopotentials can be used in these situations, that describe the core electrons in terms of a potential that accounts for electrostatic effects and often accounts for relativistic corrections as well. These pseudopotentials are constructed by describing the valence electrons with a set of nodeless pseudo orbitals and the core electrons are described by a potential, which is a functional of the distance between the nuclei and electrons.

2.3.3 Counterpoise Method

When basis sets of finite size are used, results of calculations are susceptible to basis set superposition error (BSSE), which can result in the artificial lowering of energies. If the basis set used is not infinite the electronic density near a nucleus can be described by functions that are located on other nuclei and this is the source of the error. As infinite or complete basis sets are not practical for systems, the route taken to correct BSSE is that of the counterpoise method.[9] The counterpoise corrected interaction energy of a complex, AB, can be calculated by equation 2.22.

$$\Delta E = E^{ab}(AB)_{AB} - E^{ab}(A)_{AB} - E^{ab}(B)_{AB} \quad \text{(eq. 2.22)}$$

The superscript values denote the basis set, whilst the subscript denotes the geometry of the fragment. This method calculates the interaction energy of AB, by calculating the energies of the monomers, A and B, while they retain their geometry from the complex and including the basis functions of the other monomer in the form of ghost functions, without the nuclei being present.

2.4 Post Hartree-Fock Methods

2.4.1 Correlation Energy

In section 2.2.4, the HF method was described, in that by using the Fock operator it is possible to average electron-electron interactions by use of a potential. This energy calculated by use of the HF method results in a value that is not exact as it does not take into account the energy arising from correlation (EC). The exact energy can therefore be calculated by summation of the HF energy and the EC energy.

$$E_{exact} = E_{EC} + E_{HF} \quad \text{(eq. 2.23)}$$

The correlation energy is usually of great importance within chemical systems and is vital in the accurate treatment of systems that contain dispersion based interactions, which form as a result of the correlated motion of electrons. The HF method has been advanced to yield more computationally accurate and more expensive methods, which can be collectively called the post-HF methods.

2.4.2 Configuration Interaction

Configuration interaction (CI) allows calculation of energies that account for electron correlation by constructing the total wavefunction as a linear combination of the ground state determinant and excited state determinants, equation 2.24. The HF single determinant wavefunction is represented by Φ_{HF} and the excited determinants are represented by Φ_i, where S, D, and T are the singly, doubly, and triply excited determinants respectively.

$$\Psi = c_0\Phi_{\mathrm{HF}} + \sum_{\mathrm{S}} c_S\Phi_{\mathrm{S}} + \sum_{\mathrm{D}} c_D\Phi_{\mathrm{D}} + \sum_{\mathrm{T}} c_T\Phi_{\mathrm{T}} + \ldots = \sum_{\mathrm{i}=0}^{\infty} c_i\Phi_{\mathrm{i}} \qquad \text{(eq. 2.24)}$$

Inclusion of all excited states leads to the method termed full configuration interaction, which is the most accurate quantum chemical method. As would be expected from a technique that yields such a high level of accuracy, the large number of integrals, as each term in equation 2.24 is a product of basis functions, means that this method is only applicable to the smallest of systems. This approach also has the disadvantage of being size-inconsistent when the series is truncated, such that the energy of a system with monomers at large separation does not equal the sum of the isolated monomer energies.

2.4.3 Coupled Cluster

The coupled cluster (CC) method calculates the excited states by use of an excitation operator, T (eq. 2.25), that is applied to a reference HF wavefunction. T_1 is the operator of single excitations; T_2 is the operator of double excitations and so forth. This allows the coupled cluster wavefunction to take the form of equation 2.26.

$$T = T_1 + T_2 + \ldots T_N \qquad \text{(eq. 2.25)}$$

$$\Psi = e^T\Phi_0 \qquad \text{(eq. 2.26)}$$

$$e^T = 1 + T + \frac{1}{2}T^2 + \frac{1}{6}T^3 + \ldots = \sum_{k=0}^{\infty}\frac{1}{k!}T^k \qquad \text{(eq. 2.27)}$$

If all possible excitations are taken into account by the cluster operator it is possible to calculate a wavefunction equivalent to that of the full CI method, in practice this is computationally expensive and the excitation operator is truncated to give more practical methods such as CCSD and CCSDT, where $T=T_1+T_2$ and $T= T_1+T_2+T_3$ respectively.

2.4.4 Many-Body Perturbation Theory

An alternate route to account for correlation energy, is that of many-body perturbation theory, as proposed by Møller and Plesset. This method is based on Rayleigh-Schrödinger perturbation theory, where the true Hamiltonian is expressed in the form of the zeroth order Hamiltonian, H_0, and a perturbation term, V, which includes a parameter, λ, which takes a value that varies between zero and one.

$$H = H_0 + \lambda V \qquad \text{(eq. 2.28)}$$

The eigenfunction, $\Psi_i^{(0)}$, and eigenvalue, $E_i^{(0)}$, represent the zeroth order Hamiltonian and therefore in a system where λ equals zero these represent the true Hamiltonian. In instances where λ does not equal zero, the eigenfunctions and eigenvalues of the Hamiltonian can be expressed as Taylor expansions of λ, where n denotes the order to which corrections to energy and wavefunction apply.

$$\Psi_i = \Psi_i^{(0)} + \lambda\Psi_i^{(1)} + \lambda^2\Psi_i^{(2)} + \ldots = \sum_{n=0} \lambda^n \Psi_i^{(n)} \qquad \text{(eq. 2.29)}$$

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \ldots = \sum_{n=0} \lambda^n E_i^{(n)} \qquad \text{(eq. 2.30)}$$

The first order correction to the energy is $E_i^{(1)}$, the second order correction to the energy is $E_i^{(2)}$, and so forth. These energies can be calculated from the eigenfunctions, as shown by equations 2.31-2.34 below, showing the zeroth order correction to the third order correction.

$$E_i^{(0)} = \int \Psi_i^{(0)} H_0 \Psi_i^{(0)} d\tau \qquad \text{(eq. 2.31)}$$

$$E_i^{(1)} = \int \Psi_i^{(0)} V \Psi_i^{(0)} d\tau \qquad \text{(eq. 2.32)}$$

$$E_i^{(2)} = \int \Psi_i^{(0)} V \Psi_i^{(1)} d\tau \qquad \text{(eq. 2.33)}$$

$$E_i^{(3)} = \int \Psi_i^{(0)} V \Psi_i^{(2)} d\tau \qquad \text{(eq. 2.34)}$$

In Møller-Plesset perturbation theory the zeroth order Hamiltonian is the sum of the one electron Fock operators, so to show an improvement over the HF method, equation 2.33 shows that at least a second order correction to the energy needs to be applied, this method is referred to as MP2.

Compared to the truncated CI methods, Møller-Plesset perturbation methods have shown to be size consistent for all orders, however these methods do not obey the variational theorem, *i.e.* the energy calculated from the approximation of the true wavefunction might not necessarily be greater than the true energy, whilst the HF method does obey this principle. Another issue in using this method is, that along with CI and CC methods, the calculations can be expensive. Perturbation methods are not as impractical as CI and CC, however their use is still limited to relatively small system sizes. However further variations have been applied to these methods in order to make them more applicable for larger scale systems. One of these is the use of localized orbitals, where excitations are restricted to sets of virtual orbitals, termed domains, that are spatially close to the occupied orbitals. Application of this local correlation method to MP2 is termed LMP2 and due to use of localized orbitals BSSE can be effectively eliminated when a large enough basis set is used.[10] Another route to decrease computational cost is to replace the expensive calculation of the 4-index-2-electron integrals with a cheaper combination of 2- and 3-electron integrals. This is termed density fitting, DF, and in instances has shown to reduce computational cost of MP2 by an order of magnitude.[11,12] Combination of density fitting and local correlation methods has been used throughout this work and when used with MP2 is termed DF-LMP2.[13]

2.5 Density Functional Theory

Density functional theory (DFT) has become a widespread method that is more computationally inexpensive than the post-HF methods mentioned above, but can improve on results from the HF method, by giving good geometries and it can account for electron correlation. Instead of calculating the wavefunction, DFT calculates the total electronic density distribution as a means to calculate the electronic energy.

2.5.1 Hohenberg-Kohn Theorems

The origins of DFT are found in the Thomas-Fermi model from the late 1920s, where the electronic energy was attempted to be calculated from the electron density, but flaws in this approach, such as the inability to describe molecular bonding, rendered this method impractical for any real system. However the first real advancement in the use of DFT for molecules came from the two Hohenberg-Kohn theorems, developed in 1964. Their first theorem demonstrates that the ground state energy and other system properties can be uniquely defined by the electron density, which can be expressed in a way that the energy is a unique functional of the density, ρ(r).

$$E[\rho(r)] = \int V_{ext}(r)\rho(r)dr + F[\rho(r)] \qquad \text{(eq. 2.35)}$$

The first term of equation 2.35 represents the interaction of electrons with an external potential, $V_{ext}(r)$, which is usually due to Coulombic interactions with nuclei. The second term, F[ρ(r)], is the sum of the electron's kinetic energy and interelectronic interactions. The second theorem demonstrates that the density obeys the variational principle, such that any trial density will result in an energy that is greater than or equal to the exact energy.

2.5.2 Kohn-Sham Equations

The second significant advancement in DFT came in 1965, where Kohn and Sham proposed a practical way to solve the Hohenberg-Kohn theorems for interacting electrons. The issue with the expression for the first theorem in equation 2.35 is that the term F[ρ(r)] is not known, Kohn and Sham approached this as a sum of terms to represent contributions to the total function.

$$F[\rho(r)] = E_{KE}[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)] \qquad \text{(eq. 2.36)}$$

$E_{KE}[\rho(r)]$ is the kinetic energy, $E_H[\rho(r)]$ is the electron-electron Coulombic repulsion energy, and $E_{XC}[\rho(r)]$ contains contributions from both exchange and correlation. The kinetic energy term is defined as the kinetic energy of a system of non-interacting electrons that has the same density as the real system and the electron-electron repulsion

term is the Hartree electrostatic energy, using these terms the full expression for the energy for the Kohn-Sham method can be expressed as equation 2.37.

$$E[\rho(r)] = \sum_{i=0}^{N} \int \psi_i(r) \left(-\frac{\nabla^2}{2}\right) \psi_i(r) dr + \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 + E_{XC}[\rho(r)] - \sum_{A=1}^{M} \int \frac{Z_A}{|r - R_A|} \rho(r) dr \quad \text{(eq. 2.37)}$$

2.5.3 Local Density Approximation

The difference between most DFT methods lies in the exchange-correlation functional, $E_{XC}[\rho(r)]$, if this term was exact it would be possible to find a perfect solution for any system, but as this is not known approximations are made and one of the advantages of DFT is that even relatively simple approximations give favourable results. The simplest of these is local density approximation, LDA, which is based on the uniform electron gas model, in which the electron density is constant throughout space, therefore integration over all space will give the total exchange-correlation energy, where the term $\varepsilon_{XC}[\rho(r)]$, is the exchange-correlation energy of a uniform electron gas.

$$E_{XC}[\rho(r)] = \int \rho(r) \varepsilon_{XC}[\rho(r)] dr \quad \text{(eq. 2.38)}$$

LDA assumes that spin is zero at all points, so in instances where α and β densities are not equal the more general method local spin density approximation, LSDA, is used.

2.5.6 Generalised Gradient Approximation

The use of an ideal model, such as the uniform electron gas model, has inherent errors. To overcome errors the exchange-correlation energies should not only depend on the density but the derivatives of the density should also be considered. These methods are known as gradient corrected or as generalised gradient approximation, GGA. These methods are considered as corrections to LDA and LSDA, such as the popular functional proposed by Becke[14], B88.

$$E_x^{B88} = E_x^{LDA} - \gamma \sum_s \int \frac{\rho_s(r)^{\frac{4}{3}} x_s^2}{1 + 6\gamma x_s sinh^{-1} x_s} dr \qquad \text{(eq. 2.39)}$$

$$x_s = \frac{|\nabla \rho_s(r)|}{\rho_s(r)^{\frac{4}{3}}} \qquad \text{(eq. 2.40)}$$

2.5.7 Hybrid DFT Methods

Hybrid DFT methods are those that use a combination of the exact HF exchange energy with LSDA exchange energy, these can also include gradient correction terms. One of the most widely used DFT methods is that of Becke's three-parameter functional, B3LYP.[15] This method uses a combination of exact, LSDA, and B88 exchange with a combination of LSDA and Lee-Yang-Parr, LYP,[16] correlation energy, where the terms are weighted using three parameters by fitting to experimental data. Another hybrid method, that has been used extensively throughout this work, is that of Becke's Half and Half functional, BHandH.[17] This functional uses half exact and half LDA exchange coupled with the LYP correlation term, as shown by equation 2.41.

$$E_{XC}^{BHandH} = \frac{1}{2} E_X^{HF} + \frac{1}{2} E_X^{LDA} + E_C^{LYP} \qquad \text{(eq. 2.41)}$$

2.5.8 Dispersion Corrected DFT

The use of DFT shows that electron correlation is accounted for and as these methods are much more computationally efficient than the post-HF methods above, it is a useful technique to get favourable geometries and energies on systems beyond the capability of the more accurate correlated methods. But whilst electron correlation is accounted for, the inexact treatment of the correlation term means that systems containing van der Waals forces, such as dispersion based interactions, do not give proper descriptions of the r^{-6} relationship at longer distances. However this is not the case for all functionals, for example BHandH has been shown to be effective for dispersion based systems, through what is assumed as a cancellation of errors.[18]

There are other approaches to account for dispersion in DFT methods, functionals can be highly parameterised to account for these interactions, such as Truhlar's Minnesota

family of functionals.[19,20] Another route which has been used throughout this work is that of an empirical correction, in the form of an atom pairwise sum over r^{-6} potentials, to the Kohn-Sham energy to account for dispersion, this technique is termed DFT-D and in the form below was formulated by Grimme.[21]

$$E_{DFT-D} = E_{KS-DFT} + E_{disp} \quad \text{(eq. 2.42)}$$

$$E_{disp} = -s_6 \sum_{i=1}^{N_{at}-1} \sum_{j=i+1}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}) \quad \text{(eq. 2.43)}$$

$$f_{dmp}(R_{ij}) = \frac{1}{1 + e^{-d(R_{ij}/R_r - 1)}} \quad \text{(eq. 2.44)}$$

Where N_{at} is the number of atoms in the system, R_{ij} is the distance between two atoms i and j, C_6^{ij} is the dispersion coefficient for the two atoms, s is a global scaling factor which is dependent on the functional used, and f_{dmp} is a damping function as shown by equation 2.44. The dispersion correction in equation 2.42 is termed D2 and has been refined for a broader range of applicability and higher accuracy. This newer approach is termed D3 and also takes into account r^{-8} terms as well as r^{-6} terms, as shown by equation 2.45.[22]

$$E_{disp}^{D3} = -\frac{1}{2} \sum_{j \neq i} \left(s_6 \frac{C_6^{ij}}{R_{ij}^6 + [f_{dmp}(R_{ij})]^6} + s_8 \frac{C_8^{ij}}{R_{ij}^8 + [f_{dmp}(R_{ij})]^8} \right) \quad \text{(eq. 2.45)}$$

2.6 Quantum Mechanics/Molecular Mechanics

The methods above can be used to great effect when carrying out calculations on small to medium size systems, but issues can arise when the desired system is too large, this is particularly prevalent in biological systems, such as DNA, or systems that require the presence of explicit solvent molecules. In these cases a combination of quantum mechanics and molecular mechanics (QM/MM) can be used, where the part of the system that is of greatest interest is treated at the QM level and the remainder is treated with a force field method, which was first introduced by Warshel and Levitt in 1976.[23]

The total energy of the system can be written as the sum of the QM energy, E_{QM}, the MM energy, E_{MM}, and the energy of the interaction between the two parts of the system, $E_{QM/MM}$.

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM} \quad \text{(eq. 2.46)}$$

2.6.1 Molecular Mechanics

Molecular Mechanics (MM), sometimes known as force field methods, differ from quantum mechanical methods in a way that electronic motion is neglected and energy is calculated as a function of nuclear positions. As a result of this MM can be used for systems that contain a large amount of atoms and can yield results of reasonable accuracy when compared to *ab initio* methods, at a fraction of the computational cost. A typical force field calculates energy through the summation of five potential terms as shown in equation 2.47.

$$\begin{aligned} E = & \frac{1}{2}\sum_{bonds} k_i(l_i - l_{i,0})^2 + \frac{1}{2}\sum_{angles} k_i(\theta - \theta_{i,0})^2 \\ & + \frac{1}{2}\sum_{torsions} V_n(1 + cos(n\omega - \gamma)) \\ & + \sum_{i=1}^{N}\sum_{j=i+1}^{N}(4\varepsilon_{ij}[(^{\sigma_{ij}}/_{r_{ij}})^{12} - (^{\sigma_{ij}}/_{r_{ij}})^{6}] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}) \end{aligned} \quad \text{(eq. 2.47)}$$

The first term represents a summation of the interaction between pairs of bonded atoms, modelled as a harmonic potential that shows an increase in energy as bond length deviates from the reference. The second term models all intermolecular bond angles and is also represented by a harmonic potential. The third term is periodic potential to model energy change of bond rotation. The fourth term combines the non-bonding interactions, through use of the Lennard-Jones potentials, and the electrostatic interactions through Coulombic potentials, respectively.

The AMBER force field has been used within this work and stands for Assisted Model Building and Energy Refinement[24]. It has been calibrated for use in biological simulations of nucleic acids and proteins, against diffraction studies and quantum mechanical studies. This model has some differences from the standard form, to be

effective in its treatment of biological systems, such as a potential term in the form r^{-12}-r^{-10}, being included to explicitly model hydrogen bonds and inclusion of explicit lone pairs in sulphur hydrogen bonding.

2.6.2 ONIOM

ONIOM is a QM/MM scheme that stands for Our own N-layered Integrated Orbital Molecular mechanics, that was developed by Morokuma and co-workers[25-28]. This method allows use of multiple layers but will be restricted to two layers, high and low for QM and MM respectively, within this work. The extrapolated ONIOM energy is written in the form of equation 2.48, where the term real relates to the entire system and the term model relates to the part of the system that is of interest.

$$E_{high,real}^{ONIOM} = E_{high,model} + E_{low,real} - E_{low,model} \quad \text{(eq. 2.48)}$$

2.7 Semi-Empirical Methods

Semi-empirical methods are approximate molecular orbital theories which attempt to reduce the computational cost of the traditional Hartree-Fock method. The source of the greatest proportion of time spent carrying out a HF calculation is the calculation and manipulation of integrals so in order to decrease computational effort the aim is to approximate or neglect some of these integrals. This can be achieved in part by only considering the valence electrons explicitly, whilst the core electrons are accounted for in the nuclear charge or with functions that model repulsion from the combined nuclei and electrons.

Many semi-empirical methods are based on the zero differential overlap approximation, which sets the overlap between pairs of different orbitals to zero for all volume elements. This results in a simplification of the Roothaan-Hall equations, as the overlap matrix is reduced to a unit matrix, and all three and four centre integrals being set to zero. These approximations are compensated by including parameterisation of the remaining integrals, which can be based on experiments or more accurate calculations.

2.8 Quantum Theory Atoms In Molecules

2.8.1 Electron Density

Quantum Theory Atoms In Molecules (QTAIM) is a method to partition the electron density between the atoms in a molecule as a population analysis scheme and was developed by Bader.[29] This approach is based on the electron density, which can be defined in terms of the wavefunction, as shown in equation 2.49.

$$\rho(r) = \int \psi^*(r)\psi(r)dr \qquad \text{(eq. 2.49)}$$

Analysis of the topology of the electron density can be used to define key points in the system specifically through the gradient of the density. This gradient vector forms a curve around the molecule that at all points is perpendicular to the contours of electron density and must have a beginning, which is usually infinity, and an end where it disappears, which are referred to as a critical points, as represented for ethene in figure 2.8.1.1.

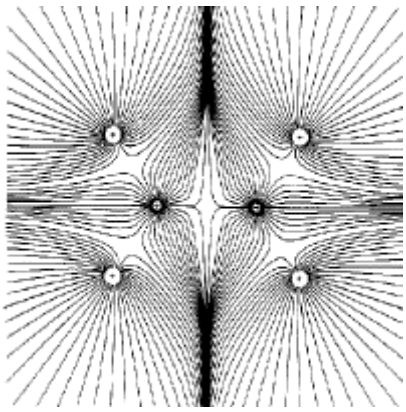
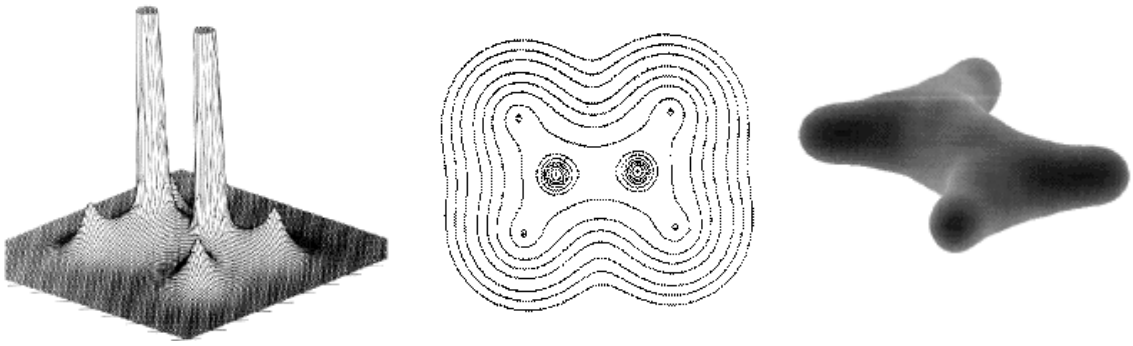


Figure 2.8.1.1 Gradient vector paths associated with an ethene molecule.[6]

The electron density can also be used to observe molecules in other ways such as 3D maps of the density in relation to planes of molecules, the contours of the density can be shown in 2D plots, and an isosurface can be mapped to show molecular shape and the surface of the molecule, as shown by representations of ethene in figure 2.8.1.2.



a) b) c)

Figure 2.8.1.2 Electron density of ethene represented as a) 3D relief, b) 2D contour map, and c) isodensity surface.[6]

2.8.2 Critical Points

Critical points occur when the gradient vector equals zero and they can be maxima, minima, or saddle points in relation to the curvature of the electron density. This can be found by diagonalising the 3x3 Hessian matrix, the second derivatives, of the electron density and the eigenvalues from this show the curvature in each direction. Each critical point can be classified by rank, which equals the number of non zero eigenvalues and in the case where all three eigenvalues are non zero there are four possible types.

1. (3,-3): This type has all three eigenvalues negative and shows a local maximum in the density, this is indicative of nuclear positions.
2. (3,-1): This type has two of the eigenvalues negative with the density maximum in the plane that corresponds to the axes of the negative eigenvalues, this is indicative of a chemical bond and is termed a bond critical point (BCP).
3. (3,+1): This type has two of the eigenvalues positive with the density minimum in the plane that corresponds to the axes of the positive eigenvalues, this is found within rings and is termed a ring critical point (RCP).
4. (3,+3): This type has all three eigenvalues positive and shows a local minimum in the density, this if found within cages and is termed a cage critical point (CCP).

The BCP is usually of most interest as it defines a bond path between the critical point and the nuclei along which the electron density is a maximum. This bond path is indicative of an interaction between two atoms but this does not necessarily indicate a

covalent interaction. At these points along with the density, the Laplacian of the electron density should also be considered, $\nabla^2\rho$.

$$\nabla^2\rho(r) = \frac{\partial^2\rho(r)}{\partial x^2} + \frac{\partial^2\rho(r)}{\partial y^2} + \frac{\partial^2\rho(r)}{\partial z^2} \qquad \text{(eq. 2.50)}$$

$\nabla^2\rho$ shows a measure of the local concentration of the charge density, *i.e.* when $\nabla^2\rho > 0$ there is a local depletion of electron density and when $\nabla^2\rho < 0$ there is locally concentrated charge. Using this it is possible to characterise the nature of interactions at the BCPs. An example being hydrogen bonds, which have an electron density at critical points an order of magnitude smaller than covalent bonds as well as a positive Laplacian of the electron density.[30]

2.9 Practical Considerations

2.9.1 Geometry Optimisation

The Potential Energy Surface, PES, of a system is a hypersurface of the potential energy which is a function of the nuclear positions. Where there are N nuclei in a system there are 3N-6 coordinate dimensions to describe the nuclear arrangement, as a result of this to visualise the PES in polyatomic systems cannot be done without taking into account derivatives of the energy. Stationary points can be found at places on the surface where the first derivative of the energy is equal to zero, these are usually the points of greatest interest in chemical systems as they equate to minima, maxima, and saddle points.

As there can be several of these stationary points on the surface the goal of geometry optimisation is to find the lowest minimum, which is also referred to as the global minimum. The global minimum corresponds to the most stable conformation and the lowest energy, whilst other local minima correspond to lower energy areas in relation to adjacent areas on the surface but are not the most stable configurations of the system. It is possible for systems to move between minima with the addition of energy to the system, this is achieved by passing over the surrounding barriers through another type of stationary point, the saddle point, which in chemical systems is also known as the transition state. To distinguish between minima and saddle points, where the first derivative is zero for both, the second derivative is taken, by calculation of the Hessian

matrix, where negative eigenvalues are transition states and zero eigenvalues are minima.

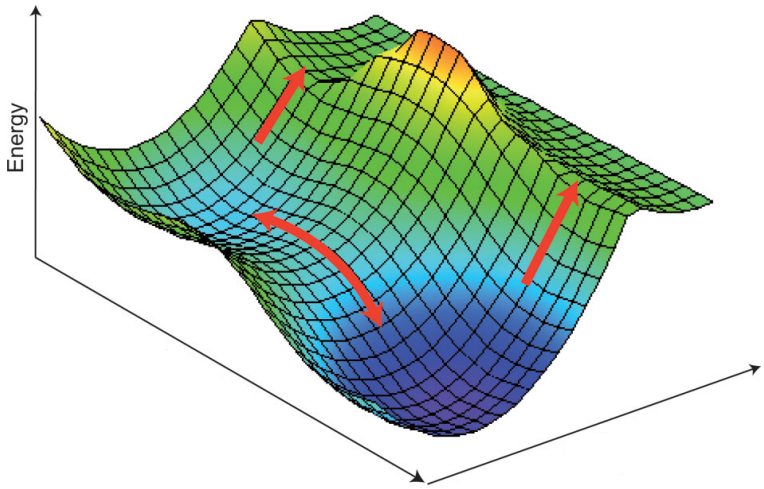


Figure 2.9.1.1 An example of a potential energy surface, taken from reference 31.

Optimisation of chemical systems aims to find these minima and in some cases the transition states, this is achieved through optimisation algorithms and one of the most common and simplest methods to include second derivatives is the Newton-Raphson method. This approach writes a function, f(x), as a Taylor expansion around the point x_0, where the first derivative can be written in the form of equation 2.51.

$$f'(x) = xf'(x_0) + (x - x_0)f''(x_0) \qquad \text{(eq. 2.51)}$$

If the function, f(x), is purely quadratic the second derivative is equal everywhere and so the second derivative of x is equal to the second derivative of x_0. So at the minimum where the first derivative of f(x) equals zero, x can be calculated by equation 2.52.

$$x = x_0 - \frac{f'(x_0)}{f''(x_0)} \qquad \text{(eq. 2.52)}$$

If x_0 is a hypothetical solution to f´(x)=0, then x in equation 2.52 is a new solution, this shows the equation is used to provide an iterative solution. However this operation can be time-consuming as for each step the Hessian matrix must be calculated and inverted. Variants on this methods have been used which aim to reduce computational effort by only requiring first derivatives and gradually constructing the inverted Hessian matrix in successive iterations, these methods a termed quasi-Newton methods.

Combination of the Newton-Rhapson method with Pulay's direct inversion of iterative subspace, DIIS, which is used to aid convergence within self consistent field reactions has yielded a method termed geometry DIIS, GDIIS, which has shown to be more efficient in the optimisation of large molecules.[32] A further improvement to this method has also been presented by Li and Frisch[33] which uses an energy-represented DIIS, this is termed GEDIIS and has shown improvements over the other methods here.

2.9.2 Solvent Models

To carry out accurate simulations on molecules or chemical systems the effect of the solvent present should be accounted for. This can be done in two ways, the solvent can be accounted for explicitly, but in these cases the addition of the necessary amount of extra molecules can render a calculation impractical. The other approach is to account for the effect of the solvent implicitly by describing it as a uniform polarisable medium with a dielectric constant, ε, with a cavity in which the solute or molecule of interest is placed, this method is termed the self consistent reaction field, SCRF. Interaction with the solvent results in a perturbation of the molecule, which in turn perturbs the reaction field, this can be accounted for in the Hamiltonian.

$$H_{total} = H_0 - \mu^T \frac{2(\varepsilon - 1)}{(2\varepsilon + 1)\alpha^3} < \Psi|\mu|\Psi > \qquad \text{(eq. 2.53)}$$

In equation 2.53 H_0 is the unperturbed Hamiltonian, μ is the dipole moment, and α is the radius of the cavity. In this approach there is an energy penalty that needs to be accounted for which arises from formation of the cavity, ΔG_{cavity}, however along with this there is some stabilisation resulting from electrostatic and van der Waals interactions between the solute and solvent, ΔG_{elec} and ΔG_{disp} respectively. In this way the free energy change of solvation can be written as equation 2.54.

$$\Delta G_{solvation} = \Delta G_{cavity} + \Delta G_{elec} + \Delta G_{disp} \qquad \text{(eq. 2.54)}$$

Two solvent models have been used throughout this work, the polarisable continuum model, PCM,[34] which uses a van der Waals surface to model the cavity, and the conductor like screening model, COSMO.[35] COSMO differs from the dielectric medium methods in that the cavity is embedded in a conductor with infinite dielectric constant.

The advantage of this is that screening effects for a conductor are easier to handle and small corrections to results of the conductor can give appropriate values for solvents with high dielectric constants, such as water. Another advantage is that the potential on the surface of the conductor from the solvent and solute charges is set to zero which results in convenient boundary conditions for determining surface charges.

2.9.3 Structural Analysis of DNA

To analyse the structures of DNA helices that have been studied throughout this work, base pair parameters have been utilised. These show a combination of rotational and translational parameters that can be used to distinguish geometry around a base pair step and allow effective comparison between different structures. Figure 2.9.3.1 exhibits the six parameters considered, alongside a reference coordinate frame. The shift, rise, and slide parameters are translational and the roll, tilt, and twist are rotational. This form of analysis was carried out using the x3DNA software.[36]

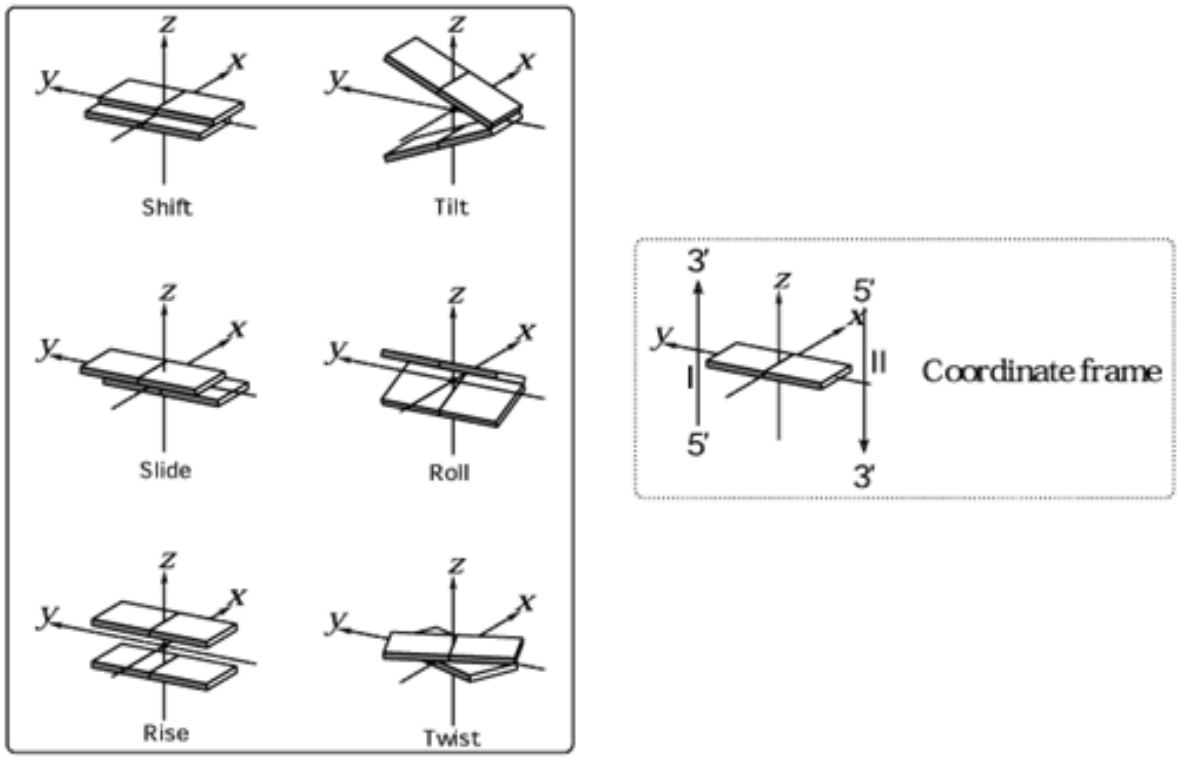


Figure 2.9.3.1 Base pair parameters considered in this work.[37]

2.9.4 Programs and Methods

All DFT calculations have been performed using Gaussian 03,[38] or Gaussian 09.[39] Unless specified the basis set 6-31+G** has been used and where required the Stuttgart-Dresden ECP, SDD,[40] has been used for transition metals. All ONIOM calculations have been carried out using Gaussian 09, with the AMBER force field. Where needed the PCM solvent model has been used.

All DFT-D calculations have been carried out using turbomole v5.10,[41] exclusively using the B97-D[42] functional with the def2-TZVP[43] basis set and the def2-ECP pseudopotential[40] where required. Unless specified B97-D was carried out using the D2 correction. Density fitting in the form of resolution of identity, (ri) has been utilised. D3 dispersion corrections were calculated using the DFT-D3 program.[22] Where needed the COSMO solvent model has been used.

All local MP2 and local CCSD calculations have been carried out using the Molpro suite of programs[44] using aug-cc-pVTZ basis sets for heavy atoms and cc-pVTZ for hydrogen.[45] The density fitting approximation has been employed for all calculations, where the aug-cc-pVTZ fitting basis set has been used for HF and MP2 where available.[46,47]

Semi-empirical calculations using the PM6[48] and PM6-DH2[49] models were carried out using the MOPAC software.[50] Where needed the COSMO solvent model has been used.

2.10 References

1. Jensen, F. "Introduction to Computational Chemistry", 2nd edition, Wiley (2007)
2. Cramer, C.J. "Essentials of Computational Chemistry", 2nd edition, Wiley (1997)
3. Leach, A.R. "Molecular Modelling: Principles and Applications", 2nd edition, Pearson (2005)
4. Hinchchliffe, A. "Molecular Modelling for Beginners", 2nd edition, Wiley (2008)
5. Bader, R.F. "Atoms in Molecules: A Quantum Theory", Oxford University Press (2003)
6. Popelier, P.L. "Atoms in Molecules: An introduction", Prentice Hall (2000)
7. Hehre, W.J.; Stewart, R.F.; Pople, J.A. *J. Chem. Phys.* **1969**, 51, 2657-2664.
8. Ditchfield, R.; Hehre, W.J.; Pople, J.A. *J. Chem. Phys.* **1971**, 54, 724.
9. Boys, S. F.; Bernardi, F. *Mol Phys* **1970**, 19, 553-.
10. Pulay, P. *Chem. Phys. Lett.* **1983**, 100, 151-.
11. Vahtras, O.; Almlöf, J.; Feyereisen, M.W. *Chem. Phys. Lett.* **1993**, 213, 514-.
12. Polly, R.; Werner, H.-J.; Manby, F.R.; Knowles, P.J. *Mol. Phys.* **2004**, 102, 2311
13. Werner. H.-J.; Manby, F.R.; Knowles, P.J. *J. Chem. Phys.* **2003**, 118, 8149-8160.
14. Becke. A.D. *Phys. Rev. A* **1988**, 37, 785-789.
15. Becke A.D. *J.Chem.Phys.* **1993**, 98, 5648.
16. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
17. Becke A.D. *J.Chem.Phys.* **1993**, 98, 1372.
18. Waller M.P.; Robertazzi, A.; Platts, J.A.; Hibbs, D.E.; Williams, P.A. *J. Comput. Chem.* **2006**, 27, 491.
19. Zhao, Y.; Truhlar, D.G. *Theor. Chem. Acc.* **2008**, 120, 215.
20. Zhao, Y.; Truhlar, D.G. *Chem. Theory. Comput.* **2008**, 4, 1849.
21. Grimme, S. *J Comput Chem* **2004**, *25*, 1463-1473.
22. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H.; *J.Chem.Phys.* **2010**, 132, 154104.
23. Warshel, A.; Levitt, M. *J.Mol.Biol.* **1976**, 103, 227-249.
24. Cornell, W.D. et al. *J. Am. Chem. Soc.* **1995**, 117, 5179-5197.
25. Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, 16, 1170-1179.
26. Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. J. *Phys. Chem.* **1996**, 100, 19357-19363.
27. Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct.* **1999**, 461, 1-21.
28. 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.
29. Bader, R. F. W. *Pure Appl. Chem.* **1988**, *60*, 145-155.
30. Koch, U.; Popelier, P.L.A. *J. Phys. Chem.* **1995**, 99, 9747-9754.
31. Glowacki, D.R.; Harvey, J.N.; Mulholland, A.J. *Nature Chemistry* **2012**, 4, 169-176.
32. Csaszar, P.; Pulay, P. *J. Mol. Struct.* **1984**, 114, 31-.
33. Li, X.; Frisch, M.J. *J. Chem. Theory Comput.* **2006**, *2*, 835–839.
34. Miertus, S.; Scrocc, E.; Tomasi, J. *Chem. Phys.* **1981**, 55, 117.
35. Klamt, A.; Schüümann, G. *J. Chem. Soc. Perkin Trans. 2* **1993**, 799-805.
36. Lu, X.-J.; Olson, W.K. *Nucleic Acids Res.* **2003**, 31(17), 5108-21.
37. 3dna.rutgers.edu
38. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao,

O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, **2004**.

39. Gaussian 09, Revision A.1, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.
40. Andrae, A.; Hausserman, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor.Chim. Acta* **1990**, *77*, 123–141
41. Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem Phys Lett* **1989**, *162*, 165-169.
42. Grimme, S. *J Comput Chem* **2006**, *27*, 1787-1799.
43. Weigend, F.; Ahlrichs, R. *Phys.Chem.Chem.Phys.* **2005**, 7, 3297.
44. Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Mitrushenkov, A.; Rauhut, G.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hetzer, G.; Hrenar, T.; Knizia, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, M. E. M. W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pflüger, K.; Pitzer, R.; Reiher, M.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. MOLPRO, version 2008.1, a package of ab initio programs, see http://www.molpro.net.
45. Kendall, R.A.; Dunning Jr, T.H.; Harrison, R.J. *J. Chem. Phys.*, **1992**, 96, 6796-806.
46. Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, 116, 3175
47. Hill, J.G.; Platts, J.A. *J. Chem. Phys.* **2008**, 128, 044104.
48. Stewart, J.J.P. *J. Mol. Model.* **2007**, 13, 1173-213.
49. Korth, M.; Pitonák, M.; Rezác, J.; Hobza, P. *J. Chem. Theory Comp.* **2010**, 6, 344-352.
50. Stewart, J.J.P. *J. Computer-Aided. Mol Design.* **1990**, 4, 1-103.

Chapter 3: Non-Covalent Interactions of Transition Metal Complexes

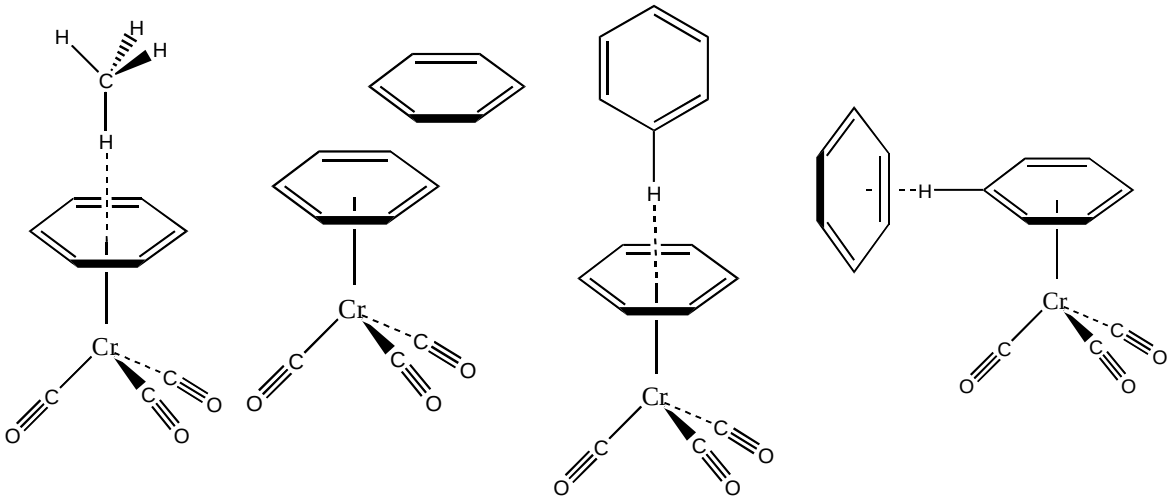
3.1 Introduction

The work in this chapter aims at testing the suitability of two DFT functionals, BHandH and B97-D, for calculating non-covalent interactions of transition metal complexes containing coordinated aromatic systems. Both methods show potential as BHandH has been used to calculate stacking interactions of benzenes, pyridines, and nucleobases[1] and B97-D has been constructed with an empirical dispersion correction to account for non-covalent interactions.[2] Model systems have been studied to validate methods which have then been used on ruthenium arene 'piano-stool' anti-cancer complexes, to study interactions in 'real' biological systems. The content in sections 3.2.1 and 3.2.2 comes from work within our group, published in references 3 and 4 respectively. All the work presented here from these publications was carried out by myself, with the exception of the CBS(T) calculations in section 3.2.1, which were carried out by my supervisor Dr Jamie Platts.

3.2 Interactions within Intermolecular Complexes

3.2.1 Non-Covalent Interactions of Ferrocene and $\eta^6(C_6H_6)Cr(CO)_3$

Transition metal complexes that contain coordinated aromatic rings have the potential to form non-covalent interactions in the same way benzene or analogous molecules do. Directly coordinated arenes can form CH...π interactions with methane and π...π stacking interactions with benzene. These interactions have been probed with two test systems, ferrocene, which contains iron with two cyclopentadienyl ligands coordinated in the η^5 fashion and $(C_6H_6)Cr(CO)_3$ which contains chromium with benzene coordinated in the η^6 fashion and three carbonyl ligands. These were chosen as a mix of sandwich and half-sandwich neutral complexes containing two types of coordinated aromatic ring. Four interaction types have been examined with these two complexes: methane, with a CH pointing toward the aromatic ring; and three benzene formations, parallel displaced (PD), T-shaped (T), and inverted-T as shown in figure 3.2.1.1.



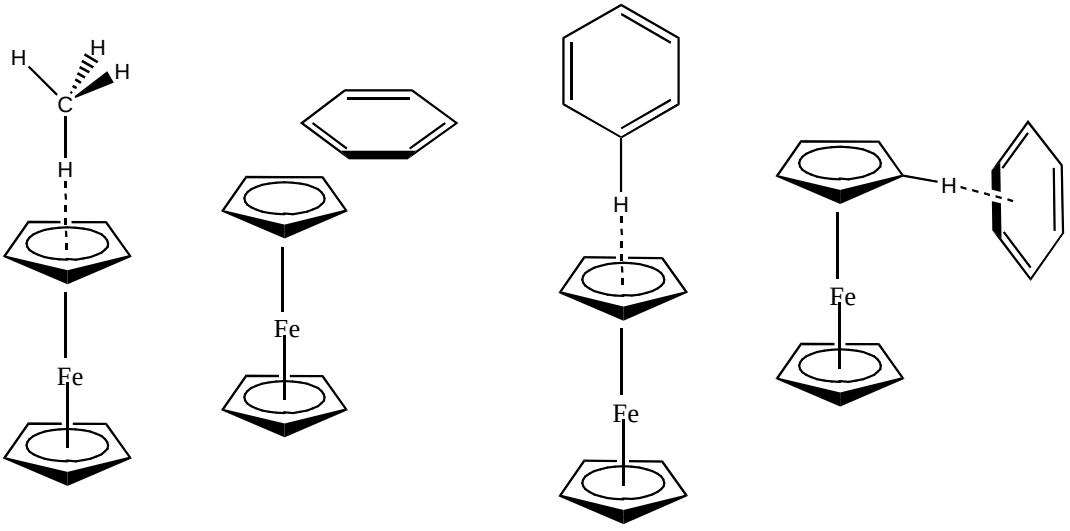


Figure 3.2.1.1 Schematic representation of intermolecular complexes of methane and benzene with ferrocene and $\eta^6(C_6H_6)Cr(CO)_3$.

Ferrocene and $\eta^6(C_6H_6)Cr(CO)_3$ were constructed and initially optimised using BP86/ 6-31+G**, which has shown to be a suitable functional for optimisation of systems containing transition metals.[5] Both structures confirm the validity of this, as comparisons against experimental crystal structure[6] and gas phase electron diffraction data[7] are in excellent agreement, as shown in table 3.2.1.1.

Table 3.2.1.1 Comparison of experimental and BP86 optimised geometries for $\eta^6(C_6H_6)Cr(CO)_3$ and ferrocene (Å and °).

	Expt [a]	DFT
$\eta^6(C_6H_6)Cr(CO)_3$		
Cr-CO	1.834 (4)	1.839
C-O	1.173 (9)	1.174
Cr-CH	2.233 (9)	2.229
C-C	1.414 (9)	1.426
Ferrocene		
Fe-C	2.054 (3)	2.047
C-C	1.435 (2)	1.441
C-H	1.080 (7)	1.091
mean-C-H[b]	3.7 (9)	1.35

[a]Estimated standard deviations given in parenthesis.

[b]Angle between C-H and mean plane C.

The intermolecular systems shown in figure 3.2.1.1, were constructed from the optimised transition metal complexes and methane, with a CH bond of methane pointing to the centre of the coordinated ring and benzene in three formations, PD, T, and inverted-T. Starting distances and angles of the benzene molecules are based on the high level optimised equilibrium values of the benzene dimer as reported in reference 8. Similarly the methane system is based on values for the methane...benzene system reported in reference 9. The functional of interest to be used here is that of BHandH, which has been shown to be effective in calculating interaction energies of the uncoordinated analogues of the systems shown in figure 3.2.1.1.[1] To check its suitability rigid scans were carried out on the constructed intermolecular systems, by scanning along the distance between the centroid of the ring and the proton pointing towards it, in intervals of 0.1 Å, in the cases of methane, T, and inverted-T. For PD the vertical distance between the two rings was the coordinate that was scanned. Along with the BHandH scans an *ab initio* method, DF-LMP2, was also used to act as a benchmark. As a result of these scans equilibrium distances and interaction energies at these distances are reported in table 3.2.1.2 and the scan for the ferrocene methane system is shown in figure 3.2.1.2.

Table 3.2.1.2 Equilibrium distances and interaction energies of $\eta^6(C_6H_6)Cr(CO)_3$ and ferrocene systems at equilibrium distances (Å and kcal/mol).

	BHandH		DF-LMP2	
	r_{eq}	E	r_{eq}	E
$\eta^6(C_6H_6)Cr(CO)_3$				
Methane	2.60	-1.60	2.80	-1.38
T	2.50	-2.06	2.50	-2.53
PD	3.40	-3.00	3.30	-5.80
Inverted-T	4.80	-6.27	4.80	-6.55
Ferrocene				
Methane	2.50	-1.97	2.80	-1.44
T	2.40	-3.02	2.50	-2.97
PD	3.40	-1.67	3.30	-4.14
Inverted-T	4.70	-3.23	4.70	-4.70

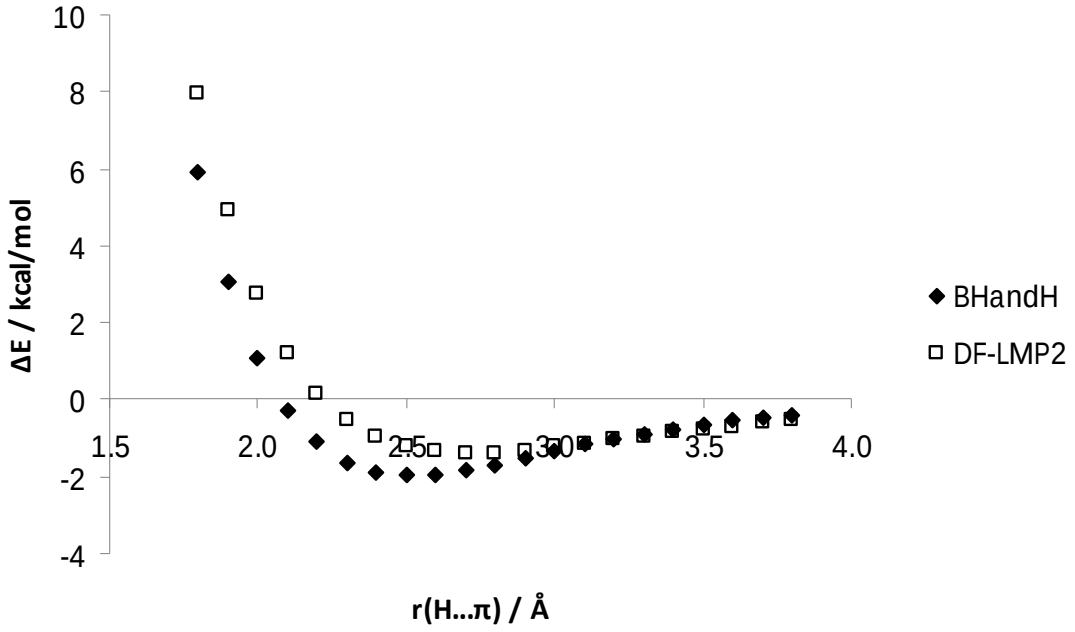


Figure 3.2.1.2 Scan profile of the Methane ferrocene system using BHandH and DF-LMP2.

The data in table 3.2.1.2 shows that the minima of the scans for BHandH and DF-LMP2 have similar equilibrium distances, with the exception of the methane complexes the distances differ by no more than 0.1 Å. The scan in figure 3.2.1.2 shows that even for the methane ferrocene system, which from the equilibrium distances has the greatest difference between BHandH and DF-LMP2, the profiles follow a similar shape, with BHandH slightly underestimating the minima and overestimating the interaction energy when compared to the DF-LMP2 values. This indicates that, in general, BHandH is as successful at locating minimum energy geometry as MP2, which means that it should be suitable at carrying out full optimisation of these intermolecular complexes.

These systems were then fully optimised without symmetry restrictions using BHandH/6-31+G** and binding energies of the metal complex methane/benzene interactions were calculated at this level, with basis set superposition (BSSE) accounted for with the counterpoise method. To supplement these values, binding energies were also calculated with another DFT functional, B97-D, which contains an empirical dispersion correction to account for electron correlation and has been used extensively throughout this work, as well as two *ab initio* techniques as benchmarks, DF-LMP2 and CBS(T). CBS(T) is considered the current state of the art method for calculating non-covalent interactions, this entails extrapolating to the basis set limit with a method such as MP2 and then correcting with a coupled cluster method, CCSD(T), with a smaller basis set, in this case cc-pVDZ.[3,10-12] Rigid scans to check the accuracy of B97-D for optimisation were not carried out as this functional was tested against BHandH and experimental data in section 5.2.1 and tested against BHandH for these complexes below. All binding energy calculations were carried out at the geometry of the BHandH optimised structures and results are reported in table 3.2.1.3.

Table 3.2.1.3 Binding energies of $\eta^6(C_6H_6)Cr(CO)_3$ and ferrocene complex…systems calculated using DFT, MP2, and CBS(T) (kcal/mol).

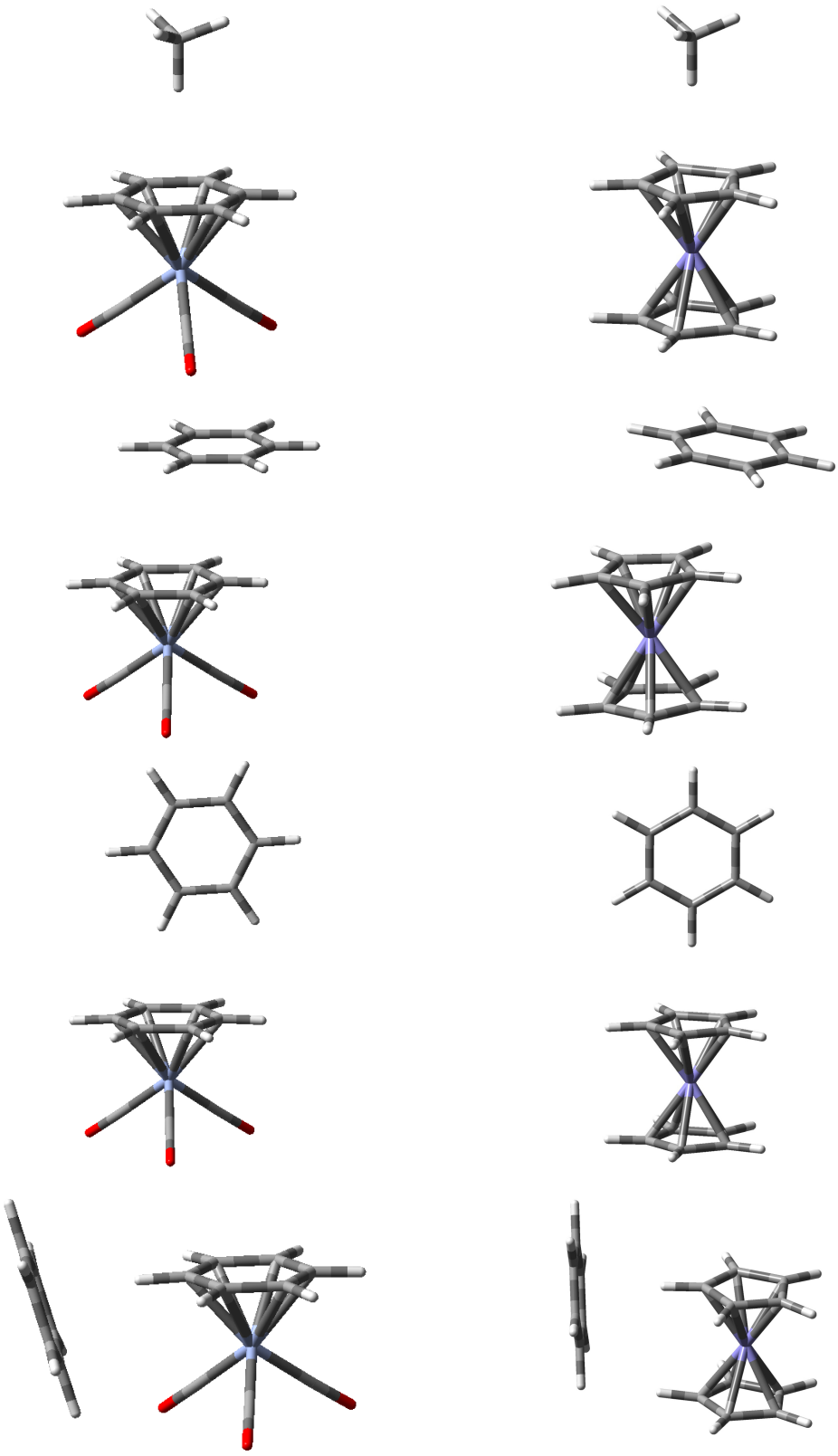
	BHandH	B97-D	DF-LMP2	CBS(T)[3]
$\eta^6(C_6H_6)Cr(CO)_3$				
Methane	-1.29	-1.25	-1.35	-1.27
T	-2.06	-2.45	-2.53	-2.38
PD	-2.92	-2.68	-5.62	-4.22
Inverted-T	-6.12	-5.40	-6.49	-5.85
Ferrocene				
Methane	-1.69	-1.42	-1.44	-1.47
T	-2.95	-2.70	-2.97	-2.86
PD	-1.60	-2.77	-3.99	-3.00
Inverted-T	-3.17	-4.03	-3.94	-3.57

For the chromium complexes all four methods give the same ranking, where inv-T > PD > T > methane. Values for the methane complex are in good agreement with no difference greater than 0.1 kcal/mol. BHandH energies for the T complex are slightly underestimated when compared to the CBS(T) values, which is to be considered the more accurate of the benchmark methods, whilst the B97-D value is slightly overestimated. The inv-T complex gives the opposite result where BHandH gives a stronger binding energy than B97-D, when compared to CBS(T) energies, this system also has significantly higher values than the corresponding T system, this is most likely due to additional interactions formed between the oxygen of carbonyl ligands and the protons on the free benzene. The most striking differences between DFT and *ab initio* are for the PD complexes, where the DFT methods give energies approximately 1.5 kcal/mol weaker than the CBS(T) values, DF-LMP2 significantly overestimates here, but this is not unexpected for this method.[13,14] However overall both DFT functionals give reasonable binding energies compared to benchmarks for this metal complex.

The ferrocene complexes aren't as conclusive in their energy ordering between methods, conversely though the energy differences between the different conformations are smaller, as shown by the small differences between the T and PD systems for CBS(T) and B97-D. Between the two functionals, B97-D generally gives values more comparable to the benchmarks, with the exception of the T complex, where BHandH performs better. The BHandH energy for the PD complex is the most anomalous result

and is significantly underestimated when compared to CBS(T), with the exception of this result and the slight differences in order both functionals still give reasonable binding energies.

As all the binding energies were calculated at the BHandH/6-31+G** optimised structures, the same systems have also been optimised using the B97-D functional to check the validity of the geometries yielded by the former method. Figure 3.2.1.3 shows the optimised structures at the BHandH level for both complexes, at a visual comparison between the two functionals significant differences cannot be discerned, hence only one set of structures are reported. To further compare the structures selected geometrical parameters have been measured and are reported in table 3.2.1.4.



$\eta^6(C_6H_6)Cr(CO)_3$ Ferrocene

Figure 3.2.1.3 Optimised structures of $\eta^6(C_6H_6)Cr(CO)_3$ and ferrocene using BHandH/6-31+G**.

Table 3.2.1.4 Selected geometrical parameters of optimised $\eta^6(C_6H_6)Cr(CO)_3$ and ferrocene complexes calculated using BHandH and B97-D (Å and °).

Complex	Method	Distance [a]	Angle between planes	Centroid...H—C
$\eta^6(C_6H_6)Cr(CO)_3$				
CH_4	BHandH	2.50		179.86
	B97-D	2.64		179.15
PD	BHandH	3.59	0.14	
	B97-D	3.80	0.77	
T	BHandH	2.49	89.35	150.10
	B97-D	2.42	79.91	157.78
inv-T	BHandH	2.29	73.90	166.40
	B97-D	2.34	74.27	163.42
Ferrocene				
CH_4	BHandH	2.53		179.20
	B97-D	2.61		177.88
PD	BHandH	3.80	9.24	
	B97-D	3.87	8.74	
T	BHandH	2.37	89.84	179.15
	B97-D	2.45	89.80	179.24
inv-T	BHandH	2.41	84.06	173.08
	B97-D	2.44	83.22	172.17

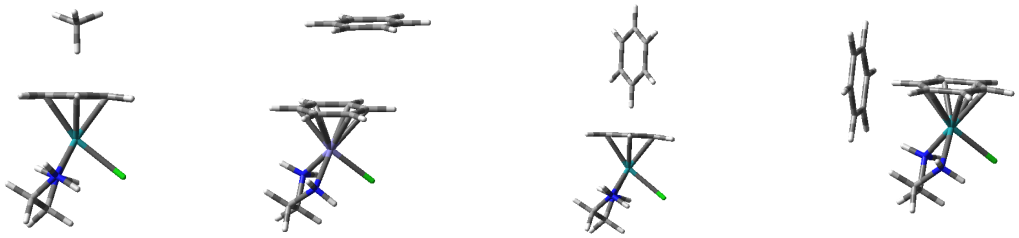
[a]Defined as distance from donor hydrogen to acceptor centroid, or centroid...centroid distance in PD, *i.e.* the distance between the centroid of the two parallel rings.

For the chromium complexes there are very small differences between the donor hydrogen/centroid to centroid distances, no larger than 0.21 Å. For the methane, PD, and inv-T systems there are insignificant differences in the reported angles. The only system that shows any real discernible difference is the T complex, where BHandH gives a structure that the two rings are perpendicular, whilst in the B97-D structure the benzene tilts slightly off the ideal by approximately 10 °, although even with this difference both methods are in good agreement. The ferrocene complexes are generally in closer agreement than the chromium complexes, with differences in the distance parameter no greater than 0.08 Å and insignificant differences in the angle

measurements. Overall, for both complexes, BHandH and B97-D give structures that are comparable so the binding energy comparison above is suitable.

3.2.2 Non-Covalent Interactions of $[\eta^6(C_6H_6)Ru(en)Cl]^+$

The transition metal complex $[\eta^6(C_6H_6)Ru(en)Cl]^+$, is based around Sadler's ruthenium arene anti-cancer drugs, containing benzene coordinated in the η^6 fashion. It is possible for it to form non-covalent interactions within intermolecular complexes including methane and benzene, specifically CH…π interactions and π…π stacking, around the coordinated benzene ring. These interactions have been probed with a comparison of the two DFT methods used above, BHandH and B97-D, and one *ab initio* approach, MP2/6-31G(0.25)d, to ascertain the effectiveness of calculating binding energies and finding optimised structures. This MP2(0.25) approach uses a basis set containing more diffuse d functions, with an exponent of 0.25 instead of 0.8, which has shown to give good results for systems containing dispersion based reactions.[15] The CBS(T) method has not been attempted here as the data above shows reasonably good agreement with the DFT methods, therefore it was deemed unnecessary to carry out these calculations on another set of systems. Complexes were constructed with one methane C-H bond pointing towards the coordinated aromatic system, and benzene in three formations, parallel displaced (PD), T-shaped (T), and "inverted" T-shaped, as shown in figure 3.2.2.1. These model systems were optimised using BHandH/6-31+G**/SDD and B97-D/def2-TZVP, with binding energies (tables 3.2.2.1 and 3.2.2.2) and geometrical parameters (table 3.2.2.3) reported below.



Methane PD T Inverted-T

Figure 3.2.2.1 Intermolecular complexes of methane and benzene with $[\eta^6(C_6H_6)Ru(en)Cl]^+$.

Due to the low symmetry of the ruthenium complex, there are several possible starting orientations for geometry optimisations: for instance, six possible inverted-T structures could be envisaged. All were explored, and systems with the lowest energy used for further analysis. During optimisation, a tendency for higher energy structures to flip to lower energy ones was noted, indicating that small barriers separate the different forms. To attain full optimisation in such cases, coordinates were extracted at the point before the flip and used as the starting point of a new optimisation wherein the system optimised to the desired local minimum.

Table 3.2.2.1 Binding energies of ruthenium complex…systems calculated using DFT and MP2 (kcal/mol).

	BHandH	B97-D	MP2(0.25)
Methane	-1.65	-1.53	-0.77
T	-2.43	-2.81	-2.27
PD	-7.04	-7.72	-8.02
Inverted-T	-11.78	-13.02	-11.72

The calculated binding energies of the ruthenium systems are reported in table 3.2.2.1. The energies calculated using all three techniques are in good general agreement, having the same ranking of: inverted-T > PD > T > methane, and values are within *ca.* 1 kcal/mol of one another. Values for the methane complex differ little from that for methane…benzene in the same orientation (literature value of -1.45 kcal/mol calculated using CCSD(T) at the basis set limit[9]). The T-shaped complex has binding energies slightly larger than, although probably within the expected errors of the methods used, a reference value for the benzene dimer in the same orientation (-2.24 kcal/mol using CCSD(T) with a triple ζ basis set[8]).

In contrast, binding energies for the parallel displaced and inverted-T forms are much larger than the analogous values in the benzene dimer. (PD reference = -2.51 kcal/mol, calculated using CCSD(T) with a triple ζ basis set[8]). The calculations in section 3.2.1 show that metal complexation increases stabilization of these forms, with CBS(T) values of -4.22 and -5.85 kcal/mol for PD and inverted-T forms of $Cr(C_6H_6)(CO)_3$ respectively. The values in table 3.2.2.1 are significantly more negative than those reported in table 3.2.1.3, a difference that may be due to the positive charge on the Ru complex rather than the neutral complexes studied above. Much of the increased

stabilization seems to stem from more favourable electrostatic interactions, especially cation...π, in the PD and inverted-T forms. Further to this, the inverted-T system is seen to form additional NH…π interactions with the ethylenediamine ligand, the other systems are unable to form analogous interactions, which could in part explain the increase in binding energy.

To try to better quantify the origins of the extra stability in these Ru-complexes, Hartree-Fock (HF) binding energies as well as the –D dispersion contribution to B97-D binding energy are reported in table 3.2.2.2. The inability of HF to calculate dispersion yields values that give a better insight into electrostatic interactions. In the isolated benzene dimer, HF binding energies are +0.80 and +3.52 kcal/mol for T-shaped and PD forms, respectively. Table 3.2.2.2 demonstrates that the T-shaped form, in which the Ru complex acts as a proton acceptor, is less stabilised by electrostatic interactions than the free benzene dimer. In contrast, the inverted-T form, in which the Ru complex is a proton donor, is actually stable at the HF level. This may be due to the proximity of the acceptor benzene to the metal centre or to formation of N—H...π interactions, possibilities that are explored in more detail below. The parallel displaced form is not bound at the HF level, but the repulsion is 2 kcal/mol smaller than in the free benzene dimer, indicating that this is indeed one source of the extra stability observed in table 3.2.2.1.

Table 3.2.2.2 Hartree-Fock binding energies and dispersion contributions (kcal/mol).

	HF/6-31+G(d,p) [a]	-D [b]
Methane	1.63	-3.35
T	3.07	-4.63
PD	1.38	-9.52
Inverted-T	-1.19	-11.13

[a]Counterpoise corrected.

[b]Dispersion contribution to B97-D binding energy.

Alongside contributions from electrostatics, the large dispersion contributions to binding energy reported in table 3.2.2.2 go some way to explaining the strong binding energy within this system. Values for the methane and T-shaped complexes are close to those for isolated benzene as a proton acceptor (-2.73 and -4.56 kcal/mol, respectively). However, in the parallel-displaced form the dispersion correction is larger than in the

free benzene dimer (-7.15 kcal/mol), such that the very large binding energy of this form is due to enhanced dispersion interactions as well as electrostatics.

Table 3.2.2.3 Selected geometrical parameters of optimised complexes (Å and °).

Complex	Method	Distance [a]	Angle between planes	Centroid...H—C
CH_4	BHandH	2.54		176.59
	B97-D	2.53		176.33
PD	BHandH	3.67	15.65	
	B97-D	3.53	8.39	
T	BHandH	2.49	88.61	148.95
	B97-D	2.53	88.68	176.86
Inverted-T	BHandH	3.73	82.17	149.06
	B97-D	3.66	84.24	126.08

[a]Defined as distance from donor hydrogen to acceptor centroid, or centroid...centroid distance in PD.

Selected geometrical parameters for all complexes are shown table 3.2.2.3. For the methane complex, both functionals give very similar structures with almost identical H…centroid distance, and root mean square deviation over all atoms of just 0.051 Å. Within this RMSD value, the difference between DFT optimized geometry of Ru complexes accounts for 0.050 Å, showing both functionals give very similar positions for the methane molecule. The centroid...H—C angles show that C—H coordinates directly above the centre of the arene. An ideal parallel-displaced complex would have parallel angle planes of benzene rings, but in both cases significant tilt is observed at the local minimum, indicating some movement towards the inverted-T structure. A rather larger tilt and slightly greater centroid...centroid distance is observed with BHandH than with B97-D, but overall agreement is good. Nevertheless, a rather large RMSD of 0.89 Å is observed, since the difference in tilt angle gives rise to differences in position of all atoms.

For the two T-shaped systems a similar pattern of broad similarity between methods is observed. Both functionals give centroid…H distances and angles between planes values that are in close agreement, however the centroid-C-H angle differs, with B97-D giving a near 180 ˚ angle, and BHandH tilting the benzene off centre, as seen previously

for the free benzene dimer.[16,17] Deviation from the ideal perpendicular orientation is again observed, an effect that is rather greater with BHandH than for B97-D. Both methods show relatively close N—H…π contacts in the optimized inverted-T complexes, again resulting in changes from the idealised form of perpendicular benzene rings. B97-D results in rather closer N-H…π contacts, such that the free benzene ring moves further from the coordinated ring compared to BHandH, this accounting for the rather large RMSD between methods of 0.59 Å.

The largest difference in the DFT binding energies and structures is found in the inverted-T system. To further probe this difference, QTAIM analysis was carried out on these complexes, with the results shown in figure 3.2.2.2. Using both methods, interactions with the ethylenediamine ligands as well with the coordinated arene are present in the minimum energy geometry: in the BHandH case, one arene...benzene and two ethylenediamine...benzene C—H...π are located, whereas in the B97-D case one N—H…π as well as two arene...benzene and two ethylenediamine...benzene C—H...π are present. Thus, the increased binding energy and less ideal geometry found with B97-D appears to be due to formation of more ethylenediamine...benzene contacts rather than any fundamental difference in description of interactions between functionals.

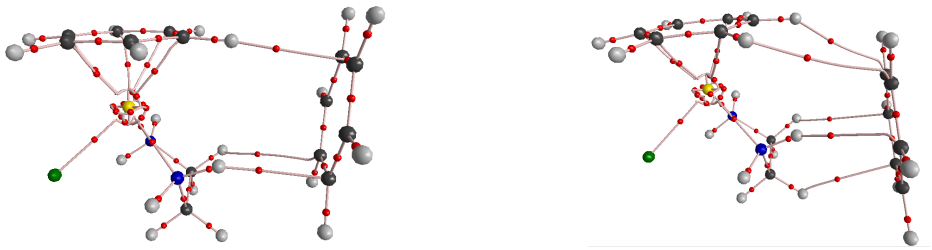


Figure 3.2.2.2 Molecular graphs of inverted-T complex calculating using BHandH (left) and B97-D (right). Red dots indicate (3, -1) or bond critical points: a cluster of such points is found around Ru due to use of ECP, but intermolecular contacts are unaffected.

3.3 Interactions within Intramolecular Complexes

The anticancer properties of these "piano stool" ruthenium arene complexes have been well documented and several different extended aromatic and partially saturated arene ligands have been employed in place of benzene.[18-22] The interactions of a series, shown in figure 3.3.1, consisting of benzene (bz), cymene (cym), biphenyl (biph), anthracene (anth), dihydroanthracene (dha), and tetrahydroanthracene (tha) have been extensively studied with DNA in chapter 4, using the B97-D functional, with a D2 type correction. To confirm the validity of this functional compared to a wide range of other methods, as well the same functional employing a newer empirical dispersion correction, tests have been carried out on the different ruthenium arene systems coordinated to guanine and adenine, through the loss of the chloride ligand. An example of these two systems for the benz arene is shown in figure 3.3.2.

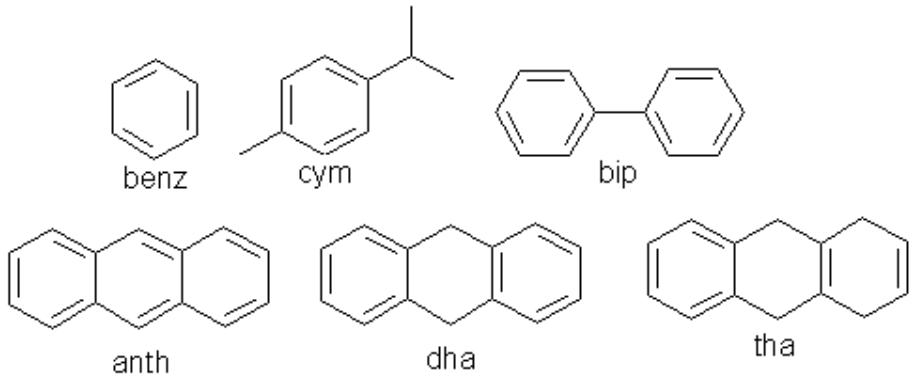
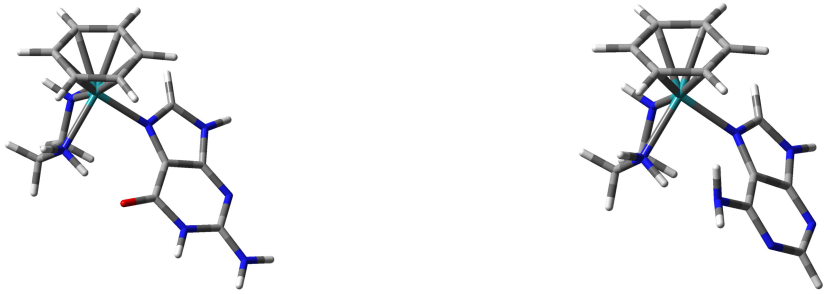


Figure 3.3.1 Structures of the arenes coordinated to ruthenium.



i) ii)

Figure 3.3.2 Structure of ruthenium benz complex coordinated to i) guanine and ii) adenine.

To attempt to increase the accuracy of this dispersion correction a more recent type, termed D3, has been formalised by Grimme.[23] This D3 correction includes r^{-8} terms for close contact interactions as well as the r^{-6} terms for longer range contacts. Ryde *et al.*, have used both DFT-D2 and DFT-D3 to calculate binding energies of a methyl group within a cobalt corrin ring, with the intent of finding whether DFT-D type functionals treat ligand binding to transition metal complexes correctly.[24] Within their system of interest the small ligand results in short range interactions that have large dispersion based contribution to the energy, when using the functional BP86 with the D3 correction; they show that the r^{-8} term becomes dominant. From comparison of binding energies to higher level DF-LMP2, which has also been shown by Hill and Platts[25] to give metal-ligand binding energies with typical errors of less than 1 kcal/mol, and DF-LCCSD they conclude that DFT-D3 provides reasonably accurate estimates. In light of these results, calculations have been carried out for binding energies of the ruthenium arenes to nucleobase systems using B97-D3, to ascertain whether D3 correction shows significant improvement over D2, even though this system does not contain relatively short metal ligand interactions. A wide variety of functionals available in the default library of Gaussian 09 have also been explored in this section, to observe whether there are any other functionals not explored so far that show good results for this application when compared to benchmark *ab initio* values.

3.3.1 Ruthenium Arene Guanine Complexes

Using the B97-D2 optimised structures from the calculations carried out on the ruthenium arene GG/CC complexes in chapter 4, systems that contain just the ruthenium complex and a single guanine, bound to ruthenium through N7 atom, were constructed by removing the additional nucleobases and sugar phosphate backbone. The additional hydrogen needed in the N9 position was added by hand, based on the geometry of an optimised B97-D2 guanine molecule. These structures were used to carry out binding energy calculations between the ruthenium and N7 nitrogen without further optimisation, using a combination of hybrid and pure DFT functionals, double hybrid methods, and post Hartree-Fock methods.

Binding energies for the ruthenium arene complexes bound to guanine have been calculated, *i.e.* the energy of the bond between the ruthenium and N7 atom of guanine, using B97-D2 and BHandH and are reported in table 3.3.1.1 and shown graphically in

figure 3.3.1.1. In addition to these values, the recent revision to the D2 correction, D3, has been applied to the B97 functional for these systems, to give B97-D3 binding energies. The damping function for these dispersion corrections would be expected to have significance on the corrections, but Grimme *et al.* have reported, in a paper examining the effect of damping in DFT-D calculations, that the Becke Johnson (BJ) damping gives negligible difference for the B97-D function on test sets containing non-covalent interactions, compared to the standard damping function, therefore the standard damping function has been used throughout.[26] A method of greater accuracy was required to benchmark these results, DF-LMP2 with the basis set aug-cc-pVTZ, was used for this and binding energies calculated are reported alongside the DFT methods in table 3.3.1.1.

Table 3.3.1.1 Binding energies of ruthenium arene complexes to guanine calculated using B97-D2, B97-D3, BHandH, and DF-LMP2, values in parenthesis are the empirical dispersion corrections (kcal/mol).

	B97-D2	B97-D3	BHandH	DF-LMP2
Benz	-93.37 (-17.07)	-90.32 (-14.02)	-98.13	-95.38
Cym	-89.02 (-20.23)	-82.51 (-16.72)	-92.22	-89.84
Bip	-85.17 (-26.73)	-81.05 (-22.61)	-91.88	-94.80
Anth	-79.30 (-27.69)	-75.35 (-23.75)	-84.46	-87.59
Dha	-91.44 (-31.20)	-86.93 (-26.69)	-96.96	-97.79
Tha	-91.05 (-30.98)	-86.52 (-26.45)	-97.71	-97.02

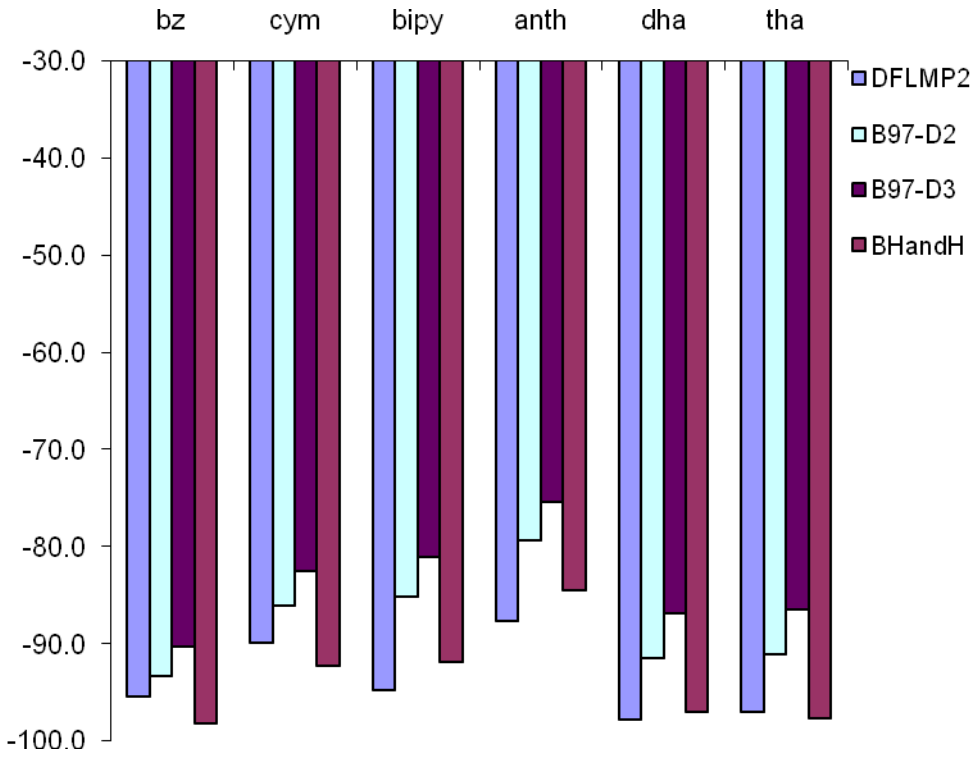


Figure 3.3.1.1 Graph illustrating the difference in binding energies (kcal/mol) for the Ru complexes with guanine calculated using DF-LMP2, BHandH, B97-D2, and B97-D3.

In an attempt to carry out more accurate benchmark calculations, DF-LCCSD calculations were carried out, with the same basis set as the DF-LMP2 calculations, due to the system size and expense of these calculations only two of the complexes were successful. The benzene and dha systems yielded binding energies of –94.40 and –97.19 kcal/mol respectively, which is within 1 kcal/mol of the DF-LMP2 values of –95.38 kcal/mol for benzene and –97.79 kcal/mol for dha. This small error between the chosen benchmark method and the more accurate calculations confirms its suitability for use here.

For the benzene complex B97-D2 and BHandH are within reasonable agreement of the DF-LMP2 values, with differences of no greater than +/- 3 kcal/mol. The high binding energy for this ligand has been noted before and explained due to the electronic nature of the benzene ligand. The cymene complex also has binding energies calculated using these two methods that are within good agreement with the more accurate method. B97-D2 performs very well for this complex with a difference of less than 1 kcal/mol, while BHandH overestimates by approximately 2.5 kcal/mol. For the remaining four ligands

B97-D2 consistently underestimates the binding energies when compared to the DF-LMP2 values, with values between 8 and 10 kcal/mol greater. BHandH however performs much better for these four, with a difference no greater than 3 kcal/mol for anth and bip, and a difference of no greater than 1 kcal/mol for the dha and tha complexes.

The binding energies calculated using B97-D3 are more positive and therefore show a weaker interaction compared to B97-D2. This difference is fairly uniform across all six complexes with a difference of approximately 4 kcal/mol. From the values of the dispersion corrections, shown in parenthesis in table 3.3.1.1, it can be confirmed that D2 corrections give a greater estimation of non-covalent interaction energies. When comparing these values to DF-LMP2 values, B97-D2 slightly underestimates the binding energies, with a further decrease in the binding energy from using the D3 correction, this shows that for these systems being studied the D2 correction is somewhat surprisingly more suitable.

Traditionally many DFT methods fail or under perform at calculating non covalent interactions, B3LYP being an example of this,[1] but instead of applying an empirical correction to account for dispersion, new functionals can be parameterised to account for dispersion. The Minnesota functionals are a family of such methods developed by Trulhar, including M06 and M06-2X.[27] A hybrid version of the B97 functional with dispersion correction has also been formalised, named wB97XD.[28] Binding energies for the ruthenium guanine complexes were calculated using these three methods to compare against the benchmark and the DFT-D values, values are reported in table 3.3.1.2.

Table 3.3.1.2 Binding energies of ruthenium arene complexes to guanine calculated using dispersion corrected DFT functionals, M06, M06-2X, and wB97XD (kcal/mol).

	M06	M06-2X	wB97XD
Benz	-85.56	-84.34	-89.62
Cym	-79.23	-78.78	-83.96
Bip	-78.76	-79.59	-85.13
Anth	-73.69	-76.72	-80.54
Dha	-83.62	-84.54	-89.92
Tha	-84.56	-84.92	-90.51

Both M06 and M06-2X perform to a similar standard with a difference between their calculated binding energies of approximately 1 kcal/mol, with the exception of anth, which has a difference of approximately 3 kcal/mol. They both underestimate binding energies when compared to the benchmark DF-LMP2 values by 10-16 kcal/mol, which is a greater difference compared to the B97-D values. The wB97XD functional performs slightly better than the Minnesota functionals used here, although it does underestimate when compared to the DF-LMP2 values. It calculates similar values to both the B97-D2 and B97-D3 results, as one would expect, but has energies that lie between the two different pure DFT-D methods. Data for the DF-LMP2, Minnesota, and hybrid B97 functionals is shown graphically in figure 3.3.1.2.

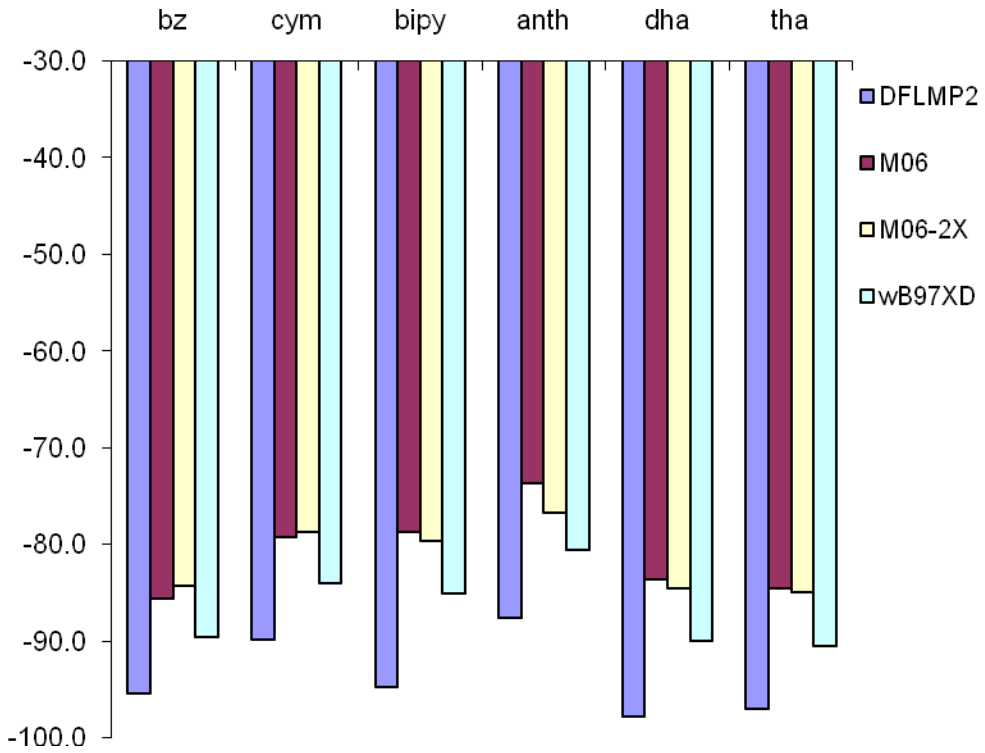


Figure 3.3.1.2 Graph illustrating the difference in binding energies (kcal/mol) for the ruthenium complexes with guanine calculated using DF-LMP2, M06, M06-2X, and wB97XD.

The graphical representations in figures 3.3.1.1 and 3.3.1.2 further show that, for this system, when compared to the DF-LMP2 values the DFT-D functionals performs better than M06 and M06-2X, and within that B97-D2 performs better than B97-D3 and wB97XD. To further examine this system an additional selection of pure and hybrid DFT functionals were used to calculate the ruthenium guanine binding energies and

compared to the benchmark DF-LMP2 values, the pure functionals are reported in table 3.3.1.3 and a graphical representation shown in figure 3.3.1.3. The hybrid functionals are reported in table 3.3.1.4 and graphical representation shown in figure 3.3.1.4.

Table 3.3.1.3 Binding energies of ruthenium arene complexes to guanine calculated using a selection of GGA DFT functionals (kcal/mol).

	PBE	BLYP	TPPS
Benz	-84.49	-75.55	-85.37
Cym	-75.01	-65.40	-74.27
Bip	-69.62	-57.70	-68.27
Anth	-62.78	-51.64	-61.12
Dha	-73.14	-60.17	-71.59
Tha	-72.63	-59.78	-71.45

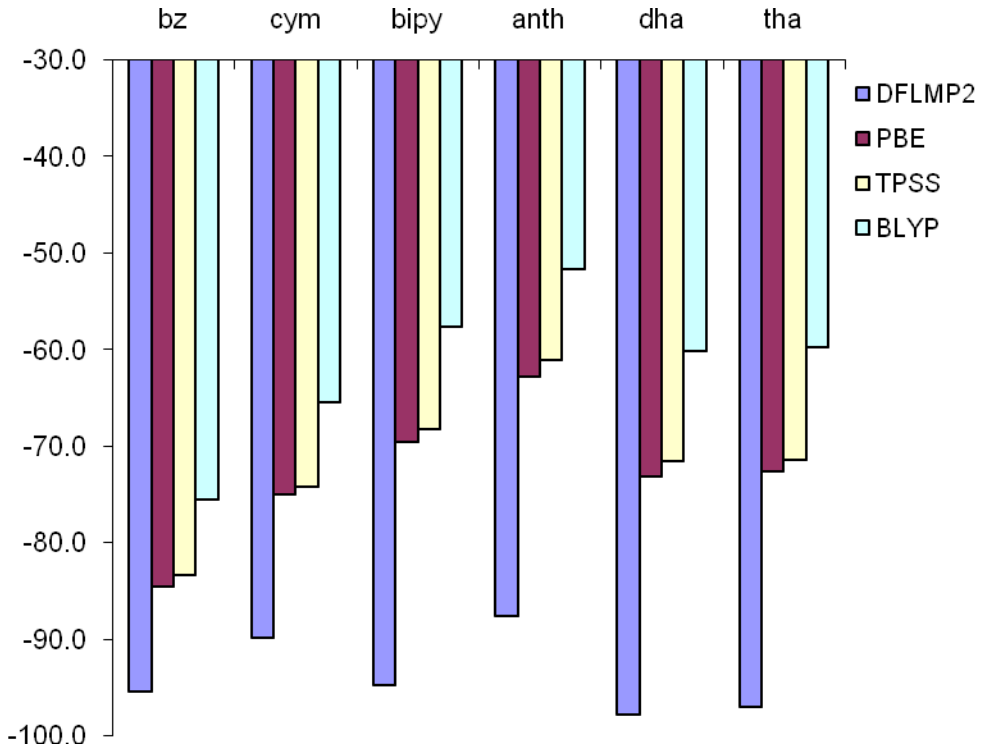


Figure 3.3.1.3 Graph illustrating the difference in binding energies (kcal/mol) for the ruthenium complexes with guanine calculated using DF-LMP2, PBE, BLYP, and TPPS.

The dispersion corrections calculated using D2 and D3, reported in table 3.3.1.1, show that non covalent interactions play a large role in these systems, more so for the bip, anth, dha, and tha ligands. This is expected as the extended aromatic regions or the C-H

groups in the partially saturated systems can form additional stabilising interactions with the guanine. As the three pure DFT functionals do not treat dispersion correctly it is not surprising that they give poor binding energy values when compared to DF-LMP2. The two systems where the pure functionals perform better, benz and cym, have smaller dispersion contributions as the ligands are smaller and have no aromatic regions over the guanine. From the data in table 3.3.1.3 and the graph in figure 3.3.1.3 it is clear that none of these functionals are suitable for the system being studied here. BLYP shows the weakest binding energies compared to DF-LMP2, PBE performs marginally better than TPSS but at its best it is still a difference of greater than 10 kcal/mol from the benchmark values.

Table 3.3.1.4 Binding energies of ruthenium arene complexes to guanine calculated using a selection of hybrid DFT functionals (kcal/mol).

	PBE0	B3LYP	TPSSh	mPW1PW91
Benz	-84.11	-77.83	-83.19	-82.09
Cym	-76.00	-69.03	-74.60	-73.74
Bip	-72.16	-63.30	-69.34	-69.29
Anth	-65.12	-57.04	-62.15	-62.09
Dha	-75.31	-65.90	-72.55	-72.02
Tha	-76.11	-66.87	-73.15	-72.79

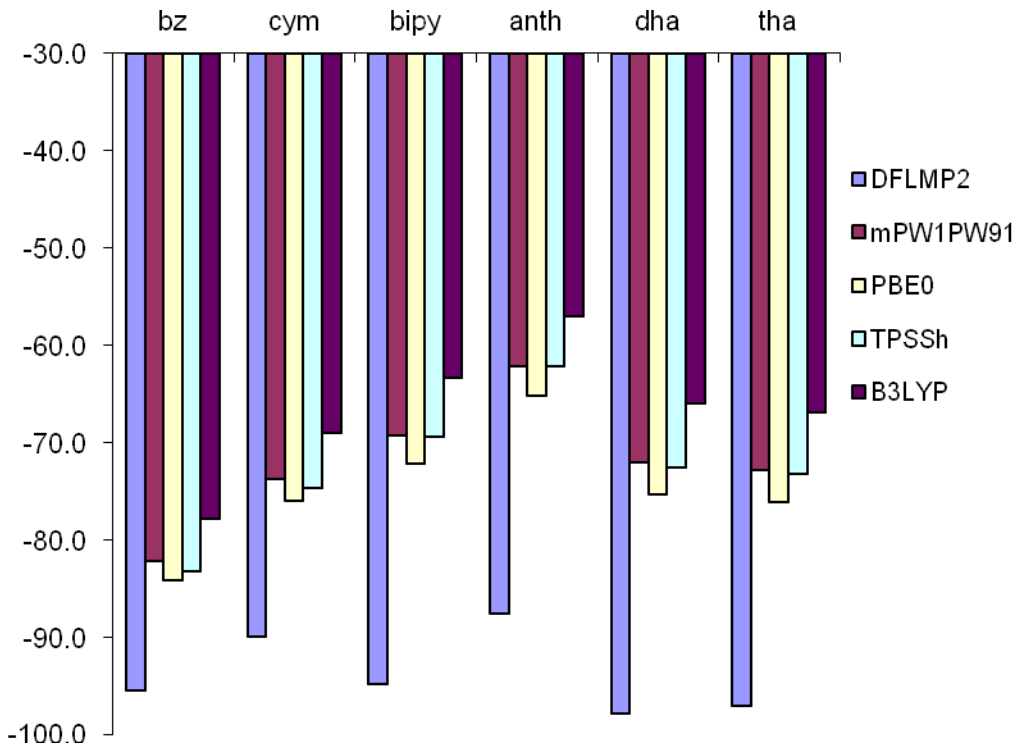


Figure 3.3.1.4 Graph illustrating the difference in binding energies (kcal/mol) for the ruthenium complexes with guanine calculated using DF-LMP2, mPW1PW91, PBE0, TPSSh, and B3LYP.

As mentioned above BHandH performs remarkably well for calculating non-covalent interactions and data from table 3.3.1.1 also confirms this as BHandH binding energies are within a small error of the benchmark DF-LMP2 values. However the other selection of hybrid DFT functionals shown in table 3.3.1.4 do not reproduce the more accurate results to an acceptable level. These functionals traditionally do not treat dispersion correctly so it is not surprising that they underestimate the energies. It is clear from both the data and graph that B3LYP performs worst here, it was also mentioned above that tests on B3LYP have shown it calculates purely dispersion based interactions as unfavourable. TPSSh and mPW1PW91 perform at a similar level, with a difference no greater than 1 kcal/mol, PBE0 performs marginally better but still significantly underestimates the binding energies in the systems where dispersion is more prominent. From observation of figure 3.3.1.4 it is clear that none of the hybrid functionals reproduce the DF-LMP2 binding energies to an acceptable level, producing values too low by a minimum of 10 kcal/mol.

With the acceptable performance of the two most suitable DFT functionals, BHandH and B97-D2, for this system it was then decided to test double hybrid and post Hartree-Fock methods, that could be implemented for a system of the size in question here. MP2 using the same basis set as for the DFT calculations was tested, along with MP2(0.25). A double hybrid method with empirical dispersion correction of the type D2, B2PLYP-D with the same basis set used for the MP2 and DFT calculations, was also tested; binding energies are reported in table 3.3.1.5 and represented graphically in figure 3.3.1.5.

Table 3.3.1.5 Binding energies of ruthenium arene complexes to guanine calculated using MP2, MP2(0.25), and B2PLYP-D (kcal/mol).

	MP2	MP2(0.25)	B2PLYP-D
Benz	-87.61	-92.00	-89.68
Cym	-81.82	-85.82	-83.30
Bip	-83.82	-88.47	-83.02
Anth	-78.46	-83.40	-77.65
Dha	-87.75	-92.61	-88.64
Tha	-87.56	-91.97	-89.19

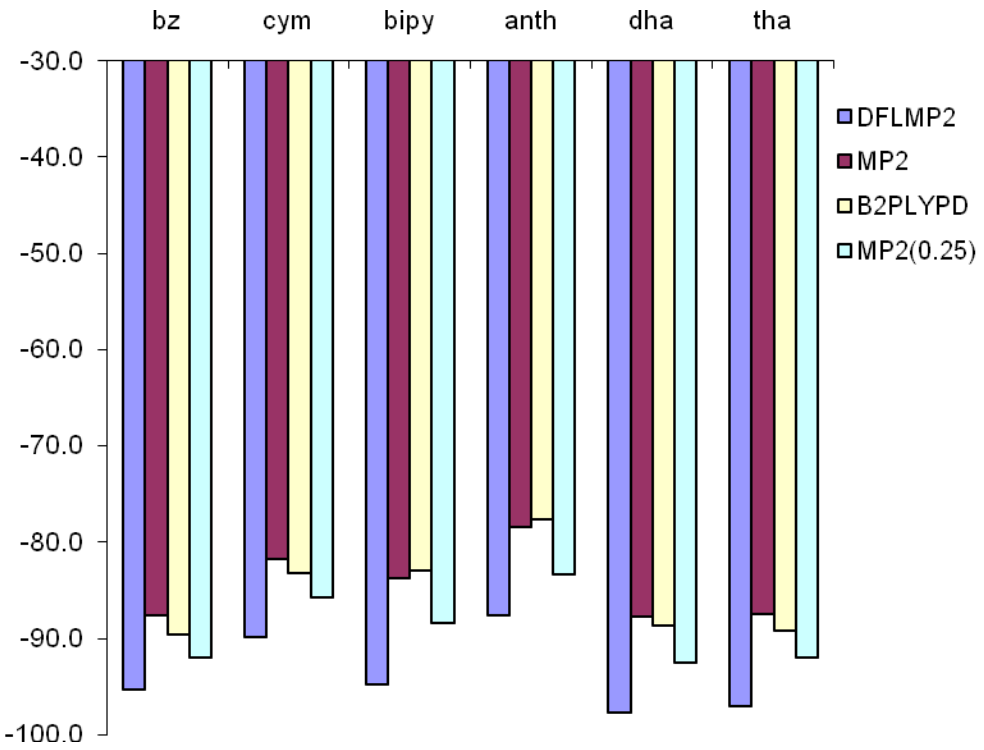


Figure 3.3.1.5 Graph illustrating the difference in binding energies (kcal/mol) for the ruthenium complexes with guanine calculated using DF-LMP2, MP2, MP2(0.25), and B2PLYP-D.

MP2 calculations do take into account dispersion interactions, but in cases have been shown to have errors up to 100%.[13] The binding energies calculated using MP2 with the 6-31+G(d,p) and SDD basis sets, are underestimated when compared to the DF-LMP2 energies but still have an acceptable error, however using MP2(0.25) with the smaller basis set with the more diffuse d functions gives much more favourable binding energies. B2PLYP-D binding energies are of a similar accuracy to the MP2 values, which can be seen in figure 3.3.1.5. Observation of figure 3.3.1.5 also confirms that MP2(0.25) is the best method of these for calculations involving the binding of ruthenium guanine complexes.

Table 3.3.1.6 The mean unsigned error (MUE) in kcal/mol, Spearman correlation rank, and Pearson correlation of all methods used for the ruthenium arene guanine systems against the benchmark DF-LMP2 method.

	MUE	R^2 Spearman Rank	R^2 Pearson Correlation
MP2	9.23	0.99	0.91
MP2(0.25)	4.69	0.94	0.94
B2PLYPD	8.49	0.71	0.78
BHandH	2.12	0.71	0.77
B97-D2	6.01	0.77	0.72
B97-D3	9.96	0.77	0.63
wB97XD	7.12	0.94	0.87
M06	12.83	0.71	0.73
M06-2X	12.25	0.94	0.83
PBE0	18.94	0.49	0.34
TPSSh	21.24	0.31	0.26
B3LYP	27.07	0.31	0.20
mPW1PW91	21.73	0.31	0.28
PBE	20.79	0.37	0.23
TPSS	22.06	0.37	0.21
BLYP	32.03	0.37	0.10

Table 3.3.1.6 shows error values of all the methods used above compared to the benchmark DF-LMP2 values. These include the R^2 Pearson correlation, which is a measure of the linear dependence of two variables; the R^2 Spearman rank correlation, which is a non-parametric version of the Pearson correlation; and the mean unsigned error (MUE), which is the average of the magnitude of the difference of the method with the benchmark. From analysis of the MUE values the best method is BHandH, with a MUE of 2.12 kcal/mol, then MP2(0.25), with a MUE of 4.69 kcal/mol. The B97-D functionals also give favourable MUE values, with B97-D2, wB97XD, and B97-D3 giving errors of 6.01, 7.12, and 9.96 kcal/mol respectively. The double hybrid method, B2PLYPD, and the two post Hartree-Fock methods also give reasonable errors with MUE values less than 10 kcal/mol. The R^2 values for the MP2 and MP2(0.25) are both very close to a value of one, which is indicative of perfect correlation, this would be expected as they are both very similar calculations to those carried out for the benchmark DF-LMP2 method. The other methods that have favourable MUEs, have R^2

values that range 0.63 to 0.94, indicating that the energy ranking of several of the methods is different for two complexes, as can be seen in the figures above.

The two Minnesota functionals perform moderately well, with relatively low MUEs and good R^2 values, but under perform compared to the aforementioned DFT functionals and the post Hartree-Fock methods. The other hybrid DFT functionals perform significantly poorer, with MUE values ranging from 18.94 kcal/mol for PBE0 to 27.07 kcal/mol for B3LYP, they also have poor R^2 values, with the highest at 0.49. The pure DFT functionals also give poor binding energies when compared to the benchmarks, with relatively high MUE values and very poor R^2 values. However the poor performance of these methods is not unexpected as it is known that they don't treat dispersion correctly, which is why the more accurate methods and the dispersion corrected methods give much more favourable values compared to the benchmark used here, which is the best available.

3.3.2 Ruthenium Arene Adenine Complexes

In order to be sure that the methods, that show promise, are not specific to the systems in question, another system was constructed where the nucleobase being studied was adenine instead of guanine. The ruthenium arene guanine systems above were modified by hand to transform the guanine into adenine, these were then fully optimised using B97-D2. For these optimised systems binding energies were calculated with all the same methods as shown above. Benchmark DF-LMP2 values are reported in table 3.3.2.1, along with the better performing methods shown above, B97-D2, B97-D3, BHandH, and MP2(0.25), these are displayed graphically in figure 3.3.2.1.

Table 3.3.2.1 Binding energies of ruthenium arene complexes to adenine calculated using DF-LMP2, MP2(0.25), BHandH, B97-D2, and B97-D3 (kcal/mol).

	B97-D2	B97-D3	BHandH	MP2(0.25)	DF-LMP2
Benz	-69.16	-65.56	-71.83	-67.68	-72.43
Cym	-62.85	-58.68	-66.40	-62.33	-67.55
Bip	-62.68	-58.17	-67.50	-64.47	-71.07
Anth	-60.54	-56.37	-65.90	-61.92	-68.27
Dha	-70.93	-66.98	-74.75	-69.89	-76.72
Tha	-70.04	-66.02	-74.64	-69.08	-75.97

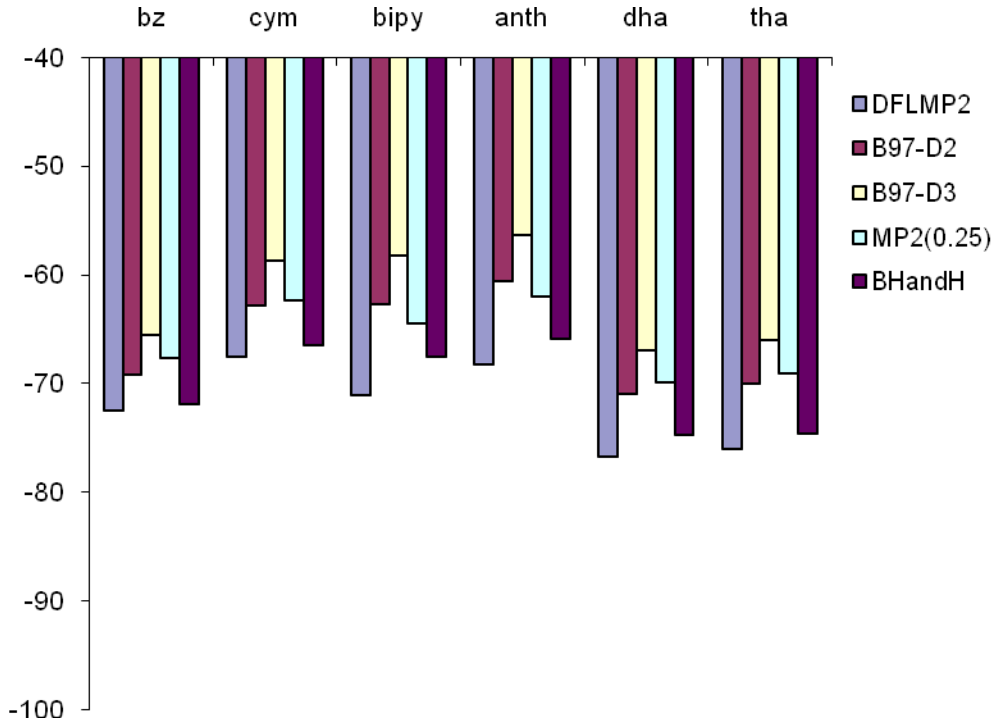


Figure 3.3.2.1 Graph illustrating the difference in binding energies for the ruthenium complexes with adenine calculated using DF-LMP2, B97-D2, B97-D3, MP2(0.25), and BHandH.

Analysis of the calculated binding energies using several methods for the guanine ruthenium arene systems has shown that the methods that best reproduce the benchmark values were BHandH, followed by MP2(0.25), and B97-D2. For the adenine systems once again BHandH gives the best results, with a difference from DF-LMP2 values of no greater than 2.5 kcal/mol, with the exception of the biphenyl complex at a difference of approximately 4.6 kcal/mol. Unlike the guanine system, where BHandH

overestimates the binding energies, in this case the energies are underestimated. B97-D2 also performs relatively well, with an average difference from the benchmark of 6.0 kcal/mol, once again the biphenyl system has the greatest error. For the guanine system MP2(0.25) performed marginally better than B97-D2, but in this case they both calculate similar values. Out of the methods shown in table 3.3.2.1 and figure 3.3.2.1, B97-D3 performs the worst, compared to the B97-D2 values it underestimates the binding energies by approximately 4.0 kcal/mol, as can also be seen for the guanine complexes.

DF-LCCSD calculations were also carried out to confirm the use of DF-LMP2 as benchmark, but as above, due to expense and time only the benzene and dha systems were studied. Using the same basis set as DF-LMP2, the DF-LCCSD binding energies were calculated as –69.18 kcal/mol for benzene and -73.56 kcal/mol for dha. These values, while not in as close agreement as the guanine DF-LCCSD values, are still in reasonable agreement with the DF-LMP2 values of –72.43 kcal/mol and -76.72 kcal/mol for benzene and dha respectively. The binding energies were also calculated for all the DFT and post Hartree-Fock methods used above, MUE values and R^2 values are reported in table 3.3.2.2.

Table 3.3.2.2 The mean unsigned error (MUE) in kcal/mol, Spearman correlation rank, and Pearson correlation of all methods used for the ruthenium arene adenine systems against the benchmark DF-LMP2 method.

	MUE	R^2 Spearman Rank	R^2 Pearson Correlation
MP2	10.65	0.94	0.96
MP2(0.25)	6.11	0.94	0.95
B2PLYPD	9.12	0.94	0.89
BHandH	1.83	0.94	0.93
B97-D2	5.97	0.83	0.83
B97-D3	10.04	0.83	0.80
wB97XD	8.08	0.89	0.96
M06	13.88	0.89	0.86
M06-2X	13.38	1.00	0.95
PBE0	18.25	0.66	0.64
TPSSh	20.09	0.66	0.52
B3LYP	25.90	0.66	0.46
mPW1PW91	20.76	0.66	0.58
PBE	19.49	0.66	0.48
TPSS	20.62	0.66	0.45
BLYP	30.03	0.66	0.30

Analysis of the data in table 3.3.2.2 shows that the methods used in table 3.3.2.1 and figure 3.3.2.1, are the ones that reproduce the benchmark data best. The wB97XD functional also performs well and has a lower MUE than B97-D3. All these methods give reasonable MUE values and good R^2 values, with once again BHandH performing best, followed by B97-D2 and MP2(0.25). The Minnesota functionals once again have very good correlation with the DF-LMP2 values, but have slightly worse MUEs than the DFT-D methods. The remaining hybrid functionals and the non-dispersion corrected pure DFT functionals perform poorly, to a similar standard as in the guanine system, giving large errors in energy and low R^2 values.

3.3.3 Comparison of Nucleobases

The binding energies calculated for the guanine and adenine systems above are not directly comparable, as the adenine was fully optimised, while the guanine was a fragment of a larger DNA structure that retained its geometry from that form. To allow direct comparisons the guanine system used has been fully optimised at the B97-D2 level, with the def2-TZVP basis set, as used for optimisation of the adenine complexes. For these newly optimised guanine complexes binding energies have been calculated, using the methods highlighted above as being able to best reproduce the benchmarks, as well as the benchmark methods. The differences between the energies of the reoptimised guanine and adenine complexes are reported in table 3.3.3.1 below, the optimised guanine binding energies are reported in the appendix.

Table 3.3.3.1 Guanine-Adenine binding energy difference (kcal/mol).

	B97-D2	BHandH	MP2(0.25)	DF-LMP2
Benz	-23.87	-25.67	-23.16	-22.05
Cym	-23.07	-25.22	-22.96	-22.03
Bip	-23.15	-25.63	-23.34	-22.12
Anth	-19.91	-21.18	-20.43	-18.75
Dha	-20.57	-20.68	-21.46	-20.04
Tha	-20.60	-22.16	-21.70	-20.27

The newly optimised guanine complexes have binding energies that have relatively small errors from the guanine structures above, that were taken from the conformation within a larger DNA fragment. From visual observation of the start and end points of the optimisations the structures had little rearrangement, leading to these minor differences in binding energies.

The preference of binding of ruthenium arene complexes to nucleobases, has been widely reported in publications,[29-35] with a clear preference for the N7 position of guanine being accepted. This effect has also been seen for several other anti-cancer metal complexes, particularly platinum (II) based drugs such as cisplatin.[36] Analysis of data in table 3.3.3.1 shows a greater binding energy of approximately 20 kcal/mol for guanine over adenine, both at the N7 position, across all six arene systems being

studied. This is in good agreement with results by Gkionis *et al.*,[32] and Futera *et al.*,[37] using BHandH/M05 and MP2 respectively.

3.4 Conclusions

The aim of these calculations was to confirm whether using the GGA functional B97 with an empirical dispersion correction of the type D2, was suitable to calculate binding energies within transition metal complexes, in light of recent publications using the newer D3 correction. Benchmark binding energies were calculated using DF-LMP2 and warranted by carrying out selected DF-LCCSD calculations. As well as the two B97-D functionals, several pure DFT, hybrid DFT, and post Hartree-Fock methods were tested and results have shown that the best methods are BHandH, B97-D2, and MP2(0.25). With wB97XD, B2PLYPD, and B97-D3 also performing adequately, but not to the desired level of accuracy. Analysis of the B97-D2 binding energies and through comparison to other more reported methods such as DF-LMP2 and BHandH, it is clear that this method can suitably reproduce values of significantly more expensive calculations. It has also been shown that it can reproduce sensible values of binding energies for systems that contain ruthenium arene based complexes binding to purine type nucleobases.

3.5 References

1. Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. *J. Comput. Chem.* **2006**, *27*, 491-504.
2. Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787-1799.
3. Mutter, S. T.; Platts, J. A. *Chem-Eur J* **2010**, *16*, 5391-5399.
4. Mutter, S.T.; Platts, J.A. *J. Phys. Chem. A* **2011**, 115, 11293-11302.
5. Dong, H.; Meng, Q.; Chen, B.-Z.; Wu, Y.-B. *J. Organomet. Chem.* **2012**, 717, 108-115.
6. Wang, Y.; Angermund, K.; Goddard, R.; Kruger, C. *J. Am. Chem. Soc.* **1987**, 109, 587.
7. Haaland, A.; Nilsson, J. *Acta Chem. Scand.* **1968**, 22, 2653.
8. Hill, J. G.; Platts, J. A.; Werner, H. J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4072-4078.
9. Tsuzuki, S.; Fujii, A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2584-2594.
10. Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985-1993.
11. Boese, A. D.; Martin, J. M. L.; Klopper, W. *J. Phys. Chem. A* **2007**, *111*, 11122-11133.
12. Morgado, C. A.; Jurecka, P.; Svozil, D.; Hobza, P.; Sponer, J. *J. Chem. Theory Comput.* **2009**, *5*, 1524-1544.
13. Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem-Us* **1996**, *100*, 18790-18794.
14. Jurecka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89-94.
15. Sponer, J.; Leszczynski, J.; Hobza, P. *J Phys Chem-Us* **1996**, *100*, 5590-5596.
16. Wang, W. Z.; Pitonak, M.; Hobza, P. *Chem. Phys. Chem.* **2007**, *8*, 2107-2111.
17. Gkionis, K.; Hill, J. G.; Oldfield, S. P.; Platts, J. A. *J. Mol. Model.* **2009**, *15*, 1051-1060.
18. 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.
19. Chen, H. M.; Parkinson, J. A.; Parsons, S.; Coxall, R. A.; Gould, R. O.; Sadler, P. J. *J Am Chem Soc* **2002**, *124*, 3064-3082.
20. Chen, H. M.; Parkinson, J. A.; Morris, R. E.; Sadler, P. J. *J. Am. Chem. Soc.* **2003**, *125*, 173-186.
21. Habtemariam, A.; Melchart, M.; Fernandez, R.; Parsons, S.; Oswald, I. D. H.; Parkin, A.; Fabbiani, F. P. A.; Davidson, J. E.; Dawson, A.; Aird, R. E.; Jodrell, D. I.; Sadler, P. J. *J. Med. Chem.* **2006**, *49*, 6858-6868.
22. Peacock, A. F. A.; Sadler, P. J. *Chem-Asian J.* **2008**, *3*, 1890-1899.
23. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, 132, 154104.
24. Ryde, U.; Mata, R.A.; Grimme, S. *Dalton Trans.* **2011**, 40, 11176.
25. Hill, J.G.; Platts, J.A. *J. Chem. Phys.* **2008**, 129, 134101.
26. Goerigk, L.; Kruse, H.; Grimme, S. *Chem. Phys. Chem.* **2011**, 12, 3421-3433.
27. Zhao, Y.; Truhlar, D.G. *Theor. Chem. Acc.* **2008**, 120, 215.
28. Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, 10, 6615-20.
29. Liu, H. K.; Berners-Price, S. J.; Wang, F. Y.; Parkinson, J. A.; Xu, J. J.; Bella, J.; Sadler, P. J. *Angew. Chem. Int. Edit.* **2006**, *45*, 8153-8156.
30. Liu, H. K.; Parkinson, J. A.; Bella, J.; Wang, F. Y.; Sadler, P. J. *Chem. Sci.* **2010**, *1*, 258-270.
31. Wing, R. M.; Pjura, P.; Drew, H. R.; Dickerson, R. E. *Embo. J.* **1984**, *3*, 1201-1206.

32. Gkionis, K.; Platts, J. A.; Hill, J. G. *Inorg. Chem.* **2008**, *47*, 3893-3902.
33. Rothlisberger, U.; Gossens, C.; Tavernelli, I. *J. Am. Chem. Soc.* **2008**, *130*, 10921-10928.
34. Rothlisberger, U.; Gossens, C.; Tavernelli, I. *J. Chem. Theory Comput.* **2007**, *3*, 1212-1222.
35. Deubel, D. V.; Lau, J. K. C. *Chem. Commun.* **2006**, 2451-2453.
36. Gkionis, K.; Hicks, M. Robertazzi, A.; Hill, J.G.; Platts J.A. *Quantum Biochem.* **2010**,
37. Futera, Z.; Klenko, J.; Sponer, J. E.; Sponer, J.; Burda, J. V. *J. Comput. Chem.* **2009**, *30*, 1758-1770.

Chapter 4: Ruthenium (II) Arene Complexes and their Interactions with DNA

4.1 Introduction

The ruthenium (II) complexes being studied in this chapter are based on those synthesized by the Sadler group[1-5] and studied through coordination to nucleobases and non-covalent interactions in sections 3.3 and 3.2.2 respectively. The ruthenium has a chelating bi-coordinate ethylenediamine (en) ligand and a variety of arene systems coordinating in an η^6 fashion. These arenes being studied are benzene (benz), cymene (cym), biphenyl (bip), anthracene (anth), dihydroanthracene (dha), and tetrahydroanthracene (tha), as shown in figure 4.1.1.

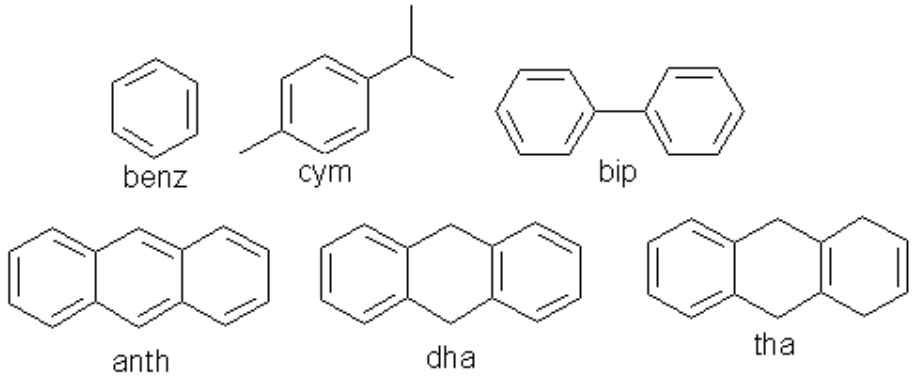


Figure 4.1.1 Structures of the arenes coordinated to ruthenium.

Due to these ruthenium complexes showing potential as anti-cancer drugs, their interactions with DNA are of interest, especially as the systems containing extended aromatic region have the capability to act as intercalators and can therefore form additional π-interactions with the aromatic regions of nucelobases. To attempt to probe these interactions DFT techniques have been utilised to obtain optimised structures and binding energies of this selection of complexes with the DNA duplex GpC·CpG, which has been extensively studied in analogous systems containing platinum anti cancer agents. The relatively low computational cost of DFT, compared to techniques that more accurately model non-covalent interactions, is key here, with two routes being explored, that of DFT with an empirical dispersion correction and a QM/MM method combining BHandH with the AMBER forcefield.

4.2 Optimisation of Base Pair Steps

Study of metal-DNA interactions is a challenge, even with the relatively low computational cost of DFT. For the complexes here, all but the benzene model are in excess of 150 atoms when including the metal complex, two base pairs and sugar-phosphate backbone. Previous work to make such calculations viable has been carried out by utilising hybrid quantum mechanics/molecular mechanics method (QM/MM) methods,[6] in which only a small portion of the overall structure is modelled at the full DFT level and the remainder with a molecular mechanics (MM) approach. Although this approach has been used with some success, the suitability of standard MM methods for transition metal systems is far from guaranteed. Transition metals generally exhibit strong d-orbital effects that can be hard to capture in molecular mechanics form.[7]

In order to circumvent the need for empirical MM parameters for Ru, the aim of this chapter is to use full DFT to examine the binding of ruthenium(II) complexes of the form $[\eta^6(\text{arene})\text{Ru(II)(en)}]^{n+}$ to DNA, by obtaining optimised geometries and binding energies at this level. This data should not only give new insight into the interactions between metal complex and DNA and within the DNA fragment, but will also form a valuable benchmark for testing faster but more approximate methods, such as QM/MM, that could be used to examine larger fragments of DNA. Calculations and results within this chapter are based on those published by our group within reference 8.

4.2.1 Optimisations using B97-D

Interactions of ruthenium complexes shown above with DNA were studied in three different forms: 1) with an “open” DNA structure and the arene ligand acting as an intercalator between base pairs; 2) with the same open DNA structure, but with the arene pointing away from DNA; 3) with DNA in its canonical B- form and the arene group pointing away from DNA (figure 4.2.1.1). The “open” DNA structure was constructed by modifying the crystal structure of a DNA-intercalator complex with the nucleic acid database ID DRBB19.[9] This is a dinucleoside monophosphate structure with UpA·ApU sequence; this was modified to give a suitable representation of a GpC·CpG fragment with an open structure allowing space for intercalator. The canonical B-DNA was constructed using the w3DNA web interface.[10] The guanine coordinated to ruthenium, through the N7 site, will be referred to as G1, forming a base

pair with cytosine, referred to as C1. The non-coordinated base pair step will be referred to as C2 and G2 for cytosine and guanine, respectively as shown in figure 4.2.1.1. The efficiency of the pure DFT B97-D method, when implemented with resolution of identity (RI) approximation, allows study of the entire system of the ruthenium arene complex with two base pairs and associated sugar phosphate backbone and Na^+ counterions at the full DFT level.

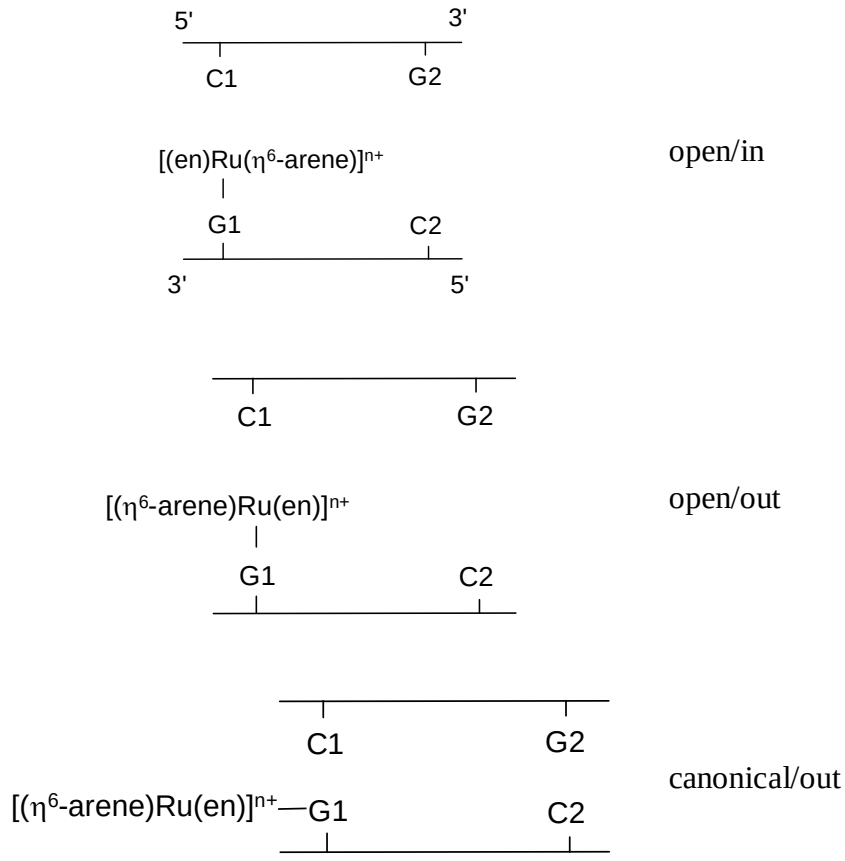


Figure 4.2.1.1 Numbering sequence for base pairs in model DNA systems and schematic form of the three DNA systems.

Following optimisation of all systems using B97-D/def2-TZVP (with an ECP on ruthenium), the binding energy of the ruthenium fragment to DNA was calculated, as were selected geometrical parameters. For some of the systems several possible orientations of arene were identified, relating to orientation with respect to ruthenium, twist of biphenyl along the phenyl-phenyl bond, and tilt of the tha and dha about the saturated carbons. Optimisations were run for several different orientations for twenty cycles, after which the lowest energy systems were taken through to full optimisation and reported below.

Table 4.2.1.1 Binding energies of the ruthenium arene complex to GC DNA fragment in the three forms (kcal/mol).

DNA / arene	open / in [a]	canonical / out	open / out
Benz	-174.28	-168.92	/
	-52.48		
Cym	-172.41	-169.89	-161.19
	-61.28		
Bip	-169.70	-157.86	-172.97
	-67.63		
Anth	-167.27	-154.84	-169.06
	-72.45		
Dha	-177.91	-154.84	-169.98
	-77.91		
Tha	-179.55	-154.92	-169.34
	-79.07		

[a]First line gas phase, second line aqueous PCM solvent.

Table 4.2.1.1 reports binding energies of the ruthenium fragments to three different forms of DNA, and figure 4.2.1.2 shows the most important data graphically. Values are not corrected for basis set superposition error (BSSE), since counterpoise correction was carried out on the benz, anth and tha open / in systems, finding that correction was less than 0.1 kcal/mol in all cases, *i.e.* essentially negligible relative to other errors in functional, basis set *etc*. Grimme *et al*[11] have also shown when a suitably large basis set is employed in DFT-D methods results become essentially free of BSSE. Table 4.2.1.1 shows the large stabilization that stems from formation of covalent Ru—N7 bonds, along with various non-covalent interactions, in the gas-phase. The partially saturated dha and tha complexes are particularly strongly bound when the arene acts as an intercalator systems (denoted "open / in"), whereas the fully aromatic anthracene complex is the most weakly bound complex to this form of DNA. In this context, we note that Sadler and co-workers have not considered anthracene in experimental studies: it is included here to examine the effects of rigidity and aromaticity relative to the saturated analogues. From this data, it appears that the increased rigidity of the anthracene ligand does not lend itself to efficient intercalation between base pairs. The biphenyl complex is slightly more strongly bound than anthracene, apparently due to the

flexibility of rotation around the central C—C bond which allows formation of closer and therefore stronger stacking and C—H…π interactions.

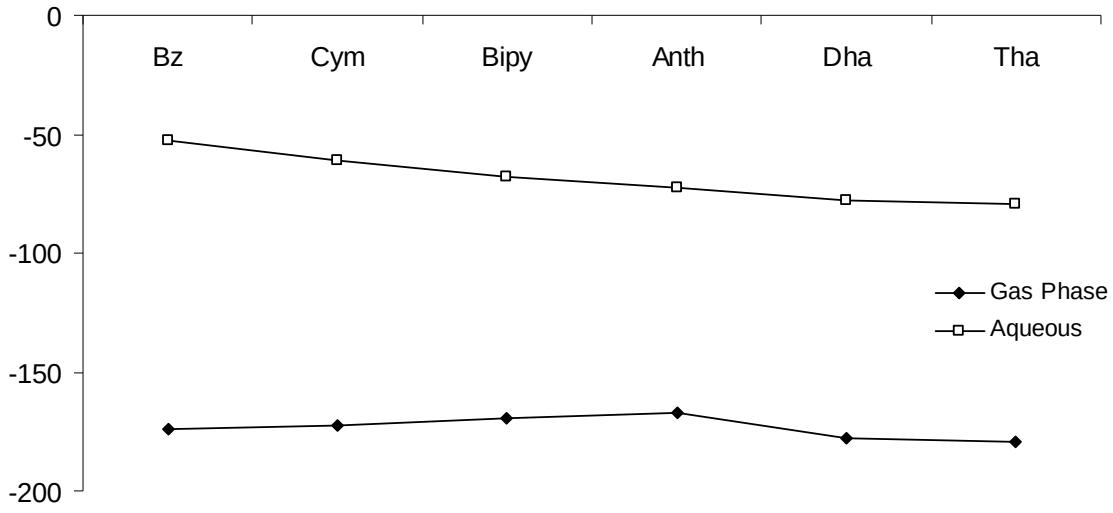
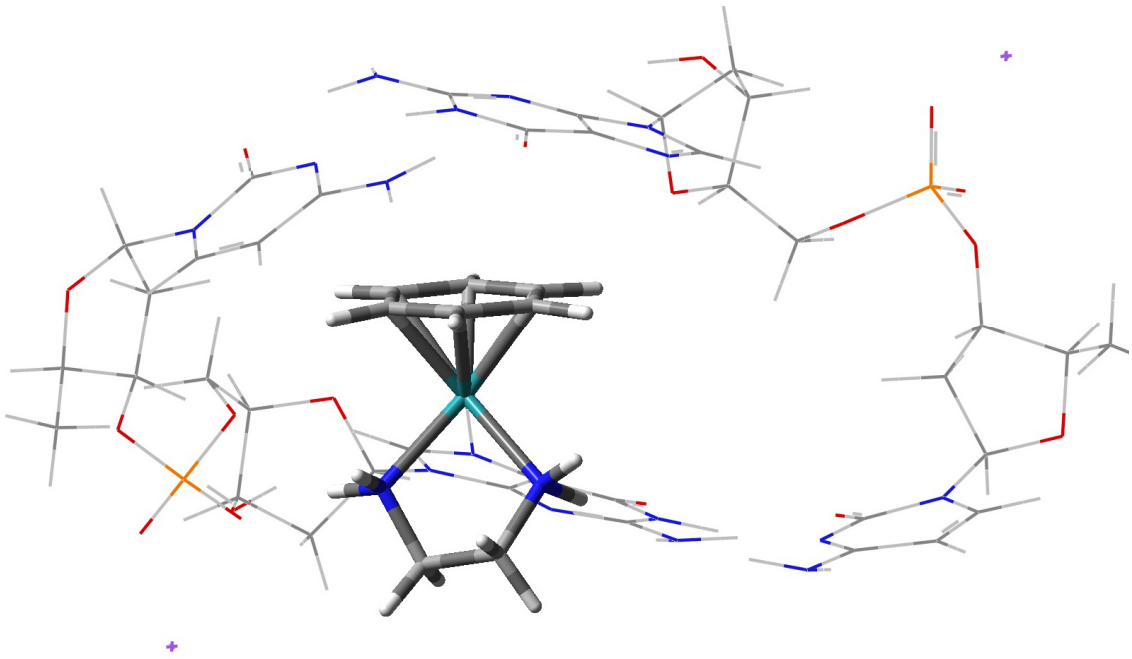


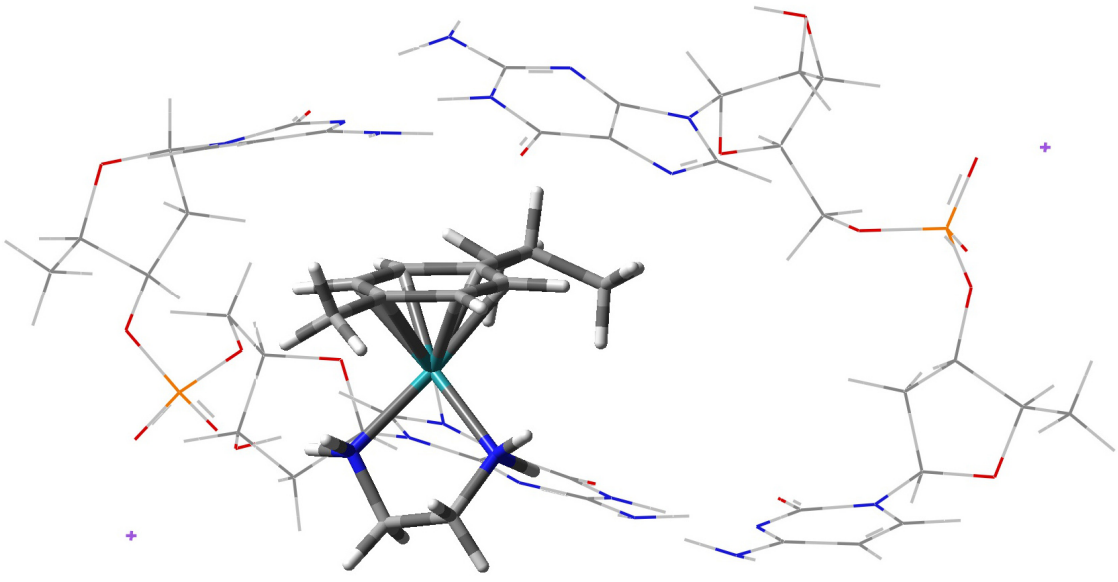
Figure 4.2.1.2 Gas- and aqueous-phase binding energies for open/in complexes (kcal/mol).

The large binding energy of the benzene complex has been noted before,[12] and was assigned to differences in the electronic structure of this ligand relative to others in the series, giving rise to stronger Ru—N covalent bonding. This is also observed in the cymene system, which has a large binding energy. The importance of the electronic structure of the benzene and cymene complexes is evident in their binding energies to DNA in its canonical B- form, for which the arene ligand points away from DNA. Here, both complexes are found to be *ca.* 10 kcal/mol more strongly bound than the remaining complexes. The remaining systems have very similar binding energies, consistent with the lack of significant non-covalent interactions between arene and DNA in this orientation. Binding energies are rather larger, but still similar for the orientation of arene away from DNA in its open form. The cymene system for the open/out form however has a significantly lower binding energy than the other ligands. In this orientation the bulky isopropyl substituent points into the gap in DNA, increasing unfavourable steric interactions and lowering binding energy. For the latter, no value is reported for benzene as optimization leads to the same geometry as for the arene pointing into DNA.

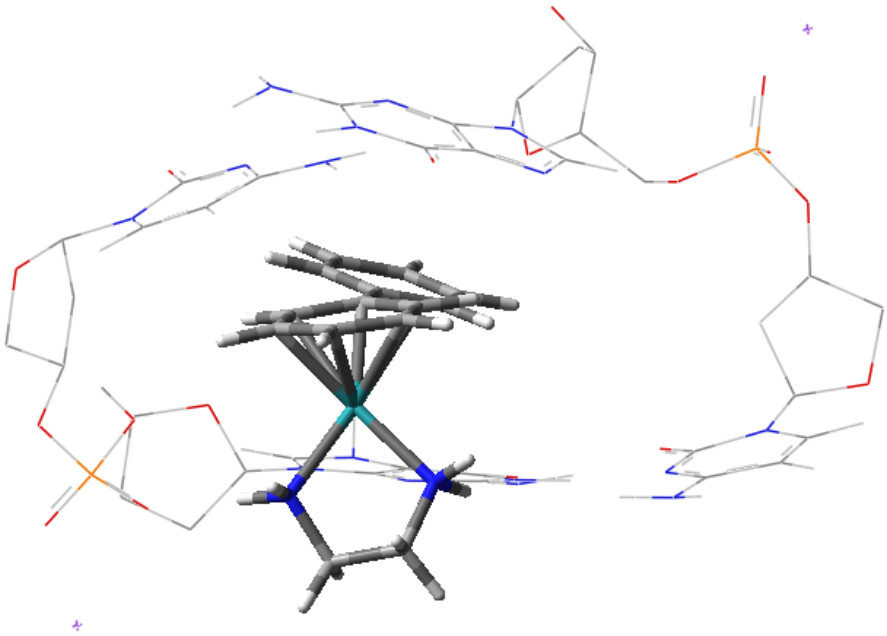
Table 4.2.1.1 and figure 4.2.1.2 also report binding energies corrected for the relative solvation of complexes and fragments, using a PCM model of aqueous solvent. Solvation corrections are relatively uniform for the larger arenes, at between 95 and 100 kcal/mol, but the corrections for benzene and cymene are much larger at 122 kcal/mol and 111 kcal/mol respectively, reducing the magnitude of the binding energies substantially. Following this correction, the benzene complex is by some distance the most weakly bound of the complexes considered, followed by the cymene complex. Inclusion of PCM correction also inverts the order of stability of biphenyl and anthracene complexes, and confirms that the greatest stability stems from the large but flexible dha and tha arene ligands.



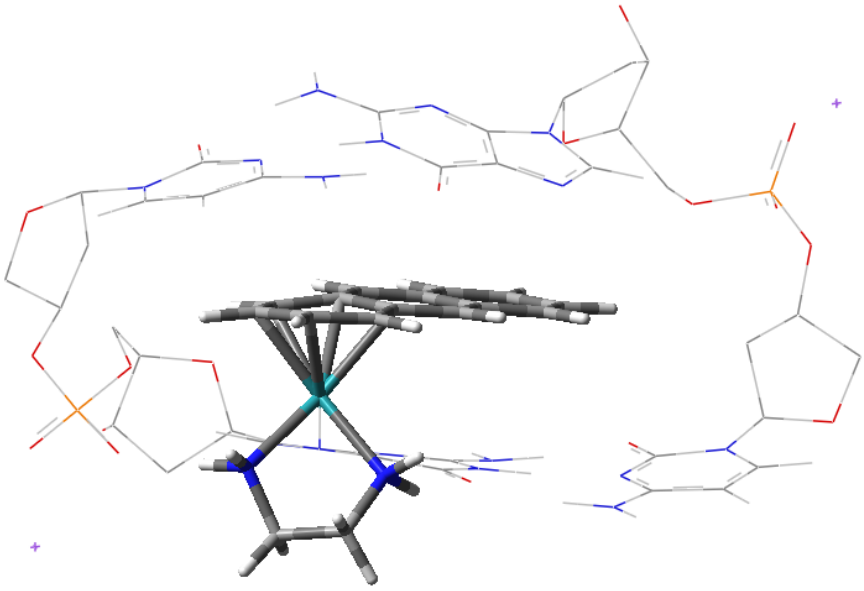
i)



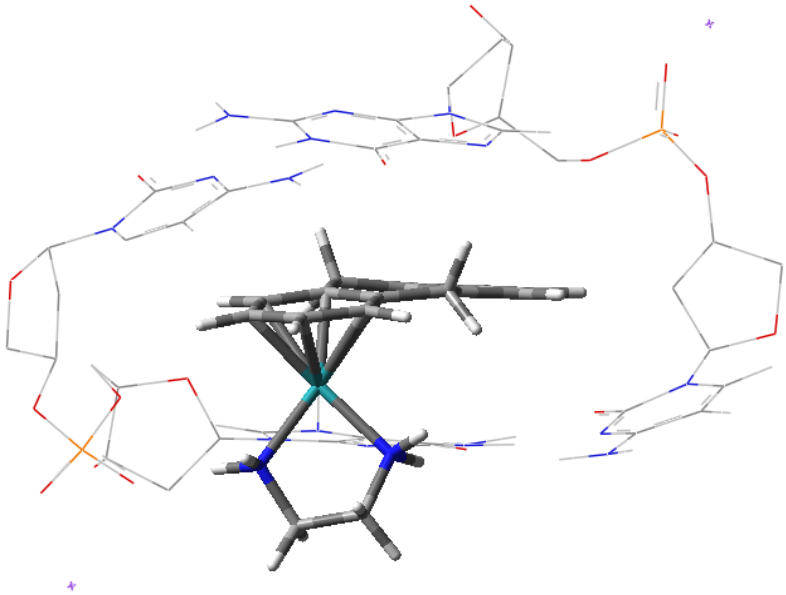
ii)



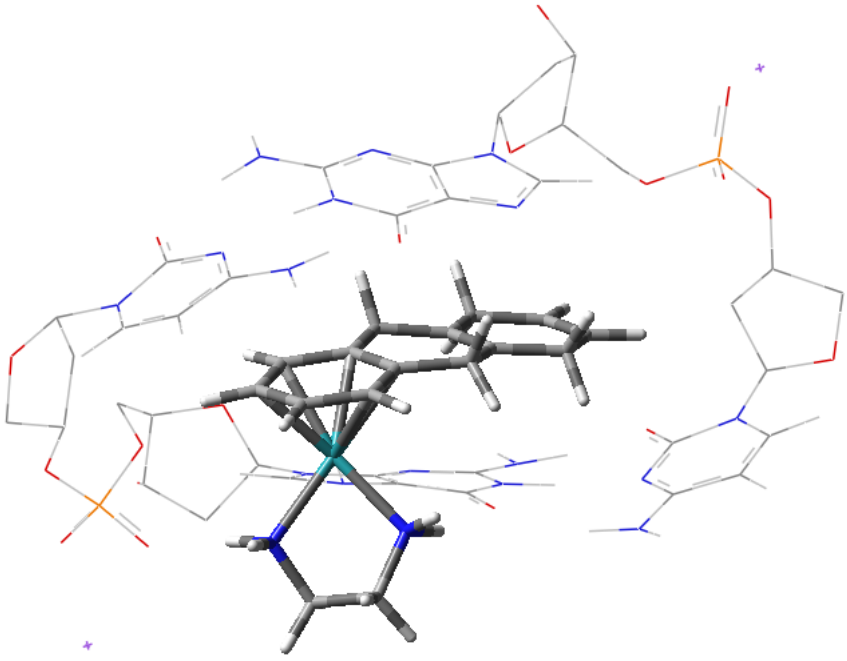
iii)



iv)



v)



vi)

Figure 4.2.1.3 Optimised structures of ruthenium arene (i) bz, ii) cym, iii) bip, iv) anth, v) dha, and vi) tha) complexes with "open-in" DNA systems. Ruthenium-arene fragment is shown in bold, with DNA as wireframe; hydrogens on sugar-phosphate backbone omitted for clarity.

In addition, the enhanced π-stacking interaction expected for arenes bound to positively charged metal centres, as discussed in section 3.2.2, apparently also plays a role here. This is evident in the optimized geometry, shown in figure 4.2.1.3, in which a relatively close stacking interaction between the coordinated benzene and G2 is observed. Without the arene in the open region of the DNA the coordinated benzene can form a closer interaction to the G2 guanine, as shown in table 4.2.1.2. The benzene complex and all the open/out systems have much closer interactions with G2, while the presence of an arene within the DNA gap results in closer C2 interactions. The greater freedom of the smaller benzene complex compared to the larger arene systems allows for formation of relatively close interactions with both G2 and C2 contributing to its large binding energy.

Table 4.2.1.2 Distance between centroid of coordinated arenes to centroid of C2 cytosine and G2 guanine for open/in and open/out systems (Å) [a].

	Open/in		Open/out	
	C2	G2	C2	G2
Benz	4.25	3.35	-	-
Cym	4.18	4.13	4.73	3.31
Bip	3.93	4.24	4.77	3.33
Anth	4.21	4.22	5.12	3.20
Dha	4.04	4.26	5.05	3.20
Tha	3.99	4.31	4.83	3.22

[a]Defined as the distance between the centroid of the coordinated aromatic ring and the centroid of the 6-membered ring of the relevant base.

Table 4.2.1.3 Geometrical parameters of hydrogen bonds (Å and °).

DNA / arene	open / in	canonical / out	Open / out
Benz			N/A
r(N—H...O6)	1.70	1.63	
∠(N—H...O6)	168.5	165.8	
r(N—H...OP)	1.83	-	
∠(N—H...OP)	163.9	-	
Cym			
r(N—H...O6)	1.69	1.65	1.72
∠(N—H...O6)	171.5	164.0	167.5
r(N—H...OP)	1.944	-	1.864
∠(N—H...OP)	159.0	-	162.9
Bip			
r(N—H...O6)	1.75	1.66	1.69
∠(N—H...O6)	152.4	166.4	174.6
r(N—H...OP)	1.87	-	1.89
∠(N—H...OP)	167.4	-	162.3
Anth			
r(N—H...O6)	1.80	1.64	1.69
∠(N—H...O6)	157.4	165.6	171.8
r(N—H...OP)	1.90	-	1.89
∠(N—H...OP)	169.9	-	164.0
Dha			
r(N—H...O6)	1.75	1.65	1.69
∠(N—H...O6)	162.4	163.4	172.9
r(N—H...OP)	1.86	-	1.89
∠(N—H...OP)	168.4	-	164.1
Tha			
r(N—H...O6)	1.74	1.64	1.69
∠(N—H...O6)	164.5	165.8	171.8
r(N—H...OP)	1.87	-	1.89
∠(N—H...OP)	168.2	-	164.4

In all systems studied, a hydrogen bond is formed between one of the N—H of the ethylenediamine ligands and the O6 of G1, as detailed in table 4.2.1.3. In the open/in and open/out systems, another H-bond is formed between the second N—H of the ligand and the formally O^- of a phosphate group on the sugar phosphate backbone; however, the canonical form does not possess this hydrogen bond. By forming the intercalator gap, the backbone deforms from its canonical orientation moving the phosphate group into a suitable orientation to form H-bonds with the coordinated complex. The benzene complex forms the shortest hydrogen bond to OP in the open/in system and cymene forms the shortest H-bond to O6, the presence of the smaller arenes apparently results in more freedom of movement during optimisation allowing the ruthenium complex to form these closer interactions. The larger arenes form slightly longer H-bonds, perhaps due to unfavourable steric interactions formed by the larger ligands, or due to competition between stacking and H-bonding. However this may have little effect as the change in distance overall is small, changing by *ca.* 0.1 Å across the series. In the other forms considered, variation in H-bond length is much less, presumably due to the lesser role of non-covalent interactions of arenes, which point away from DNA.

To form ideal stacking interactions, the planar constituents should be exactly parallel. The angles between the planes of the extended arene regions and of the 6-membered rings of the nucleobases for the open/in system are reported in table 4.2.1.4. The benzene complex forms a stacking interaction with G2, with an angle between the two planes of 11.88 °. The angle between the planes of benzene and C1 is similar, but the intramolecular distance is too large to form an interaction. The cymene complex forms a stacking interaction between the coordinated ring and C2 with angle between planes of 11.6 °, and like for the benzene there is a near parallel interaction with C1, albeit with a large distance. For the biphenyl complex, freedom of rotation about the central C—C bond allows the non-coordinated ring to form an almost parallel (4.64 °) interaction with G2. The anthracene system has the lowest binding energy, attributed to the rigidity of the ligand not allowing efficient intercalation. The angles formed between the planes of anthracene and C1, G2, and C2 support this view, having relatively large values, although the anthracene ligand is almost parallel to G1. A similar pattern is found for dihydroanthracene and tetrahydroanthracene complexes, but these have less reliance on π…π stacking interactions, as CH…π interactions prevail in the intercalative region.

Table 4.2.1.4 Angle between the planes of the nucleobases and the average plane of the arene ligand (˚).

DNA / arene	G1-arene	C1-arene	G2-arene	C2-arene
Benz	23.9	11.6	11.9	30.7
Cym	28.3	4.5	22.4	11.6
Bip	18.1	27.5	4.6	36.9
Anth	2.6	16.4	19.6	20.0
Dha	3.0	27.7	14.7	31.7
Tha	6.0	29.0	13.4	31.5

Further analysis was carried out on the intercalated complexes by comparing the DNA structure to a standardised canonical structure to obtain the shift, slide, rise, tilt, roll, and twist values, as reported in table 4.2.1.5, using the x3DNA program. The most striking difference lies in the value of rise, *i.e.* the vertical separation of base pairs, which ranges from 7.5 to 8.5 Å and is therefore more than twice the equivalent value in B-DNA. Slide values are also much larger than in the canonical form, and are uniformly negative in the optimized Ru-complexes. Variations in shift from the canonical value are also large, but for this parameter both positive and negative values are found. Roll angles are all significantly more positive than the standard value, indicating substantial opening of the base pairs induced by ruthenium complexation, while twist values are uniformly smaller than the standard value.

Table 4.2.1.5 Base-pair parameters for optimized intercalated complexes, along with standard B-DNA values for GC (Å and °).

	Shift	Slide	Rise	Tilt	Roll	Twist
Benz	-2.11	-4.38	8.53	-12.1	12.0	20.3
Cym	-2.16	-4.08	8.00	-5.5	12.2	10.3
Bip	0.84	-3.87	7.75	3.2	8.9	20.9
Anth	-1.17	-3.29	7.50	-3.0	8.6	18.9
Dha	-1.07	-3.49	7.98	-4.5	10.3	22.8
Tha	1.06	-3.42	8.06	4.0	8.3	22.0
B-DNA	0.00	0.54	3.38	0.0	1.7	38.0

As well as calculating binding energies, quantum theory atoms in molecules (QTAIM) analysis was carried out on the open/in system to quantify interactions within the intercalator region, using the AIM2000 package. We are particularly interested in the (3,-1) bond critical points (BCP), minima in the bond direction and maxima in perpendicular directions, which are informative about any interactions present (*cf*. section 2.8.2).

Table 4.2.1.6 Number and types of interactions for optimized intercalated complexes (open/in), calculated using AIM.

	coordinated ring		free ring(s)	
	π…π	CH…π	π…π	CH…π
Benz	6	0	-	-
Cym	2	7	-	-
Bip	3	1	7	0
Anth	3	1	7	0
Dha	4	0	4	4
Tha	4	0	5	8

The numbers of non-covalent interactions calculated for the open/in system, as reported in table 4.2.1.6, are in broad alignment with binding energy data. Bip and anth both have eleven interactions, less than dha, with twelve, and tha, with seventeen. This agrees with the ordering of binding energies, but is not indicative of the overall strength of the interactions, as shown by the benz and cym systems which have fewer interactions than bip and anth, but more π…π stacking from the directly coordinated ring, where the additional cation…π interactions may have an effect. As expected, increased saturation of the arene increases the number of CH…π interactions, shown by the increase through anth to tha.

4.2.2 Optimisations using ONIOM

For the greatest level of accuracy when simulating interactions between transition metal complexes and DNA the highest level of theory available is ideal. However, as mentioned above, these ruthenium arene GC adducts are approaching the limit of what is possible with full DFT, at the current level of computing. Whilst the simulations carried out using B97-D give relatively good indications of the interactions occurring at

the area within DNA, where these complexes bind, there is also an interest in larger molecules, where full DFT is impractical and would be too computationally expensive to carry out. In these cases, where there is still a focus on accurate calculation of interactions and electronics in a local area of the system, a hybrid approach combining quantum mechanics and molecular mechanics (QM/MM) is undertaken.

In QM/MM calculations a small area of the entire molecule that is of greatest interest, in this case the ruthenium, arene, en ligand, and G1 nucleobase, is treated with quantum mechanics, dubbed the high layer, whilst the remainder is treated with molecular mechanics, the low layer. This is explained in greater detail in chapter 2. This approach enables accurate treatment of the binding site of these ruthenium arene complexes and the potential to increase the size of the DNA fragment to which it is bound, to that beyond the capabilities of full DFT.

Whilst it would be possible here to increase the size of these ruthenium DNA adducts, this approach will use QM/MM to optimise and calculate binding energies of the same systems examined in section 4.2.1, with an aim to test if using this more approximate method can yield similar results to those obtained with B97-D, for systems of the size associated with drug dimer adducts. The QM/MM method chosen for these calculations is ONIOM, where the high layer includes the entire ruthenium drug and the single guanine to which it is bound, the low layer includes the remaining nucleobases, sugar phosphate backbone, and the sodium counter ions, the link atom is replaced with a hydrogen on the N9 of guanine for calculations on the high layer. The chosen method is that used by Gkionis *et al.*, in studies of platinum-DNA complexes, where the high layer is treated with the DFT functional BHandH and the low layer with AMBER molecular mechanics.[13] The interest is still focusing on non-covalent interactions occurring between the drug and DNA so the use of a functional that can model these is key here, hence the choice of BHandH, which has shown to be effective in section 3.2.

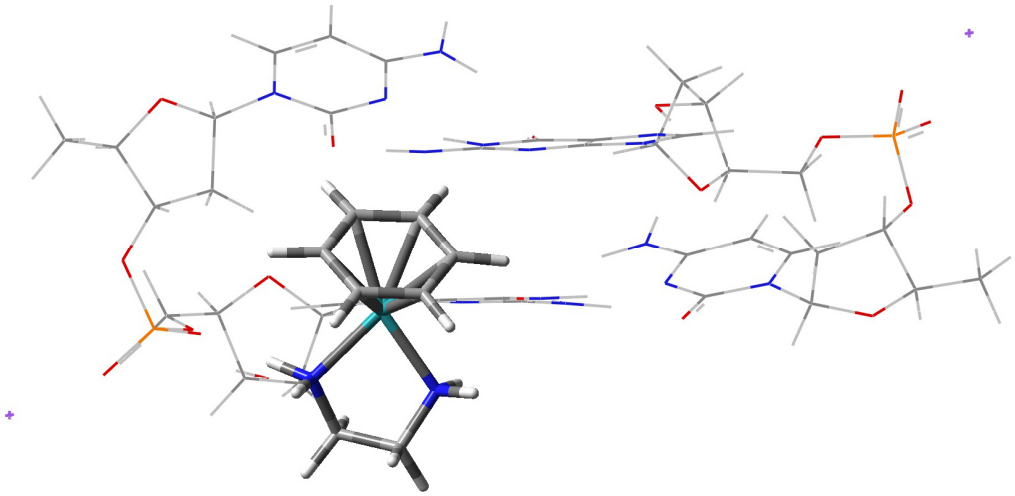
Calculations were only carried out on the open/in systems, as this was the configuration of greatest interest so deemed best suited for comparison. The starting points for optimisation were taken as the optimised B97-D structures. The optimisations were carried out initially using the 'geometry optimisation using direct inversion in the iterative subspace' (GDIIS) algorithm and with the parameters defined in reference 13, used for analogous systems containing platinum complexes. The structures failed to

fully optimise from this step, so from these outputs the lowest energy structure was taken and further optimised in Cartesian coordinates until the forces and gradient of the energy was deemed low enough that the structure was optimised. Within systems that have 'floppy' structures and a large degree of freedom, like the ones in question here, the potential energy surface can be quite flat, so it can be difficult to find the global minima, hence why these requirements were used to confirm optimisation. Binding energies of these optimised structures were calculated and are reported in table 4.2.2.1.

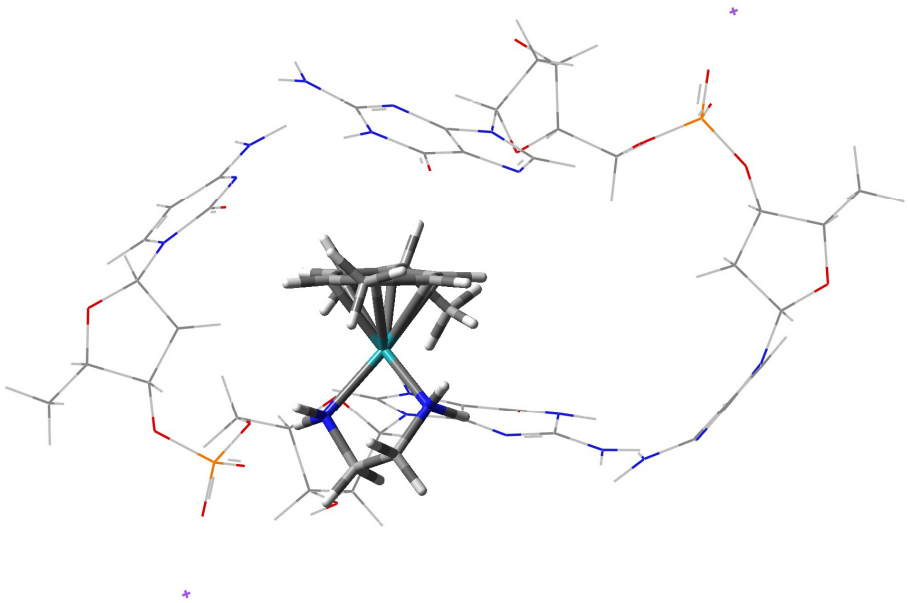
Table 4.2.2.1 Gas Phase Binding energies of the ruthenium arene complex to GC DNA (kcal/mol).

DNA / arene	Binding Energy
Benz	-189.43
Cym	-182.44
Bip	-185.19
Anth	-186.03
Dha	-191.66
Tha	-195.89

The binding energies reported in table 4.2.2.1 have been calculated using BHandH on the entire molecule, with BSSE accounted for using the counterpoise method. This approach however would not be suitable for larger molecules but in this case was carried out to enable a better comparison with the results obtained using B97-D. The tha complex yields the greatest binding energy, as expected due to it being able to form the greatest amount of stabilising non covalent interactions with the adjacent nucelobases. Dha gives the second greatest binding energy, approximately 4 kcal/mol less than tha, as explained above likely due to increased rigidity of the aromatic region acting as intercalator, this can also be attributed to the lower binding energy of the anth ligand. The unusually high binding energy of the benz system has been noted above, as well as the bip system having a lower than expected value, due to the smaller amount of interactions it can form even though it has more flexibility through rotation of the C-C bond between the phenyl rings. In the case of cymene it has comparatively lower binding energy than what would be expected from the values in table 4.2.1.1, to explain this observation of the structures is required, with the optimised systems that show significant difference shown in figure 4.2.2.1 and the associated base pair parameters reported in table 4.2.2.2.



i)



ii)

Figure 4.2.2.1 Optimised structures of ruthenium arene i) bz, ii) cym complexes with "open-in" DNA systems. Ruthenium-arene fragment is shown in bold, with DNA as wireframe.

Table 4.2.2.2 Base-pair parameters for optimized intercalated complexes, along with standard B-DNA values for GC (Å and °).

	Shift	Slide	Rise	Tilt	Roll	Twist
Benz	-0.61	-1.74	3.49	7.01	-6.15	24.31
Bip	-1.34	-2.48	6.27	13.56	26.42	25.94
Anth	0.77	-2.68	6.85	-1.64	-5.04	19.14
Dha	-0.56	-2.46	6.96	4.89	2.84	23.20
Tha	-1.16	-3.29	7.72	0.61	4.81	18.16
B-DNA	0.00	0.54	3.38	0.00	1.70	38.00

As mentioned above, the binding energy for cymene is lower than would be expected as for the ONIOM structures it has the lowest value, whilst for B97-D it is almost comparative with benzene in the gas phase. Observation of the cymene structure in figure 4.2.2.1 ii) shows that the base pair has undergone severe distortion from the expected structure, with the hydrogen bonds between bases broken. The movement of the nucleobases away from the isopropyl group would result in fewer stabilising interactions as unlike the B97-D case, the benzene ring coordinated to ruthenium does not form stacking interactions with above cytosine, this is likely the reason for the lower binding energy.

The distortion of the nucleobase configuration within cymene resulted in the x3DNA program not recognising a base pair step, hence base pair parameters were not obtained for this system, explaining its absence in table 4.2.2.2. For the remainder of the systems there is a small change in shift when compared to B-DNA, with all, apart from anth, being a negative change. For the slide parameter there is uniformly negative change, being greatest for tha and smallest for benz, whilst comparable for the remaining systems. The rise shows the greatest difference, as would be expected due to the arenes acting as intercalators, with the exception of benzene, which is comparable to B-DNA. The optimised benzene structure in figure 4.2.2.1 i) shows that unlike the B97-D case where the cytosine moved above the coordinated benzene to form a stacking interaction, the gap within the DNA closes.

4.2.3 Comparison between B97-D and ONIOM

It has been noted above that the benzene and cymene complexes have different structures for the optimisations with B97-D and ONIOM. Benzene forms a stacking interaction with the C2 cytosine when optimised with B97-D and the intercalating region partially closes when optimised with ONIOM and the cymene structure, when optimised with ONIOM, results in a base pair step too distorted to obtain parameters. From visual observation of the remaining structures, the adducts with the larger arene ligands (anth, dha, and tha) have comparable structures. This is further reinforced by the base pair parameters with no difference between them greater than 2.06 Å for shift, slide, and rise. However there are more significant differences within the tilt and roll parameters, notably for the roll value for anth and tilt for dha, but overall the two optimisation methods are in reasonable agreement for the structures of these systems.

When comparing the base pair parameters for the biphenyl system, optimised using the two methods, there are some stark differences between the tilt and roll parameters. This is not particularly evident from a comparison between the structures, but is most likely attributed to the C2 cytosine, which has a smaller angle with the plane of the G1 guanine, at 8.09 °, within in the ONIOM optimised structure, compared to 30.29 ° in the B97-D optimised structure. This large difference is not necessarily indicative of a significant difference between the structures and whilst not in as good agreement as the anth, dha, and tha, the bip is somewhat comparable between the two optimisation techniques.

From the graphical comparison between the gas phase binding energies of the B97-D and ONIOM optimised structures, in figure 4.2.3.1, the ordering for anth, dha, and tha is in good agreement, although there is some minimal difference between the magnitude of the change between dha and tha. The change in strength of binding energy ordering of the cym and benz systems can be attributed to the difference in the structures. The smaller geometrical change between the bip systems is also likely to account for BHandH resulting in a weaker binding energy than anth for the ONIOM optimised, whilst bip is stronger than anth in the B97-D optimised structure.

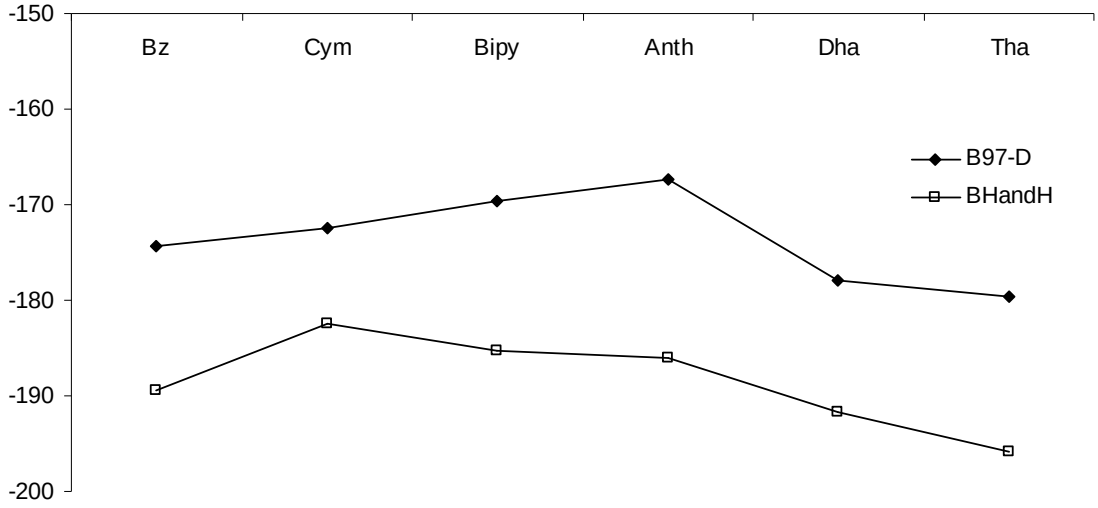


Figure 4.2.3.1 B97-D and BHandH gas phase binding energies for open/in complexes (kcal/mol)

As it is possible to carry out calculations on the systems, within this chapter, at the full DFT level, that is clearly the most desirable approach, so more credence should be given to the results obtained this way. Although with this method being realistically limited to a single base pair step, probing the effectiveness of the methods able to carry out calculations on the larger systems at this system size is a good test. The ruthenium drugs containing the smaller arene ligands, or in the case of cymene with fewer opportunities to create additional non-covalent interactions, show the greatest difference in optimised structure between the two methods. However the larger ligands are in relatively good agreement between B97-D and BHandH/AMBER optimised structures, likely due to less flexibility as they form the most non-covalent interactions. From these conclusions, and as increasing the size of the DNA increases the rigidity of the structure and so would likely yield results closer to those of the more comparable larger arene systems, the use of QM/MM methods for larger analogous DNA drug systems can be deemed to give reasonable structures.

4.3 References

1. 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.
2. Chen, H. M.; Parkinson, J. A.; Parsons, S.; Coxall, R. A.; Gould, R. O.; Sadler, P. J. *J. Am. Chem. Soc.* **2002**, *124*, 3064-3082.
3. Chen, H. M.; Parkinson, J. A.; Morris, R. E.; Sadler, P. J. *J. Am. Chem. Soc.* **2003**, *125*, 173-186.
4. Habtemariam, A.; Melchart, M.; Fernandez, R.; Parsons, S.; Oswald, I. D. H.; Parkin, A.; Fabbiani, F. P. A.; Davidson, J. E.; Dawson, A.; Aird, R. E.; Jodrell, D. I.; Sadler, P. J. *J. Med. Chem.* **2006**, *49*, 6858-6868.
5. Peacock, A. F. A.; Sadler, P. J. *Chem-Asian J.* **2008**, *3*, 1890-1899
6. Gkionis, K.; Platts, J. A. *J. Biol. Inorg. Chem.* **2009**, *14*, 1165-1174.
7. Deeth, R. J.; Anastasi, A.; Diedrich, C.; Randell, K. *Coordin. Chem. Rev.* **2009**, *253*, 795-816.
8. Mutter, S.T.; Platts, J.A. *J. Phys. Chem. A* **2011**, 115, 11293-11302.
9. Tsai, C.; Jain, S. C.; Sobell, H. M. *J. Mol. Biol.* **1977**, *114*, 301-315.
10. Zheng, G. H.; Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2009**, *37*, W240-W246.
11. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
12. Gkionis, K.; Platts, J. A.; Hill, J. G. *Inorg. Chem.* **2008**, *47*, 3893-3902.
13. Gkionis, K.; Platts, J. A. *J. Biol. Inorg. Chem.* **2009**, *14*, 1165-1174.

Chapter 5: Platinum (II) Anti-Cancer Drugs and their Interactions with DNA

5.1 Introduction

Platinum anticancer drugs, such as cisplatin and oxaliplatin, have been relatively well explored within the literature, both experimentally and computationally (*c.f.* section 1.3). The work in this chapter focuses on a relatively new drug kiteplatin or (cis-1,4-DACH)$PtCl_2$,[1] which has an isomeric form of oxaliplatin ligand, where the 1,2 diaminocyclohexane is replaced with one in a cis-1,4 conformation, as shown in figure 5.1.1, and its interactions with differing sizes of DNA adducts. The biological interest in this drug lies in its greater potency against platinum resistant cancer cell lines and besides oxaliplatin, is the only drug in advanced clinical development that is active against one of the most common cancers worldwide, colorectal cancer.[2,3]

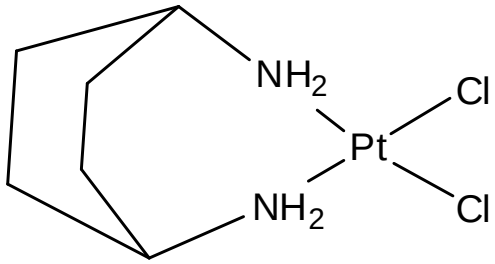


Figure 5.1.1 Structure of kiteplatin.

This work has been done in collaboration with Dr Nico Margiotta at Università degli Studi di Bari and Dr Paride Papadia at Università del Salento, who have been carrying out experimental studies on kiteplatin, particularly in regards to NMR experiments where the drug has been coordinated to several DNA fragments, single and double strand, of differing sizes. These platinum DNA adducts have been studied here, using a variety of DFT, QM/MM, and semi-empirical techniques, to elucidate structural and NMR data with the aim to complement data from the ongoing experiments.

5.2 Single Strand DNA Platinum Adducts

5.2.1 GpG

The kiteplatin DNA adduct (cis-1,4-DACH)Pt(d(GpG)) was constructed from the PDB entry 1PGC,[4] through truncation to the two coordinated guanines, with the sugar phosphate backbone. The oxaliplatin of the original PDB entry was then converted into cis-1,4-diaminocyclohexane by hand, the position of the platinum and the coordinating nitrogens was retained and the cyclohexane was adapted into a boat like conformation. This conformation was optimised within the MOPAC program, using the PM6 semi-empirical method in the gas phase, to give the head to head conformation (HH), where the two guanines point in approximately the same way. A relaxed scan was carried out at the PM6 level, with this structure as the starting point, around the dihedral angle 1-2-3-4, as shown in figure 5.2.1.1, at -5˚ intervals for 72 steps, using COSMO to model the effect of water as a solvent, with the aim to identify any other possible conformers that could be present. The profile of the first thirty steps of the scan is shown in figure 5.2.1.2, beyond this point steric hindrance between the two six-membered rings of the guanines caused the structure to reoptimise to the original HH starting conformation.

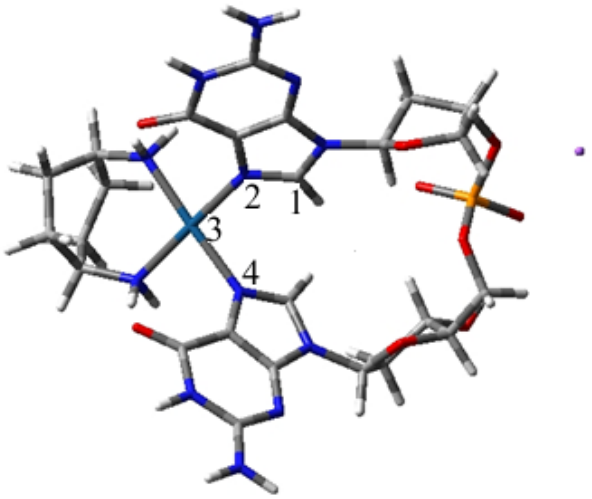


Figure 5.2.1.1 Head to head conformation of (cis-1,4-DACH)Pt(d(GpG)) showing the dihedral angle 1-2-3-4.

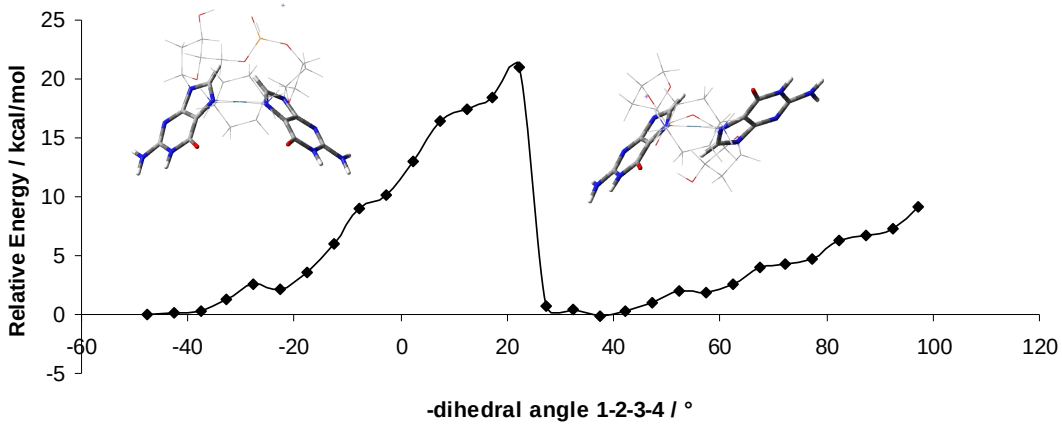


Figure 5.2.1.2 Profile of the first 30 points of the scan of (cis-1,4-DACH)Pt(d(GpG)) around the dihedral angle 1-2-3-4, energy is relative to the starting structure of the scan.

From observation of the scan profile in figure 5.2.1.2, there is a shallow minimum at the dihedral angle 37.33 °, and from visual observation of the calculation output this point was confirmed to be the head to tail conformation (HT), where the guanines are pointing in opposite directions. The dihedral angle that is scanned is a convenient variable to change, to be able to interconvert the structures, but is not directly the reaction coordinate, which likely leads to the sharp drop in energy as seen in figure 5.2.1.2. As a result of this the structure at the maxima is unlikely to be the transition state and the estimation of the rotation barrier, of 21.00 kcal/mol, is likely to be a significant overestimation. To further probe this, the same scan was carried out, but starting at the HT conformer and scanning toward the HH conformer as shown by the scan profile in 5.2.1.3.

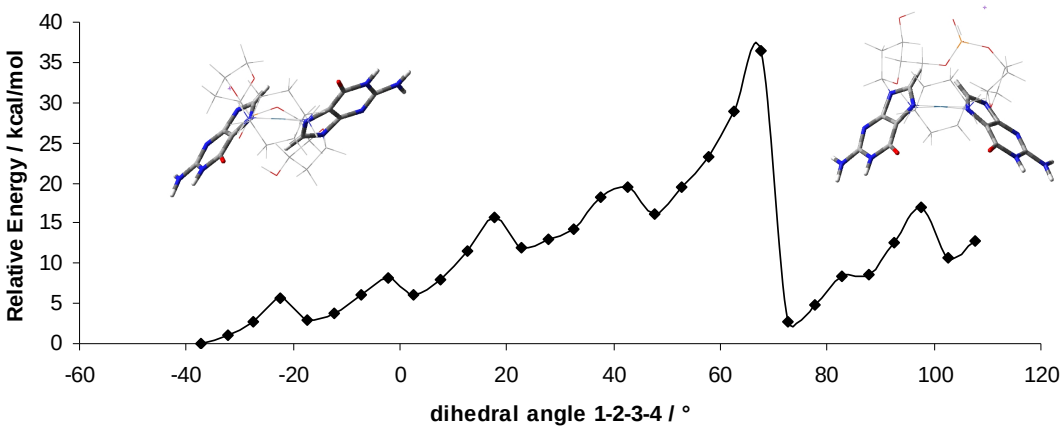


Figure 5.2.1.3 Profile of the first thirty points of the scan of (cis-1,4-DACH)Pt(d(GpG)) around the dihedral angle 1-2-3-4, starting from HT, energy is relative to the starting structure.

Observation of the minima at 72.67 ° shows this structure to be the HH conformer, albeit with a significantly larger dihedral angle than that of the PM6 optimised structure, used as the starting point for the initial scan, at 47.67 °. Whilst it retains the general form of the guanines in the head to head conformation, there is rearrangement in the sugar phosphate backbone compared to the preceding minima, suggesting the strain induced by the backbone likely causes the increase in dihedral angle. This minimum also has a greater energy difference from the HT conformer, at 2.74 kcal/mol, compared to 0.11 kcal/mol in figure 5.2.1.2. However after a PM6 optimisation, of HH, the dihedral angle becomes 47.25 °, which is more comparable to the original HH structure. An estimation of the barrier of rotation here is also likely an overestimation at 28.87 kcal/mol. A better estimation of the energy of the barrier for transition from HH to HT would be the point at which the two scans overlap, which occurs at the approximate dihedral angle of 8 ° with an approximation of the barrier at 8 kcal/mol. A transition state search was attempted in order to obtain an accurate value for the energy barrier but this was not possible as the calculation failed to complete.

The HH conformation, used as the initial scan starting point, and the HT conformation found were further optimised using the PM6-DH2 method, this improves on PM6 by adding an empirical correction for dispersion based and hydrogen bond interactions. Optimisations were carried out in the gas phase and single point COSMO corrections for water taken at the optimised gas phase geometry. Relative energies, within the chosen phase, are reported in table 5.2.1.1.

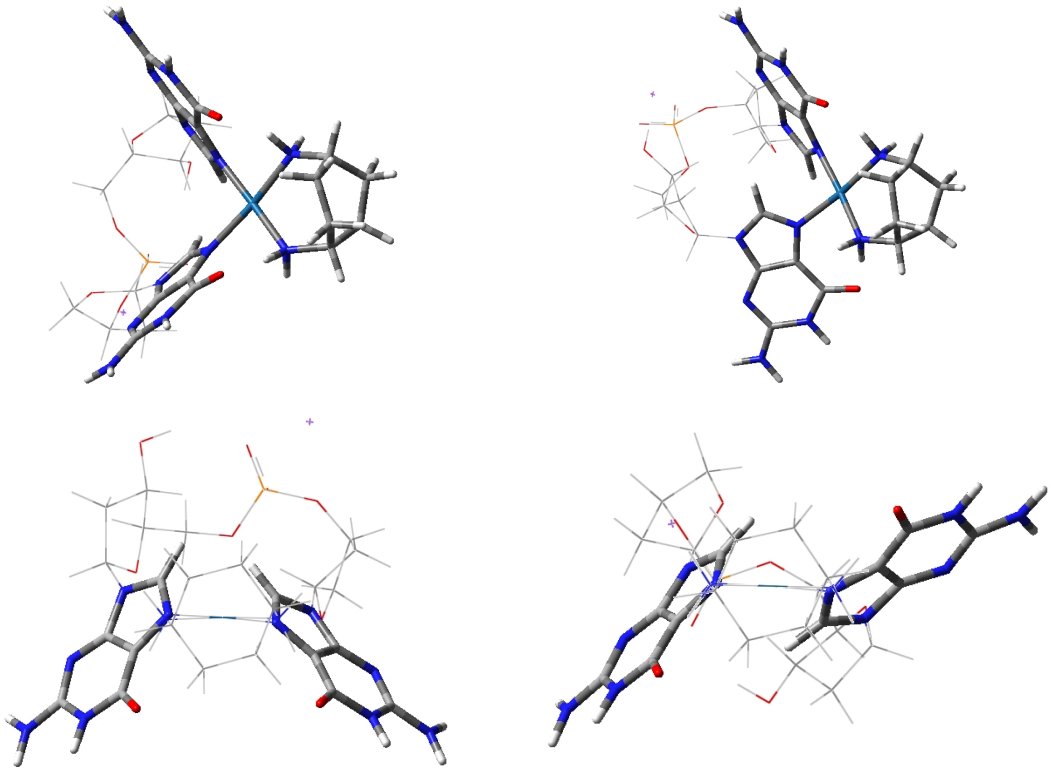
Table 5.2.1.1 Relative energies between conformations of HH and HT (cis-1,4-DACH)Pt(d(GpG)) calculated using PM6-DH2 with and without COSMO corrections for water (kcal/mol).

	Gas phasePM6-DH2 relative E	COMSO PM6-DH2 relative E
HH	0.00	0.00
HT	5.05	0.11

The relative energies in table 5.2.1.1 above show that in both gas phase and with the implicit solvent model the head to head conformation is the more stable, albeit for the gas phase the difference is significantly greater, at 5.05 kcal/mol, than with the COSMO correction at 0.11 kcal/mol. The addition of the solvent model stabilises the HT conformation greatly, to give a near comparable energy to that of the HH conformation.

These numbers offer a slight difference to the values shown in the scan profile in figure 5.2.1.2, which shows the minimum after the barrier to be more stable than the starting conformation; however the scan was carried out with PM6 without the dispersion correction, so some difference can be expected, and the two methods using the COSMO model show that HH and HT are very close in energy.

The two PM6-DH2 optimised structures were then further optimised in the gas phase, using the dispersion corrected DFT functional B97-D, employing the def2-TZVP basis set, with ECP on the platinum atom, and enabling resolution of identity (RI) within Turbomole v5.10. The optimised structures, at this level of theory, of HH and HT are shown in figure 5.2.1.4, from two perspectives. Selected geometrical parameters for both the semi-empirical and DFT structures are reported in table 5.2.1.2, along with the structural parameters for a cisplatin GpG adduct in a close to head to head conformation, from the Cambridge structure database,[5] which is included for comparison, to ensure that the methods used give suitable structures when compared to crystal structures of analogous complexes.



HH HT

Figure 5.2.1.4 B97-D optimised structures of HH and HT conformations, from two perspectives. Sugar phosphate backbone, in both, and (cis-1,4-DACH)Pt, in the lower diagrams, shown as wireframe for clarity.

Table 5.2.1.2 Geometrical parameters for optimised HH and HT structures using B97-D and PM6-DH2, and for the experimental crystal structure (Å or °).

Parameter	B97-D HH	B97-D HT	PM6-DH2 HH	PM6-DH2 HT	Cisplatin HH Expt.
Angle between guanines[a]	85.5	61.79	85.04	59.83	81.20
1-2-3-4 dihedral angle	41.12	-48.36	47.68	-37.33	69.21
N(DACH)-Pt-N(DACH) angle	97.00	96.32	90.39	90.81	91.70
N(d(GpG))-Pt-N(d(GpG)) angle	92.02	89.95	89.19	87.46	88.30
Pt-N(DACH) bond length[b]	2.08	2.07	2.05	2.04	2.05
Pt-N(d(GpG)) bond length[b]	2.08	2.08	1.98	1.99	2.00

[a]Angle between the planes of the six-membered rings of the two guanines.

[b]Bond lengths reported are average of the two present.

The geometrical parameters reported in table 5.2.1.2 show that there is little difference between the optimised structures using both methods. The platinum-nitrogen bond lengths are shorter for the PM6-DH2 optimised structures, but not by a significant amount, and the angle between the planes of the guanines are in good agreement. The dihedral angle 1-2-3-4 and the N-Pt-N angles differ a little more, but overall the structures by both methods are in reasonable agreement. For comparisons with the geometrical parameters of the crystal structure and the B97-D HH conformation, the bond lengths are in reasonably good agreement, as well as the angles between the two guanines. The N-Pt-N angles differ slightly more as well as the dihedral angle, but as the crystal structure has different ligands on the platinum this is not unexpected. Overall though, B97-D and PM6-DH2 give reasonable structures when compared to each other, so using the semi-empirical method as a starting point for calculations can be justified, and B97-D performs suitably well for the HH conformer when compared to crystal structure data of analogous complexes.

To compare the effect of solvent models on the energies of the two conformations, calculated with B97-D, single point solvent corrections were carried out on the optimised gas phase structures using the COSMO model with water as the solvent. The continuous cavity option, within Turbomole, was tested and for the level of accuracy reported was found to have no difference on the results and therefore it was not used. The relative energies in the gas phase and with solvent corrections are reported in table 5.2.1.3.

Table 5.2.1.3 Relative energies between conformations of HH and HT (cis-1,4-DACH)Pt(d(GpG)) calculated using B97-D and with COSMO solvent correction (kcal/mol).

	B97-D relative energy	B97-D with COSMO
HH	3.63	8.65
HT	0.00	0.00

The B97-D relative energies in table 5.2.1.3 differ greatly to the analogous PM6-DH2 values. For both the gas phase and solvent, B97-D reports that HT is the more stable conformer. However addition of the solvent model in this case also stabilises HT by approximately 5 kcal/mol. This implies that instead of making the two conformers closer in energy, as thought for the PM6-DH2 case above, it actually stabilises the HT conformer by a greater amount than HH.

To further examine the stabilising effect of the implicit solvent model analogous calculations were carried out using cisplatin and oxaliplatin in place of kiteplatin. Starting points were constructed from the truncated PDB entry 1PGC used to construct the 1,4-DACH adduct. For oxaliplatin no modification, other than the truncation of the DNA, was needed, whilst for cisplatin the cyclohexane of oxaliplatin was removed by hand, and protons were added to the nitrogens to reach valency, whilst retaining the positions of the nitrogen and platinum atoms.

Relaxed scans were carried out on the cisplatin and oxaliplatin adducts in the same manner as above, using the PM6 optimised structures of the modified PDB file as a start point, and rotating around the same dihedral angle. The HH and HT conformers were then optimised using PM6-DH2 and B97-D, with single point COSMO corrections. The

relative energies for cisplatin are reported in table 5.2.1.4 and the oxaliplatin values are reported in table 5.2.1.5.

Table 5.2.1.4 Relative energies between conformations of HH and HT for a cisplatin GpG adduct calculated using PM6-DH2 and B97-D, with COSMO solvent corrections for both (kcal/mol).

	PM6-DH2	PM6-DH2 with COSMO	B97-D	B97-D with COSMO
HH	5.87	7.90	0.00	7.59
HT	0.00	0.00	0.07	0.00

Table 5.2.1.5 Relative energies between conformations of HH and HT for a oxaliplatin GpG adduct calculated using PM6-DH2 and B97-D, with COSMO solvent corrections for both (kcal/mol).

	PM6-DH2	PM6-DH2 with COSMO	B97-D	B97-D with COSMO
HH	6.78	11.64	3.15	9.07
HT	0.00	0.00	0.00	0.00

From the data for cisplatin and oxaliplatin in the tables above it is clear they both follow a similar trend to the kiteplatin adducts. Upon inclusion of a solvation model the HT conformer is stabilised more than HH, in all cases. This was further validated by the differences in energy between the gas phase and solvated structures as shown in table 5.2.1.6, which shows that the HT conformer is more stabilised by between 2 and 8 kcal/mol when solvated compared to the HH conformer. Compared to the kiteplatin data, PM6-DH2 gives a larger difference between the relative energies of the two conformers for cisplatin and oxaliplatin. Overall the B97-D values with COSMO corrections are in reasonable agreement, regarding relative energies, for all three platinum complexes.

Table 5.2.1.6 Difference in energy between gas phase and solvated HT conformers minus the difference in energy between gas phase and solvated HH conformers for the three platinum drugs at the PM6-DH2 and B97-D levels (kcal/mol)

	PM6-DH2	B97-D
Cisplatin	-2.03	-7.66
Oxaliplatin	-5.06	-5.92
Kiteplatin	-4.94	-5.02

For the chosen kiteplatin adduct (cis-1,4-DACH)Pt(d(GpG)) there are two more conformations than the ones mentioned above, as it is possible to have two different head to head and head to tail orientations, through further rotation about the dihedral angle 1-2-3-4, as shown in figure 5.2.1.5.

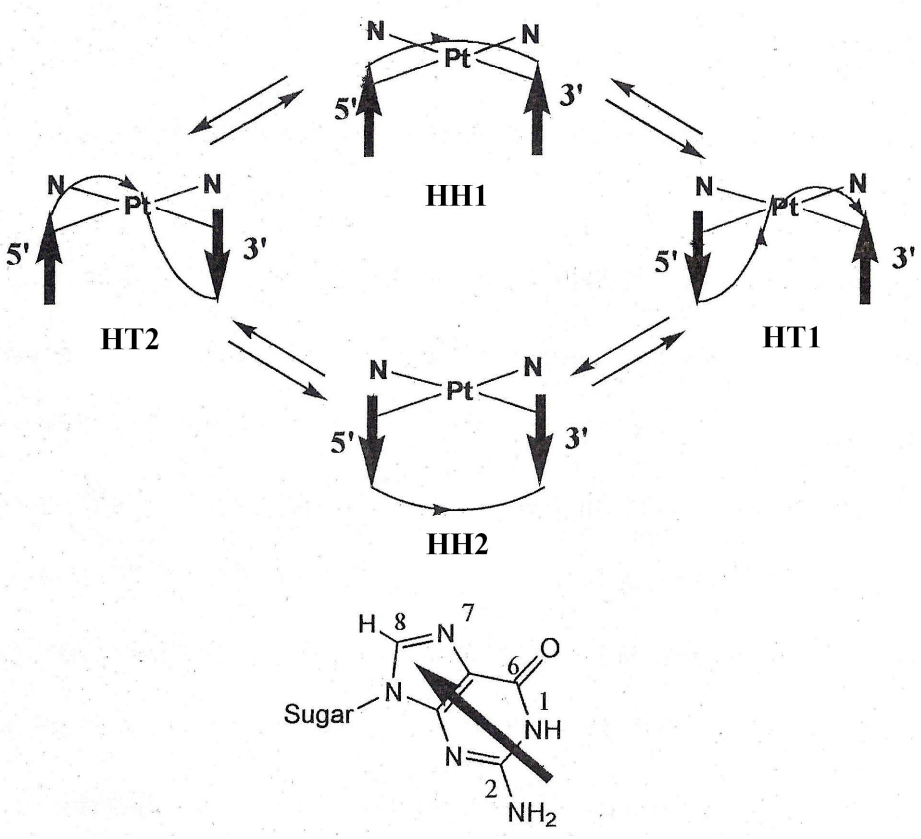


Figure 5.2.1.5 The four conformations of the (cis-1,4-DACH)Pt(d(GpG)), arrows represent the guanine bases, while the curved line linking the arrows represents the backbone.[6]

The HH1 conformer in figure 5.2.1.5, is the one that is most representative of B-DNA and this structure is equivalent to HH used in all calculations above and from here on will be designated HH1. The head to tail conformations adapt lambda/delta type isomerisation, with the Λ isomer the one used in calculations above and represented by HT1 in figure 5.2.1.5.

To obtain the second head to head conformation, HH2, another relaxed scan was carried out using PM6, this one using the HT1 B97-D optimised structure as the starting point, and once again rotating around the dihedral angle 1-2-3-4, shown in fig 5.2.1.1. From visual observation of the structures at the minima on the scan, the HH2 conformer was able to be indentified. Several other minima present on the scan, were identified as structures where the guanines adopted a conformation between the head to head and head to tail, where they were close to perpendicular. To obtain a structure for the ΔHT isomer, HT2, the same scan was carried out as for finding HT1, but the dihedral angle was increased by increments of 5 ˚ instead of decreased. From visual observation of the structures represented in the scan, the desired conformer, HT2, was able to be discerned.

The two structures obtained from the scans were then further optimised using PM6-DH2, and single point COSMO corrections were carried out at the optimised geometry. These optimised coordinates were then used as starting points for B97-D optimisations, which also used COSMO corrections at the optimised geometries. Relative energies of each method, for the four conformers, are reported in table 5.2.1.7 below.

Table 5.2.1.7 Relative energies between conformations of HH1, HT1, HH2, and HT2 (cis-1,4-DACH)Pt(d(GpG)) calculated using PM6-DH2 and B97-D, with COSMO solvent corrections for both (kcal/mol).

	PM6-DH2	PM6-DH2 with COSMO	B97-D	B97-D with COSMO
HH1	0.00	0.00	3.63	8.65
HT1	5.05	0.11	0.00	0.00
HH2	18.09	6.51	5.34	9.08
HT2	18.89	12.48	8.52	13.38

From the PM6-DH2 optimisations of HH2 and HT2, it is shown that they are significantly less stable than the HH1 conformer. Addition of the solvent model results in the same pattern as the gas phase, with the order of stability for the PM6-DH2 method HH1 > HT1 > HH2 > HT2. The energies calculated at the B97-D level in the gas phase show less disparity between the previous structures, HH1/HT1, and the new structures HH2/HT2. Although solvent corrections to PM6-DH2 lower the relative energies closer to the most stable conformer, HH1, for B97-D the three structures get a smaller change in energy compared to the relative value of the most stable HT1. At the B97-D level, the more accurate of the two methods, the order of stability is HT1 > HH1 > HH2 > HT2.

Table 5.2.1.8 B97-D binding energies of kiteplatin to four GpG conformations corrected for BSSE using counterpoise (kcal/mol).

	Gas Phase Binding Energy	COSMO Binding Energy
HH1	-238.41	-125.28
HT1	-230.29	-124.61
HH2	-228.61	-121.36
HT2	-235.22	-121.20

Binding energies for kiteplatin to GpG were calculated for all four conformations. This was carried out using B97-D, basis set superposition error (BSSE) was corrected for using the counterpoise method, values are reported for the gas phase and with the inclusion of the COSMO model in table 5.2.1.8. Counterpoise correction was not used for similar calculations on ruthenium complexes carried out on the larger base pair steps in chapter 3 as the computational expense meant some calculations were unable to be completed and for the ones that did the correction was deemed negligible. However, for these smaller adducts the computational time is significantly less so it was possible to obtain values for all systems here. The binding energies show that there is no great difference between the four structures, with an overall difference of less than 10 kcal/mol between them. The two head to head conformations have the largest and smallest values, this indicates that the geometry of the guanines and sugar phosphate backbone around 1,4-DACH does not have a significant effect on the binding energies. With the solvent model the difference between the conformers is reduced with a difference between greatest and smallest of 5.08 kcal/mol. The platinum complex remains most strongly bound to HH1, but the overall trend changes with HT2 becoming

the most weakly bound complex. The binding energies with COSMO corrections show HH1 and HT1 are clearly the more favoured conformers in accordance with relative energies shown in table 5.2.1.7.

5.2.2 Single Strand TGGT

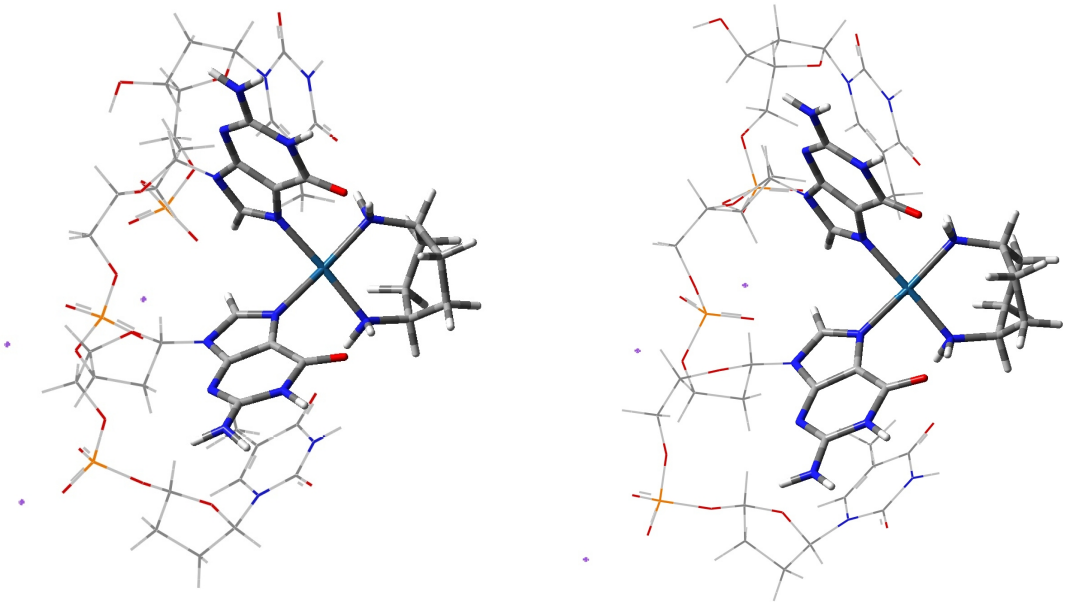
GpG is a very small fragment of DNA, to get a better insight into how kiteplatin binds to DNA, larger adducts are needed. Another single strand for which experimental data has been collected by our collaborators, is the TGGT adduct, where the single strand is expanded upon by including a terminal thymine on both ends of the DNA. From the B97-D optimised (cis-1,4-DACH)Pt(d(GpG)) in the HH1 conformation, thymines were added by hand. To give a reasonable starting point for further optimisations, the system was optimised using PM6-DH2, but with the platinum, DACH, and the two coordinated guanines frozen to allow the thymines and sugar phosphate backbone to relax. From this structure full optimisation was carried out using PM6-DH2, the resulting structure was then further optimised using B97-D, selected geometrical parameters are reported in table 5.2.2.1, and optimised structures are shown in figure 5.2.2.1.

Table 5.2.2.1 Geometrical parameters for optimised TGGT kiteplatin adducts in the HH1 conformation using B97-D and PM6-DH2 (Å or °).

Parameter	B97-D	PM6-DH2
Angle between guanines[a]	87.42	82.16
1-2-3-4 dihedral angle	49.82	41.44
N(DACH)-Pt-N(DACH) angle	97.43	96.07
N(d(TGGT))-Pt-N(d(TGGT)) angle	90.27	92.04
Pt-N(DACH) bond length[b]	2.07	2.07
Pt-N(d(TGGT)) bond length[b]	2.06	2.07

[a]Angle between the planes of the six membered rings of the two guanines.

[b]Bond lengths reported are average of the two present.



B97-D PM6-DH2

Figure 5.2.2.1 B97-D and PM6-DH2 optimised structures of HH1 conformation of TGGT kiteplatin adducts. Sugar phosphate backbone shown as wireframe for clarity.

The geometrical parameters in table 5.2.2.1, and visual observation of the structures in figure 5.2.2.1, further confirm the conclusions above that PM6-DH2 gives suitable structures compared to the B97-D method. A relaxed PM6 scan was attempted in the same fashion as used above for the GpG adduct, in order to get a structure for the head to tail conformation. Due to the increased rigidity of the structure, by the presence of an additional two nucleobases, it was not possible to find a second conformation, at the point where changing the dihedral brought the guanine too close to the adjacent thymine the structure returned to the starting head to head conformation.

Optimised cisplatin and oxaliplatin TGGT structures were also obtained using the PM6-DH2 and B97-D methods. Using the coordinates of the kiteplatin complex where the thymines and sugar phosphate backbone were allowed to relax while everything else was frozen, the 1,4-DACH Pt part was converted into the corresponding cisplatin and oxaliplatin parts by hand. These structures were then optimised fully using PM6-DH2 and B97-D. Binding energies were calculated, in the gas phase and with the COSMO model representing water, for all three platinum complexes to the TGGT DNA fragment using the counterpoise method at the B97-D level, values are reported in table 5.2.2.2.

Table 5.2.2.2 B97-D binding energies of cisplatin, oxaliplatin, and kiteplatin to TGGT, corrected for BSSE using counterpoise (kcal/mol).

Platinum Complex	Gas Phase Binding Energy	COSMO Binding Energy
Cisplatin	-307.22	-135.51
Oxaliplatin	-280.82	-135.81
Kiteplatin	-278.95	-134.18

Cisplatin has the largest gas phase binding energy to the TGGT DNA fragment, approximately 27 kcal/mol more than both oxaliplatin and kiteplatin, which of both have very similar values, with a difference less than 2 kcal/mol. With the addition of a solvent model the difference between the binding energies becomes significantly smaller, with an overall difference of less than 2 kcal/mol between all of them. Kiteplatin is still the weakest, but the trend changes with oxaliplatin becoming the most strongly bound, with a binding energy of 0.3 kcal/mol greater than cisplatin, although this difference is small and within the likely error of the calculations so it would be difficult to conclusively say that one has a greater binding energy than the other.

5.2.3 Comparison of Single Strand Adducts

A comparison between the two single strand kiteplatin DNA adducts, GpG and TGGT has been carried out, to ascertain whether lengthening the strand, by addition of two extra thymine nucleobases, has a significant effect on the geometry around the metal and coordinated bases and the binding energies. A comparison of selected geometrical parameters is reported in table 5.2.3.1 and a graphical comparison of binding energies is reported in figure 5.2.3.1.

Table 5.2.3.1 Comparison of geometrical parameters for optimised GpG and TGGT kiteplatin adducts in the HH1 conformation using B97-D (Å or °).

Parameter	GpG	TGGT
Angle between guanines[a]	85.5	87.42
1-2-3-4 dihedral angle	41.12	49.82
N(DACH)-Pt-N(DACH) angle	97.00	97.43
N(d(TpGpGpT))-Pt-N(d(TpGpGpT)) angle	92.02	90.27
Pt-N(DACH) bond length[2]	2.08	2.07
Pt-N(d(TpGpGpT)) bond length[b]	2.08	2.06

[a] Angle between the planes of the six membered rings of the two guanines

[b] Bond lengths reported are average of the two present

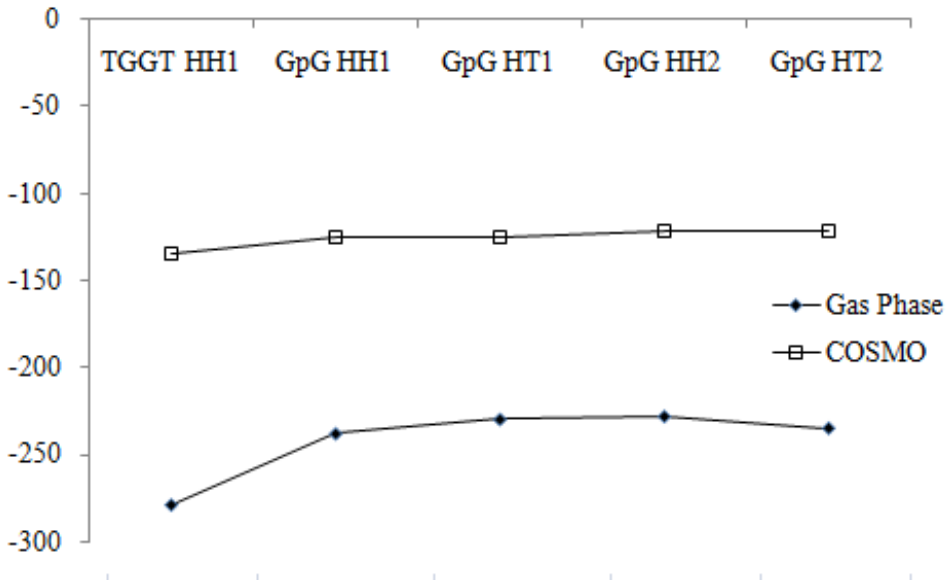


Figure 5.2.3.1 Comparison of B97-D binding energies of kiteplatin to TGGT and the four GpG conformations in the gas phase and with COSMO solvent model (kcal/mol).

Comparing the geometrical parameters shows that both structures are very similar around the coordination sites and for the guanines. This can also be confirmed from the structures shown in figures 5.2.1.4 and 5.2.2.1. The addition of thymines changes the plane angle between the guanines by an insignificant amount of 2 °, this may have been expected to be the parameter with the greatest difference between the two structures due to the extra nucleobases causing the angle between the guanines to reduce through sterics. Although as this is only a single strand adduct, there are no interstrand interactions holding the structures together and therefore the thymines have more

freedom of movement, so form stacking interactions with the guanines, which maintain a similar formation as in the smaller GpG adduct. The angles and bond lengths around the platinum differ negligibly between the two structures; the dihedral angle 1-2-3-4 has the largest difference, but is still small enough to be considered negligible, especially in light of the small difference between the plane angles.

The gas phase binding energy of kiteplatin to the TGGT strand is approximately 40 kcal/mol more than the analogous GpG conformer, this is attributed to additional favourable electrostatic interactions arising from the addition of two more negatively charged phosphate groups. Positive sodium counter-ions have been included in the system, but it is assumed they will not significantly negate this extra stability due to their relatively long range from the platinum. Inclusion of the solvent model decreases this difference to approximately 9 kcal/mol, this can be expected as the electrostatic potential from the COSMO model will screen the favourable interactions between the differing charged parts of the complex.

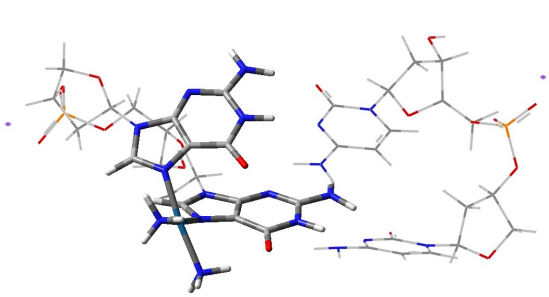
5.3 B97-D Calculations on Dual Strand DNA Platinum Adducts

5.3.1 GpG·CpC Base Pair Step

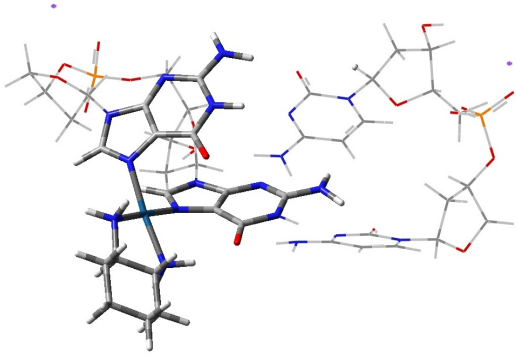
The original GpG adduct was also expanded on by increasing from single strand to dual strand DNA, to make a base pair step by the addition of CpC group. From the PDB entry 1PGC containing the analogous platinum drug oxaliplatin bound to a DNA duplex, the desired base pair step was extracted and the back bone was terminated at the 5’ end with a hydrogen on the C4’ atom and terminated at the 3’ end with an OH group on the C3’ atom. The oxaliplatin group was then converted to DACH and cisplatin, by hand, to give starting coordinates for GpG·CpC adducts of the three platinum drugs. Optimisations were carried out using PM6-DH2 and the resulting structures were further optimised with B97-D. Instead of the geometrical parameters used for the adducts in section 5.2.1 and 5.2.2 above, x3DNA analysis has been carried out to get base pair parameters, this analysis cannot be carried out for single strand DNA, values are reported in table 5.3.1.1 along with corresponding B-DNA values, and optimised structures shown in figure 5.3.1.1.

Table 5.3.1.1 Base pair parameters for three platinum drug adducts with GpG·CpC, along with standard B-DNA values for GG base pair step (Å or °).

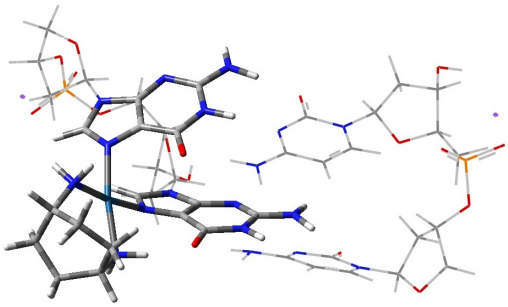
	Shift	Slide	Rise	Tilt	Roll	Twist
Cisplatin	-1.39	-1.90	3.51	1.07	28.14	30.44
Oxaliplatin	-1.37	-1.93	3.54	1.03	28.06	31.16
Kiteplatin	-1.39	-1.78	3.46	1.11	27.88	28.69
B-DNA	0.01	0.47	3.36	-0.03	1.71	35.96



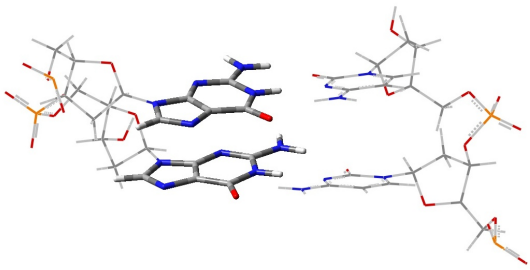
Cisplatin



Oxaliplatin



Kiteplatin



B-DNA

Figure 5.3.1.1 B97-D optimised structures of the three platinum drug adducts with GpG·CpC, along with B-DNA structure obtained from w3DNA.[7] Sugar phosphate backbone shown as wireframe for clarity.

The x3DNA data in table 5.3.1.1 gives an indication of the distortion of the base pair step through coordination of the platinum anti-cancer drugs. The roll values are where the largest difference, between the B-DNA and platinum adducts lie, with a significant increase from 1.71 ° to approximately 28 °. Coordination of the platinum to the two N7 sites on the adjacent guanines causes the roll angle to increase to avoid the huge strain if

they remained nearly co-planar, as they exist in standard B-DNA. The rise distance is fairly consistent across all four systems, with a slight increase for the platinum adducts, their large roll angle means a large increase in rise can be avoided. There are relatively minor differences between the platinum adducts and the B-DNA for the remaining parameters. All three platinum complexes give very comparable results and this shows that the base pair geometry is distorted by equivalent amounts for them.

While the x3DNA calculations give good representations of the geometry around the nucleobases, it is less conclusive regarding the geometry of the sugar phosphate backbone. While the x3DNA values are comparative for the three platinum adducts, observation of the structures in figure 5.3.1.1 shows that there is some disparity between the backbones of the strands that have the coordinated guanines, this is most clear between the cisplatin and oxaliplatin optimised structures. To quantify this torsional parameters of the backbone were obtained, as shown in figure 5.3.1.2, values are reported in table 5.3.1.2 along with average values for standard B-DNA.

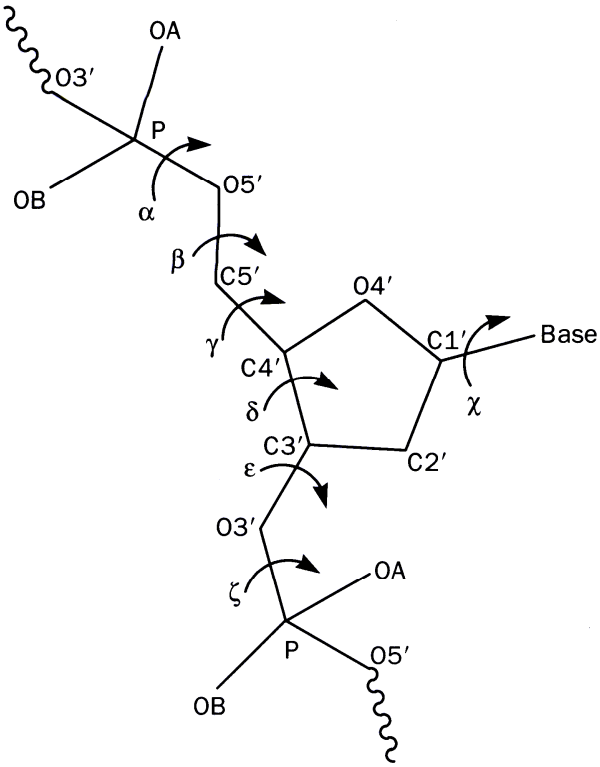


Figure 5.3.1.2 Torsional parameters of sugar phosphate backbone.[8]

Table 5.3.1.2 Torsional parameters for coordinated strand of cisplatin, oxaliplatin, and kiteplatin adducts, using B97-D, along with standard values for B-DNA (˚).

Parameter	B-DNA	Cisplatin	Oxaliplatin	Kiteplatin
α	-40.00	-75.44	-75.27	-65.52
β	-150.00	-163.29	-163.71	-175.24
γ	30.00	47.97	47.61	56.60
χ^a	-95.00	-46.87	-55.40	-32.13
χ^b	/	9.60	-15.38	19.41
δ	150.00	127.57	127.17	137.28
ε	150.00	167.31	-176.55	153.66
ζ	-100.00	-86.18	-86.11	-92.63

[a]This value represents this parameter for the pentose sugar at the 3'end.

[b]This value represents this parameter for the pentose sugar at the 5' end, due to distortion of the DNA adduct through coordination of platinum the two ends give different values, which is not seen in standard B-DNA.

Table 5.3.1.2 contains the calculated torsional parameters of the sugar phosphate backbone of the DNA strand that also contains the coordinated platinum complexes. From visual observation of figure 5.3.1.1 the cisplatin and oxaliplatin structures appear to yield different backbone structures, comparatively the kiteplatin structure is similar to the cisplatin structure. The large differences between oxaliplatin and the other two complexes lie within the ε parameter and the second χ parameter. The difference and opposite sign for the ε dihedral, likely affects the second χ dihedral, as they are both related toward the 5' end of the DNA strand. This difference is probably a result of the ε parameter having more rotational freedom in these adducts as it is close to the terminus of the strand. If the equivalent internal parameter was measured in larger DNA strands this effect might not be observed, due to increased rigidity from increasing the length of the backbone. The remainder of the parameters are in reasonable agreement with each other and with the standard B-DNA values, with the moderate differences that are a result of distortion through coordination of platinum.

Binding energies were calculated for the three complexes, at the B97-D level, using the counterpoise method, to compare to the energies for the single strand adducts in table 5.2.2.2 and to observe whether this distortion has a significant effect on the values, or in this case can be largely ignored.

Table 5.3.1.3 Binding energies calculated for cisplatin, oxaliplatin, and kiteplatin to GG base pair step using B97-D and corrected for BSSE using counterpoise (kcal/mol).

	Gas Phase Binding Energy	COSMO Binding Energy
Cisplatin	-283.90	-129.58
Oxaliplatin	-249.98	-122.60
Kiteplatin	-249.83	-121.97

The binding energies of oxaliplatin and kiteplatin are comparable with a difference of 0.15 kcal/mol and 0.63 kcal/mol for the gas phase and COSMO respectively. These two Pt drugs are isomers of each other and the binding energies show that the comparative distortion of the oxaliplatin has little significance. The binding energies calculated for TGGT are in agreement with this as in that case there is little difference between the binding energies of the two complexes. The trend, for all three drugs, is in clear agreement for both the gas phase and with solvent, with cisplatin the most strongly bound and kiteplatin the most weakly bound. As seen above the COSMO binding energies show a smaller range than gas phase, with the difference between weakest and strongest bound changing to 7.61 kcal/mol from 34.07 kcal/mol. The gas phase trend is in good agreement with those calculated for the single strand TGGT, but the COSMO trend differs, with this case having better agreement with gas phase. Also the cisplatin binding energy remains clearly the strongest for COSMO on GG, while for TGGT it is brought into a closer similarity with the other two drugs.

5.4 ONIOM Calculations on Dual Strand DNA Adducts

So far in this chapter DFT calculations have been carried out on DNA platinum adducts and the entire molecule in question has been treated at the same level. The calculations carried out using B97-D in section 5.3.1 and on the base pair steps in chapter 4 are the largest systems within this work that have been achieved. Even with the addition of resolution of identity the calculation times could take several weeks, which makes this approach somewhat impractical when wanting to increase the system size beyond this. Larger DNA fragments and systems containing explicit solvent molecules are of particular interest, so the route of multilayer QM/MM techniques has been approached. By treating the areas of the molecule that are of greatest interest with full quantum mechanical methods, in this case the platinum drug and the directly coordinated nucleobases, and treating the remainder of the molecule with lower level molecular

mechanics methods, calculations can be carried out on much larger and therefore more realistic biological systems.

The chosen methodology to carry out these QM/MM calculations has recently been outlined by Gkionis and Platts,[9] for the optimisation of analogous systems containing cisplatin and explicit water molecules, instead of the complex of interest here, kiteplatin. This was achieved using the ONIOM method, with the high layer treated with BHandH/6-31+G**; with SDD basis set and ECP on Pt, and the low layer with the AMBER forcefield. Optimisations were carried out in Gaussian 09 using the GEDIIS algorithm[10] and micro iterations. With the entire system free to move it is difficult to achieve a fully optimised structure, due to large molecules with many degrees of freedom having relatively flat potential energy surfaces. To overcome this, optimisations were carried out in stages, with parts of the molecule frozen while others were free to move. By optimising parts at a time the forces for those can be reduced while maintaining the general structure from the initial construction of the coordinates. If the system has complete freedom from the start undesirable end points can be reached whereby the separate strands of the DNA move long distances apart to reduce high initial forces.

In these calculations the QM layer is the platinum complex and the two coordinated guanines, while the MM layer is the remainder of the DNA fragment being studied, the sodium counterions, and any explicit water molecules present. For calculations on the QM layer the bonds the guanines form with the backbone are replaced with link atoms, which in this case are hydrogens. The first step is to allow just the QM layer and atoms in the MM layer that are part of the base pair step containing the coordinated guanines, including the sugar phosphate backbone; here terminated at the C4' position of the pentose sugar, to relax, while the remainder of the DNA and water is frozen. The second step is to allow just the water to relax, with the entire DNA platinum adduct frozen. The third step is to allow the DNA in the low layer, minus the ones contained in the base pair step that were not frozen in step one, to relax with the water and QM layer frozen. The fourth and final step is to allow the entire system to optimise freely. Step one does not use micro iterations while the remaining three steps do.

5.4.1 Optimisation of Dimer Adduct

The dimer is the base pair step GpG·CpC, that was small enough to be optimised fully using B97-D in section 5.3. Although an optimised structure has already been obtained for this system, another optimisation has been carried out using the ONIOM method with BHandH and AMBER to allow for more accurate comparisons between the series of kiteplatin DNA adducts studied with this method. The optimised B97-D structure was used as the starting point for the optimisation, with the platinum complex and two guanines in the QM layer. The calculation was initially carried out in the gas phase, the partial optimisation steps used above were deemed unnecessary due to the relatively small system size, and due to the issues using microiterations for the tetramer outlined in section 5.4.2, these were not used. The full optimisation was carried out using BHandH and AMBER and the optimised structure is shown in figure 5.4.1.1.

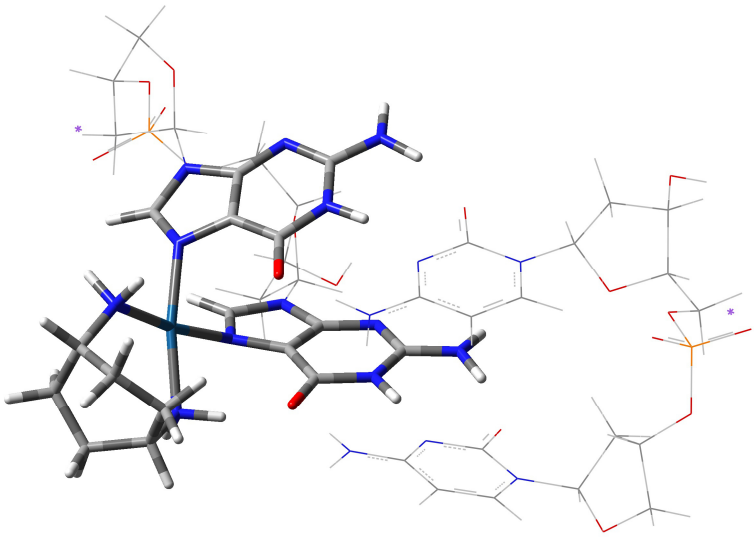


Figure 5.4.1.1 Optimised structure of kiteplatin dimer, optimised using ONIOM BHandH/AMBER. Low layer shown as wireframe for clarity.

An optimisation of this system containing explicit water molecules was also attempted. A water soak was carried out on the gas phase optimised structure using MOE and several optimisation cycles were attempted, but as with the issues for the tetramer below, while the energy was slowly dropping the forces did not appear to be close to convergence, so this process was not further attempted.

If the ability to carry out optimisations using full DFT is available then this is clearly preferable to the hybrid QM/MM approach, therefore it would be expected that the

equivalent system optimised using B97-D in section 5.3 is more accurate than the ONIOM one here. Whilst some small differences would be expected, ideally the two structures should not have huge deviations from each other. Table 5.4.1.1 shows the base pair parameters for the structures optimised using both methods, there are some noticeable differences especially the sign change between shift and tilt, and the relatively large increase of 0.28 Å in rise from the ONIOM structure to the B97-D one. Despite these differences and from a visual comparison between figure 5.4.1.1 and 5.3.1.1, the two methods give broadly similar results in terms of the optimised structure of this system.

Table 5.4.1.1 Base pair parameters of kiteplatin dimer optimised with B97-D and ONIOM (Å or °).

	Shift	Slide	Rise	Tilt	Roll	Twist
ONIOM	1.17	-1.59	3.18	-4.03	21.65	25.37
B97-D	-1.39	-1.78	3.46	1.11	27.88	28.69

5.4.2 Optimisation of Tetramer Adduct

The four nucleobase sequence, the tetramer, has also been studied as an extension from the dimer. The sequence chosen was d(TGGT), as the single strand version was studied above in section 5.2.2. The starting point was constructed from the PDB entry 1AU5,[11] which is a cisplatin octamer adduct. This was truncated to the required sequence, the cisplatin was converted to 1,4-DACH by hand while retaining the positions of the platinum and nitrogen atoms, by removing one of the bound hydrogens off each nitrogen and adding in the cyclohexane ring in a boat like conformation. Sodium counterions were placed in the vicinity of each phosphate group by hand and explicit water molecules were added with a water soak using the MOE program[12] to give approximately one hundred water molecules.

Calculations were carried out in the order expressed above, with the initial optimisation of the QM layer and additional atoms in the base pair step, carried out using the smaller basis set, 6-31G*, with BLYP and a density fitting basis set. This combination of smaller basis set and pure DFT functional with density fitting was used to speed up the initial steps, this was deemed acceptable as any inaccuracies arising would be alleviated in the later steps where the full system is optimised using BHandH/6-31+G**. This step

was successful, but the proceeding attempt at optimising the water molecules was unsuccessful, with upwards of seventy five of the water molecules being ejected from the vicinity of the adduct. The partial optimisation steps were carried out in several different orders in attempt to get the water to optimise to a sensible structure, but these were also unsuccessful, with the same result.

With the water optimisation step failing to give reasonable structures the optimisation was carried out in the gas phase, with all the explicit water molecules deleted. The initial optimisation of the QM layer and base pair step was carried out successfully but the optimisation of the remainder of the MM layer failed to give a sensible structure, with the two strands of DNA splitting. This was attributed to the microiterations as the same step carried out with the option for these turned off, but still using the GEDIIS algorithm, gave a more reasonable structure, whilst it did not fully optimise the energy and forces were decreasing. The attempt at optimisation of the entire system using microiterations failed, but with no microiterations the system fully optimised. The structure was then reoptimised using the desired BHandH functional with the 6-31+G** basis set, which is shown in figure 5.4.2.1.

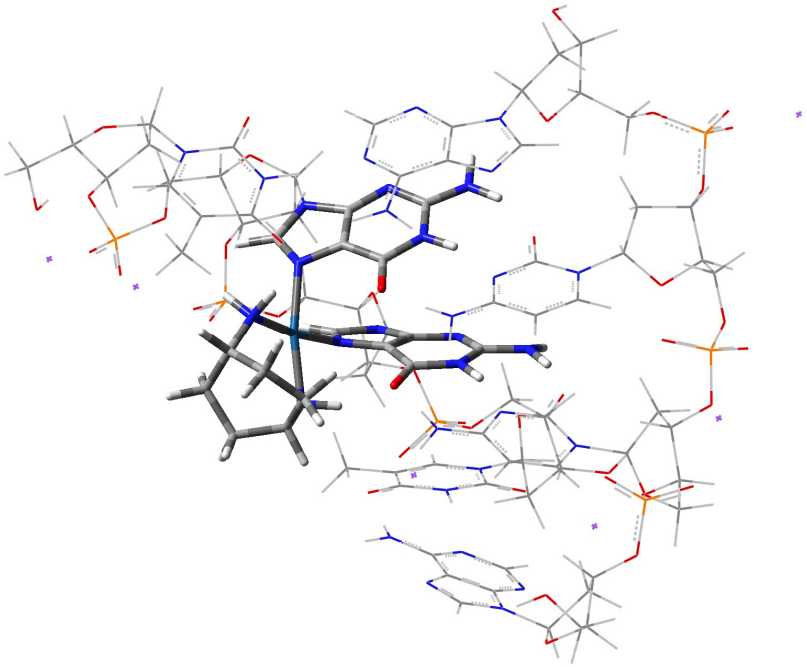


Figure 5.4.2.1 Optimised structure of kiteplatin tetramer, optimised using ONIOM BHandH/AMBER. Low layer shown as wireframe for clarity.

Every attempt at using microiterations for the tetramer adduct, either failed or gave undesirable structures, so it is reasonable to assume from this and that they have worked for some steps in the optimisations of the larger DNA adducts below, that microiterations are not suitable for the smaller systems. Knowing this another attempt at solvating the tetramer was carried out, the optimised structure shown in figure 5.4.2.1 was water soaked using MOE to give one hundred and eight water molecules. The use of microiterations was avoided and the optimisation was attempted entirely in Cartesian coordinates, whilst throughout multiple optimisation steps the energy was continuously dropping, the forces never came close enough to consider the system converged, which was also observed for the dimer. Due to this, after a significant amount of computer time, this attempted optimisation was halted.

Table 5.4.2.1 Base pair parameters for kiteplatin tetramer optimised in the gas phase (Å or °).

	Shift	Slide	Rise	Tilt	Roll	Twist
TG/CA	0.24	-1.98	2.78	9.91	6.28	30.02
GG/CC	0.86	-1.33	2.73	-5.52	19.26	22.89
GT/AC	0.89	-2.50	3.63	6.98	5.65	39.91
Average	0.66	-1.93	3.05	3.79	10.40	30.94

The base pair parameters for this optimised tetramer were obtained using x3DNA and are reported in table 5.4.2.1. As can be seen for the parameters for the octamer in table 5.4.3.1 below, the GG base pair step where the kiteplatin complex is coordinated shows a large deviation from the remaining base pair steps in several parameters. This is most prevalent for the roll parameter, where the platinated bases result in a roll of 19.26 °, significantly greater than the value of 6.28 ° and 5.65 ° for the remaining steps. The twist and tilt parameters also show reasonable decreases for the central base pair step compared to the outside steps, whilst shift, slide, and rise are more comparable across the entire adduct.

5.4.3 Optimisation of Octamer Adduct

The dual strand octamer of the sequence (5'-D(CCTG*GTCC)-3') is one of the systems that is of interest here and is also the one that was studied by Gkionis and Platts[9] when testing this optimisation method. Two initial structures were used for this octamer to test how sensitive the final structure is to the starting point. One was the NMR structure of the cisplatin octamer adduct, with the above sequence, with the PDB entry 1AU5, the second was the optimised structure from reference 9 which also used 1AU5 as its starting point. Both of these structures contain cisplatin, these were converted into kiteplatin by hand. For the starting point constructed from 1AU5, sodium counterions were placed in the vicinity of each phosphate group by hand and a water soak was carried out on the system to give a solvation shell of ninety eight water molecules, using the MOE program.

Once again the pure DFT functional BLYP was used with a density fitting basis set as well as the smaller basis set 6-31G* to increase the efficiency of the initial steps. In some steps loose convergence criterion was also specified. The first step for optimising from the first starting point was the optimisation of the base pair step and the kiteplatin complex, this was successful. With the entire DNA platinum adduct frozen, the water molecules were then optimised with microiterations, an optimised structure was reached but the calculation caused eighteen water molecules to be moved to extreme distances away from the DNA, these were deleted by hand and the modified system was then used in the next step. The MM region was optimised, with the base pair step, QM layer, and water frozen. With all the separate parts relaxed while the remainder of the system was frozen, a full optimisation was attempted, this failed with the resulting structure one where the DNA had split into two strands at distance, the cause of this was thought to be too high forces combined with the use of microiterations. To test this, the failed step was rerun but with the microiterations turned off and the functional and basis set changed to BHandH and 6-31+G**. The resulting calculation exited due to wall time restrictions, but yielded a lower energy structure with lower average forces but also had moved an additional eight water molecules to extreme distances. The eight waters were removed and another optimisation of the entire system, using microiterations, was attempted but this time at the BHandH level and not with loose convergence criterion. This calculation once again failed with a similar result of "unwound" DNA. A further

attempt at getting a lower energy structure was carried out by using an optimisation within Cartesian coordinates, instead of the GEDIIS algorithm.

Several cycles of Cartesian optimisations, followed by attempted optimisations using GEDIIS with microiterations were carried out. The lowest energy structure of the Cartesian optimisations was used as the starting point for the GEDIIS/micro optimisation, if that failed due to not converging in 25 cycles, which is the limit within the Gaussian 09 program, but the structure was sensible, the end point was then used for a further Cartesian optimisation. After three cycles of this the structure optimised fully using microiterations, the optimised structure is shown in figure 5.4.3.1.

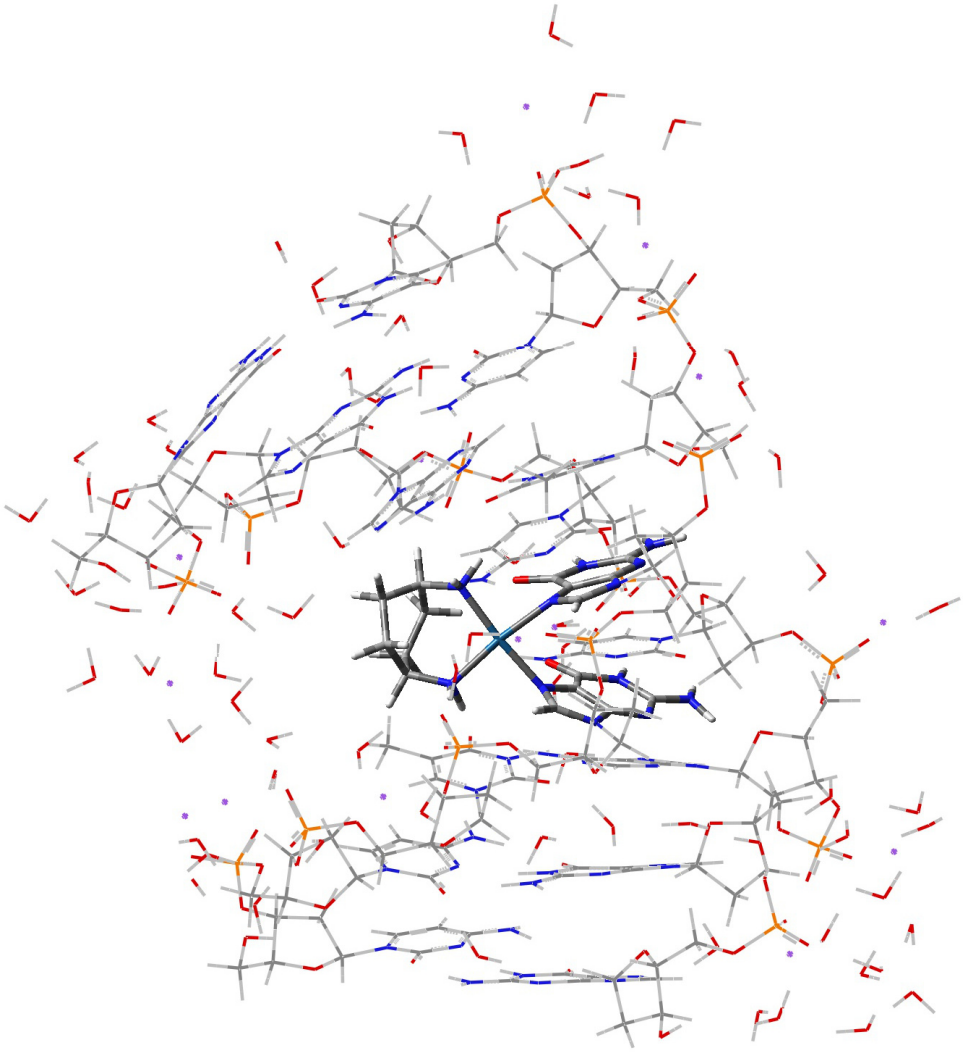


Figure 5.4.3.1 Optimised structure of kiteplatin octamer from 1AU5 crystal structure starting point, optimised using ONIOM BHandH/AMBER. Low layer and water molecules shown as wireframe for clarity.

Carrying out the same calculation but from the second starting point was slightly more straightforward. After the initial steps of optimising sections while freezing others with BLYP, the first attempt at a full optimisation failed, but one use of a BHandH optimisation step in Cartesian coordinates gave a structure that fully optimised using BHandH/AMBER with microiterations and the GEDIIS algorithm. No water molecules were ejected in this optimisation, presumably due to the starting point being based on a structure that had already been optimised with this approach, therefore the waters were already relaxed, so there were no instances where high initial forces caused the waters to move to large distances. A comparison of the extrapolated ONIOM energies for the two optimised structures, with the explicit waters removed, shows that the first starting point gives a structure that is approximately 18 kcal/mol lower in energy. The optimised structure from the starting point in reference 9 is shown in figure 5.4.3.2.

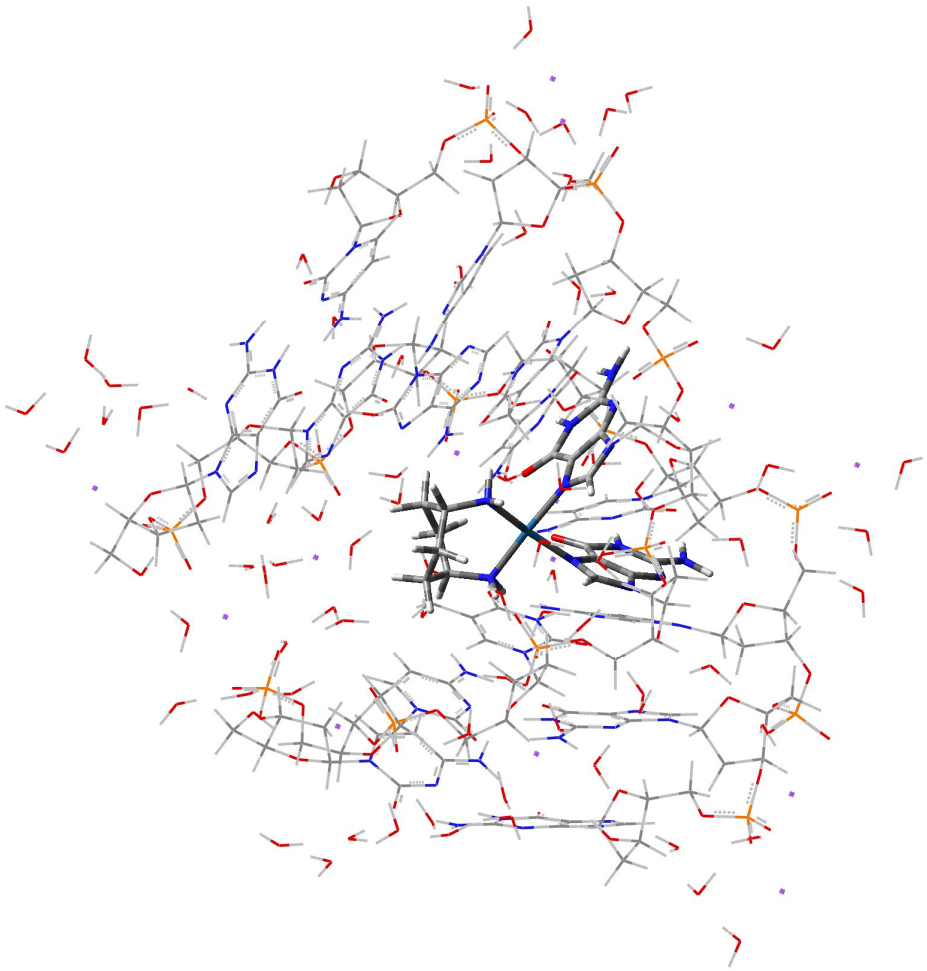


Figure 5.4.3.2 Optimised structure of kiteplatin octamer from reference 9 starting point, optimised using ONIOM BHandH/AMBER. Low layer and water molecules shown as wireframe for clarity.

To compare the two structures in figures 5.4.3.1 and 5.4.3.2, to see how sensitive the optimised structure is to different starting points, base pair parameters have been calculated using x3DNA. These values are reported below in table 5.4.3.1, with the parameters from the second starting point of the optimisation in parenthesis.

Table 5.4.3.1 Base pair parameters for kiteplatin octamer from 1AU5 optimisation starting point, values in parenthesis are from reference 9 starting point (Å or °).

Step	Shift	Slide	Rise	Tilt	Roll	Twist
CC/GG	-0.55	-0.52	3.03	5.38	-2.28	32.37
	(0.30)	(-0.44)	(3.80)	(2.78)	(5.65)	(44.83)
CT/AG	0.66	-0.90	3.76	1.49	-1.09	41.05
	(0.50)	(-1.01)	(3.88)	(1.14)	(4.43)	(37.53)
TG/CA	-1.33	0.31	2.94	-0.27	-5.24	36.34
	(-1.17)	(1.41)	(2.95)	(-1.90)	(5.52)	(33.68)
GG/CC	1.10	-0.99	2.80	-3.04	23.56	25.27
	(0.44)	(0.25)	(4.03)	(-2.05)	(50.38)	(15.47)
GT/AC	1.03	-1.14	3.71	13.94	5.34	47.25
	(0.35)	(-0.25)	(2.68)	(-3.88)	(2.97)	(28.10)
TC/GA	0.39	-0.60	3.38	-3.52	7.78	29.80
	(0.89)	(-0.14)	(3.13)	(8.40)	(2.88)	(43.73)
CC/GG	0.17	-1.99	3.81	3.53	-7.70	35.72
	(0.53)	(-1.79)	(3.76)	(-3.51)	(10.49)	(26.16)
Average	0.21	-0.83	3.35	2.50	2.91	35.40
	(0.26)	(-0.28)	(3.46)	(0.14)	(11.76)	(32.79)

The most important base pair step to examine is that of the central GG, to which the platinum is coordinated as this is likely to show the greatest differences between the two starting points. With platinum complexes that form 1,2-intrastrand crosslinks, as exhibited in the adducts of interest here, the rise and roll parameters are of particular interest, as the platinum attempts to retain the desired square planar geometry the roll angle between the platinated guanines is increased and therefore the rise is increased too. For the starting point from the PDB the roll is 23.56 °, in comparison the second starting point gives a much larger roll of 50.38 °, as can be observed from the two structures in figures 5.4.3.1 and 5.4.3.2. Due to this the rise for the second starting point is also significantly larger at 4.03 Å compared to 2.80 Å for the first starting point.

The structures of DNA adducts that have been elucidated from experimental NMR data, such as 1AU5, usually result in a family of structures that all satisfy proton distances found. As a result of this the base pair parameters can vary for the same adduct. For the PDB entry 1AU5 there is only one structure reported, so to get an experimental range of values Gkionis[9] used the PDB entry 2NQ0,[13] a dodecamer with cisplatin coordinated to a central GG step, which reports fifteen structures, to compare the range of the values of the parameters of the coordinated central step. This data is reported in table 5.4.3.2.

Table 5.4.3.2 Minimum and maximum values for base pair parameters from 2NQ0 PDB structure, from reference 9.

	Minimum	Maximum	Max-Min
Shift	0.67	1.12	0.45
Slide	-2.06	-0.94	1.12
Rise	3.90	4.61	0.71
Tilt	-3.07	-1.39	4.46
Roll	24.44	34.45	10.01
Twist	19.75	28.20	8.45

Although there is significant difference in the central base pair step for the two octamers, observation of the averages of all the base pair parameters for both starting points show that, with the exception of roll, they are for the most part comparable, with relatively small differences in the distances and angles. Along with this, the data in table 5.4.3.2, shows that parameters can actually vary by significant amounts within different experimental interpretations of the same structure. Even though this is the case the part of greatest interest still shows some stark differences, that are greater than the range in experiments, between the two starting points, but as systems of this size have many degrees of freedom it is not necessarily surprising that using two different starting points can result in this. The optimisation from the PDB entry starting point will be used for comparison in section 5.4.5, as it results in the lowest absolute energy of the two.

5.4.4 Optimisation of Dodecamer Adduct

After the successful optimisations of the octamer and tetramer, the significantly larger dodecamer, with the sequence (5'-D(CCTCTG*GTCTCC)-3'), has also been attempted. The starting point was constructed from the corresponding PDB entry 3LPV,[14] with the

desired sequence and a coordinated oxaliplatin to the central GG base pair step. The present cyclohexane ligand was converted into 1,4-DACH by hand and sodium counterions were placed in the vicinity of every phosphate moiety on the DNA backbone, in the place of the mixed sodium and magnesium atoms present in the PDB structure. To solvate the adduct, a water soak was carried out using MOE to give seventy-nine explicit water molecules. This amount of water molecules is perhaps not as many as should be present, especially as the smaller octamer has a larger number, but a small difference in the radii of the water soaks and the removal of close contacts through use of a script, meant either the small amount of water used or a significant amount which would have increased computational time, especially as the water optimisation steps had proved troublesome for the smaller complexes.

The same optimisation method employed above was utilised, initially optimising just the platinum drug and the atoms in the coordinated base pair step, using the faster, less accurate BLYP with the smaller basis set. The explicit water molecules were then optimised, although for this system an initial step using Cartesian coordinates was undertaken to relax the water molecules and avoid the potential issue of the solvation shell 'exploding', as seen for the octamer and tetramer, when this step was carried out with microiterations. The same step using microiterations was attempted from the structure yielded from the Cartesian optimisation, whilst there were no issues arising from molecules being ejected from the solvation shell, the step did not complete due to the microiterations failing to converge. Another cycle of Cartesian optimisation followed by use of microiterations was carried out giving the same result, due to this the structure from the Cartesian step was taken forward to be used in further steps.

The low level was then optimised but with the water molecules frozen and the base pair step to which the platinum is coordinated frozen also, this was initially attempted with microiterations but as for the water set it failed to converge within the maximum set of cycles. Once again optimisation within Cartesian coordinates was carried out instead, several iterations of using Cartesian optimisations followed by microiterations were carried out but every microiteration step failed to fully converge. The lowest energy structure, from the resultant Cartesian step, was used as the starting point for full system optimisation where both layers are free to move, for this step BHandH was used instead of BLYP, with the larger basis set 6-31+G**. The issues presented above, that plagued the partial optimisation steps were still present, with every attempted microiterations

step failing to converge, from a preceding Cartesian optimisation lowest energy structure. Ultimately due to the increasing time taken to carry out this optimisation, after seven restarted Cartesian steps the forces were deemed low enough to consider the structure optimised, with the RMS value of 0.000215 with the threshold at 0.000300 Hartrees/Bohr, and the maximum value at 0.003227 with the threshold at 0.000450 Hartrees/Bohr. The optimised structure is shown in figure 5.4.4.1.

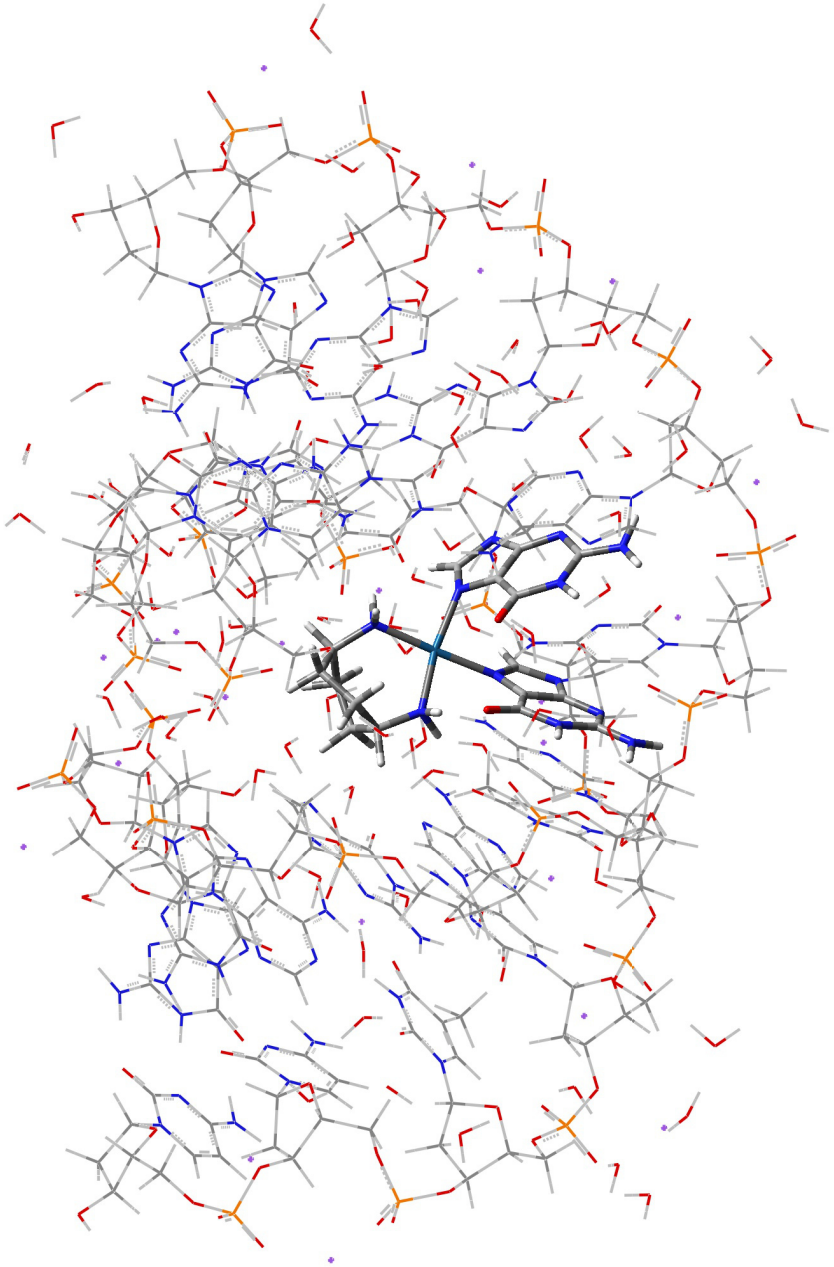


Figure 5.4.4.1 Optimised structure of kiteplatin dodecamer, optimised using ONIOM BHandH/AMBER. Low layer shown as wireframe for clarity.

Table 5.4.4.1 Base pair parameters for kiteplatin dodecamer (Å or °).

	Shift	Slide	Rise	Tilt	Roll	Twist
CC/GG	-0.37	-2.06	3.16	1.23	7.50	26.72
CT/AG	0.20	-1.71	3.30	1.42	5.25	29.99
TC/GA	0.09	-1.95	3.28	-2.12	9.94	32.30
CT/AG	-0.14	-2.01	3.14	-0.25	4.76	31.54
TG/CA	-0.38	-1.38	2.92	4.10	8.12	27.07
GG/CC	1.16	-2.15	3.01	-2.45	34.76	30.22
GT/AC	-0.32	-1.03	3.06	3.12	8.64	33.95
TC/GA	1.11	-0.30	3.33	3.41	10.52	40.70
CT/AG	-0.31	-0.03	2.91	2.73	13.25	22.75
TC/GA	0.32	-1.26	4.34	-8.73	11.40	45.34
CC/GG	0.25	-0.10	2.86	3.11	-4.96	33.03
Average	0.15	-1.27	3.21	0.51	9.93	32.15

The base pair parameters for the optimised dodecamer were calculated and are reported above in table 5.4.4.1. As expected the central GG/CC step, where the platinum is coordinated, shows the greatest distortion compared to the remainder of the helix. This is especially prevalent in the roll parameter, which is significantly larger, at 34.76°, than the other base pair steps, this is to be expected as through coordination, the ideal angle of the two guanines is forced larger to meet with the square planar conformation of the platinum, as can be seen for the other DNA adducts.

5.4.5 Comparison of ONIOM Optimised Adducts

To carry out a comparison between the four adducts optimised using the QM/MM method, several further calculations were undertaken. The binding energy of the platinum complex to the DNA is a good value to compare between different size of DNA fragments, as well as being very important when comparing a series of different drugs to infer which bind the most strongly. Due to this, binding energies were calculated for the platinum DNA adducts in the gas phase and in the aqueous phase, as reported in table 5.4.5.1.

Table 5.4.5.1 Comparison of ONIOM binding energies of kiteplatin to DNA Adducts, in gas phase and with PCM model (kcal/mol).

	Gas Phase Binding Energy	PCM Binding Energy[a]
Dimer	-271.60	-33.78 (-85.10)
Tetramer	-429.79	-186.01 (-95.96)
Octamer	-473.47	-282.42
Dodecamer	-445.50	-277.24

[a]Values in parenthesis are calculated using ONIOMPCM=X, whilst all others are calculated using ONIOMPCM=C.[15]

To calculate the gas phase binding energies above, the explicit water molecules were removed from the systems, where present and three single point energy calculations were carried out, on the platinum complex (1), the remainder of the DNA (2), and the entire adduct (3), where binding energy was calculated by (3)-(2)-(1). While a counterpoise correction would be ideal, in the current iteration of Gaussian it is not possible to do a counterpoise correction on an ONIOM calculation.

An implicit solvent model was chosen, instead of the already present explicit solvent molecules, to avoid the issue of assigning the water molecules in close proximity to the platinum to one of the two fragments. Differing numbers of solvent molecules close to the drugs between the different adducts would mean the binding energies are not directly comparable. For the binding energy in the aqueous phase PCM has been used to model the presence of water, with the use of the ONIOMPCM=C keyword.[15] This option calculates the reaction field for only the real system at the low level whilst all other sub calculations are treated in the gas phase. Whilst this is not the ideal method, it is most efficient for calculations on the larger systems. The ONIOMPCM=X keyword, which calculates the reaction field for all sub calculations, would be the ideal method for this, but was only obtainable for the dimer and tetramer, which are reported in the parenthesis in table 5.4.5.1, calculations using this on the larger fragments failed to complete within the available timeframes.

Within the gas phase binding energies the dimer has the weakest value of -271.60 kcal/mol, there is then a relatively large increase to -429.79 kcal/mol for the tetramer. There are then smaller increases to -473.47 kcal/mol and -445.50 kcal/mol for the octamer and dodecamer respectively. Although the binding energy decreases going

from the octamer to dodecamer, the values still appear to be converging, which is the outcome expected, as beyond a point adding further base pairs will have little effect on the binding energy as the distance between the additional bases and platinum drug increases.

Binding energies in the aqueous phase have a similar pattern as the gas phase, where the dimer is the weakest and the values start to converge as the size of the DNA adducts increase, also the dodecamer in this case once again has a weaker binding energy than the octamer. This may be erroneous due to the dodecamer having not fully optimised, but without carrying out calculations on the decamer or increasing the size of the adducts further to create a larger series of complexes, a conclusion cannot be reached. Although, in this relatively short series, the energies still appear to converge with increasing size, which was the desired outcome.

For the two smallest adducts, the dimer and tetramer, it was also possible to calculate the binding energies with PCM modelling the reaction field for all sub calculations. For the dimer the value calculated this way is significantly larger than using ONIOMPCM=C, increasing to -85.10 kcal/mol from -33.78 kcal/mol. The opposite is the case for the tetramer, where there is a significant decrease to -95.96 kcal/mol from -186.01 kcal/mol. The use of the less accurate ONIOMPCM=C method is likely the cause of the initial value for the dimer, which could be considered lower than expected at -33.78 kcal/mol, compared to those calculated using B97-D reported in table 5.3.1.3.

The exposed surface areas of the platinum and DACH ligand have been calculated using the Molvol program,[16] in the absence of the explicit water and have been reported in table 5.4.5.2 below. As would be expected the dimer has the largest exposed area, at 164.22 $Å^2$, due to having the smallest DNA fragment. As the size of the DNA increases in the tetramer the exposed area decreases to 135.12 $Å^2$, as the larger DNA helix reduces the area of the ligand that is exposed. The octamer has the lowest exposed area at 120.72 $Å^2$, whilst the dodecamer has a slightly higher value at 124.10 $Å^2$, similar to what was observed for the binding energies of these adducts. As above the values start to converge as the size of the DNA helix increases, which is what would be expected as the greater size of the DNA reduces the exposed surface of the drug.

Table 5.4.5.2 Exposed surface area of kiteplatin in DNA complexes (Å^2).

	Exposed Area
Dimer	164.22
Tetramer	135.12
Octamer	120.72
Dodecamer	124.10

A comparison of the base pair parameters of the platinated central GG step, for each of the adducts is reported below in table 5.4.5.3 and represented graphically in figure 5.4.5.1. Some differences are apparent in the parameters between adducts, specifically roll, which ranges from 19.26 ° to 34.76 ° for the tetramer and dodecamer respectively. Although this a relatively large difference the remainder of the parameters are broadly comparable, which concludes that changing the size of the DNA fragment still results in fairly comparable distortion around the coordination site of the platinum drug.

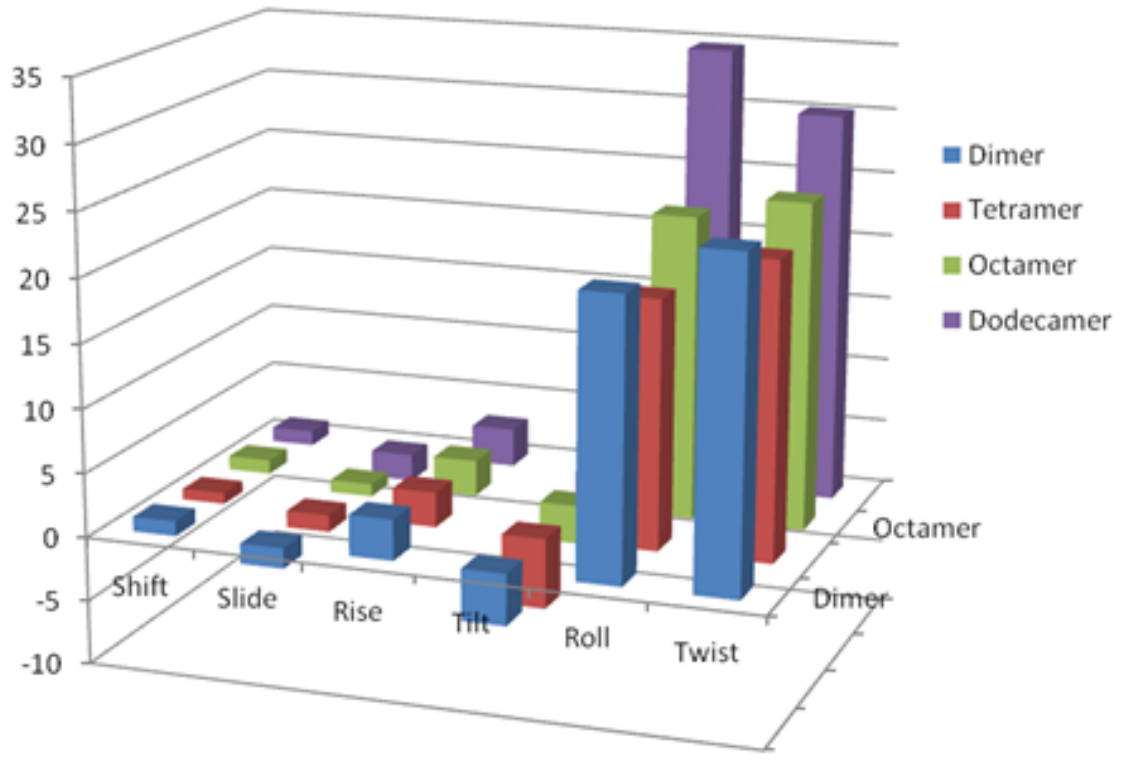


Figure 5.4.5.1 Graphical representation of base pair parameters for the GG base pair step of kiteplatin DNA complexes (Å or ˚).

Table 5.4.5.3 Base pair parameters for the GG base pair step of kiteplatin DNA complexes (Å or °).

	Shift	Slide	Rise	Tilt	Roll	Twist
Dimer	1.17	-1.59	3.18	-4.03	21.65	25.37
Tetramer	0.86	-1.33	2.73	-5.52	19.26	22.89
Octamer	1.10	-0.99	2.80	-3.04	23.56	25.27
Dodecamer	1.16	-2.15	3.01	-2.45	34.76	30.22

5.5 NMR Calculations on Kiteplatin GpG Adducts

One of the most powerful techniques of structure resolution for biological systems is the use of nuclear magnetic resonance (NMR) spectroscopy. Through measurement of chemical shifts and spin-spin coupling constants within single and multi-dimensional NMR, structures of large biological systems can be obtained in situations where it is difficult to obtain crystal structure data. Quantum chemical calculations can be used in conjunction with these techniques, as it is possible to calculate isotropic shielding values for nuclei and coupling constants using DFT and a variety of other methods.

5.5.1 ^{1}H NMR Shift Calculations

To aid experiments being carried out by our collaborators, the calculation of certain NMR values was carried out for the four conformers of the GpG adduct from section 5.2.1. The B97-D optimised structures were used as the starting points the calculation of the ^{1}H isotropic shifts, which were carried out with the BP86 functional with the TZVP basis sets. Calculations have also been carried out with the smaller basis sets SVP and PCS1, as shown by Platts and Gkionis[17] to be effective in modelling changes in ^{1}H NMR shifts due to intermolecular interactions despite their relatively small size: these values are reported in the appendix.

Carrying out NMR calculations gives results as isotropic shielding values for each nuclei, while experimental NMR gives chemical shift values, which are relative to the ^{1}H shielding value of tetramethylsilane (TMS). To get comparable results, calculations were also carried out on TMS, the structure was optimised using B97-D with the def2-TZVP basis set, the same as used for the optimisation of the platinum DNA adducts being studied here. The NMR calculation was carried out within Gaussian 09 using the

gauge invariant atomic orbital (GIAO) method,[18-22] with the BP86 functional and an isotropic shielding value for the protons in TMS was found at 31.34, 31.30, and 31.60 for the basis sets SVP, PCS1, and TZVP respectively. Using the same GIAO method, NMR calculations were carried out on the four GpG adducts, chemical shift values were obtained by subtracting the isotropic shielding values from the one calculated for TMS, results for the TZVP basis set are reported in table 5.5.1.1, with the shifts considered most significant in bold. The trends are very similar for all three basis sets, with differences usually no greater than 0.30 ppm across all shifts.

Only the shifts for the protons located on the guanine rings and sugar phosphate backbone have been reported as the chemical shifts of protons on the 1,4-DACH ligand give little structural information of the DNA, which is the part of greater interest. Some difference in shift would be expected for these protons, as by changing the conformation the local environment changes. The maximum difference across conformers is less than 1.10 ppm and usually significantly less, therefore structural conclusions are difficult to ascertain from this data.

The proton shifts associated with the nitrogen and oxygen atom types can be largely ignored as in experimental conditions solvent exchange leads to peak broadening and it is difficult to get exact and consistent values. In table 5.5.1.1 below, the values that are considered to be most significant to the structure are the protons on the C8 atoms on the guanines and the protons on the C1' and C2' atoms on the pentose sugars. These shifts will be most affected between conformers as they are closest to the part of the complex that changes. A graphical representation of the difference between these shifts is shown for the 3' end and 5' end in figures 5.5.1.1 and 5.5.1.2 respectively.

Table 5.5.1.1 ^{1}H NMR shifts for the protons in the DNA for the four GpG adducts, calculated at BP86 level with TZVP basis set and SDD basis set and ECP on Pt (ppm).

Atom Type[a]	End of DNA	HH1 Shift	HT1 Shift	HH2 Shift	HT2 Shift
Backbone					
C1'	**3'**	**6.34**	**6.11**	**5.93**	**6.43**
C2'	**3'**	**2.15**	**2.66**	**2.29**	**3.43**
C2'	**3'**	**2.67**	**2.21**	**0.61**	**2.68**
C3'	3'	5.56	3.79	4.33	5.02
C4'	3'	3.65	4.82	3.47	4.29
C5'	3'	4.22	2.16	4.27	3.98
C5'	3'	4.55	4.48	4.74	4.06
O3'	3'	2.42	7.08	5.65	0.89
C1'	**5'**	**7.25**	**6.53**	**7.10**	**6.94**
C2'	**5'**	**3.36**	**3.48**	**4.76**	**3.63**
C2'	**5'**	**2.06**	**3.34**	**1.69**	**2.69**
C3'	5'	4.67	4.85	4.89	4.56
C4'	5'	4.47	3.89	4.48	3.90
C4'	5'	4.56	4.54	4.67	4.47
Guanine					
C8	**3'**	**7.83**	**6.48**	**6.33**	**7.54**
N1	3'	6.93	7.14	7.08	7.13
N2	3'	4.78	4.91	4.64	4.85
N2	3'	4.21	4.34	4.21	4.29
C8	**5'**	**7.96**	**6.70**	**7.95**	**6.44**
N1	5'	7.03	7.13	6.46	7.07
N2	5'	4.69	4.79	4.98	4.56
N2	5'	4.18	4.34	3.97	4.21

[a]Atom type is the one which the proton of interest is bonded to, with the numbering system shown in figure 1.2.2.2.

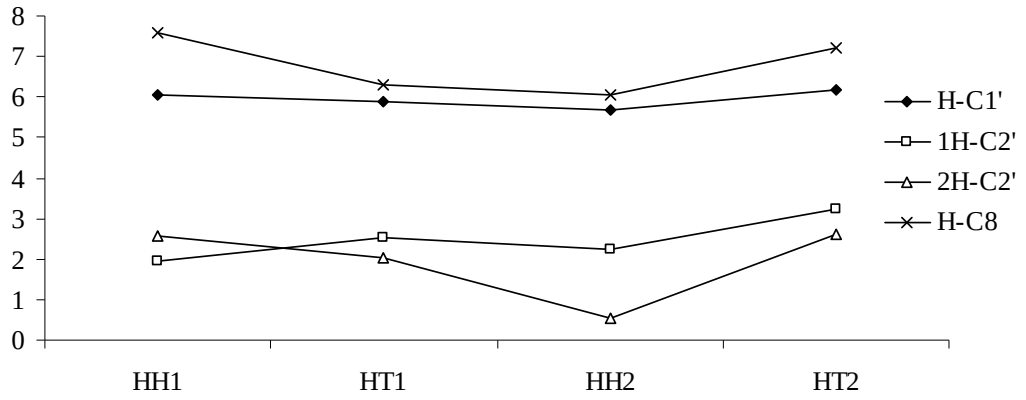


Figure 5.5.1.1 Comparison of ^{1}H NMR shifts at the 3' end of the four GpG conformations (ppm).

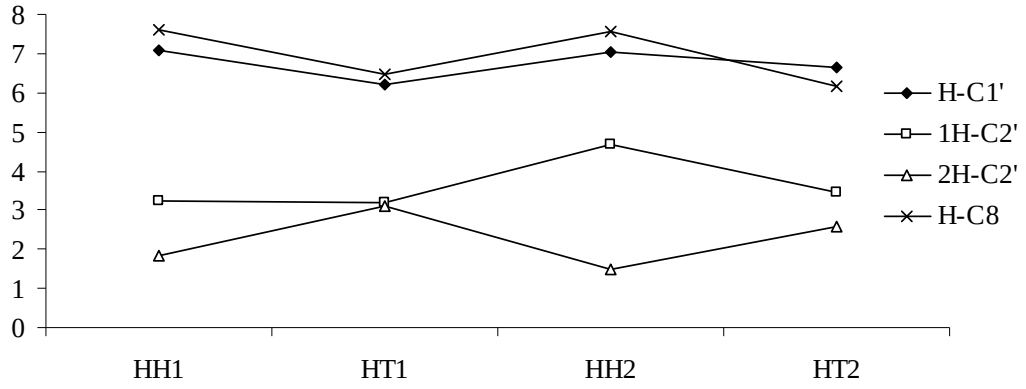


Figure 5.5.1.2 Comparison of ^{1}H NMR shifts at the 5' end of the four GpG conformations (ppm).

The graphs in the figures above show that there is a difference between the proton shifts of the conformers, as expected. At the 3' end, as shown in figure 5.5.1.1, the shift for the proton bonded to the C8 atom is greatest for HH1, then changes by 1.32 ppm for HT1, this difference is also observed at the 5' end but with a smaller change of 1.13 ppm. The shifts for H-C8 and H-C1' have a similar trend at the same ends of the DNA, with the HH2 conformer being the one that differs as at the 3' end its shifts are closer to that of HT1, as opposed to 5' where it is comparable to HH1. The shifts of H-C1' follow a similar trend to those of H-C8, however the exact difference between the conformers is less.

For the two protons bonded to C2' there is a greater difference between the conformers than observed with the C1' and C8, however the trends once again are similar between

the ends of the DNA. To conclude on the significance of this calculated data and how proton shifts change between conformers, comparison to experimental data is ideal and this work is currently being undertaken at Università degli Studi di Bari.

5.5.2 ^{13}C NMR Shift Calculations

Isotropic shielding values for the carbon nuclei are also of interest and have been calculated using the same method as above for the proton values. The SVP basis set is likely too small to give reasonable accuracy for the ^{13}C nuclei, unlike the case for the protons, so the larger basis set TZVP has been utilised. TZVP is still not as big a basis set as would be desirable, but carrying out NMR calculations is very memory intensive and computational costs can increase steeply with large increases in the number of basis functions, so a trade off between accuracy and cost has to be made. The ^{13}C isotropic shielding value for TMS was calculated at 183.77 for the TZVP basis set. ^{13}C chemical shifts with the TZVP basis set are reported in table 5.5.2.1, where once again bold indicates shifts considered most significant.

Table 5.5.2.1 ^{13}C NMR shifts for the carbons in the DNA for the four GpG adducts, calculated at BP86 level with TZVP basis set and ECP on Pt (ppm).

Atom Type	End of DNA	HH1 Shift	HT1 Shift	HH2 Shift	HT2 Shift
Backbone					
C1'	**3'**	**87.66**	**97.15**	**91.18**	**100.91**
C2'	**3'**	**40.67**	**41.83**	**45.07**	**45.54**
C3'	3'	74.05	78.57	75.91	81.80
C4'	3'	95.03	93.14	93.80	101.51
C5'	3'	68.80	71.85	67.14	70.36
C1'	**5'**	**96.16**	**97.32**	**97.58**	**100.49**
C2'	**5'**	**41.35**	**38.67**	**39.41**	**37.63**
C3'	5'	85.92	84.18	85.70	86.94
C4'	5'	79.43	82.97	82.33	83.27
Guanine					
C2	3'	151.84	151.51	151.86	151.65
C2	5'	151.51	151.52	152.88	151.32
C4	3'	152.72	153.00	150.68	151.66
C4	5'	152.82	152.99	152.85	154.76
C5	3'	118.85	118.61	119.41	120.25
C5	5'	119.28	119.65	122.22	123.55
C6	3'	155.65	156.23	155.39	156.09
C6	5'	155.28	155.43	152.52	156.53
C8	**3'**	**139.64**	**139.70**	**136.40**	**140.67**
C8	**5'**	**142.02**	**142.78**	**138.58**	**147.15**

Once again only the shifts of the carbon nuclei within the DNA part of the adduct have been reported. The ^{13}C shifts for the carbon atoms in the DACH ring have a less significant difference than the associated proton shifts, with a difference no greater than 5.00 ppm. While this value is larger than the difference for the proton NMR, the overall range of shifts where carbon nuclei can appear is considerably larger than protons, so a difference of shift of this magnitude can still be considered insignificant. The ^{13}C shifts deemed most significant here are those for C1', C2', and C8, from observation of the other values above there seems to be very little change between the four conformers, with most of the atom types having a change no larger than 5 ppm. A graphical

representation of the difference between the important shifts is shown for the 3' end and 5' end in figures 5.5.2.1 and 5.5.2.2 respectively.

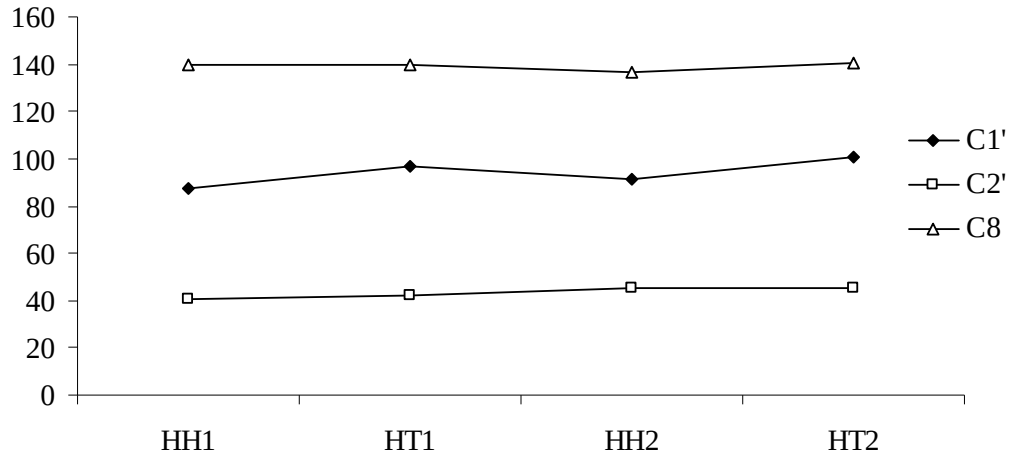


Figure 5.5.2.1 Comparison of ^{13}C NMR shifts at the 3' end of the four GpG conformations (ppm).

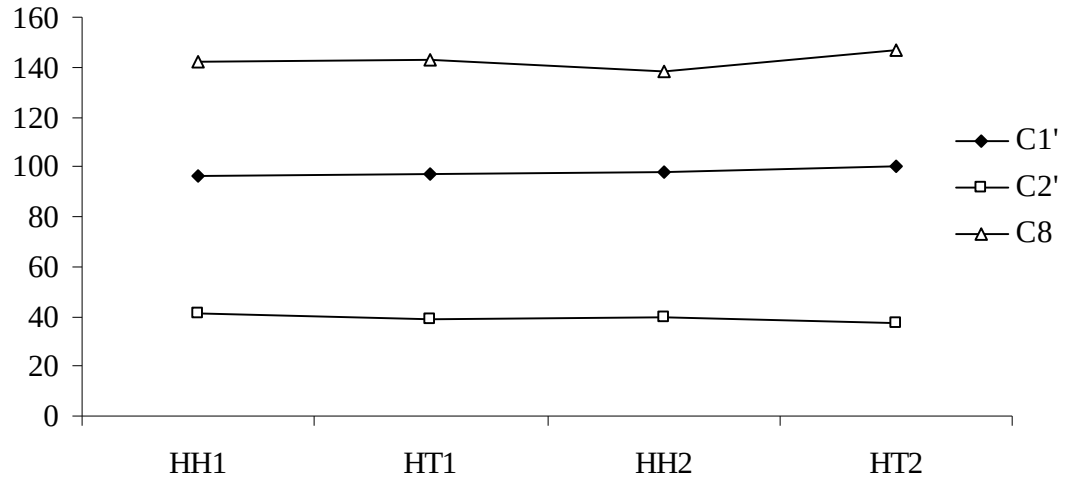


Figure 5.5.2.2 Comparison of ^{13}C NMR shifts at the 5' end of the four GpG conformations (ppm).

For the shifts of the C2' nuclei there is little difference across the four conformers, at both ends of the DNA, so it appears changing conformation has a relatively small effect on the shift at this position. This also appears to be the case for C1' shifts at the 5' end, however there is a greater difference at the 3' end of 13.25 ppm. C8 shifts at the 3' end have little difference between conformers, the 5' shifts have some more variety but there are still no significant changes. From this data it is clear that the shifts associated with carbon nuclei have a much smaller dependence on the conformation than the protons.

Once again to get more conclusive answers this data needs to be compared to corresponding experimental values, which are still in the process of being obtained.

5.5.3 $^{3}J_{HH}$ Coupling Constant Calculations

The indirect dipole dipole coupling, or *J* coupling, values within ^{1}H NMR can be of great use in resolving structures as they can give valuable insights into dihedral angles, if the coupling occurs over three bonds. There is a direct relationship between the ^{3}J coupling values and the dihedral angle of the two nuclei, as shown by the Karplus equation.[23] These values have been calculated for the four GpG adducts using the BP86 functional with the SVP and TZVP basis sets. Results are reported for the TZVP basis set in tables 5.5.3.1, 5.5.3.2, 5.5.3.3, and 5.5.3.4 for HH1, HT1, HH2, and HT2 respectively. The nomenclature H- followed by the atom type to which it is bound has been used, where an atom has two protons they have been labelled 1H and 2H.

Table 5.5.3.1 *J* coupling constants for HH1 conformation, calculated at the BP86 TZVP level (Hz).

3' DNA End							
	H-C1'	1H-C2'	2H-C2'	H-C3'	H-C4'	1H-C5'	2H-C5'
H-C1'							
1H-C2'	8.84						
2H-C2'	1.28						
H-C3'		9.50	7.56				
H-C4'				6.51			
1H-C5'					1.30		
2H-C5'					2.64		

5' DNA End						
	H-C1'	1H-C2'	2H-C2'	H-C3'	1H-C4'	2H-C4'
H-C1'						
1H-C2'	9.11					
2H-C2'	3.53					
H-C3'		5.80	0.35			
1H-C4'				5.74		
2H-C4'				0.96		

Table 5.5.3.2 *J* coupling constants for HT1 conformation, calculated at the BP86 TZVP level (Hz).

3' DNA End							
	H-C1'	1H-C2'	2H-C2'	H-C3'	H-C4'	1H-C5'	2H-C5'
H-C1'							
1H-C2'	5.21						
2H-C2'	0.12						
H-C3'		8.27	6.22				
H-C4'				4.68			
1H-C5'					6.63		
2H-C5'					8.43		

5' DNA End						
	H-C1'	1H-C2'	2H-C2'	H-C3'	1H-C4'	2H-C4'
H-C1'						
1H-C2'	3.76					
2H-C2'	8.63					
H-C3'		1.29	3.52			
1H-C4'				3.21		
2H-C4'				0.02		

Table 5.5.3.3 *J* coupling constants for HH2 conformation, calculated at the BP86 TZVP level (Hz).

3' DNA End							
	H-C1'	1H-C2'	2H-C2'	H-C3'	H-C4'	1H-C5'	2H-C5'
H-C1'							
1H-C2'	2.57						
2H-C2'	7.96						
H-C3'		8.77	8.64				
H-C4'				6.89			
1H-C5'					0.40		
2H-C5'					4.55		

5' DNA End						
	H-C1'	1H-C2'	2H-C2'	H-C3'	1H-C4'	2H-C4'
H-C1'						
1H-C2'	8.39					
2H-C2'	4.64					
H-C3'		6.24	0.36			
1H-C4'				5.44		
2H-C4'				0.81		

Table 5.5.3.4 *J* coupling constants for HT2 conformation, calculated at the BP86 TZVP level (Hz).

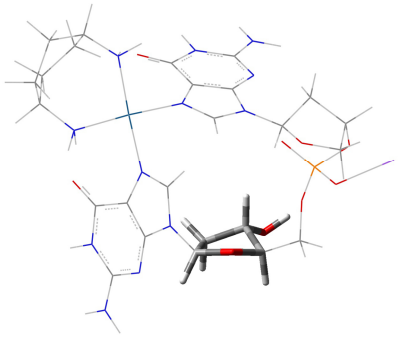
3' DNA End							
	H-C1'	1H-C2'	2H-C2'	H-C3'	H-C4'	1H-C5'	2H-C5'
H-C1'							
1H-C2'	4.72						
2H-C2'	7.57						
H-C3'		0.55	5.21				
H-C4'				0.04			
1H-C5'					1.62		
2H-C5'					2.02		

5' DNA End						
	H-C1'	1H-C2'	2H-C2'	H-C3'	1H-C4'	2H-C4'
H-C1'						
1H-C2'	6.33					
2H-C2'	7.44					
H-C3'		0.52	5.17			
1H-C4'				3.08		
2H-C4'				0.15		

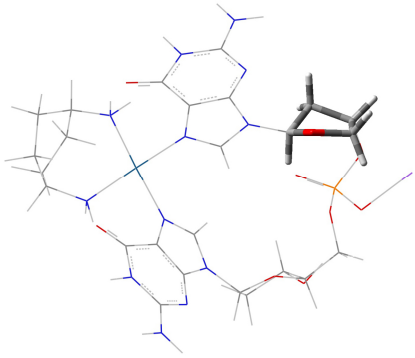
The trends between the SVP and TZVP basis sets used are the same for the coupling constants, but the difference between the values is not consistent for the two of them. The values calculated using the larger basis set, TZVP, which can be assumed to give the more accurate results, have been reported. The coupling constants calculated using the SVP basis set are reported in the appendix.

The coupling constants calculated above can be used to give structural insight into the pentose sugars of the sugar phosphate backbone of DNA. Due to steric hindrance the five membered rings cannot be planar, so one of the atoms moves out of the plane, which atom does this and in which direction has a dramatic effect on the overall structure of DNA, giving rise to the A and B types of DNA. For the four GpG complexes studied here, the C1'-O4'-C4' atoms always remain in the plane and C2' or C3' puckers out of the fixed plane. If the sugar puckers toward the O5' atom on the backbone this is endo puckering, while the opposite direction, toward O3', is exo

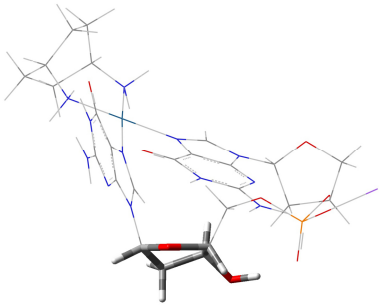
puckering. Information on which type of puckering is occurring is reported in table 5.5.3.5 and diagrams of the pentose sugars in the four complexes are shown in figure 5.5.3.1.



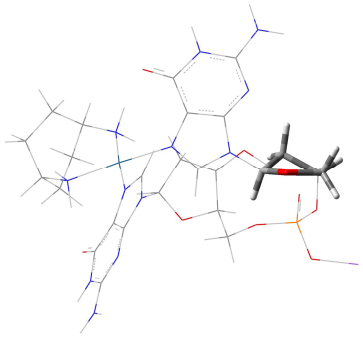
HH1 3' end



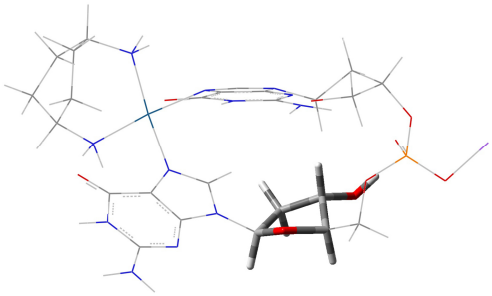
HH1 5' End



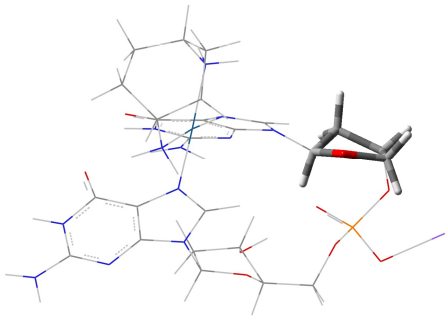
HT1 3' End



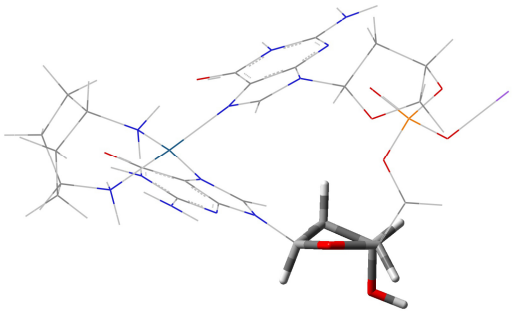
HT1 5' End



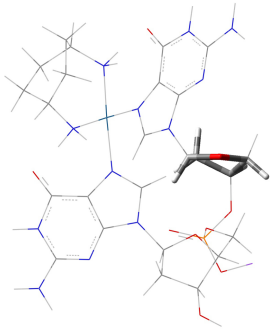
HH2 3' End



HH2 5' End



HT2 3' End



HT2 5' End

Figure 5.5.3.1 Diagrams showing pentose sugars of kiteplatin GpG complexes along the plane C1'-O4'-C4', remainder of complexes shown as wireframe for clarity.

Table 5.5.3.5 Endo/exo puckering of pentose sugars of kiteplatin GpG complexes.

	DNA End	Atom out of Plane	Geometry
HH1	3'	C3'	Endo
	5'	C2'	Endo
HT1	3'	C2'	Exo
	5'	C2'	Endo
HH2	3'	C3'	Endo
	5'	C2'	Endo
HT2	3'	C2'	Endo
	5'	C3'	Exo

From the data in table 5.5.3.5 and figure 5.5.3.1, the most common type of sugar pucker is C2' endo, which is present for all four conformers. This type of puckering is the one observed in B-DNA, so it is not unexpected to be the most common, as the original adduct was constructed from a PDB structure of this type of DNA. Both HH conformers, at the 3' end, however have a C3' endo pucker, which is more commonly observed in A-DNA. They also have similar coupling constants for H-C4' and H-C3', with a difference of 0.38 Hz The HT conformers, at the 3' end, have a significantly different corresponding coupling constant, with values of 4.68 Hz and 0.04 Hz for HT1 and HT2 respectively. HT1 at 3' has a C2' exo pucker and HT2 has a C2' endo, their difference in structure is the result of this change in coupling constant, also for sugars that pucker at the C2' position, coupling constants between protons of C4' and C3' will be of less significance. The C3' to C2' coupling constants for the two HT conformers greatly differ, at 8.27/6.22 Hz and 0.55/5.21 Hz for HT1 and HT2 respectively, due to HT1 being exo and HT2 endo, a similar effect can be observed for the C1'- C2' couplings.

Three of the four conformers have a C2' endo pucker at the 5' end, with HT2 being the exception. The coupling between protons on C2' - C3' are in reasonable agreement for the HH conformers, with values of 5.80/0.35 Hz and 6.24/0.36 Hz for HH1 and HH2 respectively, however the values HT1 differ at 1.29/3.52 Hz. This leads to the conclusion that coupling constants for protons on the carbon atom that puckers are comparable when the same type of puckering is occurring, but only within the same conformation, as can be seen by the difference in head to head and head to tail coupling constants when they both have C2' endo sugar pucker. To be able to get more

conclusive relationships between coupling constants and DNA structure, for these cases, NMR data from experiments being carried out needs to accompany the data from the calculations in this section.

5.5.4 Effect on NMR Calculations of DNA Termination

For calculations regarding small DNA fragments a choice of where to terminate the sugar phosphate backbone has to be made while creating the starting structure for further optimisations. For the four GpG adducts this was carried out by terminating the 3' end at the O3' atom with an additional hydrogen, and terminating the 5' end at the C4' with an additional hydrogen. The justification for terminating the 5' end this way was that having a relatively free moving -CH_2OH moeity instead would have little structural effect on the rest of the backbone or the overall optimisation and as there is a CH_2 between the pentose sugar and the O5' atom, there would be negligible effects on the electronics of the ring. Also terminating this way reduces the number of basis functions and gives fewer geometrical degrees of freedom, which will lead to a slight increase in computational speed.

The experimental results obtained for these structures were taken with the additional methylene at the C5' position. To check if the addition of this group has a significant effect on the calculated NMR values, beyond the C4' atom at the 5' end, a comparison of the HH1 structure has been carried out. The B97-D optimised HH1 structure was modified by hand to add a -CH_2OH at the C4' position and the structure was then optimised using B97-D, the same NMR calculations as carried out above were then carried out on the newly optimised structure. A graphical comparison of some of the important ^{1}H shifts at the BP86/SVP level is shown in figure 5.5.4.1 and ^{13}C shifts at the BP86/TZVP level in figure 5.5.4.2.

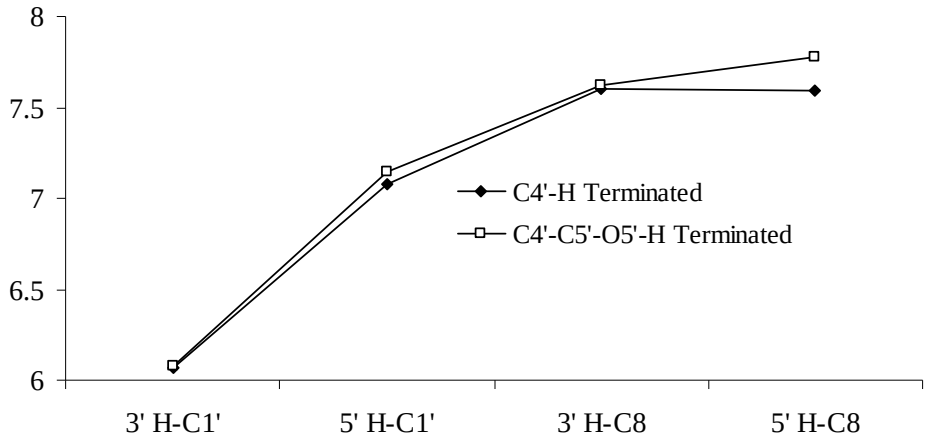


Figure 5.5.4.1 A comparison of ^{1}H NMR shifts for HH1 terminated with -H and -CH_2OH, at the C4' position (ppm).

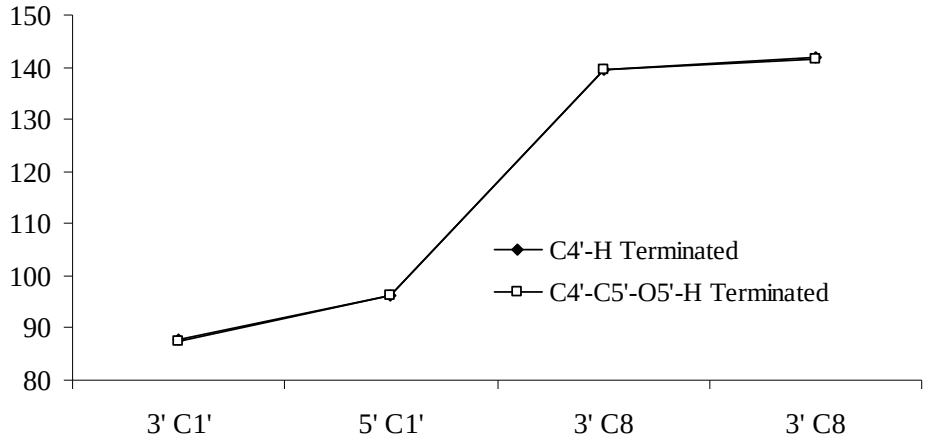


Figure 5.5.4.2 A comparison of ^{13}C NMR shifts for HH1 terminated with -H and -CH_2OH, at the C4' position (ppm).

The full list of ^{1}H and ^{13}C shifts are reported in the appendix, but a comparison between the two ways of terminating is shown in figures 5.5.4.1 and 5.5.4.2, for some of the shifts deemed most important, those of the C8 and C1' carbon and protons. In both graphs there is significant overlap of the lines, which clearly indicates that changing the termination site has little to no effect on these shifts, the only one with any real visible difference is the 5' H-C8 shift, which has a small difference of 0.19 ppm. As it is clear that the calculated NMR shifts at these sites have little dependence on where the 5' end is terminated, a similar comparison has been done for the ^{1}H and ^{13}C shifts on the pentose sugar. The comparison for the ^{1}H shifts for the 5' pentose sugar are shown in

figure 5.5.4.3 and the ^{13}C shifts in figure 5.5.4.4. A comparison for the ^{3}J couplings of the protons, calculated at the BP86/TZVP level, at the 5' end is reported in table 5.5.4.1.

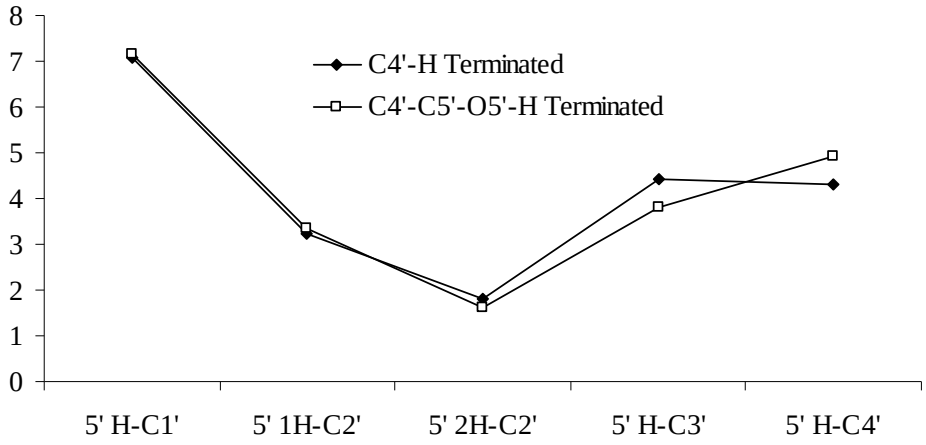


Figure 5.5.4.3 A comparison of ^{1}H NMR shifts of 5' pentose sugar for HH1 terminated with -H and -CH_2OH, at the C4' position (ppm).

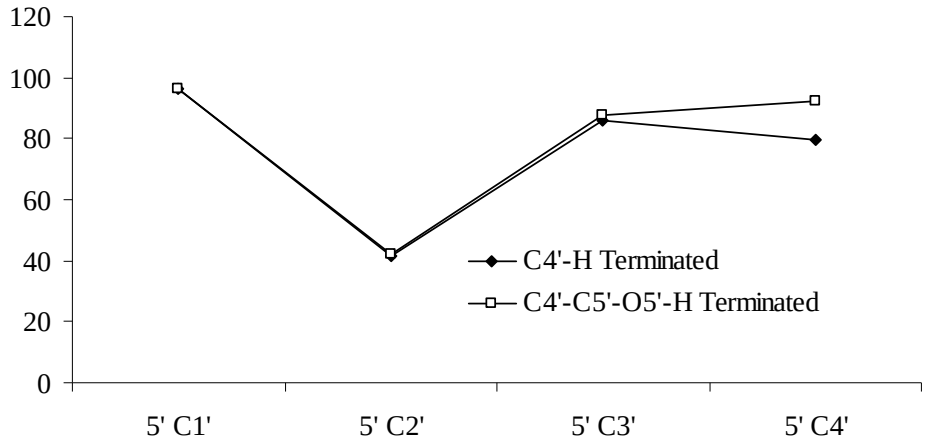


Figure 5.5.4.4 A comparison of ^{13}C NMR shifts of 5' pentose sugar for HH1 terminated with -H and -CH_2OH, at the C4' position (ppm).

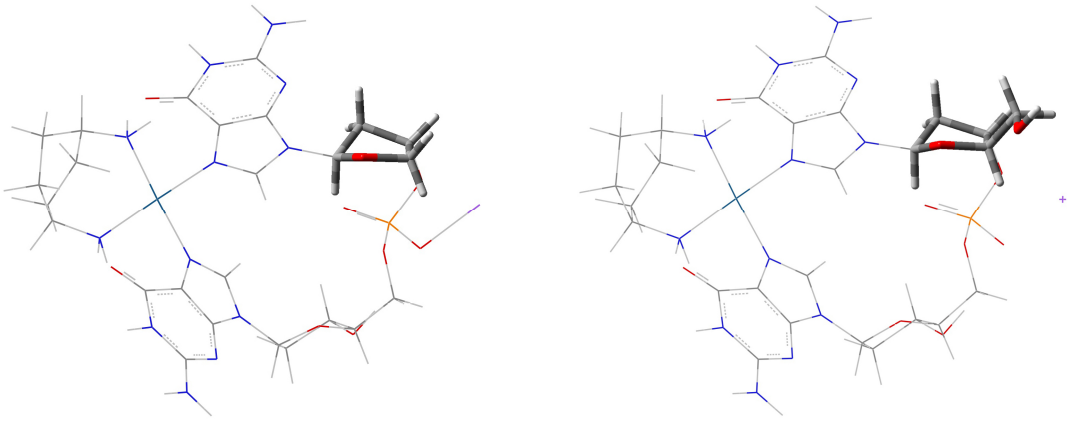
Both graphs, in figures 5.5.4.3 and 5.5.4.4, show that there is little difference in shifts for the protons and carbons of the C1' and C2' atoms. For the shifts of C3' some difference can be observed but the only significant differences on both are for the C4' atom. This confirms that terminating the 5' end at C4' with a hydrogen is suitable for calculating isotropic shieldings, with the exception of the C4' atom of which it is bound to.

Table 5.5.4.1 A comparison of the *J* coupling constants for 5' pentose sugar for HH1 terminated with -H and -CH_2OH, at the C4' position, calculated at the BP86/TZVP level (Hz).

-C4'-C5'-O5'-H Termination							
	H-C1'	1H-C2'	2H-C2'	H-C3'	H-C4'	1H-C5'	2H-C5'
H-C1'							
1H-C2'	9.23						
2H-C2'	3.33						
H-C3'		6.07	0.35				
H-C4'				1.47			
1H-C5'					8.59		
2H-C5'					2.22		

-C4'-H Termination						
	H-C1'	1H-C2'	2H-C2'	H-C3'	1H-C4'	2H-C4'
H-C1'						
1H-C2'	9.11					
2H-C2'	3.53					
H-C3'		5.80	0.35			
1H-C4'				0.96		
2H-C4'				5.74		

A comparison of the data in table 5.5.4.1, leads to the same conclusion as that of comparing the proton and carbon NMR shifts. The coupling constants for the protons on the pentose sugar at the 5' end are comparable, for the values associated with C1', C2', and C3', for the two different terminations, with the largest difference of 0.27 Hz, for the value between H-C3' and 1H-C2'. The coupling constants associated with C4' do differ between -H and -CH_2OH, but this is expected as a proton has been removed from that atom in the latter case and there are additional couplings with protons on the added C5' atom. This likely gives rise to the larger difference, of 0.51 Hz between H-C3' and H-C4'. The effect on the sugar puckering has also been tested, with diagrams shown in figure 5.5.4.5.



-C4'-H Terminated -C4'-C5'-O5'-H Terminated

Figure 5.5.4.5 Diagrams showing 5' pentose sugars of HH1 terminated with -H and -CH_2OH at C4', along the plane C1'-O4'-C4', remainder of complexes shown as wireframe for clarity.

The pentose sugars, shown in figure 5.5.4.5, both have the same puckering, regardless of termination site. Both exhibit C2' endo puckering, with some distortion causing C3' to also be out of the plane, albeit by a marginal amount. This effect is slightly more prevalent in the sugar terminated by -CH_2OH, but this can likely be attributed to the addition of this extra functional group.

A comparison of calculated ^{1}H and ^{13}C NMR shifts, along with $^{3}J_{HH}$ coupling constants, show that by choosing to terminate the DNA at the 5' end with a proton on the C4' atom, instead of -CH_2OH, has little to no effect on values associated to atoms further than one bond away. For the HH1 GpG 1,4-DACH Pt complex, the sugar puckering is also the same in both cases of termination and from visual observation of the whole structures in figure 5.5.4.5 it is also clear that there is little change to the overall geometry. As a result of this the calculations carried out on the remaining conformers above, all terminated with a proton at C4', can be assumed to be accurate to those if the termination was changed to the more biologically correct -CH_2OH, with the exception of value for the C4' atom and to a much lesser extent the C3' atom.

5.6 References

1. Hoeschele, J.D.; Hollis Showalter, H.D.; Kraker, A.J.; Elliott, W.L.; Roberts, B.J.; Kampf, J.W. *J. Med. Chem.* **1994**, 37, 2630-2636.
2. Van Cutsem, E.; Peeters, M.; Siena, S.; Humblet, Y.; Hendlisz, A.; Neyns, B.; Canon, J.-L.; Van Laethem, J.-L.; Maurel, J.; Richardson, G.; Wolf, M.; Amado, R.G. *J. Clin. Oncol.* **2007**, 25, 1658-1664.
3. Margiotta, N.; Marzano, C.; Gandin, V.; Osella, D.; Ravera, M.; Gabano, E.; Platts, J.A.; Petruzzella, E.; Hoeschele, J.D.; Natile, G. *J. Med. Chem.* **2012**, 55, 7182-7192.
4. Leduc, A.M.; Trent, J.O.; Wittliff, J.L.; Bramlett, K.S.; Briggs, S.L.; Chirgadze, N.Y.; Wang, Y.; Burris, T.P.; Spatola, A.F. *Proc. Natl. Acad. Sci. USA* **2003**, 100, 11273-11278.
5. Sherman, S.E.; Gibson, D.; Wang, A.H.-J.; Lippard, S.J. *J. Am. Chem. Soc.* **1988**, 110, 7368.
6. Margiotta, N. *Personal Communication*
7. http://w3dna.rutgers.edu/
8. http://www.atdbio.com/content/5/Nucleic-acid-structure
9. Gkionis, K.; Platts, J.A. *Comp. Theor. Chem.* **2012**, 993, 60-65.
10. Li, X.; Frisch, M.J. *J. Chem. Theory and Comput.* **2006**, 2, 835-839.
11. Yang, D.; van Boom, S.S.; Reedijk, J.; van Boom, J.H.; Wang, A.H. *Biochemistry* **1995**, 34, 12912-12920.
12. *Molecular Operating Environment (MOE)*, 2012.10; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, **2012**.
13. 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. *Biochemistry* **2007**, 46, 6477-6487.
14. Todd, R.C.; Lippard, S.J. *J. Inorg. Biochem.* **2010**, 104, 902-908.
15. Vreven, T.; Mennucci, B.; da Silva, C.O.; Morokuma, K.; Tomasi, J. *J. Chem. Phys.* **2001**, 115, 62-72.
16. Dodd, L.R.; Theodorou, D.N. *Molecular Physics* **1991**, 72, 1313-1345.
17. Platts, J.A.; Gkionis, K. *Phys. Chem. Chem. Phys.* **2009**, 11, 10331-10339.
18. London, F. *J. Phys. Radium* **1937**, 8, 397-409.
19. McWeeny, R. *Phys. Rev.* **1962**, 126, 1028.
20. Ditchfield, R. *Mol. Phys.* **1974**, 27, 789-807.
21. Wolinski, K.; Hilton, J.F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, 112, 8251-8260.
22. Cheesman, J.R.; Trucks, G.W.; Keith, T.A.; Frisch, M.J. *J. Chem. Phys.* **1996**, 104, 5497-5509.
23. Karplus, M. *J. Am. Chem. Soc.* **1963**, 85, 2870-2871.

Chapter 6: Conclusions

The work presented in this thesis summarises the efforts at modelling intra- and inter-molecular non-covalent interactions of organometallic transition metal complexes. Several of the complexes were chosen as they have shown the potential to act as antitumour agents and for these interactions with differing sizes of DNA adducts have been explored. This was achieved using a variety of DFT and QM/MM approaches and where available calculations have been validated by experimental data and high level *ab initio* benchmark calculations.

The interactions of two test transition metal complexes, namely ferrocene and $\eta^6(C_6H_6)Cr(CO)_3$, with methane and benzene have been probed. Intermolecular complexes were constructed to represent the T-shaped methane...benzene interactions and several of the conformations of the benzene dimer, to model a variety of CH...π and π...π interactions. Two DFT methods were tested for their suitability at modelling the energies of non-covalent interactions in these systems, BHandH, a hybrid functional that has shown promise in similar applications and B97-D, a pure functional with an empirical dispersion correction. Both of these methods were benchmarked against two *ab initio* methods, DF-LMP2 and CBS(T), where the latter is considered the most accurate approach possible for systems of the size studied here. Results were in good agreement for all four approaches with the exception of the parallel displaced complexes, which had larger differences but were still in reasonable agreement. When testing the suitability of using both DFT functionals for optimisation of complexes containing non-covalent interactions, both gave comparable structures and fixed scans at the BHandH and DF-LMP2 levels show that this DFT method gives equilibrium distances comparable to the higher level method.

With the suitability of BHandH and B97-D ascertained, similar intermolecular interactions of the ruthenium arene complex of the form $[\eta^6(C_6H_6)Ru(en)Cl]^+$ with methane and benzene were studied. The *ab initio* method MP2(0.25) was also utilised and all three methods were in good agreement, showing that C-H...π complexes in which the Ru-arene acts as a hydrogen bond acceptor have binding energies and geometries similar to those of the free benzene. In contrast, π-stacking interactions of the coordinated arene are found to be 7 to 8 kcal/mol stronger than those in the benzene dimer (*ca.* 2.5 kcal/mol). Similarly, the inverted-T form in which the Ru-arene acts as a

hydrogen bond donor to the π-system of benzene is stronger again. The increased strength of these interactions is shown to be due to more favourable electrostatics as well as significantly increased dispersion forces. In the latter case, formation of N-H...π as well as C-H...π interactions plays a role in stabilisation, as demonstrated by QTAIM analysis.

These ruthenium arene complexes of the general form $[\eta^6(\text{arene})\text{Ru(en)Cl}]^+$ have been shown to have potential anticancer activity, so were further studied with a series of extended aromatic and partially saturated arene ligands. Intramolecular interactions occurring when these complexes bond to the DNA nucleobases adenine and guanine, have been modelled and used to test the suitability of a variety of hybrid DFT, pure DFT, and post Hartree-Fock methods, with a focus on confirming whether using the GGA functional B97 with an empirical dispersion correction of the type D2, was suitable to calculate binding energies within transition metal complexes, in light of recent publications using the newer D3 correction. Comparing against benchmark DF-LMP2, that has been further warranted by calculation of selected DF-LCCSD values, the best methods were found to be BHandH, B97-D2, and MP2(0.25), with wB97-D, B2PLYPD, and B97-D3 also performing adequately, but not to the desired level of accuracy. Through this analysis of the B97-D2 binding energies and through comparison to higher level methods, it been shown that this method suitably reproduces values of significantly more expensive calculations.

The ruthenium arene guanine/adenine complexes were expanded upon by increasing DNA to a GpC·CpG base pair step, where the extended arenes have the potential to act as intercalators. Geometry optimisation using B97-D allowed the importance of these interactions to be determined as well as any hydrogen bonds between coordinated N-H groups to potential acceptors on DNA. Sensitivity of binding energies to the nature of the arene groups was found, such that larger and more flexible arenes are able to form more stabilising interactions with DNA bases. The need for flexibility to maximise interactions is shown by the relatively low binding energy of the anthracene complex, whereas dihydro- and tetrahydro-anthracene bind particularly strongly. The importance of non-covalent interactions between arene and DNA bases was demonstrated by QTAIM analysis, in which larger numbers of π-stacking and C—H...π hydrogen bonding critical points are found for the more strongly bound complexes. The anomalously strong binding of the benzene complex is due to the detailed electronic

structure of this complex, which has a much lower energy LUMO than the others considered. This is most evident from complexes in which DNA is in its canonical B-form, with the arene pointing out from the double helix. The energy difference between the intercalated and canonical forms also highlights the importance of non-covalent interactions in the former.

Platinum anticancer drugs have also been explored, with a particular focus on the relatively new complex kiteplatin, which contains an isomeric form of the diaminocyclohexane ligand of oxaliplatin. Adducts of this drug with GpG were studied and conformational searches carried out using relaxed scans with the semi-empirical method PM6 to ascertain four possible conformations. These conformations were studied using PM6-DH2 and B97-D, and results compared against analogous calculations on cisplatin and oxaliplatin. To aid assignment of experimental data, ^{1}H isotropic shielding values, ^{13}C isotropic shielding values, and $^{3}J_{HH}$ coupling constants have been calculated at the BP86 level, for kiteplatin GpG adducts. The type of sugar puckering occurring on the pentose sugars of the DNA backbone was also recorded and related to calculated coupling constants.

Larger adducts of kiteplatin, cisplatin, and oxaliplatin with DNA have also been modelled using B97-D, including the base pair step GpG·CpC. Base pair parameters for these adducts have shown that in general the nucleobases are in similar conformations, however torsional parameters of the sugar phosphate backbone show some disparity between oxaliplatin and kiteplatin/cisplatin. Binding energies show that cisplatin is most strongly bound in the gas phase and with an implicit solvent model, whilst oxaliplatin and kiteplatin are comparable and give interaction energies approximately 34 kcal/mol and 7 kcal/mol weaker in the gas phase and solvent phase respectively.

Throughout this work the B97-D functional has been used to great effect in optimising relatively large systems and calculating interaction energies, but to model larger than the base pair steps using this method would pose major problems. For the larger systems studied, namely kiteplatin adducts with a tetramer, octamer, and dodecamer, the QM/MM approach was used. Optimisation of these adducts with explicit water molecules was achieved using ONIOM to combine BHandH with AMBER molecular mechanics. This approach was also used for the base pair step and results compared against those calculated using only DFT. Whilst there were some noticeable differences

in base pair parameters, including a sign change between shift and tilt, and a relatively large increase of 0.28 Å in rise, the methods do give broadly similar structures. ONIOM calculated binding energies for the set of adducts show that the energies converge as the size of the DNA fragment increases, with a large increase in strength of interaction between the dimer and tetramer, -271.60 kcal/mol and -429.79 kcal/mol respectively, while this change decreases significantly to the octamer with a binding energy of -473.47 kcal/mol. This trend is also exhibited in the calculated exposed surface areas of kiteplatin in the adducts. The base pair parameters around the central platinated guanines for all the adducts were also compared and exhibit some differences, particularly in the range of values for the roll parameter. Despite this difference the remainder of the parameters are broadly comparable, concluding that changing the size of DNA fragment still results in fairly comparable distortion around the coordination site of the drug.

Chapter 7: Appendix

7.1 Appendices for Chapter 3

Table 7.1.1 Binding energies of ruthenium arene complexes with guanine after reoptimisation (kcal/mol).

	B97-D2	BHandH	MP2(0.25)	DF-LMP2
Benz	-93.03	-97.50	-90.84	-94.48
Cym	-85.92	-91.62	-85.29	-89.58
Bip	-85.83	-93.13	-87.81	-93.19
Anth	-80.45	-87.08	-82.35	-87.02
Dha	-91.50	-95.43	-91.35	-96.76
Tha	-90.64	-96.80	-90.78	-96.24

7.2 Appendices for Chapter 5

Table 7.2.1 ^{1}H NMR shifts for the protons in the DNA for the four GpG adducts, calculated at BP86 level with SVP basis set and ECP on Pt (ppm).

Atom Type[a]	End of DNA	HH1 Shift	HT1 Shift	HH2 Shift	HT2 Shift
Backbone					
C1'	3'	6.07	5.88	5.69	6.19
C2'	3'	1.96	2.53	2.23	3.24
C2'	3'	2.57	2.04	0.54	2.63
C3'	3'	5.36	3.76	4.27	4.89
C4'	3'	3.42	4.66	3.27	4.18
C5'	3'	4.14	2.06	4.16	3.81
C5'	3'	4.36	4.30	4.65	3.80
O3'	3'	2.08	6.69	5.35	0.62
C1'	5'	7.08	6.19	7.03	6.64
C2'	5'	3.24	3.18	4.66	3.45
C2'	5'	1.82	3.09	1.50	2.57
C3'	5'	4.44	4.56	4.60	4.32
C4'	5'	4.32	3.80	4.42	3.86
C4'	5'	4.52	4.39	4.46	4.32
Guanine					
C8	3'	7.60	6.28	6.04	7.23
N1	3'	6.74	6.88	6.86	6.88
N2	3'	4.61	4.73	4.49	4.68
N2	3'	4.15	4.24	4.16	4.23
C8	5'	7.59	6.46	7.57	6.16
N1	5'	6.78	6.90	6.27	6.83
N2	5'	4.49	4.61	4.76	4.40
N2	5'	4.10	4.26	3.92	4.13

[a]Atom type is the one which the proton of interest is bonded to

Table 7.2.2 ^{1}H NMR shifts for the protons in the DNA for the four GpG adducts, calculated at BP86 level with PCS1 basis set and ECP on Pt (ppm).

Atom Type[a]	End of DNA	HH1 Shift	HT1 Shift	HH2 Shift	HT2 Shift
Backbone					
C1'	3'	6.42	6.17	5.96	6.46
C2'	3'	2.20	2.81	2.47	3.46
C2'	3'	2.86	2.25	0.83	2.84
C3'	3'	5.68	3.88	4.52	5.14
C4'	3'	3.63	4.96	3.42	4.38
C5'	3'	4.40	2.26	4.43	4.01
C5'	3'	4.61	4.63	4.87	3.99
O3'	3'	2.62	7.26	5.85	1.16
C1'	5'	7.27	6.50	7.28	6.91
C2'	5'	3.43	3.51	4.97	3.69
C2'	5'	2.05	3.32	1.72	2.78
C3'	5'	4.67	4.82	4.85	4.58
C4'	5'	4.66	4.12	4.72	4.17
C4'	5'	4.83	4.63	4.82	4.61
Guanine					
C8	3'	7.85	6.56	6.37	7.49
N1	3'	6.95	7.07	7.09	7.10
N2	3'	4.88	4.98	4.76	4.95
N2	3'	4.46	4.55	4.47	4.54
C8	5'	7.81	6.75	7.86	6.47
N1	5'	6.98	7.10	6.51	7.07
N2	5'	4.75	4.87	5.06	4.67
N2	5'	4.41	4.57	4.25	4.46

[a]Atom type is the one which the proton of interest is bonded to

Table 7.2.3 *J* coupling constants for HH1 conformation, calculated at the BP86 SVP level (Hz).

3' DNA End							
	H-C1'	1H-C2'	2H-C2'	H-C3'	H-C4'	1H-C5'	2H-C5'
H-C1'							
1H-C2'	6.82						
2H-C2'	0.89						
H-C3'		7.27	5.76				
H-C4'				4.96			
1H-C5'					1.03		
2H-C5'					1.76		

5' DNA End						
	H-C1'	1H-C2'	2H-C2'	H-C3'	1H-C4'	2H-C4'
H-C1'						
1H-C2'	7.06					
2H-C2'	2.03					
H-C3'		4.43	0.23			
1H-C4'				4.32		
2H-C4'				0.54		

Table 7.2.4 *J* coupling constants for HT1 conformation, calculated at the BP86 SVP Level (Hz).

3' DNA End							
	H-C1'	1H-C2'	2H-C2'	H-C3'	H-C4'	1H-C5'	2H-C5'
H-C1'							
1H-C2'	4.15						
2H-C2'	0.03						
H-C3'		6.34	4.47				
H-C4'				3.59			
1H-C5'					5.04		
2H-C5'					6.33		

5' DNA End						
	H-C1'	1H-C2'	2H-C2'	H-C3'	1H-C4'	2H-C4'
H-C1'						
1H-C2'	2.41					
2H-C2'	6.56					
H-C3'		1.11	2.54			
1H-C4'				2.19		
2H-C4'				-0.15		

Table 7.2.5 *J* coupling constants for HH2 conformation, calculated at the BP86 SVP level (Hz).

3' DNA End							
	H-C1'	1H-C2'	2H-C2'	H-C3'	H-C4'	1H-C5'	2H-C5'
H-C1'							
1H-C2'	1.93						
2H-C2'	5.74						
H-C3'		6.63	6.62				
H-C4'				5.26			
1H-C5'					0.32		
2H-C5'					3.42		

5' DNA End						
	H-C1'	1H-C2'	2H-C2'	H-C3'	1H-C4'	2H-C4'
H-C1'						
1H-C2'	6.43					
2H-C2'	2.97					
H-C3'		4.68	0.24			
1H-C4'				3.96		
2H-C4'				0.44		

Table 7.2.6 *J* coupling constants for HT2 conformation, calculated at the BP86 SVP level (Hz).

3' DNA End							
	H-C1'	1H-C2'	2H-C2'	H-C3'	H-C4'	1H-C5'	2H-C5'
H-C1'							
1H-C2'	3.06						
2H-C2'	5.68						
H-C3'		0.45	3.70				
H-C4'				-0.11			
1H-C5'					1.03		
2H-C5'					1.55		

5' DNA End						
	H-C1'	1H-C2'	2H-C2'	H-C3'	1H-C4'	2H-C4'
H-C1'						
1H-C2'	4.48					
2H-C2'	5.62					
H-C3'		0.38	3.75			
1H-C4'				2.06		
2H-C4'				0.02		

Table 7.2.7 A comparison of the ^{1}H NMR shifts for the protons in the DNA for HH1 terminated with a -H and -CH_2OH, at the C4' position, calculated at BP86 level with SVP basis set and ECP on Pt (ppm).

Atom Type[a]	End of DNA	-C4'-H Terminated HH1 Shift	-C4'-C5'-O5'-H Terminated HH1 Shift
Backbone			
C1'	3'	6.07	6.08
C2'	3'	1.96	2.59
C2'	3'	2.57	1.94
C3'	3'	5.36	5.37
C4'	3'	3.42	3.46
C5'	3'	4.14	4.34
C5'	3'	4.36	4.21
O3'	3'	2.08	2.22
C1'	5'	7.08	7.15
C2'	5'	3.24	1.61
C2'	5'	1.82	3.34
C3'	5'	4.44	3.79
C4'	5'	4.32	4.92
C4'	5'	4.52	
C5'	5'		3.83
C5'	5'		3.87
O5'	5'		0.15
Guanine			
C8	3'	7.60	7.62
N1	3'	6.74	6.75
N2	3'	4.61	4.61
N2	3'	4.15	4.15
C8	5'	7.59	7.78
N1	5'	6.78	6.79
N2	5'	4.49	4.48
N2	5'	4.10	4.06

[a]Atom type is the one which the proton of interest is bonded to

Table 7.2.8 A comparison of the ^{13}C NMR shifts for the carbons in the DNA for HH1 terminated with a -H and -CH_2OH, at the C4' position, calculated at BP86 level with TZVP basis set and ECP on Pt (ppm).

Atom Type	End of DNA	-C4'-H Terminated HH1 Shift	-C4'-C5'-O5'-H Terminated HH1 Shift
Backbone			
C1'	3'	87.66	87.50
C2'	3'	40.67	41.17
C3'	3'	74.05	73.86
C4'	3'	95.03	94.48
C5'	3'	68.80	68.73
C1'	5'	96.16	96.38
C2'	5'	41.35	42.04
C3'	5'	85.92	87.80
C4'	5'	79.43	92.36
C5'	5'		71.52
Guanine			
C2	3'	151.84	152.19
C2	5'	151.51	151.60
C4	3'	152.72	156.04
C4	5'	152.82	155.25
C5	3'	118.85	118.39
C5	5'	119.28	119.18
C6	3'	155.65	152.36
C6	5'	155.28	152.78
C8	3'	139.64	139.41
C8	5'	142.02	141.39