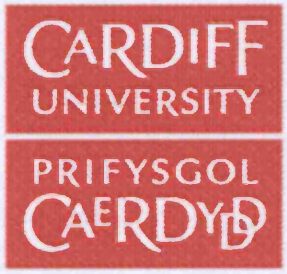


QUANTUM CHEMICAL STUDIES OF DNA AND METAL-DNA STRUCTURES

H-BONDING AND π-STACKING

Arturo Robertazzi

Ph.D. Thesis, 2006

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

VENI, VIDI, VICI...

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Jamie Platts for all his help and enthusiasm during the three years of my PhD. I would like also to thank all the people in my lab and all those I have met during this beautiful time at Cardiff University.

Well... I am Italian, so I cannot escape from this: "*Grazie Mamma Jole* for all the nice food you cooked every time I came back to Italy and *Grazie Papà Giorgio*, your credit card has always been available for me!". A big *grazie* goes to my sister *Rusenè* too, the *whole family* and all my friends from the deep south of Europe.

Last but not least, a *dankjewel* to Eveline, who supported and helped me during all this time, enduring my frenzied discussions of quantum mechanics.

ABSTRACT

Metal complexes interact with many different sites of nucleic acids, stabilising the structure or, in some cases, leading to severe distortion or non-canonical forms of DNA such as triplexes, quadruplexes, junctions etc. Remarkably, several transition metals are considered potentially active anti-cancer drugs. Among these the most studied is certainly cis-diamminodichloroplatinum(II), or cisplatin, which after an activation process, attacks DNA in guanine-rich regions leading to strong distortion of DNA structure. Theoretical and experimental works suggest π-stacking disruption and GC pair distortion as the most relevant effects. In this work, *ab initio* and DFT calculations are extensively employed in order to explore the role of basic forces in DNA and metal-DNA adducts. To do so, Atoms in Molecules (AIM) theory has been used as a tool to decompose binding energies into contributions of covalent bond, H-bond and π-stack energies, leading to a clearer picture of the studied systems. Firstly, DFT methods were employed to investigate the hydrolysis mechanism of cisplatin, a key step in the activation of the drug. Subsequently, an AIM based approach has been proposed to estimate H-bond energies in metal-DNA complexes. This allowed us to investigate the effect of platination on GC pair and, more generally, the role of H-bonding in such systems. A large study of transition metal-purine complexes, from Ti to Hg, has been discussed, providing a systematic analysis of the effect of metallation on the GC pair. As well as H-bonding, π-stacking plays a fundamental role in DNA and metal-DNA structures. In order to avoid use of expensive and, in certain cases, prohibitive methods such as MP2 and CCSD calculations, a new hybrid DFT functional, BHandH, has been applied to stacked model complexes (from benzene to DNA nucleobases). In addition, AIM analysis was shown to be useful tool in estimating π-stacking interactions in these systems. Thus, QM/MM calculations (QM = BHandH, MM = AMBER) were employed in order to investigate the role of H-bonding and π-stacking in DNA and cisplatin-DNA adducts, the interplay between those being our main focus. One example of realistic platinated octamer was also studied with the BH&H/AMBER/AIM approach, leading to general agreement with experimental data.

PUBLICATIONS

Chapter 3

i. Robertazzi, A. and J.A. Platts, "Hydrogen bonding, solvation, and hydrolysis of cisplatin: A theoretical study". *J. Comput. Chem.*, **2004**, 25(8), 1060-1067

ii. Robertazzi, A. and Platts, J. A., "Hydrogen bonding and covalent effects in binding of cisplatin to purine bases: *ab initio* and atoms in molecules studies". *Inorg. Chem.* **2005,** 44(2), 267-274

iii. Robertazzi, A. and Platts, J. A., "Binding of transition metal complexes to guanine and guanine-cytosine: hydrogen bonding and covalent effects". *J. Biol. Inorg. Chem.* **2005,** 10(8), 854-866

Chapter 4

iv. Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A., "Hybrid density functional theory for π-stacking interactions: application to benzenes, pyridines, and DNA bases". *J. Comput .Chem.*, **2006**, 27(4), 491-504.

v. Robertazzi, A. and Platts, J. A., "Gas-phase DNA oligonucleotide structures. A QM/MM and Atoms in Molecules study". *J. Phys. Chem. A* (*in press*).

vi. Robertazzi, A. and Platts, J. A., "A QM/MM study of cisplatin-DNA oligonucleotides: from simple models to realistic systems". *Chem. Eur. J.* (*in press*).

CONTENTS

Chapter 1

Literature Review

Chapter 2

Theory and Methodology

Chapter 3

The chemistry of cisplatin and other transition metal ligands

From hydrolysis to DNA bases complexes

Chapter 4

DNA and cisplatin-DNA structures

H-bonding and π-stacking

1 Literature Review

1.1 Preface

Nucleic acids (DNA or RNA) represent the biological code that regulates vital processes in organisms, determining the features and activities of single cells. In particular, DNA stores genetic information that, after the *transcription* process, regulates the nature of the amino acids in proteins' structure. Also, DNA plays a fundamental role in transmitting the biological code from the parent cell to the daughter cell through *replication*. On the other hand, perturbation of DNA structure can lead to critical consequences; for instance, metal ions, depending on their nature, can either stabilise the DNA structure or disrupt its fundamental properties. The aim of this chapter is to illustrate the essential features of the DNA macromolecule and the effect on its structure of transition metals, particularly cisplatin. The basic concepts on DNA structures are available from Voets's *Fundamentals of Biochemistry,* the reader is therefore directed to this text book for further details.[1]

1.2 DNA structures

More than 50 years ago, Watson and Crick proposed the structure of DNA, a two-stranded macromolecule of deoxynucleotides connected *via* a chain of phosphodiester bonds, see Figure 1.1.[2] The most common form is B-DNA, which shows the following features:

i. the two polynucleotide strands present a common axis producing a double helix with a diameter of *ca.* 20 Å;

ii. the nucleobases (guanine, adenine, thymine and cytosine, or G, A, T, C) lie in planes *quasi* perpendicular to the main axis, occupying the inner region of the double helix, while the sugar-phosphate backbone winds around the outside, forming the major and minor groves;

iii. the nucleobases are paired as follows: guanine with cytosine (GC) and adenine with thymine (AT);

iv. the nucleobases can replace each other causing no rearrangements in the phosphate backbone. In contrast, any other combination of bases would significantly distort the final structure.

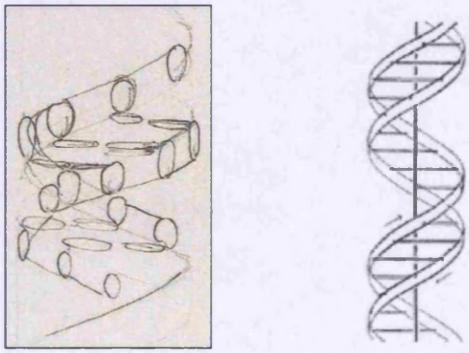


Figure 1.1 First Watson and Crick's schemes of DNA.[2]

1.2.1 Flexibility of DNA: conformational properties of nucleotide units

B-DNA is not the only known form of DNA: depending on the conditions, DNA can arrange in A- and Z-DNA structures which significantly differ from B-DNA (Figures 1.2a-b). A-DNA prevails under dehydrating conditions, whereas Z-DNA needs high salt concentrations. Thus, DNA chains have a high degree of flexibility, with single DNA

residues being able to adopt very different conformations or, for instance, each base deviating from ideal geometry by rolling and twisting. These conformational changes appear to be fundamental in many processes involving nucleic acid structures.

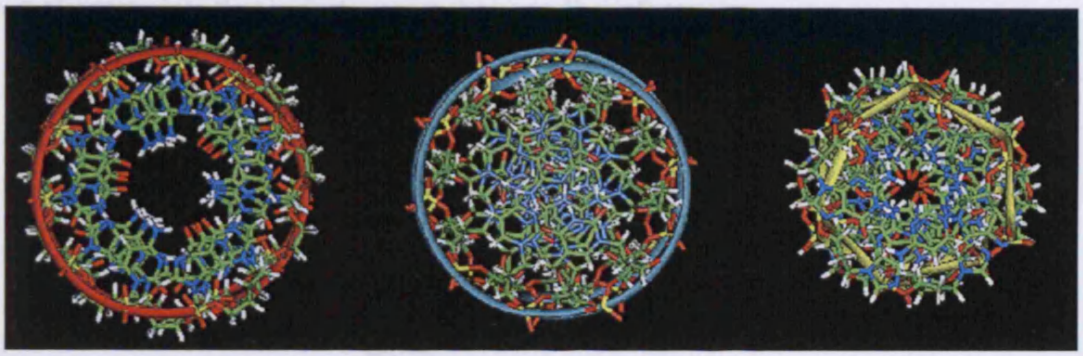


Figure 1.2-a A-DNA, B-DNA and Z-DNA, top view.

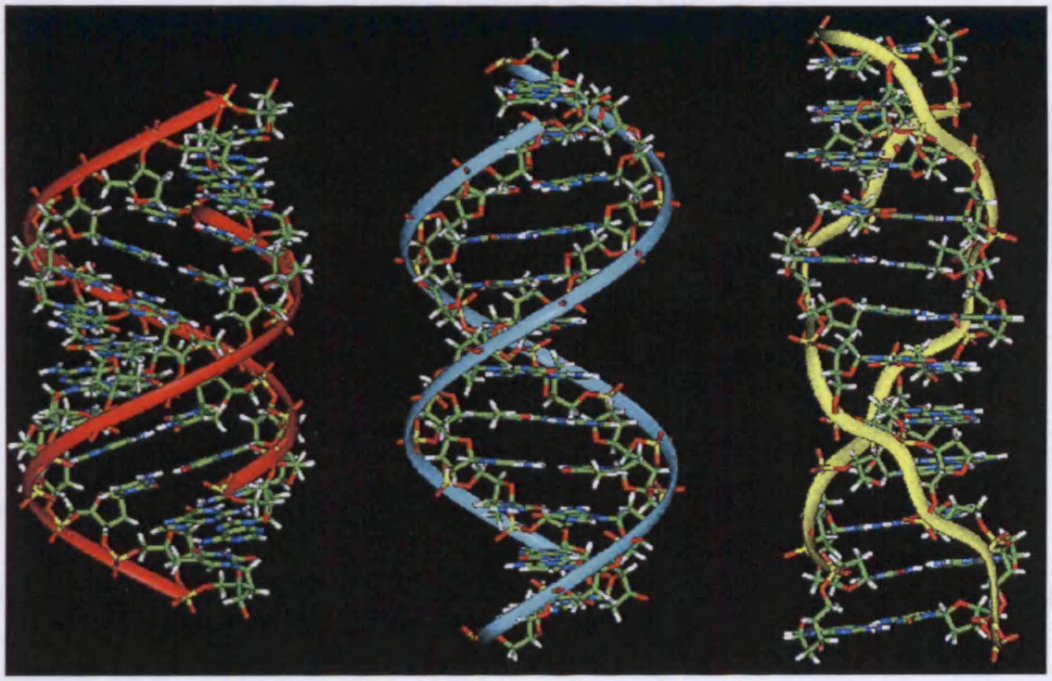


Figure 1.2-b A-DNA, B-DNA and Z-DNA, side view (pictures from http://en.wikipedia.org).

Despite the high degree of flexibility of DNA, geometrical distortion of the structure is limited by the nucleotide units. As Figure 1.3 displays, the conformation of the backbone is described by six torsion angles of the sugar-phosphate chain and the torsion angle indicating the orientation of the base around the glycosidic bond (C1' to the base). Although the are seven degrees of freedom represented by these angles, there are numerous internal constraints that in fact restrict the overall conformation.

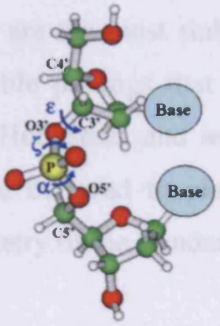


Figure 1.3 Torsion angles of the phosphate backbone.

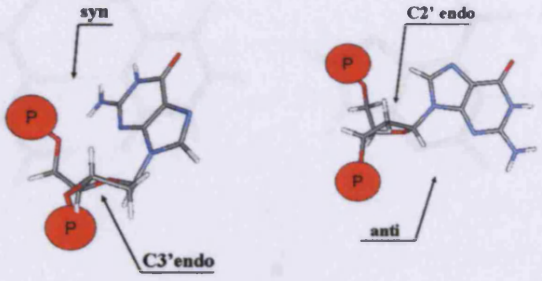


Figure 1.4 *Syn* and *anti* conformation in DNA bases.

The permissible conformations of the nucleotide units show the following features:

i. purine residues have two sterically allowed positions: *syn* or *anti* as in Figure 1.4, the *anti* being the most adopted conformation. For pyrimidines only the *anti* conformation is possible;

ii. typically, the four atoms of the ribose ring are coplanar while the fifth atom, generally the C2' or C3', is out of plane to relieve steric hindrance. The most known conformations are *C3'-endo* and *C2'-endo*, where the out of plane atom is on the same side of the ring as C5';

iii. the torsion angles of the sugar-phosphate backbone are not free to rotate because of the non-covalent interactions between the ribose ring and the phosphate groups and steric hindrance between residues.

1.2.2 Base pairing and π-stacking

Figure 1.5a displays the Watson-Crick base pairing of GC and AT. Guanine and cytosine are bonded *via* three strong hydrogen bonds, while adenine and thymine by two. Although the Watson-Crick pairs are the most stable, other pairing combinations are known. There are in fact many possible pairings that involve at least two hydrogen bonds, including reverse Watson-Crick, Hoogsteen and *wobble* (or mismatched) base pairs (Figure 1.5b).[3] These combinations are believed to play an important biological role especially if not constrained in the geometry of the standard double helix.[4]

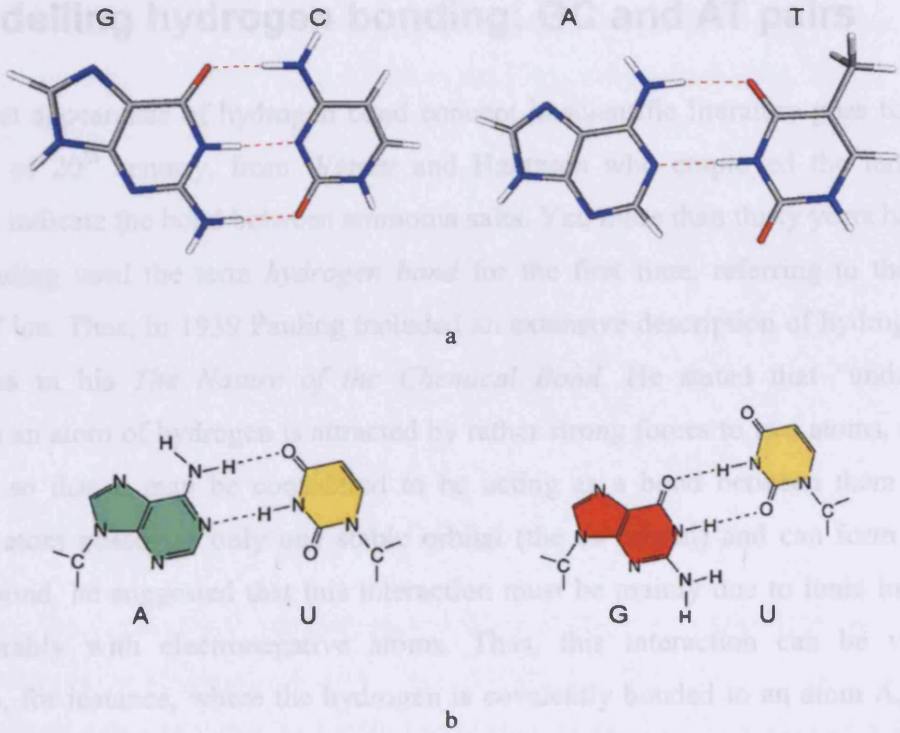


Figure 1.5 (a) GC and AT Watson-Crick pairs; (b) AU and GU wobble pairs.

H-bonding is not the only interaction that nucleobases undergo. In fact, while it is understood that H-bonding is required for the specificity of base pairing, it is also believed that these interactions contribute little to the final stability of nucleic acid structure. For instance, when denaturing DNA polymers, the hydrogen bonds between bases are replaced by contacts to water molecules. Therefore, another kind of intermolecular force must intervene to stabilise the structure of nucleic acids: these forces are the π-stacking interactions between aromatic rings of nucleobases. The analysis of nucleic acid structures has shown that both purine and pyrimidine residues tend to form extended stacks of planar parallel molecules, and it is believed that such interactions are far stronger between stacked G and C than A and T. Also, different sets of base pairs have different stacking energies, suggesting that these interactions are sequence-dependent. More details on the nature of hydrogen bonding and π-stacking are reported in following sections.

1.3 Modelling hydrogen bonding: GC and AT pairs

The earliest appearance of hydrogen bond concept in scientific literature goes back to the beginning of 20th century, from Werner and Hantzsch who employed the term *second valence* to indicate the bond between ammonia salts. Yet, more than thirty years had to pass before Pauling used the term *hydrogen bond* for the first time, referring to the residual entropy of ice. Thus, in 1939 Pauling included an extensive description of hydrogen bonds interactions in his *The Nature of the Chemical Bond.* He stated that "under certain conditions an atom of hydrogen is attracted by rather strong forces to two atoms, instead of only one, so that it may be considered to be acting as a bond between them". As the hydrogen atom possesses only one stable orbital (the 1s orbital) and can form only one covalent bond, he suggested that this interaction must be mainly due to ionic interactions and preferably with electronegative atoms. Thus, this interaction can be viewed as A—H...B, for instance, where the hydrogen is covalently bonded to an atom A, the acid, and interacts with B, the base, *via* electrostatics. Further details on hydrogen bonding are available from Steiners's book, *The Weak Hydrogen Bond.*[5]

1.3.1 Nature of hydrogen bonding

In the last decades, hydrogen bonding has been extensively studied using both experimental and theoretical methods. Although Pauling's definition is generally considered valid, the hydrogen bond is not a simple interaction but a complex combination of several contributions, such as electrostatics, polarization, exchange repulsion, charge transfer and dispersion. While the electrostatics, charge transfer and polarisation are directional, the exchange repulsion and dispersion forces are isotropic. These terms are attractive, apart from the exchange repulsion contribution which follows an r^{-12} function: it is very weak at long distances and becomes rapidly strongly repulsive at short distances, this being a manifestation of the Pauli exclusion principle. The dispersion interactions, on the contrary, are always attractive, arising from mutual polarisation of the electronic clouds. Charge transfer involves motion of electrons from occupied orbitals of one molecule to unoccupied orbitals of the other: this term gains importance with very strong hydrogen bonds which show *quasi*-covalent character.[6] Electrostatics is dominant in strong hydrogen bonds and represents the interaction between the partial positive charge of hydrogen and the electronegative base B. The latter is dominant in strong hydrogen bonds, where it contributes around 70% of the attractive terms and the interaction is highly directional. In weak hydrogen bonds, the relative contribution due to electrostatics is less important and the bond becomes more and more isotropic. For an overview of the theory of intermolecular forces, see section 2.9.

1.3.2 *Ab initio* and DFT calculations on hydrogen bonds

The first *ab initio* calculations appear almost 30 years ago, when Kollman *et. al.* studied several hydrogen bonded systems with single-determinant Hartree-Fock (HF) calculations and small basis sets,[7] proving that crucial information could be gained from a theoretical approach. Since then, rapid progress in theoretical chemistry and computer technology has been made. Thus, nowadays, theoretical calculations can be considered a valid alternative to experimental techniques in order to describe hydrogen bond interactions.

Since hydrogen bonding results from the interplay of several components, modelling these interactions is not an easy task. Although electrostatics are the essential forces keeping together the two groups A—H and B, dispersion forces and other subtle intermolecular interactions intervene. Thus, the most adequate theoretical tools are those that can take into account instantaneous interactions between electrons, such as high level calculations (for details on theoretical methods, see Chapter 2). Therefore in order to model properly hydrogen bonded systems, MP2 or higher level of theory is needed. However, the post-HF methods are expensive in terms of cost of calculations and they become rapidly unfeasible with the size of the system. Providentially, lower methods such as HF and density functional theory (DFT) are considered 'good enough' in many cases.

HF calculations are inexpensive and practical for systems with large number of atoms, although not capable of taking into account dispersion forces. However, it has been shown that HF level of theory is able to provide reasonable geometries for hydrogen bonded systems. Therefore, the following method is extensively used: a) the optimisation is carried out at HF level, b) a single point calculation is performed on the HF geometry at higher level of theory, such as MP2 or DFT. [8-13]

Density functional theory is considered an essential tool in modelling large systems as it is relatively cheap and reasonably accounts for electron correlation. Therefore, DFT is extensively used in treating hydrogen bonded systems.[14-18] In fact some authors consider the B3LYP functional superior to MP2 calculations, with the advantage that DFT calculations are much cheaper.[19,20]

1.3.3 GC and AT pairs

In 1979, Yanson *et al.* published a work that was destined to become a classic in the study of Watson-Crick base pairs.[21] They estimated the formation enthalpies between nucleobases in the crystalline phase and *in vacuo via* field mass spectrometry, providing the binding energies of both GC and AT, which, to date, are the reference experimental values. The binding energies, according to Yanson's work, are 21.0 kcal/mol and 12.1 kcal/mol for GC and AT respectively.

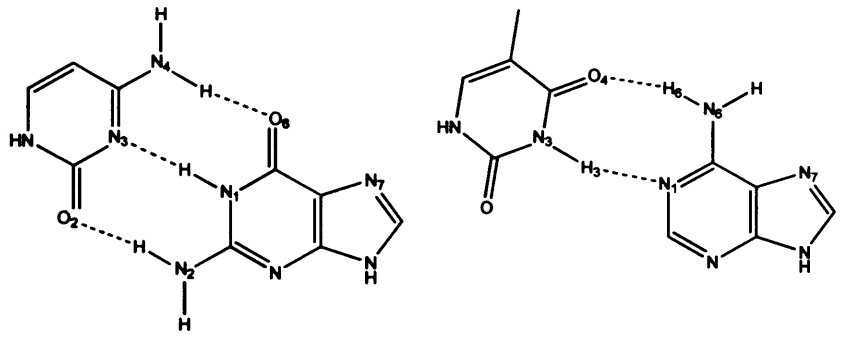


Figure 1.6 Numbering scheme for GC and AT.

Since then, much attention has been paid in order to rationalise the hydrogen bond interactions in DNA base pairs and obtain credible energies and geometries using theoretical methods.[22] The first reliable results were achieved by Hobza and co-workers optimising at HF/6-31G(d,p) and running a single point calculation at MP2 level.[13] Few years later they tested this method by fully optimising the systems at MP2 level,[23,24] suggesting that the HF/6-31G(d,p) is capable of providing reasonable accuracy for GC and AT, especially for a subsequent evaluation of molecular properties *via* single point calculations at higher level. Recently, they proposed the ultimate binding energies of GC and AT that represent, to date, the most accurate and reliable calculations.[25] Firstly they explored the potential energy surface of the base pairs with molecular dynamics simulations, localising the most important minima of the system and then evaluated them at HF/6-31G(d,p) and MP2/6-31G*(0.25).[26,27] Subsequently, they optimised those final structures at RI-MP2, using large basis sets, TZVPP ([5s3p2d1f/3s2p1d]). The binding energies for these minima were estimated as the sum of the complete basis set limit of the MP2 energy and a correction term, representing the difference between the MP2 and CCSD(T) stabilization energies, ($E_{CCSD(T)} - E_{MP2}$). Finally, the zero-point vibrational energy (ZPVE) and temperature-dependent enthalpy terms were estimated. Thus, the ultimate binding energies of GC and AT estimated via high level *ab initio* calculations are 28.8 kcal/mol and 15.4 kcal/mol, somewhat larger than experimental values.

Hobza and co-workers in another recent work, studied all the possible conformations that the canonical base pairs can adopt, finding twelve structures for the GC pair along the Potential Energy Surface (PES).[28] The most stable, about 4-10 kcal/mol lower than other minima, was indeed the Watson-Crick pair; however, many diverse hydrogen bonds

patterns were found. Thus, most GC molecules adopted Watson-Crick conformation, but other conformations were populated as well, up to 10% of the entire range of complexes. Hobza's final suggestion is that the Yanson's binding energies are too small and refer to a set of conformations, rather than just to the Watson-Crick pairing mode.

Post-HF ab initio calculations are still quite expensive, therefore there is high interest in cheaper techniques, such as HF and DFT methods. Table 1.1 reports crucial parameters of the GC and AT pairs from Guerra[29] and Hobza's calculations.[25] It is evident that, while the optimisation carried out at HF level underestimates the binding, leading to longer hydrogen bond distances, DFT calculations have the opposite drawback and overestimate the bonding. Thus, these cannot replace higher level methods as reference calculations: nonetheless, they do provide a fairly good analysis. In particular, DFT functionals reproduce qualitatively higher level calculations in computing binding energies of GC and AT. Also, the MP2//HF energies are quite close to CCSD energies: this means that HF geometries are reasonable and the approach DFT//HF or MP2//HF also provides good results and a practical performance.[13]

Table 1.1 GC and AT binding energies (kcal/mol) and geometries (Å).

	GC				AT		
	ΔE	$O_6...N_4$	$N_1...N_3$	$N_2...O_2$	ΔE	$N_6...O_4$	$N_1...N_3$
Expt.[a]	21.0	2.91	2.95	2.86	12.1	2.95	2.82
BP86/TZ2P[b]	23.8	2.73	2.88	2.87	11.8	2.85	2.81
PW91/TZ2P[b]	26.3	2.72	2.88	2.87	14.0	2.85	2.79
B3LYP/6-31G(d,p)[b]	24.0	2.79	2.93	2.92	13.2	2.94	2.84
MP2/DZP//HF/6-31G(d)[b]	25.4	2.93	3.05	3.01	11.9	3.08	3.01
MP2/CCSD[c]	28.5				15.4		

a: Yanson[21] and Rosenberg's[30,31] experimental values; b: Guerra's[29] calculations; c: Hobza's[25] calculations.

1.3.4 –NH_2 groups of purine molecules and hydrogen bonding

The high flexibility of amino groups and their involvement in hydrogen bonds is considered one of the major outcomes of quantum chemical studies of interactions of DNA bases. *Ab initio* calculations strongly suggest that the —NH_2 groups are intrinsically non-planar, with a partial sp^3 character.[32-34] Thus, the —NH_2 groups are pyramidal: at least one of the hydrogen atoms is out of plane and the nitrogen moves in the opposite direction earning a partial negative charge. In a study dated 1994, Sponer *et al.* performed HF and MP2 calculations on numerous purine and pyridine bases in order to find out the non-planar character of amino groups.[35] They suggested that the pyramidalization is greater for guanine molecules than cytosine and adenine: non-planar guanine is favoured about 1 kcal/mol over planar, while for the cytosine and adenine the energy difference drops under 1 kcal/mol.

The non-planarity of the —NH_2 groups promotes important interactions involving the nucleobases: N—H groups become able to interact *via* out-of-plane hydrogen bonds to other bases in the DNA structure. Also, nitrogen atoms, because of their partial negative charge, are able to act as hydrogen bonding acceptors. Both aspects are fundamental for the acid nucleic structure and have been recently studied *via* crystallographic analysis.[36-38]

1.4 Modelling π-stacking

Aromatic π...π stacking interactions are generally defined as the attractive interactions that occur between the π-clouds of aromatic systems in a parallel, face-to-face orientation. They play a fundamental role in many aspects of chemistry and biochemistry,[39-41] *e.g* molecular recognition,[42] self-assembly,[43,44] supramolecular chemistry and general host-guest interactions.[45-48] In biology π...π-stacking is often integral to the structure and function of proteins, co-factors and substrates.[49] While individually weak, the additive power of these interactions has large effects, DNA structure being the quintessential example.[2] In such intricate scenarios, very often the π...π interaction is considered as some sort of *deus ex machina*, intervening in reactions, stabilising complexes and influencing structure. Therefore, being able to estimate the energetic and structural features of these interactions would be extremely useful in modelling and understanding many important phenomena.

A myriad of experimental and theoretical methods have been employed to investigate π stacking interactions.[50-54] State-of-the-art electronic structure methods such as Møller-Plesset perturbation and coupled-cluster methods show that dispersive forces play the primary stabilizing role in π-stacked complexes,[52,55] as well as electrostatic and exchange-repulsion forces. Dispersion is a result of electron correlation, therefore methods that approximate or ignore electron correlation are deemed unsuitable. However, the computational resources required for correlated post-HF methods increase rapidly with molecular size, and hence are practically limited to relatively small model systems. Therefore there is considerable demand for a computationally efficient electronic structure method capable of modelling π... π stacking.

1.4.1 Nature of π-stacking

A very simple model was proposed by Hunter and Sanders[39] in order to explain the stacking behaviour of aromatic systems. They placed point charges of +1 on carbon atoms and also two associated charges of -½ above and below them. Then the interaction between the two aromatic rings was computed by summing the charge and using the Coulomb's law. Following this approach, Hunter and Saunders summarised the results of their studies in three rules:

i. π...π repulsion prevails in eclipsed conformations;

ii. π...σ attraction prevails in T-shaped conformations;

iii. π...σ attraction prevails in parallel-displaced conformations.

Several other models based upon the charge distribution have been proposed, however the aromatic interactions are very complex and a quantum mechanical approach is needed. In particular, the final structure of benzene dimers and, more generally, any aromatic stacked systems originates from the subtle interplay of electrostatics, dispersion, inductive and exchange repulsion forces (see Chapter 2 for an overview about intermolecular forces).[52] Induction and dispersion attractions are always present, both at long and short distances. At short distance, overlap between electronic clouds may occur, leading to the *so-called*

damped induction, exchange-induction, damped dispersion and exchange-dispersion contributions, proportional to the overlap. For more details on this matter, the reader is directed to Jeziorski's review, where these concepts are rigorously illustrated.[56]

1.4.2 *Ab initio* methods for modelling π-stack: benzene stacked dimers

The benzene dimer is the prototypical example of an aromatic π…π-stacked complex and a vast number of publications are available in the literature.[50-54] Two minima on the PES are found, corresponding to "T-shaped" and "parallel-displaced" geometries, with an "eclipsed" sandwich geometry transition state, as displayed in Figure 1.7. [23,57,58]

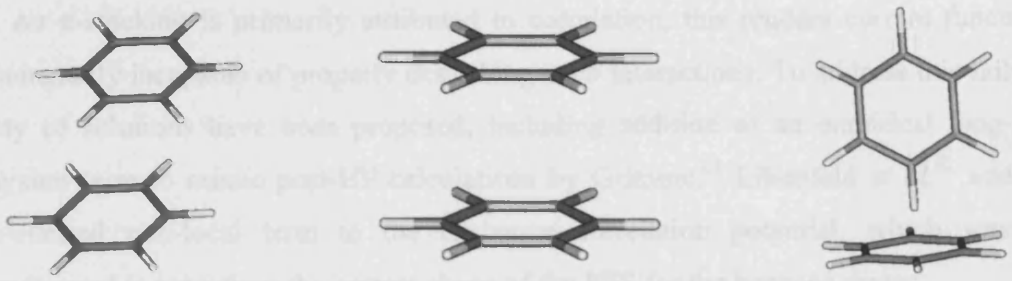


Figure 1.7 Benzene dimers: (a) parallel-displaced, (b) eclipsed and (c) T-Shaped conformation.

Sherrill *et al.* used high level *ab initio* calculations, including extrapolation to the MP2 basis set limit and inclusion of a CCSD(T) correction, to show that the T-shaped and parallel-displaced configurations are virtually isoenergetic, with binding energies of 2.74 and 2.78 kcal/mol respectively, whereas the sandwich structure is less stable at 1.81 kcal/mol.[55] These theoretically-calculated values are consistent with Grover's measurements of the benzene dimer binding energy, 2.40 ±0.41 kcal/mol.[59]

Sherrill *et al.* also found that substituted benzene dimers bind more strongly than unsubstituted benzene,[60] regardless of electron withdrawing or donating character, an

intriguing trend ascribed to the subtle interplay of electrostatic and dispersion forces. They concluded that MP2 qualitatively reproduces more accurate PES data, but consistently overestimates the binding energy of stacked complexes. Moreover, large basis sets such as aug-cc-pVQZ are required for convergence of MP2 interaction energy, and basis set superposition error (BSSE) is significant even with these.

Density functional theory has been extensively used to study many intermolecular interactions, including hydrogen bonding and C–H...π interactions, making an attractive choice due to its computational efficiency. DFT is an exact theory, and therefore can in principle model aromatic π...π-stacking. However, current approximations of the exact exchange-correlation functional prevent accurate modelling of π...π-stacking. This is due to the energy being either a function of the local density (LDA) or of the gradient of the density (GGA), hence long range electron correlation is not implicitly included (see section 2.4). As π-stacking is primarily attributed to correlation, this renders current functionals fundamentally incapable of properly describing such interactions. To address this failure, a variety of solutions have been proposed, including addition of an empirical long-range dispersion term to mimic post-HF calculations by Grimme.[61] Lilienfeld *et al.*[62] added an atom-centred non-local term to the exchange correlation potential, which was then parameterised to reproduce the correct shape of the PES for the benzene dimer.

Despite the known shortcomings, there has been much recent interest in applying DFT to π-stacking interactions. Hybrid functionals contain adjustable parameters, so it is feasible that such parameters might be appropriately adjusted to reproduce the results of higher level calculations, if only due to a cancellation of errors. This offers an attractive way to parameterise DFT with a view to mimicking more expensive high level calculations. Meijer and Sprik[63] showed that the local density approximation (LDA) functional reproduces the PES of the benzene dimer with reasonable accuracy. However, Fan *et al.*[64] found that all the density functionals tested failed to locate the energy minimum for the π-stacked benzene dimer. This conclusion was confirmed by the work of Johnson *et al.*, who reported a study of the benzene dimer using a wider variety of pure and hybrid density functionals, reaching the conclusion that none are adequate.[65] Cerny and Hobza have shown the new X3LYP functional[66] to fail for stacking of DNA bases even though it was parameterised

with a training set including dispersion bound systems.[67]

Perez-Jorda *et al.*,[68] and more recently Walsh,[69] have shown that a combination of Hartree-Fock theory and the Wilson-Levy correlation functional, the so-called HF+WL method, performs impressively in predicting the binding of a range of intermolecular interactions in weakly-bonded systems, from rare-gas dimers to π-stacked complexes. Stressing that this approach does not reproduce the known r^{-6} behaviour for dispersion forces, it is instead suggested that these results are due to non-zero overlap between interacting molecules, such that these are no longer purely dispersion-driven interactions.

1.4.3 π-stack interaction of DNA bases

The distance between DNA bases in stacked complexes is *ca.* 3.3 Å, and results from a balance of the dispersion forces and the short-range exchange-repulsion forces,[22] the mutual orientation of the bases being essentially determined by electrostatic effects.[70,71] Thus, the interaction between DNA bases is complex, and, although force field based methods qualitatively reproduce such forces, high level calculations are needed in order to describe properly stacked DNA bases. Since such calculations are extremely demanding in terms of computational costs, DNA base stacked complexes represent a great challenge for theoretical/computational chemists. Nevertheless, several data are available in the literature on π-stacking of such systems. Hobza and co-workers, for instance, have tested *ab initio* methods, MP2 and CCSD(T), for model systems such as the benzene dimer,[35] and extensively applied them to interactions between DNA bases.[25] These authors showed that the deficiencies of MP2 can be counteracted by use of a medium-sized basis set with a more diffuse polarisation function than normal, a method they termed MP2/6-31G(0.25)*.[26,27] This approach has been applied to many π-stacked systems with encouraging results, and also critically compared to empirical force-field approaches in DNA structural modelling.[72] Table 1.2 reports Hobza's[25,73] binding energies of GG, CC, GC, UU(p) dimers along with parallel UU(p) and anti-parallel UU(ap) dimers from Nielsen's work.[74]

The largest binding energy is for GC, equal to 16.90 kcal/mol, almost 6 kcal/mol more than the uracil dimers, while for GG and CC the energies are 12.90 and 10.40 kcal/mol. However, Hobza's GC and Nielsen's UU complexes are fully optimised, whereas the remainder are simply treated as rigid monomers separated at a certain distance, normally around 3.3 Å, typically the distance between two nucleobases in DNA structures. Thus, a direct comparison between fully optimised and rigid monomers is not rigorously practicable because of a) the hydrogen bonding between $—NH_2$ groups of one base and heavy atoms of the other base and b) deformation of the aromatic ring occurring in the optimised complexes.

Table 1.2 Binding energies of stacked DNA bases.

	Method	Binding Energy kcal/mol
UU(p)[a]	MP2/CCSD	11.60
GG[a]	MP2/CCSD	12.90
CC[a]	MP2/CCSD	10.40
GC[a]	MP2/CCSD[c]	16.90
AT[a]	MP2/CCSD[c]	11.60
UU(p)[b]	MP2/CCSD[c]	9.70
UU(ap)[b]	MP2/CCSD[c]	8.80

a: Hobza's[25,75] calculations; b: Nielsen's[74] calculations;
c: fully optimised at MP2 level.

A- Hydrogen bonding and deformation of the aromatic rings

As seen in section 1.3.4, $—NH_2$ groups of nucleobases are not planar and are able to interact via hydrogen bonds to the heavy atoms (N or O) of other bases. Also, the nitrogen atom can act as hydrogen bond acceptor. For instance, the full optimisation of GC performed by Hobza and co-workers clearly shows that the stabilisation energy is not simply due to π-stacking interactions, but also to hydrogen bonding, see Figure 1.8a-b. This means, firstly, that the hydrogen bonding might be able to modulate the π-stacking

interactions, but also that, in order to obtain a more realistic model of DNA bases, full optimisation should be employed.

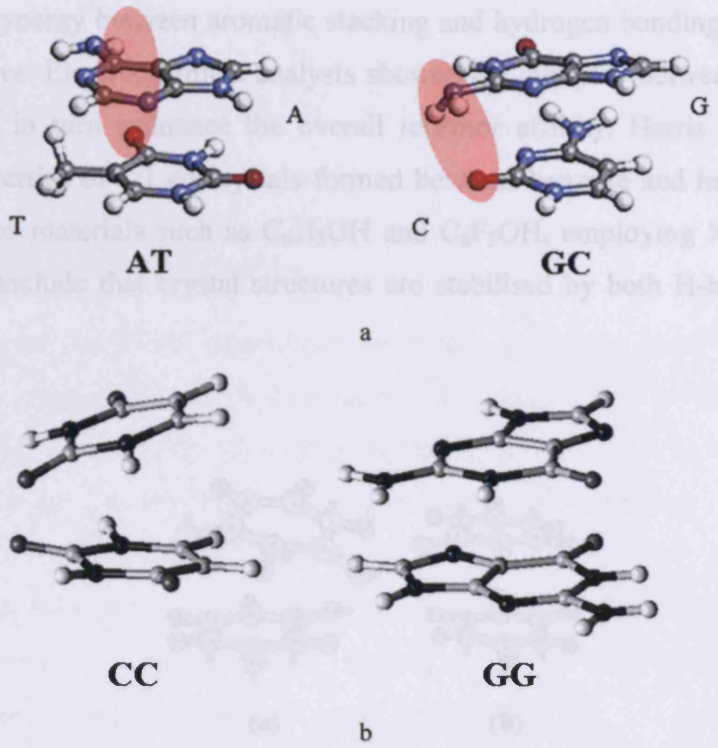


Figure 1.8 (a) Hydrogen bonding and π-stacking in fully optimised GC and AT stacked complexes[25] and (b) planar nucleobases purely π-stacked structures.[75]

Another important aspect of the fully optimised structures is the deformation of aromatic rings: Hobza and co-workers studied at MP2/6-31G*(0.25) level the stacked structures of several nucleobases, showing that significant structural deformation of the monomers occurred.[23] In particular, the deviation from the plane involves not only —NH_2 groups, but also the atoms of the aromatic rings of bases, see Figure 1.9. The degree of deformation can be monitored in terms of the difference in energy of the planar geometry and the fully relaxed one. These authors found that the cytosine deformation energy is *ca.* 1.5 kcal/mol, while is quite smaller for the AU complex, 0.7 kcal/mol. Interestingly, the comparison between the UU dimer energy of Hobza and Nielsen is excellent if the deformation energy is added to Hobza's calculations, confirming that full optimisation is needed in order to obtain a complete picture of the DNA base stacked complexes.

B - Interplay between hydrogen bonding and π-stacking

Experimental evidence of the interplay between hydrogen bond and π-stacking has been provided by numerous groups. For instance, Gray *et al.*[76] studied model systems to investigate the synergy between aromatic stacking and hydrogen bonding in the binding of a flavin derivative. Electrochemical analysis showed the interplay between H-bonding and stacking, which in turn influence the overall receptor affinity. Harris and co-workers[77] studied the properties of 1:1 co-crystals formed between benzene and hexafluorobenzene, as well as related materials such as C_6H_5OH and C_6F_5OH, employing X-ray and neutron diffraction to conclude that crystal structures are stabilised by both H-bond and stacking interactions.

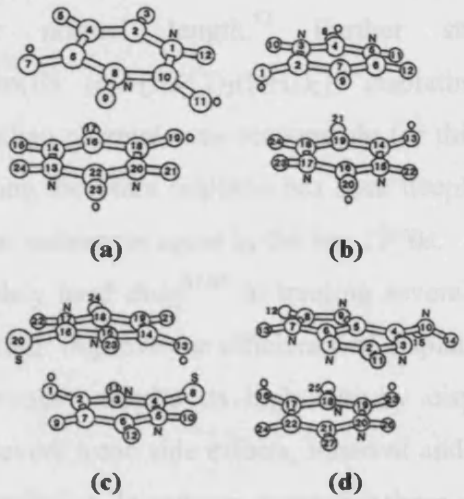


Figure 1.9 Deformation in DNA bases stacked complexes:[78]
(a) UU(p), (b) UU(ap), (c) Thiouracil-U, (d) GC.

Theoretical studies have led to similar conclusions: Geerlings and co-workers showed that in stacked complexes of pyridine and benzene, the H-bonding capacity of the pyridine nitrogen is closely related to the interaction between the aromatic rings. In particular, they suggested that electron donating substituents on benzene lead to charge transfer to pyridine, and hence to a more basic nitrogen.[79] More recent work[80] on the influence of stacking on the H-bonding ability of cytosine showed similar results: the substituted benzene was able to modulate the donor/acceptor characters of N and O atoms on the pyrimidine base. Guo *et*

al.[81] studied the effects of π-stacking on multiply H-bonded dimers of ureidopyrimidinone, showing that both the strength of H-bonds and the stability of tautomers is influenced by π-stacking. This was explained in terms of charge-transfer enhancement between the H-bonded partners.

1.5 Cisplatin-DNA adducts

At the beginning of the 1960s, Rosenberg observed a surprising phenomenon. When an electric field was applied to an aerobic solution of *Escherichia Coli* cells through platinum electrodes, the organisms did not reproduce normally, but they grew in forms of filaments 300 times longer their normal length.[82] Further studies suggested that cis-diamminodichloroplatinum(II) (cis-$[Pt(Cl)_2(NH_3)_2]$, cisplatin, or *cis*-DDP), firstly synthesised in 1844 by an Italian chemist, was responsible for this curious behaviour (see Figures 1.10 and 1.11).[83] Along the years cisplatin has been deeply studied and tested,[84-86] leading to FDA approval as an anticancer agent in the late 1970s.

Nowadays, cisplatin is a widely used drug[87,88] in treating several different tumours. Yet, research is fervent on this matter: improve the efficiency of cisplatin and reduce its toxicity are the main aims. In particular, despite its high activity cisplatin has some critical drawbacks as, for instance, severe toxic side effects, inherent and acquired resistance, and limited solubility in aqueous solution. In order to overcome these drawbacks, the search for new platinum drugs is intense. Many such platinum(II) complexes have been synthesized and tested as potential drugs, including direct analogues of the general form cis-$[PtX_2A_2]$ and more recently platinum(IV) compounds. However, only three more platinum drugs have been registered for clinical use, namely oxaliplatin,[89] carboplatin,[90,91] and nedaplatin (Figure 1.10).[92] The mechanism of platinum drugs has been studied for decades,[85,93,94] with DNA identified as the main target. In this section a literature overview has been reported.

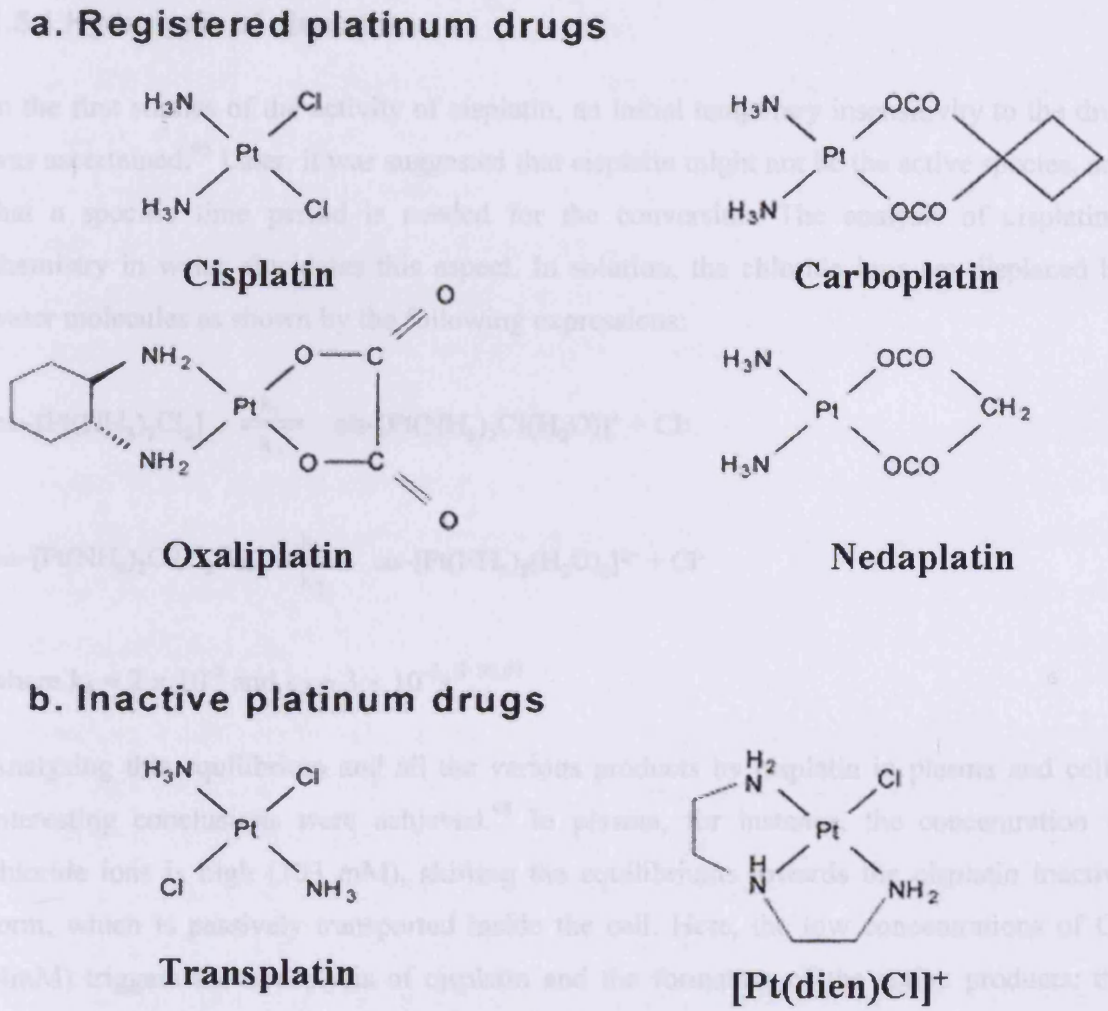


Figure 1.10 Structures of (a) the four platinum antitumor drugs currently registered for clinical use (marketed drugs) and of (b) two biologically inactive platinum compounds.[94]

a b c

Figure 1.11 Cisplatin: (a) structure (C_{2V}), (b) commercial drug, (c) effect on *Escherichia coli B* grown overnight in media containing the drug.

1.5.1 Hydrolysis of cisplatin

In the first studies of the activity of cisplatin, an initial temporary insensitivity to the drug was ascertained.[95] Later, it was suggested that cisplatin might not be the active species, and that a specific time period is needed for the conversion. The analysis of cisplatin's chemistry in water elucidates this aspect. In solution, the chloride ions are displaced by water molecules as shown by the following expressions:

$$cis\text{-}[Pt(NH_3)_2Cl_2] \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} cis\text{-}[Pt(NH_3)_2Cl(H_2O)]^+ + Cl^- \quad (1)$$

$$cis\text{-}[Pt(NH_3)_2Cl(H_2O)]^+ \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} cis\text{-}[Pt(NH_3)_2(H_2O)_2]^{2+} + Cl^- \quad (2)$$

where $k_1 \approx 2 \times 10^{-5}$ and $k_2 \approx 3 \times 10^{-5} s^{-1}$.[96,97]

Analysing this equilibrium and all the various products by cisplatin in plasma and cells, interesting conclusions were achieved.[98] In plasma, for instance, the concentration of chloride ions is high (103 mM), shifting the equilibriums towards the cisplatin inactive form, which is passively transported inside the cell. Here, the low concentrations of Cl^- (4mM) triggers the hydrolysis of cisplatin and the formation of the active products: the aqua species of mono- and bi-charged forms of cisplatin are the most likely to interact with DNA.[99,100] It is believed that the approaching step is dominated by electrostatics and, therefore, quite fast. Thus, the limiting step for reaction of cisplatin with DNA is hydrolysis, which then controls the time dependence of platinum binding to DNA.[101,102] Interestingly, the rate of dissociation of cisplatin-DNA adducts is extremely slow at body temperature, suggesting that the binding of cisplatin to DNA is kinetically rather than thermodynamically controlled.[101]

1.5.2 The mode of activated cisplatin to DNA

Once cisplatin is activated by the hydrolysis reactions, it is ready to interact with DNA. Because of the complexity of the double helix, the binding sites suitable for transition metals and, particularly, for cisplatin are numerous.[103] For instance, the positive charge of

metal could be attracted by the electron rich phosphate oxygen atoms or to the nitrogen or oxygen of the DNA bases. Also, planar molecules, such as cisplatin could interact *via* intercalation, being positioned into the space between two base pairs.[104] Spectroscopic studies of cisplatin interacting with nucleic acid macromolecules showed that a shift of absorbance of DNA from 259 to 264 nm occurred, typically indicating a direct interaction of the metal to the bases.[105] Moreover, as early studies suggested, cisplatin binds DNA noncompetitively with intercalators, confirming the covalent nature of Pt-base bond.[106]

Scheme 1 of Figure 1.12 displays the generally believed mechanism for the formation of the cisplatin-DNA adducts: one water molecule displaces the chloride ion and the first platinum-N7 bond to guanine molecule occurs; the second chloride ion leaves and cisplatin is able to bind again to either guanine or adenine, leading to ~65% of cis-GG adducts and ~20% of cis-AG adducts.[86]

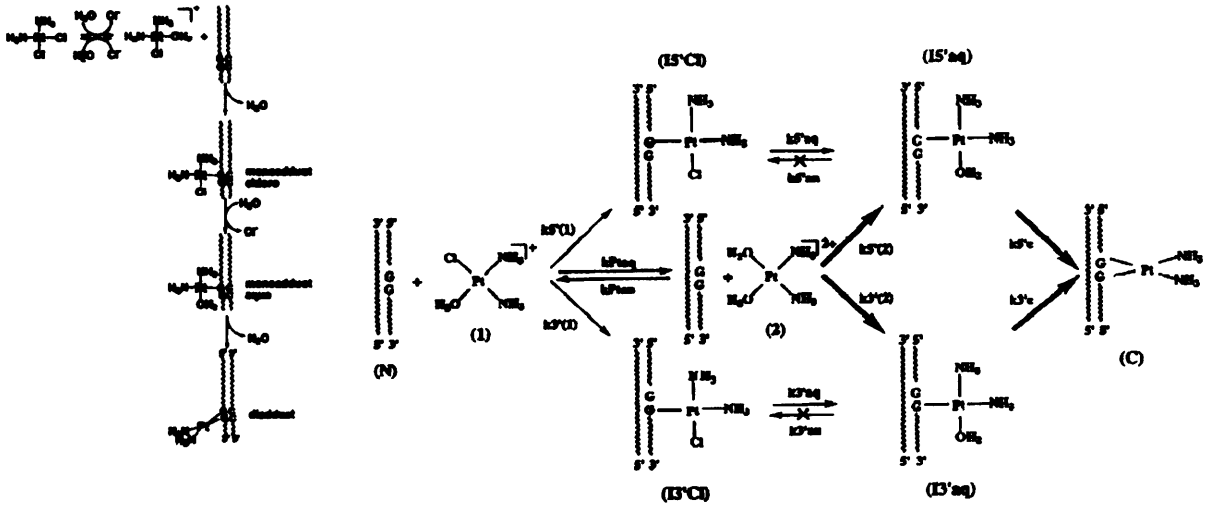
Nevertheless, recent results indicate that the picture might be more intricate. For instance, Kozelka and co-workers[107] suggested that the general assumption that the monoaquated cisplatin form is the main species reacting with DNA *in vivo*, might be incorrect. For instance, cisplatin could undergo the second hydrolysis step before binding DNA, as the electron rich side of the macromolecule could favour such process (scheme 2 of Figure 1.12).[108] In fact, Jestin *et al.* showed that this is possible, at least *in vitro*.[109] Therefore more studies are needed in order to solve this intricate mechanism and provide a better understanding of cisplatin's biochemistry.

1.5.3 Binding mode of cisplatin with DNA

Although, in principle, cisplatin could bind the phosphate oxygen atoms of the DNA structure,[110,111] the purine and pyrimidine nitrogen atoms show a greater affinity for platinum(II) complexes.[112] In particular, under neutral conditions, studies of unpaired DNA bases showed that the platinum binds the purine and pyridine in the following order: N7 of guanine > N7, N1 of adenine > N3 of cytosine.[86] Further studies of Raman difference spectrophotometry on nucleoside monophosphates interacting with cisplatin, confirm a similar order of stability: GMP > AMP >> CMP > UMP.[113] Figure 1.13 displays all the

most stable binding modes of bi-functional cisplatin to DNA: a) cisplatin binds to two adjacent purine bases, such as cisGG or cisGA, the 1,2-intrastrand cross-link,[114] b) cisplatin binds two purine molecules separated by a third molecule, such as cisGxG, the 1,3-intrastrand cross-link[115], c) cisplatin binds purine molecules belonging to two different double helix structures, the interstrand cross-link.[116]

In vivo studies[117,118] showed that the 1,2-intrastrand adducts are the major products formed from the interaction between cisplatin and DNA. By using immunochemical methods, the cisplatin-DNA complexes were taken from cells and subsequently examined, providing evidence that for one patient the ratio of cisGG and cisGA adducts is roughly 3:1, 65% and 20% respectively. Thus, these facts would suggest that the 1,2-intrastrand adduct may be important for cisplatin's anticancer activity: in fact, it is not only the major product both *in vivo* and *in vitro*, also compounds unable to bind in such way are clinically inactive. Therefore, a volume of work has been done in order to characterize the 1,2-intrastrand adduct.



scheme 1 *scheme 2*

Figure 1.12 Reaction paths of activated cisplatin.[107]

1.5.4 Platination effect on DNA structure: cell death

Experimental studies of cisplatin-DNA adducts showed that the effect of platination is strong: the structure of DNA is severely shortened and bent,[119] leading to a loss of stability of the double helix.[120] In order to obtain a better understanding of the effect of platination, several NMR and X-ray studies were used to examine cisplatin-DNA adducts. Figure 1.14 illustrates the *cis*-[$Pt(NH_3)_2$[d(pGpG)]] determined by X-ray analysis,[93] where the two guanine molecules are oriented in a "head-to-head" configuration, the O6 atoms on the same side of the platinum coordination plane. Interestingly, the dihedral angle between the guanine planes is between 76° and 87°, indicating a severe disruption in the π-stacking interaction. Another relevant element of such structure is the hydrogen bond interaction, N—H...O, occurring between the NH_3 group of cisplatin and the O6 atom, which might play an important role in stabilising such complexes.

Several other groups have focused on the structural aspects of cisplatin-DNA adducts,[121-125] in particular, high resolution X-ray studies confirm the large distortion degree of the DNA after platination, although suggesting that previous data might have overestimated the extent of the effect.[114,126] Also, these studies confirm that N—H...O hydrogen bonds between cisplatin and guanine molecules occur. Figure 1.15 displays a DNA dodecamer containing an 1,2-intrastrand d(GpG) cisplatin adduct determined by NMR.[127] The resulting structure is still a B-form conformation, bent about 78°, while the dihedral angle between the guanine molecules is 47°. The base pairs retain their hydrogen bond structure, although severe distortion of geometry occurs.

The effect of cisplatin depends on several factors, such as the nature of the DNA oligonucleotides, the position where cisplatin attacks the structure etc. However, it is clear that the main effects caused by platination are generally severe bending of the entire DNA chain. This triggers structural rearrangements, preventing DNA transcription activity and/or inducing recognition by damage repair proteins,[128] ultimately resulting with cell death through apoptosis, necrosis or both.[129] For further details the reader is directed to Lippard's recent review.[130]

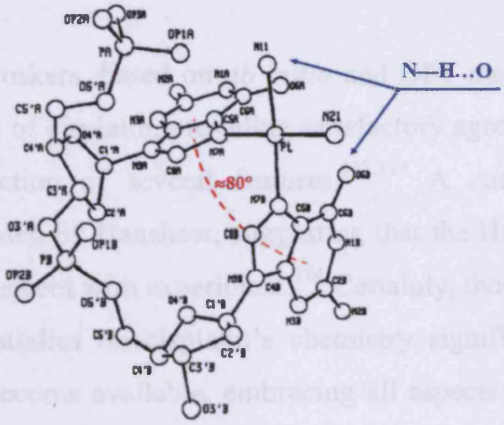


Figure 1.14 Representation of the X-ray crystal structure of *cis*-[Pt(NH3)2[d(pGpG)]], from ref. 93 and ref. 131.

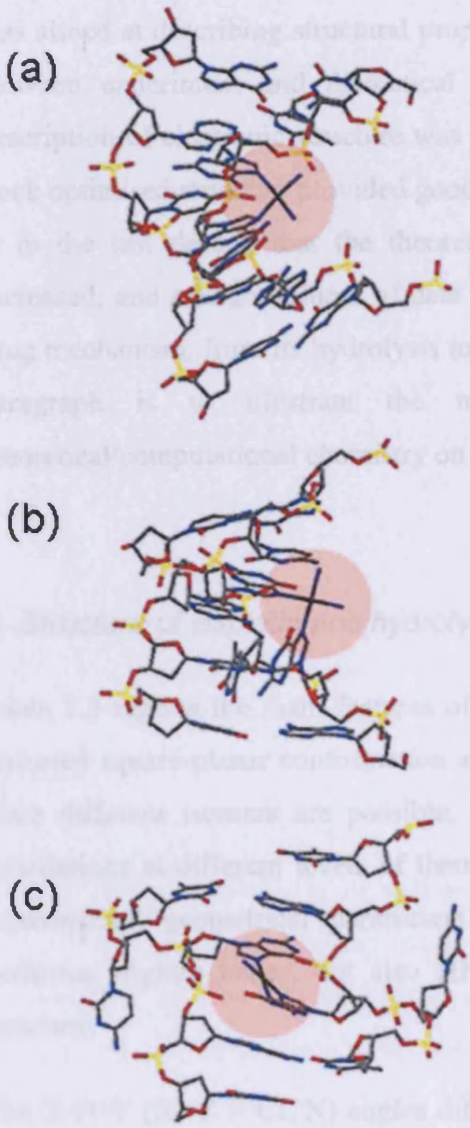


Figure 1.13 Binding mode of bi-functional cisplatin:[130] a) 1,2-intrastrand cross-link,[114] b) 1,3-intrastrand cross-link[115], c) interstrand cross-link.[116]

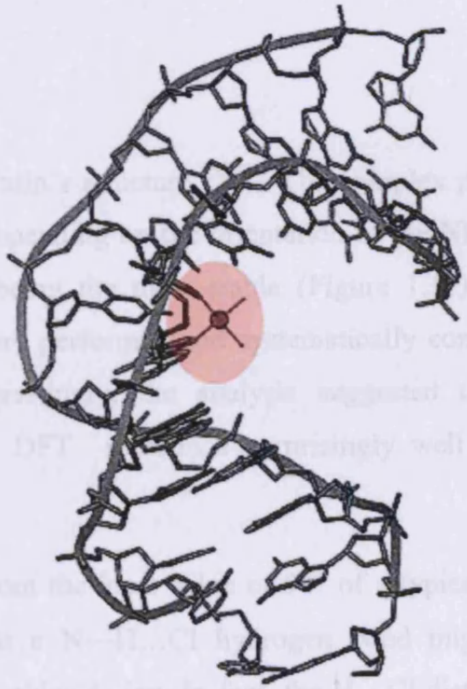


Figure 1.15 Representation of the NMR solution structure of platinated d(CCTCTG*G*TCTCC)·d(GGAGACCAGAGG) structure.[127]

1.5.5 Theoretical studies of cisplatin

Initial work performed by Carloni and co-workers, based on *ab initio* and DFT methods, was aimed at describing structural properties of cisplatin, providing satisfactory agreement between experiment and theoretical prediction of several features.[132,133] A complete description of electronic structure was proposed by Hausheer, suggesting that the Hartree-Fock optimised structure provided good agreement with experiment.[134] Certainly, though, it is in the last decade that the theoretical studies on cisplatin's chemistry significantly increased, and a large volume of data has become available, embracing all aspects of the drug mechanism, from its hydrolysis to effect on DNA structure. Therefore, the aim of this paragraph is to illustrate the major outcomes achieved by the means of theoretical/computational chemistry on platinum based-drugs.

A- Structure of cisplatin and hydrolysis

Table 1.3 reports the main features of cisplatin's structure.[133,134] The complex presents a distorted square-planar conformation and, depending on the orientation of the NH_3 group, three different isomers are possible, C_{2V} being the most stable (Figure 1.11). Several calculations at different levels of theory were performed and systematically compared to experimental geometrical parameters. Interestingly, the analysis suggested that MP2 performs slightly better, but also HF and DFT reproduce surprisingly well cisplatin structure.

The X-Pt-Y (X, Y = Cl, N) angles differ from the ideal value of 90° of a typical square-planar complex. It has been proposed that a N—H...Cl hydrogen bond might occur between the NH_3 group of cisplatin and the chloride ion. In fact, the H...Cl distance was 2.39 Å, from Carloni's calculations, indicating a possible hydrogen bond. Similarly, Hausheer found that the H...Cl distance ranges between 2.33 and 2.57 Å, depending on the theoretical method. In the original crystal structure, the intra-molecular distance N...Cl is 3.01,[135] which is close to the theoretical values and within a weak N—H...Cl hydrogen bond.[136] Hausheer also performed a charge analysis on the atoms involved in this supposed interaction, confirming the hypothesis of N—H...Cl hydrogen bond.

Table 1.3 Theoretical geometrical parameters of cisplatin (C_{2v} symmetry).

	HF[a]	MP2[a]	DFT[b]	Expt.[c]
Pt—N (Å)	2.139	2.090	2.065	2.01±0.04
Pt—Cl (Å)	2.348	2.312	2.315	2.33±0.01
∠N-Pt-N (°)	95.0	96.5	98.0	
∠N-Pt-Cl (°)	84.7	84.9	83.0	
∠Cl-Pt-Cl (°)	95.6	93.8	95.5	

a: Hausheer's;[134] b: Carloni's;[133] c: experimental values in the solvent-free crystal.[135]

Several studies focused on the analysis of cisplatin's hydrolysis,[137-140] among these, Zhang and co-workers[141] employed the recently developed mPW1PW91/SDD[142] functional. Figure 1.16 displays the hydrolysis scheme as reported in Zhang's work: cisplatin (1) approaches a water molecule, forming a hydrogen bonded complex (2), the intermediate (I1); the structure evolves towards a five-coordinate TBP-like transition state (TS), Figure 1.17; the first chloride ion leaves, forming the second intermediate (I2): the product is formed (P). The second step of the hydrolysis follows a similar path, with two intermediates ($I1_2$, $I2_2$), a transition state (TS_2) and the product (P_2).

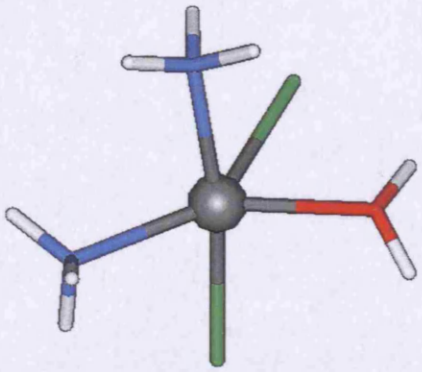


Figure 1.17 Transition state (TS1).

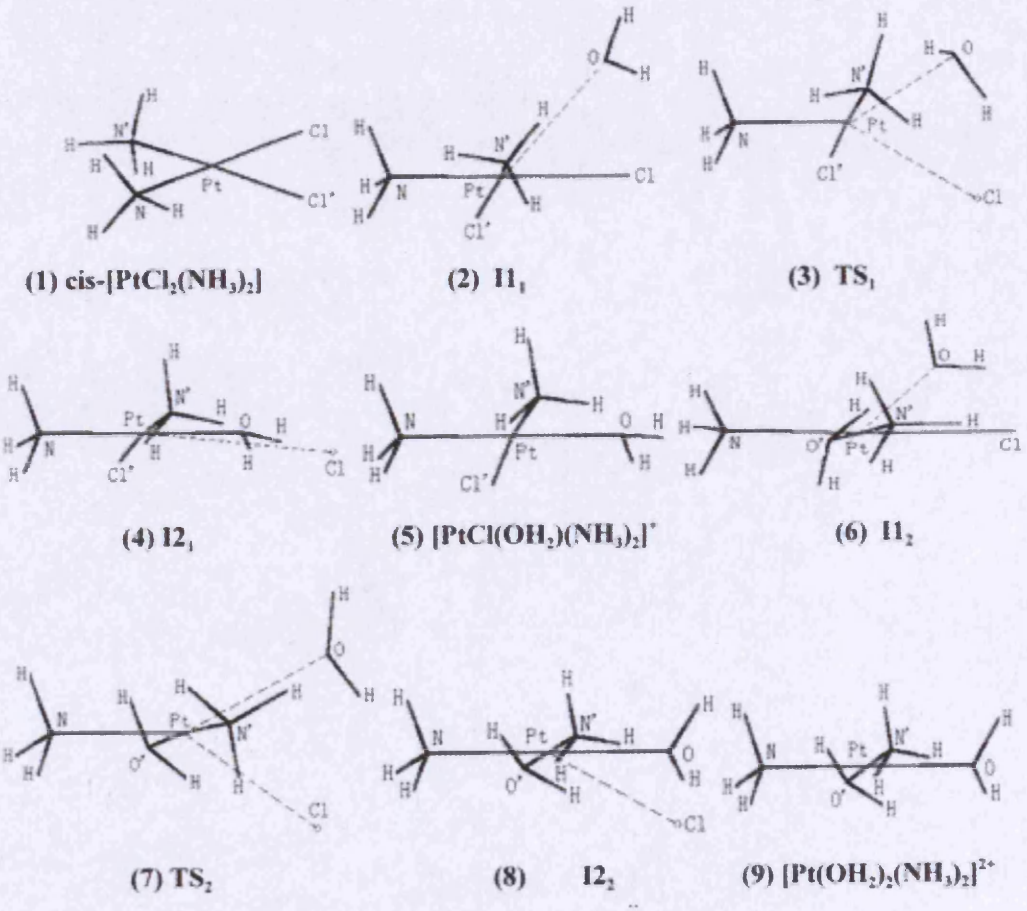
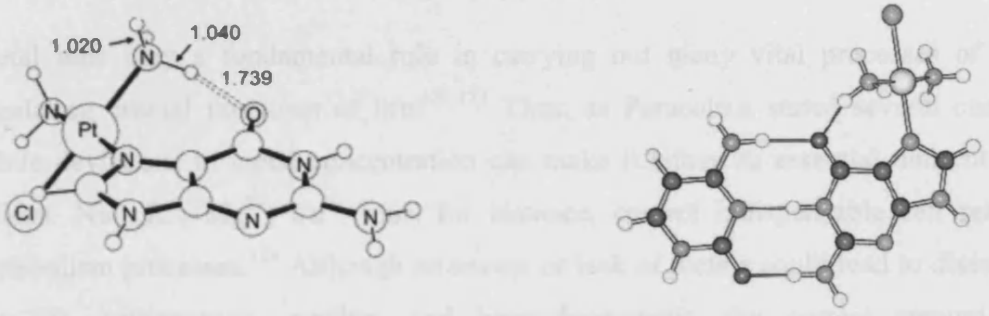


Figure 1.16 Cisplatin hydrolysis scheme.[141]

The results available in literature suggest that in order to reproduce the experimental reaction rate, the inclusion of the solvent is needed. For further details on this, the reader is directed to Zhang[141] and Carloni's[143] publications.

B- Cisplatin-purine complexes

The interaction between cisplatin and DNA has been widely studied with several methods in order to glean crucial information regarding the known specificity of cisplatin for certain sites, and the effect of platination on nucleic acids.[144-147]



a b

Figure 1.18 Cisplatin DNA bases complexes: (a) cisplatin...guanine[146] and (b) cisplatin...GC adducts.[147]

Leszczynski *et al.*[147] have performed extensive DFT and *ab initio* calculations on complexes of platinum with one and two DNA bases, in order to examine the fundamental properties of platinum-DNA interaction. As expected, they found the G-Pt-G structure to be the most stable, along with the mixed complex A-Pt-G. In order to clarify the preference of cisplatin for guanine over adenine, Lippard and co-workers[146] carried out DFT studies of adenine and guanine complexes with $[Pt(Cl)(NH_3)_2]^+$: both thermodynamics and kinetics of the complexes were taken into account, confirming that guanine is up to 20 times more reactive than adenine toward cisplatin. Furthermore the ability of Pt to bind to purines was studied to elucidate the features of the Pt-purine interaction, suggesting a lack of π-back donation between metal and base.[148]

Carloni's group[143] used Car-Parrinello MD methods to study the Pt-DNA bond, along with some thermodynamic aspects of hydrolysis of cisplatin: good agreement with reported experimental data confirmed the success of this method in treating cisplatin biochemistry. Furthermore, several studies indicate that although severely distorted, the hydrogen bond pattern in GC pair is essentially retained.[149-151]

1.6 Transition metal-DNA adducts

Metal ions play a fundamental role in carrying out many vital processes of organisms, regulating crucial functions of life.[152,153] Thus, as Paracelsus stated several centuries ago, subtle deviations in metal concentration can make it either an essential nutrient or a lethal poison. Na^+, K^+, Mg^{2+}, Ca^{2+} ions, for instance, control indispensable cell reactions and metabolism processes.[154] Although an excess or lack of metals could lead to diseases such as anaemia, osteoporosis, sterility, and hemochromatosis, the correct amount might be extremely helpful in curing other diseases.[155,156] In particular, anti-tumour activity is a useful property of metals, especially transition metals. The important role of platinum based drugs is well known, but other metals such as gold, ruthenium, titanium, and vanadium also show promise in this area.[87,157]

Metal ions may interact with many different sites of nucleic acids, generally those with local concentrations of electron density. They can bind the canonical DNA structure or lead to noncanonical forms, such as triplexes, quadruplexes, junctions etc. As seen above, the favoured sites are the negatively charged phosphate groups and the electron rich N and O atoms of the bases.

Metallocene complexes of vanadium and titanium are believed to have spermicidal activity: Uckun and co-workers[158] argued that although the effect on cells is similar to cisplatin, the mechanism might be different, namely, the metal-generated radicals such as OH· could lead to apoptosis. Ruthenium complexes appear to act in a similar manner as platinum drugs, penetrating the cell and attacking nucleic acids at the N7 of guanine.[159] Interesting

compounds include Sadler's[160,161] [(η^6-arene)Ru(X)(Y)(Z)] complexes, where arene is benzene or substituted benzene, and X, Y, and Z are halide, acetonitrile or isonicotinamide, which inhibit cell growth in a similar manner to platinum drugs.

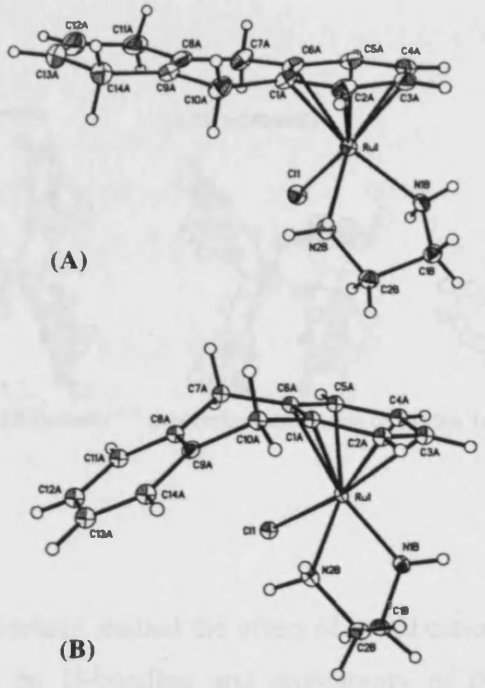


Figure 1.19 X-ray structures and atom numbering schemes for Sadler's[162] [(η^6-$C_{14}H_{14}$)Ru(en)Cl]$^+$ (A) and [(η^6-$C_{14}H_{12}$)Ru(en)Cl]$^+$ (B).

Gold complexes are promising candidates for cytotoxic and anti-cancer agents.[163] Gold(I) complexes are found to be active against various types of cancer, killing cells via damage to mitochondria.[164] In vitro studies indicate that some gold(III)[165] complexes are highly cytotoxic toward cultured human tumour cell lines, and are able to overcome resistance to platinum. Significant differences in the mode of action are observed compared with cisplatin, with generally weaker binding to DNA in vitro. Among other transition metals, Rh,[166] Ir,[167] Cu[168] and Co[169] are considered potentially active against tumours. Thus, transition metals represent a fundamental source of possible new drugs for treating cancer or other diseases.

Several theoretical/computational studies of metals other than platinum have been reported: Hobza *et al.* showed that metallation (M = Mg^{2+}, Hg^{2+}) distorts the Watson Crick pair in a

similar manner to cisplatin, weakening the N—H...O_6 H-bond and strengthening the O_2...H—N_2 H-bond.[170] They also found that Hartree-Fock metal-guanine binding energies do not differ significantly from MP2 values, confirming that the main source of bonding is electrostatic, and further that Hartree-Fock methods can be reasonably applied to such complexes.

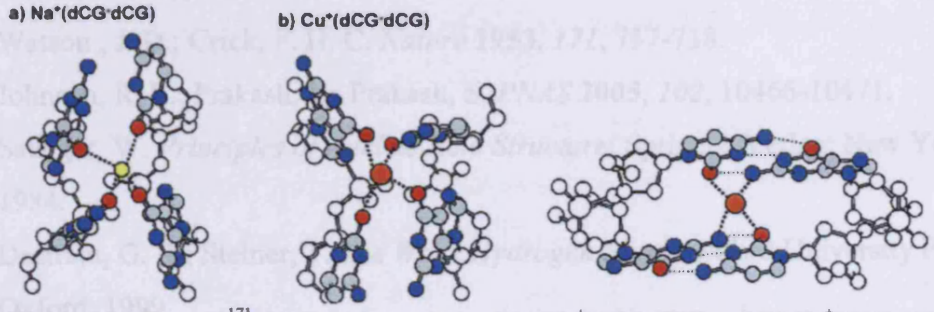


Figure 1.20 Bowers'[171] theoretical structures of (a) Na^+(dCG·dCG) and (b) Cu^+(dCG·dCG).

Poater and co-workers studied the effect of metal cations (M = Cu^+, Ca^{2+} and Cu^{2+}) bound to N7 of guanine on H-bonding and aromaticity of the guanine-cytosine pair, using DFT methods.[172] They explained the distortion of the Watson Crick pair in terms of modification of donor-acceptor character of the atoms involved in H-bonding. Also, they suggested the increase of the aromaticity of the guanine and cytosine six-membered rings as a major factor in the redistribution of H-bond energy. Recently, Bowers *et al.*[171] reported experimental and theoretical work on this matter (Figure 1.20). Employing ion mobility mass spectrometry techniques, they analysed the behaviour of several metal cations in gas phase, suggesting that "soft" metal acids such as Pt bond to guanine N7, and hence stabilize the Watson-Crick structure of dCG·dCG. In contrast, "hard" cations, such as Li^+, Na^+, K^+, Cr^{2+}, Mn^{2+} etc., prefer O6 metallation, and so promote a globular structure of the dinucleotide duplex in which the GC pair is disrupted.

1.7 References

(1) Voet, D.; Voet, J. G.; Pratt, C. W. *Fundamentals of Biochemistry*; Wiley & sons, Inc: New York, 1999.

(2) Watson , J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737-738.

(3) Johnson, R. E.; Prakash, L.; Prakash, S. *PNAS* **2005**, *102*, 10466-10471.

(4) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(5) Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond*; Oxford University Press: Oxford, 1999.

(6) Gilli, P.; Bertolasi, V.; Ferretti, V.; Gilli, G. *J. Am. Chem. Soc* **1994**, *116*, 909-915.

(7) Kollman, P. A.; Allen, L. C. *Chem. Rev.* **1972**, *72*, 283-&.

(8) Florian, J.; Leszczynski, J. *J. Am. Chem. Soc.* **1996**, *118*, 3010-3017.

(9) Giuliano Alagona, C. G. *Int. J. Quantum Chem.* **2002**, *90*, 641-656.

(10) Gu, J.; Leszczynski, J. *J. Phys. Chem. A* **2000**, *104*, 1898-1904.

(11) Koo, J. C. P.; Lam, J. S. W.; Salpietro, S. J.; Chass, G. A.; Enriz, R. D.; Torday, L. L.; Varro, A.; Papp, J. G. *J. Mol. Struct.: THEOCHEM* **2002**, *619*, 143-194.

(12) Nagy, P. I.; Alagona, G.; Ghio, C. *J. Am. Chem. Soc.* **1999**, *121*, 4804-4815.

(13) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965-1974.

(14) Novoa, J. J.; Sosa, C. *J. Phys. Chem.* **1995**, *99*, 15837-15845.

(15) Christoph Maerker; Paul Von R. Schleyer; Klaus R. Liedl; T. K. Ha; Martin Quack ; Suhm, M. A. *J. Comput. Chem.* **1997**, *18*, 1695-1719.

(16) Dong, X. L.; Zhou, Z. Y.; Tian, L. J.; Zhao, G. *Int. J. Quantum Chem.* **2005**, *102*, 461-469.

(17) Gong, X.-l.; Zhou, Z.-y.; Zhang, H.; Liu, S.-z. *J. Mol. Struct.: THEOCHEM* **2005**, *718*, 23-29.

(18) Lozynski, M.; Rusinska-Roszak, D.; Mack, H. G. *J. Phys. Chem. A* **1998**, *102*, 2899-2903.

(19) Rablen, P. R.; Lockman, J. W.; Jorgensen, W. L. *J. Phys. Chem. A* **1998**, *102*, 3782-3797.

(20) Chung, G.; Kwon, O.; Kwon, Y. *J. Phys. Chem. A* **1998**, *102*, 2381-2387.

(21) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149-1170.

(22) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Mol. Struct.: THEOCHEM* **2001**, *573*, 43-53.

(23) Hobza, P.; Sponer, J. *Chem. Phys. Lett.* **1998**, *288*, 7-14.

(24) Hobza, P.; Sponer, J.; Cubero, E.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2000**, *104*, 6286-6292.

(25) Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608-15613.

(26) 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.

(27) Sponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76-87.

(28) Kabelac, M.; Hobza, P. *J. Phys. Chem. B* **2001**, *105*, 5804-5817.

(29) FonsecaGuerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117-4128.

(30) Rosenberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 145-167.

(31) Seeman, N. C.; Rosenberg, J. M.; Suddath, F. L.; Parkkim, J. J.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 109-144.

(32) Leszczynski, J. *Int. J. Quantum Chem.* **1992**, *44*, 43-55.

(33) Bludsky, O.; Sponer, J.; Leszczynski, J.; Spirko, V.; Hobza, P. *J. Chem. Phys* **1996**, *105*, 11042-11050.

(34) Sponer, J.; Hobza, P. *Int. J. Quantum Chem.* **1996**, *57*, 959-970.

(35) Sponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161-3164.

(36) Sponer, J.; Hobza, P.; Florián, J.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **1996**, *13*, 827-833.

(37) Luisi, B.; Orozco, M.; Sponer, J.; Luque, F. J.; Shakked, Z. *J. Mol. Biol.* **1998**, *279*, 1123-1136.

(38) Vlieghe, D.; Sponer, J.; Meervelt, L. V. *Biochemistry* **1999**, *38*, 16443-16451.

(39) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc* **1990**, *112*, 5525-5534.

(40) Burley, S. K. P., G. A *Science* **1985**, 23.

(41) Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 837-846.

(42) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem.-Int. Edit.* **2003**, *42*, 1210-1250.

(43) Claessens, C. G.; Stoddart, J. F. *J. Phys. Org. Chem.* **1997**, *10*, 254-272.

(44) Fyfe, M. C. T.; Stoddart, J. F. *Acc. Chem. Res.* **1997**, *30*, 393-401.

(45) Askew, B.; Ballester, P.; Buhr, C.; Jeong, K. S.; Jones, S.; Parris, K.; Williams, K.; Rebek, J. *J. Am. Chem. Soc* **1989**, *111*, 1082-1090.

(46) Smithrud, D. B.; Diederich, F. *J. Am. Chem. Soc* **1990**, *112*, 339-343.

(47) Hunter, C. A. *Chem. Soc. Rev.* **1994**, *23*, 101-109.

(48) Rebek, J. *Chem. Soc. Rev.* **1996**, *25*, 255-&.

(49) McGaughey, G. B.; Gagne, M.; Rappe, A. K. *J. Biol. Chem.* **1998**, *273*, 15458-15463.

(50) Tsuzuki, S.; Uchimaru, T.; Tanabe, K. *Theochem-J. Mol. Struct.* **1994**, *113*, 107-118.

(51) Tsuzuki, S.; Luthi, H. P. *J. Chem. Phys.* **2001**, *114*, 3949-3957.

(52) Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. *Chem. Rev.* **2000**, *100*, 4145-4185.

(53) Steed, J. M.; Dixon, T. A.; Klemperer, W. *J. Chem. Phys.* **1979**, *70*, 4940-4946.

(54) Arunan, E.; Gutowsky, H. S. *J. Chem. Phys.* **1993**, *98*, 4294-4296.

(55) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc* **2002**, *124*, 10887-10893.

(56) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887-1930.

(57) Jaffe, R. L.; Smith, G. D. *J. Chem. Phys.* **1996**, *105*, 2780-2788.

(58) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790-18794.

(59) Grover, J. R.; Walters, E. A.; Hui, E. T. *J. Phys. Chem.* **1987**, *91*, 3233-3237.

(60) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc* **2004**, *126*, 7690-7697.

(61) Grimme, S. *J. Comput. Chem* **2004**, *25*, 1463-1473.

(62) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, art. no.-153004.

(63) Meijer, E. J.; Sprik, M. *J. Chem. Phys.* **1996**, *105*, 8684-8689.

(64) Ye, X. Y.; Li, Z. H.; Wang, W. N.; Fan, K. N.; Xu, W.; Hua, Z. Y. *Chem. Phys. Lett.* **2004**, *397*, 56-61.

(65) Johnson, N. P.; Hoeschele, J. D.; Rahn, R. O. *Chem.-Biol. Interact.* **1980**, *30*, 151-169.

(66) Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673-2677.

(67) Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624-1626.

(68) Perez-Jorda, J. M.; San-Fabian, E.; Perez-Jimenez, A. J. *J. Chem. Phys.* **1999**, *110*, 1916-1920.

(69) Walsh, T. R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 443-451.

(70) Poltev, V. I.; Shulyupina, N. V. *J. Biomol. Struct. Dyn.* **1986**, *3*, 739-765.

(71) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590-5596.

(72) Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1997**, *18*, 1136-1150.

(73) Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802-11808.

(74) Leininger, M. L.; Nielsen, I. M. B.; Colvin, M. E.; Janssen, C. L. *J. Phys. Chem. A* **2002**, *106*, 3850-3854.

(75) Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802-11808.

(76) Gray, M.; Goodman, A. J.; Carroll, J. B.; Bardon, K.; Markey, M.; Cooke, G.; Rotello, V. M. *Org. Lett.* **2004**, *6*, 385-388.

(77) Meejoo, S.; Kariuki, B. M.; Harris, K. D. M. *ChemPhysChem* **2003**, *4*, 766-769.

(78) Hobza, P.; Sponer, J. *Chem. Phys. Lett.* **1998**, *288*, 7-14.

(79) Mignon, P.; Loverix, S.; Geerlings, P. *Chem. Phys. Lett.* **2005**, *401*, 40-46.

(80) Mignon, P.; Loverix, S.; Steyaert, J.; Geerlings, P. *Nucleic Acids Res.* **2005**, *33*, 1779-1789.

(81) Guo, D. W.; Sijbesma, R. P.; Zuilhof, H. *Org. Lett.* **2004**, *6*, 3667-3670.

(82) Rosenberg, B.; Van Camp, L.; Krigas, T. *Nature* **1965**, *205*, 698.

(83) Rosenberg, B.; VanCamp, L.; Grimley, E. B.; Thomson, A. J. *J. Biol. Chem.* **1967**, *242*, 1347.

(84) Rosenberg, B.; VanCamp, L.; Trosko, J. E.; Mansour, V. H. *Nature (London)* **1969**, *222*, 385.

(85) Wong, E.; Giandomenico, C. M. *Chem. Rev.* **1999**, *99*, 2451-2466.

(86) Sherman, S. E.; Lippard, S. J. *Chem. Rev.* **1987**, *87*, 1153-1181.

(87) Gordon, M.; Hollander, S. *J. Med.* **1993**, *24*, 209-265.

(88) Weiss, R. B.; Christian, M. C. *Drugs* **1993**, *46*, 360-377.

(89) Cvitkovic, E. *Semin. Oncol.* **1998**, *25*, 1-3.

(90) Reedijk, J. *Chem. Commun.* **1996**, 801-806.

(91) Eastman, A.; Barry, M. A. *Biochemistry* **1987**, *26*, 3303-3307.

(92) Lebwohl, D.; Canetta, R. *Eur. J. Cancer* **1998**, *34*, 1522-1534.

(93) Sherman, S. E.; Gibson, D.; Wang, A. H. J.; Lippard, S. J. *J. Am. Chem. Soc.* **1988**, *110*, 7368-7381.

(94) Fuertes, M. A.; Alonso, C.; Perez, J. M. *Chem. Rev.* **2003**, *103*, 645-662.

(95) Harder, H. C.; Rosenber.B *Int. J. Cancer* **1970**, *6*, 207-&.

(96) Aprile, F.; Martin, D. S. *Inorg. Chem.* **1962**, *1*, 551-557.

(97) Reishus, J. W.; Martin, D. S. *J. Am. Chem. Soc.* **1961**, *83*, 2457-2462.

(98) Lim, M. C.; Martin, R. B. *J. Inorg. Nucl. Chem.* **1976**, *38*, 1911-1914.

(99) Basolo, F.; Gray, H. B.; Pearson, R. G. *J. Am. Chem. Soc.* **1960**, *82*, 4200-4203.

(100) Gray, H. B.; Olcott, R. J. *Inorg. Chem.* **1962**, *1*, 481-485.

(101) Ushay, H. M.; Tullius, T. D.; Lippard, S. J. *Biochemistry* **1981**, *20*, 3744-3748.

(102) Johnson, N. P.; Hoeschele, J. D.; Rahn, R. O. *Chem.-Biol. Interact.* **1980**, *30*, 151-169.

(103) Barton, J. K.; Lippard, S. J. *Nucleic Acid-Metal Ion Interactions*; Wiley: New York, 1980.

(104) Shieh, H. S.; Berman, H. M.; Dabrow, M.; Neidle, S. *Nucleic Acids Res.* **1980**, *8*, 85-97.

(105) Horacek, P.; Drobnik, J. *Biochim. Biophys. Acta* **1971**, *254*, 341-&.

(106) Howegrant, M.; Wu, K. C.; Bauer, W. R.; Lippard, S. J. *Biochemistry* **1976**, *15*, 4339-4346.

(107) Kozelka, J.; Legendre, F.; Reeder, F.; Chottard, J. C. *Coord. Chem. Rev.* **1999**, *192*, 61-82.

(108) Mills, P.; Anderson, C. F.; Record, M. T. *J. Phys. Chem.* **1985**, *89*, 3984-3994.

(109) Jestin, J. L.; Lambert, B.; Chottard, J. C. *J. Biol. Inorg. Chem.* **1998**, *3*, 515-519.

(110) Reily, M. D.; Marzilli, L. G. *J. Am. Chem. Soc* **1986**, *108*, 6785-6793.

(111) Bose, R. N.; Cornelius, R. D.; Viola, R. E. *J. Am. Chem. Soc* **1986**, *108*, 4403-4408.

(112) Peloso, A. *Coord. Chem. Rev.* **1973**, *10*, 123-181.

(113) Mansy, S.; Chu, G. Y. H.; Duncan, R. E.; Tobias, R. S. *J. Am. Chem. Soc* **1978**, *100*, 607-616.

(114) Takahara, P. M.; Rosenzweig, A. C.; Frederick, C. A.; Lippard, S. J. *Nature* **1995**, *377*, 649-652.

(115) Van Garderen, C. J.; Van Houte, L. P. A. *Eur. J. Biochem.* **1994**, *225*, 1169-1179.

(116) Huang, H.; Zhu, L.; Reid, B. R.; Drobny, G. P.; Hopkins, P. B. *Science* **1995**, *270*, 1842-1845.

(117) Eastman, A. *Biochemistry* **1986**, *25*, 3912-3915.

(118) Fichtingerschepman, A. M. J.; Vanderveer, J. L.; Denhartog, J. H. J.; Lohman, P. H. M.; Reedijk, J. *Biochemistry* **1985**, *24*, 707-713.

(119) Cohen, G. L.; Bauer, W. R.; Barton, J. K.; Lippard, S. J. *Science* **1979**, *203*, 1014-1016.

(120) Poklar, N.; Pilch, D. S.; Lippard, S. J.; Redding, E. A.; Dunham, S. U.; Breslauer, K. J. *PNAS* **1996**, *93*, 7606-7611.

(121) Sip, M.; Schwartz, A.; Vovelle, F.; Ptak, M.; Leng, M. *Biochemistry* **1992**, *31*, 2508-2513.

(122) Herman, F.; Kozelka, J.; Stoven, V.; Guittet, E.; Girault, J. P.; Huynhdinh, T.; Igolen, J.; Lallemand, J. Y.; Chottard, J. C. *Eur. J. Biochem.* **1990**, *194*, 119-133.

(123) Rice, J. A.; Crothers, D. M.; Pinto, A. L.; Lippard, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 4158-4161.

(124) Denhartog, J. H. J.; Altona, C.; Vanboom, J. H.; Vandermarel, G. A.; Haasnoot, C. A. G.; Reedijk, J. *J. Biomol. Struct. Dyn.* **1985**, *2*, 1137-1155.

(125) Denhartog, J. H. J.; Altona, C.; Vandermarel, G. A.; Reedijk, J. *Eur. J. Biochem.* **1985**, *147*, 371-379.

(126) Takahara, P. M.; Frederick, C. A.; Lippard, S. J. *J. Am. Chem. Soc.* **1996**, *118*, 12309-12321.

(127) Gelasco, A.; Lippard, S. J. *Biochemistry* **1998**, *37*, 9230-9239.

(128) Kruidering, M.; van de Water, B.; Zhan, Y.; Baelde, J. J.; de Heer, E.; Mulder, G. J.; Stevens, J. L.; Nagelkerke, J. F. *Cell Death Differ.* **1998**, *5*, 601-614.

(129) Gonzalez, V. M.; Fuertes, M. A.; Alonso, C.; Perez, J. M. *Mol. Pharmacol.* **2001**, *59*, 657-663.

(130) Jamieson, E. R.; Lippard, S. J. *Chem. Rev.* **1999**, *99*, 2467-2498.

(131) Sherman, S. E.; Gibson, D.; Wang, A. H. J.; Lippard, S. J. *Science* **1985**, *230*, 412-417.

(132) Tornaghi, E.; Andreoni, W.; Carloni, P.; Hutter, J.; Parrinello, M. *Chem. Phys. Lett.* **1995**, *246*, 469-474.

(133) Carloni, P.; Andreoni, W.; Hutter, J.; Curioni, A.; Giannozzi, P.; Parrinello, M. *Chem. Phys. Lett.* **1995**, *234*, 50-56.

(134) Pavankumar, P. N. V.; Seetharamulu, P.; Yao, S.; Saxe, J. D.; Reddy, D. G.; Hausheer, F. H. *J. Comput. Chem.* **1999**, *20*, 365-382.

(135) Milburn, G. H. W.; Truter, M. R. *J. Chem. Soc. A* **1966**, 1609.

(136) Mareque Rivas, J. C.; Brammer, L. *Inorg. Chem.* **1998**, *37*, 4756-4757.

(137) Zhu, C. B.; Raber, J.; Eriksson, L. A. *J. Phys. Chem. B* **2005**, *109*, 12195-12205.

(138) Raber, J.; Zhu, C. B.; Eriksson, L. A. *J. Phys. Chem. B* **2005**, *109*, 11006-11015.

(139) Raber, J.; Zhu, C. B.; Eriksson, L. A. *Mol. Phys.* **2004**, *102*, 2537-2544.

(140) Burda, J. V.; Zeizinger, M.; Leszczynski, J. *J. Comput. Chem.* **2005**, *26*, 907-914.

(141) Zhang, Y.; Guo, Z.; You, X. Z. *J. Am. Chem. Soc.* **2001**, *123*, 9378-9387.

(142) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664-675.

(143) Carloni, P.; Sprik, M.; Andreoni, W. *J. Phys. Chem. B* **2000**, *104*, 823-835.

(144) Pelmenschikov, A.; Zilberberg, I.; Leszczynski, J.; Famulari, A.; Sironi, M.; Raimondi, M. *Chem. Phys. Lett.* **1999**, *314*, 496-500.

(145) Zilberberg, I. L.; Avdeev, V. I.; Zhidomirov, G. M. *J. Mol. Struct.: THEOCHEM* **1997**, *418*, 73-81.

(146) Baik, M. H.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2003**, *125*, 14082-14092.

(147) Burda, J. V.; Leszczynski, J. *Inorg. Chem.* **2003**, *42*, 7162-7172.

(148) Baik, M. H.; Friesner, R. A.; Lippard, S. J. *Inorg. Chem.* **2003**, *42*, 8615-8617.

(149) Sigel, R. K. O.; Lippert, B. *Chem. Commun.* **1999**, 2167-2168.

(150) Burda, J. V.; Sponer, J.; Leszczynski, J. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4404-4411.

(151) Sigel, R. K. O.; Freisinger, E.; Lippert, B. *J. Biol. Inorg. Chem.* **2000**, *5*, 287-299.

(152) Hanzlik, R. P. *Inorganic Aspects of Biological and Organic Chemistry*; Academic Press: New York, 1976.

(153) Frausto da Silva, J. J. R.; Williams, R. J. P. *The Biological Chemistry of the Elements. The Inorganic Chemistry of Life*; Clarendon Press: Oxford, 1994.

(154) Bray, A.; Lews, J.; Walter, R. R. *Essential Cell Biology. An introduction to the Molecular Biology of the Cell*; Garland Publishing: New York, 1998.

(155) Shier, D.; Buttler, J.; Lewis, R. *Hole's Human Anatomy & Physiology*; McGraw-Hill: Boston, 1996.

(156) DiDonato, M.; Zhang, J. Y.; Que, L.; Sarkar, B. *J. Biol. Chem.* **2002**, *277*, 13409-13414.

(157) Clarke, M. J.; Zhu, F. C.; Frasca, D. R. *Chem. Rev.* **1999**, *99*, 2511-2533.

(158) Ghosh, P.; D'Cruz, O. J.; Narla, R. K.; Uckun, F. M. *Clin. Cancer Res.* **2000**, *6*, 1536-1545.

(159) Sava, G.; Zorzet, S.; Giraldi, T.; Mestroni, G.; Zassinovich, G. *Eur. J. Cancer* **1984**, *20*, 841-847.

(160) 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.

(161) 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.

(162) Chen, H.; Parkinson, J. A.; Parsons, S.; Coxall, R. A.; Gould, R. O.; Sadler, P. J. *J. Am. Chem. Soc.* **2002**, *124*, 3064-3082.

(163) Bernersprice, 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.

(164) 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.

(165) Calamai, P.; Carotti, S.; Guerri, A.; Mazzei, T.; Messori, L.; Mini, E.; Orioli, P.; Speroni, G. P. *Anti-Cancer Drug Des.* **1998**, *13*, 67-80.

(166) Katsaros, N.; Anagnostopoulou, A. *Crit. Rev. Oncol./Hematol.* **2002**, *42*, 297-308.

(167) Sava, G.; Giraldi, T.; Mestroni, G.; Zassinovich, G. *Chem.-Biol. Interact.* **1983**, *45*, 1-6.

(168) Yang, P.; Wang, H. F.; Gao, F.; Yang, B. S. *J. Inorg. Biochem.* **1996**, *62*, 137-145.

(169) Jung, M.; Kerr, D. E.; Senter, P. D. *Arch. Pharm.* **1997**, *330*, 173-176.

(170) Burda, J. V.; Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1997**, *101*, 9670-9677.

(171) Baker, E. S.; Manard, M. J.; Gidden, J.; Bowers, M. T. *J. Phys. Chem. B* **2005**, *109*, 4808-4810.

(172) Poater, J.; Sodupe, M.; Bertran, J.; Sola, M. *Mol. Phys.* **2005**, *103*, 163-173.

2 Theory and Methodology

2.1 Preface

The aim of this chapter is to describe the basic elements of the theoretical tools that have been employed in this work. Most methodologies are based on quantum mechanical techniques; in few cases molecular mechanics has been employed. Initially these methods were feasible only for very small systems such as atoms and diatomic molecules, whereas, nowadays, thanks to the increases in computing power, systems of real interest can be modelled. The basic principles of the theories reported in this chapter are easily accessible, for instance, from several text books and references therein.[1-5]

2.2 Hartree-Fock theory

The starting point of Hartree-Fock theory is the non-relativistic time-independent Schrödinger equation:

$$H\Psi(r,R) = E\Psi(r,R) \tag{2.2.1}$$

The Schrödinger equation belongs to the category of equations called *partial differential eigenvalue equations,* in which an operator (H, the Hamiltonian operator) acts on a function (Ψ, the wave function) and returns the function multiplied by a scalar value (E, the energy). To solve the equation, the values of E and functions Ψ need to be determined. Let us consider a molecular coordinate system where the distance between the i_{th} and j_{th} electrons is $r_{ij}=|r_i - r_j|$; the distance between the A_{th} and B_{th} atoms is $R_{AB}=|R_A-R_B|$; finally, the distance between the i_{th} electron and the A_{th} atom is $R_{iA}=|r_i-R_A|$. In atomic units, the Hamiltonian for a generic system with N electrons and M nuclei is:

$$H = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{2.2.2}$$

where

i. Z_A is the atomic number of the nucleus A;

ii. M_A is the ratio of the mass of nucleus A to the mass of an electron;

iii. $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$.

The first two terms of the Hamiltonian represent the kinetic energy operators of electrons and nuclei, respectively. The third term expresses the coulomb attraction between electrons and nuclei. The last two terms represent the repulsion between pairs of electrons and between pairs of nuclei, respectively.

2.2.1 The Born-Oppenheimer approximation

The full Schrödinger equation can be solved only for a small number of simple systems, such as the particle in a box, the harmonic oscillator and the hydrogen atom. The problem in finding the exact solution arises when dealing with more than one electron, *e.g.* helium atom, which has two electrons and one nucleus. This is known as the *three-body problem*. So far, for such systems no exact solution has been found, therefore the solution to the Schrödinger equation for polyelectronic problems can only be an approximation.

As nuclei are much heavier than electrons, their motion is much slower: electrons can almost instantaneously adjust their position depending on the position of nuclei. The Born-Oppenheimer approximation takes advantage of this phenomenon and states that the wave function of a system can be decomposed as following:

$$\Psi_{TOT} = \Psi_{elec}\Psi_{nuc} \tag{2.2.3}$$

and

$$H_{TOT} = H_{elec} + H_{nuc} \tag{2.2.4}$$

In particular, the electronic Hamiltonian can now be written as:

$$H_{elec} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \tag{2.2.5}$$

and the Schrödinger equation involving the electronic Hamiltonian as:

$$H_{elec}\Psi_{elec} = E_{elec}\Psi_{elec} \tag{2.2.6}$$

If one has solved the electronic problem, it is possible to solve the nuclear Schrödinger equation. As the electrons move much faster than nuclei, it is reasonable to replace the instantaneous electronic coordinates by their average values. This generates a nuclear

Hamiltonian for the motion of nuclei in the average field of the electrons. Thus, the total energy can be calculated from:

$$(T_{nuc} + E_{elec}(R))\psi_N(R) = E_{TOT}\psi_N(R) \qquad (2.2.7)$$

2.2.2 Interpretation of the wave function

The wave function Ψ (from now on *elec* and *nuc* subscripts will be omitted for clarity as only electron wave functions are discussed) describes the motion and distribution of electrons in a generic system. However, its deep physical meaning is not that obvious. The universally accepted interpretation of the wave function was proposed by Born and it is associated with the quantity:

$$|\Psi_i|^2 = \Psi_i^*\Psi_i \qquad (2.2.8)$$

where Ψ_i^* is the complex conjugate of Ψ_i. $|\Psi_i|^2 d\tau$ represents the probability of finding an electron in a certain volume of space $d\tau$. This implies that integrating the probability over all space should give as result the number of the electrons in the system, N:

$$\int \Psi_i^* \Psi_i d\tau = N \qquad (2.2.9)$$

Moreover, if $\hat{A}\Psi = A\Psi$, where $\hat{A}$ is a generic operator, Ψ an eigenvector of $\hat{A}$, it can be written:

$$\int \Psi_i^* \hat{A} \Psi_i d\tau = A \qquad (2.2.10)$$

where A is the value obtained from the wave function through the operator $\hat{A}$. Thus, depending on the chosen operator, many properties can be determined from the wave function Ψ. For instance, in the case of the Hamiltonian operator, the property obtained would be the energy of the system.

2.2.3 Molecular orbital approximation

The simplest many-electron wave function is the Hartree product of spatial single electron components $\Phi_i(x,y,z)$. However, spin functions must be included to take account of electronic spin (α or β) as well as the indistinguishability of electrons. Therefore, the correct wave function must be antisymmetric: exchanging the coordinates of any two electrons will cause the wave function to change sign. The antisymmetry of the wave function is achieved by using Slater determinants in which the columns are the spin orbitals and the rows permute the electron coordinates. Subsequently, the wave function obeys the Pauli exclusion principle, since if two rows of the Slater determinant are equal the determinant would be zero. Also, interchanging the coordinates of two electrons corresponds to interchanging two rows of the Slater determinant, resulting in a change of sign of the determinant. For N electrons and N spin orbitals, it has the form:

$$\Psi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_N(1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \cdots & \phi_N(N) \end{vmatrix} \qquad (2.2.11)$$

2.2.4 Hartree-Fock method and self-consistent field (SCF)

Finding approximate solutions to the Schrödinger equation has been the quantum chemists' major aim for many years: the Hartree-Fock approximation plays the central role in solving such an intricate problem. The main idea is to replace the instantaneous electron-electron repulsion with an average electronic field generated by all electrons in the molecule. The new operator acting on the wave function is an effective one-electron Hamiltonian operator called the *Fock* operator, which takes the form:

$$f = h + v^{HF} \qquad (2.2.12)$$

where h is the core-Hamiltonian operator: a simple one-electron Hamiltonian for a system with only an electron moving in the field of nuclei; h has the following expression:

$$h = -\frac{1}{2}\nabla^2 - \sum_A \frac{Z_A}{r_{iA}} \tag{2.2.13}$$

The other term in the Fock operator is v^{HF}, representing the average potential *experienced* by the i_{th} electron:

$$v^{HF} = \sum_b J_b - K_b \tag{2.2.14}$$

The first term (J_b, *Coulomb* term) takes into account the classic Columbic repulsion between electrons; the second, K_b, is the *Exchange* term, which is associated with the spin correlation. For a generic closed-shell system, Roothaan and Hall developed separately the way to derive solutions for such systems. They introduced the Linear Combination of Atomic Orbitals (LCAO) approach, splitting the molecular orbital into a combination of single electron orbitals:

$$\psi_i = \sum_v^K c_{vi}\phi_v \tag{2.2.15}$$

This approach transforms the Hartree-Fock equations into a matrix problem: the entire set of equations assumes the simple form:

$$\mathbf{Fc} = \mathbf{Sc\varepsilon} \tag{2.2.16}$$

where **c** is a M x M matrix of coefficients c_{vi}, **ε** is M x M diagonal matrix whose elements are the orbital energies. **S** is the overlap matrix with elements S_{ij}, and **F** is the Fock matrix with elements F_{ij}.

$$S_{ij} = \int \chi_i^*(r1)\chi_j(r1)dr_1 \tag{2.2.17}$$

$$F_{ij} = \int \chi_i^*(r1)f_1\chi_j(r1)dr_1 \tag{2.2.18}$$

The overlap matrix is not necessarily the unit matrix as the functions are normalised but not always orthogonal. Thus, the Roothaan-Hall equations must be diagonalised such that the following equation is satisfied:

$$|\boldsymbol{F} - \varepsilon \boldsymbol{S}| = \mathbf{0} \tag{2.2.19}$$

Since the Fock matrix itself depends on the molecular orbital expansion coefficients, the equation is non-linear and must be solved iteratively: the procedure for solving the Hartree-Fock equation is called the self-consistent field (SCF) method. One way to solve this problem is achieved by generating a set of trial solutions as an "initial guess", employed to calculate the Exchange and Coulomb operators, leading to an initial Fock matrix $\mathbf{F}^{0}$. Hence, the Hartree-Fock equations are solved giving a new set of coefficients, which in turn is used to build a new Fock matrix $\mathbf{F}^{1}$. The SCF method gradually refines the individual electronic solutions while the total energy decreases. The *variation theorem* states that the energy calculated in this way will be always greater or equal to the real solution. Therefore this iterative process continues until a convergence criterion has been achieved: for instance the coefficients of matrix **c** are unchanged and the total energy reaches a stationary value.

2.3 Basis sets

An individual molecular orbital is expressed in the form:

$$\phi_i = \sum_{\alpha}^{N} c_{\alpha i} \chi_{\alpha} \tag{2.3.1}$$

where $c_{\alpha i}$ are molecular orbital expansion coefficients and χ_{α} are the basis functions.
Thus, molecular orbitals are expressed as linear combinations of one-electron functions, which are usually centred on nuclei. In order to form a complete basis set, and therefore remove all approximations from this method, an infinite number of basis functions are

required. However such an approach is impracticable for the purpose of calculation. Therefore a finite number of atomic orbitals centred on the nuclei of the molecule are typically used as basis functions. This method is called LCAO. Two types of basis function are usually employed: the Slater Type Orbitals (STO) and Gaussian Type orbitals (GTO).

2.3.1 STOs and GTOs

When the Slater Type orbitals are written in polar coordinates they take the form:

$$\chi(r,\theta,\varphi) = NY_{lm}(\theta,\varphi)r^{n-1}e^{-\zeta r} \tag{2.3.2}$$

where N is a normalisation constant; $Y_{l,m}$ are spherical harmonic functions of polar angles θ and φ; r is the distance between nucleus and electron; the exponent ζ depends on the angular momentum quantum number (l) and the effective nuclear charge.
STOs possess a number of features that make them well-suited as basis functions: in particular showing correct decay with increasing r ($r\rightarrow\infty$), and the correct form for 1s orbitals, as at the nucleus the STO has a cusp. However, computation of the integrals for the products of STOs is very expensive, so that STOs can feasibly be used only for atomic and diatomic systems where high accuracy is required.

The form of the GTO is:

$$\chi(x,y,z) = N\, x^{l_x} y^{l_y} z^{l_z} e^{-\zeta r^2} \tag{2.3.3}$$

and may be written in polar as well as Cartesian coordinates. The sum of l_x, l_y and l_z indicates the type of orbital (an effective total angular momentum). As shown in Figure 2.1, GTOs are generally inferior to STOs. The reason for that resides in the behaviour of the Gaussian approaching nuclei ($r\rightarrow 0$): the GTO has zero slope (no cusp) at this point. Also they tend to zero too rapidly with increasing r. For these reasons more GTOs are required to achieve equivalent accuracy to STOs. Nevertheless, the increase in number

of basis functions is counterbalanced by the ease with which the integrals can be calculated.

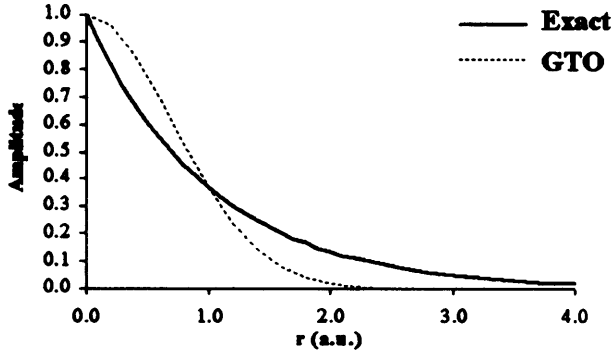


Figure 2.1 Exact and GTO behaviour as distance from the nucleus increases.

The smallest number of basis functions possible is known as a minimal basis set, which includes only enough basis functions to describe each atom using the filled orbitals in the ground state. In order to better describe molecules, the number of basis functions has to be increased, *e.g.*, using a fixed combination of Gaussian-Type orbitals called primitive functions. For instance, double-ζ (DZ) type basis set have functions of each type for each atom, improving the description of the electron distribution. However, since chemical bonding involves mainly valence orbitals, the core orbitals are virtually independent from the chemical environment. Therefore is possible to design basis sets that describe valence electrons with high accuracy, avoiding an extra calculation on the core electrons. Such basis sets are called split-valence basis sets. In particular, the Pople basis sets employs a segmented contraction, *i.e.* the primitive Gaussians are only used for one basis function and not for another of the same angular momentum. For example, the 2s and 3s basis functions in a generic atom do not share the same primitives. The most popular is perhaps the 6-31G basis set, in which the core orbitals are described by six contracted GTOs and the valence orbitals by three contracted GTOs and a more diffuse primitive GTO.

In order to provide for a better model in systems where higher angular momentum functions are significant, polarisation functions can be added. This is achieved by adding *p*-functions for hydrogen and *d*-functions for first-row, *f*-, *g*- and so on. Those extra functions provide for a better description of anisotropy of the electron distribution. Similarly, in order to better describe lone pairs of electrons, anions and excited states, diffuse s and p functions may be added to heavy atoms.

2.3.2 Effective core potential

Treating heavy elements through quantum chemistry represents a difficult challenge due to two critical problems. Firstly, heavy elements, such as transition metals, possess a large number of electrons, which in turn need a large number of basis functions to describe them. These electrons are mostly core electrons and thus one could think that a minimal basis set would be sufficient for them. Unfortunately, even a small basis set becomes intractable for systems of this size. Secondly, the core electron of these heavy elements reach very high velocities sufficiently close to the speed of light to manifest relativistic effects that standard methods cannot treat efficiently.

In 1935, Hellmann proposed a quite radical idea to solve this problem: replace the basis set with an analytical function able to reproduce accurately and more efficiently the movement of the core electrons in heavy elements. In this way Coulomb repulsion effects, Pauli exchange and relativistic effects would be included in such functions called Effective Core Potentials (ECPs), or also Pseudopotentials.

There are a few essential steps in order to properly design such functions. First, an electron wave function must be created for a specific atom from an HF or a relativistic Dirac-HF calculation. Then, the pseudo-orbitals are constructed such that in the core region the nodeless pesudo-orbitals replace the standard orbitals and match the Hartree-Fock orbitals in the valence region.

In other words, the pseudo-orbitals behave as the regular valence orbitals in the outer part, but they do not have a nodal structure in the core region (see Figure 2.2). In fact, the core electrons are replaced by a potential, which also includes relativistic effects. In the final step this potential is fitted as a set of Gaussian functions:

$$U_{ECP}(r) = \sum_i a_i r n^i e^{-\alpha_i r^2} \qquad (2.3.4)$$

where a_i, n_i and α are determined by least squares fit and depend on the angular momentum (s-, p-, d- etc.). The more Gaussians are used, the better the results, although the cost of calculation rapidly increase.

Throughout this work, Los Alamos double zeta basis set, LANL2DZ,[6-8] and Stuttgart/Dresden SDD[9] basis set have been extensively employed for treating platinum and other transition metals.

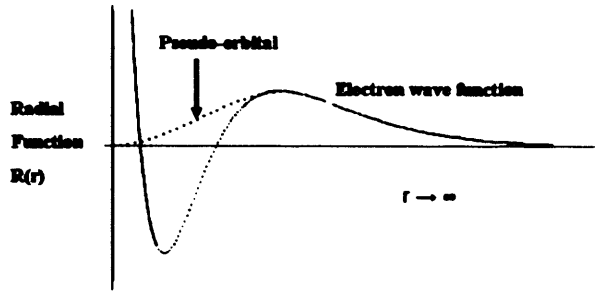


Figure 2.2 A schematic illustration of pseudopotential functions.

2.4 Post-HF methods and Density Functional Theory

Although Hartree-Fock theory produces only approximate solutions, the calculated geometries and energies for equilibrium structures are often in good agreement with experiment or with higher level calculations. One of the reasons is that HF, using an anti-symmetric wave function, incorporates *exchange correlation*. This means that the motion of two electrons with parallel spins is correlated. On the other hand, Hartree-Fock neglects a significant property of electrons in the system: the electron correlation. In fact, in Hartree-Fock electrons are treated as moving in an average potential generated by other electrons and the instantaneous position of an electron is not influenced by the exact position of its neighbour. This means that the motion of electron with anti-parallel spins remains uncorrelated. In real systems, though, the motion of electrons is *correlated* and they avoid each other more than HF theory predicts. The difference between the HF energy and the *real* energy is called the *correlation energy*. Neglecting the electron correlation represents the most significant drawback of the HF theory.

Many post-HF theories have been developed in order to treat the correlation energy explicitly. Although limited by their computational cost, such as Møller-Plesset perturbation theory and coupled cluster theory (CC) are among the most used. The aim of this section is not to provide a detailed description of such methods, but to illustrate the basic principles on which they are based.

The Møller-Plesset theory is principally based upon Rayleigh-Schrödinger theory, in which the *true* Hamiltonian operator H is expressed as the sum of a 'zeroth-order' Hamiltonian H_0 and a perturbation term, V:

$$H = H_0 + \lambda V \tag{2.4.1}$$

where λ is a parameter that can vary from 0 and 1.

In this way the Hamiltonian, and subsequently the wave function and the energy, can be expressed as powers of λ. The zeroth-order energy represents the one-electron Fock operators for N electrons. Also, it can be shown that the sum of the zeroth and first-order energies is equal to the Hartree-Fock energy. Thus, in order to estimate the contribution of electron correlation, further terms have to be computed. MP2 corresponds to the second-order perturbation terms, MP3 to the third, etc. Although an expensive method, MPn calculations generally provide high quality results in terms of geometries and molecular properties.

The idea in Coupled Cluster methods is to include all corrections of a given type to infinite order. The CC wave function is written as $\Psi_{CC} = e^T \phi_0$, where T is the operator that acts on a HF reference wave function generating all the i_{th} excited Slater determinants. Thus the Schrödinger equation becomes:

$$He^T \phi_0 = Ee^T \phi_0 \tag{2.4.2}$$

The wave function is then expanded and the coefficients of the single terms are called 'amplitudes'. The coupled cluster correlation energy is therefore determined by the

singles and doubles amplitudes (CCSDT, CCSDTQ etc). The CC approach is even more expensive than MP2, although it provides high accuracy and very reliable data.

In the past two decades, Density Functional Theory (DFT) has been extensively employed to take into account the correlation energy as an alternative 'cheap' theory. For instance, DFT is less computationally demanding than post-HF methods and, in some cases (large systems with more than 100 atoms) even less than HF itself.

2.4.1 Energy Functional

Unlike Hartree-Fock, DFT is based on the relationship between the total electron energy and the overall electron density. Fermi and Thomas first developed a model in which they suggested that the ground state energy of a system is connected to the overall electron density of the system itself. The breakthrough came with the work of Hohenberg and Kohn about forty years later. They showed that the ground-state energy of a system is *uniquely* defined by the overall electron density. In other words, the Energy (E) is a unique *functional* of the electron density, ρ(r). This means that for a given function ρ(r) there is a single corresponding energy. In DFT, the energy functional is written as following:

$$E[\rho(r)] = \int V_{ext}(r)\rho(r)dr + F[\rho(r)] \qquad (2.4.3)$$

The first term is an external potential and is associated with the Columbic interaction between electrons and nuclei. $F[\rho(r)]$ involves both kinetic and interaction energy of the electrons of the system. Kohn and Sham suggested that $F[\rho(r)]$ should be the sum of several terms, *i.e.*

$$F[\rho(r)] = E_{KE}[\rho(r)] + E_{H}[\rho(r)] + E_{XC}[\rho(r)] \qquad (2.4.4)$$

where

- $E_{KE}[\rho(r)] = \sum_{i=1}^{N} \int \psi_i(r) \left(\frac{-\nabla^2}{2} \right) \psi_i(r) dr$ expresses the kinetic energy of electrons

in the non-interacting reference system.

- $E_H[\rho(r)] = \frac{1}{2} \int\int \frac{\rho(r1)\rho(r2)}{|r_1 - r_2|} dr_1 dr_2$, known as Hartree electrostatic energy. It

represents the electrostatic energy generated by all the possible interactions between pairs of charge densities.

- $E_{XC}[\rho(r)]$ contains contributions from the exchange and correlation energy and

therefore is called the *exchange-correlation* energy functional. However, its analytical form is not known and must be approximated.

Combining these expressions and adding the electron-nuclear interaction gives the Kohn-Sham expression of the ground-state electronic energy:

$$E[\rho(r)] = \sum_{i=1}^{N} \int \psi_i(r) \left(\frac{-\nabla^2}{2} \right) \psi_i(r) dr + \frac{1}{2} \int\int \frac{\rho(r1)\rho(r2)}{|r_1 - r_2|} dr_1 dr_2 + \\ + E_{XC}[\rho(r)] - \sum_{A=1}^{M} \int \frac{Z_A}{|r - R_A|} \rho(r) dr \qquad (2.4.5)$$

The ground-state electron density can be determined by summing the square moduli of the occupied one-electron orthonormal orbitals as following displayed:

$$\rho(r) = \sum_{i=1}^{N} |\psi_i(r)|^2 \qquad (2.4.6)$$

Introducing 2.4.6 in 2.4.5 and applying the variatonal principle to the electronic energy $E[\rho(r)]$ the Kohn-Sham equation will assume the form:

$$\left\{ \left(\frac{-\nabla^2}{2} \right) - \left(\sum_{A=1}^{M} \frac{Z_A}{r_1 - R_A} \right) + \int \frac{\rho(r_2)}{r_1 - r_2} dr_2 + V_{XC}(r_1) \right\} \psi_i(r_1) = \varepsilon_i \psi_i(r_1) \qquad (2.4.7)$$

To solve the Kohn-Sham equations a self-consistent procedure similar to HF is needed. An initial guess of the electron density $\rho^{(0)}$ is taken and from it a set of orbitals can be derived. Then, an improved electron density $\rho^{(1)}$ is calculated from the set of orbitals, used in the second iteration to calculated a new set of orbitals and so on, until convergence criteria are achieved.

2.4.2 The exchange-correlation functional

One reason that makes DFT such an appealing approach is that even very simple approximations to the exchange-correlation functional can give satisfactory results. Although E_{XC} is a functional of the electron density and could in principle be entirely determined, no analytical form is known. Thus, E_{XC} has to be approximated as an integral involving only the electron density and possibly its gradient.

A- LDA approximation

The simplest approach is the *local density approximations* (LDA) in which the E_{XC} is a function only of the electron density at current point. Thus, we can write:

$$E_{XC}[\rho(r)] = \int \rho(r)\varepsilon_{XC}(\rho(r))dr \tag{2.4.8}$$

where ε_{XC} is the exchange-correlation per electron. By differentiating this expression we would obtain the exchange-correlation potential as

$$V_{XC}[r] = \rho(r)\frac{d\varepsilon_{XC}(\rho(r))}{d\rho(r)} + \varepsilon_{XC}(\rho(r)) \tag{2.4.9}$$

Several forms of the LDA E_{XC} exist, some of which contain both the exchange and the correlation expression, while others are separated. The following is commonly used only for exchange energy under the LDA:

$$E_X = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3} \int \rho(r)^{4/3} dr \qquad (2.4.10)$$

Similarly there are forms for the correlation energy, such as Perdew and Zunger's and Vosko, Wilk and Nusair's functional. Because of their complexity, no detailed form is given here.

B- Gradient corrected functionals

The LDA functionals are known to perform quite well, despite their relative simplicity. However, their results are totally inadequate for some specific problems. Therefore, a more complex approach has been suggested, the gradient-corrected approximations. Within the gradient-corrected approach, the *non-local* functionals depend upon the gradient of the electron density at each point in the space, as well as its value. These functionals are typically divided into the exchange and correlation energy contributions. A typical example of the exchange functional is the one proposed by Becke:

$$E_X^{B88} = E_X^{LDA} - \gamma \sum_s \int \frac{\rho_s(r)^{\frac{4}{3}} x_s^2}{1 + 6\gamma x_s \sinh^{-1} x_s} dr \qquad (2.4.11)$$

with $\quad x_s = \rho_s(r)^{-\frac{4}{3}} \left|\nabla \rho_s(r)\right|$

where γ is a parameter chosen to fit the known exchange energies of the inert gas atoms. The 2.4.11 equation is considered as a correction to LDA approach and is extensively used in computational chemistry.

C- Hybrid HF/DFT functionals

As well as the pure DFT functionals, Hybrid Hartree-Fock/DFT methods are available. In particular, Becke has recently formulated functionals that include a mixture of HF and DFT exchange along with DFT correlation energy's expressions. For instance, the B3LYP functional is extensively used for very diverse chemical systems:

$$E_{XC}^{B3LYP} = E_X^{LDA} + 0.20(E_X^{HF} - E_X^{LDA}) + 0.72E_X^{B88} + E_C^{VWN} + 0.81(E_C^{LYP} - E_C^{VWN}) \tag{2.4.12}$$

Becke determined the parameter values by fitting several molecular properties, such as atomisation energies, ionisation potentials etc. Thus, the HF and LDA exchange expressions are combined and adjusted by Becke's gradient-corrected exchange. In a similar manner, the local VWN energy correlation is used and corrected by the Lee, Yang, Parr's (LYP) which gives both local and non-local components. The original form was expressed as follows:

$$E_C = -a\int \frac{1}{1+d\rho^{-1/3}}\left(r + b\rho^{-2/3}\left(C_F\rho^{5/3} - 2t_W + \left(\frac{1}{9}t_W + \frac{1}{18}\nabla^2\rho\right)e^{-cr^{-1/3}}\right)\right)dr \tag{2.4.13}$$

with $t_W(r) = \sum_{i=1}^{N}\frac{|\nabla\rho_i(r)|^2}{\rho_i(r)} - \frac{1}{8}\nabla^2\rho$ and $C_F = \frac{3}{10}(3\pi^2)^{\frac{2}{3}}$

where a, b, c and d are constants with values 0.049, 0.132, 0.2533 and 0.349, respectively.

Another example of hybrid functionals is the BHandH (BH&H, half-and-half), which has the following expression:

$$E_{XC} = 0.5*E_X^{HF} + 0.5*E_X^{LDA} + E_C^{LYP} \qquad (2.4.14)$$

This BH&H functional has been widely used during this work and extensively applied to π-stacked systems.

2.5 PES: search of maxima and minima

The way in which the potential energy of a system varies as a function of nuclear coordinates is usually referred to as the *Potential Energy Surface* (PES). For a system with *N* atoms the energy is a function of *3N-6* internal or *3N* Cartesian coordinates. Figure 2.3 displays the PES of a generic system that can exist in two different conformations, for instance, the reactant and the product, both minima of the PES. In particular, the product is the global minimum, the lowest in energy, whereas the reactant is a local minimum. They are connected through a saddle point, the transition state. Both minima and maxima are stationary points on the energy surface and can be determined through the first and second derivatives of the potential energy function.

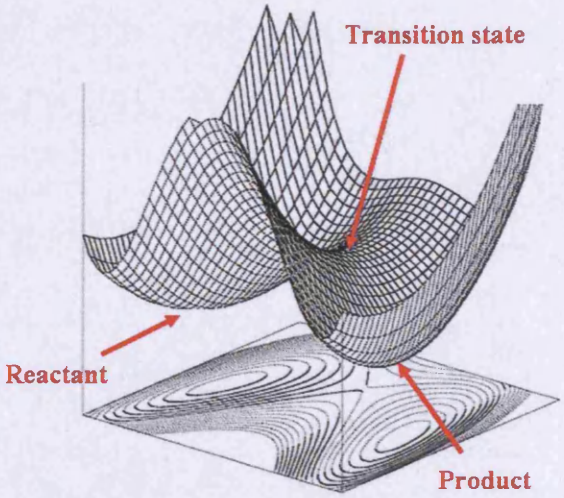


Figure 2.3 PES: reactant, transition state and product.

2.5.1 First and second derivatives of the energy

Derivatives of the energy with respect to the geometrical coordinates provide information that are essential in order to determine minima and maxima along the PES. The direction of the first derivative of the energy (the gradient), for instance, indicates where the minimum lies, and the magnitude of the gradient indicates the steepness of the local slope. Thus, as the forces acting on the atoms of the system is equal to the negative of the gradient, the energy of the system can be lowered by moving each atom in such a way that the forces are reduced. Second derivatives indicate the curvature of the potential energy function, which is extremely important in order to predict where the function will change direction, for instance, passing *via* a certain stationary point and also useful in order to distinguish maxima from minima.

The derivatives methods are usually divided in two main categories, depending on the order of the derivatives employed. The first-derivatives methods, are based on the analysis of the gradient of the energy, whereas the second-derivatives algorithms use both first and second-order derivative of the function. The *Newton-Raphson* method is the simplest method which includes the calculation of the inverse of the Hessian Matrix of second derivatives of energy with respect to the geometrical coordinates. However, this procedure is often computationally prohibitive. Besides, one may not be able to calculate analytical derivatives, which are generally preferable. Thus, the *Quasi-Newton* methods are an alternative to the Newton-Raphson approach. In particular the

Gaussian03 package, extensively used in this work, employs the Berny algorithm, which is a *Quasi-Newton Method.* Without getting into details (for further details the reader is directed to text books reported in the reference list),[1-5] Quasi-Newton methods are second-order derivative methods that provides better convergence properties and less computational effort as they gradually build up the inverse Hessian matrix in successive iterations.

2.5.2 Minima and Maxima search

The analysis of the Hessian matrix gives essential information about the properties of the stationary points. In particular, at minima, the first derivatives of the potential energy are all zero. Also, the Hessian matrix present no negative eigenvalues. On the contrary, at saddle points, while the first derivatives are all zero, the Hessian matrix present n negative eigenvalues, also called *imaginary* frequencies. The order of the saddle points depends on n: an n_{th}-order saddle point has n negative eigenvalues. In particular, the first order saddle points are *transition states*, where the energy passes through a maximum (the transition state) connecting two minima (reactants and products).

2.6 QM/MM calculations: ONIOM approach

Ab initio methods are known to be expensive in terms of cost of calculations. The time of the calculation depends on several parameters such as the method itself (see Table 2.1 for scaling of different methods), the number of basis functions and of course the size of the system.

Table 2.1 Scaling of theoretical methods.[10]

Method	scaling[a]
HF	N^3
DFT	N^3
MP2	N^5
CCSD	N^7

a: N is the number of basis functions.

In some cases to treat the entire molecule at *ab initio* level becomes prohibitive, thus many solutions have been found for this problem. One approach to the simulation of such large systems is the combination of quantum mechanics and molecular mechanics (QM/MM). In this way the system can be divided in two (or three) regions where different levels of theory can be applied. For instance, the 'reacting' part of the system is described quantum mechanically, whereas the remainder using a cheaper force field based method. The total energy of the system can be written as:

$$E_{TOT} = E_{QM} + E_{MM} + E_{QM/MM} \qquad (2.6.1)$$

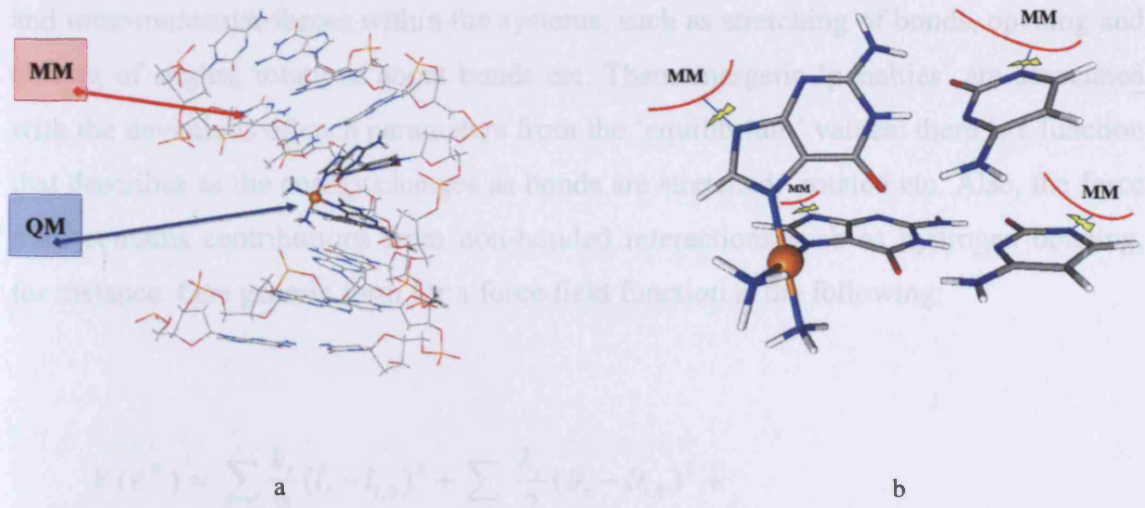
where:

i. E_{QM} is the energy of the region treated quantum mechanically;

ii. E_{MM} is the energy of the region treated with molecular mechanics;

iii. $E_{QM/MM}$ is the interaction energy between the quantum mechanics and molecular mechanics regions.

In some cases, $E_{QM/MM}$ is due entirely to non-bonded interactions, for instance, in solvated systems where the solute is treated at QM level and the solvent at MM. In all the systems studied in this project, though, the QM and MM regions are in the same molecule, as Figure 2.4 displays. Thus the system is *cut* along the interface between the two regions and $E_{QM/MM}$ contains terms that describe the interaction between the QM and MM regions. This is a non-trivial problem and renders QM/MM approach difficult to implement. Depending on the way that the interaction is modelled, several methods have been suggested.



a b

Figure 2.4 Cisplatin-DNA adduct: (a) QM and MM regions; (b) detail of QM region.

For instance, one should avoid half filled orbitals or unphysical open-shell orbitals for the QM region. Two general approaches have been proposed. In one, a sp^2 orbital containing one electron is established along the QM-MM interface. Another way is to employ link atoms: typically a hydrogen atom is added so that the original valence is preserved. In this case interactions between atoms are reduced in magnitude or even completely neglected.

Throughout this work, the ONIOM approach has been used.[11-15] This technique employs hydrogen atoms as link wherever the QM region has been cut and the interaction between the QM and MM regions is purely electrostatic, as the atoms treated at QM level feel the charge of the MM regions. In particular, ONIOM is a subtractive method where QM is calculated for a small region and MM for both small and full regions as following:

$$E_{ONIOM} = E_{small,QM} + E_{full,MM} - E_{small,MM} \qquad (2.6.2)$$

2.6.1 Molecular Mechanics: AMBER force field

Unlike *ab initio* methods, molecular mechanics ignore completely the motions of electrons and describe systems on the basis of classical physics. Many of the force fields on which molecular mechanics is based can be interpreted as a combination of the intra-

and inter-molecular forces within the systems, such as stretching of bonds, opening and closing of angles, rotations about bonds etc. Then, energetic 'penalties' are associated with the deviations of such parameters from the 'equilibrium' values: there is a function that describes as the energy changes as bonds are stretched, rotated etc. Also, the force field contains contributions from non-bonded interactions such as hydrogen bonding, for instance. One generic form for a force field function is the following:

$$V(r^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\vartheta_i - \vartheta_{i,0})^2 +$$

$$+ \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \tag{2.6.3}$$

The first term in equation 2.6.3 models the interaction between pairs of bonded atoms assuming the harmonic potential approximation that gives an increasing of the energy as the length l_i deviates from the equilibrium distance $l_{i,0}$. The second term is a sum over all the valence angles, again in the approximation of the harmonic potential. The third term describes the torsional potential that models the changing of the energy with the rotation around a bond. The fourth contribution arises from non-bonded interactions, modelled using a simple Coulomb potential for the electrostatics and a Lennard-Jones for van der Waals contributions (see 2.9.4).

Throughout this work, the Amber force field (*parm96.dat*) as implemented in G98/03 package has been widely used. Several publications suggest that this force field is able to model properly DNA structure as both hydrogen bonding and dispersion forces are well modelled.[16] Details about this force field can be obtained from Weiner and Cornell's publications.[17,18]

2.7 Atoms in Molecules (AIM) Theory

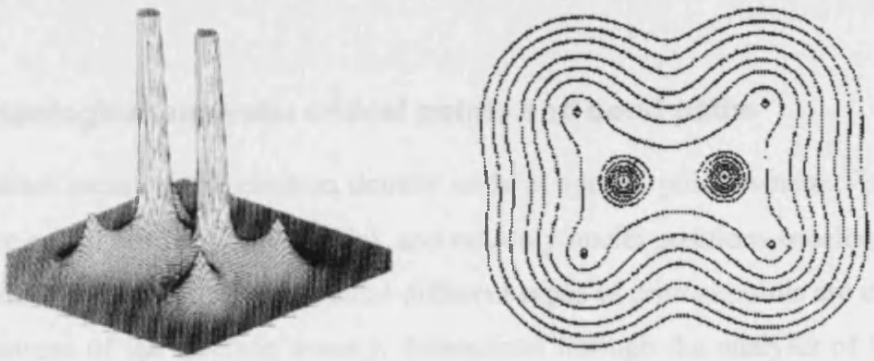
All the properties that can be determined about a molecular system are calculated through an appropriate operator from the wave function Ψ. The quantity $|\Psi|^2$ corresponds to the electron density of a system. Thus, among several approaches aimed to obtain chemical information from the wave function, Bader's Atoms in Molecules (AIM) theory is based on the analysis of the electron density.[19]

2.7.1 Electron density and the gradient vector

AIM theory is based on the analysis of the electron density, ρ, defined as follows:

$$\rho(r) = \int \psi^*(r)\psi(r)dr \qquad (2.7.1)$$

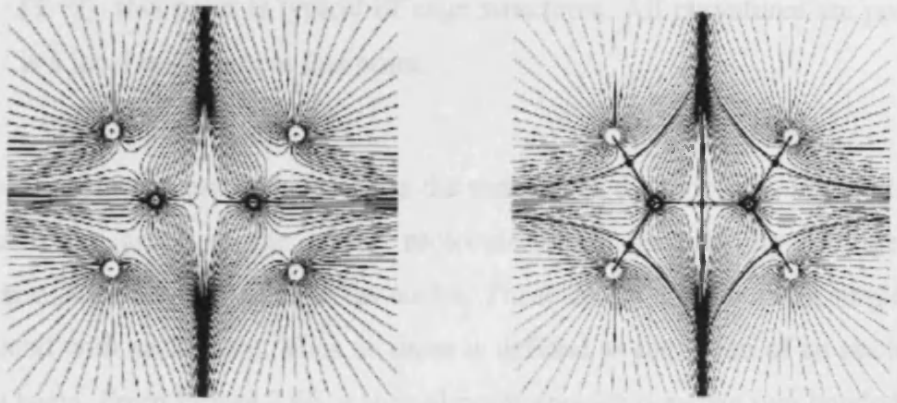
Figures 2.5 display the spatial distribution of the electron density for the ethene molecule in two (2.5b) and three (2.5a) dimensions: the electron density is a maximum at the position of each nucleus, while decays rapidly away from these positions. Furthermore, the cloud of electrons is denser at nuclear positions and become more diffuse as one moves away. As expected, local minima appear between the nuclei representing a covalent bond.



a b

Figure 2.5 The electron density of the ethene molecule, in 3D (a) and in the plane (b).

A crucial tool employed in AIM is the gradient of the electron density $\nabla\rho$, which is a vector pointing in the direction of greatest increase in ρ, Figure 2.6. The gradient vector is everywhere perpendicular to a surface of electron density. From Figure 2.6 it is obvious that most of the gradient paths are attracted by nuclei, called *nuclear attractor* in the frame of AIM theory. Some of these end at special points in between nuclei, called critical points. The analysis of the ensemble of these points and path reveals essential information on molecular properties of the system.



a b

Figure 2.6 A display of the trajectories that terminate at the nuclei for the ethene molecule.

2.7.2 Topological analysis: critical points and bond paths

The gradient paths of the electron density ends at special points where $\nabla\rho = 0$. Such points are called critical points (CP's), and not only nuclei positions (nuclear attractors) possess this characteristic. In particular different types of critical points are classified by the curvatures of the electron density, determined through the analysis of the Hessian matrix built from the second derivatives of ρ with respect to x, y and z. Thus, critical points are denoted as (n, m), where n (the rank of CP) is the number of non-zero curvatures, generally three, and m (the CP's signature) is the sum of their signs. Usually for stable systems and for all the cases studied in this work, n is equal to three, leading to four different types critical point:

i. (3,-3): this CP is found at nuclei positions, where all the curvatures are negative. ρ is a local maximum at this point;

ii. (3,-1): this CP is characteristic of a chemical bond. Here two curvatures are negative, and ρ is a maximum in the plane defined by the two corresponding axes; one curvature is positive, and ρ is a local minimum along the axis;

iii. (3,+1): this point is found within rings. Two curvatures are positive, here ρ is a minimum in the plane defined by the corresponding axes; one curvature is negative, and ρ is a maximum along this axis;

iv. (3,+3): this point is typical of cage structures. All curvatures are positive and ρ is a local minimum at this point.

Crucial information can be learned via the analysis of the topology. Figure 2.6b displays the molecular graph of the ethane molecule. There is a family of trajectories that originates at infinity and ends at the nuclei. These define a basin and a single attractor is associated with each basin. Thus an *atom* is defined as the union of an nuclear attractor and its basin. From Figure 2.6b, a special point appears between two bonded atoms: this is the *Bond Critical Point* (BCP), (3,-1). Electron density at this point gives important information about the corresponding bond. Also, there is a unique pair of trajectories originating from the BCP and ending at the neighbouring nuclei. They define a line along which the electron density is a maximum in space. At equilibrium geometry, this line is the *Bond Path*, which faithfully recovers the network of chemical bonds that are assigned on the basis of chemical considerations. Thus, the lines of maximum electron density linking bonded nuclei form the *molecular graph*, see Figure 2.7 for some examples.

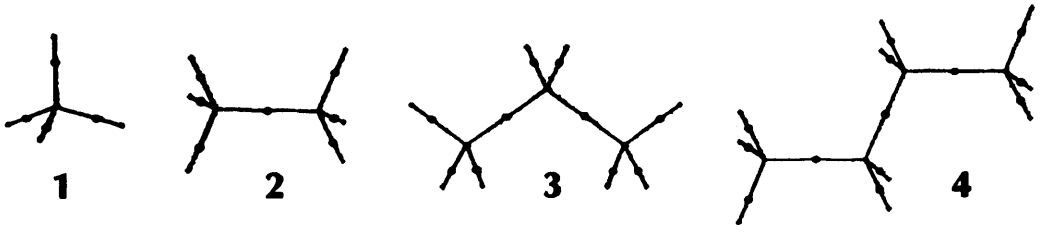


Figure 2.7 Molecular graphs of methane (1), ethane (2), propane (3) and butane (4).

A- Laplacian of the electron density

The laplacian of electron density ($\nabla^2\rho = \frac{\partial^2\rho}{\partial x\partial x} + \frac{\partial^2\rho}{\partial y\partial y} + \frac{\partial^2\rho}{\partial z\partial z}$) is a measure of local concentrations of the charge density. In particular, regions where $\nabla^2\rho > 0$ present a local depletion of electron density, whereas regions where $\nabla^2\rho < 0$ are locally concentrated charge. In practice, in order to analyse properties of molecules, the quantity $L = -\nabla^2\rho$ is typically employed.

2.7.3 AIM theory and chemical bonds

Perhaps the most important benefit of AIM theory is characterising molecular properties from the analysis of the electron density, its gradient and Laplacian, in particular at the BCP's. Thus, interactions between bonded atoms can be classified in two categories:

i. shared interactions, such as covalent bonds: large ρ and $L > 0$;

ii. closed-shell interactions, such as ionic bonds: low ρ and $L < 0$.

For instance, covalent bonds are characterised by high electron density at the BCP and $L > 0$. Bader calculated the electron density at the bond critical points for a series of emblematic diatomic molecules, such as H_2, B_2, N_2 and O_2, suggesting that $0.2728 < \rho < 0.5513$ au.[20,21] Similarly the values of L range between 0.1983 and 1.3784 au. Moreover, Howard and Larmache showed that the electron density at the BCP is a good measure of the strength of the covalent bond at least for small families of molecules: larger electron density values correspond to stronger bonds.[22]

Thus, it is possible *via* simple analysis to characterise H-bonds and van der Waals interactions. As they are *closed shell* interactions, $L < 0$, a depletion of charge occurs at the BCP. Also the electron density is much lower than covalent bond, *ca.* 10 times and 100 times smaller for H-bond and van der Waals interactions respectively.

A- AIM theory and H-bonding

Carroll and Bader studied H-bonded systems and their topological properties. By comparing plots of charge density between monomers and complexes, they noticed a mutual penetration of the van der Waals envelopes. This property allowed them to unambiguously decide whether atoms are linked *via* weak,[23] bifurcated,[24] or intramolecular[25] H-bonds. Koch and Popelier confirmed that for a generic H-bond, electron density must be low and $L < 0$.[26] They also noticed that the non-bonded radii of the atoms involved in the interaction is smaller than in the free monomers. This is a consequence of the mutual penetration as suggested previously by Carroll and Bader. Few more criteria can be taken into account such as the decrease of the volume of the hydrogen atom or, for instance, dipolar polarization.

B- AIM theory and van der Waals interactions

Similarly, Bader characterised a diverse collection of van der Waals complexes confirming that AIM was able to provide a description of these complexes "beyond geometry". In particular *via* AIM it is possible to identify the "van der Waals" bonds between heavy atoms, such as in the argon dimer, for instance, where the electron density is 0.00288 au and L = -0.0122 au. They suggested that ρ might be a good descriptor of the strength of such bonds as a correlation between the electron density and the degree of penetration of adjacent atoms occurs.[27]

2.7.4 AIM in this study: practical considerations

Throughout this work all the AIM properties were computed using AIMPAC and EXTREME programs.[21] Generally the analysis is based on the electron density calculated at Bond Critical Points, (3,-1). From now on, we will refer to Bond Critical Points simply as CP's and the electron density collected at these points as ρ_{CP}.

2.8 Statistical analysis of collected data

From the analysis of molecular properties of systems studied in this project, a vast number of data is available. In most of the cases the Standard Least Squares Method and the Partial Least Squares Method have been used to handle the computed properties. In this work all the data were analysed with the JMP 4.02 package (www.JMPdiscovery.com).

2.8.1 Least Squares Methods

A- Standard least squares method

In order to predict responses for an *exact* model of form $Y = \alpha + \beta X + \varepsilon$ (where ε is a random-error term), a predictive equation can be developed:

$$\hat{Y} = \hat{\alpha} + \hat{\beta} X \qquad (2.8.1)$$

where $\hat{\alpha}$ and $\hat{\beta}$ are estimates of the true intercept and gradient, respectively, whereas $\hat{Y}$ is the predicted estimate of Y. The quantity called residual, $r = Y - \hat{Y} = Y - \hat{\alpha} + \hat{\beta}X$ (2.8.2) is a measure of how well $\hat{Y}$ predicts the response variable Y. Ideally, the residual should be zero, but no prediction fits the data points exactly: therefore, the smaller the residual, the better.

In least squares analysis, the coefficients $\hat{\alpha}$ and $\hat{\beta}$ are chosen so that the sum of squared residuals is as close to zero as possible (equation 2.8.3).

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\alpha} - \hat{\beta} X_i\right)^2 \qquad (2.8.3)$$

In particular, the sum of the squared residuals, the quantity in 2.8.3, is minimised by adjusting the least squares coefficients, $\hat{\alpha}$ and $\hat{\beta}$.

B- Multiple linear regression (MLR)

The basic principles of multiple linear regression is like simpler linear regression, but the space of fit is three or more dimensions. Typically, a response variable (Y) is fitted using a linear combination of p independent variables (X).

Thus exact multiple regression model is:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 ... + \beta_p X_p + \varepsilon \qquad (2.8.4)$$

where β_p are the regression coefficients, X_p the independent variables of system i, Y the response variable and ε the random-error term.

2.8.2 Statistical tools

Typically various tools are used in order to test the quality of fit, the predictive ability of the model and significance of the descriptors.

A- Squared correlation

The squared correlation coefficient (R^2 or r^2) indicates how much the variance in the Y data is accounted. Equation 2.8.5 shows that it is based on the ratio of the Residual Sum of Squares (*RSS)*, which describes the deviation from the regression line to the Total Sum of Squares (*TSS)*:

$$R^2 = 1 - \frac{RSS}{TSS} \qquad (2.8.5)$$

where

$$RSS = \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2$$

and

$$TSS = \sum_{i=1}^{N} (Y_i - < Y >)^2$$

Thus, if R^2 approaches unity, a perfect correlation between observed and predicted Y values occurs, on the other hand, if close to zero, no relationship subsists.

B- Root Mean Square Error and Standard Deviation

Equation 2.8.6 defines the Root Means Square Error (rms error), which is another measure of the quality of fit:

$$rms = \sqrt{\frac{RSS}{n}} \tag{2.8.6}$$

In order to calculate the rms error, the mean of squares is divided by n, which is the number of data points. A similar quantity is the standard deviation (sd), where the means square is divided by n - *1*. For $n > 20$, the difference between sd and rms is small, and either may be used to estimate variation in the data.

2.8.3 Partial least squares method (PLS)

The standard multivariate linear regression cannot cope with highly correlated independent variables and fails when the number of independent variables exceeds the number of observables. Thus, an alternative statistical method employed in this work is the Partial Least Squares method (PLS), which becomes very useful when analysing data with many, noisy, collinear variables. Without getting into details, the basic idea behind the PLS approach is that the dependent variable is modelled as arising from a small set of so called 'latent variables', where all the independent measured variables are modelled as linear combinations of these latent variables.

A- Q^2 *and* R^2_{cv}

In order to further test the predictive power of a model, cross-validation is needed. This requires removing one (R^2_{CV}) or more (Q^2) observations from the dataset and fitting the model to the remaining data. The model is then employed in order to predict the *Y* values of the omitted observations.

For instance, the expression for Q^2 is the following:

$$Q^2 = 1 - \frac{PRESS}{TSS} \tag{2.8.7}$$

where

$$PRESS = \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2$$

It is important to remark that R^2 continues to increase as new descriptors are added, whereas Q^2 does not and, in fact, when a certain degree of complexity is reached in the model, the predictive ability may decrease.

B- F-*ratio*

Equation 2.8.8 shows the expression for the F-ratio, which indicates the probability that a real relationship exists in a multivariate model. In other words, a large value of F indicates that the probability that equation derived from the data fit is valid is greater than by chance:

$$F = \frac{R^2 \cdot (n - K - 1)}{K \cdot (1 - R^2)} \tag{2.8.8}$$

where

n is the number of data points and *K* is the number of variables.

2.9 Intermolecular forces

Atoms and molecules can interact via non-bonded forces that play, as well as covalent and ionic bonds, an important role in determining the final structure of molecular systems. In this work, a brief description of intermolecular forces is given.

2.9.1 Electrostatics

A- Permanent multipole moments

In a complex system such as a molecule, there are often electronegative atoms that attract electrons more than other less electronegative ones, leading to an unequal distribution of charge in the system. This can be represented as an arrangement of point charges within the molecule. Thus, using Coulomb's law, the electrostatic interaction can be written as:

$$V = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad (2.9.1)$$

where N_A and N_B are the numbers of point charges in the molecules A and B.

In order to model a molecule as single entity with a specific distribution of charges, the *central multipole expansion* can be used. This approach is based on the electric moments such as the charge (q), the dipole (μ), the quadrupole (θ), the octopole (Φ) and higher-order terms.

The simplest electric moment, is the dipole, defined as $\sum q_i r_i$, where q_i are the charges, located at the positions r_i. The dipole moment, which is a vector with the components along the three Cartesian axes, has contributions from electrons and nuclei, as following:

$$\mu_{nuc} = \sum_{A=1}^{M} Z_A R_A \tag{2.9.2}$$

and

$$\mu_{elec} = \int d\tau \Psi_0 (\sum_{i=1}^{N} -\hat{r}_i) \Psi_0 \tag{2.9.3}$$

where 2.9.2 is calculated involving discrete charges, and 2.9.3 from a continuous function of electron density, using the appropriate operator $\hat{r}_i$.

The quadrupole is a non-spherically symmetrical distribution of charge: for instance, four charges that sum to zero, so that they do not lead to a dipole, but spatially arranged in a specific way, resulting in non-zero quadrupole, see Figure 2.8.

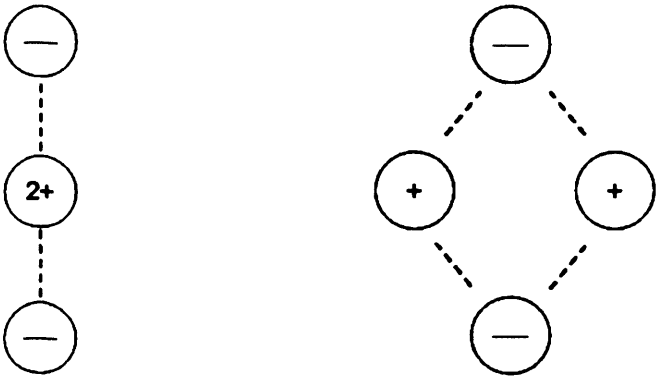


Figure 2.8 Two spatial arrangements of charges that lead to a quadrupole moment.

Whereas the dipole moment has three components along the Cartesian axes, the quadrupole has nine, arising from the pairwise combinations of x, y and z, such as:

$$\Theta = \begin{pmatrix} \sum q_i x_i^2 & \sum q_i x_i y_i & \sum q_i x_i z_i \\ \sum q_i y_i x_i & \sum q_i y_i^2 & \sum q_i y_i z_i \\ \sum q_i y_i z_i & \sum q_i z_i y_i & \sum q_i z_i^2 \end{pmatrix} \tag{2.9.4}$$

Higher electric moments such as hexapole, octopole and so on, are similarly defined.

In order to model the interaction between two molecules, all the combinations such as charge-charge, charge-dipole, charge-quadrupole, charge-octopole, etc., have to be taken into account. All these terms depend on different inverse powers of the separation r: for instance, a dipole-dipole interaction decays with r^{-3}, while charge-charge interaction with r^{-1}.

It has to be emphasised that this approach is valid only if the separation r between the two molecules involved is much larger than the internal dimensions of the molecules. Yet, all the information about the multipole moments can be gained from the wave function and therefore can be computed using quantum mechanics.

B- Polarisation (Induction)

The electrostatic interactions arise not only from *permanent* charge distributions, but also from changes induced by, for instance, an external field: this process is called *polarisation*. The main effect of the external field on a charge distribution is to induce a dipole moment in the molecule. The magnitude of the induced dipole, μ_{ind}, is proportional to the electric field E, with the constant of proportionality being the polarisability α, as $\mu_{ind} = \alpha E$. In turn, the induced dipole is able to generate an electric field that decays as r^{-3}. For instance, if a polar molecule A induces a dipole on a molecule B, then molecule B will successively be able to affect the distribution of B itself, or another molecule C.

Polarisation is a cooperative effect and is usually modelled employing a set of coupled equations solved iteratively. Initially, the induced dipoles are set to zero. Then, a first guess of the induced dipole is calculated from the permanent charges. The electric field generated by the induced dipoles is added into the permanent electric field and used to refine the estimation of the induced dipoles. The calculation continues until convergence is reached.

2.9.2 Exchange repulsion forces

If two atoms approach, at short distances the potential energy increases rapidly even for very small decrease of the separation. This behaviour has quantum mechanical origin, connected to the Pauli principle which allows no two electrons to possess the same set

of quantum numbers. In other words, when two atoms approach, the electrons with the same spin tend to avoid each other, the final effect being a stronger repulsive interaction between the positive charges of the nuclei. At very short distances the repulsion varies with *1/r*, but at larger *r* the energy decays exponentially as exp(-2r). These interactions are, therefore, called *exchange* repulsion forces (see Figure 2.11).

2.9.3 Dispersion forces

Dispersive forces are attractive long-range interactions between instantaneous dipoles, which arise from fluctuations in the electrons clouds.

A- London dispersion formula

London was the first one to illustrate how these forces could be explained in the frame of quantum mechanics theory. The dispersion arises from the mutual polarisability of the electronic clouds of the systems involved in the interaction. Thus, the energy due to these interactions can be simply be written as:

$$E_{disp} = C\frac{\alpha_1\alpha_2}{r^6} \qquad (2.9.5)$$

where C is a constant of proportionality, α_1 and α_2 are the polarisabilities of atoms or molecules and *r* is the distance between the two interacting systems.

This simple model gives reasonable results, although improved models are available.

B- Drude's model

Drude proposed a model to rationalise the source of such forces. Figure 2.10 displays two molecules with two charges, +q and –q separated by a distance *r*.

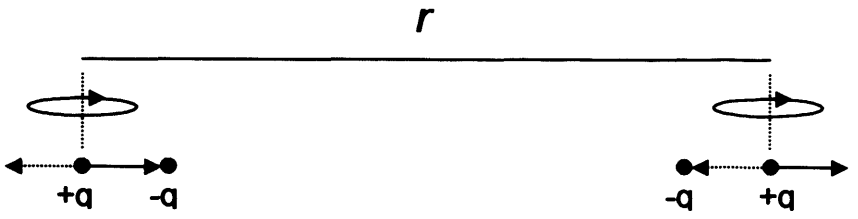


Figure 2.10 Drude's model (adapted from ref. 3).

The negative charge fluctuates around the equilibrium in accord to an harmonic potential, with angular frequency ω along the z axis. Let us consider a second molecule identical to the first one, with the positive charge on the z axis and the negative charge oscillating following an harmonic potential. At infinite distance, the two molecules are not interacting and the total ground-state energy is just twice the zero-point energy of a single molecule, $\hbar\omega/2\pi$. On the contrary, when the two molecules approach, moving along the z axis, an interaction occurs: the dispersion forces, which varies with r^{-6} in the Drude two-dimensional model (see Figure 2.11).

This simple model takes into account only dipole-dipole interactions. However, higher-order terms such as dipole-quadrupole, quadrupole-quadrupole, etc., give important contributions to the final interaction. Thus, Drude's model can be extended as a series expansion:

$$V(r) = \frac{C_6}{r^6} + \frac{C_8}{r^8} + \frac{C_{10}}{r^{10}} + \ldots + \frac{C_n}{r^n} \qquad (2.9.6)$$

All the C_n coefficients are negative and they indicate an attractive interaction.

2.9.4 Van der Waals interaction

Van der Waals interactions are those forces intervening between two neutral systems not directly bonded. At large distance the van der Waals energy (E_{vdw}) is equal to zero; by reducing the distance, a minimum in the potential energy surface is reached; then the

interaction becomes repulsive. A popular potential, which describes this behaviour is the Lennard-Jones (LJ, see Figure 2.11), which has the following form:

$$E_{LJ}(r) = \frac{C_1}{r^{12}} - \frac{C_2}{r^6} \tag{2.9.7}$$

The first term is mainly due to the exchange repulsion forces that prevail at very short distances between the two interacting systems, while the second term depends on the dispersion forces, which are always attractive and prevail at long distances.

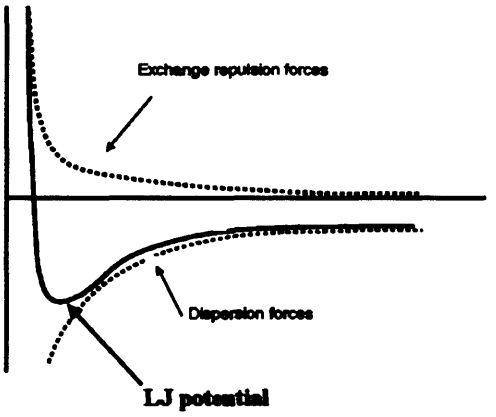


Figure 2.11 The Lennard-Jones potential constructed from exchange repulsion and dispersion forces.

2.10 References

(1) Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond*; Oxford University Press: Oxford, 1999.

(2) Jensen, F. *Introduction to Computational Chemistry*; Wiley & Sons: Chichester, 1999.

(3) Leach, A. R. *Molecular Modelling. Principles and applications*; Prentice Hall: Harlow, 2001.

(4) Atkins, P. W.; Friedman, R. S. *Molecular Quantum Mechanics*; Oxford, University Press: Oxford, 1997.

(5) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to advanced electronic structure theory*; Macmillan publishing Co. Inc.: New York, 1982.

(6) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270-283.

(7) Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284-298.

(8) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299-310.

(9) Andrae, D.; Haussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123-141.

(10) Friesner, R. A. *PNAS* **2005**, *102*, 6648-6653.

(11) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170-1179.

(12) Matsubara, T.; Sieber, S.; Morokuma, K. *Int. J. Quantum Chem.* **1996**, *60*, 1101-1109.

(13) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357-19363.

(14) Svensson, M.; Humbel, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 3654-3661.

(15) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959-1967.

(16) Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350-364.

(17) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc* **1984**, *106*, 765-784.

(18) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc* **1995**, *117*, 5179-5197.

(19) Bader, R. F. W. *Atoms in Molecules-A Quantum Theory*; Oxford, University Press: Oxford, 1990.

(20) Bader, R. F. W.; Essen, H. *J. Chem. Phys.* **1984**, *80*, 1943-1960.

(21) Bieglerkonig, F. W.; Bader, R. F. W.; Tang, T. H. *J. Comput. Chem.* **1982**, *3*, 317-328.

(22) Howard, S. T.; Lamarche, O. *J. Phys. Org. Chem.* **2003**, *16*, 133-141.

(23) Calhorda, M. J. *Chem. Commun.* **2000**, 801-809.

(24) Louit, G.; Hocquet, A.; Ghomi, M.; Meyer, M.; Suhnel, J. *Physchemcomm* **2002**, 94-98.

(25) Hocquet, A. *Phys. Chem. Chem. Phys.* **2001**, *3*, 3192-3199.

(26) Koch, U.; Popelier, P. L. A. *J. Phys. Chem.* **1995**, *99*, 9747-9754.

(27) Bone, R. G. A.; Bader, R. F. W. *J. Phys. Chem.* **1996**, *100*, 10892-10911.

3 The chemistry of cisplatin and other transition metal ligands.

From hydrolysis to DNA bases complexes

3.1 Preface

The aim of this study is to describe the cisplatin's chemistry, from aspects of its hydrolysis mechanism to interaction with DNA bases, focusing on the effect of platination over the GC pair. Also, a systematic study of the binding with guanine and the effect on the guanine...cytosine pair of all the transition metals has been discussed.

3.2 Cisplatin's hydrolysis: solvation and H-bonding

As understood from chapter 1, solvation is critical in activating cisplatin to its active forms, $[Pt(NH_3)_2(OH_2)Cl]^+$ and $[Pt(NH_3)_2(OH_2)_2]^{2+}$, wherein chloride ions are displaced by water molecules.[1,2] In this mechanism, hydrogen bonds are believed to play a fundamental role: for instance, direct hydrogen bonds between water and the metal centre in platinum complexes have been observed in accurate ab initio calculations.[3] Hydrogen bonding has also been implicated in the recognition of cisplatin's active site. X-ray crystallographic observations reveal strong N—H...O contacts between ammine and carbonyl groups,

which have been proposed as a mechanism for cisplatin's preference for binding at the N7 of guanine in DNA and its selectivity for intrastrand guanine-phosphate-guanine (GpG) linkages.[4-8] As a small molecule capable of forming H-bonds, cisplatin is very hydrophilic, with an octanol-water partition coefficient (log P_{oct}) of less than -2.0.[9,10] This leads to poor intestinal absorption of cisplatin, and hence, to the necessity of intravenous administration: better oral and intestinal uptake has therefore become a key factor in the search for new platinum drugs.[11]

3.2.1 Calculation method

All the calculations were performed at DFT level, using Gaussian03.[12] Following the work of Wysokinski and Michalska,[13] we have made extensive use of Adamo and Barone's modified PW91 functional (denoted mPW1),[14] which in combination with Stuttgart-Dresden (SDD) pseudopotential/basis set,[15] has been shown to give excellent results for platinum complexes. For consistency with previous work,[16] electrostatic potentials and complexes with HF and NCH were also computed using the B3LYP functional[17,18] with a mixed basis set consisting of 6-31+G(d,p)[19-22] on all light atoms and SDD on Pt. Although cisplatin itself has C_{2v} symmetry, no symmetry constraints were applied to any H-bonded complexes. All minima and transition states were confirmed as such via harmonic frequency calculation.

Also, we extensively used the molecular electrostatic potential (MEP), which is crucial to predict sites and strengths of hydrogen bonding.[23] It has been recently shown that the MEP, in combination with the topological energy densities, can accurately predict Abraham's acidity (**A**) and basicity (**B**) scales.[24] We therefore made use of an in-house C-program to extract minima and maxima of the MEP on the van der Waals surface (defined as the 0.001 au isosurface). These quantities, denoted $V_{S,Min}$ and $V_{S,Max}$, were combined with bond CP energy densities according to the regressions set out by Wysokinski[13] to predict overall hydrogen bond acidity and basicity scales. Also the Atoms in Molecules theory has been extensively employed in order to characterise bonding and molecular properties.

3.2.2 Results and discussion

A- Electronic structure of cisplatin: AIM analysis

Initial optimization of cisplatin at the mPW1/SDD level agreed with the findings of myriad theoretical and experimental studies,[25,26] giving a C_{2v} minimum with a distorted square-planar coordination about Pt (see also section 1.5). Therefore, we will not discuss this structure in any depth, except to note that the in-plane N—H groups are aligned toward Cl, resulting in N—H...Cl distances and angles of 2.40 Å and 113.2°. These values compare well with the data reported in section 1.5, and place the intra-molecular contacts within Steiner's geometrical parameters for weak hydrogen bonding, namely 2.0-3.0 Å and 90-180°, respectively.[27] The strain induced by constraining all angles around Pt to be exactly 90° is 2.70 kcal mol^{-1}, that is around the value expected for two weak H-bonds. Natural bond orbital (NBO) charges are also suggestive of H-bonding, with values of +0.447 au for the in-plane H's, falling to +0.428 au for the out-of-plane Hs. AIM provides an unambiguous definition of atoms and bonds directly from the electron density, ρ. Topological analysis reveals (3, -1) CP's in the expected position for all covalent/dative bonds, but no bonding CP's in the intramolecular N—H...Cl region. Thus, the key criterion for assigning this interaction as a hydrogen bond is missing here, and one can summarize that this is simply a weak electrostatic attraction, rather than a direct H-bond. Properties of the bond CP's that are located in cisplatin are reported in Table 3.1.

Table 3.1 Bond Critical Point properties in Cisplatin (au).[a]

	ρ_{CP}	$\nabla^2\rho$	ε	E
Pt—Cl	0.080	0.199	0.091	-0.018
Pt—N	0.106	0.388	0.100	-0.032
N—H_{in}	0.302	-1.347	0.017	-0.381
N—H_{out}	0.313	-1.384	0.018	-0.394

a: ρ_{CP} is the electron density at the bond CP, $\nabla^2\rho$ the Laplacian of the density here, ε the bond ellipticity, and E the energy density at the bond CP.

B – The electrostatic potential

The electrostatic potential that results from this optimized structure of cisplatin is shown in Figure 3.1. This reveals a picture of a highly polar molecule containing large areas of both positive and negative potential. In the positive (red) region, four distinct maxima are present, one near each out-of-plane ammine NH, with a small area of less positive potential between them. The negative (blue) region, by contrast, contains a single minimum between the chlorides. Interestingly, the surface above and below Pt is close to neutral, indicating that its formal charge of +2 is not reflected on the molecular surface.

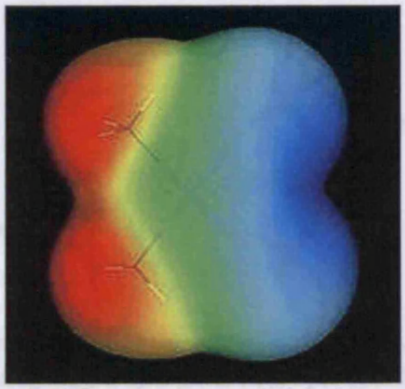


Figure 3.1 Electrostatic potential (red is positive, green neutral, and blue negative).

These observations can be quantified by searching for the maximum and minimum values of the MEP on this surface, which results in values of -0.092 au for $V_{S,Min}$ and +0.099 au for $V_{S,Max}$. These values suggest that cisplatin is a strong H-bond acid and base (compare with values of -0.072 and +0.058 for acetamide, one of the strongest H-bond acids and bases considered in Platts' studies).[24] It is a general observation that hydrogen bonding occurs at these sites of maximal and minimal surface MEP. This was checked for this case by optimizing the geometry of complexes of cisplatin with HF and NCH at the same B3LYP/SDD-6-31+G(d,p) basis set. Figure 3.2 shows the results of these optimizations, from which we can confirm that this general rule is indeed followed here. NCH forms a symmetrical complex between two out-of-plane NHs to form the C_s complex shown below,

in a fashion reminiscent of urea and related compounds. Clearly, two such arrangements are possible above and below the coordination plane, which will act to further enhance acidity. We were unable to locate a stable energy minimum for NCH complexed to the in-plane NH, confirming the impression gained from Figure 3.1 that this group is unlikely to form external H-bonds. The geometry of the HF complex also follows the MEP, and forms the C_s complex shown below. Here, the most stable position of the HF is in the coordination plane between the Cls, but associated with a single Cl rather than symmetrically between them (Cl…H distance are 2.241 and 3.301 Å).

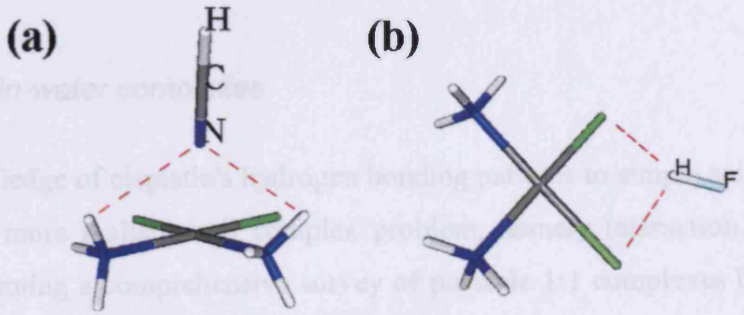


Figure 3.2 B3LYP/SDD-6-31+G(d,p) optimized geometries of complexes of cisplatin with (a) hydrogen cyanide and (b) hydrogen fluoride.

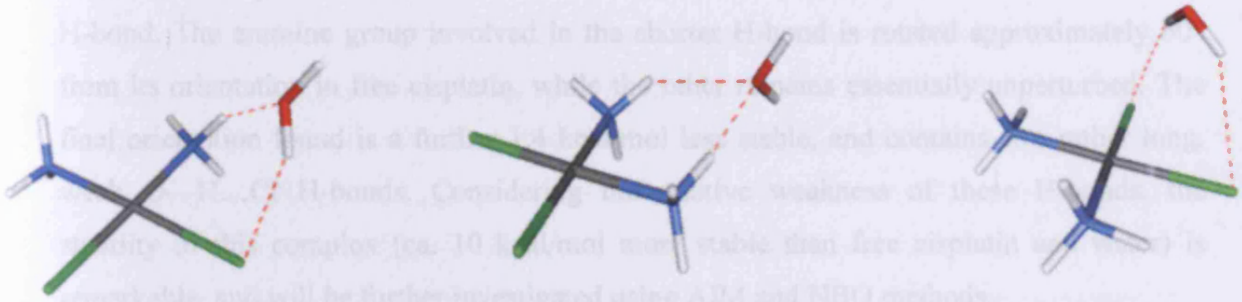
C- Abraham's acidity and basicity of cisplatin

As well as providing insight into the geometry of H-bonding, these complexes allow us to predict cisplatin's place on Abraham's **A** and **B** acidity and basicity scales, *via* the properties of the H-bond CP's therein. In the NCH complex, two distinct N—H…N CP's are present, that is, a bifurcated H-bond is formed. Each of these CP's has a rather low kinetic energy density of 0.0047 au, less than half the value found in the complex of acetamide with NCH. Combining these values with $V_{S,Max}$, and accounting for the second symmetrical complex below the plane, leads us to a prediction of **A** = 0.70, that is, stronger than most monofunctional organic acids, and around five times more acidic than

uncomplexed ammonia. H-bond basicity follows a similar pattern: two F—H...Cl CP's are present, both with low-energy densities. Combined with the deep $V_{S,Min}$ noted above, this predicts **B** = 0.84, again rather stronger than most monofunctional bases and an order of magnitude greater than organic chlorides. These calculations also reveal that cisplatin's large acidity and basicity are mostly due to electrostatic effects, that is, it is a hard acid/base, with almost negligible covalent overlap. Cisplatin's octanol/water partition coefficient, log P_{oct}, has been measured to be -2.53.[9,10] Combining our calculated **A** and **B** values with previously determined size and polarity/polarizability descriptors using Abraham's Linear Solvation Energy Relationship (LSER) for log P_{oct} yields a predicted value of -2.58, giving some independent verification of our results.

D- 1:1 *cisplatin-water complexes*

With this knowledge of cisplatin's hydrogen bonding patterns to simple acids and bases, we now turn to a more realistic and complex problem, namely interaction with water. We begin by performing a comprehensive survey of possible 1:1 complexes between cisplatin and water, the results of which are summarized in Table 3.2 and Figure 3.3. Only three distinct complexes were obtained from this search, as all other starting geometries resulted in one of these structures when fully optimized. The most stable of these (Figure 3.3-1) combines two H-bonds, forming a bridge from N—H to Cl, both of which are relatively short and strong. An interesting feature of this complex is the rotation of the ammine group involved in the N—H...O H-bond, such that the in-plane H now points away from Cl, indicating that the Pt—N bond is able to rotate to adopt the optimum geometry for H-bonding. The overall stabilization energy of this complex is 16.61 kcal/mol, or 14.63 kcal/mol after correction for BSSE.[28]



1 2 3

Figure 3.3 Optimized geometries of 1:1 cisplatin…H_2O complexes.

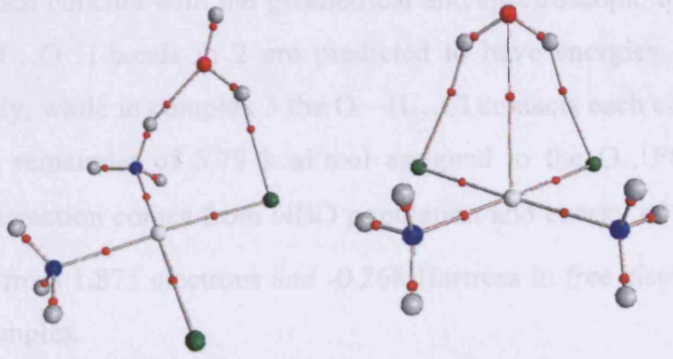
Table 3.2 Geometrical and energetic properties of 1:1 Cisplatin…H_20 complexes.

	Relative energy (kcal/mol)	H-bonds	B…H (Å)	A…B (Å)	B …H—A (°)	$\Delta\lambda^a$ (cm^{-1})
1	0.0	N—H…O	1.709	2.696	155.6	263.8
		O—H…Cl	2.230	3.087	144.6	211.2
2	+5.06	N—H…O	1.895	2.912	169.6	38.7
		N—H…O	2.373	3.264	142.9	0.5
3	+6.45	O—H…Cl	2.644	3.476	143.7	50.5
		O—H…Cl	2.645	3.478	143.8	50.5

a: this is the change in harmonic stretching frequency of the A—H donor.

The next most stable 1:1 complex is more than 5 kcal/mol higher in energy, and echoes the geometry found with NCH, with two N—H...O H-bonds to separate ammine groups (Figure 3.3-2). This complex is no longer symmetrical, and contains one short and one long H-bond. The ammine group involved in the shorter H-bond is rotated approximately 60° from its orientation in free cisplatin, while the other remains essentially unperturbed. The final orientation found is a further 1.4 kcal/mol less stable, and contains two rather long, weak O—H...Cl H-bonds. Considering the relative weakness of these H-bonds, the stability of this complex (ca. 10 kcal/mol more stable than free cisplatin and water) is remarkable, and will be further investigated using AIM and NBO methods.

In the most stable complex, (3, -1) CP's are found for both H-bonds in the expected positions. It is well established that H-bond strength approximately correlates with the value of the electron density at the H-bond CP, ρ_{CP} (see section 2.7) In this case, it is clear that the N—H...O H-bond is rather stronger (ρ_{CP} = 0.044 au) than the O—H...Cl (0.024 au), confirming the pattern seen in $\Delta\lambda$ in Table 3.2. A similar situation is seen in the second complex, where the short N—H...O H-bond has ρ_{CP} = 0.030 au compared to 0.010 au in the longer contact. The final complex presents a more intricate picture: two H-bond CP's are again found, with ρ_{CP} = 0.011, corresponding to weak interactions. However, a third intramolecular bond CP is also present, this time linking O with Pt directly (see Figure 3.4). Thus, it appears that the stability of this complex is not due solely to H-bonding, but also to overlap between water lone pairs and the empty d_{z^2} orbital on Pt, such that it could be thought of as the first step toward chemical oxidation to a Pt^{IV} complex.



1 3

Figure 3.4 Molecular graph of cisplatin...H_2O complexes 1 and 3.

Because all 1:1 complexes contain several H-bond interactions, it is difficult to separate out the contribution of each to the overall stabilization energy. We therefore turn to the AIM theory, and specifically bond CP properties, to decompose the overall interaction into individual contributions. Following Grabowski,[29] we use changes in electron density on formation of hydrogen bonds to yield a more accurate prediction of H-bond energies (E_i) than using the electron density at the H-bond CP. The analysis of all the cisplatin complexes studied (n = 18), namely, with 1, 2, 3...up to 10 water molecules, suggested that the variation of electron density at the donor H nucleus ($\Delta\rho$ /ρ_0) of these complexes provides a more accurate correlation (r^2 = 0.98) to the purely H-bond interaction energy (E_{TOT}) than the variation at the donor X—H bond CP (r^2 = 0.93) as reported from Grabowski. Thus the overall interaction energy can be decomposed in single H-bond energies as follows:

$$E_i = (\Delta\rho / \rho_0) \times E_{TOT} \qquad (3.1)$$

It has to be stressed that this formula is only valid for cisplatin-water complexes because of its high family-dependence. Eq. 3.1 estimates the H-bond energies in complex 1 to be 9.13 and 7.47 kcal/mol for N—H...O and O—H...Cl, respectively, that is, both are reasonably strong H-bonds, which concurs with the geometrical and spectroscopic quantities in Table 3.2. The two N—H...O H-bonds in 2 are predicted to have energies of 8.80 and 2.78 kcal/mol, respectively, while in complex 3 the O—H...Cl contacts each contribute just 2.18 kcal/mol, leaving a remainder of 5.79 kcal/mol assigned to the O...Pt contact. Further evidence for this interaction comes from NBO population and energy of the d_{z^2} orbital on Pt, which increases from 1.875 electrons and -0.268 Hartrees in free cisplatin to 1.928 and -0.284 in this 1:1 complex.

E- First solvation sphere of cisplatin

Although the patterns and motifs of H-bonding in these 1:1 complexes reveal some interesting properties of cisplatin, it is obvious that interactions with water as a solvent will be much more intricate. We therefore set out to study how several water molecules might interact simultaneously with cisplatin, progressively adding more solvent molecules until saturation is reached. Figure 3.5 shows our estimate of a full first solvation sphere of cisplatin, in which the metal complex is surrounded by 10 water molecules, every H-bond donor and acceptor site is saturated with H-bonds to water, and addition of further water molecules gives rise only to water…water contacts. The placement of these water molecules was guided by the surface electrostatic potential of cisplatin (Figure 3.1), the contacts observed for 1:1 complexes (Figure 3.3), and by a good deal of trial and error, in which optimised structures containing only water…water contacts were rejected.

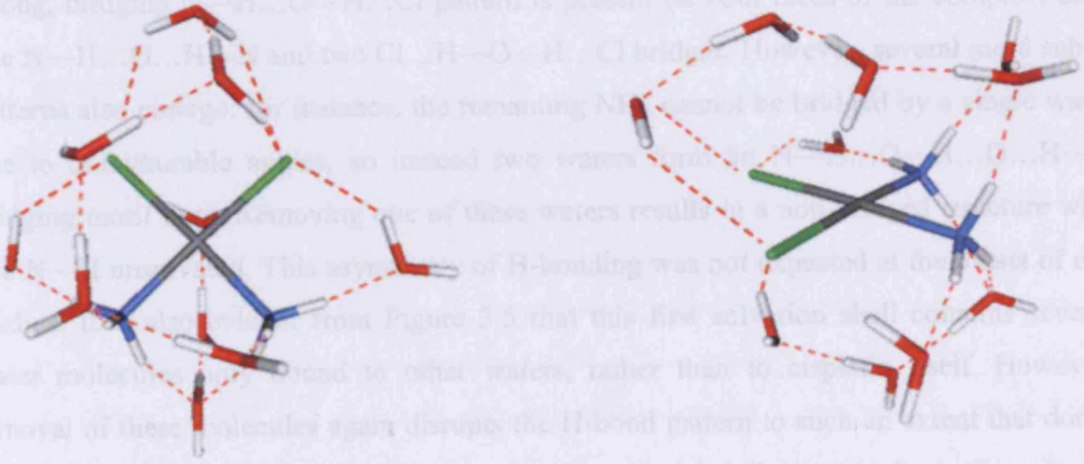


Figure 3.5 Two views of the optimized cisplatin…$10H_2O$ complex.

This complex has an overall stabilization energy of -192.02 kcal/mol at the mPW1/SDD level, or ca. 19 kcal/mol per molecule of water. That this figure is higher than the largest stabilization of 1:1 complexes (16.61 kcal/mol before BSSE correction), indicates that some extra source of stabilization is present. Solvation substantially distorts the internal geometry of cisplatin: the energy of cisplatin frozen at the solvated geometry shown in Figure 3.5 is 12.71 kcal/mol higher than at equilibrium. Pt—Cl and Pt—N distances undergo large changes: the former are weakened by strong interactions with water molecules (average value 2.457 vs. 2.368 Å in the gas phase), while the latter are shorter than in the gas phase (2.039 vs. 2.126 Å). These changes are also reflected in bond CP properties, where the density in Pt—N bonds increases by around 0.02 au while that in Pt—Cl bonds falls by around 0.01 au. Large changes are also evident in the angles around Pt, for instance the N—Pt—N and Cl—Pt—Cl angles fall by -13.5° and -7.5°, respectively, while N—Pt—Cl angles open out by 9.1°, such that the coordination geometry is much closer to the ideal square-planar values.

Several familiar H-bonding motifs are apparent in the structure shown in Figure 3.5. The strong, bridging N—H...O—H...Cl pattern is present on both faces of the complex, as is one N—H...O...H—N and two Cl...H—O—H...Cl bridges. However, several more subtle patterns also emerge: for instance, the remaining NHs cannot be bridged by a single water due to unfavourable angles, so instead two waters form an N—H...O—H...O...H—N bridging motif here. Removing one of these waters results in a non-bridged structure with one N—H unsolvated. This asymmetry of H-bonding was not expected at the outset of our studies. It is also evident from Figure 3.5 that this first solvation shell contains several water molecules only bound to other waters, rather than to cisplatin itself. However, removal of these molecules again disrupts the H-bond pattern to such an extent that donor or acceptor atoms are no longer solvated, and cannot be discounted from the primary solvation shell.

Table 3.3 Geometry and individual H-bond energies in 10:1 complex, and comparison with 1:1 complexes.

	B...H (Å)	A...B (Å)	B...H—A (°)	E_{HB} in 10:1complex (kcal/mol)	E_{HB} in 1:1 complex (kcal/mol)
N—H...O	3.032	3.718	125.4	[a]	9.13
N—H ...O	1.803	2.833	173.0	6.72	9.13
N—H ...O	1.973	2.120	151.3	4.78	8.80
N—H ...O	1.773	2.528	174.7	8.85	8.80
N—H ...O	1.679	2.724	172.1	11.14	[b]
N—H ...O	1.800	2.838	173.2	8.97	[b]
O—H ...Cl	2.237	3.092	144.9	6.40	7.47
O—H ...Cl	2.183	3.152	167.6	8.12	7.47
O—H ...Cl	2.478	3.310	142.6	5.17	2.18
O—H ...Cl	2.427	3.300	147.6	5.77	2.18
O—H ...Cl	2.383	3.268	149.9	5.50	2.18
O—H ...Cl	2.410	3.235	141.8	5.33	2.18
O—H ...O	1.548	2.564	178.2	17.35	[b]
O—H ...O	1.502	2.523	174.7	19.07	[b]
O—H ...O	1.537	2.548	171.0	16.66	[b]
O—H ...O	1.310	2.414	174.6	27.77	[b]
O—H ...O	1.526	2.535	170.9	17.85	[b]
O—H ...O	1.551	2.559	172.3	16.00	

a: No H-bond critical point present; b: No such interaction in any 1:1 complex.

In total, the complex contains six N—H...O, six O—H...Cl, and a further six O—H...O H-bonds, all of whose presence was confirmed by the existence of a bond CP and intermolecular bond path. As noted above, eq. 3.1 can be used to decompose the overall interaction energy into contributions of single H-bonds, giving a better insight of the system. All such results are reported in Table 3.3, and where appropriate H-bond strengths in the 10:1 and 1:1 complexes are compared. This analysis reveals that the strong N—H...O and O—H...Cl H-bonds that stabilize complex 1 are weaker in the 10:1 complex. Conversely, the bridging Cl... H—O—H...Cl interactions are much stronger here, presumably due to the increased polarization of Pt—Cl and O—H bonds. However, perhaps the most striking feature of Table 3.3 is the dominant role played by O—H...O H-bonds in determining the overall stabilization of the complex. 105.40 kcal/mol, or 59% of the overall stabilization, comes from these six contacts, whose strength ranges from 13.02 to 23.14 kcal/mol (*cf.* 9.51 kcal/mol stabilization of the isolated water dimer at the same

level). Thus, it is apparent that, while most direct H-bonds to cisplatin are slightly weaker in this solvated shell than in isolation, water...water interactions are significantly enhanced by their proximity to cisplatin.

F- Cisplatin's hydrolysis

As noted above, solvation plays a key role in the activation of cisplatin, in which chloride ions are successively replaced by water to form the active mono- and di-aqua species. The reaction is thought to proceed via a five-coordinate trigonal-bipyramidal transition state, with a reported barrier of around 24 kcal/mol,[2,30,31] well reproduced in a recent in vacuo DFT study by Zhang and co-workers.[32] Costa *et al.*[33] have recently reported a similar study of the hydrolysis of cis-dichloro(ethylenediamine)platinum(II), demonstrating the importance of solvent effects. Because we have observed H-bonding to water to be important for the ground state of cisplatin, we have reexamined the potential energy surface for this reaction in the presence of explicit water molecules. Indeed, the most stable 1:1 complex found here is essentially identical to that reported by Zhang *et al.* as the first stage of hydrolysis (see 1.5.5).

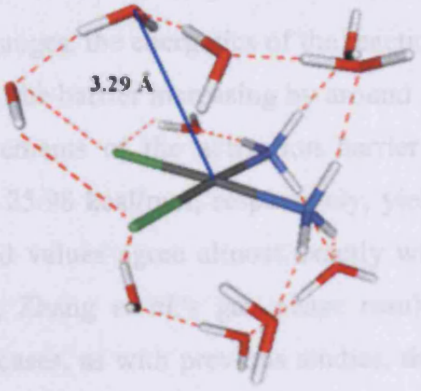
Table 3.4 compares selected geometrical and energetic parameters for this reaction in the gas phase and within the solvation shell described above, and the corresponding geometries are shown in Figure 3.6. Interestingly, a different water molecule to that found by Zhang *et al.* is involved in the hydrolysis reaction. Instead of the water involved in strong bridging H-bonds, we find that the solvent molecule corresponding to the least stable 1:1 complex found above reacts most easily. This observation may have its root in the cisplatin's ability of forming direct Pt—O bonds, or may simply be due to the fact that this water molecule loses less stabilisation due to H-bonding than others.

Table 3.4 Energy and selected geometrical properties of Hydrolysis reaction of cisplatin.

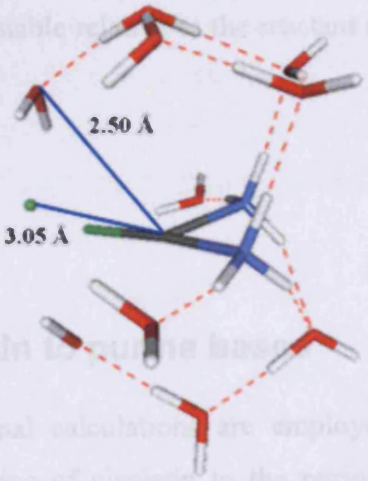
		Reactant	TS	Product
Pt—Cl: (Å)	*in vacuo*	2.38	2.74	4.03
	10:1 H_2O	2.46	3.05	3.99
Pt—O: (Å)	*in vacuo*	3.59	2.40	2.12
	10:1 H_2O	3.29	2.50	2.10
Cl—Pt—O: (°)	*in vacuo*	58.1	68.3	135.7
	10:1 H_2O	68.5	64.2	45.2
Cl—Pt—N: (°)	*in vacuo*	177.8	143.4	47.0
	10:1 H_2O	175.4	148.5	138.3
O—Pt—N: (°)	*in vacuo*	123.8	148.3	176.7
	10:1 H_2O	109.8	147.2	176.2
Energy[a]	*in vacuo*	0.0	22.92	7.29
(kcal/mol)	10:1 H_2O	0.0	24.28	7.62

a: values relative to the energy of reactant.

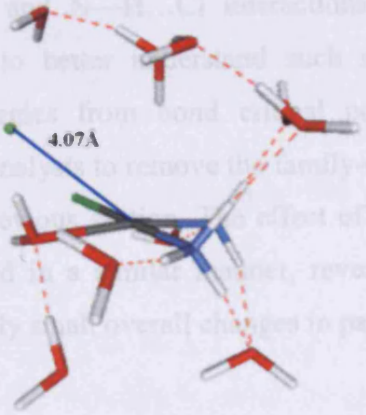
Solvation has a large effect on the reaction path, leading both to longer Pt—Cl and Pt—O lengths at the TS, although the angle of approach (as measured by the Cl—Pt—O angle) is essentially conserved. Variations in bond lengths on moving to the TS show even larger changes: Pt—Cl stretches by just 0.36 Å in the gas phase, but by 0.59 Å in solvation, while Pt—O falls by 1.19 Å without and 0.78 Å with solvation. Thus, it appears that the presence of explicit solvent water leads to a later TS, that is, one that has moved further towards products than would otherwise be predicted. Considerable changes are found in the final product structures, most notably in the Cl—Pt—O and Cl—Pt—N angles. This is due to the hydrogen bonding properties of the liberated chloride ion, which *in vacuo* must form H-bonds to the ammine NHs, but in the solvated model prefers to H-bond to water molecules, thereby having more freedom to move from the metal centre.



Reactant



TS



Product

Figure 3.6 Reaction path for hydrolysis of cisplatin with 10 explicit water molecules.

Despite these rather large changes, the energetics of the reaction are only slightly altered by our explicit solvation model, the barrier increasing by around 1 kcal/mol to 24.28 kcal/mol. Three experimental measurements of the activation barrier have been reported, giving values of 22.96, 23.97, and 25.98 kcal/mol, respectively, yielding an average value 24.30 kcal/mol. Thus, our solvated values agree almost exactly with this average, although all predicted values (including Zhang *et al.*'s gas phase results) are within the estimated experimental error. In both cases, as with previous studies, the overall reaction is found to be endothermic, presumably due to the separation of charge involved. It is somewhat surprising that, despite the availability of many more H-bond donor groups, the product of this reaction is marginally less stable relative to the reactant in solution as compared to the gas phase.

3.3 Binding of cisplatin to purine bases

Ab initio and density functional calculations are employed to investigate the role of hydrogen bonding in the binding of cisplatin to the purine bases guanine and adenine. Through the use of the theory of AIM, it is shown that hydrogen bonds are ubiquitous in such systems, with N—H...N and N—H...Cl interactions present in addition to the expected N—H...O. In order to better understand such systems, a new method for predicting hydrogen bond energies from bond critical point properties is proposed, employing partial least squares analysis to remove the family-dependence of simple models such as the one illustrated in previous section. The effect of platination on the pairing of guanine with cytosine is studied in a similar manner, revealing large redistributions of hydrogen bonding but surprisingly small overall changes in pairing energy.

3.3.1 Calculation methods: the hydrogen bond model

As seen in section 1.3, very high level calculations, including extrapolation to the complete basis set limit and treatment of correlation using *e.g.* coupled cluster methods, are required for quantitatively accurate results on DNA base pairing. Such calculations are unfeasible for the large systems studied here, and in any case our goal is to explore qualitative trends rather than achieve quantitative accuracy. Therefore, we have taken an alternative route, and attempted to test this against experimental or higher level theoretical results wherever possible. All geometry optimisations were carried out without symmetry constraints at the HF level using the 6-31G(d,p) basis set[19,20] on C, H, O, and N atoms and the SDD basis set and ECP[15] on Pt. Following harmonic frequency calculation confirmation as minima or transition state, subsequent single point energy and electron density calculations were performed using the standard B3LYP density functional[17,18] with a DGDZVP basis set[34] on C, H, O, and N and SDD on Pt. An essentially equivalent method has recently been shown to accurately reproduce the pairing energy of guanine with cytosine.[35]

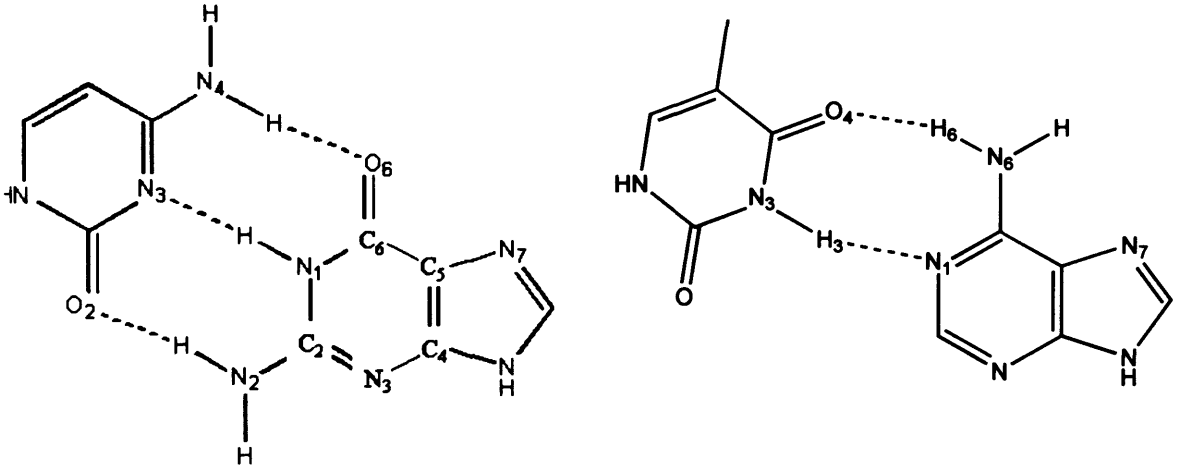


Figure 3.7 Numbering scheme for guanine...cytosine and adenine...thymine pair.

The H-bond model illustrated in previous section is highly family-dependent and is valid only for cisplatin-water complexes, therefore it cannot be applied to complexes of purine molecules with cisplatin. As discussed in paragraphs 2.7.3 and 3.2.2, many studies have demonstrated approximately linear relations between H-bond stabilisation energy and both the increase in density at H...B bond CP and the decrease at A—H for a wide range of

A—H…B systems. For instance, a recent study[29] showed a high-quality, family-independent relation between E_{HB} and $(\rho_0 - \rho)/\rho_0$, where ρ is the density at the A—H bond CP in the H-bonded complex and ρ_0 is the equivalent value in the uncomplexed A—H donor. In order to check how best to model the H-bonding interactions of cisplatin-DNA models, and to re-train such models at the theoretical level used, we extended the training set used previously[29] to encompass a much wider range of hydrogen bonded, including complexes of cisplatin with water, HF *etc.* taken from section 3.2. Models of counterpoise corrected[28] hydrogen bond stabilisation energy, E_{HB}, were then re-trained, with all properties evaluated at the B3LYP/DGDZVP level (see Table 3.5, Graph 1 and Figure 3.8). Overall, $\rho_{H\ldots B}$ gave the best single parameter linear fit to E_{HB} (r^2 = 0.96, rms error = 1.74 kcal/ mol), notably better than $[(\rho_0 - \rho)/\rho_0]_{A—H}$ (r^2 = 0.92, rms error = 2.36 kcal/mol) and the variation at the donor hydrogen nucleus as shown in previous section (r^2 = 0.85 only, and therefore not taken into further consideration here).

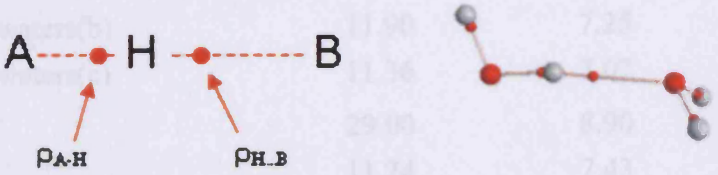
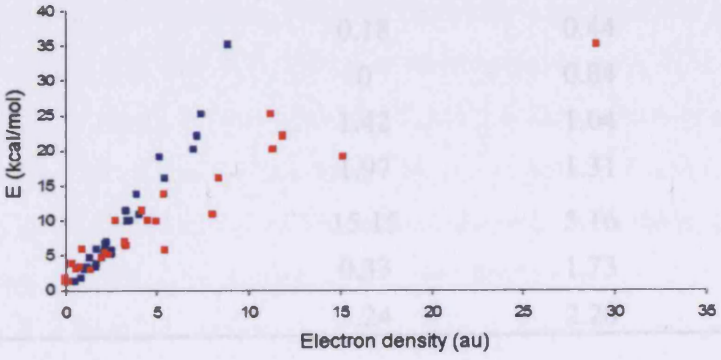


Figure 3.8 Electron densities in the A—H…B hydrogen bond.



Graph 1 Binding energies vs. electron densities: ■$\rho_{H\ldots B}$ and ■$[(\rho_0 - \rho)/\rho_0]_{A—H}$

Table 3.5 Binding energy and electron density properties of H-bonded complexes.

	$[(\rho_0-\rho)/\rho_0]_{A-H}$	$\rho_{H...B}\times10^{-2}$	ΔE_{bsse}
$(HF)_2$	2.36	2.50	5.22
$(H_2O)_2$	2.17	2.14	5.53
$ClH...OH_2$	5.42	2.55	5.63
Cisplatin...CH_3F	0.90	1.72	5.83
Cisplatin...HCl	3.33	2.16	6.43
FH...NCH	3.24	2.20	6.78
$FH...OH_2$	4.97	3.37	9.83
Cisplatin...NH_2CH_3	4.50	3.40	10.04
Cisplatin...NH_3	4.20	3.31	11.31
Cisplatin... H_20	2.70	3.55	10.03
Cisplatin...HF	8.04	4.07	10.86
$FH...NH_3$	5.36	3.91	13.8
Cisplatin+2 waters(a)	8.36	5.43	15.98
Cisplatin+2waters(b)	11.90	7.25	22.08
Cisplatin+2waters(c)	11.36	7.07	20.13
$HOH...OH^-$	29.00	8.90	35.23
GC	11.24	7.43	25.14
$(H_2S)_2$	0	0.53	1.20
HOH... FCH_3	0.83	1.68	3.30
$NCH...OH_2$	0.82	1.21	3.30
$HCCH...H_2O$	0.62	1.28	3.03
$HCCH...SH_2$	0.18	0.44	1.24
$NCH...H_2S$	0	0.84	1.64
HCl... H_2S	1.42	1.04	2.89
FH... H_2S	1.97	1.31	4.69
$NH_3...OH^-$	15.15	5.16	19.11
$NH_3...$ H_2S	0.33	1.73	3.81
$NH_3...H_2O$	3.24	2.28	6.91

However, both descriptors $\rho_{H...B}$ and $[(\rho_0 - \rho)/\rho_0]_{A—H}$ showed some family dependence, with slightly different fits for H-bonds involving organics or inorganics. We therefore carried out a partial least squares (PLS) analysis to incorporate both density properties into a single model: using just one latent variable, PLS yielded a much less family-dependent fit, shown below (E_{HB} in kcal/mol, density properties in au):

$$E_{HB} = 0.38 + 187.01\ \rho_{H...B} + 65.98\ [(\rho_0 - \rho)/\rho_0]_{A—H} \quad (3.2)$$

n=28; $r^2 = 0.974$; $Q^2 = 0.972$; rms error = 1.36 kcal/mol

Thus, by combining density properties from A—H and H...B bonds, we are able to reduce the overall error of fitting by around 0.4 kcal/mol, and to produce a model equally applicable to organics or inorganics. To the best of our knowledge, combining closely related density properties with PLS to improve on simple linear fits to E_{HB} is a new approach, and one that appears worthy of further applications.

3.3.2 Results and discussion

A- Monofunctional platinum adducts

Initial optimisation of complexes of *cis*- and *trans*-$[Pt(Cl)(Pur)(NH_3)_2]^+$ (Pur = adenine or guanine) identified two stable binding sites for platinum complexes on guanine (O6 and N7) and two on adenine (N1 and N7): all other starting points (*e.g.* N3) for optimisation either reverted to one of these, or was unstable. Table 3.6 shows that, as expected, the N7 guanine site is favoured over the N7 of adenine by *ca.* 15 kcal/mol and O6 of guanine by *ca.* 12 kcal/mol. Complexation at N1 of adenine is relatively favourable, but as this site is blocked by hydrogen bonding in duplex DNA, this binding mode is not typically seen experimentally, and is therefore not considered further in this work. Furthermore, cisplatin forms consistently more stable complexes than its trans analogue. While the affinity of cisplatin for guanine N7 is well established by many previous studies,[36-38] several features of Table 3.6 are worthy of further comment. Firstly, the calculated binding energies are in

excellent agreement with literature values, where available, supporting our choice of theoretical method.[8]

Secondly, the difference in binding energy of cis- and transplatin is remarkably constant across three different binding sites, ranging from 12.0 kcal/mol for G_{N7} to 9.6 for A_{N7}, such that the preferred binding site of transplatin is also G_{N7}. The lesser stability of the trans-complexes is well known, and widely rationalised as a manifestation of the "trans-effect".[39] However, that this difference is approximately constant is significant, because transplatin is much less able to form hydrogen bonds to guanine or adenine than is cisplatin (see below). Such hydrogen bonds have been proposed as the means by which cisplatin shows a preference for G_{N7}, but the results in Table 3.6 suggest that hydrogen bonding can play only a partial role in determining this preference. It is also evident that binding to G_{O6} is considerably weaker than to G_{N7}, reflecting the lesser importance of such carbonyl binding modes.

Table 3.6 Monofunctional platinum adducts.

	Binding energy (kcal/mol) [a]		r (Pt—X) (Å)	ρ_{CP} (Pt—X) (au)
cisPt-G_{N7}	80.45	(80.69) †	2.092	0.1025
transPt-G_{N7}	68.54	(67.29) ‡	2.118	0.0964
cisPt-G_{O6}	68.69	(67.41) ‡	2.059	0.0933
transPt-G_{O6}	57.95	(59.22) ‡	2.129	0.0799
cisPt-A_{N7}	65.51	(65.47) †	2.077	0.1058
transPt-A_{N7}	56.04	(53.69) ‡	2.103	0.0996
cisPt-A_{N1}	71.56		2.062	0.1127

a: Values in parenthesis: † from ref. 8, ‡ calculated value at the same level as ref. 8.

Table 3.6 also contains distance and bond CP data for all Pt—X bonds. Neither shows any clear relation with the total binding energy – the shortest Pt—N bond is found in cisPt-A_{N7}, while the most strongly bound complex, cisPt-G_{N7}, contains a Pt—N bond of intermediate

length, albeit with rather high electron density. Moreover, the Pt—O bond in cisPt-G_{O6} is very short, but as measured by ρ_{CP} is weaker than any Pt—N bond. This suggests that overall binding energy must be considered as a sum of covalent and hydrogen bonding effects, and hence that properties of Pt—X bonds should not be expected to correlate with overall binding energy, but only with the covalent contribution to this.

As shown in Table 3.7, six of the seven complexes considered contain intramolecular base-ligand H-bonds, as evidenced by the presence of a (3, -1) CP and accompanying bond path. CisPt-G_{N7} contains the shortest intramolecular hydrogen bond of all mono-functional adducts studied (N—H...O = 1.892 Å), an interaction which also has the highest electron density and Laplacian at the H-bond CP. Only cisPt-A_{N1} contains no such H-bonds, instead adopting a conformation in which the planes of Pt-coordination and base are almost orthogonal (dihedral = 74.9°). Again, the trend of binding energies in Table 3.6 cannot be explained solely by this data: for instance, the N—H...N interaction in transPt-G_{O6} is shorter than the N—H...O of transPt-G_{N7}, but the latter complex is considerably more stable. It is notable that the complex with the highest overall binding energy, cisPt-G_{N7}, contains both a relatively strong Pt—N bond and the strongest N—H...O interaction, as measured by ρ_{CP}.

Table 3.7 Geometrical and electron density properties of hydrogen bond interactions.

	A—H...B	ρ(H...A) (au)	$\nabla^2\rho$(H...A) (au)	ρ_{A-H} (au)	r(H...A) (Å)	E_{HB} [a] (kcal/mol)
cisPt-G_{N7}	N—H...O	0.0293	0.1164	0.323	1.892	7.46
transPt-G_{N7}	N—H...O	0.0183	0.0748	0.328	2.104	4.21
cisPt-G_{O6}	N—H...Cl	0.0213	0.0694	0.330	2.278	5.74
transPt-G_{O6}	N—H...N	0.0204	0.0708	0.327	2.130	4.80
cisPt-A_{N7}	N—H...N	0.0175	0.0574	0.327	2.232	4.48
transPt-A_{N7}	N—H...N	0.0087	0.0313	0.330	2.568	0.52

a: Calculated from eq. 3.2.

Values of $\rho_{A—H}$ and $\rho_{H...B}$ may be used, *via* eq. 3.2, to estimate hydrogen bond strengths, E_{HB}, also reported in Table 3.7. This analysis shows that the N—H...O contact in cisPt-G_{N7} is significantly stronger than any other present, but that all complexes except transPt-A_{N7} undergo substantial stabilisation (4 – 6 kcal/mol) due to H-bonding. Thus, the extra stability conferred upon cisPt-G_{N7} by H-bonding is insufficient to explain the overall stability of this complex, echoing the conclusions of Lippard *et al.*[40] The complex transPt-A_{N7} contains the same N—H...N6 contact as its *cis-* analogue, but the steric requirements of *trans-*coordination mean that in this case the H-bond is far from linearity (143.7°), leading to much lower stabilisation due to H-bonding here.

Our estimate of the N—H...N interaction in cisPt-A_{N7} (4.48 kcal/mol) agrees well with Friesner's result of *ca.* 5 kcal/mol.[8] Several studies, including those of Hobza[41] and Burda,[42] have shown significant pyramidalization of adenine $-NH_2$ groups on complexation to metals, a result supported by our calculations on this complex (sum of angles around N7 = 336.5°). The complex cisPt-G_{O6} is stabilised by a Pt—Cl...H interaction, the presence of which is perhaps unsurprising given our findings on the acceptor strength of Pt—Cl groups, section 3.2. Thus, even in these relatively simple cases, the abundance of donor and acceptor groups mean that almost all complexes are significantly stabilised by hydrogen bonding. Only in transPt-A_{N7} does this not hold: here also a (3, -1) CP corresponding to a hydrogen bond is found, but with such low properties that its energy is estimated at just 0.52 kcal/mol, *i.e.* effectively zero given the RMS error on equation 3.2.

Having estimated the H-bond energy in each complex, we can estimate the stabilisation due to covalent binding of platinum to O6 or N7, E_{Cov}, as the difference between overall stabilisation, and E_{HB}, *i.e.* binding energy = $E_{HB} + E_{Cov}$. These values are reported in Table 3.8, along with density properties of the Pt—X (X=O6 or N7) in each complex. Since eq. 3.2 is approximate, and since this approach ignores any cooperativity between E_{HB} and E_{Cov}, such values are necessarily only estimates. However, it is clear that cisPt-G_{N7} contains the strongest Pt—X bond in this series, approximately 9 kcal/mol greater than that in transPt-G_{N7}, while Pt—O bonds to guanine and Pt—N bonds to adenine are weaker again. Encouragingly, there is a linear relationship ($r^2 = 0.96$) between E_{Cov} and ρ(Pt—X) for the four guanine complexes, though this does not hold for adenine complexes. This finding is

tested for more complexes below, but further work is required to establish whether such a relation is general, since the complexes studied here cover only a narrow range of binding energies. This approach therefore suggests that the extra stability of cisPt-G_{N7} over cisPt-A_{N7} is due to both covalent and hydrogen bonding effects, with the former dominating.

Table 3.8 Covalent contribution to binding energy, and properties of Pt—X (X=O6 or N7) bonds.

	E_{Cov} (kcal/mol)	ρ(Pt—X) (au)	$\nabla^2\rho$(Pt—X) (au)
cisPt-G_{N7}	72.99	0.1025	0.374
transPt-G_{N7}	64.33	0.0964	0.365
cisPt-G_{O6}	62.95	0.0933	0.471
transPt-G_{O6}	53.15	0.0799	0.397
cisPt-A_{N7}	61.03	0.1058	0.390
transPt-A_{N7}	55.52	0.0996	0.379

B- Bifunctional platinum adducts

It is known that when cisplatin binds to DNA, the major products are 1,2 intrastrand GG and AG adducts,[43] where platinum binds to both bases at the N7 position. We have therefore investigated a number of bifunctional adducts using the same methods as above, simply by replacing the chloride ion in the monofunctional complexes with an appropriate base. Table 3.9 contains binding energies and selected geometrical parameters of these bifunctional adducts, and the optimised geometry of a representative compound, G_{N7}-cisPt-G_{N7}, shown in Figure 3.9.

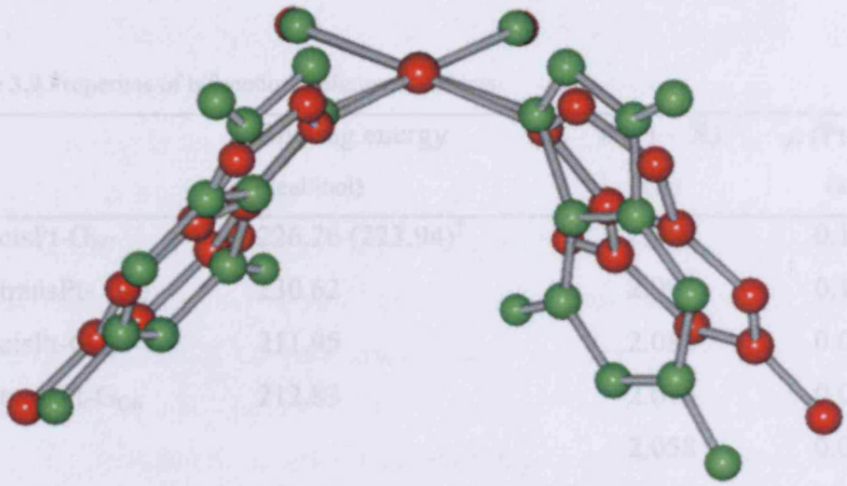


Figure 3.9 Comparison of X-ray[44] (green) and optimised (red) structures of G_{N7}-cisPt-G_{N7}.

Further support for the choice of method comes from the overall good agreement of optimised geometry of cisPt-G_{N7} with a structure of *cis*-[$Pt(NH_3)_2$] complexed to GpG obtained by Sherman *et al. via* X-ray diffraction,[44] as shown in Figure 3.9 (phosphate and sugar groups have been omitted from the X-ray structure for clarity). Coordination about the Pt centre and internal geometry of each guanine is almost exactly reproduced, as is the geometry and orientation of one guanine. The orientation of the second ring is shifted by *ca.* 12° from the X-ray geometry, as measured by the dihedral angle between the planes of each ring. However, differences in orientation of similar magnitude are also found between the four independent molecules within the crystalline unit cell, so such a difference can probably be ascribed to crystal packing forces. Such forces would also explain why the optimised geometry is very close to C_s symmetry, unlike the X-ray structures which are all substantially asymmetrical.

Table 3.9 Properties of bifunctional platinum adducts.

	Binding energy (kcal/mol)	r (Pt—X) (Å)	ρ (Pt—X) (au)
G_{N7}-cisPt-G_{N7}	226.26 (223.94)[†]	2.099	0.101
G_{N7}-transPt- G_{N7}	230.62	2.093	0.103
G_{O6}-cisPt-G_{O6}	211.95	2.082	0.088
G_{O6}-transPt-G_{O6}	212.83	2.076	0.089
		2.058	0.095
A_{N7}-cisPt-A_{N7}	190.64	2.060	0.112
A_{N7}-transPt-A_{N7}	196.14	2.060	0.111
		2.063	0.111
A_{N7}-cisPt-G_{N7}	208.86	2.077[a]	0.107
		2.084[b]	0.105
cisPt-G "chelate"	165.56 (164.16)[†]	2.140 (N)	0.091
		2.117 (O)	0.087

[†]:Calculated value at the same level as ref. 8; a: (Pt—A); b: (Pt—G).

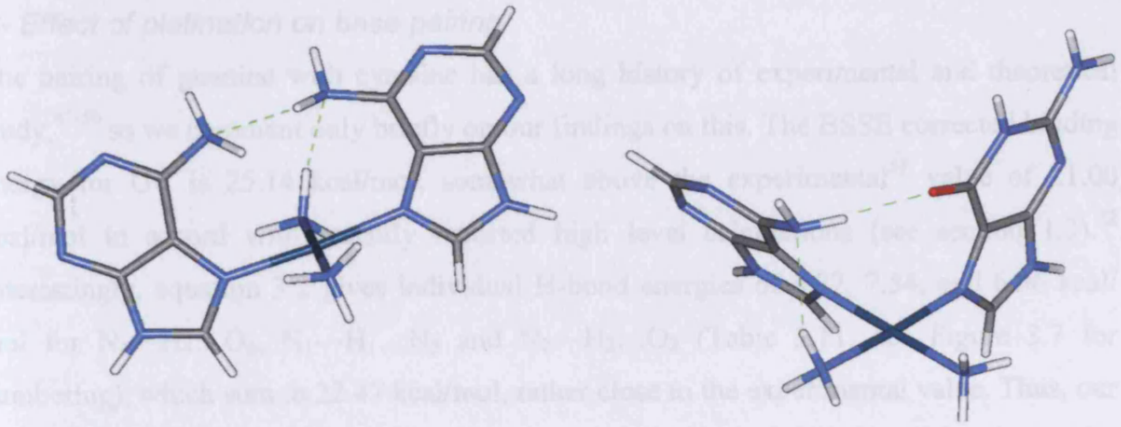
As expected, complexes at the N7 site of guanine are most stable, though interestingly the complex of transplatin is more stable than that of cisplatin, perhaps due to decreased steric repulsion between bases, a hypothesis explored further below. Indeed, all *trans-* complexes considered are more stable than their *cis-* isomers. Such complexes are unlikely to form in a single strand of DNA due to the constraints of the backbone, but could conceivably form across strands. This is in accord with the hypothesis that cisplatin's activity is related more to its ability to form 1,2 intrastrand linkages than simply to the strength of binding. Complexes through the O6 site of guanine are less stable, and show less difference between *cis-* and *trans-* complexes, while adducts of adenine are less stable still, and the mixed

complex A_{N7}-cisPt-G_{N7} has intermediate stability. Our calculations also corroborate previous findings[45,46] that a "chelate" complex with $[Pt(NH_3)_2]^{2+}$ bound to O6 and N7 of a single guanine is stable, albeit with rather lower binding energy and hence less experimental importance than the more conventional bi-functional complexes.

Table 3.10 contains details of hydrogen bonds within bifunctional complexes. For the bis-guanine complexes, values are broadly similar to those in Table 3.7 for monofunctional adducts, *i.e.* replacement of Cl^- with guanine does not strongly affect the pattern of H-bonding. However, each N—H...O H-bond in G_{N7}-cisPt-G_{N7} is *ca.* 3 kcal/mol weaker than that in cisPt-G_{N7}, perhaps due to strain resulting from the proximity of two large bases. The two bis-adenine complexes reported in Table 3.10 form substantially asymmetric complexes. In both complexes, hydrogen bonds form a Pt—N—H...N_6—H...N_6—C ring structure (see Figure 3.10), in which Pt—N—H...N_6 is considerably shorter and stronger than N_6—H...N_6. Indeed, the former interaction in A_{N7}-transPt-A_{N7} is the strongest found in any complex considered in this study. Attempts to re-optimise this complex to the more expected symmetrical structure reverted to this structure in all cases. A similar pattern is seen in A_{N7}-cisPt-G_{N7}, where N—H...N and N—H...O H-bonds form an analogous ring structure, though the energy of these contacts is considerably lower than in the bis-adenine complexes. While the formation of such a motif would be hindered by a DNA backbone, the variety of H-bonds found in such apparently straightforward complexes is nonetheless remarkable. In contrast, however, no intramolecular H-bonds are present in the chelate structure, since both NH_3 groups are too remote from the guanine to form such interactions.

Table 3.10 Hydrogen bonding in bifunctional adducts.

	A—H…B	$\rho_{H...A}$ (au)	$\nabla^2 \rho_{H...A}$ (au)	r(H…A) (Å)	E_{HB} (kcal/mol)	E_{Cov} (kcal/mol)
G_{N7}-cisPt-G_{N7}	N—H…O (x2)	0.0207	0.0606	2.038	4.26	217.74
G_{N7}-transPt- G_{N7}	N—H…O (x2)	0.0232	0.0946	1.992	4.72	221.18
G_{O6}-cisPt-G_{O6}	N—H…N (x2)	0.0239	0.0816	2.054	4.93	202.09
G_{O6}-transPt-G_{O6}	N—H…N	0.0253	0.0860	2.021	5.12	202.59
A_{N7}-cisPt-A_{N7}	N—H…N	0.0224	0.0702	2.110	8.29	177.71
	N—H…N	0.0140	0.0442	2.317	4.64	
A_{N7}-transPt-A_{N7}	N—H…N	0.0289	0.0890	1.989	9.84	182.37
	N—H…N	0.0122	0.0389	2.393	3.93	
A_{N7}-cisPt-G_{N7}	N—H…O	0.0163	0.0698	2.088	5.19	200.22
	N—H…N	0.0164	0.0532	2.271	3.45	



a b

Figure 3.10 Optimised geometry of (a) A_{N7}-transPt-A_{N7} and (b) A_{N7}-cisPt-G_{N7}.

Again, we can estimate the contribution from covalent Pt—X bonds to overall binding energies by subtracting the sum of E_{HB} for all H-bonds in each complex – trends in E_{Cov} for monofunctional adducts are conserved here. These results confirm that the stability of G_{N7}-cisPt-G_{N7} and G_{N7}-transPt-G_{N7} is largely due to covalent effects, since in both cases the H-bonding characteristics are unremarkable. The extra stability of *trans-* complexes also appears to be due largely to covalent bonding, rather than to reduced steric repulsion. While the trends noted for monofunctional complexes are conserved in Tables 3.9 and 3.10, *i.e.* binding energy of $G_{N7} > G_{O6} > A_{N7}$, values for bifunctional complexes are considerably more than twice the values for mono-functional adducts throughout. This effect is largest for G_{O6}-transPt-G_{O6} (97 kcal/mol), falling to 47 kcal/mol for G_{N7}-cisPt-G_{N7}. This appears to result from increased covalent binding of bases to the doubly charged $[Pt(NH_3)_2]^{2+}$ centre, since the presence of extra H-bonds contribute at most around 10 kcal/mol.

C- Effect of platination on base pairing

The pairing of guanine with cytosine has a long history of experimental and theoretical study,[47-50] so we comment only briefly on our findings on this. The BSSE corrected binding energy for GC is 25.14 kcal/mol, somewhat above the experimental[51] value of 21.00 kcal/mol in accord with recently reported high level calculations (see section 1.3).[52] Interestingly, equation 3.2 gives individual H-bond energies of 8.37, 7.54, and 6.56 kcal/mol for N_4—H_4...O_6, N_1—H_1...N_3 and N_2—H_2...O_2 (Table 3.11, see Figure 3.7 for numbering), which sum to 22.47 kcal/mol, rather close to the experimental value. Thus, our chosen method appears to be capable of providing accurate, BSSE-free H-bond energies even in cases where multiple H-bonds are present.

Table 3.11 Hydrogen bond energies in free and platinated GC pairs (kcal/mol).

	N_4—H_4...O_6	N_1—H_1...N_3	N_2—H_2...O_2	E_{HB}	ΔE
GC	8.37	7.54	6.56	22.47	0.00
cisPt-G_{N7}≡C	4.89	8.76	8.95	22.60	+0.13
transPt-G_{N7}≡C	4.90	9.09	9.13	23.12	+0.65
cisPt-G_{O6}≡C	1.52	6.84	9.45	17.81	-4.66
G_{N7}-cisPt-G_{N7}≡C	3.78	9.66	10.63	24.07	+1.60
G_{N7}-transPt-G_{N7}≡C	3.80	9.38	10.52	23.70	+1.23
A_{N7}-cisPt-G_{N7}≡C	3.57	9.80	10.98	24.35	+1.88

Table 3.11 details how platination at various sites affects the pattern of GC pairing. It is clear that in all cases significant changes result from the covalent binding of platinum, and that the pattern of changes is broadly conserved. Throughout, the strongest H-bond in free GC is weakened by between 40 and 80% of its original value. This might be expected where the H-bond acceptor atom for this interaction, G_{O6}, is involved directly in platination, and indeed the largest changes are seen where this is the case. However, substantial disruption of G_{O6}'s acceptor ability also stems from coordination at G_{N7}. It is not clear whether this is due to the inductive effect of the positive metal centre, or to the direct effect of the Pt—N—H...O H-bond already present. Some evidence for the former scenario may come from the fact that the N_4—H_4...O_6 H-bond is weaker in the dicationic bifunctional adducts than the monofunctional ones. Further studies on this point are reported in next section.

In contrast, H-bonds in which guanine acts as an H-bond donor are generally stronger in platinated complexes than in free GC. This is more pronounced for N_2—H_2...O_2, wherein increases of 35-65% are observed, whereas smaller increases are seen in N_1—H_1...N_3 in most cases, and even a small decrease is found for cisPt-G_{O6}≡C. This decrease is clearly seen in electron density properties, but is not apparent from consideration of geometrical properties alone: neither H...N nor N...N distances (not reported) change significantly from their free GC values in this case. Also, that a larger increase in donor strength is found at

N_2—H_2, *i.e.* furthest from the site of platination, goes against electrostatic arguments which would suggest that N_1—H_1 should be affected more.

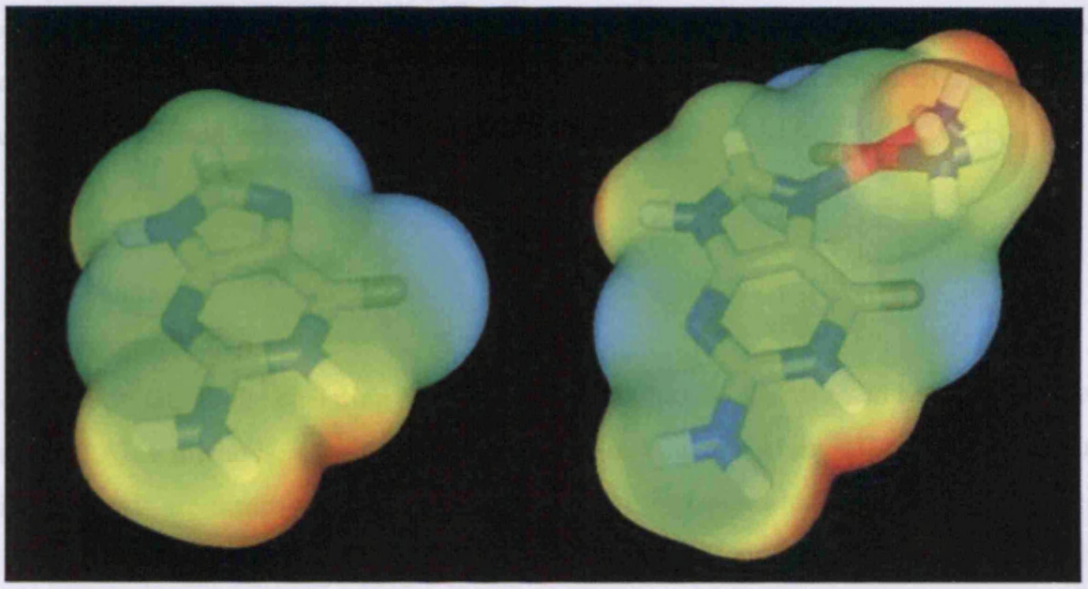


Figure 3.11 0.001 au iso-surface MEP of guanine and cisPt-G_{N7} (blue is negative, red positive).

Clearly, platination substantially changes the bonding and electron distribution within the guanine, yielding more subtle changes in the pattern of H-bonding than might initially be expected. Figure 3.11 shows the molecular electrostatic potential (MEP) on the 0.001 au isodensity surface for guanine and cisPt-G_{N7}. Drastic changes are evident throughout the molecule, most notably at O6 but also at most other donor and acceptor sites. The acceptor ability of O6 is almost completely lost, with only a very small region of negative potential associated with this atom, while N3's negative MEP is significantly enhanced, such that this site becomes the global minimum. Donor strengths of N_1—H_1 and N_2—H_2 are less apparently affected by platination at N7, nor is any major difference between these two sites evident in Figure 3.11.

Despite these changes, the overall strength of GC pairing is remarkably insensitive to platinum binding at G_{N7}, the largest change being +1.9 kcal/mol in the case of A_{N7}-cisPt-

G_{N7}≡C, and just +0.13 kcal/mol for cisPt-G_{N7}≡C. Binding to G_{O6}, on the other hand, reduces the overall stabilisation by 4.7 kcal/mol, due to a massive reduction in the strength of N_4—H_4...O_6, offset slightly by an increase in the strength of N_2—H_2...O_2. These results contradict previous suggestions[42] that platination enhances the hydrogen bonding between guanine and cytosine, suggesting a more subtle redistribution of stabilisation. An alternative explanation for the observed[53] increase in formation constant of cisPt-G_{N7}≡C over GC is discussed below.

D- Distortion of GC pair

The effect on GC of the chelating bifunctional adduct has been addressed before,[45,46] albeit not using AIM methods, so only a brief discussion is given. Here, all classical Watson-Crick H-bonds are destroyed, and the mutual planarity of bases lost, with a dihedral angle of 63°. However, there remains substantial overall stabilisation: only one intermolecular H-bond CP was found in this case, a very short, strong C=O...H-N contact (H...O = 1.642Å), predicted to have an energy of 26.43 kcal/mol. Thus if present, such a chelate would have a drastic effect on base pairing and DNA structure, though the results in Table 3.9 suggest this is energetically unlikely.

This redistribution of H-bond energy leads to geometrical changes in the GC pair, characterised in Table 3.12 as (a) the angle between C=O vectors in G & C; and (b) the dihedral angle between the mean planes of each base. The free GC pair is exactly planar, and the C=O vectors are almost exactly anti-parallel. This arrangement is broadly conserved in all complexes other than cisPt-G_{O6}≡C, with less than 2° change in C=O vectors and up to 7.4° in the dihedral between mean planes, the largest changes being found for cisPt-G_{N7}≡C. The geometry of this complex appears to show that the N—H...O H-bond from cisplatin induces this change in dihedral by "attacking" the bottom face of guanine's C=O, leading the N—H...O from cytosine to shift round to the top face, such that the two H-bonds to O6 are approximately collinear (161.2°). In contrast, platination at O6 leads to large changes in geometry, with *ca.* 15° change in C=O...C=O angle and almost 30°

between the mean planes of G and C, both of which can be ascribed to the almost complete loss of the N_4—H_4…O_6 interaction (see Figure 3.12).

Table 3.12 Effect of platination on geometry of GC pairing (see Figure 3.7 for definitions).

	Angle between C=O vectors in G and C (°)	Dihedral between mean planes of G and C (°)
GC	172.42	0.00
cisPt-G_{N7}≡C	170.85	7.42
transPt-G_{N7}≡C	172.96	2.18
cisPt-G_{O6}≡C	158.66	29.07
G_{N7}-cisPt-G_{N7}≡C	175.35	1.33
G_{N7}-transPt-G_{N7}≡C	173.96	1.38
A_{N7}-cisPt-G_{N7}≡C	175.25	2.57

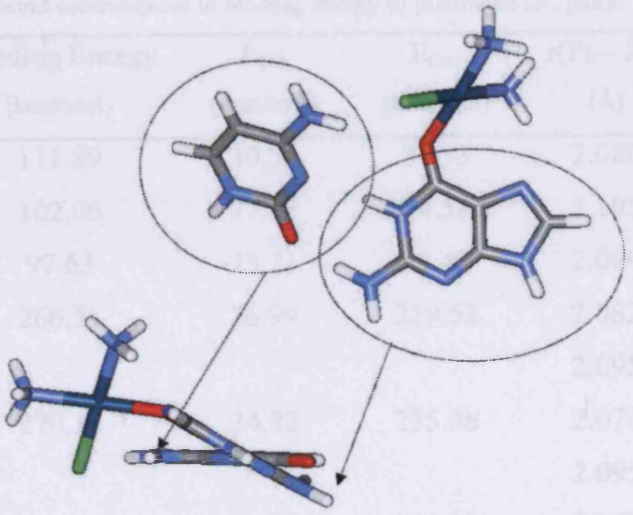


Figure 3.12 Schematic of cisPt-G_{O6}≡C, showing the dihedral between planes of G and C.

E- Pt—N(O) bonds

In general, intramolecular H-bond types and strengths in the GC paired complexes do not differ greatly from those reported in Tables 3.7 and 3.10, and so are not reported. Estimation of all inter- and intra-molecular H-bond energies gives the covalent contribution to the overall binding energy (Table 3.13). Combining these values with those in Tables 3.8 and 3.9 further confirms the excellent linear relation between E_{Cov} and ρ(Pt—X) for guanine complexes ($r^2 = 0.99$). Moreover, this analysis indicates an increase in E_{Cov} in the GC paired complexes compared with their unpaired analogues. For instance, the simple G_{N7} adduct of cisplatin shows an increase of 8.3 kcal/mol when paired with cytosine, with a corresponding decrease in bond length and increase in CP density. This effect is even more pronounced in other adducts, such that the average increase in E_{Cov} on addition of cytosine is 11 kcal/mol, reflected in bond lengths and electron densities throughout. Thus, although the formation energy of platinated GC pairs is greater than of isolated guanine, our analysis suggests this is due to the formation of stronger Pt—X bonds rather than to enhanced hydrogen bonding between guanine and cytosine, as proposed previously.[42]

Table 3.13 Covalent and H-bond contributions to binding energy of platinated GC pairs.

	Binding Energy (kcal/mol)	E_{HB} (kcal/mol)	E_{Cov} (kcal/mol)	r(Pt—X)[a] (Å)	$\Sigma\rho$(Pt—X) (au)
cisPt-G_{N7}≡C	111.89	30.56	81.33	2.080	0.106
transPt-G_{N7}≡C	102.06	27.48	74.58	2.105	0.100
cisPt-G_{O6}≡C	97.63	23.21	74.42	2.069	0.106
G_{N7}-cisPt-G_{N7}≡C	266.51	36.99	229.52	2.082	0.208
				2.095	
G_{N7}-transPt-	270.18	34.22	235.96	2.076	0.209
G_{N7}≡C				2.095	
A_{N7}-cisPt-G_{N7}≡C	245.74	36.23	209.51	2.077	0.212
				2.088	

a: Where two values given, the first corresponds to the base involved in a GC pair.

3.4 A systematic study of transition metals-GC interaction

As seen in chapter 1, many transition metals interact with DNA leading to potential anti cancer activity. In this section, the role of covalent and hydrogen bonding in transition metal complexes to guanine has been investigated. As for section 3.3, the effects on GC pairing are explored by the means of density functional calculations and Atoms in Molecules analysis.

3.4.1 Results and discussion

A- Metal...guanine complexes

Table 3.14 and Figure 3.13 report ligands, coordinations and electron configuration for each metal group studied. Ligands were chosen to obtain singly charged metals and, where possible, closed shell ions. In particular, all metals of the titanium group are $[M^{IV}Cl(Cp)_2]^+$, since such complexes are known to have anti-tumour activity.[54] The vanadium group elements considered are in the fifth oxidation state, as $[M^{V}=O(NH_3)Cl_2]^+$ in approximate square pyramidal conformation. The Cr group metals are d^3, octahedral as $[M^{III}Cl_3(NH_3)_2]^+$. Fe, Zn and Mn group elements share the same ligands, as octahedral $[MCl(NH_3)_4]^+$ complexes. The Co group metals are in the oxidation state I, with three NH_3 ligands in square planar orientation. The Ni group metals are derived from cisplatin as square planar $[M^{II}Cl(NH_3)_2]^+$. For the Cu group, two different ligands were employed, namely square planar $[M^{I}(NH_3)_3]^+$ for Cu and Ag and linear $[M^{I}NH_3]$ for Au. In this way, we aim to eliminate the gross effect of changing the overall charge on the complex, and hence to study the more subtle effects of metal and ligand change.

Table 3.14 Metal group elements features.

Metal/ oxidation state	Electron configuration/ spin state	Ligand(s)	Conformation
Ti(IV)	d^0 / singlet	$[Cl(Cp)_2]$	Tetrahedral
V(V)	d^0 / singlet	$[(Cl_2NH_3)=0]$	Square Pyramidal
Cr(III)	d^3 / quartet	$[Cl_3(NH_3)_2]$	Octahedral
Mn(II)	d^5 / sextet	$[Cl(NH_3)_4]$	Octahedral
Fe(II)	d^6 / singlet	$[Cl(NH_3)_4]$	Octahedral
Co(I)	d^8 / singlet	$[(NH_3)_3]$	Square planar
Ni(II)	d^8 / singlet	$[Cl(NH_3)_2]$	Square planar
Cu(I)	d^{10} / singlet	$(NH_3)_3$	Square planar
		(NH_3)	Linear
Zn(II)	d^{10} / singlet	$[Cl(NH_3)_4]$	Octahedral

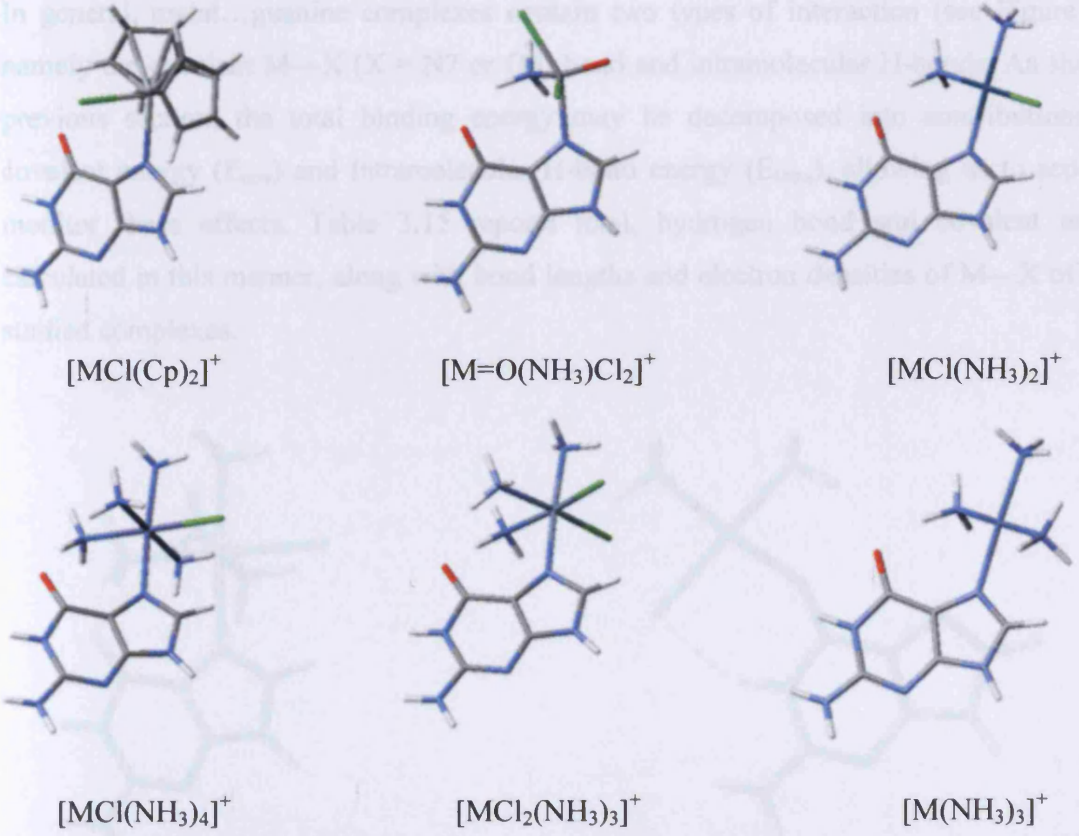


Figure 3.13 Ligands from Table 3.14, shown as complexes to N7 of guanine.

Table 3.15 contains energetic, geometrical, and electron density properties of all metal...guanine complexes studied in this work. As shown in previous section, Hartree-Fock theory provides reasonable geometries for metal...guanine adducts.[55] To further test this, Table 3.15 contains data from B3LYP/6-31G(d,p) optimisations (Lanl2DZ basis set and ECP on metal) as reported in ref. 8. The agreement between DFT and our combined HF optimised, DFT single point (HF/DFT) calculations is satisfactory for all the complexes. Only in the case of the Mn...guanine adduct does HF significantly underestimate the binding energy, probably due to the fact that the correlation becomes important in the case of the Mn(II) sextet state. Despite this, both methods agree that the binding energy of this complex is very low. As a further test, S^2 values for any non-singlet complexes were calculated, and in all cases are close to the ideal values (between 3.76 and 3.78 for the Cr group, and all exactly 8.75 for the Mn group), indicating little or no contamination from other spin-states.

In general, metal...guanine complexes contain two types of interaction (see Figure 3.14), namely the covalent M—X (X = N7 or O6) bond and intramolecular H-bonds. As shown in previous section, the total binding energy may be decomposed into contributions from covalent energy (E_{cov}) and intramolecular H-bond energy (E_{intra}), allowing us to separately monitor these effects. Table 3.15 reports total, hydrogen bond and covalent energies calculated in this manner, along with bond lengths and electron densities of M—X of all the studied complexes.

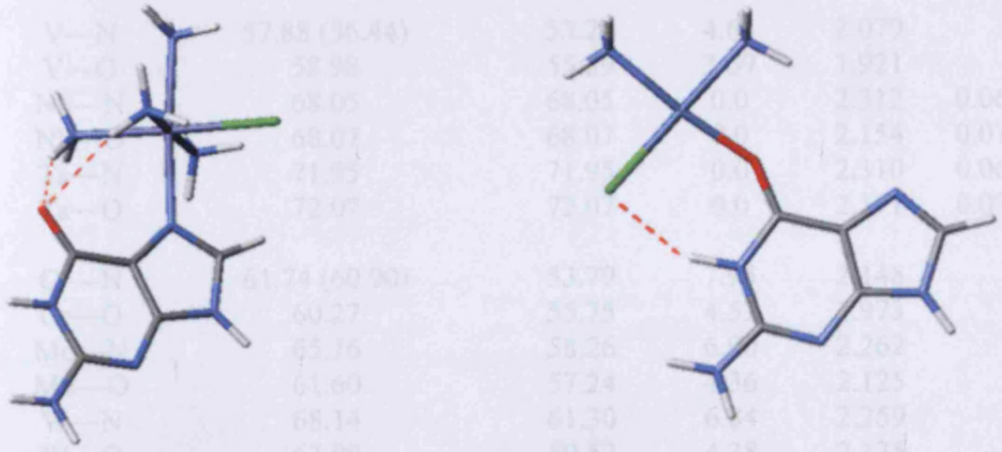


Figure 3.14 Intramolecular hydrogen bonds in Fe—N7 and Ni—O6 complexes.

Let us discuss briefly the nature of the intramolecular hydrogen bonding. As seen previously, bonds such N—H...O, N—H...Cl, and N—H...N are ubiquitous in metal...guanine complexes. The strength of these interactions depends on both the nature of the metal and the ligand, ranging from 3 to 8 kcal/mol. For example, the $[MCl(Cp)_2]$...guanine complexes (M = Ti, Zr, Hf) have weak C—H...N and C—H...O H-bonds, with energies no larger than *ca.* 4 kcal/mol. Similarly, N—H...O and N—H...N interactions for the Co group complexes average around 4 kcal/mol. In contrast, N—H...O and N—H...Cl H-bonds in the Fe group contribute 7 kcal/mol on average, while in the Ni group, the H-bond motifs noted for cisplatin are conserved, with energies between 5.7 and 7.5 kcal/mol, the strongest being cisplatin's. Thus, generally H-bonding contributes *circa* 10% to the total binding energy on average. However, in a few cases, no H-bonding was found, *e.g.* the binding energies of the Nb, Ta and Au complexes are purely from the metal-guanine interaction.

Table 3.15 Energies, electron densities and bond length in the metal...guanine complexes.

Group	Complex	Binding Energy[a] (kcal/mol)	E_{cov} (kcal/mol)	E_{intra} (kcal/mol)	r(M—X) (Å)	ρ(M—X) (au)
Ti	Ti—N	36.40(35.67)	33.20	3.17	2.342	0.048
	Ti—O	51.88	48.43	3.45	2.000	0.085
	Zr—N	48.52	46.08	2.44	2.424	0.049
	Zr—O	60.75	57.58	3.18	2.133	0.074
	Hf—N	50.35	47.91	2.44	2.116	0.076
	Hf—O	63.26	58.95	4.31	2.401	0.051
V	V—N	57.88 (56.44)	53.24	4.64	2.079	0.087
	V—O	58.98	55.29	3.69	1.921	0.105
	Nb—N	68.05	68.05	0.0	2.312	0.062, 0.029[b]
	Nb—O	68.07	68.07	0.0	2.154	0.077, 0.016[b]
	Ta—N	71.95	71.95	0.0	2.310	0.063, 0.035[b]
	Ta—O	72.07	72.07	0.0	2.171	0.075, 0.024[b]
Cr	Cr—N	61.74 (60.90)	53.79	7.95	2.148	0.063
	Cr—O	60.27	55.75	4.52	1.973	0.081
	Mo—N	65.16	58.26	6.90	2.262	0.064
	Mo—O	61.60	57.24	4.36	2.125	0.074
	W—N	68.14	61.30	6.84	2.259	0.068
	W—O	63.90	59.52	4.38	2.135	0.074

Mn	Mn—N	34.05 (44.01)	27.27	6.78	2.430	0.034
	Mn—O	36.76	31.86	4.90	2.203	0.046
	Tc—N	29.47	26.17	3.3	2.613	0.035
	Tc—O	29.61	25.56	4.03	2.408	0.043
	Re—N	26.07	19.03	7.03	2.866	0.025
	Re—O	21.30	17.09	4.21	2.627	0.032
Fe	Fe—N	46.18 (46.11)	38.78	7.40	2.222	0.043
	Fe—O	46.96	40.05	6.91	2.086	0.048
	Ru—N	53.20	46.34	6.86	2.249	0.058
	Ru—O	49.52	42.80	6.72	2.189	0.055
	Os—N	57.99	50.61	7.38	2.225	0.067
	Os—O	52.49	45.71	6.78	2.200	0.060
Co	Co—N	43.33 (43.20)	43.33	3.41	2.218	0.050
	Co—O	41.15	41.15	3.83	2.145	0.040
	Rh—N	50.75	50.75	3.92	2.218	0.063
	Rh—O	44.83	44.83	4.58	2.252	0.046
	Ir—N	59.06	59.06	4.27	2.153	0.082
	Ir—O	49.70	49.70	5.04	2.204	0.058
Ni	Ni—N	66.85 (67.60)	61.21	5.64	1.907	0.083
	Ni—O	62.32	56.44	5.88	2.017	0.077
	Pd—N	71.54	64.16	7.38	2.050	0.085
	Pd—O	62.28	56.73	5.55	2.081	0.094
	Pt—N	80.45 (80.69)	72.99	7.46	2.092	0.103
	Pt—O	68.69 (67.41)	63.00	5.69	2.059	0.093
Cu	Cu—O	25.83 (23.70)	21.10	4.73	2.522	0.023
	Ag—O	24.96	20.56	4.40	2.604	0.026
	Au—N	80.42	80.42		2.108	0.102
	Au—O	70.13	70.13		2.125	0.086
Zn	Zn—N	29.11 (30.67)	21.60	7.51	2.419	0.034
	Zn—O	30.61	24.34	6.27	2.178	0.046
	Cd—N	32.75	26.93	5.92	2.532	0.036
	Cd—O	33.27	28.58	4.69	2.362	0.043
	Hg—N	42.82	36.94	5.88	2.472	0.048
	Hg—O	34.04	31.11	2.93	2.376	0.050

a: calculated using combined HF/DFT approach, except for values in brackets, which are the binding energies calculated at B3LYP/6-31G**(lanl2dz). Positive values indicate stabilisation; b: reported as the value for the principal interaction, then any secondary interaction (see text).

Table 3.15 also reports the covalent energies of the metal...guanine bonds, (estimated as in previous section) as the total binding energy less the contribution from H-bonding. The results present a wide diversity of cases, mainly depending on the position of the metal in the periodic table. Thus, the covalent energies of the 54 studied complexes range between 20 and 80 kcal/mol, and the M—X length between 1.900 and 2.900 Å. Depending on the metal, the relative stability of M—O and M—N also varies. In such an intricate picture, it is still possible to spot interesting trends and tendencies that can provide important information on the interaction of transition metals with DNA.

Stable complexes of the titanium group metals are found binding to both O6 and N7, having covalent energies between 40 and 80 kcal/mol. As evident in Table 3.15, complexes at O6 are more strongly bound than at N7, with the average energy difference around 13 kcal/mol, in accord with the known preference of Ti^{IV} for hard oxygen ligands.[39] Moreover, the M—O bonds are shorter than corresponding M—N. Interestingly, the total and covalent energies increase down the group, such that the Ti complexes are *ca.* 20 kcal/mol more weakly bound than the Hf complexes.

In the vanadium group, the total binding and the covalent energy increases compared to Ti, an effect that appears to be stronger for the N7 complexes, leading to almost equivalence of the energy of V—N and V—O, which differ by just 2 kcal/mol The behaviour of Nb and Ta is intriguing, with M—N7 and M—O6 virtually isoenergetic (the difference is less than 0.1 kcal/mol on average). The metals are asymmetrically bound to N7 or O6, but also contain a secondary covalent interaction to the other nucleophilic site on guanine, *i.e.* Nb—N and Ta—N complexes contain an interaction to O6. AIM confirms this point (see Figure 3.15), and allows us to quantify each interaction: for instance, the electron density for the Ta—N complex at the Ta...O interaction is roughly half that of the Ta...N bond, while the Ta—O complex is contains a weaker Ta...N interaction, with *ca.* 30% of the strength of the Ta...O bond. This is particularly interesting, as such chelating complexes of cisplatin, although weakly bound, are very active in destabilising DNA pairing bases.[45,46,55]

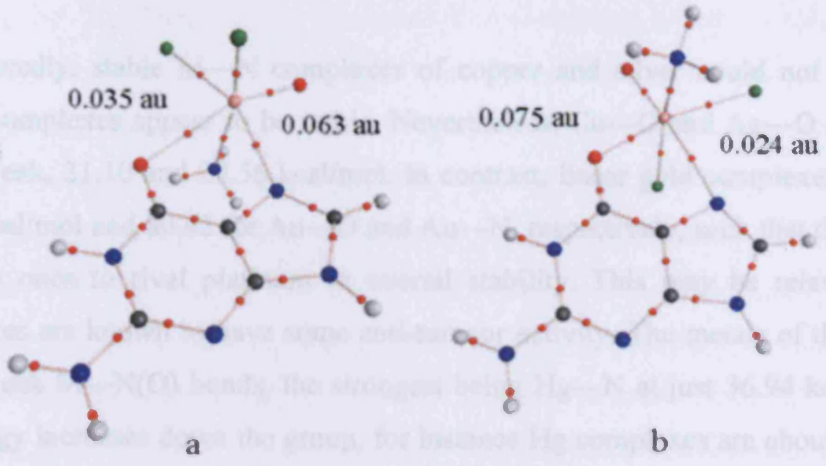


Figure 3.15 Molecular graphs of the Ta—N complex and Ta—O.

Cr group metals show similar binding energies to both N7 and O6 of guanine, with covalent energies ranging between 54 and 60 kcal/mol. Again, these energies increase down the group: here the difference between, for instance, W—N and Cr—N is 7 kcal/mol. Also, the M...O6 bonds are clearly shorter than corresponding M...N7. Metals of Mn group are the most weakly bound to guanine, with total binding energies less then 30kcal/mol and bond length greater than 2.5 Å on average. Interestingly, unlike all the other groups studied, here the total binding energy and the covalent energy decrease down the group. Thus, Re—N and Re—O are the weakest complexes considered here, *i.e.*, the covalent bonds are the longest and the covalent energy the smallest, 2.866 and 2.627 Å, and 19.03 kcal/mol and 17.09 kcal/mol, respectively.

Complexes of the Fe and Co group present similar features, *e.g.* the covalent energies of Fe—N and Co—N differ by less 3 kcal/mol on average. While Fe—O and Fe—N are essentially isoenergetic, Co shows a slight preference for N7. The heavier metals, though, show larger binding energies to N7, about 4 kcal/mol on average. Here, as seen before, the energy increases along the groups and thus larger metals are more strongly bound to guanine. The nickel group complex energies are clearly larger than any previous group and, perhaps not surprisingly, platinum complexes are the most strongly bound, about 10 kcal/mol more than nickel and palladium. All complexes follow platinum's known behaviour, and are bonded to N7 rather than O6.

Unexpectedly, stable M—N complexes of copper and silver could not be found, so only M—O complexes appear to be stable. Nevertheless, Cu—O and Ag—O covalent bonds are rather weak, 21.10 and 20.56 kcal/mol. In contrast, linear gold complexes are highly stable, 70.13 kcal/mol and 80.42 for Au—O and Au—N, respectively, such that these complexes are the only ones to rival platinum in overall stability. This may be relevant, as such gold complexes are known to have some anti-tumour activity. The metals of the zinc group form rather weak M—N(O) bonds, the strongest being Hg—N at just 36.94 kcal/mol. As before, the energy increases down the group, for instance Hg complexes are about 10 kcal/mol more stable than the corresponding Zn complexes.

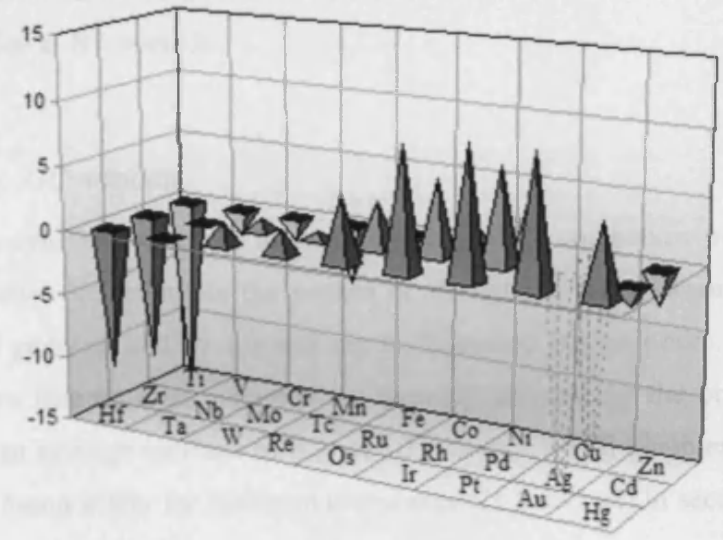


Figure 3.16 $\Delta_{N\text{-}O}$ (kcal/mol) of transition metals: positive values indicate a preference for N7.

An important property is the ability of the metal to form bonds to O6 of guanine, as it has been shown both experimentally and theoretically that metallation at this site may drastically disrupt the GC pairing, leading to the formation of globular DNA conformation.[45,46,55,56] Figure 3.16 contains a schematic drawing of the relative preference for O6 and N7 for the complexes reported in Table 3.15. The height of the pyramids is proportional to $\Delta_{N\text{-}O}$, the energy difference between M—N and M—O complexes, with positive values indicating

preference for N7. Therefore, as discussed above, titanium group complexes clearly prefer O6, with Δ_{N-O}= -13 kcal/mol on average. Moving to vanadium, the energy difference is much less pronounced, the pyramids disappearing as the values approach to zero. The situation is intriguing for both Mn and Cr groups, where the first row metals prefer O6 but the later rows showing the reversed trend and positive Δ_{N-O}. Most other metals prefer the N7 position, albeit with a few interesting exceptions. For instance, Fe—O is slightly more stable than Fe—N by about 2 kcal/mol. Cu and Ag, unlike Au, do not bond to N7 at all, although the M-O6 complexes are rather weak. Finally, the Zn group elements prefer O6, except Hg, which prefers N7. From Figure 3.16, another aspect becomes clear: in all cases, Δ_{N-O} values increase down the group, and in some cases (*e.g.* Mn, Zn groups) change sign. This is easily explained in terms relative hardness: descending down the group, the number of valence electrons is constant, while the size increases. Thus, the metals become relatively softer, and tend prefer coordination at N7 over O6.

B- Metal...GC adducts

Adding a cytosine molecule to the metal...guanine complexes in the standard Watson-Crick position hardly alters the pattern or strength of intramolecular H-bonding between metal and guanine, and so this will not be discussed in detail here. On the contrary, Table 3.16 shows that covalent energies are strongly affected by the presence of cytosine: we calculate an average increase in E_{cov} of 10 kcal/mol for all complexes considered, close to the value found solely for platinum complexes (11 kcal/mol) in section 3.3. As the increase in the energy is virtually constant, all trends discussed above are repeated. Strengthening of M—O and M—N bonds is also evident from AIM analysis, with both types of bond in general richer in electron density by 10 to 15%.

While the total energy is close to that of the free GC pair, those H-bonds in which guanine acts as a base are weaker, and those for which guanine is the proton donor are stronger, leading to significant deformation of the pair, as depicted in Figure 3.17. The analysis of data in Table 3.16 and Figures 3.17 and 3.18 reveal a similar pattern for all complexes considered here.

Table 3.16 E_{cov} and hydrogen bond energies in the Metal...GC complexes.

	Complex	E_{cov} (kcal/mol)	$H_4...O_6$ (kcal/mol)	$H_1...N_3$ (kcal/mol)	$H_2...O_2$ (kcal/mol)	E_{GC} (kcal/mol)
	GC		8.37	7.54	6.56	22.47
Ti	Ti—N	41.33	5.27	8.58	8.82	22.67
	Ti—O	61.10	2.67	2.12	10.91	15.70
	Zr—N	55.36	5.14	8.93	8.82	22.89
	Zr—O	73.45	0.00	5.29	10.80	16.09
	Hf—N	57.07	5.18	8.95	8.96	23.09
	Hf—O	77.67	1.27	5.23	5.68	12.18
V	V—N	63.81	4.45	10.02	11.48	25.95
	V—O	69.67	1.94	8.86	10.00	20.80
	Nb-chelate	81.46	3.08	10.65	9.28	23.01
	Ta-chelate	85.09	3.58	10.95	9.32	23.85
Cr	Cr—N	62.04	4.85	8.70	8.82	22.37
	Cr—O	63.52	1.64	4.57	11.3	17.34
	Mo—N	65.83	4.80	8.42	8.95	22.16
	Mo—O	65.78	1.95	5.00	10.61	17.57
	W—N	70.96	4.79	8.57	7.07	20.44
	W—O	67.96	1.95	5.02	10.75	17.70
Mn	Mn—N	32.50	5.39	8.31	8.56	22.26
	Mn—O	39.38	2.25	6.60	8.85	17.70
	Tc—N	29.19	5.37	8.19	8.42	21.99
	Tc—O	32.82	2.83	6.84	8.61	18.28
	Re—N	24.5	5.24	8.17	8.43	21.84
	Re—O	28.6	4.56	7.3	8.61	20.50
Fe	Fe—N	46.20	5.18	8.26	8.76	22.20
	Fe—O	49.89	1.90	5.71	9.25	16.86
	Ru—N	53.20	5.14	8.30	8.66	22.10
	Ru—O	51.44	2.51	6.50	8.92	17.93
	Os—N	57.69	5.18	8.02	8.75	21.96
	Os—O	53.41	2.87	6.71	8.88	18.46
Co	Co—N	47.57	5.85	8.38	8.35	22.58
	Co—O	45.58	4.70	7.97	7.42	20.09
	Rh—N	54.70	5.64	8.36	8.19	22.19
	Rh—O	48.60	4.90	8.00	7.52	20.42
	Ir—N	62.30	5.61	8.34	8.52	22.47
	Ir—O	53.62	4.61	7.89	7.60	20.10
Ni	Ni—N	70.00	4.89	8.79	8.91	22.58
	Ni—O	69.66	3.41	7.62	9.84	20.87

	Pd—N	74.02	4.90	8.73	8.89	22.52
	Pd—O	65.83	3.28	6.92	9.14	19.34
	Pt—N	81.33	4.89	8.76	8.95	22.60
	Pt—O	74.58	1.52	6.84	9.45	17.81
Cu	Cu—O	20.78	4.41	7.95	7.42	20.78
	Ag—O	25.26	5.48	8.74	8.67	22.89
	Au—N	89.22	4.07	8.34	9.00	21.41
	Au—O	79.36	5.54	7.95	7.42	20.91
Zn	Zn—N	28.71	5.22	8.34	8.54	22.10
	Zn—O	34.97	2.76	6.59	8.82	18.17
	Cd—N	31.56	5.20	8.23	8.46	21.90
	Cd—O	38.66	3.12	6.90	8.61	18.63
	Hg—N	44.79	4.74	8.47	8.73	21.94
	Hg—O	43.76	3.66	8.77	9.49	21.92

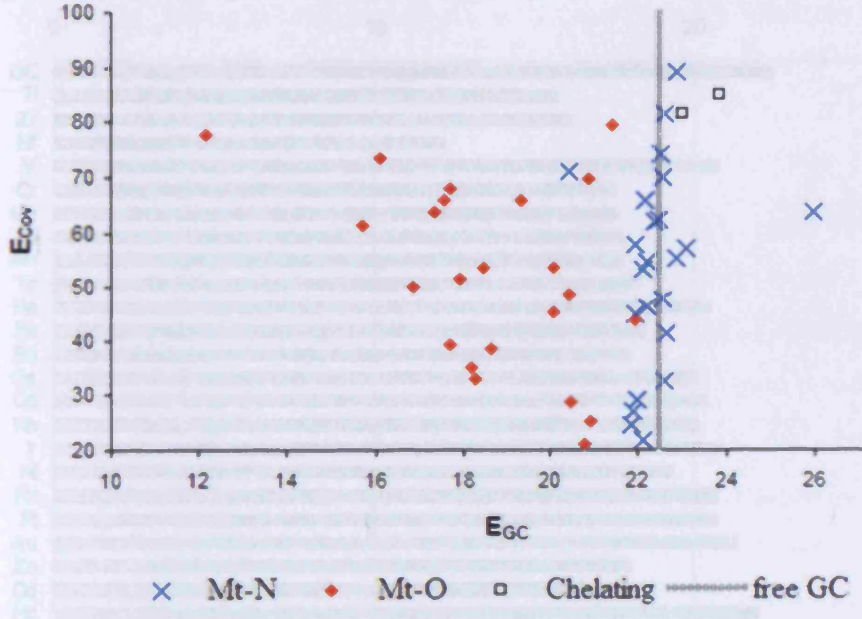


Figure 3.17 Plot of E_{GC} vs E_{cov} (kcal/mol).

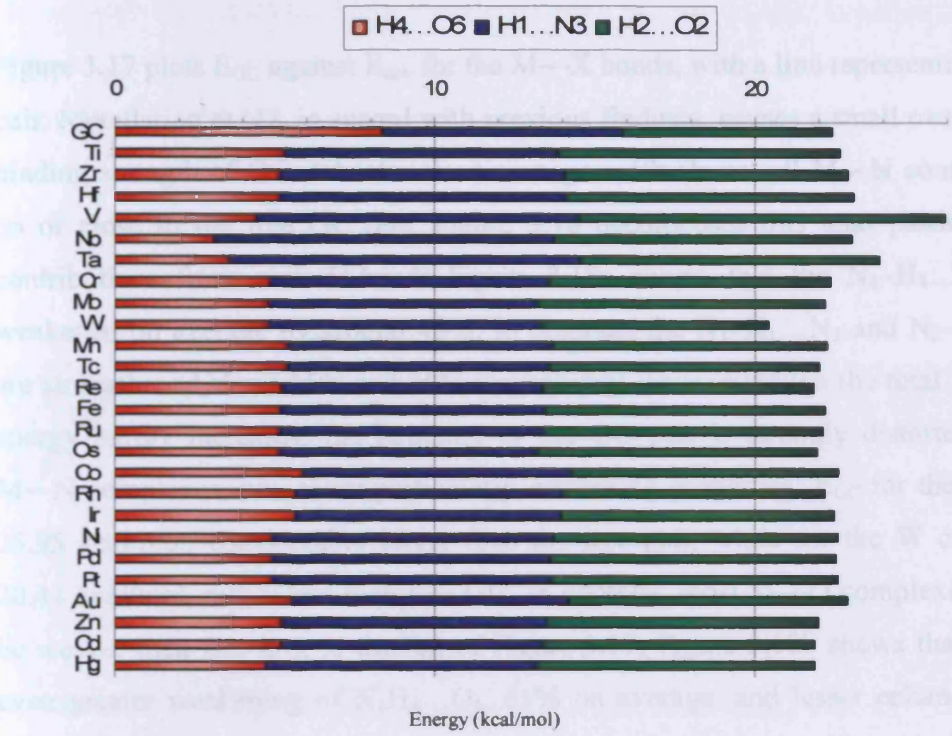


Figure 3.18a GC hydrogen bonds in the M-N_{GC} complexes.

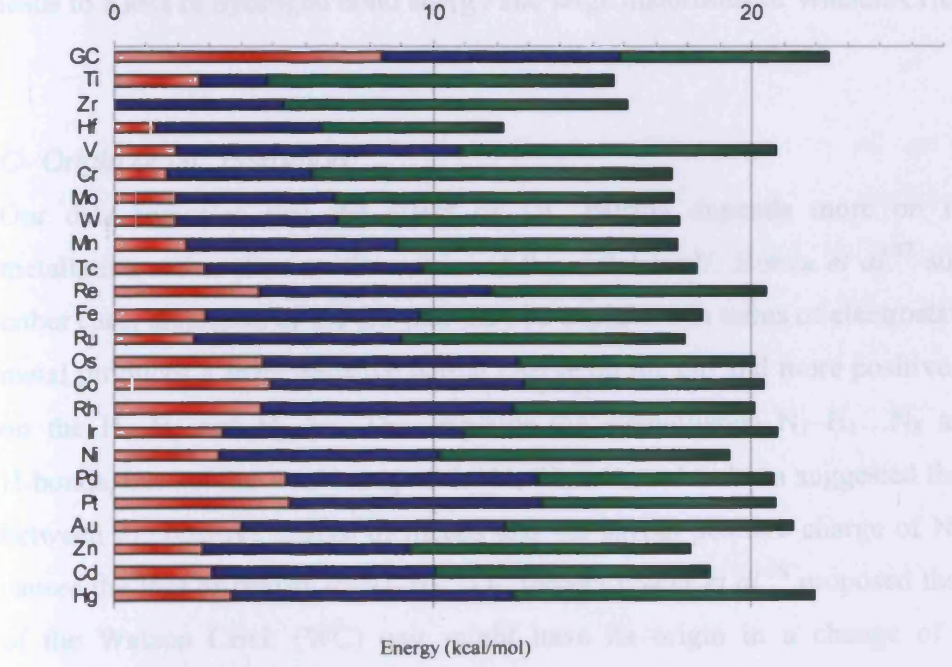


Figure 3.18b GC hydrogen bonds in the M-O_{GC} complexes.

Figure 3.17 plots E_{GC} against E_{cov} for the M—X bonds, with a line representing the free GC pair. Metallation at N7, in accord with previous findings, causes a small overall gain in the binding strength of GC, 1 kcal/mol on average, with almost all M—N complexes located on or close to the free GC line. Figure 3.18 decomposes this total pairing energy into contributions from each H-bond: Figure 3.18a shows that the N_4–H_4...O_6 bonds are weakened on average by around 40%. In contrast, the N_1–H_1...N_3 and N_2–H_2...O_2 bonds are strengthened about 15% and 30% respectively. Thus, although the total hydrogen bond energy hardly increases, the structure of the GC pair is strongly distorted. Among the M—N complexes, two show particularly interesting properties: E_{GC} for the V complex is 25.95 kcal/mol, considerably larger than the free pair, while for the W complex E_{GC} is 20.44 kcal/mol, rather less than free GC. In contrast, most M—O complexes are found to be weaker than free GC, to the left of Figure 3.17. Figure 3.18b shows that this is due to even greater weakening of N_4H_4...O_6, 65% on average, and lesser enhancement of N_2–H_2...O_2 when compared to metallation at N7. Complexation at O6 therefore uniformly leads to a loss of hydrogen bond energy and large distortions of Watson-Crick pairing.

C- Origin of GC distortion

Our data indicates that the effect on GC pairing depends more on the position of metallation rather than on the nature of the metal itself. Hobza *et al.*[57] suggested that in either case, distortion of the GC pair may be explained in terms of electrostatics. Firstly, the metal produces a more negative partial charge on the O6 and more positive partial charges on the H_1-N_1 and H_2-N_2. This explains the strengthened N_1-H_1...N_3 and N_2-H_2...O_2 H-bonds, but not the weakening of the N_4-H_4...O_6. It has been suggested that the repulsion between the positive charge of metals and the partial positive charge of N-H on cytosine causes the loss of energy in N_4-H_4...O_6, though Poater *et al.*[58] proposed that the distortion of the Watson Crick (WC) pair might have its origin in a change of donor/acceptor capability rather than electrostatic interactions.

To further investigate this, the molecular electrostatic potential (MEP) on the 0.001 au isodensity surface for some metal...guanine complexes were calculated (see Figure 3.19).

In general, the maximum MEP on this surface (V_{max}) is located between the two N-Hs of guanine, and is related to the strengthening of N_1–H_1…N_3 and N_2–H_2…O_2 H-bonds. For example, in the Ti—O complex, the sum of the energies of those H-bonds is 13.03 kcal/mol and V_{max} = 0.1876 au, while in the Pt—N complex these values are 17.71 kcal/mol and 0.1991 au, rising to 18.06 kcal/mol and 0.2176 au for V—O. Thus, these data support the electrostatic argument noted above.

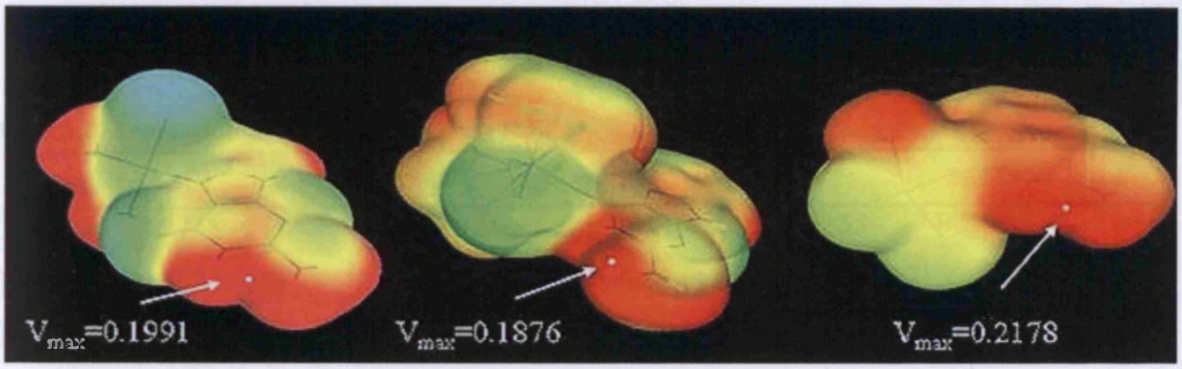


Figure 3.19 0.001 au isosurface electrostatic potential of Pt—N, Ti—O and V—O complexes.

AIM analysis (Table 3.17) reveals some interesting properties of the metal…guanine complexes, which may explain the differences of O and N metallation. As noted previously,[58] metallation affects all bonds of the six-membered ring of guanine, as well as C2-N2 and C6-O6 (see page 93 for numbering). In particular, C6-O6 bonds are longer and have reduced electron density than in isolated guanine, while all other bonds except the C4-C5 are strengthened by the presence of the metal. The latter is virtually unchanged in the N complexes, and slightly weakened in O complexes. The position of the metal does not affect C2-N2, but strengthening of the ring bonds and weakening of C6-O6 are more pronounced for O rather than N metallation. O metallation decreases C6-O6 ρ_{CP} by between 15-20% and increases N1-C6 and C5-C6 ρ_{CP} by between 10-15%, changes which are uniformly less than 5% for N metallation.

This suggests an explanation of the different effects of O and N metallation on GC pairing. Complexation at N7 only slightly affects the electron density of guanine near the

H-bonding position. Thus, distortion might be mainly due to the electrostatic effects reported above. In contrast, along with the electrostatic repulsion, O metallation induces large redistributions of density within the six-membered ring: the direct M—O interaction shifts density from the C6-O6 bond toward the metal, weakening this bond and reducing donor ability of O6. This further weakens $N_4H_4...O_6$, and leads to overall loss of stability of the GC pair. Interestingly, the V and Ti group elements, those most strongly bound to O6, promote the largest shifts in density and H-bond energy among the complexes considered.

Table 3.17 Topological analysis of guanine ring (au).[a]

	$\rho_{N1\text{-}C6}$	$\rho_{C6\text{-}O6}$	$\rho_{C5\text{-}C6}$	$\rho_{N3\text{-}C4}$	$\rho_{C2\text{-}N2}$
Guanine	0.2758	0.4174	0.2934	0.3255	0.3234
Ti—N	0.2924 (+6)	0.4082 (-2)	0.2956 (+1)	0.3481 (+7)	0.3366 (+4)
Ti—O	0.3106 (+12)	0.3504 (-16)	0.3259 (+11)	0.3444 (+5)	0.3386 (+5)
V—N	0.2936 (+6)	0.4090 (-2)	0.3062 (+4)	0.3418 (+5)	0.3379 (+5)
V—O	0.3107 (+13)	0.3497 (-16)	0.3257 (+11)	0.3456 (+6)	0.3395 (+5)
Fe—N	0.2913 (+5)	0.4029 (-4)	0.2986 (+2)	0.3362 (+3)	0.3335 (+3)
Fe—O	0.3041 (+10)	0.3760 (-10)	0.3100 (+6)	0.3361 (+3)	0.3330 (+3)

a: Values in parenthesis are percentage variations from free guanine.

D- Some interesting metals

From the data reported here, the early metals are perhaps the most interesting examples. Firstly, they show a clear preference for O over N complexation, most evident for the Ti group, where $\Delta_{N\text{-}O}$= -13 kcal/mol on average. Furthermore, the covalent energies are among the largest of the metals: Hf—O and V—O bonds are as strong as most platinum complexes reported in previous section. This is especially interesting, as M—O complexes appear to disrupt completely the Watson-Crick GC pair, with Ti, Zr and Hf reducing the GC pair binding by *ca.* 10 kcal/mol.

In addition, AIM analysis reveals an odd property of the Ta and Nb complexes. In both cases, as shown in Figure 3.15, a "chelating" complex results: while in the isolated guanine

complexes the bonding is asymmetric, addition of a cytosine leads to much more symmetrical complexes, with similar M—N and M—O electron densities and bond lengths (the ratio $\rho_{M\text{-}O}/\rho_{M\text{-}N}$ approaches unity). Therefore, as both bonds are relatively strong, this "chelating" mode strengthens the metal...guanine interaction. However, the effect of these complexes on GC pairing does not follow the pattern seen for chelating platinum complexes, with both Ta and Nb complexes having similar pairing energy to the free GC pair.

Although the mechanism is still unknown, several studies suggest Ru complexes to be potentially active against cancer. From Tables 3.15 and 3.16, Ru...guanine adducts are not particularly strong, nor indeed are any Fe group metals, with covalent energies of between 46 and 51 kcal/mol. Although data from Table 3.15 and 3.16 suggest a preference for N7, metallation on O6 is also intriguing: for instance, the Fe group causes some of the strongest effects on pairing. From Figure 3.17, the Fe group collects away from the GC line and in the upper region, close to the Ti group. Here, along with electrostatic and electron density shift reasons, further repulsion may originate from interaction between the NH groups of the ligand and cytosine, which are extremely close.

It has been shown experimentally that gold complexes are potentially active in treating cancer.[59,60] *In vitro* studies showed that Au-DNA adducts are weaker than corresponding Pt-DNA ones. In contrast, our data suggests that Au forms the most stable monofunctional adducts considered, with Au—N binding energy of 89.22 kcal/mol, around 8 kcal/mol larger for Pt. This apparent incongruence is resolved when we consider that, unlike the linear Au complexes, Pt-based drugs bind DNA in a bifunctional way, where the total binding energy is more than the double that of monofunctional complexes. Interestingly, Au hardly changes the GC pairing energy, being 21.41 and 20.91 kcal/mol for the Au—N and Au—O, respectively.

E- The effect of changing ligand

Table 3.18 reports properties of Ti complexes in which the ligand set was varied, allowing us to investigate the importance of ligands in metal...guanine (and GC) interactions. Along

with the metallocene sandwich (Cp_2) discussed above, $[Ti(NH_3)_2Cl_3]^+$ and $[Ti{=}O(NH_3)_2Cl]^+$ were considered, *i.e.* keeping the mono-charged d^0 nature of the complexes constant. Table 3.18 indicates that, independent of the ligands, Ti complexes prefer the O6 position of guanine. The $\Delta_{N\text{-}O}$ is on average equal to −10.7 kcal/mol, although this trend is reversed for $[Ti{=}O(NH_3)_2Cl]^+$ complexes, where $\Delta_{N\text{-}O}$ is slightly positive, a surprising finding that may be resolved by AIM analysis (see Figure 3.20). In the Ti—N7 complex to guanine, an extra bond CP is present, representing a weak Ti...O interaction (distance = 2.926Å, electron density = 0.011 au). In contrast, the Ti—O6 contains an N—H...N H-bond, in common with all other complexes considered, including the Ti—N_{GC} complex in which the secondary Ti...O interaction is no longer present (distance = 3.481Å). Thus, the $[Ti{=}O(NH_3)_2Cl]^+$ complex presents some peculiar bonding modes, and analysis of the topology of the electron density allows explanation of its apparently anomalous behaviour.

Table 3.18 Effect of ligand variation on Ti...G and Ti...GC complex.

	Ligand	Bonding	Binding Energy	E_{intra}	E_{GC}	E_{cov}
Ti—G	Cp_2Cl	Ti—N7	36.40	3.17		33.20
	Cp_2Cl	Ti—O6	51.88	3.45		48.43
	$(NH_3)_2Cl_3$	Ti—N7	48.51	8.33		40.18
	$(NH_3)_2Cl_3$	Ti—O6	55.01	2.91		52.10
	$=O(NH_3)_2Cl$	Ti—N7	59.35	0.0		59.35
	$=O(NH_3)_2Cl$	Ti—O6	60.16	3.65		56.51
Ti—GC	Cp_2Cl	Ti—N7	67.34	3.33	22.67	41.33
	Cp_2Cl	Ti—O6	78.33	1.53	15.70	61.10
	$(NH_3)_2Cl_3$	Ti—N7	78.62	8.71	22.92	47.00
	$(NH_3)_2Cl_3$	Ti—O6	87.91	8.70	19.51	58.70
				1.00[a]		
	$(NH_3)_2ClO$	Ti—N7	91.31	4.09	22.66	64.56
	$(NH_3)_2ClO$	Ti—O6	95.49	1.90	21.23	72.36

a: this energy corresponds to cytosine-NH...Cl-Metal interaction.

H-bonding results in Table 3.18 include both intramolecular and GC results. The presence of the polar hydrogens of ammine ligands increases the strength of the former, especially in the N7 complexes. Similarly to other N7 complexes, individual H-bonds are distorted but the overall pairing energy is conserved. Ti—O6 complexation reduces the pairing energy, with Δ_{GC} (the difference between free and metallated GC) varying between 1 and 7 kcal/mol. These data indicate that the ligand modulates the behaviour of the metal to some extent, but ultimately all Ti complexes reported in Table 3.18 behave qualitatively in a similar manner.

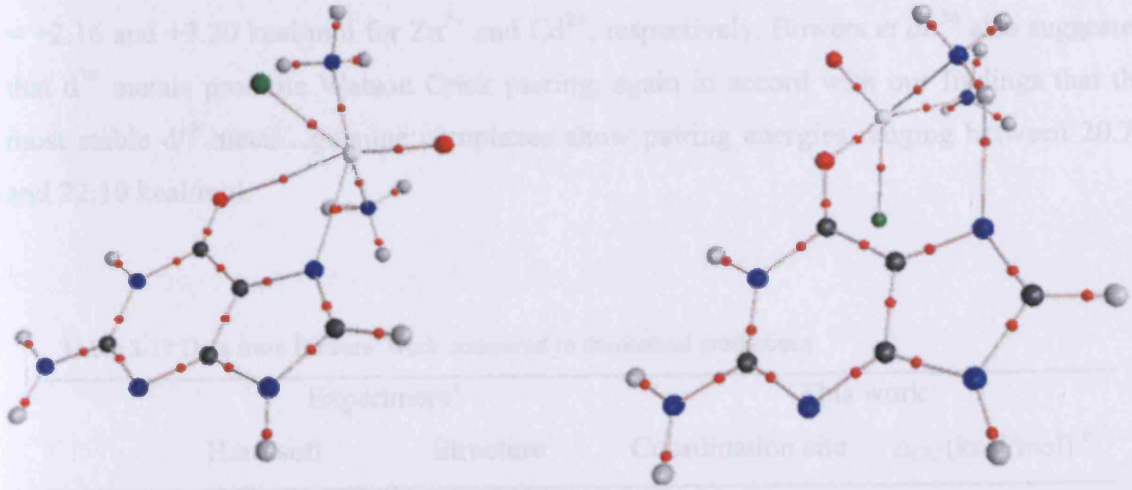


Figure 3.20 Molecular graphs of $[Ti{=}O(NH_3)_2Cl]$—N7 and $[Ti{=}O(NH_3)_2Cl]$—O6 complexes.

F- Comparison with literature

Bowers and co-workers[56] recently analysed gas-phase binding properties of metals to the dinucleotide duplex dCG·dCG using ion mobility mass spectrometry. They showed that metals classified as hard acids led to globular structures with disruption of the GC pair, whereas soft metals stabilised the Watson-Crick structure. Table 3.19 compares our data with the results reported by Bowers. It should be noted that the experimental data are for bare M^{n+}, *i.e.* without any ligands, in contrast with our calculations, thus any comparison can only be qualitative. Nevertheless, we find broad agreement between the two

approaches, with all the cases in which a globular structure is found showing substantial disruption of the GC pair ($\Delta_{GC} << 0$), while those found to maintain pairing show either little or no disruption.

This agreement is not perfect, however: our results suggest complexation of Zn^{2+} and Cd^{2+} at O6, and hence substantial disruption. We therefore also performed optimisation of complexes of *naked* Zn^{2+} and Cd^{2+} complexes to the Watson-Crick GC pair. Both metals are doubly charged and adopt a chelating conformation, binding to both O6 and N7. Data obtained for these chelating structures agree much better with experiment: in both cases AIM analysis confirms that metallation leads to a stabilisation of the GC pairing, with Δ_{GC} = +2.16 and +3.20 kcal/mol for Zn^{2+} and Cd^{2+}, respectively. Bowers *et al.*[56] also suggested that d^{10} metals promote Watson Crick pairing, again in accord with our findings that the most stable d^{10} metal...guanine complexes show pairing energies ranging between 20.78 and 22.10 kcal/mol.

Table 3.19 Data from Bowers' work compared to theoretical predictions.

	Experiment[a]		This work	
	Hard/soft	Structure	Coordination site	Δ_{GC} (kcal/mol) [c]
Cu^{+}	Soft	WC	O6	-1.70
Ag^{+}	Soft	WC	O6	+0.47
Cd^{2+}	Soft	WC	O6	-3.84
			N7/O6[b]	+2.16[b]
Pt^{2+}	Soft	WC	N7	+0.13
Zn^{2+}	Soft	WC	O6	-4.3
			N7/O6[b]	+3.20[b]
Fe^{2+}	Borderline	Globular	O6	-5.61
Ni^{2+}	Borderline	Globular	N7/O6	-1.8
Cr^{2+}	Hard	Globular	O6	-5.13
Mn^{2+}	Hard	Globular	O6	-4.77

a: from ref. 56; b: the calculations are referred to metal...guanine complexes with no ligand on the metal; c: variation in GC pairing energy in complex from free GC.

Finally, it is well known that the electron density at the bond CP is related to bond strength, at least for series of similar bonds,[61] including our results for Pt—N and Pt—O bonds. Here we are able to extend that outcome to a much more diverse set of complexes, a total of 108 data for 27 metals. We find that the close linear relation established for Pt—N and Pt—O bonds is rather less apparent for this data, with ρ_{CP} for M—X bonds in metal...guanine and metal....GC complexes of all the transition metals correlates fairly well to E_{cov}, with $r^2 = 0.80$.

3.5 Concluding remarks

3.5.1 Solvation and hydrolysis of cisplatin

We have shown, by means of density functional calculations and analysis of resulting electron densities and electrostatic potentials, that cisplatin is capable of forming strong hydrogen bonds as both a donor and acceptor. This is best exemplified by our estimates of its position on Abraham's acidity and basicity scales, wherein values of $\boldsymbol{A}$ = 0.70 and $\boldsymbol{B}$ = 0.84 indicate cisplatin is comparable to the strongest monofunctional organic acids and bases. Interactions with water bring out this amphiphilic nature, as the most stable 1:1 complex contains a N—H...O—H...Cl bridge between cisplatin's acidic and basic ends. Such motifs are preserved as further molecules are added to form small clusters, though water...water interactions soon come to dominate the overall stabilisation of such clusters. These interactions are enhanced by the polarization of water molecules due to their proximity and H-bonding to cisplatin, such that most interactions are notably stronger than in the isolated water dimer. The effect of this explicit solvation on the hydrolysis of cisplatin, a key step in its activation, is also explored. Large differences in the geometry and position of the trigonal bipyramidal transition state are found between solvated and gas phase, but the energetics of the reaction are less affected, with only a 1 kcal/mol difference in activation energy.

3.5.2 Cisplatin (and transplatin) complexes with DNA bases

Our results on cisplatin and transplatin DNA bases complexes have shown that calculations at the B3LYP/DGDZVP(SDD)//HF/6-31G(d,p)(SDD) level are capable of reproducing literature (where available) or higher-level theoretical geometries and binding energies of complexes of cisplatin with purine bases. Also, we have shown that electron density properties and partial least-squares analysis can be used to form an accurate, family-independent model of H-bond strength, which can then be used to decompose the total binding energy of cisplatin-purine complexes into covalent and H-bond contributions. In particular, hydrogen bonds are ubiquitous in such complexes, with N—H...O, N—H...N, and N—H...Cl contacts all observed. Thus, the known preference of cisplatin for the N7 position of guanine cannot be explained on the basis of H-bonding alone, even though this complex contains one of the strongest H-bonds found.

Interestingly, complexes of cisplatin with two purine bases follow the same general trend of stability as do single base complexes, although trans complexes are generally more stable than their cis counterparts. An even greater variety of H-bond motifs is present in such bifunctional complexes, with purine···purine contacts dominating structures containing adenine in particular.

Platination at the N7 position of guanine has a dramatic effect on the hydrogen bonds involved in pairing to cytosine, weakening N_4—H_4...O_6 but strengthening N_1—H_1...N_3 and N_2—H_2...O_2, leading to large changes in the geometry of the GC pair, but only small differences in the total binding energy. Platination at O6 or chelation to N7 and O6, on the other hand, destroys the normal Watson-Crick pattern of H-bonding, though substantial pairing energy remains.

3.5.3 Systematic studies of transition metals and DNA bases complexes

Analysis of more than one hundred metal...guanine(cytosine) complexes leads to interesting conclusions. Firstly, as shown for cisplatin complexes, calculations at B3LYP/DGDZVP(SDD)//HF/6-31G*(SDD) are capable of reproducing higher-level

binding energies of transition metal purine complexes. In these complexes, hydrogen bonds such as N—H...O, N—H...N, N—H...Cl, C—H...O and C—H...N contribute *ca.* 10% of overall stabilisation. The analysis of covalent bond shows that relative stability of M—O6 and M—N7 depends more on the nature of the metal than the choice of ligand: early metals prefer O6, where more electron rich metals prefer N7. Also, across rows relative stability decreases as M—N7 energies become larger than M—O6. Interestingly, the effect of metallation on the GC pair is similar to that observed in platinum complexes, *i.e.* N_4—H_4...O_6 is weakened, while N_1—H_1...N_3 and N_2—H_2...O_2 are enhanced, leading to large changes in the geometry of the GC pair. The overall pairing energy of GC is unchanged for the M—N complexes, but significantly reduced for M—O. Among the transition metals, titanium and vanadium group metals show particularly large covalent bond energies and strongly affect the GC pair, where the energy falls more than 10 kcal/mol. The analysis of topology and charges suggests that the effect of metallation at N7 on GC pairing can be explained in terms of electrostatics, but metallation at O6 also induces large redistributions of electron density within guanine, leading to a loss of H-bond donor capability of O6. Finally, the electron density at the metal...guanine bond correlates with the covalent energy, though without the accuracy found previously for Pt complexes alone.

3.6 References

(1) Aprile, F.; Martin, D. S. *Inorg. Chem.* **1962**, *1*, 551-557.

(2) Reishus, J. W.; Martin, D. S. *J. Am. Chem. Soc.* **1961**, *83*, 2457-2462.

(3) Kozelka, J.; Bergès, J.; Attias, R.; L, F. *Angew. Chem. Int. Ed.* **2000**, *39*, 198-201.

(4) Spingler, B.; Whittington, D. A.; Lippard, S. J. *Inorg. Chem.* **2001**, *40*, 5596-5602.

(5) Baik, M. H.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 4495-4503.

(6) Deubel, D. V. *J. Am. Chem. Soc.* **2002**, *124*, 5834-5842.

(7) Burda, J. V.; Sponer, J.; Hrabakova, J.; Zeizinger, M.; Leszczynski, J. *J. Phys. Chem. B* **2003**, *107*, 5349-5356.

(8) Baik, M. H.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2003**, *125*, 14082-14092.

(9) Platts, J. A.; Hibbs, D. E.; Hambley, T. W.; Hall, M. D. *J. Med. Chem.* **2001**, *44*, 472-474.

(10) Screnci, D.; McKeage, M. J.; Galettis, P.; Hambley, T. W.; Palmer, B. D.; Baguley, B. C. *Br. J. Cancer* **2000**, *82*, 966-972.

(11) Eastman, A.; Barry, M. A. *Biochemistry* **1987**, *26*, 3303-3307.

(12) Frisch, M. J. T., 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.; 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.; and Pople, J. A. *Gaussian, Inc., Pittsburgh PA* **2003**.

(13) Wysokinski, R.; Michalska, D. *J. Comput. Chem.* **2001**, *22*, 901-912.

(14) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664-675.

(15) Andrae, D.; Haussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123-141.

(16) Lamarche, O.; Platts, J. A.; Hersey, A. *J. Chem. Inf. Comp. Sci.* **2004**, *44*, 848-855.

(17) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-1377.

(18) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789.

(19) Hehre, W. J.; Ditchfie.R; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257-&.

(20) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265-3269.

(21) Harihara.Pc; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213-222.

(22) Ditchfie.R; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724-&.

(23) Politzer, P.; Truhlar, D. G. *Chemical Applications of Atomic and Molecular Electrostatic Potentials*: New York, 1981.

(24) Lamarche, O.; Platts, J. A. *Phys. Chem. Chem. Phys.* **2003**, *5*, 677-684.

(25) Carloni, P.; Sprik, M.; Andreoni, W. *J. Phys. Chem. B* **2000**, *104*, 823-835.

(26) Pavankumar, P. N. V.; Seetharamulu, P.; Yao, S.; Saxe, J. D.; Reddy, D. G.; Hausheer, F. H. *J. Comput. Chem.* **1999**, *20*, 365-382.

(27) Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond*; Oxford University Press: Oxford, 1999.

(28) Boys, S. F.; Bernardi, F. *Mol. Phys* **1970**, *19*, 553.

(29) Grabowski, S. J. *Chem. Phys. Lett.* **2001**, *338*, 361-366.

(30) Basolo, F.; Pearson, R. G. *Mechanism of Inorganic Reactions*; Wiley: New York, 1967.

(31) Bose, R. N.; Viola, R. E.; Cornelius, R. D. *J. Am. Chem. Soc* **1984**, *106*, 3336-3343.

(32) Zhang, Y.; Guo, Z.; You, X. Z. *J. Am. Chem. Soc.* **2001**, *123*, 9378-9387.

(33) Luiz Antonio Sodre, C.; Willian, R. R.; Wagner, B. D. A.; Helio, F. D. S. *J. Chem. Phys.* **2003**, *118*, 10584-10592.

(34) Sosa, C.; Andzelm, J.; Elkin, B. C.; Wimmer, E.; Dobbs, K. D.; Dixon, D. A. *J. Phys. Chem.* **1992**, *96*, 6630-6636.

(35) Richardson, N. A.; Wesolowski, S. S.; Schaefer, H. F. *J. Am. Chem. Soc.* **2002**, *124*, 10163-10170.

(36) Zeizinger, M.; Burda, J. V.; Leszczynski, J. *Phys. Chem. Chem. Phys.* **2004**, *6*, 3585-3590.

(37) Mansy, S.; Chu, G. Y. H.; Duncan, R. E.; Tobias, R. S. *J. Am. Chem. Soc* **1978**, *100*, 607-616.

(38) Yang, X. L.; Wang, A. H. J. *Pharmacol. Ther.* **1999**, *83*, 181-215.

(39) Cotton, A. F.; Wilkinson, G.; Gaus, P. L. *Basic Inorganic Chemistry*; John Wiley & Sons, Inc: New York, 1995.

(40) Baik, M. H.; Friesner, R. A.; Lippard, S. J. *Inorg. Chem.* **2003**, *42*, 8615-8617.

(41) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247-3276.

(42) Burda, J. V.; Sponer, J.; Leszczynski, J. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4404-4411.

(43) Fichtinger-Schepman, A. M. J.; Vanderveer, J. L.; Denhartog, J. H. J.; Lohman, P. H. M.; Reedijk, J. *Biochemistry* **1985**, *24*, 707-713.

(44) Sherman, S. E.; Gibson, D.; Wang, A. H. J.; Lippard, S. J. *Science* **1985**, *230*, 412-417.

(45) Pelmenschikov, A.; Zilberberg, I.; Leszczynski, J.; Famulari, A.; Sironi, M.; Raimondi, M. *Chem. Phys. Lett.* **1999**, *314*, 496-500.

(46) Zilberberg, I. L.; Avdeev, V. I.; Zhidomirov, G. M. *J. Mol. Struct.: THEOCHEM* **1997**, *418*, 73-81.

(47) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem.-Eur. J.* **1999**, *5*, 3581-3594.

(48) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737-738.

(49) Popelier, P. L. A.; Joubert, L. *J. Am. Chem. Soc* **2002**, *124*, 8725-8729.

(50) Joubert, L.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4353-4359.

(51) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149-1170.

(52) Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608-15613.

(53) Sigel, R. K. O.; Freisinger, E.; Lippert, B. *J. Biol. Inorg. Chem.* **2000**, *5*, 287-299.

(54) Ghosh, P.; D'Cruz, O. J.; Narla, R. K.; Uckun, F. M. *Clin. Cancer Res.* **2000**, *6*, 1536-1545.

(55) Robertazzi, A.; Platts, J. A. *Inorg. Chem.* **2005**, *44*, 267-274.

(56) Baker, E. S.; Manard, M. J.; Gidden, J.; Bowers, M. T. *J. Phys. Chem. B* **2005**, *109*, 4808-4810.

(57) Burda, J. V.; Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1997**, *101*, 9670-9677.

(58) Poater, J.; Sodupe, M.; Bertran, J.; Sola, M. *Mol. Phys.* **2005**, *103*, 163-173.

(59) Bernersprice, 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.

(60) 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.

(61) Howard, S. T.; Lamarche, O. *J. Phys. Org. Chem.* **2003**, *16*, 133-141.

4 DNA and cisplatin-DNA structures

H-bonding and π-stacking

4.1 Preface

In this chapter we focus on the fundamental forces, *i.e.* π-stacking and H-bonding that stabilise free and platinated DNA structures. A new approach based on DFT is proposed in order to describe π-stacking, and AIM methods are extensively employed to analyse intermolecular forces in such structures. These studies cover a large variety of examples, from simple benzene dimers up to large systems such as free and platinated DNA oligonucleotides.

4.2 Hybrid HF/DFT for π-stacking interactions

As reported in section 1.3, high level calculation, such as MP2 and CCSD are needed to describe properly π-stack interactions. However the cost of these calculations becomes rapidly prohibitive when considering large systems. Therefore strong interest in the last years has been shown towards cheaper methods, such as DFT (see chapter 1 and 2).

4.2.1 Calculation method

Throughout this work, we have made extensive use of Becke's "half-and-half" functional, BH&H.[1] This is an *ad hoc* mixture of exact (HF) and local density approximation exchange, coupled with Lee, Yang, and Parr's expression[2] for the correlation energy (see section 2.4.2 and equation 2.1.14). Our interest in this particular functional was initially sparked by Pérez-Jordá and Becke's observation[3] that it reproduces the geometry and binding energy of rare gas dimers with reasonable, albeit not quantitative, accuracy. BH&H was used with a range of basis sets, including Pople's 6-31G[4] and 6-311G[5] families with varying levels of polarisation and diffuse functions, and Dunning's correlation consistent cc-pV*n*Z family,[6,7] with and without diffuse functions. The performance of BH&H was also compared to a wide range of more commonly used functionals, most notably Becke's 3-parameter exchange functional B3LYP, the recently reported X3LYP,[8] as well as MP2 results with a range of basis sets including Hobza's 6-31G(0.25)*.[9,10] All binding energies reported have been corrected for BSSE using the counterpoise method of Boys and Bernardi.[11] Throughout this work, all dimer geometries were fully optimised using BH&H without symmetry constraint, unless otherwise stated. All polarisabilities were calculated in Gaussian03 as energy derivatives at the BH&H/6-311++G(d,p) level.

4.2.2 The BH&H functional applied to benzene stacked dimers

Counterpoise-corrected potential energy surfaces for the "parallel-displaced" (C_{2h}) benzene dimer using a number of methods are shown in Figure 4.1 (compare section 1.4). The rigid potential energy scans were generated by varying the interplane distance with the "slip" coordinate fixed at the value found by Sherrill and co-workers (see ref.12 for definitions of these coordinates). Monomer geometries were fully optimised using BH&H/6-311++G(d,p), then frozen during scans. As expected, the Hartree-Fock PES is repulsive at all separations, as is that from the B3LYP functional, while the MP2 PES is attractive at separations greater than 3Å, with greatest stabilisation of 2.97 kcal mol^{-1} at 3.40Å.

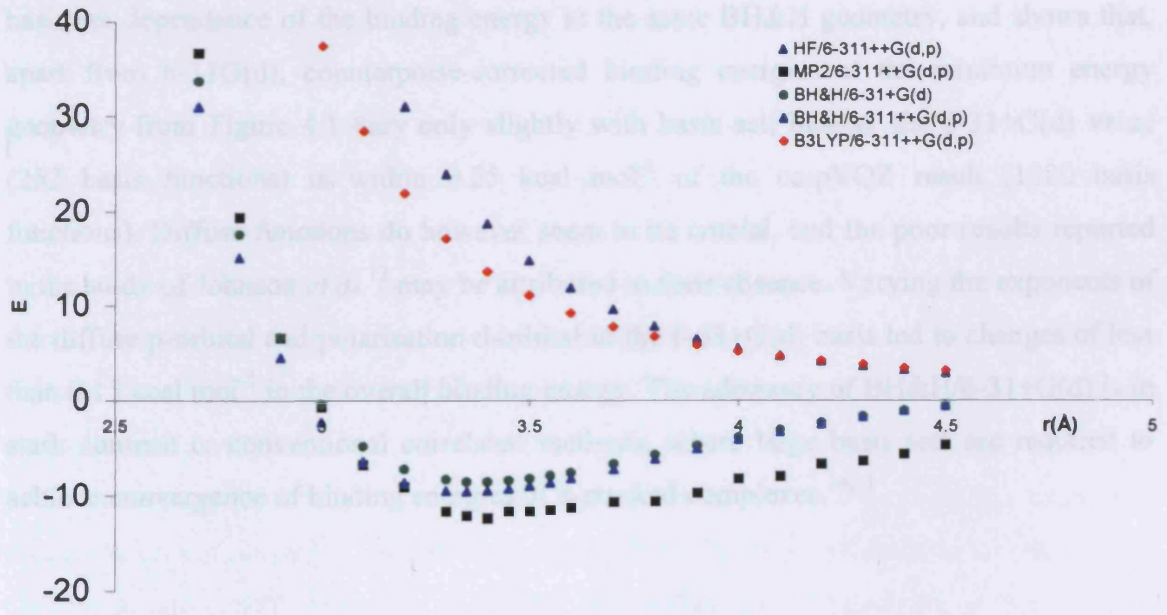


Figure 4.1 Counterpoise-corrected relative energies (kcal/mol) for parallel-displaced (C_{2h}) benzene dimer.

Remarkably, the BH&H/6-311++G(d,p) PES qualitatively reproduces the MP2 results, giving overall stabilisation at separations greater than 3 Å, and a minimum at 3.35 Å with a binding energy of 2.31 kcal mol^{-1}. Perhaps unsurprisingly, given the lack of long-range electron correlation, the BH&H binding energy decays to zero rather more quickly than MP2, but around the equilibrium separation we observe close agreement between the methods. Thus, our BH&H and MP2 estimates of binding energy with this basis set bracket Sherrill *et al.*'s value of 2.78 kcal mol^{-1} for the CCSD(T) basis set limit. As no charge transfer is possible in this symmetrical system, and HF fails to recover any binding, implying the importance of dispersion, this suggests that BH&H contains a fortuitous cancellation in the exchange-correlation functional that mimics the dispersive part of methods such as MP2.

Moreover, the BH&H potential energy surface is very similar using the smaller 6-31+G(d) basis set (maximum binding energy 2.04 kcal mol^{-1} at 3.35Å). Table 4.1 further explores basis set dependence of the binding energy at the same BH&H geometry, and shows that, apart from 6-31G(d), counterpoise-corrected binding energies at the minimum energy geometry from Figure 4.1 vary only slightly with basis set. Indeed, the 6-31+G(d) value (252 basis functions) is within 0.25 kcal mol^{-1} of the cc-pVQZ result (1020 basis functions). Diffuse functions do however seem to be crucial, and the poor results reported in the study of Johnson *et al.*[13] may be attributed to their absence. Varying the exponents of the diffuse p-orbital and polarisation d-orbital in the 6-31+G(d) basis led to changes of less than 0.15 kcal mol^{-1} in the overall binding energy. The adequacy of BH&H/6-31+G(d) is in stark contrast to conventional correlated methods, where large basis sets are required to achieve convergence of binding energies of π-stacked complexes.[14,15]

Table 4.1 Variation of BH&H binding energy with basis set.

Basis Set	Binding energy (kcal mol^{-1})
6-31G(d)	1.19
6-31+G(d)	2.04
6-311++G(d,p)	2.31
6-311++G(2df,2p)	2.29
cc-pVDZ	1.92
aug-cc-pVDZ	2.28
cc-pVTZ	2.14
aug-cc-pVTZ	2.15
cc-pVQZ	2.17

Table 4.2 compares the performance of a range of functionals at the same BH&H geometry, and demonstrates that all gradient-corrected functionals considered completely fail to recover the positive binding. By contrast, functionals based solely on LDA give a positive binding energy. The BH&H functional contains exactly the same correlation functional as SLYP, which considerably overestimates binding. This suggests that the HF exchange reduces the extent of overestimation from the LDA exchange functional, thereby causing the happy accident of simulating the higher-level results. In this respect, BH&H is similar to the recently proposed X3LYP functional,[8] which also combines two exchange functionals that alone do not properly describe dispersion forces to achieve satisfactory binding of van der Waals complexes. However, Table 4.2 shows that X3LYP does not predict positive binding energy for the benzene dimer at this geometry, perhaps unsurprisingly in the light of Hobza's recent study of stacking on DNA base pairs.[16]

Table 4.2 Binding energies using various functionals with the 6-31+G(d) basis set.

Method	Binding energy (kcal mol^{-1})
B3LYP	-3.84
BLYP	-4.89
PBE	-1.82
HCTH/407	-3.75
PW91	-1.51
mPW1PW91	-2.80
X3LYP	-3.29
LDA	2.47
SLYP	5.50
Hartree-Fock	-5.26

A- BH&H compared with post-HF methods

We then investigated whether BH&H can adequately model the subtle variations in electrostatic and dispersion forces that result from substitution of benzene. Following Sinnokrot and Sherrill,[15] four substituents with differing electron donating or withdrawing character were added, and the resulting stacked complex fully optimised using BH&H/6-311++G(d,p): Table 4.3 compares binding energies with CCSD(T)/aug-cc-pVTZ calculations. In general agreement is excellent: the average difference is just 0.20 kcal mol^{-1} in binding energy, and 0.06Å in inter-plane distance (not reported). The largest discrepancy is found for cyanobenzene, for which errors are less than 0.4 kcal mol^{-1} and 0.1 Å. In all four cases, BH&H performs slightly better than MP2/6-31G(0.25)*, which tends to overestimate the CCSD(T) values. In addition, both BH&H and MP2/6-31G(0.25)* perform much better than the MP2/aug-cc-pVDZ calculations reported in ref. 15, where binding energy is overestimated on average by of 1.26 kcal mol^{-1}. Sinnokrot and Sherrill ascribed the trends in binding energy along this series to the subtle interplay of electrostatic, dispersion, and exchange forces. It seems remarkable that an approximate method as BH&H is able to capture this trend, lending further validity to our use of this functional.

Table 4.3 BH&H, MP2 and CCSD(T) binding energies of stacked benzene, pyridine and DNA base complexes.

	Binding Energy (kcal mol^{-1})		
	BH&H/ 6-311++G(d,p)	CCSD(T)/ basis limit	MP2/ 6-31G(0.25)*
Benzene dimer	2.31	2.78[a]	2.46[b]
Benzene-toluene	2.22	2.27[c]	2.44[b]
Benzene-phenol	2.31	2.17[c]	2.40[b]
Benzene-fluorobenzene	2.56	2.29[c]	2.65[b]
Benzene-cyanobenzene	2.74	3.05[c]	3.38[b]
Pyridine-benzene	2.88		2.78[d]
Pyridine-toluene	2.83		3.34[d]
Pyridine-phenol	2.92		2.69[d]
Pyridine-aniline	3.78		3.20[d]
Pyridine-chlorobenzene	3.21		3.35[d]
Pyridine-nitrobenzene	3.47		3.79[d]
Pyridine-formylbenzene	3.05		3.49[d]
Pyridine-fluorobenzene	3.06		2.89[d]
Pyridine-benzoic acid	3.19		3.49[d]
Pyridine-cyanobenzene	3.46		4.13[d]
Pyrimidine dimer	2.86	3.50[e]	3.14[b]
UU(*p*)	10.64	9.70[f]	7.90[b]
UU(*ap*)	9.33	8.80[f]	8.11[b]
CC	10.61	10.40[e]	8.37[b]
TT	9.79		6.87[b]
GG	13.77	12.90[e]	12.49[b]
AA	7.28		8.06[b]

a: Ref 14; b: This work, at BH&H optimised geometry; c: ref. 15; d: ref. 17; e: ref. 18; f: ref. 19.

Table 4.3 includes analogous data for 10 pyridine/benzene complexes, previously studied by Mignon *et al.*[17] at the MP2/6-31G(0.25)* level. Once again, good agreement is obtained with BH&H/6-311++G(d,p), with an average discrepancy of 0.35 kcal mol^{-1}, and a maximum difference of 0.67 kcal mol^{-1} for the pyridine/cyanobenzene complex. Thus, BH&H captures the interactions between these polarised species to similar accuracy as for the benzene dimer.

A more realistic test of this DFT method lies in its ability to model the stacking interactions between DNA and RNA bases. Therefore, we have included six such dimers in Table 4.3: reference post-HF binding energies are reported at the CCSD(T) or MP2 level for co-planar dimers at the monomer MP2 geometry: following a similar approach with BH&H generally leads to excellent agreement. For the anti-parallel pyrimidine dimer, following Hobza and Sponer[9] we use rigid monomers and an inter-plane separation of 3.30Å within the C_i point group. BH&H calculations yield counterpoise corrected binding energies of 2.86 and 2.80 kcal mol^{-1}, using 6-311++G(d,p) and 6-31+G(d), respectively. This compares well with the best estimate of 3.50 kcal mol^{-1}, obtained via extrapolation of MP2 energies to the basis set limit and CCSD(T) correction.

Nielsen and co-workers recently reported an analysis of the uracil dimer (UU) PES at the MP2 level,[19] finding two minima for the parallel (*p*) and anti-parallel (*ap*) isomers. CCSD(T) binding energies for these isomers were then calculated to be 9.70 and 8.80 kcal mol^{-1}, respectively. Our BH&H calculations are able to reproduce both minima with essentially identical geometries, while slightly over-estimating the binding energy of each, with errors of *ca.* 1 and 0.5 kcal mol^{-1}, respectively. Not only are the binding energies of UU reproduced with small errors, but BH&H also correctly predicts the relative stability of the two orientations, unlike MP2. Even greater accuracy results for the anti-parallel cytosine dimer (CC), for which the BH&H binding energy is within 0.25 kcal mol^{-1} of Hobza's CCSD(T) value.

For purine bases, the difference between DFT and post-HF energies is slightly larger but still acceptable: the guanine dimer (GG) result is within 1 kcal mol^{-1} of Hobza's CCSD(T) value, again from rigid PES scans at the monomer geometry. To the best of our knowledge,

no analogous calculation has been reported for the adenine dimer (AA), so the only comparable binding energy is at MP2 level, which again is reasonably well-reproduced by our DFT calculations, though unlike most entries for DNA bases in Table 4.3, the BH&H value is actually slightly smaller than the MP2/6-31G(0.25)* result.

Table 4.3 contains 10 comparisons of BH&H with CCSD(T) binding energies, and 22 with MP2/6-31G(0.25)* values, thereby allowing us to quantify the performance of this method. From this data, the average absolute error (*aae*) in BH&H binding energies is just 0.45 kcal mol^{-1} when referenced against CCSD(T) values and 0.63 kcal mol^{-1} against MP2, with a maximum deviation from CCSD(T) of 0.94 kcal mol^{-1}. By comparison, the MP2 has an *aae* from CCSD(T) of 0.67 kcal mol^{-1}. That such accuracy can result from the drastic cancellation of errors discussed above is remarkable. However, we note that the success of MP2/6-31G(0.25)* is also due to some extent to error cancellation: here, overestimation of binding due to incomplete treatment of correlation is balanced by underestimation due to the small basis set. Such tests are necessarily limited to complexes that may be tackled using post-SCF methods: the purine dimers GG and AA are at the limit of current technology. Therefore, we have tested the performance of BH&H for modelling π-stacking as well as is currently possible, and found it to be successful in all cases.

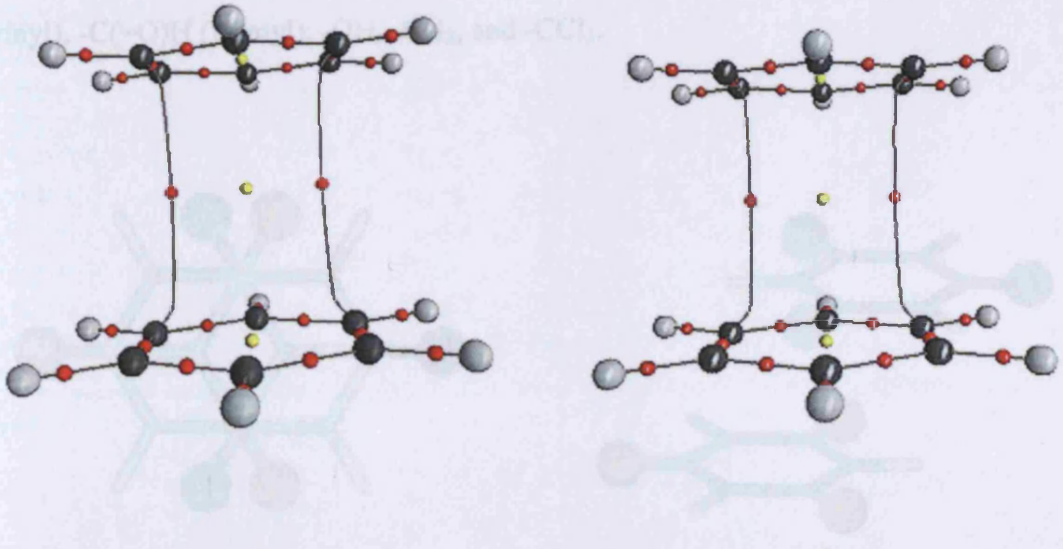
B- Advantages of using BH&H

Use of a DFT method to describe π-stacking interactions has a number of advantages over conventional post-HF methods, not least of these being speed of calculation. As a typical example, a counterpoise calculation on the thymine dimer using BH&H/6-311++G(d,p) required slightly less than three hours, compared with over four for an analogous MP2/6-31G(0.25)* calculation. Superficially, this may not seem a great speed-up, but the formal scaling of DFT as N^3 compared with N^5 for MP2 means this should become more evident for larger systems. Moreover, the lesser basis set requirements of a single-determinant Kohn-Sham DFT calculation mean that BSSE in the former calculation is only 1.75 kcal mol^{-1}, much less than the value of 10.80 kcal mol^{-1} for the latter. As it is a laborious process to correct for BSSE at each step of a geometry optimisation, it should therefore prove feasible to go beyond the frozen monomer approach widely used in studies of stacking, and instead fully optimise stacked complexes to yield accurate geometries in

realistic timescales. The inherently superior scaling of Kohn-Sham DFT with respect to size, combined with the use of smaller basis sets, synergistically reduces computational time and increases scope to larger systems compared with post-SCF techniques. Thus, there is much potential for this approach to be applied to larger systems, such as DNA duplexes or oligonucleotides, for which conventional correlated methods rapidly become unfeasible (see sections 4.3 and 4.4).

4.2.3 Electron density properties of stacked complexes

While the PES is the key property in any theoretical calculation, our interests also lie in characterising intermolecular interactions using electron densities. We then turned to AIM to analyse the electron densities: the MP2 and BH&H molecular graphs are shown in Figure 4.2. These graphs are remarkably similar, with Cartesian coordinates of all CP's agreeing to within 0.075 Å. The intramolecular interactions are entirely as expected, but *inter*-molecular contacts are revealing, consisting of two C…C bond paths and CP's, with an associated ring CP, required to satisfy the Poincare-Hopf relation.[20] This suggests that the dominant interaction in this dimer is indeed π-stacking, and not C—H…π hydrogen bonding as has been suggested previously.[21]



a b

Figure 4.2 Molecular graph of parallel-displaced benzene dimer: (a) MP2/6-311++G(d,p) and (b) BHandH/6-311++G(d,p) (red: bond CP, yellow: ring CP).

Analysis of electron density properties at the CP's reveals very similar electron distribution in the intermolecular region between the two methods. Using MP2, ρ_{CP} at bond and ring CP's are 0.0059 and 0.0052 au, respectively, compared with values of 0.0071 and 0.0063 au with BH&H. Thus, the density functional method slightly overestimates the extent of charge build-up in the intermolecular region. However, both methods agree that electron densities are very low at all intermolecular CP's, and are only marginally higher at CP's than at the RCP's, a typical feature of such weak, easily perturbed interactions. As a reference, BH&H calculated values for bond CP's in the water and argon dimers are 0.0343 and 0.0059 au, respectively.

Since substituents regulate the polarizability of the rings and the binding energy, changes in the topology of the electron density must be expected. The substituted benzene complexes initially studied by Sherrill are intriguing, indicating that both electron donating and withdrawing groups increase binding. BH&H works well for those complexes considered in Table 4.3, so we extended this approach to a wider range of substituents. A set of substituted benzene/substituted benzene complexes was constructed from the parallel-displaced benzene dimer geometry, as shown in Figure 4.3. Substituents X and Y, and where appropriate X', X'', Y', and Y'', were placed so as to not interact directly with the aromatic rings. Functional groups used were $-NO_2$, -F, -Br, -Cl, $-NH_2$, -CN, $-CH{=}CH_2$ (vinyl), -C(=O)H (formyl), -OH, $-CH_3$, and $-CCl_3$.

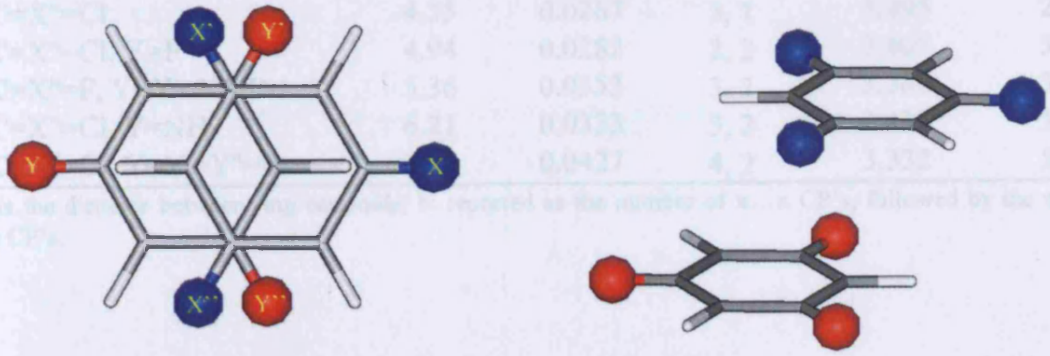


Figure 4.3 Two views of the substituted benzene dimers used.

Table 4.4 Calculated properties of substituted benzene dimers.

	ΔE (kcal mol^{-1})	$\Sigma\rho_\pi$	No CP's [b]	R (Å) [a]	$E_D/C = \alpha_1 * \alpha_2 * r^{-6}$
X=H, Y=H	2.31	0.0141	2, 0	3.758	1.45
X=OH	2.31	0.0141	2, 0	3.742	1.48
Y=OH, X=OH	2.32	0.0140	2, 0	3.746	1.48
X=Cl	2.53	0.0195	2, 0	3.688	1.75
X=CH=CH_2	2.55	0.0141	2, 0	3.725	1.65
X=F	2.56	0.0140	2, 0	3.732	1.46
X=Cl, Y=CH_3	2.62	0.0142	2, 0	3.686	1.76
X=NH_2, Y=OH	2.73	0.0141	2, 0	3.747	1.50
X=COH	2.75	0.0141	2, 0	3.760	1.47
X=Cl, Y=F	2.87	0.0197	3, 0	3.656	1.78
X=NH_2	2.87	0.0197	3, 0	3.787	1.42
X=OH, Y=COH	2.91	0.0141	2, 0	3.734	1.52
X=NH_2, Y=NH_2	2.98	0.0144	2, 0	3.711	1.64
X=NO_2, Y=COH	3.03	0.0140	2, 0	3.739	1.51
X=CN	3.07	0.0190	3, 0	3.709	1.52
X=Cl, Y=Cl	3.07	0.0141	2, 0	3.634	2.05
X=Cl, Y=Br	3.08	0.0193	3, 0	3.604	2.24
X=NO_2, Y=NO_2	3.12	0.0190	3, 0	3.602	1.86
X=NO_2	3.20	0.0192	3, 0	3.690	1.62
X=X'=F, Y=Y'=F	3.28	0.0244	4, 0	3.442	2.15
X=X'=X"=F	3.36	0.0206	2, 1	3.568	1.79
X=CCl_3	3.39	0.0197	3, 0	3.64	1.82
X=Cl, Y=NH_2	3.43	0.0200	3, 0	3.699	1.75
X=X'=Cl	3.68	0.0211	3, 0	3.581	2.25
X=X'=X"=F, Y=Y'=F	3.73	0.0224	2, 1	3.512	1.84
X=X'=X"=F, Y=Y'=Y"=F	3.79	0.0259	4, 0	3.405	2.15
X=X'=Br	3.84	0.0213	3, 0	3.576	2.44
X=NH_2, Y=NO_2	4.27	0.0214	3, 0	3.590	1.95
X=X'=F, Y=NH_2	4.32	0.0262	3, 1	3.555	1.94
X=X'=X"=Cl	4.35	0.0267	3, 1	3.495	2.81
X=X'=X"=Cl, Y=F	4.94	0.0282	2, 2	3.405	3.18
X=X'=X"=F, Y=Y'=Y"=Br	5.36	0.0353	3, 2	3.385	3.43
X=X'=X"=Cl, Y=NH_2	6.21	0.0333	3, 2	3.434	3.20
X=X'=X"=Cl, Y=Y'=Y"=Br	7.72	0.0427	4, 2	3.332	5.24

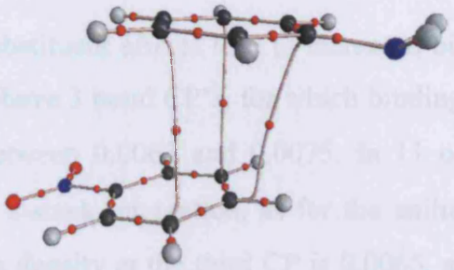
a: R is the distance between ring centroids; b: reported as the number of π...π CP's, followed by the number of X...π CP's.

Table 4.4 reports the binding energies, number of intermolecular bond CP's and the sum of electron densities here, inter-centroid distances and polarisabilities for 34 substituted benzene complexes computed at the BH&H/6-311++G(d,p) level. In line with the results in Table 4.3, we find with this larger set of molecules that *all* substitutions increase binding energy. For instance, the complex of aniline with benzene is bound by 2.87 kcal mol^{-1}, while with nitrobenzene the binding energy is 3.20 kcal mol^{-1}. This trend is repeated for all di- and tri-substituted complexes considered in Table 4.4: substitution increases binding energy between aromatic rings, with a slightly greater effect from electron withdrawing groups. The most strongly bound complex considered here is that between trichloro- and tribromo-benzene, which has a counterpoise corrected binding energy more than three times that of the unsubstituted benzene dimer. Generally, the complexes in Table 4.4 remain in the parallel-displaced geometry of the parent complex, but enhanced interaction energy may also result from X...π interactions, rather than directly from π-stacking.

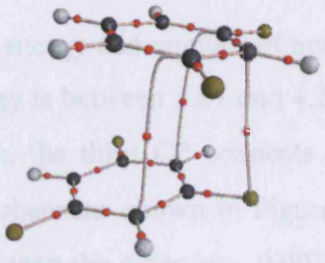
Substitution modifies geometry and electron densities of these complexes, leading to altered topologies. This intricate scenario is evident in both the number of π...π and X...π bond CP's in the intermolecular region, and the sum of the electron density at these CP's, $\Sigma\rho_{\pi}$. The number of intermolecular bond CP's varies from 2, for the parallel-displaced benzene dimer, to a maximum of 6 for the most strongly bound complexes. A total of 13 complexes are found to have the same topology as the benzene dimer (see Table 4.5), with two bond CP's connecting aromatic carbons, which can be assigned as π...π interactions. The binding energy for these complexes ranges from 2.31 to 3.07 kcal mol^{-1}, and the average electron density per CP, ρ_{ave} from 0.0071 to 0.0098 au.

Table 4.5 Summary of electron density properties from Table 4.4.

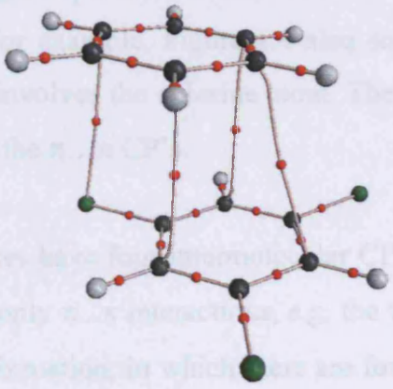
CP's	Number	ΔE Range (kcal mol^{-1})	ρ_{ave} Range (au)
2	13	2.31 – 3.07	0.0071 – 0.0098
3	13	2.87 – 4.27	0.0063 – 0.0075
4	5	3.28 – 4.94	0.0061 – 0.0075
5	2	5.36 – 6.21	0.0055 – 0.0071
6	1	7.72	0.0072



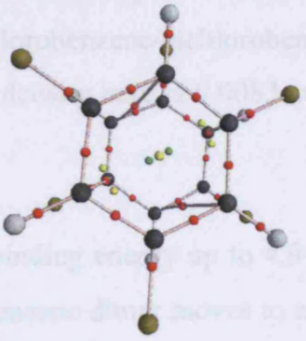
Aniline/nitrobenzene



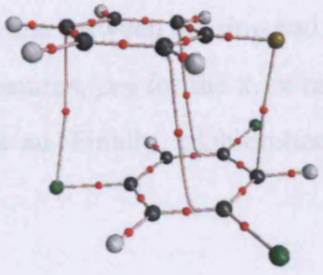
Trichlorobenzene/dichlorobenzene



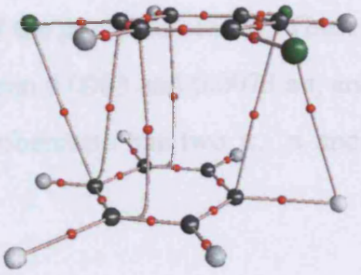
Trichlorobenzene dimer



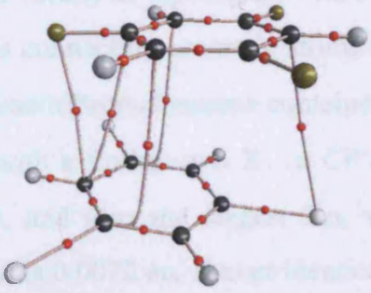
Trifluorobenzene dimer



Trichlorobenzene/fluorobenzene



Trichlorobenzene/tribromobenzene



Trifluorobenzene/tribromobenzene

Figure 4.4 Molecular graphs of some example complexes.

Stronger substituent effects lead to increased binding energy and number of bond CP's. 13 complexes have 3 bond CP's, for which binding energy is between 2.87 and 4.27 kcal/mol, and ρ_{ave} between 0.0063 and 0.0075. In 11 of these, the third CP connects two carbon atoms in a π-stack interaction, as for the aniline/nitrobenzene shown in Figure 4.4. Here, the electron density at the third CP is 0.0065, smaller than the other two, 0.0074 au. In the two remaining complexes, the third CP connects the aromatic ring to a substituent, an X...π interaction. For example, Figure 4.4 also shows trichlorobenzene/dichlorobenzene, where the third CP involves the chlorine atom. The electron density here is 0.0083 au, somewhat larger then at the π...π CP's.

Five complexes have four intermolecular CP's, with binding energy up to 4.94 kcal mol^{-1}. Two contain only π...π interactions, *e.g.* the trifluorobenzene dimer moves to an essentially eclipsed conformation, in which there are four π...π interactions, each with $\rho = 0.0065$ au. In contrast, the trichlorobenzene dimer and aniline/dichlorobenzene complex contain three π...π CP's and one between the ring and chlorine of the partner molecule. These complexes show similar features, ρ_{CP} for the π...π ranges between 0.0063 and 0.0073 au, and the X...π value is 0.0058 au. Finally, trichlorobenzene/fluorobenzene has two π...π and two X...π interactions.

The most strongly bound complexes, with binding energies higher than 5 kcal mol^{-1}, present a wide variety of topologies, with 5 or 6 intermolecular CP's, with a general pattern of 3 or 4 CP's connecting aromatic atoms with additional X...π interactions. For instance, trichlorobenzene/tribromobenzene contains six bond CP's, four of which are π...π stacking interactions, with a further two X...π CP's. This complex has the largest binding energy reported here, and also the largest $\Sigma\rho_{\pi}$ value of 0.0427 au. Interestingly, the electron density per CP is 0.0072 au, almost identical to the value for the benzene dimer, suggesting that binding energy is closely connected to the number of critical points and/or the total intermolecular electron density.

These observations indicate that substituents affect binding energy and topology in a broadly consistent manner: π-stacking interactions and associated CP's are ubiquitous, with between two and four such CP's in all complexes, with further X...π CP's in some cases. In general, more strongly bound complexes contain more CP's, such that the average energy per interaction appears to remain relatively constant. Consistent with this, we find that the total number of intermolecular CP's correlates with the energy, with an r^2 value 0.88 for the 34 complexes in Table 4.4. It has been suggested that ρ at the intermolecular ring CP decreases with increasing interaction strength,[22] and we also observe this trend, this value decreasing from 0.0063 au in the parent complex to 0.0040 au in the trifluorobenzene dimer, with this ring CP replaced by a cage or (3, +3) CP in the most strongly bound complexes, *i.e.* those bound by more than 5 kcal mol^{-1}. We note also that Zhikol *et al.* successfully used similar density properties to correlate the binding energy of the benzene dimer in a range of conformations.[23]

Also included in Table 4.4 is the sum of the electron density at all intermolecular bond CP's, $\Sigma\rho_\pi$. Inspection of these values again suggests some relation to the strength of binding: $\Sigma\rho_\pi$ for the benzene dimer is 0.0141 au from 2 CP's, while the trifluoro-benzene/benzene complex (ΔE = 3.36 kcal mol^{-1}) has a value of 0.0206 au, collected at 3 CP's. The trifluorobenzene dimer is stronger still (ΔE = 3.73 kcal mol^{-1}, $\Sigma\rho_\pi$ = 0.0224 au, 4 CP's), and replacing F with Br on one ring increases ΔE and $\Sigma\rho_\pi$ to 5.36 kcal mol^{-1} and 0.0403 au at 5 CP's. The most strongly bound complex in Table 4.4 also has the highest value of $\Sigma\rho_\pi$ present.

Table 4.6 contains the analogous data for 10 pyridine/benzene complexes, the cytosine dimer and 11 purine/benzene and purine/purine complexes. In general, pyridine complexes are more strongly bound than their benzene analogues, and contain an extra bond CP in the intermolecular region, connecting the pyridine nitrogen with the carbon directly above it. This is in addition to two C...C interactions, analogous to those for the benzene dimer. $\Sigma\rho_\pi$ values are also somewhat higher here, due both to the extra C...N and stronger C...C

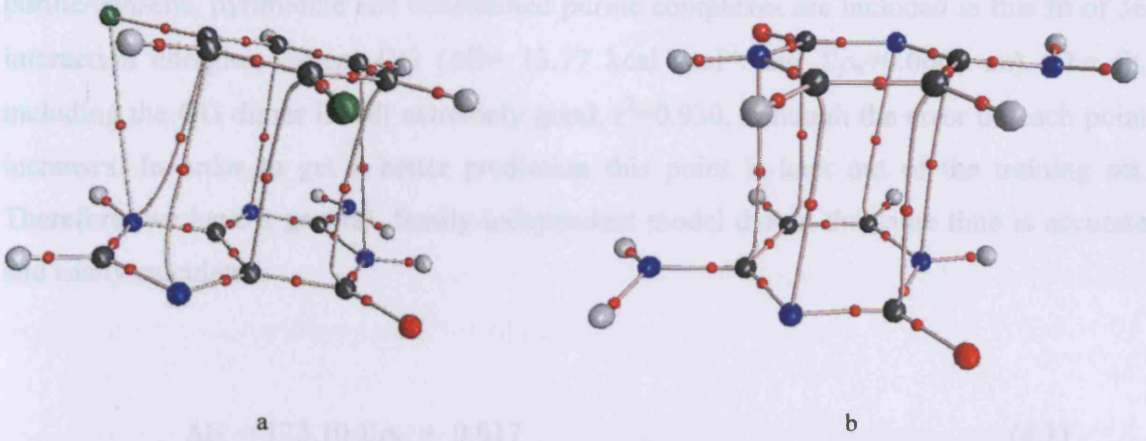
interactions. Once again, the binding energy seems to be essentially independent of electron-donating/withdrawing character.

Six purine/benzene complexes were included in this analysis, all of which have high binding energies ranging from 5 – 15 kcal mol^{-1}, and $\Sigma\rho_\pi$ between 0.0272 and 0.0447 au. As expected from the above discussion, $\Sigma\rho_\pi$ follows the trend of binding energy: a larger value generally corresponds to a greater binding energy. Figure 4.5 shows the intricacy of such interactions for the guanine/dichlorobenzene complex. 7 CP's are found in the region between the interacting partners, 6 of which are C...C and N...C π...π interactions, along with one X...π.

Table 4.6 Calculated properties of pyridine and purine complexes.[a]

	ΔE (kcal mol^{-1})	$\Sigma\rho_\pi$	No CP's	R (Å)	$E_D/C = \alpha_1 * \alpha_2 * r^{-6}$
Pyridine/X=CH_3	2.83	0.0183	3	3.730	1.40
Pyridine/X=H	2.88	0.0193	3	3.665	1.55
Pyridine/X=OH	2.92	0.0143	2	3.663	1.55
Pyridine/X=COH	3.05	0.0141	2	3.734	1.40
Pyridine/X=F	3.06	0.0163	3	3.732	1.34
Pyridine/X=COOH	3.19	0.0142	2	3.642	1.57
Pyridine/X=Cl	3.21	0.0196	3	3.616	1.80
Pyridine/X=CN	3.46	0.0195	3	3.623	1.60
Pyridine/X=NO_2	3.47	0.0195	3	3.608	1.70
Pyridine/X=NH_2	3.78	0.0203	3	3.710	1.47
Adenine/benzene	4.60	0.0272	4	3.408	3.55
Guanine/benzene	5.61	0.0383	5	3.241	5.04
Guanine/dichlorobenzene	7.12	0.0429	7	3.285	5.39
Guanine/dibromobenzene	7.20	0.0429	6	3.315	5.49
Adenine dimer [b]	7.28	0.0395	5	3.238	6.57
Adenine/dichlorobenzene	7.32	0.0447	6	3.315	4.86
Adenine/bromoadenine [b]	7.64	0.0429	6	3.315	5.22
Adenine/ dibromobenzene	7.64	0.0429	5	3.315	5.22
Guanine/bromoguanine [b]	8.60	0.044	6	3.293	7.07
Bromoadenine dimer	8.92	0.0485	6	3.267	7.42
Cytosine dimer	10.61	0.0535	6	3.070	6.02
Guanine dimer [b]	13.77	0.0600	6	3.274	6.78

a: geometries fully optimised at BH&H/6-311++G(d,p) unless otherwise stated; b: all $-NH_2$ groups constrained to be planar at monomer geometry throughout.



a b

Figure 4.5 Topology of (a) guanine/dichlorobenzene and (b) cytosine dimer.

The dimers of cytosine, adenine and guanine are included in Table 4.6, along with complexes of bromoadenine with adenine, bromoguanine with guanine, and the bromoadenine dimer. Free optimisation of these complexes, reported below, results in pyramidalisation of $-NH_2$ groups and strong hydrogen bonding between purines. For the purposes of Table 4.6, therefore, all $-NH_2$ groups were constrained to remain planar at the monomer geometry, resulting in purely π-stacked complexes, which can be directly compared to the remaining entries. The guanine dimer is strongly bound (ΔE = 13.77 kcal mol^{-1}) with associated high $\Sigma\rho_\pi$ (0.0590 au at 6 CP's), while the adenine dimer is weaker at 7.28 kcal mol^{-1} and $\Sigma\rho_\pi$ = 0.0393 at 5 CP's, following the pattern observed by Hobza *et al.*[24] The cytosine dimer was fully optimised, since no hydrogen bonds are present here so that the binding energy is entirely due to stacking. The binding energy of this dimer is large, more than 10 kcal mol^{-1}, and 6 intermolecular CP's are found, all π…π interactions, such that $\Sigma\rho_\pi$ = 0.0535 au, comparable to the most strongly bound purine complexes.

4.2.4 The electron density as a descriptor of π-stacking energy

Analysing the data reported so far, we find that $\Sigma\rho_\pi$ is linearly correlated with binding energy, with r^2 = 0.950, and a standard deviation of 0.48 kcal mol^{-1}, a fit illustrated in Figure 4.6. All data from purely stacked benzene/benzene, pyridine/benzene,

purine/benzene, pyrimidine and constrained purine complexes are included in this fit of 56 interaction energies, except GG (ΔE= 13.77 kcal mol^{-1} and $\Sigma\rho_\pi$=0.0600 au). The fit, including the GG dimer is still extremely good, r^2=0.930, although the error on each point increases. In order to get a better prediction this point is kept out of the training set. Therefore, we have a general, family-independent model that at the same time is accurate and easily calculated:

$$\Delta E = 173.10\,\Sigma\rho_\pi + 0.017 \qquad (4.1)$$

$$n= 56;\ r^2 = 0.950;\ r^2_{cv} = 0.940\ sd = 0.48\ kcal\ mol^{-1}$$

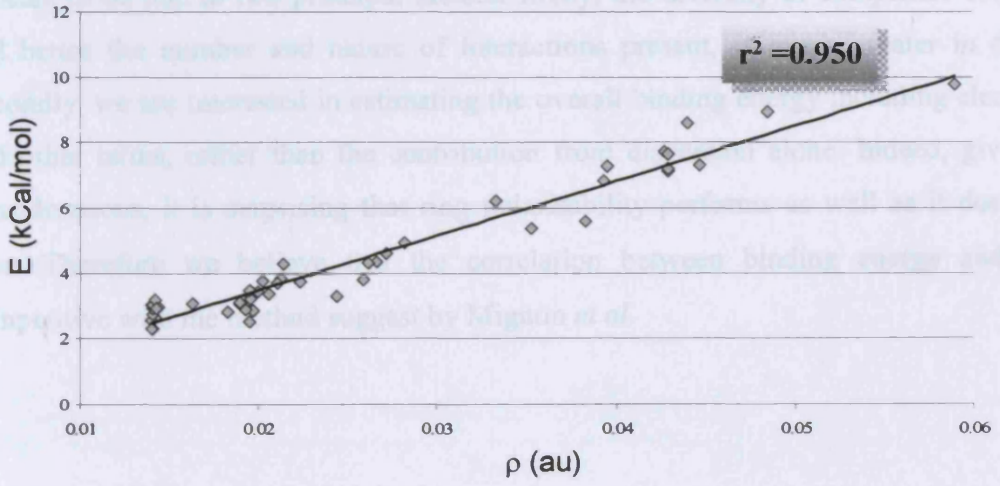


Figure 4.6 Plot of Binding Energy (kcal mol^{-1}) *vs.* $\Sigma\rho_\pi$ (au).

In contrast, the distance between aromatic rings performs relatively poorly (r^2 = 0.78) as a predictor of binding energy. Alternatively, Mignon *et al.*[17] proposed a method for prediction of π-stacking energies based on London's formula[25] for the dispersion interaction:

$$E_{disp} = C\,\alpha_1\,\alpha_2 / r^6 \qquad (4.2)$$

where α_1 and α_2 are the polarisabilities of the pyridine and benzene rings, r is distance between the interacting systems and C is a constant of proportionality. In their work, ring polarisability was calculated by subtracting the polarisability of the substituents (*e.g.* methyl radical, in the case of toluene) from that of the entire molecule, leading to excellent prediction of the correlation part of the binding energy ($r^2 = 0.95$).

We calculated ring polarisabilities in the same manner for each of the complexes in Tables 4.4 and 4.6, and combined according to equation 4.2, where C was recalculated for these DFT data. A reasonable correlation exists between this value and the overall binding energy ($r^2 = 0.90$), *i.e.* slightly worse from ref. 17 or Figure 4.6. This loss of accuracy appears to be due to two principal factors: firstly, the diversity of complexes considered, and hence the number and nature of interactions present, is much greater in our data. Secondly, we are interested in estimating the overall binding energy including electrostatic and other terms, rather than the contribution from dispersion alone. Indeed, given these considerations, it is surprising that ring polarisability performs as well as it does in this case. Therefore we believe that the correlation between binding energy and $\Sigma\rho_\pi$ is competitive with the method suggest by Mignon *et al.*

4.2.5 Hydrogen bonding and π-stacking in stacked DNA bases complexes

As seen in section 1.3 and 1.4 stacked complexes of DNA bases form hydrogen bonds as well as π-stacking interactions and undergo substantial structural deformations within each ring system. The binding energy in a fully optimised system is not therefore due solely to π-stacking, and decomposition into contributions from H-bonds and π-stacking interactions is not trivial, especially as cooperative effects between these two effects has been proposed.[26] AIM analysis is ideally suited to decomposing such instances of multiple interactions, allowing identification of π-stacking and hydrogen bonded contacts. We have shown in section 3.3 that H-bond energies can be predicted from electron density properties

with an error of *ca.* 1 kcal mol^{-1}: here, we have recalibrated the training set of section 3.3 at BH&H/6-311++G(d,p) level. However, it is well known that many DFT methods tend to overestimate H-bond strengths: BH&H optimisation of Watson-Crick paired GC and AT confirms that this is the case here. BSSE corrected pairing energies of 37.60 and 18.95 kcal mol^{-1} result for GC and AT, significantly higher than Hobza's CCSD(T) values of 28.5 and 15.4 kcal mol^{-1}, respectively,[27] but rather closer to other DFT values.[28] However, preliminary unpublished data (Steve Oldfield's Ph.D. Thesis) show that the correlation of MP2 and BH&H H-bond energies is almost perfect (r^2=0.989). In particular, H-bond energies are systematically overestimated about 1-2 kcal/mol and relative energy differences calculated at BH&H level are close to MP2 level.

In order to test our methods for decomposing the overall binding energy, we carried out full, unconstrained BH&H/6-311++G(d,p) geometry optimisation of seven DNA bases, namely, GG, GC, AA, GA, UU, TT and CC. Table 4.7 contains an analysis of π-stacking and H-bonding interactions in these complexes, while Figure 4.7 shows the resulting molecular graphs. Whereas in Table 4.6 all $-NH_2$ groups were kept planar to avoid formation of hydrogen bonds, here the molecules are completely relaxed, leading in general to increased binding energies.

Table 4.7 Analysis of DNA base interactions (au and kcal mol^{-1}).

	ΔE		π-stack		H-bond		$E_\pi + E_{HB}$
	BH&H	CCSD(T)	$\Sigma\rho_\pi$	E_π [a]	$\rho_{B...H}$	E_{HB} [b]	
GG	18.72		0.0500	8.67	0.0240	7.96	16.63
GC	17.29	16.90[c]	0.0185	3.31	0.0352	9.01	12.31
AA	9.20		0.0388	6.74	0.0090	2.69	9.43
GA	9.95		0.0344	5.97	0.0180	4.87	10.84
UU(*p*)	10.64	9.70[d]	0.0329	5.71	0.0147	4.12	9.83
TT	12.68		0.0318	5.52	0.0228	5.75	11.28
CC	10.61	10.40[e]	0.0535	9.28	0.00	0.00	9.28

a: Calculated from equation 4.1; b: Calculated as in section 3.3; c: ref. 27; d: ref. 19; e: ref. 18.

The fully optimised guanine dimer GG is by some margin the most strongly bound complex considered in this work, with a binding energy calculated at BH&H level of 18.72 kcal mol^{-1}. AIM analysis reveals five CP's corresponding to π-interactions, which sum to $\Sigma\rho_\pi$ = 0.0500 au, and a further two associated with N—H...N and N—H...O H-bonds, with a total electron density of 0.0240 au. Based on these values, we estimate a contribution of 8.67 kcal mol^{-1} from π-stacking, and 7.96 kcal mol^{-1} from H-bonding, which taken together give 16.63 kcal mol^{-1}, *i.e. ca.* 2 kcal mol^{-1} less than the BH&H supermolecule approach. Stacked GC is also strongly bound, and in good agreement with literature CCSD(T) calculations: here, H-bonding has a greater effect than π-stacking interactions, with highly non-planar $-NH_2$ groups in both molecules, and just two stacking CP's.

The adenine dimer AA shows similar topological features, with five π-stacking CP's in the intermolecular region. However, the density associated with each π...π interaction is lower than in GG, such that the stacking energy E_π is estimated at just 6.74 kcal mol^{-1}. In this dimer, only one N—H...N hydrogen bond, with an estimated energy of 2.69 kcal mol^{-1}, was found, leading to a combined estimate of binding energy marginally higher than found *via* the super-molecule approach. The guanine adenine complex GA has the lowest $\Sigma\rho_\pi$ of the purine complexes in Table 4.7, resulting in E_π = 5.97 kcal mol^{-1}. However, two relatively strong N—H...N bonds are also found between the molecules, which contribute 4.87 kcal mol^{-1}, giving of interaction energy *ca.* only 1 kcal mol^{-1} from the directly calculated value.

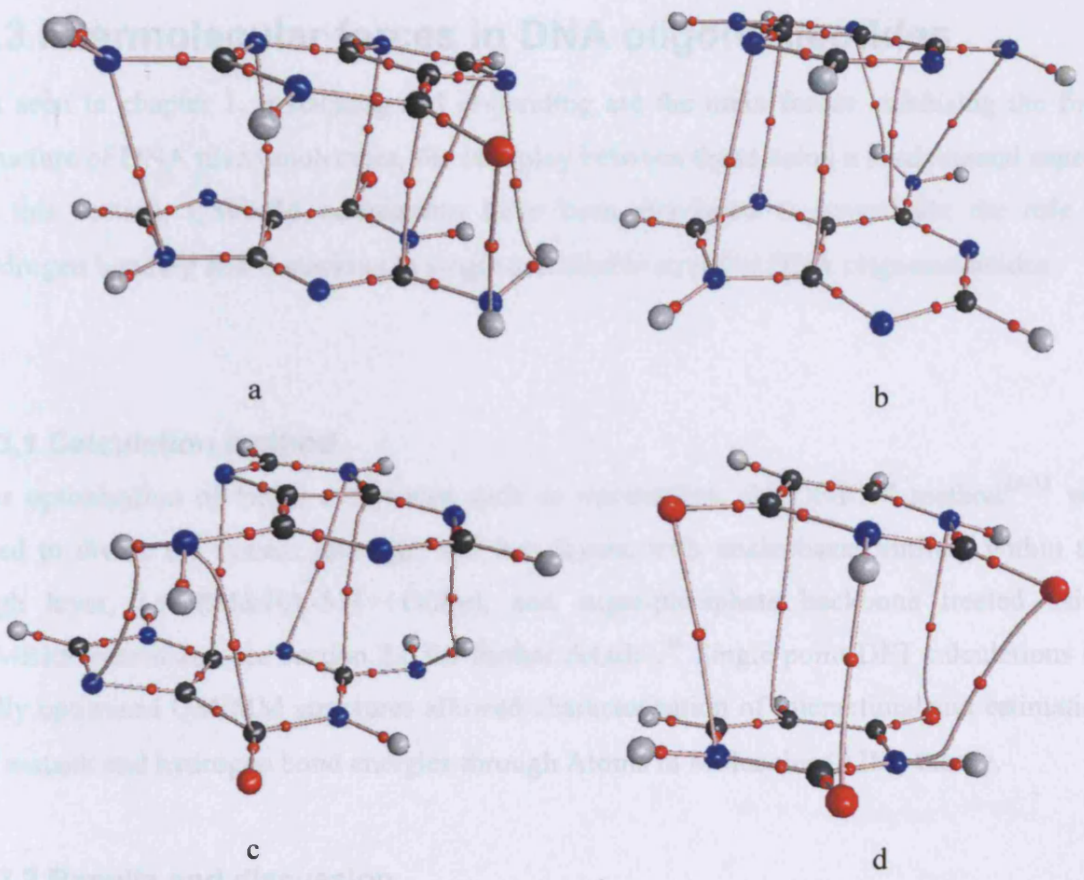


Figure 4.7 Molecular graphs of (a) GG, (b) AA, (c) GA, and (d) UU.

Table 4.7 indicates that H-bonding plays a part in pyrimidine dimers also: UU has four π-stacking CP's, for which $\Sigma\rho_\pi = 0.0329$, equivalent to 5.71 kcal mol^{-1} from eq. 4.1, and a single N—H...O CP (4.12 kcal mol^{-1}), which again sum to a value close to the directly calculated binding energy. Similarly, the TT is bound through four π-stacking and two C—H...O CP's, which contribute 5.52 and 5.75 kcal mol^{-1} from stacking and H-bonding respectively. In contrast, CC contains no H-bonding interactions, and is bound purely *via* π-stacking.

4.3 Intermolecular forces in DNA oligonucleotides

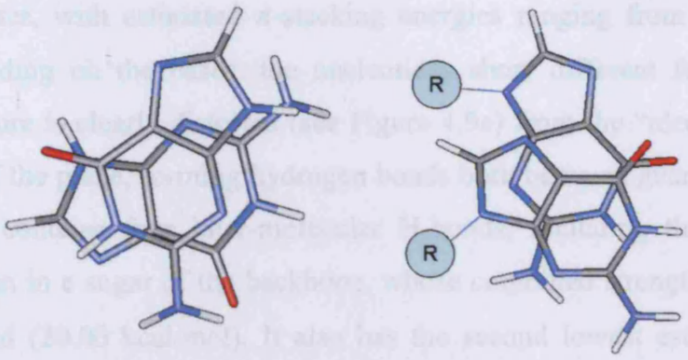
As seen in chapter 1, π-stacking and H-bonding are the main forces stabilising the final structure of DNA macromolecules, the interplay between those being a fundamental aspect. In this section, QM/MM calculations have been employed to investigate the role of hydrogen bonding and π-stacking in single and double stranded DNA oligonucleotides.

4.3.1 Calculation method

For optimisation of larger complexes such as nucleotides, the ONIOM method[29-33] was used to divide the system into high and low layers, with nucleobases entirely within the high layer, *i.e.* BH&H/6-311++G(d,p), and sugar-phosphate backbone treated using AMBER potentials (see section 2.6 for further details).[34] Single point DFT calculations on fully optimised QM/MM structures allowed characterisation of interactions and estimation of π-stack and hydrogen bond energies through Atoms in Molecules (AIM) theory.

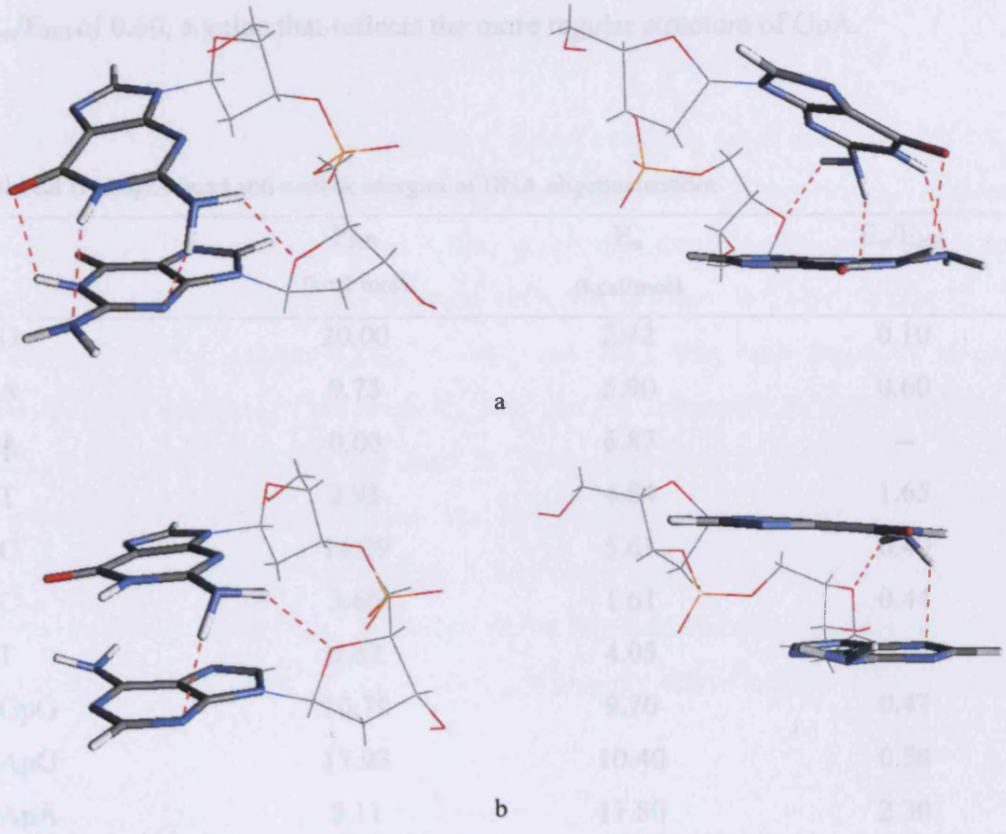
4.3.2 Results and discussion

The computational efficiency of the BH&H/6-311++G(d,p) level opens up the possibility of studying large systems: our ultimate goal in these studies is to examine the stacking of DNA bases such as di- and tri-nucleotides. In accord with literature,[35-37] the geometries and binding energies of sugar/phosphate-linked complexes differ from those of free dimers and trimers due to the phosphate backbone. The DNA frame in fact tends to keep bases in a co-planar, co-parallel orientation, whereas free complexes do prefer co-planar bases but an almost perpendicular orientation between the axes of bases, which maximizes interactions (Figure 4.8). Thus, π-stacking and hydrogen bonding in nucleotides are generally weaker compared to free structures.



a b

Figure 4.8 Conformation of guanine dimer (a) and GpG (b).



a

b

Figure 4.9 Two views of (a) GpG and (b) GpA.

A- H-bonding and π-stacking of single stranded DNA structures

Table 4.8 summarises our analysis of these nucleotides and Figures 4.9 and 4.10 illustrate some examples. Generally, all DNA chains studied keep a co-planar, co-parallel orientation

of bases, with estimated π-stacking energies ranging from 2.5 to 7.0 kcal/mol. However, depending on the bases, the nucleotides show different features. For example, the GpG structure is clearly distorted (see Figure 4.9a) from the "ideal" geometry, with N—H atoms out of the plane, forming hydrogen bonds both between guanines and to the sugar backbone. GpG contains four inter-molecular H-bonds, including three between bases and one to oxygen in a sugar of the backbone, whose combined strength is the largest of all complexes studied (20.00 kcal/mol). It also has the second lowest estimated π-stacking energy (2.40 kcal/mol), giving a ratio E_π/E_{HB} of just 0.10. In contrast, GpA is much more co-planar (Figure 4.9b), indicating that H-bonding and π-stacking are more equally shared: AIM analysis shows two H-bonds (9.80 kcal/mol) and four π-stacking interactions (5.90 kcal/mol), giving a ratio E_π/E_{HB} of 0.60, a value that reflects the more regular structure of GpA.

Table 4.8 Hydrogen bond and π-stack energies of DNA oligonucleotides.

	E_{HB} (kcal/mol)	E_π (kcal/mol)	E_π/E_{HB}
GpG	20.00	2.42	0.10
GpA	9.75	5.90	0.60
ApA	0.00	6.87	--
ApT	2.93	4.84	1.65
GpC	14.29	5.61	0.40
CpC	3.65	1.61	0.44
CpT	6.62	4.05	0.61
GpGpG	20.70	9.70	0.47
GpApG	17.93	10.40	0.58
ApApA	5.11	11.80	2.30
ApTpA	19.38	5.72	0.30
GpCpG	36.48	7.30	0.20

π-stacking is the sole intermolecular interaction in ApA, with no H-bonding but five π-stacking CP's found, corresponding to 6.87 kcal/mol. Interestingly, the $-NH_2$ nitrogen

atoms interact via π-stacking rather than H-bonding, and contribute *ca.* 40% to $\Sigma\rho_\pi$, yielding the largest π-stack energy among the dinucleotides considered. ApT shows similar properties, with just a single weak C—H...π interaction (2.93 kcal/mol) and three π-stacking CP's interaction, with the $-NH_2$ nitrogen in adenine again involved in π-stacking rather than H-bonding. Thus, for ApA and ApT, π-stacking interactions are more important than H-bonds. GpC contains three π-stacking and three H-bonding CP's, contributing 5.61and 14.20 kcal/mol, respectively (E_π/E_{HB} = 0.40) with $-NH_2$ groups of both Guanine and Cytosine involved in both H-bonds and π-stacking. Similarly, CpC is mainly stabilised by H-bonding, with just one H-bond and one stacking CP, contributing 3.65 and 1.61 kcal/mol, respectively. CpT is more strongly bound than CpC, with H-bond and stacking energy of 6.62 and 4.05 kcal/mol, each from two CP's.

Turning to the tri-nucleotides, optimisation of GpGpG yields a much more regular structure than found in GpG, in which each pair contains two N—H...N and N—H...O H-bonds, with bases almost parallel (see Figure 4.10a) suggesting that H-bonding is less dominant than in GpG. Stacking interactions between each pair of bases are also similar (5.35 and 4.30 kcal/mol), giving a ratio E_π/E_{HB} = 0.47, indicating that both forms of interaction stabilise the final structure. Thus it seems that GpG is unusually distorted by inter-base H-bonding: the H-bond energy per pair in GpGpG is around half that found in the dinucleotide. GpApG (Figure 4.10b) has a balance of π-stacking and H-bonding (E_π/E_{HB} = 0.58), with some redistribution of energy compared to GpA: H-bonding is diminished in one G...A pair and enhanced in the other. In contrast, ApApA is dominated by π-stacking just as in ApA (E_π/E_{HB} = 2.31), with only one relatively weak H-bond per pair.

Interestingly, the ApTpA and GpCpG present unique features among the complexes studied here. As shown in Figure 4.10c-d, the structure of these trinucleotides is so distorted that H-bonding between the first and third bases occurs, while the combined π-stacking energies are the smallest found among trinucleotides. On the other hand, the hydrogen bond energies are among the strongest, 19.38 for ApTpA and 36.48 kcal/mol for GpCpG, with the first base and third base (A...A for ApTpA and G...G for GpCpG) interacting via strong

N—H…N and N—H…O bonds (*ca.* 30% of the overall H-bond energy). Thus, the ratio E_{π}/E_{HB} for ApTpA and GpCpG is 0.30 and 0.20 respectively. Therefore, as the hydrogen bond interactions prevail by far over π-stacking, the overall structure assumes a globular conformation.

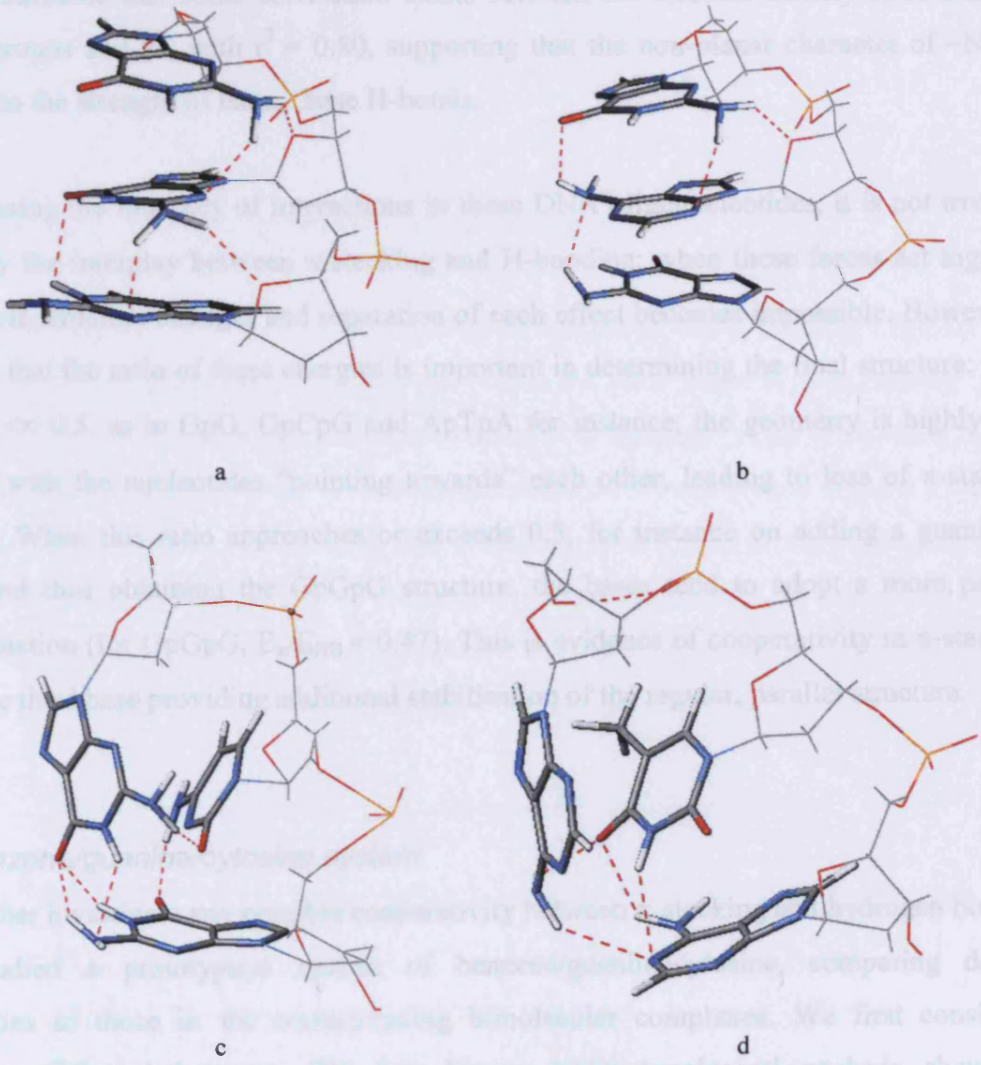


Figure 4.10 GpGpG (a), GpApG (b), GpCpG (c) and ApTpA (d) nucleotides.

These studies show that $-NH_2$ groups interact *via* H-bonding to heavy atoms of other bases, participate in π-stacking or, in some cases, both. $-NH_2$ groups of guanine are largely involved in H-bonding, mainly *via* N—H…N and N—H…O, leading to a high degree of

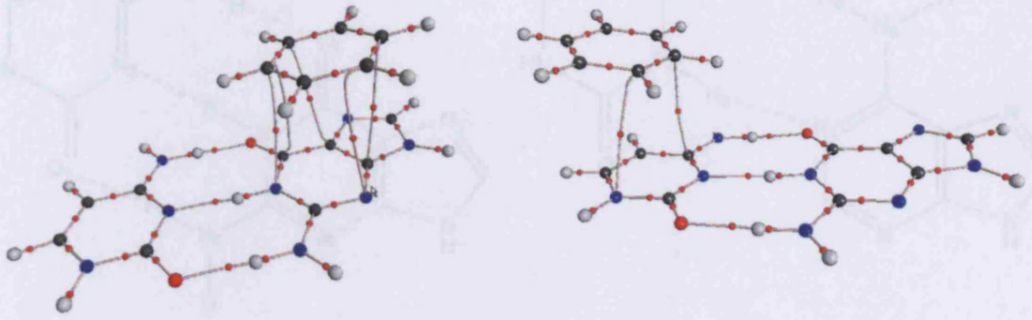
pyramidalisation with the sum of bond angles at N (Σ°) equal to *ca.* 330° on average. $-NH_2$ groups of adenine are involved in both N—H...N H-bonds and π-stacking interactions. Adenine's H-bonds are generally weaker than those in guanine, and the $-NH_2$ groups are closer to planarity (Σ° = 346.7° on average). Cytosine shows similar properties: $-NH_2$ groups interact *via* both H-bonding and stacking, and Σ° ranges between 347.0° and 352.0°. It is remarkable that some correlation exists between the electron density of H-bonds of $-NH_2$ groups and Σ°, with $r^2 = 0.80$, supporting that the non-planar character of $-NH_2$ is related to the strength of base...base H-bonds.

Considering the intricacy of interactions in these DNA oligonucleotides, it is not trivial to quantify the interplay between π-stacking and H-bonding: when these forces act together, the whole structure changes and separation of each effect becomes impossible. However, it is clear that the ratio of these energies is important in determining the final structure: when $E_\pi/E_{HB} << 0.5$, as in GpG, GpCpG and ApTpA for instance, the geometry is highly non-planar, with the nucleotides "pointing towards" each other, leading to loss of π-stacking energy. When this ratio approaches or exceeds 0.5, for instance on adding a guanine to GpG and thus obtaining the GpGpG structure, the bases tend to adopt a more parallel conformation (for GpGpG, $E_\pi/E_{HB} = 0.47$). This is evidence of cooperativity in π-stacking, with the third base providing additional stabilisation of the regular, parallel structure.

B- Benzene/guanine/cytosine system

To further investigate any possible cooperativity between π-stacking and hydrogen bonding we studied a prototypical system of benzene/guanine/cytosine, comparing density properties to those in the corresponding bimolecular complexes. We first considered benzene...GC and benzene...CG (see Figure 4.11): topological analysis shows no qualitative difference from bimolecular complexes. Moreover, in neither case do $\Sigma\rho_\pi$ nor ρ_{HB} differ significantly from values found in the analogous dimers (guanine-cytosine, benzene-guanine, and benzene-cytosine), with maximum variations of 0.001 au, suggesting that only little interplay between stacking and H-bonding occurs here. Following Geerlings *et al.*'s recent work,[38] we then considered the effect of substitution on benzene, including

groups such as $-NO_2$, -F, $-CH_3$, -CHO, -OH, and $-NH_2$. Table 4.9 reports data for these complexes.



a b

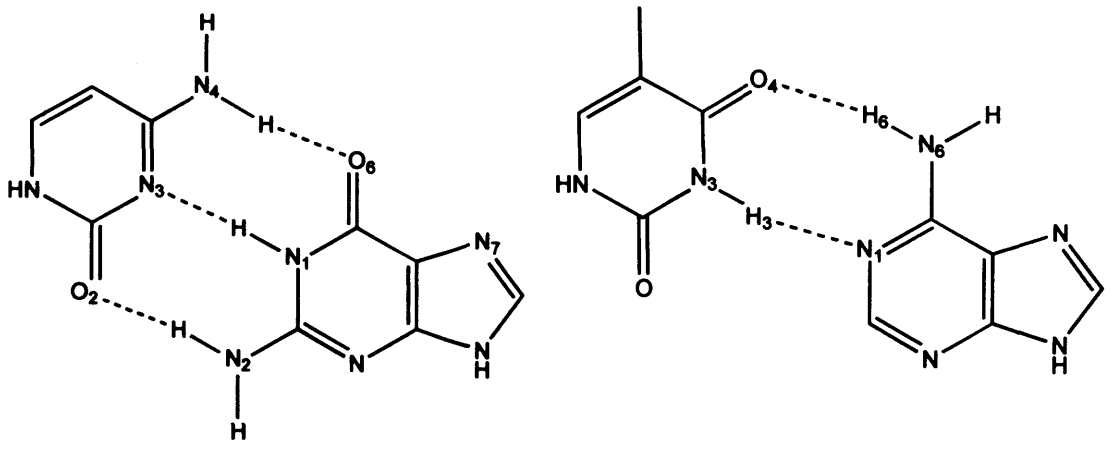
Figure 4.11 Topology of (a) benzene...GC and (b) benzene...CG.

Geerlings suggested that the mutual influence of π-stacking and H-bonding depends on the hardness of the substituted benzene, *i.e.* benzenes with electron withdrawing groups stacked over guanine lead to lower charge transfer to cytosine. Thus, from $-NO_2$ to $-NH_2$ cytosine acts as a progressively better H-bond acceptor (through N_3 and O_2) and a worse H-bond donor (through H_4), confirming that π-stacking does influence the H-bonding of GC. However, individual variations in H-bonds are small (generally no more than *ca.* 1.5 kcal/mol), and since $H_4...O_6$ has the opposite trend from $H_1...N_3$ and $H_2...O_2$, the total pairing energy hardly changes but distortion of the GC pair occurs.

Table 4.9 Electron density (ρ in au) at H-bonds CP's in benzene...G≡C.

	$H_4...O_6$[a]	$H_1...N_3$[a]	$H_2...O_2$[a]
free GC	0.0520	0.0436	0.0365
$-NO_2$	0.0512	0.0438	0.0371
-CHO	0.0512	0.0439	0.0371
-F	0.0507	0.0442	0.0375
-H	0.0506	0.0443	0.0377
$-CH_3$	0.0505	0.0443	0.0378
-OH	0.0506	0.0444	0.0378
$-NH_2$	0.0501	0.0446	0.0383

a: Numbering scheme shown in Figure 4.12.



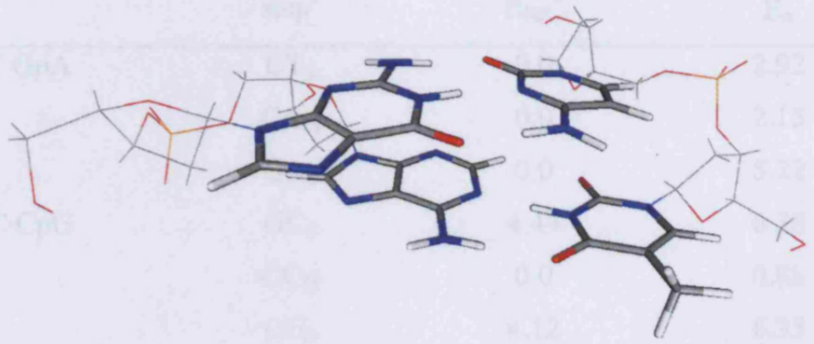
a b

Figure 4.12 Numbering scheme for (a) GC and (b) AT.

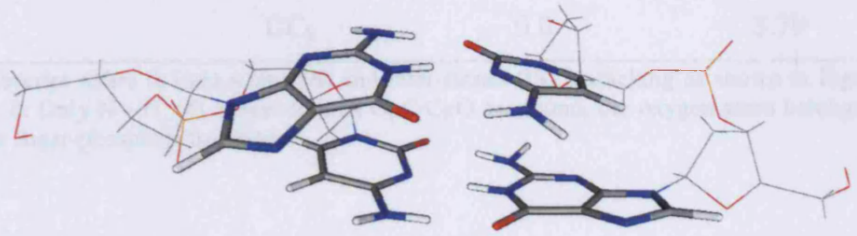
C- H-bonding and π-stacking of double stranded dinucleotides

A similar treatment of more realistic models of DNA chains, namely the double stranded dinucleotides, GpC·CpG, CpT·GpA and GpG·CpC, is shown in Figures 4.13 and 4.14. To our knowledge, this is the first attempt to fully optimise such systems using *ab initio* or DFT methods, as opposed to classical force fields. Compared with solution and crystal structures, little is known about the structure of DNA in the gas phase, but experimental and computational studies agree on some important aspects, especially that base pairing and π-stacking are preserved, but strong distortion of DNA occurs.[39]

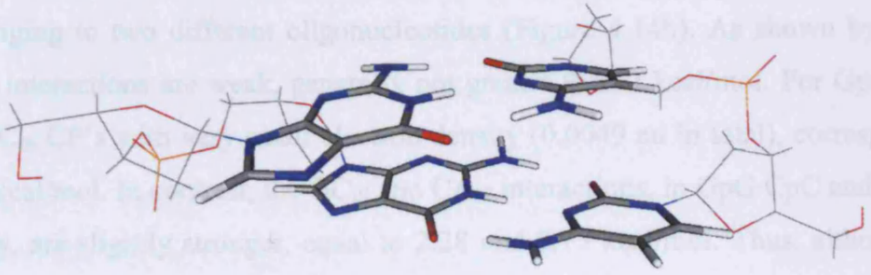
Our results are consistent with this: π-stacking and H-bonding are evident, as are large distortion of the "ideal" DNA chains structure. Table 4.10 reports data for intra- (S) and inter-strand (IS) π-stacking as well as H-bonding, as schematised in Figure 4.14a. In general, calculated energies are similar to those for the single nucleotides strands (Table 4.8). For instance, the stacking energies between C...T and G...A in CpT·GpA are 2.92 and 5.22 kcal/mol, within 1 kcal/mol of the corresponding single strand energies. Similarly, the intra-strand G...C π-stacking energies for GpC·GpC are *ca.* 6 kcal/mol, *cf.* 5.61 for single stranded nucleotides. Interestingly, in GpG·CpC, both G...G and C...C are rather larger than the corresponding single strands, 7.00 and 3.79 kcal/mol respectively, compared with *ca.* 5.00 and 1.61 kcal/mol.



a



b



c

Figure 4.13 CpT·GpA (a), GpC·CpG (b) and GpG·CpC (c) duplexes.

Table 4.10 H-bonding and π-stacking energies of the duplexes.

	step[a]	E_{HB}[b]	E_{π}
CpT·GpA	CT_S	0.0	2.92
	CA_{IS}	0.0	2.15
	GA_S	0.0	5.22
GpC·CpG	GC_S	4.44	6.38
	CC_{IS}	0.0	0.88
	CG_S	4.12	6.35
GpG·CpC	GG_S	0.0	7.06
	GC_{IS}	0.0	2.28
	CC_S	0.0	3.79

a: subscript refers to intra-strand (S) and inter-strand (IS) π-stacking as shown in Figure 4.14; b: Only N—H...O interactions in GpC·CpG are found, the oxygen atom belonging to the sugar-phosphate backbone.

Topological analysis also reveals evidence for inter-strand stack interactions (IS) between bases belonging to two different oligonucleotides (Figure 4.14b). As shown by Hobza *et al.*,[40] these interactions are weak, generally not greater than 2 kcal/mol. For GpC·CpG we find two CC_{IS} CP's with very small electron density (0.0049 au in total), corresponding to less than 1kcal/mol. In contrast, the GC_{IS} and CA_{IS} interactions, in GpG·CpC and CpT·GpA respectively, are slightly stronger, equal to 2.28 and 2.15 kcal/mol. Thus, although weak, these interactions contribute between 10 and 25% of the overall π-stacking energy, and therefore play a role in the structure of these chains.

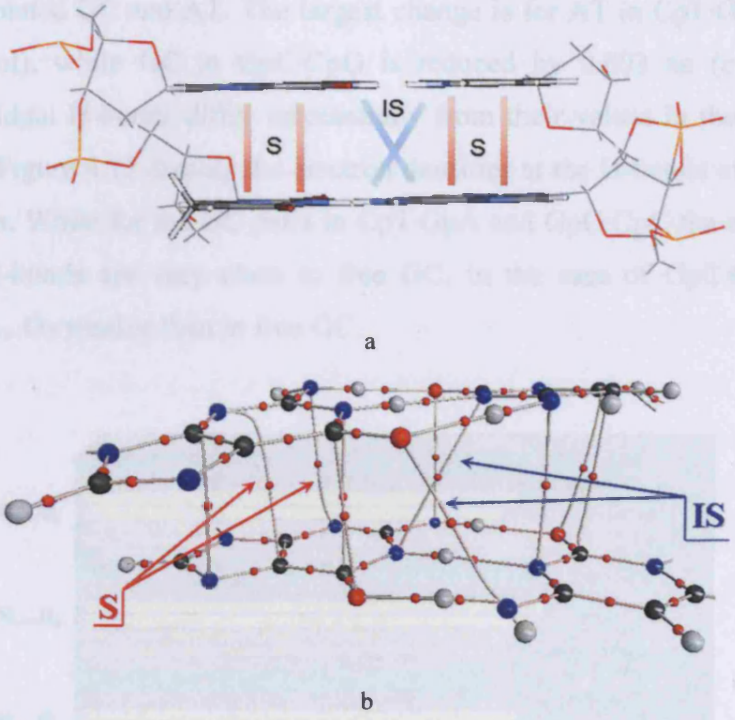


Figure 4.14 (a) Schematic drawing of the intra-strand stack (S) and inter-strand stack (IS) interactions; (b) detail of the GpG·CpC topology.

As noted above, the flexibility of $-NH_2$ groups allows H-bonds to stabilise single stranded nucleotides, especially in guanine and adenine strands. However, our analysis of the duplexes CpT·GpA, GpC·CpG and GpG·CpC suggests a different scenario: comparatively few intra-strand H-bonds are found here (though of course such points are found between strands) involving solely guanine N—H…O, with estimated energies of *ca.* 4 kcal/mol each. Moreover, $-NH_2$ groups are much closer to planarity than in the single strands, and are involved in π-stacking interactions rather than intra-strand H-bonds. This appears an effect of *inter*-strand pairing: the bases are paired via strong Watson-Crick H-bonds in the plane of the molecule, acting to constrain $-NH_2$ groups to this plane, which are thus less able to deform. This is evident in the average values of Σ°: C = 358.3° > A = 357.1° > G = 351.2°, and the increased π-stacking interactions of these nitrogen atoms seen in Table 4.10.

Although the strength of H-bonding is overestimated at BH&H level (see previous section), it is still possible to use AIM to compare GC and AT pairing in various environments (Table 4.11 and Figure 4.15). In all cases considered, the overall pairing energy is close to

that found in isolated GC and AT. The largest change is for AT in CpT·GpA at +0.008 au (*ca.* +2 kcal/mol), while GC in GpC·CpG is reduced by 0.003 au (*ca.* -1 kcal/mol). However, individual H-bonds differ substantially from their values in the free base pairs: Table 4.11 and Figure 4.15 display the electron densities at the H-bonds of GC pairs in the studied duplexes. While for the GC pairs in CpT·GpA and GpG·CpC the electron densities of the single H-bonds are very close to free GC, in the case of GpC·CpG, $H_4...O_6$ is stronger and $H_2...O_2$ weaker than in free GC.

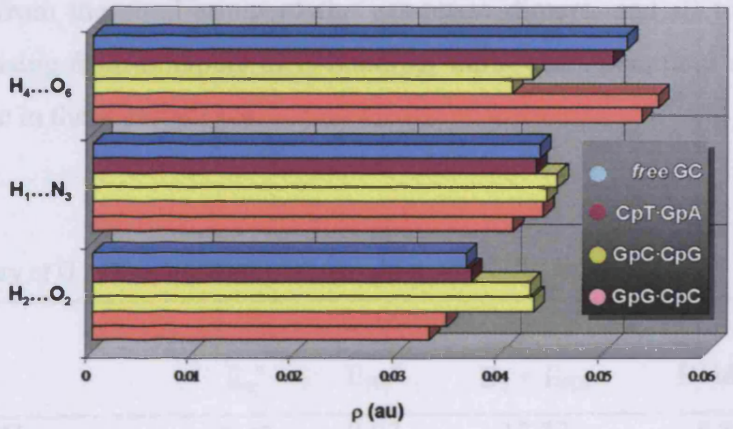


Figure 4.15 Electron density at the H-bonds of GC pair in the studied duplexes.

Table 4.11 GC and AT pairing in the duplexes. [a]

DNA	Base pair	$H_4...O_6$ ($H_3...N_1$)	$H_1...N_3$ ($H_6...O_4$)	$H_2...O_2$	ρ_{TOT}
free	GC	0.0520	0.0436	0.0365	0.132
	AT	0.0586	0.0285		0.087
CpT·GpA	GC	0.0506	0.0433	0.0370	0.131
	AT	0.0644	0.0302		0.095
GpC·CpG	GC	0.0409	0.0443	0.0431	0.129
	GC	0.0429	0.0453	0.0427	0.131
GpG·CpC	GC	0.0534	0.0410	0.0329	0.128
	GC	0.0550	0.0440	0.0346	0.134

a: see Figure 4.12 for numbering; the numbering in parenthesis refers to AT.

D- Comparison with literature

Table 4.12 summarises our results in terms of G...G, G...A and G...C interactions, excluding those from distorted GpCpG and ApTpA. Thus, we compared the estimation of the π-stack and H-bond energies to BSSE corrected binding energies calculated as shown in Hobza's works.[40,41] Nucleobase geometries were extracted from the optimised nucleotide structures and the backbone replaced by hydrogen atoms on N9; then the BSSE corrected energy was evaluated at both BH&H/6-311++G(d,p) and at MP2/6-31G(0.25)*.[9,10] This approach was used in order to *i)* test the capability of BH&H of reproducing π-stack energies away from the equilibrium of the gas-phase dimers, and *ii)* to clarify whether cooperativity arising from interplay of H-bonding, intra- and interstrand stack interactions might play a role in these complexes.

Table 4.12 Summary of G...G, G...A, and G...C interactions.

		E (kcal/mol)			ΔE^b	
		E_π[a]	E_{HB}[a]	$E_\pi + E_{HB}$	BH&H	MP2
G...G	**GpG**	2.42	9.97	12.37	8.85	10.89
	GpGpG	4.32	5.92	10.24	6.52	9.51
	Gp**GpG**	5.35	5.92	11.27	4.85	8.61
	GpG·CpC	7.06	0.0	7.06	3.21	4.25
G...A	**GpA**	5.90	4.42	10.32	11.10	12.64
	GpApG	5.57	4.10	9.67	6.75	10.07
	Gp**ApG**	4.83	8.53	13.36	9.27	11.30
	CpT·**GpA**	5.22	0.0	5.22	5.14	8.85
G...C	**GpC**	5.61	9.94	15.55	15.94	14.68
	GpC·CpG	6.38	0.0	6.38	6.20	9.57
	GpC·**CpG**	6.35	0.0	6.35	6.05	9.40

a: E_π and E_{HB} are calculated from topological analysis of the electron density; b: BSSE corrected binding energies of corresponding stacked base pairs.

BH&H binding energies are generally smaller than corresponding MP2 values by about 2 kcal/mol, except for the GpC and GpA geometry, where the agreement is close. In the other cases, two factors may be responsible for this discrepancy: *i)* as shown in previous section, MP2 overestimates π-stacking energies; *ii)* the BH&H functional may be less effective in predicting the energy of non-equilibrium geometries than was found for fully optimised species, see paragraph 4.2.2. Comparison of MP2 binding energies and the sum of $E_\pi + E_{HB}$ shows reasonable agreement (statistical errors estimated to be *ca.* 2 kcal/mol) for all single stranded oligonucleotides: the estimation of binding energy indeed differs 1.30 kcal/mol on average from MP2. However, this changes for the double stranded DNA structures, where the difference is often more then 3 kcal/mol.

Thus, while the approach of taking stacked base pairs from optimised single stranded oligonucleotides works well, it apparently fails in the case of duplexes because of the intricacy of interactions. In other words, the interaction, for instance, of G...A in CpT·GpA is strongly affected by the environment, and particularly of the complementary bases C and T which interact with G and A *via* strong H-bonds, as well as inter- and intra-strand stack interactions. AIM analysis, which takes into account the effects of environment on the electron density, therefore complements the supermolecule approach, allowing study of the subtle interplay arising from the complexity of H-bonding and stack interactions.

Hobza[40] and Geerlings[38] have performed high-level *ab initio* calculations on stacked base structures extracted from experimental DNA geometries. Despite the use of different geometries and theoretical methods, agreement between their values and our BH&H data is qualitatively good: stacking in GpG is rather weak (2.42 kcal/mol), slightly less than the Geerlings's MP2 value of 3.39 kcal/mol, but almost doubles to 5.35 and 4.33 kcal/mol in GpGpG. G...A interactions are also affected by the length of the chain, with stacking energies between 3.70 and 5.90 kcal/mol: the latter value is close to Hobza's value of 6.5 kcal/mol. Finally, the π-stacking energy estimated from the topology in GpC is rather large, between 5.61 and 6.38 kcal/mol, within *ca.* 1-2 kcal/mol of Hobza's values of 7.7 and 7.9 kcal/mol, with less pronounced differences from environmental factors.

Similarly, direct comparison with experimental studies is not possible due to difficulties in obtaining accurate experimental gas-phase DNA structures: nonetheless, our models match several known facts. Bowers and co-workers have provided many fascinating results, *e.g.* that single stranded nucleotides exist in three different conformations, with important interplay between π-stacking and H-bonding, confirming that base...base N—H...N/O H-bonding occurs, *cf.* Table 4.8 and Figures 4.9 and 4.10.[39] They also show that the conformation of di- and tri- nucleotides is largely determined by the sequence, even for such small DNA chains. Our work supports this, for instance, while the GpG structure is strongly distorted, with stacking energy of 2.42 kcal/mol, GpA is almost parallel with π-stacking energy of *ca.* 5 kcal/mol.

Several studies indicate that in the gas phase Watson-Crick pairs are better preserved in G...C than A...T,[42,43] due to the stronger H-bonds here. Orozco and co-workers[43] showed that G...C stacked pairs also are better preserved, suggesting that these interactions are largely responsible for the maintenance of the structural features of DNA in gas-phase. Our results indicate that G...C stacking interactions are strong, similar to G...G. Moreover, Orozco and co-workers stressed that although many known DNA features are retained in the gas phase, some interesting differences emerge. Molecular dynamics simulations found that T-shape π-stacking occurred in some DNA chains. Similarly, we find structures of GpCpG and ApTpA that present such features, with two bases interacting *via* parallel stacking and the third in a T-shape conformation (see Figure 4.10c-d). In particular, our analysis indicates that this conformation is principally due N—H...O/N and C—H...π hydrogen bonding interactions.

Thus, even if direct comparison of DNA geometry obtained with this hybrid functional to experiment is not practicable, literature theoretical and experimental data supports our estimations of stacking in these nucleotides, supporting the validity of this approach.

4.4 Cisplatin-DNA adducts: H-bonding and π-stacking

Following the studies reported in previous section, QM/MM calculations have been employed to investigate the role of hydrogen bonding and π-stacking in several single and double stranded cisplatin-DNA structures. The BH&H/AMBER/AIM approach was also used to study platination of a double stranded DNA octamer d(CCTG*G*TCC)·d(GGACCAGG) (platinated guanines indicated by *), for which an experimental structure is available. Comparison between theory and experiment is satisfactory, and also reproduces previous DFT based studies of similar structures.

4.4.1 Calculation method

Calculation methods are essentially the same as previous section: the ONIOM method[29-33] was used to split the system into QM and MM regions, with platinated nucleobases and cisplatin itself entirely within the high level layer, i.e. BH&H/6-311++G(d,p) (for the Pt atom the SDD[44] basis set and ECP was used), and sugar-phosphate backbone treated using AMBER potentials.[34]

4.4.2 Results and discussion

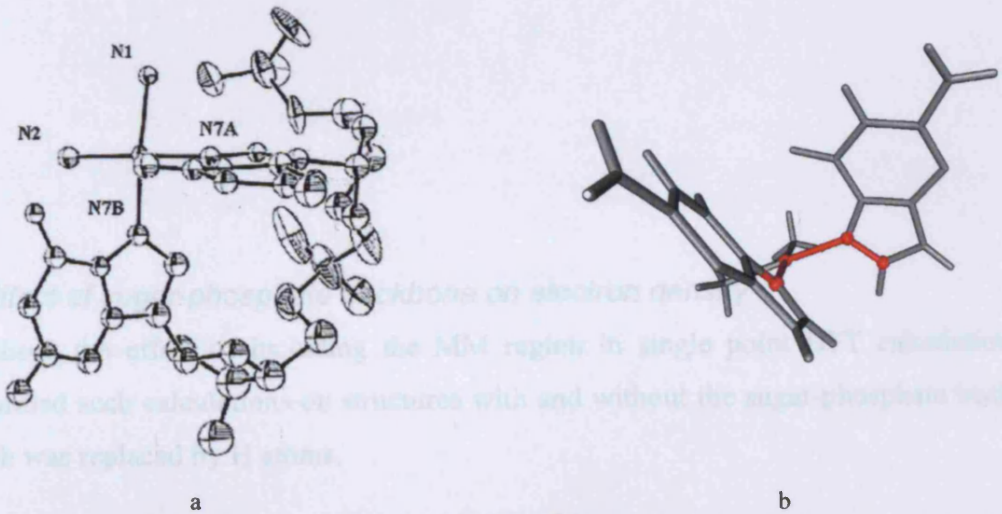
Having established in sections 4.2 and 4.3 that the hybrid BH&H density functional can account for π-stacking interactions in model systems of DNA, it is important to test its performance for cisplatin and related structures before using it to analyse the effects of platination on π-stacking. The optimised structure of cisplatin obtained at the BH&H/6-311++G(d,p)(SDD) level is reported in Table 4.13, along with experimental and various theoretical values. In general, agreement is excellent with both experiment from solvent-free crystal values[45] and those from HF,[46] MP2,[46] and DFT[47], with bond lengths within *ca.* 0.05 Å of experimental and similar bond angles to all previous theoretical estimates.

Table 4.13 Bond distances and angles of cisplatin.

	BH&H	HF/6-31G(d,p)[a]	MP2/6-31G(d)[a]	BLYP[b]	Expt.[c]
Pt—N (Å)	2.058	2.139	2.090	2.065	2.01±0.04
Pt—Cl (Å)	2.283	2.348	2.312	2.315	2.33±0.01
N-Pt-N (°)	97.9	95.0	96.5	98.0	
N-Pt-Cl (°)	83.4	84.7	84.9	83.0	
Cl-Pt-Cl (°)	95.2	95.6	93.8	95.5	

a: Hausheer *et al.*'s;[46] b: Carloni *et al.*'s;[47] c: experimental values in the solvent-free crystal.[45]

In addition, comparison between optimised and Sherman's[48] X-Ray data for cis-[Pt(NH_3)$_2$(d(pGpG))], where p indicates the sugar-phosphate backbone of DNA, is excellent (see Table 4.14). The average difference between theoretical prediction and experimental data on Pt—N bonds is just 0.01 Å, whereas bond and dihedral angles differ *ca.* 1° on average.



a b

Figure 4.16 a) Atom labelling in cis-[Pt(NH_3)$_2$(d(pGpG))],[48] and b) dihedral angles between guanine bases, in accord with Orbell's convention.[49]

Table 4.14 Geometric features of cis-[$Pt(NH_3)_2$(d(pGpG))].[a]

	BH&H	Expt.[b]
Pt—N1	2.030	2.050(0.036)
Pt—N2	2.032	2.055(0.045)
Pt—N7A	2.032	1.968(0.055)
Pt—N7B	2.031	2.015(0.063)
N7A-Pt-N1	88.5	89.6(1.3)
N7A-Pt-N2	177.0	176.8(2.5)
N7A-Pt-N7B	91.0	88.3(2.2)
N1-Pt-N2	92.2	91.7(1.01)
N1-Pt-N7B	176.7	175.9(2.5)
N2-Pt-N7B	88.2	90.3(1.8)
Gua/Gua[c]	78.0	81.2(4.3)

a: see Figure 4.16a for labelling; b: average over 4 molecules reported by Sherman *et al.*,[48] sd in parenthesis; c: Dihedral angle between guanines: see Orbell *et al.*[49] and Figure 4.16b.

A- Effect of sugar-phosphate backbone on electron density

To check the effect of including the MM region in single point DFT calculations, we performed such calculations on structures with and without the sugar-phosphate backbone, which was replaced by H atoms.

Table 4.15 Electron densities (au) of **a** and **b** structures.

		a	**b**
	Pt—N7	0.12282	0.12402
	N—H…O	0.05631	0.05651
GGs	C…N	0.00954	0.00954
	C…C	0.00911	0.00905
	N…C	0.00947	0.00946
	C…O	0.01110	0.01111
CCs	N…O	0.00866	0.00849
	C…N	0.00892	0.00892
	N…N	0.00894	0.00893
GCis	O…N	0.00846	0.00849
	N…N	0.00445	0.00445
	O…N	0.00647	0.00650
GC_{WC}†	$H_4…O_6$	0.03878	0.03881
	$H_1…N_3$	0.04687	0.04696
	$H_2…O_2$	0.05053	0.05055
GC_{WC}‡	$H_4…O_6$	0.05348	0.05347
	$H_1…N_3$	0.04164	0.04180
	$H_2…O_2$	0.03588	0.03597

†: platinated Watson-Crick GC pair;
‡: free Watson-Crick GC pair.

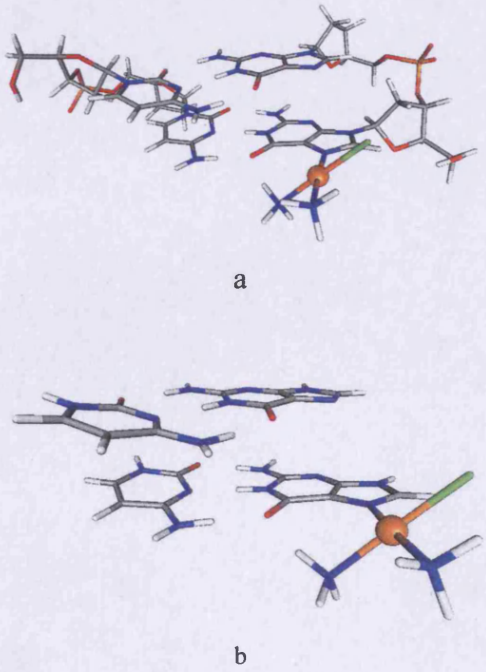


Figure 4.17 platinated GpG·CpC with (a) and without (b) sugar-phosphate backbone

Table 4.15 and Figure 4.17 illustrate the instance of the monofunctional complex of GpG·CpC (see next paragraphs for further details): identical topology (*i.e.* number and type of CP's) and almost identical electron density values were found, the largest difference being 0.0012 au, and the average just 0.0001 au, or less than 1% of a typical value. These results are reflected in all the studied systems, suggesting that topology and electron density are essentially independent from the atoms included in the MM region.

B- Interaction energies of cisplatin single stranded DNA complexes

Table 4.16 reports binding energies for all studied mono- and bifunctional complexes of cisplatin with single stranded DNA, along with bond lengths and electron densities of platinum...base bonds (Pt—X, where X = N or O). As expected from data reported in Chapter 3 and many previous studies,[50-53] guanine complexes are more stable than adenine. For instance, $cisGpG_{mono}$ has the highest binding energy of all monofunctional complexes, with cisplatin directly bound to nitrogen of one guanine. $cisGpA_{mono}$, where the guanine molecule not directly bound to the metal centre is substituted with adenine, has a binding energy 4 kcal/mol less, as well as a slightly shorter, stronger Pt—X bond (r = 2.011Å, ρ_{CP} = 0.123 au). Interestingly, AIM reveals a weak secondary interaction between Pt and N7 of adenine, with ρ_{CP} = 0.009 au: this point will be returned to below.

The binding energy for $cisGpGpG_{mono}$ (see Figure 4.18) is much larger than for the dinucleotides, approximately 30 kcal/mol more than $cisGpG_{mono}$. However, the Pt—N_G bond is similar in distance and density to platinated dinucleotides, suggesting that the extra binding energy is mostly due to electrostatic attraction between phosphate and platinum. As above, AIM reveals a secondary interaction to N7 of G2, with ρ_{CP} = 0.014 au (see Figure 4.18b).

Table 4.16 Binding energies and bonding properties of platinated adducts.

	Binding energy (kcal/mol)		r(Pt—X) (Å)	ρ_{CP}(Pt—X) (au)
$cisGpG_{mono}$	147.17	Pt—N_G	2.022	0.119
$cisGpA_{mono}$	143.45	Pt—N_G	2.011	0.123
		Pt…N_A	3.421	0.009
$cisApG_{mono}$	127.51	Pt—N_A	2.100	0.113
$cisGpGpG_{mono}$	174.40	Pt—N_G	2.014	0.120
		Pt…N_G	3.149	0.014
$cisGpG_{bi}$	312.43	Pt—N_G	2.032	0.116
		Pt—N_G	2.032	0.117
$cisGpA_{bi}$	285.22	Pt—N_G	2.017	0.122
		Pt—N_A	2.012	0.124
$cisGpG_{chel}$	284.57	Pt—N_G	2.049	0.109
		Pt—O_G	2.111	0.084
$cisGpGpG_{bi}$	397.96	Pt—N_G	2.043	0.112
		Pt—N_G	2.009	0.123
$cisGpApG_{bi}$	396.35	Pt—N_G	2.029	0.117
		Pt—N_A	2.004	0.126

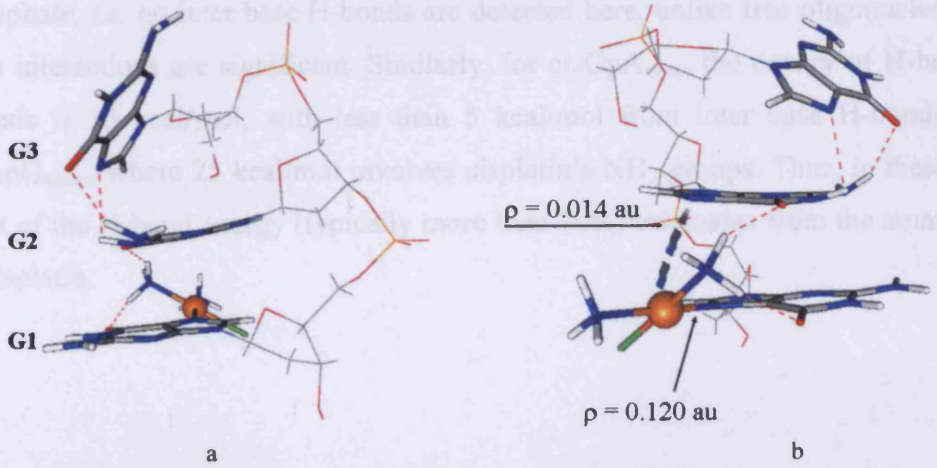


Figure 4.18 Optimised geometry of $cisGpGpG_{mono}$ showing (a) distortion of G…G interaction and (b) electron densities at the Pt—N_G bond and Pt…N secondary interaction.

As seen in section 3.3, the binding energies of bifunctional adducts are more than double those of monofunctional complexes, due to the +2 charge on these complexes. The trend observed for mono-functional complexes is preserved, with guanine complexes more strongly bound than adenine. In particular, the cisGpG$_{bi}$ is more than 25 kcal/mol more stable than cisGpA$_{bi}$, a much larger difference than observed in the monofunctional complexes above. The structure in which platinum is chelated by O and N of a single guanine has similar stability to cisGpA$_{bi}$, *i.e.* considerably less than the bifunctional adduct, in agreement with previous work (see section 3.3).[54,55] The trinucleotides show a similar trend, but here the difference in binding energy between the cisGpGpG$_{bi}$ and cisGpApG$_{bi}$ is only 2 kcal/mol. To rationalise these differences, we turn to AIM analysis to decompose into covalent bonding, H-bonding and π-stacking effects.

C- H-bonding and π-stacking of monofunctional cisplatin-DNA adducts

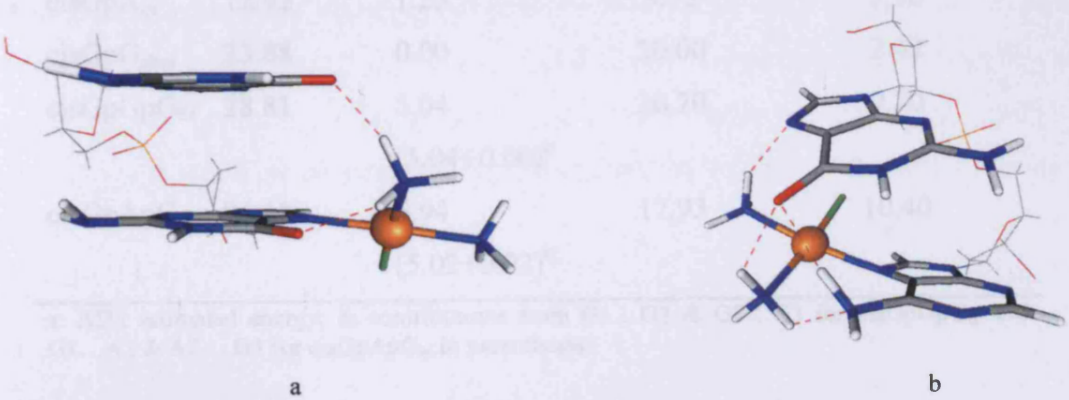
Table 4.17 compares H-bonding and π-stacking energies estimated from AIM data between free and platinated di- and trinucleotides. It is clear that H-bonding is prevalent in platinated species, with strong interactions involving Pt—N—H and Pt—Cl groups as donors and acceptors. For instance, in cisGpG$_{mono}$, the energy due to these interactions is 21 kcal/mol, with only 2 kcal/mol from a single N—H...O H-bonds between guanine and phosphate, *i.e.* no inter base H-bonds are detected here, unlike free oligonucleotides where such interactions are significant. Similarly, for cisGpA$_{mono}$ the energy of H-bonding of Pt ligands is 15 kcal/mol, with less than 5 kcal/mol from inter base H-bonding, and for cisApG$_{mono}$, where 25 kcal/mol involves cisplatin's NH_3 groups. Thus, in these complexes most of the H-bond energy (typically more than 70%) originates from the ammonia groups of cisplatin.

Table 4.17 H-bonding and π-stacking of free and platinated oligonucleotides (kcal/mol)

	platinated		*free*	
	E_{HB}[a]	E_{π}[a]	E_{HB}[a]	E_{π}[a]
$cisGpG_{mono}$	23.22	5.25	20.00	2.42
$cisGpA_{mono}$	20.83	3.00	9.75	5.90
$cisApG_{mono}$	36.28	2.13	9.75	5.90
$cisGpGpG_{mono}$	25.36	6.97 (3.18+3.79)[b]	20.70	9.70

a: AIM estimated energy; b: contributions from G1…G2 and G2…G3 in parenthesis.

In contrast, platination reduces all π-stacking energies bar one by 3-4 kcal/mol, accompanied by substantial geometrical distortion (Figures 4.18 and 4.19). It appears that strong Pt—N—H…X (X = N, O) H–bonds cause the two purines to point towards each other, leading to a loss of π-stacking energy. Only in $cisGpG_{mono}$ is π-stacking enhanced by platination, from 2.42 to 5.25 kcal/mol: however, this may be due to the initial strong distortion of free GpG structure, and the low value of E_{π} in free GpG (see previous section). In the trinucleotide $cisGpGpG_{mono}$ (Figure 4.18) both G…G stacks are of approximately equal energy, and sum to *ca.* 3 kcal/mol less than the free complex.



a b

Figure 4.19 H-bonds in (a) $cisGpG_{mono}$ and (b) $cisApG_{mono}$.

D- H-bonding and π-stacking of bifunctional cisplatin-DNA adducts

As seen in chapter 1, when cisplatin binds to DNA, the major products are 1,2-intrastrand G-Pt-G and A-Pt-G complexes: Table 4.18 reports H-bonds and π-stack energies of these bi-functional adducts. cisGpG$_{bi}$ contains two almost symmetric Pt—N—H...O interactions with energies of 10 kcal/mol each (Figure 4.20a), whereas cisGpA$_{bi}$ (Figure 4.20b) contains just one such interaction to guanine, along with a much weaker Pt—N—H...N contact with adenine. In the guanine chelate complex cisGpG$_{chel}$ (Figure 4.20c), strong H-bonds are formed between cisplatin ammine groups and O6/N7 of the non-coordinated guanine, leading to a high H-bond energy. Therefore, as for monofunctional complexes, most H-bonding energy stems from ammine groups of cisplatin. In the trinucleotide complexes, H-bonding energy is larger than in the free structure, and again this comes mainly from cisplatin. For instance, in cisGpGpG$_{bi}$ G3 interacts via Pt—N—H...O with cisplatin, while Pt—N—H...O and N—H...N are found between cisplatin and G1 (see Figure 4.20d). The importance of hydrogen bonding involving cisplatin over those between bases is also apparent in the $-NH_2$ groups of guanine and adenine, which are significantly less pyramidal than in optimisation of free DNA (average sum of angles ≈ 350°, *cf.* 330° in free DNA).

Table 4.18 H-bonding and π-stacking of free and platinated oligonucleotides (kcal/mol).

	platinated		*free*	
	E_{HB}[a]	E_{π}[a]	E_{HB}[a]	E_{π}[a]
cisGpG$_{bi}$	21.16	0.00	20.00	2.42
cisGpA$_{bi}$	12.92	1.25	9.75	5.90
cisGpG$_{chel}$	23.88	0.00	20.00	2.42
cisGpGpG$_{bi}$	28.81	5.04 (5.04+0.00)[b]	20.70	9.70
cisGpApG$_{bi}$	26.18	5.94 (5.02+0.92)[b]	17.93	10.40

a: AIM estimated energy; b: contributions from G1...G2 & G2...G3 for cisGpGpG$_{bi}$ and G1...A2 & A2....G3 for cisGpApG$_{bi}$ in parenthesis.

As seen in 1.5.4, one of the main effects of platination is to disrupt π-stacking between bases.[48,56-61] The topological results in Table 4.18 support this: in $cisGpG_{bi}$, $cisGpG_{chel}$, and $cisGpGpG_{bi}$ no bond critical points corresponding to π-stacking are located between the platinated bases, and hence $E_\pi = 0$ in all these cases. In complexes involving adenine, *i.e.* $cisGpA_{bi}$ and $cisGpApG_{bi}$ a solitary π-stacking critical point between G and A is located, corresponding to an energy contribution of just 1 kcal/mol. In the two trinucleotides considered, stacking between the non-platinated base and its neighbour is hardly disrupted from that found in the free structure. *i.e.* 5 kcal/mol for G...G and 6 kcal/mol for G...A. In this way, topological analysis using AIM is able to quantify the disruption of intra-strand stacking, showing it to be large in all cases and largest between guanines, while interactions between the remaining bases is virtually unchanged.

E- Platinated double stranded DNA complexes

As well as affecting the H-bonding and π-stacking within DNA strands, platination can also disrupt interactions between strands. To study this we optimised mono- and bi-functional and chelate GpG·CpC complexes (see Figure 4.21), denoted $cisGpG{\cdot}CpC_{mono}$, $cisGpG{\cdot}CpC_{bi}$ and $cisGpG{\cdot}CpC_{chel}$ respectively, as well as one bi-functional complex of GpA·CpT ($cisGpA{\cdot}CpT_{bi}$) . Properties of covalent Pt—N(O) bonds follow the patterns outlined above, though as in previous work the presence of cytosine in the base pair leads to a systematic strengthening of these interactions (see section 3.3). AIM analysis reveals a number of secondary interactions Pt...X (X = N, O), detailed in Table 4.19. As seen in single stranded complexes, these interactions are weak with bond lengths longer than 3Å and ρ_{CP} between 0.01 and 0.02 au, *i.e.* around 1Å longer and an order of magnitude weaker than the "direct" interactions. Hydrogen bonds within strands and involving cisplatin are almost identical to the single-stranded complexes, and so no further details are reported on these interactions.

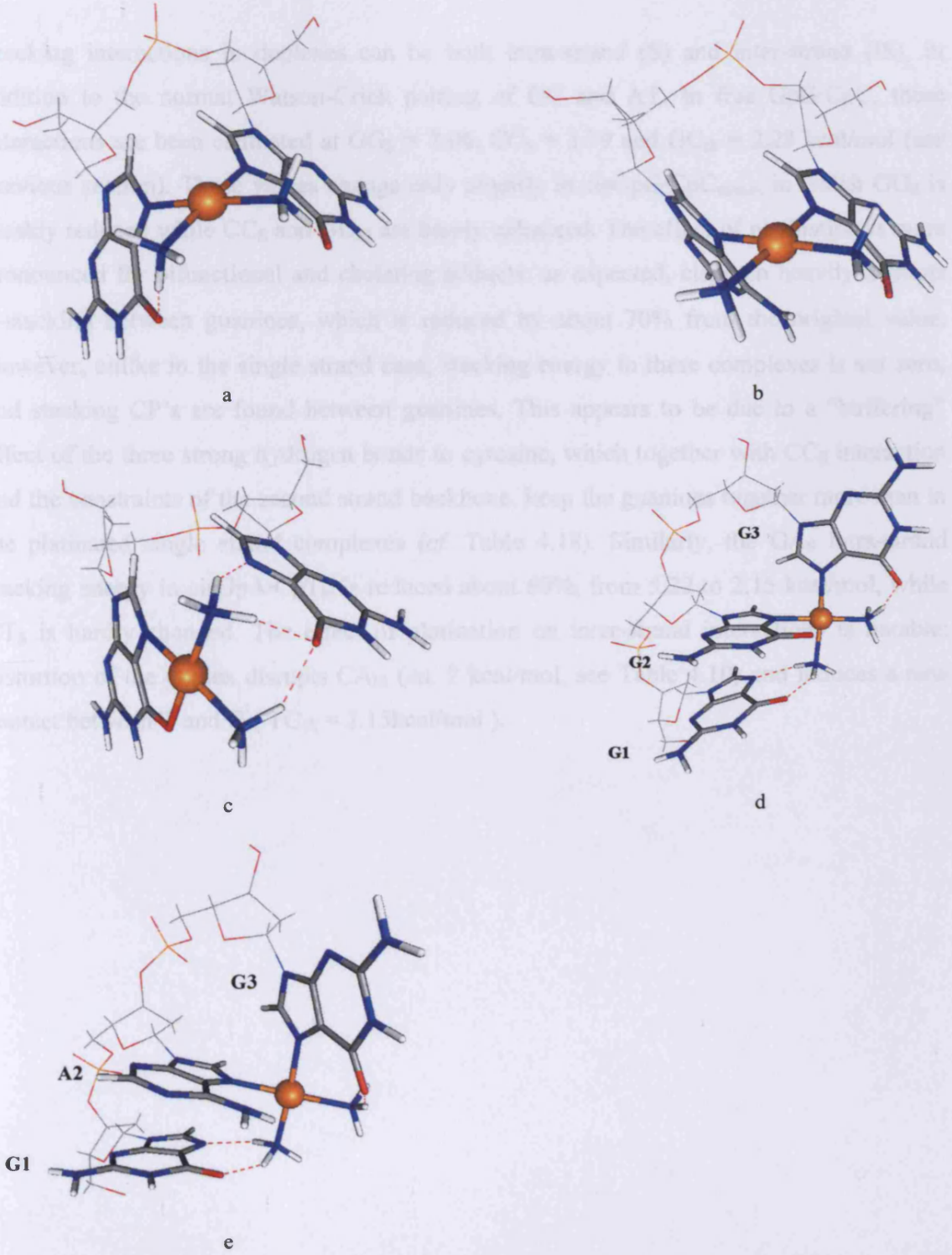


Figure 4.20 Optimised Geometries of (a) cisGpG$_{bi}$, (b) cisApG$_{bi}$, (c) cisGpG$_{chel}$, (d) cisGpGpG$_{bi}$, and (e) cisGpApG$_{bi}$.

Stacking interactions in duplexes can be both intra-strand (S) and inter-strand (IS), in addition to the normal Watson-Crick pairing of GC and AT. In free GpG·CpC, these interactions are been estimated at GG_S = 7.06, CC_S = 3.79 and GC_{IS} = 2.28 kcal/mol (see previous section). These values change only slightly in cisGpG·CpC_{mono}, in which GG_S is weakly reduced while CC_S and GC_{IS} are barely enhanced. The effect of platination is more pronounced for bifunctional and chelating adducts: as expected, cisplatin heavily disrupts π-stacking between guanines, which is reduced by about 70% from the original value. However, unlike in the single strand case, stacking energy in these complexes is not zero, and stacking CP's are found between guanines. This appears to be due to a "buffering" effect of the three strong hydrogen bonds to cytosine, which together with CC_S interaction and the constraints of the second strand backbone, keep the guanines together more than in the platinated single strand complexes (*cf.* Table 4.18). Similarly, the GA_S intra-strand stacking energy in cisGpA·CpT_{bi} is reduced about 60%, from 5.22 to 2.15 kcal/mol, while CT_S is hardly changed. The effect of platination on inter-strand interactions is notable: distortion of the duplex disrupts CA_{IS} (*ca.* 2 kcal/mol, see Table 4.10) and induces a new contact between T and G (TG_{IS} = 2.15kcal/mol).

Table 4.19 Interactions in platinated duplexes.

	Pt—X		$E_{N—H...O}$ (kcal/mol)	E_{π} (kcal/mol)		
	r (Å)	ρ (au)		GG_S (GA_S)	CC_S (CT_S)	GC_{IS} (TG_{IS})
free GpG·CpC	--	--	--	7.06	3.79	2.28
free GpA·CpT	--	--	--	5.22	2.92	2.15
cisGpG·CpC_{mono}	2.007	0.124	14.32	6.65	4.60	3.41
cisGpG·CpC_{bi}	2.022	0.120	10.82	2.28	3.10	4.31
	2.017	0.122				
	3.026[a]	0.017				
cisGpG·CpC_{chel}	2.060	0.107	13.80	2.08	4.97	3.02
	2.080	0.093				
	3.028[b]	0.018				
cisGpA·CpT_{bi}[c]	2.008	0.125	8.75	2.15	3.60	2.71
	2.019	0.121				

a: Pt…O secondary interaction; b: Pt…N secondary interaction; c: first row refers to Pt—G and second to Pt—A.

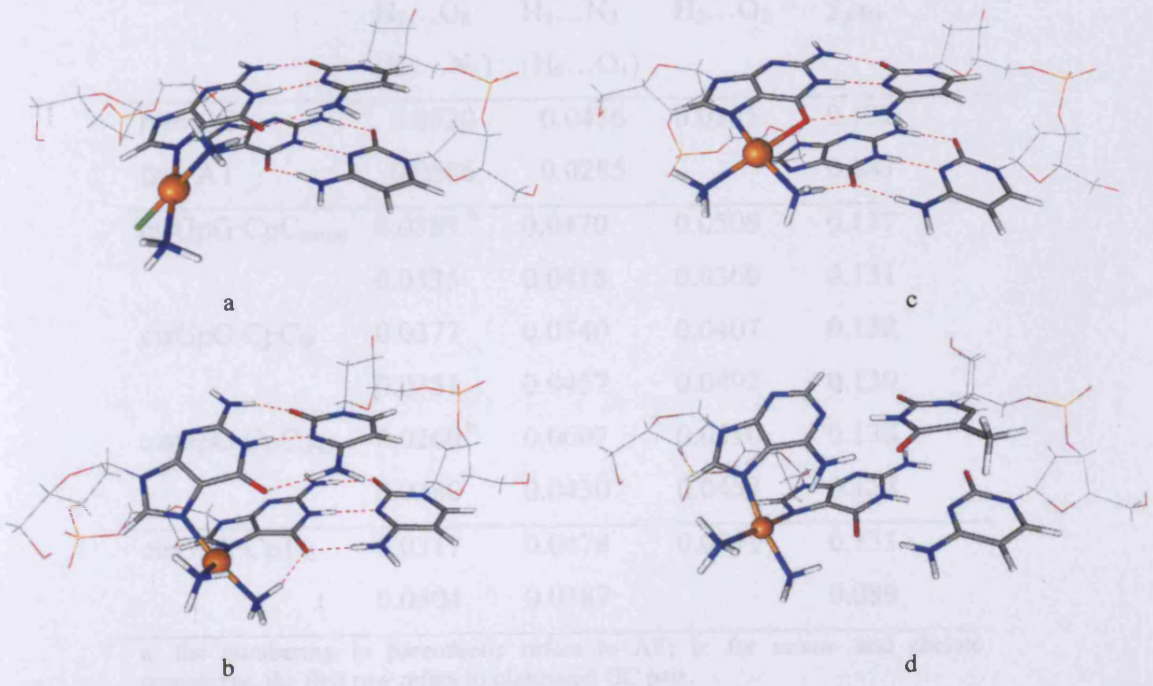


Figure 4.21 Optimised geometries of (a) cisGpG·CpC_{mono}, (b) cisGpG·CpC_{bi}, (c) cisGpG·CpC_{chel} and cisGpA·CpT_{bi}.

As well as stacking interactions, platination is known to affect the Watson-Crick pairing between G & C and A & T. Table 4.20 reports the electron density at the H-bond critical point for each interaction, which shows that platination weakens $H_4...O_6$ but strengthens $H_1...N_3$ and $H_2...O_2$, such that the overall electron density is hardly changed. This pattern is almost symmetrical in the bifunctional complex cisGpG·$CpC_{bi,}$ while in cisGpA·CpT, the H-bond in which adenine acts as a proton donor ($H_6...O_4$) is considerably stronger than in free AT, while where adenine is a proton acceptor ($H_3...N_1$) the H-bond is weakened. In cisGpG·CpC_{mono}, the non-platinated guanine's interaction with cytosine is barely affected by the presence of cisplatin, but in cisGpG·CpC_{chel} both GC pairs are affected, to the extent that the non-platinated GC is in fact the weakest found in this work.

Table 4.20 electron density (au) of GC pair in platinated duplexes.[a]

	$H_4...O_6$ ($H_3...N_1$)	$H_1...N_3$ ($H_6...O_4$)	$H_2...O_2$	$\Sigma\rho_{CP}$
free GC	0.0520	0.0436	0.0365	0.132
free AT	0.0586	0.0285		0.087
cisGpG·CpC_{mono}	0.0389 [b]	0.0470	0.0506	0.137
	0.0535	0.0418	0.0360	0.131
cisGpG·CpC_{bi}	0.0377	0.0540	0.0407	0.132
	0.0351	0.0457	0.0492	0.130
cisGpG·CpC_{chel}	0.0260 [b]	0.0607	0.0510	0.138
	0.0380	0.0430	0.0457	0.128
cisGpA·CpT_{bi}	0.0317	0.0478	0.0519	0.131
	0.0504	0.0387		0.089

a: the numbering in parenthesis refers to AT; b: for mono- and chelate complexes, the first row refers to platinated GC pair.

*F- Platinated d(CCTG*G*TCC)·d(GGACCAGG)*

In order to demonstrate what we believe to be the potential of the BH&H/AMBER/AIM approach, and to provide a better model of platination of DNA, we report calculations of larger scale complexes. Figure 4.22 shows the QM/MM optimised geometry of a cisplatin adduct of the octamer duplex d(CCTG*G*TCC)·d(GGACCAGG),[62] solvated by *ca.* 400 H_2O molecules. As shown in Figure 4.22, the QM region includes four bases, *i.e.* cisGpG·CpC_{bi} and cisplatin, with the remaining DNA bases, sugar-phosphate backbone, and water molecules treated using AMBER. The experimental NMR structure (PDB entry 1AU5) was used as the starting point for optimisation. The ability of AMBER to reproduce DNA structures is well-reported,[63] so our focus here is on the QM region.

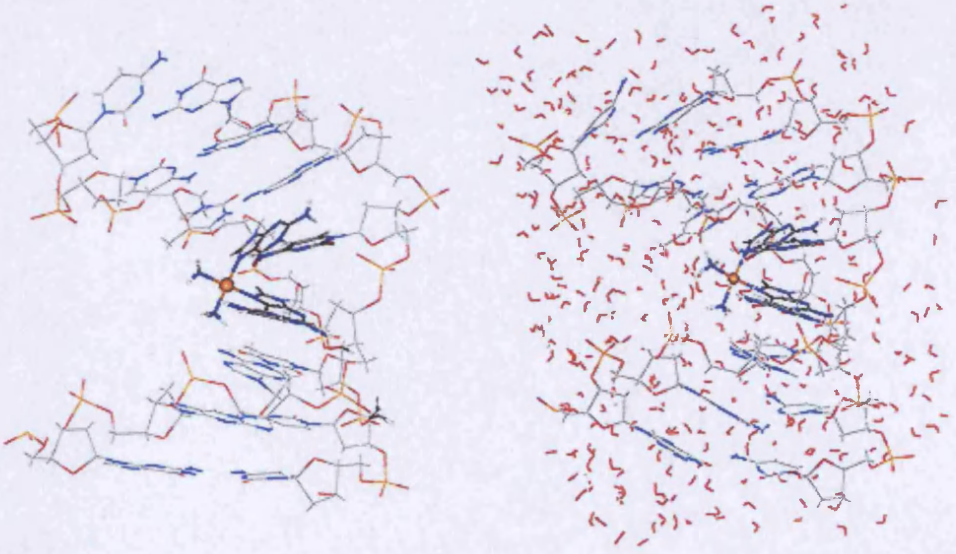


Figure 4.22 Experimental (a) and optimised (b) geometries of cisplatin-DNA adduct.

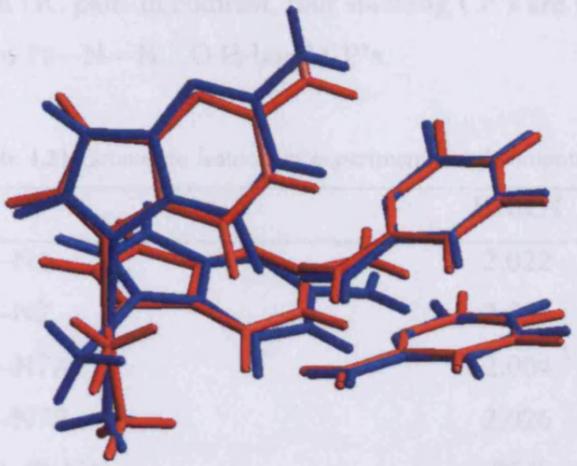


Figure 4.23 Overlay of optimised (red) and experimental (blue) platinated GpG·CpC.

Table 4.21 and Figure 4.23 indicate general agreement between optimised and NMR structures: bond lengths are slightly over-estimated in our calculations by between 0.02 and 0.05Å, while angles deviate by 2-6°. The dihedral angle between guanines, and its change from the model cisGpG complex (Table 4.15) is well reproduced, supporting our choice of an ONIOM: BH&H/AMBER method. The RMS deviation between calculated and optimised Cartesian coordinates is 2.12Å, which compares reasonably well with the values of 0.7 – 1.3Å quoted by Reedijk *et al.*[62] for differences between different refinements against NMR data, albeit for the entire octamer duplex structure.

We then evaluated the topology of electron density for both the experimental (cis_{EXP}) and calculated (cis_{QM}) geometries of platinated GpG·CpC within this octamer duplex: as noted above, inclusion of the MM region in these calculations makes essentially no difference in smaller duplexes, so here the QM region was extracted from the overall structure, link atoms replaced with hydrogen, and a single point DFT calculation carried out. Electron density at Pt—N bonds is similar in both structures (0.118 and 0.123 au in calculated structure *cf.* 0.127 and 0.133 au from experimental). Moreover, both structures contain secondary Pt…O interactions: two are present in the experimental structure but just one in the optimised geometry. Despite the similarity in geometries noted above, differences in the electron density of intermolecular interactions are more apparent: in the experimental geometry, just two CP's corresponding to π-stacking are found, along with the expected

three for each GC pair. In contrast, four stacking CP's are found in the optimised geometry, as well as two Pt—N—H...O H-bond CP's.

Table 4.21 Geometric features of experimental and computed Pt-coordination.[a]

	BH&H	Expt.[b]
Pt—N1	2.022	2.000
Pt—N2	2.030	1.987
Pt—N7A	2.004	1.984
Pt—N7B	2.026	1.963
N7A-Pt-N1	85.9	91.2
N7A-Pt-N2	172.8	177.6
N7A-Pt-N7B	89.2	87.4
N1-Pt-N2	95.2	91.2
N1-Pt-N7B	170.4	178.5
N2-Pt-N7B	88.3	90.2
Gua/Gua[c]	60.3	58.0

a: see Figure 4.16a for labelling; b: NMR data from Reedijk *et al.*;[62] c: see Orbell *et al.*[49] for convention of dihedral angles and Figure 4.16b.

The energetic consequences of this topology, and of differences between experimental and theoretical structures, are detailed in Table 4.22. Hydrogen bonds formed in the optimised structure contribute *ca.* 6-7 kcal/mol each to the stability of the complex, a similar figure to that found in model complexes. Stacking interactions between guanines is in both cases limited to a single interaction, corresponding to less than 2 kcal/mol, whereas stacking between cytosines is weaker than in smaller models, but slightly higher in the optimised structure. Inter-strand interactions are absent in the experimental structure and very weak in the optimised one. Although the effect on GC pairing follows the pattern established above, these effects are slightly more pronounced here than in smaller oligonucleotides: for instance $H_4...O_6$ bonds and $H_2...O_2$ are strongly perturbed between 30 and 50%, but the overall H-bond energy is reduced only about 2 kcal/mol from its original value.

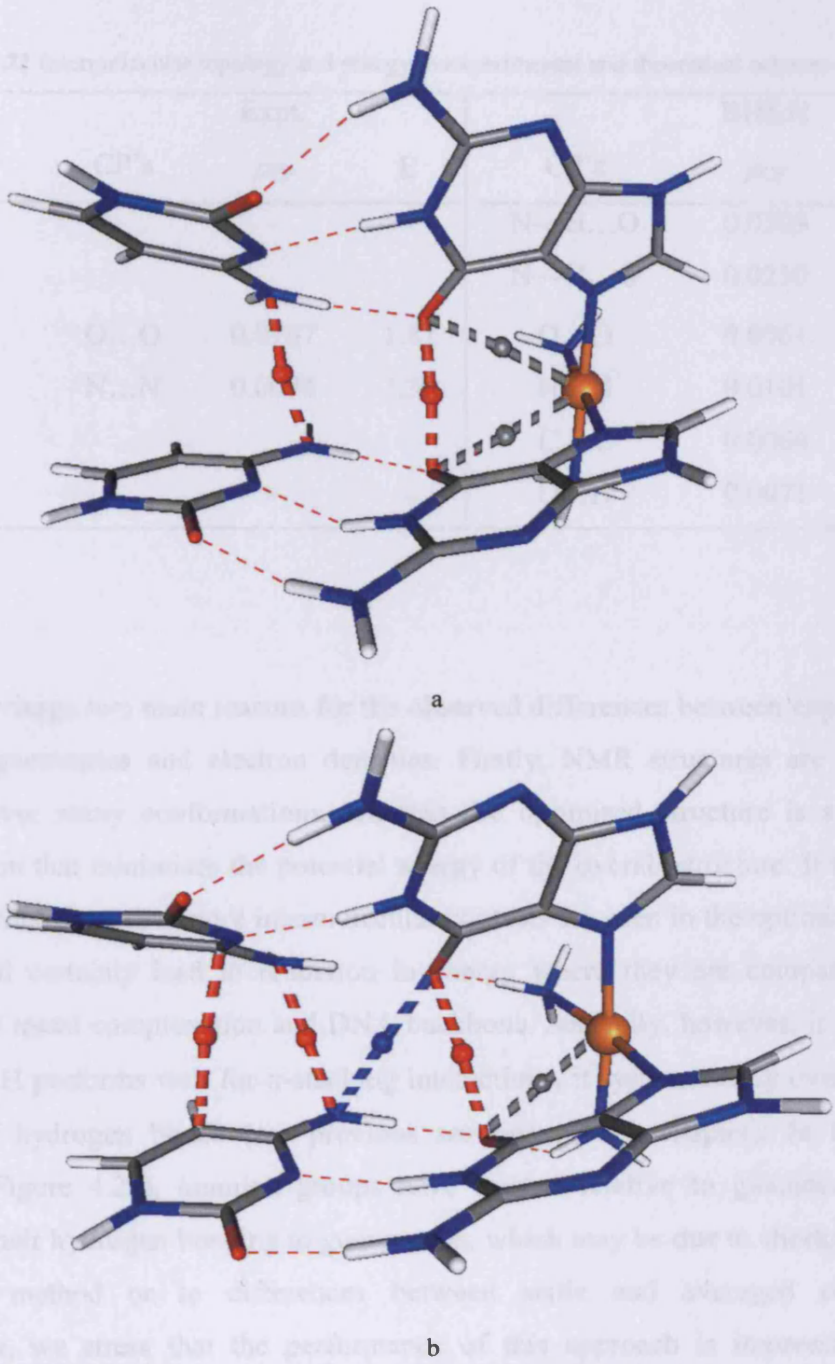


Figure 4.24 Representation of intermolecular interactions found in (a) experimental and (b) optimised geometry of platinated GpG·CpC from octamer complex.

Table 4.22 Intermolecular topology and energy in experimental and theoretical octamer structures.

	Expt.			BH&H		
	CP's	ρ_{CP}	E	CP's	ρ_{CP}	E
Pt—G	-	-	-	N—H...O	0.0309	7.65
				N—H...O	0.0210	5.90
GG_S	O...O	0.0107	1.87	O...O	0.0061	1.08
CC_S	N...N	0.0074	1.30	N...N	0.0101	2.87
				C...C	0.0064	
GC_{IS}	-	-	-	O...N	0.0071	1.23

One can envisage two main reasons for the observed differences between experimental and optimised geometries and electron densities. Firstly, NMR structures are by definition averaged over many conformations, whereas the optimised structure is a single static conformation that minimises the potential energy of the overall structure. It is perhaps not surprising, therefore, that more intermolecular contacts are seen in the optimised geometry, as they will certainly lead to reduction in energy where they are compatible with the demands of metal complexation and DNA backbone. Secondly, however, it is known that while BH&H performs well for π-stacking interactions, it systematically overestimates the strength of hydrogen bonds (see previous sections in this chapter). In the optimised structure (Figure 4.23), ammine groups have rotated relative to guanines in order to maximise their hydrogen bonding to guanine O6, which may be due to shortcomings in the theoretical method or to differences between static and averaged conformations. Nonetheless, we stress that the performance of this approach is impressive given the difficulty of simultaneous modelling of platination, hydrogen bonding and π-stacking.

Table 4.23 secondary Pt...N(O) interactions in platinated oligonucleotides.

		Pt...X	
		r (A)	ρ (au)
GpA_{mono}	Pt...N7	3.421	0.009
$GpGpG_{bi}$	Pt...O6	3.212	0.012
$GpGpG_{mono}$	Pt...N7	3.149	0.014
$GpG{\cdot}CpC_{bi}$	Pt...O6	3.026	0.017
$GpG{\cdot}CpC_{chel}$	Pt...N7	3.028	0.018
cis_{QM}	Pt...O6	3.042	0.017
cis_{EXP}	Pt...O6	3.311	0.010
	Pt...O6	3.412	0.008

G- Pt...N and Pt...O secondary interactions

Finally, throughout this work we have identified secondary interactions such as Pt...O and Pt...N *via* AIM analysis. Table 4.23 summarises all such interactions found, showing that such contacts are always longer than 3Å and rather weak, with ρ_{CP} between 0.008 and 0.017 au, while no clear difference between Pt...O and Pt...N interactions is apparent. Thus, any contribution to the stability of complexes will be small, but they might exert some influence on geometry, since our data suggests that these interactions are directed to the axial positions about platinum: such weak axial interactions have been noted before (see section 3.2) or, for instance, Kozelka's work.[64]

4.5 Concluding Remarks

4.5.1 BH&H and AIM applied to π-stacked systems

The hybrid BH&H functional combined with modest basis sets qualitatively reproduces the PES of higher-level calculations for a number of instances of π-stacking. Binding energies of complexes of substituted benzenes and pyridines, as well as pyrimidine and purine DNA

bases, are well reproduced. Our results are sufficiently encouraging to suggest that this method will allow application to many more examples of π-stacking. Specifically, the computational tractability, modest basis set requirements, and number of established computational packages for DFT mean that such a method will be a highly attractive alternative to post-SCF calculations in a great many fields. Given the interest in extending quantum chemical calculations to macromolecular systems, we suggest that this approach may be of great utility in, for example, studies of DNA oligonucleotides or metalloproteins, particularly as the QM level in hybrid QM/MM calculations.

Our data also suggest a good deal of promise for this DFT/AIM approach to analysis of stacked DNA bases. Firstly, fully unrestrained optimisation is feasible at this level, and allows the interplay of stacking and hydrogen bonding, this providing more realistic geometries than the frozen monomer approach. Then, AIM methods allow us to decompose the overall interaction into contributions from π-stacking and hydrogen bonding, yielding accurate estimates of binding energy in all cases considered.

4.5.2 H-bonding and π-stacking in gas-phase oligonucleotides

Combined BH&H/6-311++G(d,p) and AIM analysis has allowed study of the intermolecular forces and their mutual interplay in DNA chains and some model systems. Single stranded di- and tri-nucleotides show that the conformation adopted is connected to the number of and type of bases, with a balance of H-bonding and π-stacking needed to obtain regular structures. Trinucleotides in which the central base is cytosine or thymine have highly distorted structures, closer to "T-shaped" complexes rather than the more normal parallel stacking structure, in which two bases interact with the third *via* hydrogen bonds and C—H…π interactions. $-NH_2$ groups play an important role, involved in both H-bonds and π-stacking, and are found to be significantly non-planar in many structures.

Furthermore, the interplay of π-stacking and H-bonding was explored: simple models such as benzene/guanine/cytosine confirm that benzene molecules can modulate H-bonding capacity, leading to distortion in the GC pair, but barely perturbing the overall binding energy. In DNA duplexes, Watson-Crick pairing of GC and AT is hardly affected by

stacking partners, in accord with literature. Base pairing also leads to increased planarity of $-NH_2$ groups, which now interact mainly *via* π-stacking rather than H-bonding, leading overall gain of π-stacking energy of 1-2 kcal/mol per stacked pair, compared with single stranded chains. In such studies, AIM analysis is particularly useful, as the intricacy of inter- and intra-strand interactions means that pairwise analysis of base-base interactions inevitably ignores the wider environment. Where comparison is possible, energies and structures at least qualitatively match experimental and theoretical data reported in literature.

4.5.3 H-bonding and π-stacking in platinated DNA structures

The combination of BH&H/6-311++G(d,p) and AIM analysis has allowed us to investigate the role of covalent and intermolecular forces in cisplatin-DNA adducts. Comparison with experimental geometries was found to be satisfactory for both cisplatin itself and its complexes with guanine. The interaction of cisplatin with single-stranded DNA follows the pattern established experimentally, *i.e.* complexes to guanine are more stable than those with adenine. Interactions of cisplatin's ammine and chloride groups, including N—H...Cl, Pt—N—H...O and Pt—N—H...N dominate H-bond energies, and contribute significantly to overall stabilisation. Both mono- and bifunctional complexation induces strong distortion: for instance, bifunctional cisplatin-DNA complexes show major disruption of π-stacking between the bases bound to the metal. Complexes of cisplatin with DNA duplexes were also studied in order to monitor the effect of platination on both H-bonding and π-stacking. Intramolecular H-bonds and covalent Pt—N bonds are close to single stranded complexes and the effect on GC Watson-Crick pair is similar to that found in simple models such as platinated GC pair: the pattern of stabilisation is altered, but the overall stability of GC is virtually unchanged.

We have also presented data on a realistic model, namely the platinated octamer cis[d(CCTG*G*TCC)·d(GGACCAGG)], for which NMR structural data is available. QM/MM calculations reproduced the experimental structure at the platinated GpG·CpC core, with bond lengths and angles within *ca.* 0.04 Å and 4° of experimental values on average, respectively. AIM analysis shows that π-stacking interactions are seriously

disrupted by platination, being reduced by more than 80% compared to non-platinated structures. The H-bonding pattern in the GC pair is affected in a similar manner as smaller oligonucleotides, although the effect is more pronounced in the octamer structure. AIM reveals secondary Pt...O6 in both experimental and computed geometries: more studies are needed in order to clarify any biological relevance of such interactions.

4.6 References

(1) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-1377.

(2) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789.

(3) Perezjorda, J. M.; Becke, A. D. *Chem. Phys. Lett.* **1995**, *233*, 134-137.

(4) Hehre, W. J.; Ditchfie.R; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257-&.

(5) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265-3269.

(6) D. E. Woon, T. H. D. J. *J. Chem. Phys.* **1993**, *98*, 1358.

(7) R. A. Kendall, T. H. D. J., and R. J. Harrison *J. Chem. Phys* **1992**, *96*, 6796.

(8) Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673-2677.

(9) 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.

(10) Sponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76-87.

(11) Boys, S. F.; Bernardi, F. *Mol. Phys* **1970**, *19*, 553.

(12) Jaffe, R. L.; Smith, G. D. *J. Chem. Phys.* **1996**, *105*, 2780-2788.

(13) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334-338.

(14) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc* **2002**, *124*, 10887-10893.

(15) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc* **2004**, *126*, 7690-7697.

(16) Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624-1626.

(17) Mignon, P.; Loverix, S.; De Proft, F.; Geerlings, P. *J. Phys. Chem. A* **2004**, *108*, 6038-6044.

(18) Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802-11808.

(19) Leininger, M. L.; Nielsen, I. M. B.; Colvin, M. E.; Janssen, C. L. *J. Phys. Chem. A* **2002**, *106*, 3850-3854.

(20) Bader, R. F. W. *Atoms in Molecules-A Quantum Theory*; Oxford, University Press: Oxford, 1990.

(21) Pettersson, I.; Liljefors, T. *J. Comput. Chem.* **1987**, *8*, 1139-1145.

(22) Cubero, E.; Orozco, M.; Luque, F. J. *J. Phys. Chem. A* **1999**, *103*, 315-321.

(23) Zhikol, O. A.; Shishkin, O. V.; Lyssenko, K. A.; Leszczynski, J. *J. Chem. Phys.* **2005**, *122*, art. no.-144104.

(24) Kabelac, M.; Hobza, P. *J. Phys. Chem. B* **2001**, *105*, 5804-5817.

(25) London, F. *Zeitschrift fur Physik* **1930**, *63*, 1842.

(26) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Mol. Struct.: THEOCHEM* **2001**, *573*, 43-53.

(27) Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608-15613.

(28) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc* **2000**, *122*, 4117-4128.

(29) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170-1179.

(30) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357-19363.

(31) Svensson, M.; Humbel, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 3654-3661.

(32) Matsubara, T.; Sieber, S.; Morokuma, K. *Int. J. Quantum Chem.* **1996**, *60*, 1101-1109.

(33) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959-1967.

(34) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc* **1995**, *117*, 5179-5197.

(35) Burda, J. V.; Sponer, J.; Leszczynski, J. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4404-4411.

(36) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247-3276.

(37) Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. *J. Comput. Chem.* **2005**, *in press*.

(38) Mignon, P.; Loverix, S.; Steyaert, J.; Geerlings, P. *Nucleic Acids Res.* **2005**, *33*, 1779-1789.

(39) Gidden, J.; Bushnell, J. E.; Bowers, M. T. *J. Am. Chem. Soc* **2001**, *123*, 5610-5611.

(40) Dabkowska, I.; Gonzalez, H. V.; Jurecka, P.; Hobza, P. *J. Phys. Chem. A* **2005**, *109*, 1131-1136.

(41) Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1997**, *18*, 1136-1150.

(42) Ganem, B.; Li, Y. T.; Henion, J. D. *Tetrahedron Lett.* **1993**, *34*, 1445-1448.

(43) Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 8007-8014.

(44) Andrae, D.; Haussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123-141.

(45) Milburn, G. H. W.; Truter, M. R. *J. Chem. Soc. A* **1966**, 1609.

(46) Pavankumar, P. N. V.; Seetharamulu, P.; Yao, S.; Saxe, J. D.; Reddy, D. G.; Hausheer, F. H. *J. Comput. Chem.* **1999**, *20*, 365-382.

(47) Carloni, P.; Andreoni, W.; Hutter, J.; Curioni, A.; Giannozzi, P.; Parrinello, M. *Chem. Phys. Lett.* **1995**, *234*, 50-56.

(48) Sherman, S. E.; Gibson, D.; Wang, A. H. J.; Lippard, S. J. *J. Am. Chem. Soc.* **1988**, *110*, 7368-7381.

(49) Orbell, J. D.; Marzilli, L. G.; Kistenmacher, T. J. *J. Am. Chem. Soc* **1981**, *103*, 5126-5133.

(50) Burda, J. V.; Leszczynski, J. *Inorg. Chem.* **2003**, *42*, 7162-7172.

(51) Eastman, A. *Biochemistry* **1986**, *25*, 3912-3915.

(52) Fichtingerschepman, A. M. J.; Vanderveer, J. L.; Denhartog, J. H. J.; Lohman, P. H. M.; Reedijk, J. *Biochemistry* **1985**, *24*, 707-713.

(53) Davies, M. S.; Berners-Price, S. J.; Hambley, T. W. *J. Am. Chem. Soc* **1998**, *120*, 11380-11390.

(54) Zilberberg, I. L.; Avdeev, V. I.; Zhidomirov, G. M. *J. Mol. Struct.: THEOCHEM* **1997**, *418*, 73-81.

(55) Pelmenschikov, A.; Zilberberg, I.; Leszczynski, J.; Famulari, A.; Sironi, M.; Raimondi, M. *Chem. Phys. Lett.* **1999**, *314*, 496-500.

(56) Cohen, G. L.; Bauer, W. R.; Barton, J. K.; Lippard, S. J. *Science* **1979**, *203*, 1014-1016.

(57) Sip, M.; Schwartz, A.; Vovelle, F.; Ptak, M.; Leng, M. *Biochemistry* **1992**, *31*, 2508-2513.

(58) Herman, F.; Kozelka, J.; Stoven, V.; Guittet, E.; Girault, J. P.; Huynhdinh, T.; Igolen, J.; Lallemand, J. Y.; Chottard, J. C. *Eur. J. Biochem.* **1990**, *194*, 119-133.

(59) Rice, J. A.; Crothers, D. M.; Pinto, A. L.; Lippard, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 4158-4161.

(60) Denhartog, J. H. J.; Altona, C.; Vanboom, J. H.; Vandermarel, G. A.; Haasnoot, C. A. G.; Reedijk, J. *J. Biomol. Struct. Dyn.* **1985**, *2*, 1137-1155.

(61) Denhartog, J. H. J.; Altona, C.; Vandermarel, G. A.; Reedijk, J. *Eur. J. Biochem.* **1985**, *147*, 371-379.

(62) Yang, D. Z.; Vanboom, S.; Reedijk, J.; Vanboom, J. H.; Wang, A. H. J. *Biochemistry* **1995**, *34*, 12912-12920.

(63) Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350-364.

(64) Kozelka, J.; Bergès, J.; Attias, R.; L, F. *Angew. Chem. Int. Edit. Engl.* **2000**, *39*, 198-201.

GENERAL CONCLUSIONS

DFT calculations were applied in order to provide a model of first solvation sphere of cisplatin: after preliminary calculations on 1:1 cisplatin:water complexes, the effect of ten explicit water molecules was studied, yielding a slightly better energy barrier for the first hydrolysis of cisplatin compared to previous studies; also large geometric changes on the mechanism were found. Atoms in molecules (AIM) theory was extensively employed, confirming the relevance of electron density analysis for such systems. The HF/DFT level of theory was then used to investigate features of platinum-purine complexes and the effect of platination on the GC pair. An AIM based method for estimating single H-bond energies applicable to inorganic and organic complexes was also proposed. This approach allowed us to investigate the role of covalent and H-bond energies in platinated models of GC: platination on N7 of guanine weakens $H_4...O_6$ bonds but strengthens $H_1...N_3$ and $H_2...O_2$, leading to large changes in the geometry of the GC pair, but only small differences in the total binding energy. On the other hand, platination at O6 or chelation to N7 and O6 destroys the normal Watson-Crick pattern of H-bonding, though substantial pairing energy remains. A systematic analysis of all transition metals, from Ti to Hg revealed similar effects of metallation over the GC pair, with early metals preferring the O6 position, whereas more electron rich metals prefer N7. The distortion of GC pair was explained in terms of electrostatics and electron density redistributions within guanine: the combination of both, for instance in Mt—O6 complexes, leads to a strong loss of H-bond donor capability of O6.

Despite the importance of simple systems, larger scale models are needed in order to analyse the effect of metallation on the interplay between intermolecular forces, such as hydrogen bonding and π-stacking. Post-HF calculations are required in order to describe π-stack interactions, but current methods are unfeasible for such large systems. We proposed a new method based on Becke's half-and-half functional, in order to avoid such

obstacles. The hybrid functional was tested for more than 50 molecules, including substituted benzenes, pyridines and DNA bases. These data were compared to MP2 and CCSD calculations wherever available, confirming that BH&H qualitatively reproduces high-level results. AIM methods were developed to decompose total binding energy into contributions from hydrogen bonding and π-stacking, allowing us to recognize the source stabilisation in several DNA adducts of interest. QM/MM (QM = BH&H and MM = Amber) calculation of gas-phase single stranded di- and tri-nucleotides confirmed several experimental findings, suggesting that a) the conformation adopted depends to the number of and type of bases; b) a balance of H-bonding and π-stacking resulted in regular structures, whereas a lack of that led to highly distorted conformations; b) as shown in small models, $-NH_2$ groups play an important role, as they interact both via H-bonds and π-stacking, being significantly non-planar in many structures.

Similar methods were then applied to several platinated oligonucleotides. Where possible, the comparison between experimental and computed structures was found satisfactory confirming the validity of our QM/MM calculation based on the BH&H functional. Also several experimental facts were confirmed by our calculation: a) the trend of binding energies, the complexes with guanine being generally more stable than those with adenine; b) the disruption of π-stacking between platinated nucleobases; c) the distortion of GC pair; d) the relevance of H-bonds such N—H...O involving cisplatin's NH_3 groups. In addition, one example of a platinated octamer for which experimental structure is available, was presented. Our QM/MM calculations reproduced reasonably well the experimental NMR structure. As in smaller models, platination led to disruption of π-stacking interactions with a loss of energy of 80% and large redistribution of energy in GC pair, but small changes in the overall pairing energy.