UPTEC X 01 018 ISSN 1401-2138
APR 2001

JOHAN RABER

A theoretical study of the interactions between anti-cancer drug cisplatin and DNA.

Master's degree project

Molecular Biotechnology Programme
Uppsala University School of Engineering

UPTEC X 01 018	**Date of issue 2001-04**

Author

Johan Raber

Title (English)

A Theoretical Study of the Interactions Between Anti-cancer Drug Cisplatin and DNA.

Title (Swedish)

Abstract

This work is aimed at qualitatively elucidating the important chemical interactions between anti-cancer drug cisplatin and its target DNA by means of the Quantum Chemical method Density Functional Theory (DFT). The steps prior to cisplatins bonding to DNA have also been investigated and are shown to be feasible from a Quantum Chemical point of view.

Keywords

Cisplatin, DFT, Quantum Chemistry, DNA, reaction path, transition state.

Supervisors

Leif A. Eriksson
Department of Biochemistry, Uppsala University

Examiner

David van der Spoel
Department of Biochemistry, Uppsala University

Project name	Sponsors
Language **English**	Security 2000-12
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **56**

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

Contents

Introduction

Cisplatin, or *cis*-diamminedichloroplatinum(II) (*cis*-DDP), is a potent anti-cancer drug especially effective against tumours in the sex glands, head and neck[1]. The most likely target of the drug, correlating to anti-tumour activity, has in a number of investigations been shown to be cellular DNA[2], distorting the tertiary structure of DNA and thereby inhibiting the replication and transcription machinery of the cell[9,10,11]. Recently the persistent nature of the inhibitory effect has been shown, in all likelihood, to be a masking of the cisplatin induced damage by means of indigenous proteins of the cell nucleus[3], the so-called HMG-box containing proteins. Non-tumour cells are not as affected by this damage as the tumour cells because the repair system of normal cells is working properly, and their metabolic rate is not elevated as in transformed cells. As with most cytotoxins, this is the basis of its action, i.e. the toxicity of the drug is higher for the transformed cells than for non-transformed.

Cisplatin was discovered by coincidence in 1965 by Rosenberg *et al*. and is, in spite of its simple structure, one of the most potent anti-cancer drugs known to date. Even though it has been around for over thirty years, very little is known about the reasons of its efficacy, and virtually none of the modifications done to cisplatin has improved its performance versus cancer[4]. Experiments so far have revealed only the structure of some products, rudimentary data on kinetics and which products are formed, so a theoretical investigation on the mechanism of its action is well in its place.

Cisplatin is a neutral, square planar compound with Platinum in the centre of the square, coordinating four ligands, two ammine groups and two chloride groups in a *cis*-conformation. The structural isomer *trans*-Platin also attacks DNA, bonding to the same bases as cisplatin, but has not shown any clinical activity versus cancer cells[5]. The reason for this clearly lies in the structural differences of the two isomers, fig. 1, and thereby the structural differences induced at the site of platination in DNA.

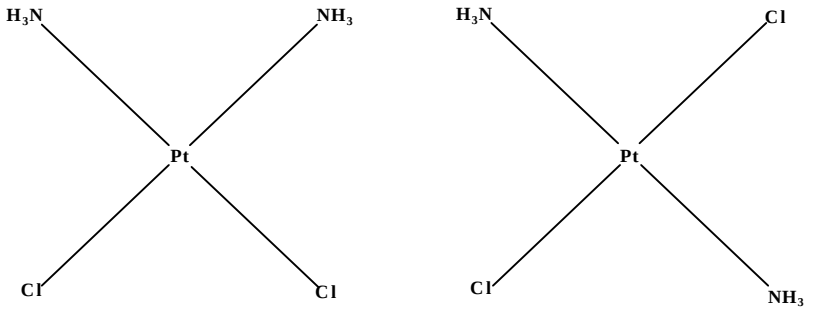


Figure 1. The two geometrical isomers *cis*-DDP and *trans*-DDP respectively.

Platinum is in its 2+-oxidation state and has the d^8 configuration of its valence shell. It forms complexes with ligands in a substitution reaction with the substituting ligand entering from either side of the plane toward the Platinum centre forming a trigonal bipyramid as a transition state, according to textbook theory[6,7]. Both entering and leaving ligand in the transition state are coordinated in an equatorial position of the bipyramid and in the subsequent step the leaving ligand is released on the other side of the original square compound (fig. 2).

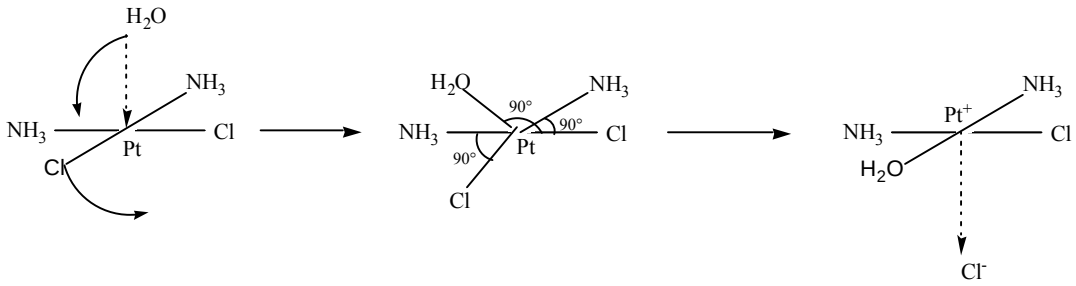


Figure 2. Idealized reaction mechanism of the first step in the aquation of cisplatin.

Several adducts[8] can be formed by cisplatin to DNA, the major ones being the intrastrand (i.e. on the same strand) adjacent 5'-GG adduct, intrastrand adjacent 5'-AG and non-adjacent intrastrand didentate adduct GXG (X = any base), at 65%, 25%, 6% percent respectively. The remaining part consists of cisplatin monofunctionally bound to G and interstrand (i.e. between the two strands) bifunctional adducts G-G at ~3% (fig. 3).

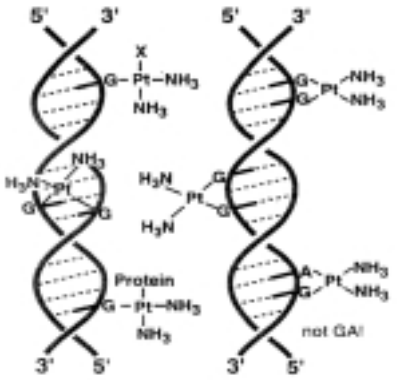


Figure 3. Schematic representation of the different adducts formed by cisplatin. Figure from ref. [5].

DNA itself offers a few leads as to why adducts are formed at certain positions of the purine bases Guanine (G) and Adenine (A). The adducts are, according to experiments[8, formed exclusively at the N7 position of the purine bases A and G exposed in the major groove of the DNA helix (fig. 4), and as can be seen in the figure the other possible sites of platination are all exposed in the minor groove or are involved in the 'base-pairing' of the two strands.

The conceivable platination sites (other than those of the major groove) would thus probably offer more sterical hindrance for the different steps of the reaction, since the width of the minor groove typically only ranges between 4-6 Å depending on the base-pair sequence. Another important feature of the DNA-helix to be considered is the twist. DNA in its native B-form twists around its axis 360° in ten base-pairs (~34 Å in length) and this renders bases with the same neighbours different chemical environments. For instance, in the sequence 5'-AGGA-3', the two G's will experience a different chemical environment even though they have the same neighbouring bases, although in a different order from the individual base's point of view.

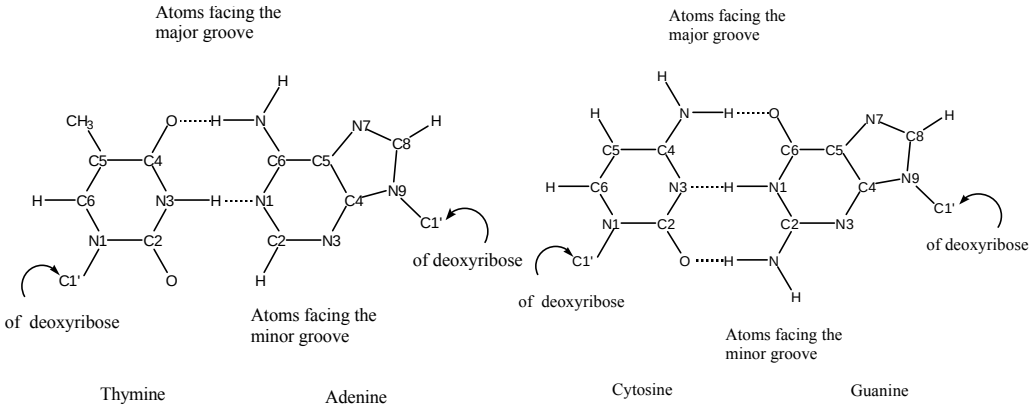


Figure 4. Base pairing of the DNA bases.

This work will focus on the different steps in the formation of the two most abundant adducts, intrastrand 5'-GG and 5'-AG (or rather, a model system thereof). A few notes are in place:

- The AG adduct is direction specific, 5'-AG-3'. No exception to this directionality has been found[8].
- No monofunctional adducts to A has been detected[8].
- The bond to the 3' base of the dimer is weaker than the 5' base bond[12].
- The bonds are subject to breaking due to water substitution. Especially 5'-GG interstrand bonds have a well ordered 'water cage' around the site of the lesion, with water molecules in a favourable position for a nucleophilic attack on the Platinum coordination centre[13].
- There is evidence that cisplatin does not enter the nucleus in its chlorinated form but rather as the doubly water substituted form $Pt[NH_3]_2[H_2O]_2^{2+}$[14].

Part of the work has dealt with revealing the energetics of the double water substitution of cisplatin yielding the Platinum complex indicated in the last item of the above list.

Structural changes in DNA due to bonding of cisplatin

The damage in tertiary structure of DNA induced by cisplatin depends on which type of adduct is formed. Only two types of structures have to my knowledge been determined by means of X-ray crystallography or NMR, the 5'-GG intrastrand adduct and the GG interstrand adduct. In these cases, however, there is a wealth of papers on their respective structures[13,15,16]. However, there is good reason to believe that 5'-AG intrastrand adducts show close structural similarity to the GG counterpart. This is mentioned because the reaction mechanism proposed in this work is correlated to experimental data of intrastrand GG adducts but parallels are made to the AG adduct as well. These parallels are by no means supported by experimental data but rather constitute 'educated guesses' founded on calculations.

The structure of the intrastrand 5'-GG adduct.

The formation of this adduct disrupts the helical structure by de-stacking the two adjacent base pairs and locally unwinding DNA at the site of the lesion, thus creating a hydrophobic pocket facing the minor groove, which is widened and flattened, see fig. 6. As a consequence, a kink in the helix axis towards the major groove is introduced over the two consecutive base pairs, a kink whose value is measured as the deviation from the native, linear state of the DNA molecule (i.e. kink angle = 0). The value of the kink angle varies in different studies depending on the methods used (e.g. NMR, X-ray-crystallography) but also on the sequence context and the length of the fragment used in the experiment. In general:

- The kink angle value increases the shorter the investigated fragment is.
- Higher values of the kink angle are obtained if the adduct is flanked by less 'rigid' base pairs (i.e. A-T base pairs).
- NMR studies give higher values of the kink angle than those from X-ray crystallography. Probably due to packing interactions between neighbouring molecules in the crystal not present in NMR experiments.

However, an average value of the kink angle is ~50° in crystal structures and ~70-80° in NMR. In crystal structures the didentate adduct causes a local unwinding of the DNA chain changing the

puckering of the 5' base ribose of the adduct to the C2'-endo conformation, thus resembling A-form DNA usually found in highly desiccated DNA, instead of the normal B-form (fig.5). This change does not occur in the 3' ribose which retains its native C3'-endo form[15]. The local unwinding of the helix is also present in comparative NMR studies but here the overall structure of the fragment has retained its native B-form, albeit with a kink present.

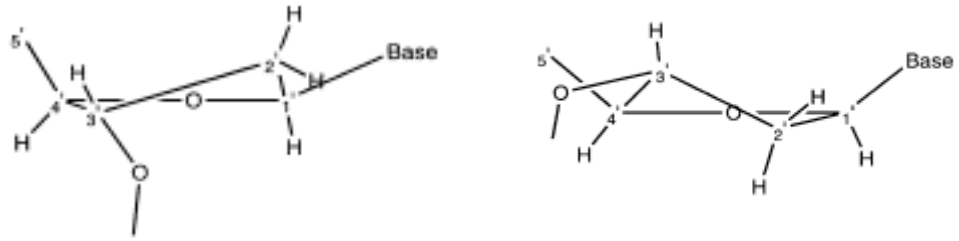


Figure 5. C2'-endo (left) and C3'-endo (right) conformations of ribose.

These distortions put stress on the adduct, displacing the centrally coordinated Platinum out of the planes of the bases by about 0.8 Å, 1.0 Å, crystal and NMR structures respectively, and places the top ammine group within hydrogen bond distance to one of the oxygen atoms of the backbone phosphate group.

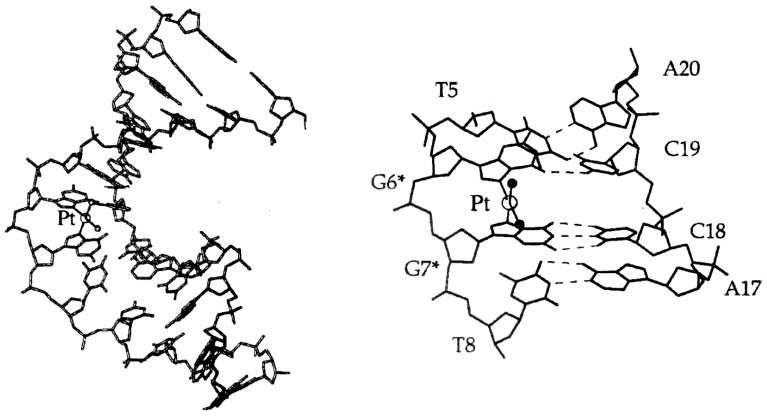


Figure 6. NMR determined structure of a platinated dodecamer duplex DNA (left) and a close up of the induced damage (right). The kink is clearly visible and centred around the site of platination. Picture from ref. [15].

The structure of the interstrand GG adduct.

Relatively few structural studies have been published on this type of adduct[13,17,18], however the data presented show a significant difference in the structural distortions compared to the intrastrand adduct. The prime feature of this adduct is the cross-linking between the two strands at GC sequences, thereby causing a kink in the double helix. In this instance though, the kink is towards the minor groove, with a value of ~47° (fig. 7).

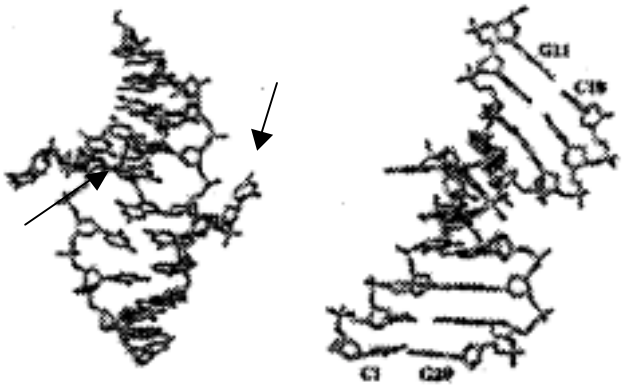


Figure 7. View from the minor groove of the interstrand lesion caused by cisplatin (left). Note the complementary cytosine bases protruding from the duplex. The figure on the right is the same structure rotated 90^{o} around the axis of the duplex. Picture taken from ref. [13]

Another feature of this adduct not present in the intrastrand adduct is the complementary cytosine bases extruding from the lesion site.

Like in the intrastrand case, stress is put on the adduct, forcing the platinum out of the plane of the bases by 0.3 and 0.6 Å, respectively. The excellent crystallographic data of one reference[13 revealed a very well ordered water structure around the site of the platinum lesion. Two water molecules in particular were shown to be located on the quaternary axis of the platinum square, well positioned for a nucleophile attack on platinum. This may account for the relative instability (compared with the intrastrand adduct) of this lesion and hence accounting for the low ratio of this adduct out of the total amount of platinum adducts, although this seems to be mostly speculations by the authors of ref. [13].

Theory

Introduction

In the following sections, a brief outline of the *ab initio* and Density Functional Theory (DFT) methods in Quantum Chemistry will be given. Neither of these is by any means complete or comprehensive, but is merely intended to serve as a starting point to highlight the fundamental differences and similarities of these approaches to solve the Schrödinger equation for many-electron systems (e.g. non-hydrogen atoms and molecules). The interested reader can consult the vast flora of textbooks[22,23] available on the subject for more detailed treatises.

Also, an outline of the theory for the chemical problem at hand is included as well as a description of the theoretical model for incorporating solvent effects in Quantum Chemistry calculations.

The Schrödinger equation

The general form of the time-independent, non-relativistic equation for N nuclei and n electrons is

$$\hat{H}\left|\Psi(\mathbf{r},\mathbf{R})\right\rangle = E\left|\Psi(\mathbf{r},\mathbf{R})\right\rangle, \qquad (1)$$

which is an eigenvalue equation, where $\left|\Psi(\mathbf{r},\mathbf{R})\right\rangle$ is a wave function depending on the positions **R** of the N nuclei and the positions **r** of the n electrons. E denotes the energy eigenvalue of the equation. $\hat{H}$ is the Hamiltonian operator:

$$\hat{H} = \hat{T} + \hat{V} = -\frac{1}{2}\left(\sum_{I}^{N}\frac{\nabla^2{}_I}{M_I} + \sum_{i}^{n}\nabla^2{}_i\right) + \left(-\sum_{I}^{N}\sum_{i}^{n}Z_I r_{Ii}^{-1} + \sum_{i<j}^{n} r_{ij}^{-1} + \sum_{I<J}^{N} Z_I Z_J R_{IJ}^{-1}\right) \qquad (2)$$

in atomic units (e.g. the electron charge and mass equalling unity). $\hat{T}$ and $\hat{V}$ are the kinetic and potential energy operators respectively. M_I is the mass of nucleus I. ∇_i^2 is a differential operator acting on electron i and ∇_I^2 is the corresponding operator acting on nucleus I. $Z_{I,J}$ is the charge of the different nuclei involved and R_{IJ}, r_{ij}, r_{Ii} denotes nucleus-nucleus, electron-electron and electron-nucleus distances respectively. Hence, the first part of the right hand side of eqn. (2) corresponds to the kinetic energy operator $\hat{T}$ of the electrons and nuclei and the second corresponds to the potential energy operator $\hat{V}$ of the system. The first term of the potential energy operator is the attractive interaction between electrons and nuclei (i.e. it lowers the energy E in eqn. (1)). The second and third term denotes the repulsion between electron-electron and nuclei-nuclei pairs respectively.

The solution of eqn. (1), in this formulation, is a daunting task indeed but, fortunately, an approximation that reduces the complexity of eqn. (1) considerably has been shown to be valid, namely the Born-Oppenheimer approximation.

The Born-Oppenheimer approximation

By noting the fact that the electron mass is many orders of magnitude smaller than the mass of the nuclei (e.g. $m_p \sim 1836\ m_e$), Born and Oppenheimer showed[19] that one can assume that the variations in the electronic wave function, Ψ_{el}, are small with respect to the nuclear wave functions, so that the first and second derivatives can be neglected. Another way of putting it is that the electrons work on a much faster timescale than the nuclei and can hence adapt to new positions of the nuclei very quickly, thereby always maintaining an equilibrium position in relation to the nuclei. Mathematically this means that (1) can be separated in two parts, a nuclear and an electronic:

$$\hat{H}_{nuc}\left|\Psi_{nuc}(\mathbf{R})\right\rangle = E_{nuc}\left|\Psi_{nuc}(\mathbf{R})\right\rangle \quad (3)$$

and

$$\hat{H}_{el}\left|\Psi_{el}(\mathbf{r},\mathbf{R})\right\rangle = E_{el}(\mathbf{R})\left|\Psi_{el}(\mathbf{r},\mathbf{R})\right\rangle \quad (4)$$

The nuclear and electronic Hamiltonian are reformulated from (2) as:

$$\hat{H}_{nuc} = -\frac{1}{2}\sum_{I}^{N}\frac{\nabla_I^2}{M_I} + E_{el}(\mathbf{R}) + \sum_{I<J}^{N} Z_I Z_J \mathbf{R}_{IJ}^{-1}$$
$$\hat{H}_{el} = -\frac{1}{2}\sum_{i}^{n}\nabla_i^2 - \sum_{i}^{n}\sum_{I}^{N} Z_I \mathbf{r}_{Ii}^{-1} + \sum_{i}^{n}\sum_{j>i}^{n}\mathbf{r}_{ij}^{-1} \qquad (5)$$

The electronic part of the wave function will now only depend on the geometrical arrangement of the nuclei and not on the actual wave functions of the nuclei, i.e. the nuclei are considered fixed. Under this approximation, the electronic energy is obtained as:

$$E_{el}(\mathbf{R}_{fixed}) = \langle \Psi_{el} | \hat{H}_{el} | \Psi_{el} \rangle \qquad (6)$$

The total energy of the system under study is then given by adding the nuclear repulsion term $\sum_{I<J}^{N} Z_I Z_J \mathbf{R}_{IJ}^{-1}$, to the electronic energy.

Simplifying the problem (1) by means of the Born-Oppenheimer approximation above still leaves the problem of finding approximate solutions to eqn. (4), since there are no analytical solutions to this problem. One way of addressing this is described in the following section.

The Hartree-Fock equations

The derivation of the HF-equations (and many other equations in quantum mechanics) relies heavily on the variational principle and states that, for any trial wave function

$$E_{trial} = \frac{\langle \Psi_{trial} | \hat{H}_{el} | \Psi_{trial} \rangle}{\langle \Psi_{trial} | \Psi_{trial} \rangle} \geq \frac{\langle \Psi_0 | \hat{H}_{el} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle} = E_0 \qquad (7)$$

Where Ψ_{trial} is a trial wave function giving the trial energy E_{trial}, which is always larger, or at best equal to, E_0, the true energy of the system.

The HF-equations constitute a one-determinant approximate solution to the Schrödinger equation for many electron systems, developed by Hartree[20] and later improved by Fock[21]. The method can be described as an independent particle model since it treats each electron individually, as if moving through a 'mean potential' created by the other electrons of the system. Each electron of the system thus has its own single particle wave function (i.e. orbital).

The single-determinant wave function is commonly represented as a Slater determinant

$$|\Psi\rangle = \frac{1}{\sqrt{n!}}\begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_n(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & & \vdots \\ \vdots & & \ddots & \\ \chi_1(\mathbf{x}_n) & \cdots & & \chi_n(\mathbf{x}_n) \end{vmatrix} \tag{8}$$

in which the spin-orbitals $\chi_i(\mathbf{x})$ are products of a spatial orbital, $\psi(\mathbf{x})$, and a spin function $\alpha(\omega)$ or $\beta(\omega)$, spin up or spin down. The factor $1/\sqrt{n!}$ is a normalization constant whose particular form is due to the fact that an unfolding of the determinant has n! terms. The determinant representation is convenient since it fulfils the anti-symmetry condition and the Pauli principle of the wave function, i.e. switching places of two electrons (rows) changes the sign of the wave function and two identical sets of quantum numbers returns a null valued wave function (i.e. non-existing).

The exact derivation of the HF-equations will not be covered here, but an interpretation of them is in place. Provided that the spin-orbitals are orthonormal, $\langle \chi_i | \chi_j \rangle = \delta_{ij}$, and the Slater determinant is normalized, the energy expectation value of the electronic Hamiltonian takes this form

$$\begin{aligned} E' &= \langle \Psi' | \hat{H}_{el} | \Psi' \rangle = \langle \Psi' | - \sum_i^n \frac{1}{2}\nabla_i^2 - \sum_i^n \sum_I^N \frac{Z_I}{\mathbf{r}_{iI}} + \sum_i^n \sum_{j>i}^n \frac{1}{\mathbf{r}_{ij}} | \Psi' \rangle \\ &= \sum_i \langle \chi_i | \hat{h}_i | \chi_i \rangle + \frac{1}{2}\sum_{ij} \left(J_{ij} - K_{ij} \right) = \sum_i \langle i | \hat{h}_i | i \rangle + \frac{1}{2}\sum_{ij} \left(\langle ij | ij \rangle - \langle ij | ji \rangle \right) \end{aligned} \tag{9}$$

The operator $\hat{h}_i$ is the one-electron core Hamiltonian of a hydrogen-like atom

$$\hat{h}_i = -\frac{1}{2}\nabla_i^2 - \sum_I^N Z_I \mathbf{r}_{iI}^{-1} \tag{10}$$

J_{ij} and K_{ij} are referred to as the Coulomb and exchange integrals respectively.

$$J_{ij} = \langle \chi_i \chi_j \,|\, \chi_i \chi_j \rangle = \langle \chi_i(1) | \hat{J}_j \,|\, \chi_i(1) \rangle = \int \chi_i^*(1) \chi_i(1) \mathbf{r}_{12}^{-1} \chi_j^*(2) \chi_j(2) d\mathbf{x}_1 d\mathbf{x}_2 \tag{11}$$

$$K_{ij} = \langle \chi_i \chi_j \,|\, \chi_j \chi_i \rangle = \langle \chi_i(1) | \hat{K}_j \,|\, \chi_i(1) \rangle = \int \chi_i^*(1) \chi_j(1) \mathbf{r}_{12}^{-1} \chi_j^*(2) \chi_i(2) d\mathbf{x}_1 d\mathbf{x}_2 \tag{12}$$

Integral (11) is an ordinary Coulomb repulsion integral and integral (12) describes a purely quantum mechanical correction to the Coulomb term which takes into account the indistinguishable nature of the different electrons.

Through variational minimization of $E'[\chi_i]$ the Hartree-Fock equation can be obtained

$$\left(\hat{h}(1) + \sum_{j \neq i} \hat{J}_j(1) - \sum_{j \neq i} \hat{K}_j(1) \right) \chi_i(1) = \varepsilon_i \chi_i(1) \tag{13}$$

Here $\hat{J}_j(1)$ and $\hat{K}_j(1)$ are operators defined in terms of their effect on $\chi_i(1)$.

$$\hat{J}_j(1)\chi_i(1) = \left(\int \chi_j^*(2) \mathbf{r}_{12}^{-1} \chi_j(2) d\mathbf{x}_2 \right) \chi_i(1) \tag{14}$$

$$\hat{K}_j(1)\chi_i(1) = \left(\int \chi_j^*(2) \mathbf{r}_{12}^{-1} \chi_i(2) d\mathbf{x}_2 \right) \chi_j(1) \tag{15}$$

By noting that $\left(\hat{J}_i(1) - \hat{K}_i(1) \right) \chi_i(1) = 0$, this allows us to define the fock operator

$$\hat{f}(1) = \hat{h}(1) + \sum_j \left(\hat{J}_j(1) - \hat{K}_j(1) \right), \tag{16}$$

which in turn enables us to write eqn.(13) as an eigenvalue equation:

$$f\left|\chi_j\right\rangle = \varepsilon_j\left|\chi_j\right\rangle \qquad (17)$$

where ε_j represents the energy of spin-orbital χ_j . The last part of eqn. (16) thus represents the averaged potential, V^{HF} , in which the independent electrons move. This also means that any calculation of electronic structure requires some prior knowledge about the nature of this potential, i.e. a starting guess. In calculations involving molecules this guess is commonly provided via some derivation of LCAO-MOs and Huckel theory of conjugated π-systems.

There are many solutions to the eigenvalue problem (17). The correct solutions to (17) can be found by forming a Slater determinant out of the solutions corresponding to the n lowest eigenvalues. The exact Hartree-Fock energy is then obtained in accord with eqn. (9). Within a one-determinant description this constitutes the best approximation to the true energy of the system.

In essence, the Hartree-Fock equations allow us to treat each electron separately, thereby reducing the computational complexity of the Schrödinger equation considerably. However, the potential in which the electrons move has to be refined since this is not exactly known *a priori*. The method of refinement is called the Self Consistent Field (SCF) procedure, originally devised by Hartree, and in short means that, after each fock eigenvalue is calculated, the HF-potential, V^{HF}, is modified accordingly which is then used to refine the next electronic wave function, and so on. This procedure is repeated until a stationary solution is found, i.e. variations in the total electronic energy are sufficiently small, most commonly. The coupling between electronic wave functions is thereby dealt with in an indirect manner, through the HF potential, and the rate of convergence of the SCF procedure is thus closely related to the accuracy of the starting guess.

Density Functional Theory (DFT)

The *ab initio* methods uses variations in wave functions for the calculation of atomic and molecular properties and, as we now turn to DFT methods, a note about the concept of functionals and their derivatives is called for. A functional can colloquially be described as a function of a function, e.g. F[f(**x**)], where f(**x**) serves as a variable of the function F[..]. As in ordinary analysis an extreme of a functional is characterized by the differential of the functional, δF[f(**x**)], equaling zero. However, the differential δF is not defined as in ordinary analysis. Consider a small change in the value of f(**x**), the corresponding change in functional value is then F[f(**x**)+δf(**x**)]-F[f(**x**)]. The part of the functional linearly depending on δf(**x**), at the point **x**, can the be expressed as

$$\frac{\delta F}{\delta f(\mathbf{x})}\delta f(\mathbf{x}) \quad . \tag{18}$$

δF/δf(**x**) is defined as the functional derivative, $\lim_{\delta\to 0}\frac{F[f(\mathbf{x})+\delta f(\mathbf{x})]-F[f(\mathbf{x})]}{\delta}$. By integrating (18) over all points **x**, the differential δF becomes

$$\delta F[f(\mathbf{x})] = \int \frac{\delta F}{\delta f(\mathbf{x})}\delta f(\mathbf{x})\delta\mathbf{x}$$

evaluated throughout space (all **x**). In this formalism, the *ab initio* methods would be considered a minimization of the energy functional E[Ψ(**x**)] and give the correct solution of the Schrödinger equation if a full minimization with all allowed wave functions were performed. In DFT this type of minimization is done while here using the electron density as its variable, i.e. E[ρ(**x**)]. References for the following sections concerning DFT can be found in [24].

The Hohenberg-Kohn theorems

The first Hohenberg-Kohn (HK) theorem states that there is a one-to-one relation between the external potential (e.g. the geometrical arrangement of the nuclei) and the electron density of the

ground state, i.e. the electron density providing the lowest energy for that potential. In other words, this means that no two external potentials can give rise to the same electronic density, and vice versa. The proof can be outlined as follows:

Assume that, for an N-electron system, there exists two external potentials, v(**r**) and v'(**r**), differing by more than a constant, that produces the same electron density ρ(**r**). The two corresponding Hamiltonians $\hat{H}$ and $\hat{H}'$, have different, normalized ground state wave functions, Ψ and Ψ', but the same electronic density. Employing the variational principle on $\hat{H}$ with the trial wave function Ψ' and likewise for $\hat{H}$ with Ψ, two inequalities can be written:

$$E_0 < \langle \Psi' | \hat{H} | \Psi' \rangle = \langle \Psi' | \hat{H}' | \Psi' \rangle + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle = E_0' + \int \rho(\mathbf{r})(v(\mathbf{r}) - v'(\mathbf{r}))d\mathbf{r} \qquad (19)$$

and

$$E_0' < \langle \Psi | \hat{H}' | \Psi \rangle = \langle \Psi | \hat{H} | \Psi \rangle + \langle \Psi | \hat{H}' - \hat{H} | \Psi \rangle = E_0 + \int \rho(\mathbf{r})(v(\mathbf{r}) - v'(\mathbf{r}))d\mathbf{r} \qquad (20)$$

Adding these expressions yields the contradiction $E_0+E_0' < E_0'+ E_0$. Thus we must have $\hat{H} = \hat{H}'$ (explicitly v(**r**) = v'(**r**)), so there cannot be two external potentials giving rise to the same electronic density ρ(**r**).

This allows us to describe T[ρ(**r**)] (kinetic energy), V_{ee}[ρ(**r**)] (electron-electron interactions), V_{en}[ρ(**r**)] (electron-nuclear interactions) and E_{tot}[ρ(**r**)] (total energy) in terms of electronic densities.

$$E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + V_{en}[\rho(\mathbf{r})] + V_{ee}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r} + T[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + V_{XC}[\rho(\mathbf{r})] \qquad (21)$$

Here, the electron-electron interactions have been split into two parts, J and V_{XC}. J[ρ(**r**)] is the Coulomb repulsion part and V_{XC}[ρ(**r**)] is the remainder, describing the exchange-correlation

behavior of the electrons. The three last terms of (21) are also referred to as $F_{HK}[\rho(\mathbf{r})]$. The main focus of present day DFT research is finding the correct $V_{XC}[\rho(\mathbf{r})]$.

The second theorem proves that the variational principle for the energy as a functional of the density holds, i.e. any trial density $\rho'(\mathbf{r})$ will always give a higher energy compared to the energy obtained using the correct ground state density $\rho_0(\mathbf{r})$:

$$E_0[\rho_0(\mathbf{r})] \leq E[\rho'(\mathbf{r})], \tag{22}$$

subject to the conditions that $\rho'>0$ and $\int \rho' d\mathbf{r} = N$. The proof of this theorem is omitted, but can be found in many textbooks on DFT. However, a very important result comes out of this: $F_{HK}[\rho]$ is a universal functional of the density. The exact $F_{HK}[\rho]$ will therefore provide the exact equation for the ground state density.

The conditions for these two theorems to hold are the N- and v-representability conditions briefly described in the next section.

Allowed densities

An infinite number of wave functions can be constructed to fit a certain density. However, the HK theorems are only valid for densities that fulfill two conditions: The N- and v-representability conditions. A v-representable density is one where the density can be associated with an anti-symmetric ground state wave function of a many-electron Hamiltonian. The conditions for v-representability are not known, luckily though the HK theorems are also valid for N-representable densities. An N-representable density fulfils the following conditions.

$$\rho(\mathbf{r}) \geq 0$$
$$\int \rho(\mathbf{r}) d\mathbf{r} = N$$
$$\int \left| \nabla \rho(\mathbf{r})^{\frac{1}{2}} \right|^2 d\mathbf{r} < \infty$$

Furthermore it requires that the density can be obtained from an anti-symmetric wave function. As can be seen, these conditions are not very restrictive and a multitude of wave functions can fulfill them, which leaves the problem of finding the correct one. The scheme employed to do this is

$$
\begin{aligned}
E_0 &= \min_{\Psi}\left\langle \Psi \left| \hat{T} + \hat{V}_{ee} + v(\mathbf{r}) \right| \Psi \right\rangle \\
&= \min_{\rho}\left[\min_{\Psi \to \rho}\left\langle \Psi \left| \hat{T} + \hat{V}_{ee} + v(\mathbf{r}) \right| \Psi \right\rangle \right] \\
&= \min_{\rho}\left[\min_{\Psi \to \rho}\left\langle \Psi \left| \hat{T} + \hat{V}_{ee} \right| \Psi \right\rangle + \int v(\mathbf{r})\rho(\mathbf{r}) d\mathbf{r} \right] \\
&= \min_{\rho}\left[F_{HK}\left[\rho(\mathbf{r})\right] + \int v(\mathbf{r})\rho(\mathbf{r}) d\mathbf{r} \right] = E_0
\end{aligned}
$$

In words this means that out of all N-representable densities picking the one that

i) Minimizes the functional expression of line four, the energy functional, and
ii) Corresponds to the ground state wave function minimizing the line one expression (or rather minimizes the expectation value of $\left\langle \Psi \left| \hat{T} + \hat{V}_{ee} \right| \Psi \right\rangle$).

The Kohn-Sham equations

The last two sections provide the means and terms for finding the form of the density energy functional but leave the problem treated in this section, finding the ground state electronic density. In 1965, Kohn and Sham proposed a solution to the problem of electronic interactions in many-electron systems by splitting the kinetic energy functional T[ρ(**r**)] into a functional of non-interacting electrons and a remainder included in the exchange-correlation functional, an idea corresponding to the approach taken by Hartree and Fock. This separation is written as:

$$G_{HK} = T_n\left[\rho(\mathbf{r})\right] + E_{XC}\left[\rho(\mathbf{r})\right] \qquad (23)$$

$T_n[\rho(\mathbf{r})]$ is the known Thomas-Fermi kinetic energy functional of non-interacting electron systems. The derivations below all assume a closed-shell situation (paired spins).

Rather than using the total density as a variable, we can decompose it into a set of single-particle orbitals, Kohn-Sham (KS) orbitals, like

$$\rho(\mathbf{r}) = \sum_{i}^{n} \sum_{s} |\Psi_i(\mathbf{r}, s)|^2 \ . \tag{24}$$

The occupancy of these orbitals is chosen to be one for the first n orbitals and zero for the rest. The orthonormality constraint is also imposed on the orbitals, i.e.

$$\int \Psi_i(\mathbf{r}, s) \Psi_j(\mathbf{r}, s) d\mathbf{r} = \delta_{ij} \ . \tag{25}$$

These orbitals exactly describe a system of n non-interacting electrons, permitting us to treat the remaining, lesser part of the kinetic energy in an indirect way. The corresponding expression for the kinetic energy functional is

$$T_n[\rho] = \sum_{i}^{n} \left\langle \Psi_i \left| -\frac{1}{2} \nabla^2 \right| \Psi_i \right\rangle ,$$

where $T_n[\rho] < T_{true}[\rho]$. The expression $F_{HK}[\rho] = T[\rho] + J[\rho] + V_{XC}[\rho]$ (from the HK theorem section) can thus be rewritten as

$$F[\rho] = T_n[\rho] + J[\rho] + E_{XC}[\rho] \quad , \tag{26}$$

with $E_{XC}[\rho]$ defined as

$$E_{XC}[\rho] = T[\rho] - T_n[\rho] + V_{ee}[\rho] - J[\rho] \ . \tag{27}$$

The total energy then becomes:

$$\begin{aligned} E[\rho] &= T_n[\rho] + J[\rho] + E_{XC}[\rho] + \int v(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \\ &= \sum_i^n \sum_s \int \Psi_i^*(\mathbf{r})\left(-\frac{1}{2}\nabla^2\right)\Psi_i(\mathbf{r})d\mathbf{r} + J[\rho] + E_{XC}[\rho] + \int v(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \end{aligned} \tag{28}$$

Imposing the orthonormality constraint on the wave functions by means of Lagrange multipliers we obtain the Euler-Lagrange equations

$$\hat{h}_i\Psi_i = \left(-\frac{1}{2}\nabla^2 + v_{eff}\right)\Psi_i = \sum_j^n \varepsilon_{ij}\Psi_j\,, \tag{29}$$

where ε_{ij} are the Lagrange multipliers obtained through the orthonormality constraint. Diagonalisation of the matrix formed by the $\{\varepsilon_{ij}\}$ yields

$$\left(-\frac{1}{2}\nabla^2 + v_{eff}\right)\Psi_i = \varepsilon_i\Psi_i\,, \tag{30}$$

i.e. the Kohn-Sham orbitals. These orbitals constitute the DFT equivalence of the HF orbitals above, and like those, they are non-linear and must be solved iteratively in an SCF procedure. V_{eff} is defined as

$$v_{eff}(\mathbf{r}) = v(\mathbf{r})\frac{\delta J[\rho]}{\delta[\rho]} + \frac{\delta E_{XC}[\rho]}{\delta\rho}. \tag{31}$$

The total energy of the system can now be aquired from the expression (28).

DFT: Pros and cons.

There has been some controversy through the years between advocates of DFT and the hard core *ab initio* people, over this method. This has largely subsided in recent years, mostly due to the many successful applications of DFT, unattainable by *ab initio* methods. One of the main objections to DFT has been the somewhat empirical nature of it. For instance the E_{XC} functional is partly constructed by parameter fitting to experimental data. Another objection raised is that there is no way of improving the calculations like the CI method or perturbation theory in *ab initio* calculations. This is certainly true, however DFT has some major advantages over *ab initio*, mainly the scalability. As a rule of thumb, DFT scales (in computational time) to the square of the number of electrons in the system, whereas a simple HF calculation scales to the power four of the number of electrons. Another advantage is the inclusion of correlation in the calculations, something that can only be reached in *ab initio* theory via extensions of the single determinant description into the multi-determinant correction domain, resulting in even poorer scalability. In addition, DFT is generally considered to give the most accurate results for transition metals, for instance Platinum. Three drawbacks of DFT must however be recognised:

i) Whereas there are gradual ways of improving the electronic structures in the *ab initio* formulation, no such methods are available in DFT. The only way of improving the calculations in DFT lies in refining and correcting the exchange-correlation functional, which is the only remaining unknown in this formalism. Naturally this renders it the main focus of today's DFT research.

ii) Since the exact form of the energy functional, $E[\rho]$, is not known, we can not be sure that the energy obtained from a calculation is the lower limit as stated by the variational principle. The variational principle does indeed apply to the KS equations, provided that all parts of the energy functional are known, but, as mentioned above, that is not the case.

iii) The Kohn-Sham orbitals have no real physical meaning like the Hartree-Fock orbitals. In practise though, the calculated energy levels of Kohn-Sham orbitals correspond well to observations.

Ligand substitution in square planar complexes[6

The common denominator of square-planar complexes, apart from the obvious one, is that they are all cationic species with the d^8 configuration of their valence shell and have a four-coordinate, planar ligand structure, fig. 8.

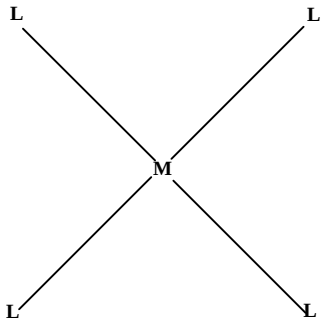


Figure 8. Idealised square-planar compound.

An explanation to the square-planar structure is provided by crystal field theory. Placing an atom with the d^8 configuration of its valence shell in an octahedral field, as shown in fig. 9, causes a split of the degenerate d-orbital energies as indicated in the figure. A removal of the point charges on the z-axis causes a further split of the degenerate energy levels, leaving the orbital highest in energy unoccupied, i.e. the x^2-y^2 orbital (fig.9).

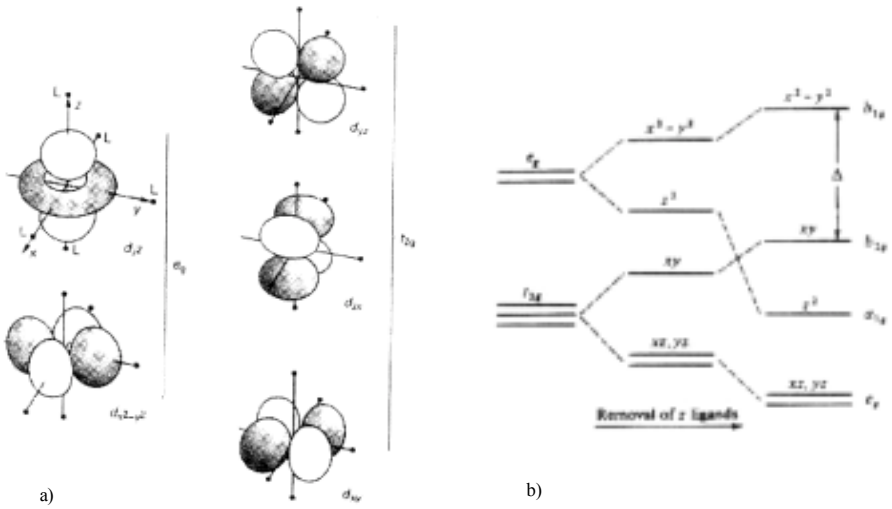


Figure 9. a) The d-orbitals and their degeneracy in, b), fields of different symmetry. Picture taken from: Andri Smith, "Cisplatin – 6. Transition Metal Chemistry", 1999. http//www.chemcases.com/cisplat/cisplat06.htm.

The most extensively studied metal ion forming this type of complex is platinum(II). However several other transition metal ions are known to form square planar complexes, apart from members of the platinum group [Ni(II), Pd(II)] also Au(III), Rh(I) and Ir(I). Unfortunately, comparatively little is known about the reactivity and the finer points of the reaction mechanism for the last three ionic species. Schematically the reaction is depicted as

$$ML_2AX+Y \rightarrow ML_2AY+X, \quad (32)$$

And is governed by the rate law

$$\text{Subst. rate}=(k_1+k_2[Y])[ML_2AX] \quad (33)$$

Furthermore, the substitution is stereo specific, i.e. a *trans*-reactant gives a *trans*-product and a *cis*-reactant gives a *cis*-product. The rate law (33) deserves a closer examination, the k_1 term in particular. Rate constant k_1 result from substitutions by the solvent molecules and the rate law should really read

$$\text{Subst. rate}=(k[\text{solvent}]+k_2[Y])[ML_2AX], \quad (34)$$

But for all practical purposes the concentration of the solvent can be considered constant (and high by definition) and the rate law simplifies to eqn. (33).

The reaction (32) is generally considered to proceed via a five-coordinated trigonal bipyramid as ideally represented in fig. 2, either as a transition state or a meta-stable intermediate, fig. 10.

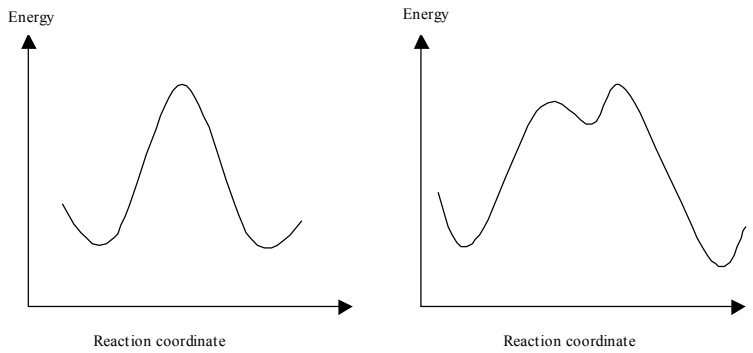


Figure 10. Schematic representation of the energetics involved in a substitution proceeding via either a transition state (left) or a metastable intermediate (right).

Most data collected about the reaction mechanism has been acquired by studying complexes of Pt(II). The reactivity order of the platinum group is

Ni(II) > Pd(II) > Pt(II).

Which ligand is due for substitution is, simplistically put, depending on two things:

i) The intrinsic 'reactivity' of the attacking ligand compared to the ligands already bonded to the metal ion. This is to some extent correlated to the nucleophilic strength, or rather the ligands' strength as a Lewis base.
ii) The nature of the ligand in *trans*-position to the leaving ligand. This effect is also termed the "*trans*-effect".

The *trans*-effect means that different ligands in *trans*-position to the leaving ligand labilize the bond of the leaving ligand in various degrees. It does not mean that a strongly *trans*-labilizing ligand automatically direct the substitution to its *trans* position, it rather means that the activation energy for a substitution in that position is reduced. A few theories on the *trans*-effect have been put forward. Among them the recognition that a strongly polarizing ligand can induce a dipole in the metal ion along its bond axis, thereby repelling the ligand opposite to it (thus reducing that bond strength), see fig. 11.

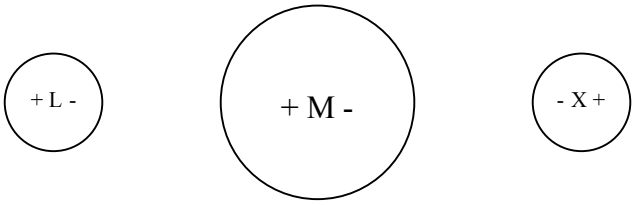


Figure 11. A strongly polarizing ligand (L) induces a dipole in the coordinating metal ion, whereby a repelling force is induced between the metal ion and the ligand (X) in *trans* position to L

A second view emphasizes the importance of good σ-bond donors, as a good donor will claim a larger share of the metal σ-orbitals, leaving less for the *trans*-ligand and thereby weakening its bond.

Quantum chemical treatment of solvent effects

The importance of the solvent in a chemical reaction cannot be neglected, hence the need for accurate treatment of solvent effects also in a quantum chemical context. Up until some years ago these problems were dealt with in a very approximate way, e.g. the Onsager model or the Born approximation, but the fast development of computer hardware and new numerical methods have cleared the way for more detailed solvation models. These computational schemes can broadly be divided in two categories:

- Discrete models, where hundreds or thousands of solvent molecules surround the solute and are treated in a classical, Newtonian way. This clearly gives only the large-scale properties of the system and may produce undesired stochastic effects due to the intrinsic dynamics of the model.
- Continuum models shifts the focus to a subsystem (i.e. the solute), treated with high accuracy using some quantum mechanical model. The remaining part is treated as a continuum in which the average electrical properties of the bulk of the system (i.e. the solvent) are used to set the environment for the solute.

In this thesis, the latter type of model has been used, here represented by the PCM model[25]. The advantages a continuum representation offer, compared to a discrete one, is computational inexpensiveness and a natural 'averaging' of solute-solvent interactions devoid of stochastic effects sometimes encountered in Molecular Dynamics simulations.

In continuum solvation models the bulk of the solvent is represented as an infinitely large continuous dielectric medium, in which a cavity is created approximating the shape of the solute so as to accommodate the species under study. The cavity is created as a number of interlocking spheres equal to the number of nuclei, centered on these. Overlapping areas are discarded and the resulting outer surface is then subdivided into small surface 'facets' called tesserae. Typically there are 60 tesserae per sphere (fig. 12)

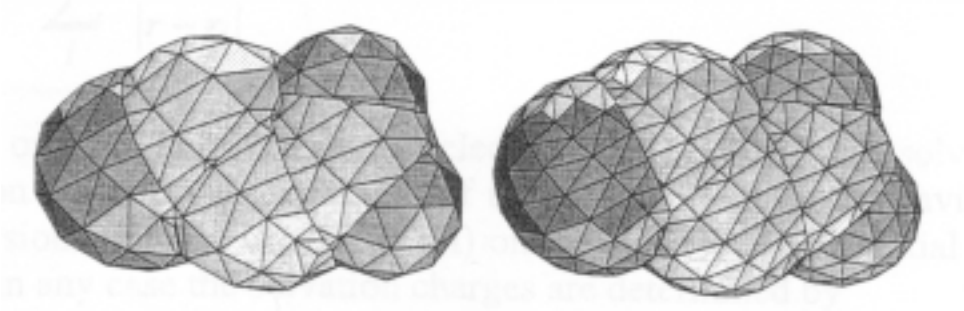


Figure 12. Cavity for glycine, created using the GePol algorithm. The area of the tesserae is 0.4 and 0.2 Å respectively. From ref [25].

The solute solvent interactions can be described as a perturbation to the *in vacuo* Hamiltonian, $\hat{H}_0$, like

$$\hat{H} = \hat{H}_0 + \hat{V}_{\text{int}} \tag{35}$$

with $\hat{V}_{\text{int}}$ comprising three types of interactions

$$\hat{V}_{\text{int}} = \hat{V}_{el} + \hat{V}_{ster} + \hat{V}_{dis} \quad . \tag{36}$$

$\hat{V}_{el}$ is a potential created as a response to the solute's electrostatic potential on the surface of the cavity. $\hat{V}_{ster}$ is a sterical, or Pauli repulsion contribution and $\hat{V}_{dis}$ represents the dispersion properties of the solvent acting on the solute. The sterical and dispersive parts can be calculated outside the SCF procedure, which is usually the case, since the main impact on the SCF calculations is exerted by the electric component of the perturbation and the mutual dependence of the different parts is small enough to be neglected.

Commonly the sterical and dispersive interactions are included in the form of a Lennard-Jones potential. It has been shown that $\hat{V}_{el}$ can be expressed in terms of an 'apparent charge density' σ appearing on the cavity surface and this forms the basis of this method. As a computational convenience one assumes σ to be constant in each tesserae as to express $\hat{V}_{el}$ as a sum of point charges $\{q_i\}$ (solvation charges) placed in the centre of each tesserae:

$$\hat{V}_{el}(\mathbf{r}) = \sum_{i}^{\#tess} \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|} \tag{37}$$

In the original version of PCM the point charges depend on the normal component of the electric field, in more recent versions on the electrostatic potential at the cavity surface. In either case the point charges $\{q_i\}$ are determined by

- The solute's nuclear and electronic densities.
- The influence of the other solvation charges.

This means that a set of coupled linear equations equal to the number of solvation charges must be solved in every SCF cycle to find the charges. The system to solve has the structure:

$$\mathbf{Dq} = -\mathbf{b}_{\mathrm{sol}} \tag{38}$$

Where $\mathbf{q}$ is a vector containing the solvation charges q_i, $\mathbf{b}_{\mathrm{sol}}$ is the vector containing either the solute electrostatic potential or, as in the original PCM version, the normal of the solute's electric

field evaluated at the tesserae centers. The matrix **D** depends on the dielectric constant of the solvent and geometrical parameters related to the cavity shape. Solving the system (38) allows us to define the perturbation operator $\hat{V}_{\text{int}}$ and solve the Schrödinger equation.

Methods

All calculations presented in this work have been performed using the Gaussian 98 program package on Alpha(433) DEC workstations or Pentium III PC's (Digital UNIX or LINUX operating systems respectively) unless where otherwise indicated.

The first part of the investigation, dedicated to the hydration of cis-DDP, employed the Becke 88[29] (B) exchange functional and Lee, Yang and Parr[30, 31] (LYP) correlation functional in conjunction with the LANL2DZ[32, 33, 34, 35] basis set for geometry optimizations *in vacuo*. The frequencies have been calculated using analytic second derivatives for the above functionals and basis set. Solvation energies were obtained from single point calculations using the PCM[36, 37, 38] method as implemented in Gaussian 98 on structures from the geometry optimizations *in vacuo* with the above exchange-correlation functionals and basis set unchanged.

In the second part, focusing on the cis-$Pt[NH_3]_2[H_2O]_2^{2+}$ + 2G → cis-$Pt[NH_3]_2[G]_2^{2+}$ + $2H_2O$ reactions, the calculations were performed using the same set-up (BLYP/LANL2DZ, analytic second derivatives) as in the first part for geometry optimizations and frequencies, although no solvation energies have been calculated here. The main reasons for this are:

i) There is no change in net charge during the different stages of this reaction as in the first part, meaning that for the rather similar structures encountered here there would, in all likelihood, simply be a shift in free energy in the same order of magnitude at all stages of the reaction.

ii) It is uncertain which dielectric constant to use, since the reaction *in vivo* takes place on a surface (i.e. the major groove of DNA) rather than in a bulk solution and the stacking interactions between bases is not accounted for in the PCM model.

Effective Core Potential (ECP) basis sets

The Hartree-Fock and Kohn-Sham equations as presented above are clearly non-relativistic and for heavier elements (such as Platinum) with highly charged nuclei they produce erroneous results in geometry optimizations, due to an underestimation of the shielding effect by inner shell

electrons. The larger shielding by these electrons is a relativistic effect since the inner shell orbitals contract as their electrons reach velocities where relativity must be accounted for (This contraction is by the way the reason for the relatively unchanged atomic size of the elements in their respective transition metal group.). Although corresponding relativistic versions of the HF and KS equations exist, explicitly calculating the orbital contractions, less computationally intensive alternatives have been developed, suitable for solving problems of the kind covered in this work (e.g. geometry optimizations, frequencies). The method employed here[26] relies on implicitly, via the basis set, accounting for the inner shell electrons by replacing the inner shell orbitals with an effective core potential, mimicking the true potential of the inner shell electrons. The outer shells can be described by Gaussian functions in the same manner as ordinary basis sets. The basis set used in this study, LANL2DZ, is an ECP well represented in published articles on theoretical chemistry problems and is furthermore the most practical choice in the Gaussian program package for this particular problem.

Linear Combination of Atomic Orbitals

The most common model for describing the chemical bond in a quantum chemical context is the LCAO molecular orbital model. The basic idea behind this description is that the orbitals of different atoms sharing certain features can be linearly combined to make up new *molecular orbitals* (MO), in a sense constituting the 'glue' that holds the molecule together. In this description each MO can be written as a summation in this form.

$$\psi_i = \sum_{\mu=1}^{K} c_{\mu i}\phi_\mu \qquad (39)$$

ψ_i is a molecular orbital, $c_{\mu i}$ is a coefficient and ϕ_μ is one of K atomic orbitals. In the informative case of the hydrogen molecule, there are two atomic 1s orbitals (denoted $1s_a$ and $1s_b$) that can be combined, subject to some conditions, in two ways to give

$\psi_+ = 1\sigma_g = N_+^{-1/2}\,(1s_a + 1s_b)$

$\psi_- = 1\sigma_u^* = N_-^{-1/2}\,(1s_a - 1s_b)$

$N_{+/-}^{-1/2}$ are normalisation constants. These linear combinations result in new energy levels of the molecule compared to those of the individual atoms as shown in fig. 13.

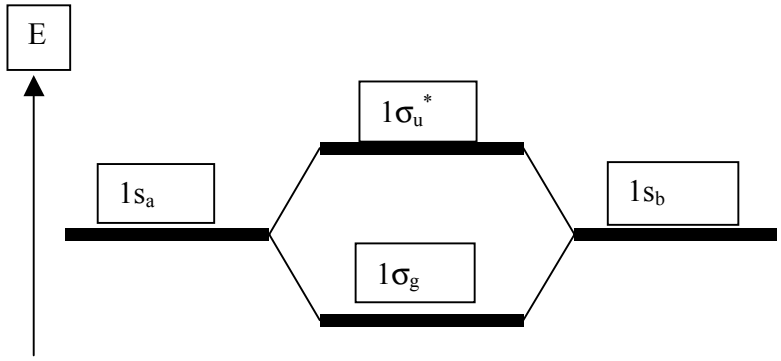


Figure 13. Linear combinations of atomic orbitals of two hydrogen atoms comprising the molecular orbitals a hydrogen molecule.

The MOs are then filled with electrons in accord with the *aufbau* principle. In larger molecules the linear combinations are of course more complex and difficult to find but it is still only a matter of solving systems of linear equations, which is an easy task for modern computers.

Geometry optimizations and frequency calculations

A geometry optimization can be viewed as the search for a configuration of the nuclei that correspond to the closest local minimum with respect to the forces acting upon them. In practical calculations this means that after convergence of the SCF procedure is achieved, the forces on each nucleus is calculated by differentiating the potential at each nuclear point and a step is taken in the direction of the gradient calculated. This is done until a stationary point is reached, hopefully the local minimum you are interested in, characterized by the requisites of a minima in multidimensional analysis.

Usually the forces are transformed into *normal coordinates* and collected in a Hessian, or force constant matrix. In essence, the normal mode transformation constitutes a decoupling of the internal vibrations of the molecule. The eigenvalues of the Hessian in the region of a minimum should all be positive. By contrast, in regions connecting two minima on a potential surface, i.e. transition state regions, one eigenvalue should be negative and the rest positive, thus constituting a saddle point on the potential surface.

In a *normal mode* frequency calculation, a frequency corresponding to a negative eigenvalue would then be imaginary, cf. the situation of a mass loaded spring with a negative force constant. References to 'negative' frequencies in this work (and many others) do actually mean imaginary frequencies by custom.

Thermodynamic stabilities

One measure of the feasibility of a reaction is obtained by comparing the calculated energies of the different stages in a reaction pathway as

$$\Delta E = \sum_{prod=1}^{\#prod} E_{prod} - \sum_{reac=1}^{\#reac} E_{reac} \qquad (40)$$

These calculations naturally require an unchanged stoichiometry at the different points along the reaction path to be comparable. The energies are however merely an indication of the feasibility of a reaction, as there are other factors to be accounted for, e.g. sterical hindrance and thermodynamic effects. This is mentioned since no direct observations of energies can be done at some of the states calculated and comparisons will instead be made to kinetic data, which is a bit hazardous. This is the only alternative, though, to relate some of the calculated data to reality.

Results

1.0 Aquation of *cis*-DDP

The complete aquation of cis-DDP is a two-step substitution reaction in which the two chlorides of cis-DDP are sequentially replaced by water molecules:

$$\begin{aligned} &cis-Pt(\mathrm{II})\left[NH_3\right]_2\left[Cl\right]_2 + H_2O \Leftrightarrow cis-Pt(\mathrm{II})Cl\left[NH_3\right]_2 H_2O^+ + Cl^- \Leftrightarrow \\ &cis-Pt(\mathrm{II})Cl\left[NH_3\right]_2 H_2O^+ + H_2O \Leftrightarrow cis-Pt(\mathrm{II})\left[NH_3\right]_2\left[H_2O\right]_2^{2+} + Cl^- \end{aligned} \qquad (41)$$

From this it can be seen that the charge separation of the first substitution is one unit positive and one unit negative. The corresponding charge separation of the second substitution is two units positive and one unit negative. If the geometries of the respective transition states (TS) were assumed to be reasonably similar, especially referring to Pt-Cl(leaving) distance, one would expect an increase in the energy barrier of the second substitution.

1.1 Optimization of reactants and products.

These optimizations were fairly straightforward. Input geometries were obtained using the Spartan semi-empirical package with Pd(II) as a substitute for Pt(II), since no semi-empirical parameters for Platinum have yet been derived and implemented. In the subsequent, more rigorous, DFT calculations the setup were as described in the methods section with Pt(II) back instead of Pd(II). The figure below (fig. 14) shows the final, optimized structures, of cis-DDP and the diaquated species as well as an X-ray structure of cis-DDP.

Figure 14. Optimized structures of cis-DDP (left) and the diaquated species (middle). The right figure is a geometry determined by means of X-ray crystallography.

Energies of these optimized species and other along the reaction pathway are summarized in table 1. In all of the above structures the angles between coordinated ligands sum to a total of 360.0^{o}, and they are perfectly planar. However, when comparing individual angles it is obvious that the optimized structures are not perfectly square whereas the X-ray structure is, see the angular values in table 2. This is not surprising since the optimizations were performed in absence of a solvent model quenching the electrostatic interactions. The electrostatic repulsions are evident in the calculated structure of cis-DDP with a Cl-Pt-Cl angle of ~97^{o} whereas the experimental structure is right-angled. The Cl-Pt-N angle, ~81^{o}, is then, by the same line of reasoning, a result of electrostatic attraction between the negatively charged chlorides and the partial positive charges of the ammine group hydrogens. Implicitly, this explanation model is strengthened by the optimized structure of the diaquated cis-DDP derivate being close to right-angled.

Bond lengths are increased in the calculated structure of cis-DDP compared with the experimental one, by 7-8%. A possible explanation for this difference is that the experimental structure was determined by means of X-ray crystallography. When comparing X-ray structures to structures determined by NMR, distances are consistently shorter in X-ray structures. For instance, the Pt-NH_3 distance in X-ray structures of cisplatin's bifunctional adduct to DNA, is 2.00 Å, and the corresponding distance determined by NMR is 2.05 Å.

Solv.\Species	Reac.	RC1	TS1	IMP	RC2	TS2	Product
PCM	0	-1.8	10.8	8.0	-1.3	18.9	2.2
In vacuo	0	-11.8	2.1	123.7	99.4	123.3	111.7

Table 1. Relative energies (in kcal/mol) along the reaction path for the aquation of cisplatin.

1.2 Optimization of the first Transition State (TS1) and its Reactant Complex (RC1).

Optimization of TS1 was done in reverse reaction order, i.e. the intermediate product (IMP) cis-$Pt[NH_3]_2[Cl][H_2O]^+$ was used as a starting point and a relaxed potential energy scan was performed by forcing the chloride ion towards the Pt(II) center at fixed intervals ranging from 3.00 Å to 2.70 Å. A structure from the optimization at Pt-Cl=2.80 Å were picked and the subsequent transition state optimization(s) finally yielded the structure of fig. 15. The forward reaction order was unsuitable due to severe convergence problems as the attacking water molecule was very volatile, i.e. it did not find the reaction path. Important geometrical parameters are collected in table 2.

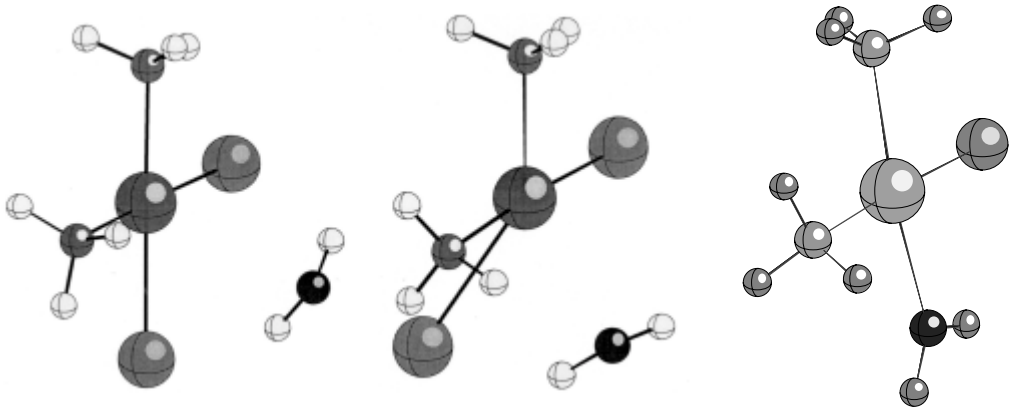


Figure 15. Geometries of the reactant complex, transition state and the resulting intermediary product for the first aquation of cisplatin.

In TS1, the geometry of the equatorially oriented ligands (NH_3(top in fig.), H_2O (entering) and Cl^- (leaving)) is planar as the sum of the angles is ~360^o and the individual ligands are closely right-angled to the axial ones, thus forming a trigonal bipyramid as predicted by the theory for this kind of reaction. The optimized TS1 has a single negative frequency of –166.0 cm^{-1}. This is a very low

magnitude of negative frequency for an optimized transition state, indicating a flat potential energy surface around the point of the transition state. The vibrational mode corresponding to this negative frequency in the forward reaction direction is characterized by an increase in bond length for the leaving ligand, a decrease in bond length for the entering ligand and an increased H_2O(equatorial)-Pt-NH_3(equatorial) angle as predicted by theory. The reactant complex was found by standard procedures, following the intrinsic reaction coordinate indicated by the normal mode vector corresponding to the negative eigenvalue of the Hessian at the TS. The lowest vibrational frequency of RC1 is positive and is characterized by a variation of the water-Pt distance. Limited time has prevented an optimization to the product complex, which should have been done.

The second TS (TS2), its reaction complex and the final product complex.

As in the previous section, the localization of the second transition state (denoted TS2) was done in the reverse reaction order, for the reasons mentioned above. The same type of coordinate driving was performed and a promising geometry was taken as a starting point for the transition state optimization, the result of which is shown in fig. 16 along with its adherent reactant and product complexes.

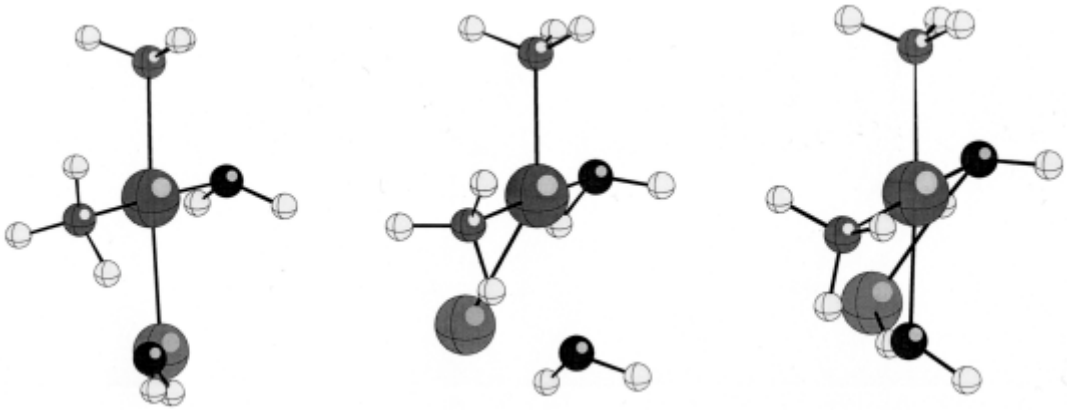


Figure 16. Reactant complex, transition state and product complex for the second substitution in the aquation of cisplatin.

TS2 is skewed in comparison to TS1 with the leaving chloride ligand symmetrically spaced in relation to the two water ligands (HOH-Cl = 2.13 Å (axial) and 2.14 Å (equatorial) respectively).

The plane of the equatorial ligands is twisted in comparison with TS1 and the axial ligands are no longer perpendicular to the equatorial plane, thus not resembling the idealized trigonal bipyramid. Geometrical details are listed in table 2.

The frequency for the optimized TS2 was calculated to be –164.5 cm^{-1}. Again indicating a flat potential energy surface in the vicinity of the transition state geometry. The corresponding vibrational mode has the same characteristics as TS1. The reactant and product complexes were obtained by the standard procedure outlined above. The lowest vibrational modes of the respective complexes were both positive, i.e. minima, and corresponded to nucleophile attack in either reaction order, forward and reverse respectively.

1.4 The potential surface of the total reaction

Energies at the different stages of the reaction have been calculated with and without a solvent model (PCM) but using the geometries calculated *in vacuo*. The PCM was not included in the actual geometry optimizations due to oscillatory behavior, i.e. alternations between two geometries (seemingly indefinite), in the optimization of TS2, leaving this data set incomplete. However, the single point calculations on the *in vacuo* geometries with the PCM included should provide a good measure of the energy at each stage of the reaction, as the inclusion of a solvation model did not alter the vacuum geometries of the other states significantly.

Although the stoichiometry is constant in the energy diagram below (fig. 17), the stochiometry has not been constant throughout all geometry optimizations of the total reaction (41). This has in part been a practical consideration but is also motivated by kinetic data and thermodynamics. In one article[28] the time lapse between the two substitutions was estimated to be three hours on average, ample time for the first chloride ion to diffuse far away from cisplatin in the chloride depleted cytosol of the cell, thus leaving the situation described in section 1.3.

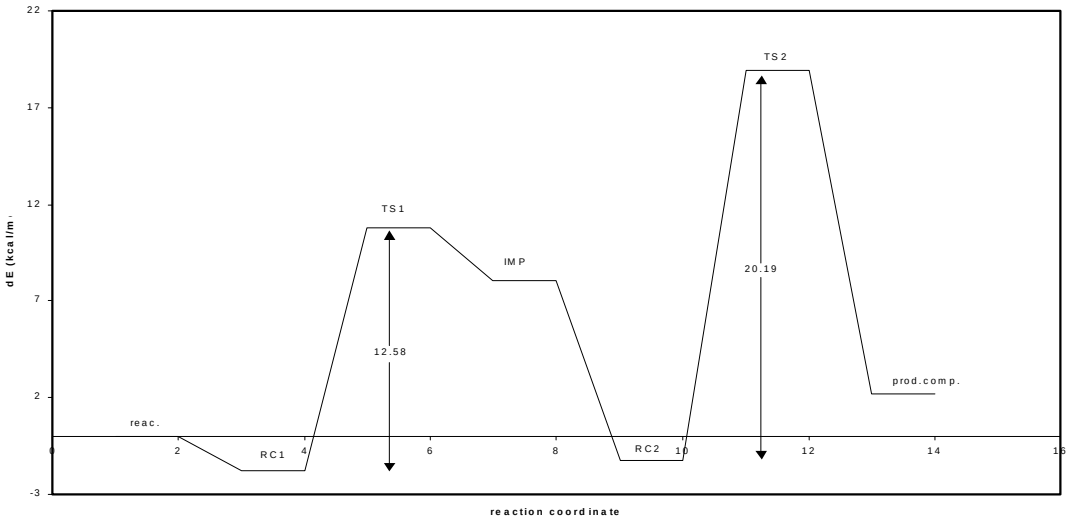


Figure 17. Energy diagram for the aquation of cisplatin.

2.0 The attack of diaquated cisplatin on purine bases Guanine and Adenine

The final product of the aquation of cis-DDP has a large number of hydrogen bond donors and in combination with a hydrogen bond acceptor, such as O6 on Guanine (G) or N6 of Adenine, a number of possible attack modes arise. Three of them are schematically represented in fig. 18. Two more can easily be pictured, the single hydrogen bond to a water ligand or to an ammine group of cisplatin. In contrast to the approach used for the preceding calculations on the aquation of cisplatin these optimizations have been performed in the forward reaction order.

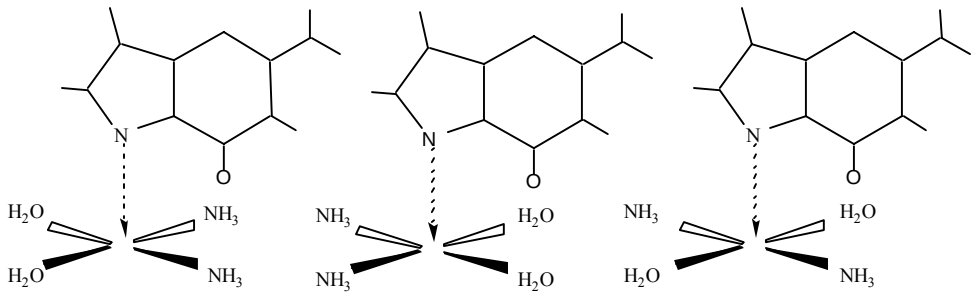


Figure 18. Alternative attack modes of purine base Guanine on aquated cisplatin.

The model system used for localization of the first TS consisted of a purine base and the diaquated cis-DDP, which should suffice to give a qualitatively accurate picture of the formation

of a monofunctional adduct to DNA. The inclusion of a much larger piece of the DNA molecule is well beyond the capacity of present day computers and computer codes within an all quantum mechanical description (Hybrid QM-MM approaches may well work, though. See Discussion section.). No PCM energy calculations has been performed on the structures presented in this section since the reactions do not take place in a homogenous continuum but rather on the heterogeneous surface of the major groove of DNA, and while energy diagrams are presented, naturally these are only accurate in the context of this model system. Hence, this section must be viewed as a qualitative investigation of the true interactions.

2.1 Optimization of the first transition state (TS1) and reactant complex (RC1) for A and G as entering ligands.

In item 2.0 above, five modes of attack were suggested. All of these have been investigated but only two resulted in transition states: The single hydrogen bond of O6- H_2O(in the case of G as the attacking ligand) or N6-NH_3(in the case of A as the attacking ligand) to either water or ammine group respectively in *cis*-coordination to the leaving water ligand. The lowest vibrational modes of the transition states presented correspond to the transition state modes in the cisplatin aquation section, i.e. a decrease of N7(entering)-Pt distance, increase of H_2O(leaving)-Pt distance and an increase of the N7(equatorial)-Pt-NH_3(equatorial) angle. N7 is defined in fig. 2 in the introduction section. Important geometrical parameters, energies and frequencies are collected in table 3. Figure 19 below show the reactant complexes and transition states with either Adenine or Guanine as entering ligands.

The standard procedure, with the leaving ligand still present, described above was employed for RC optimizations. No product complexes have been optimized. However, intermediary products (IMP, right in fig. 19) were optimized to provide a starting point for the subsequent calculations on the second substitution. The following localization of the second transition states have been done using the Guanine intermediary product as a starting point with both Adenine and Guanine as entering ligands, thus resembling the actual situation *in vivo*.

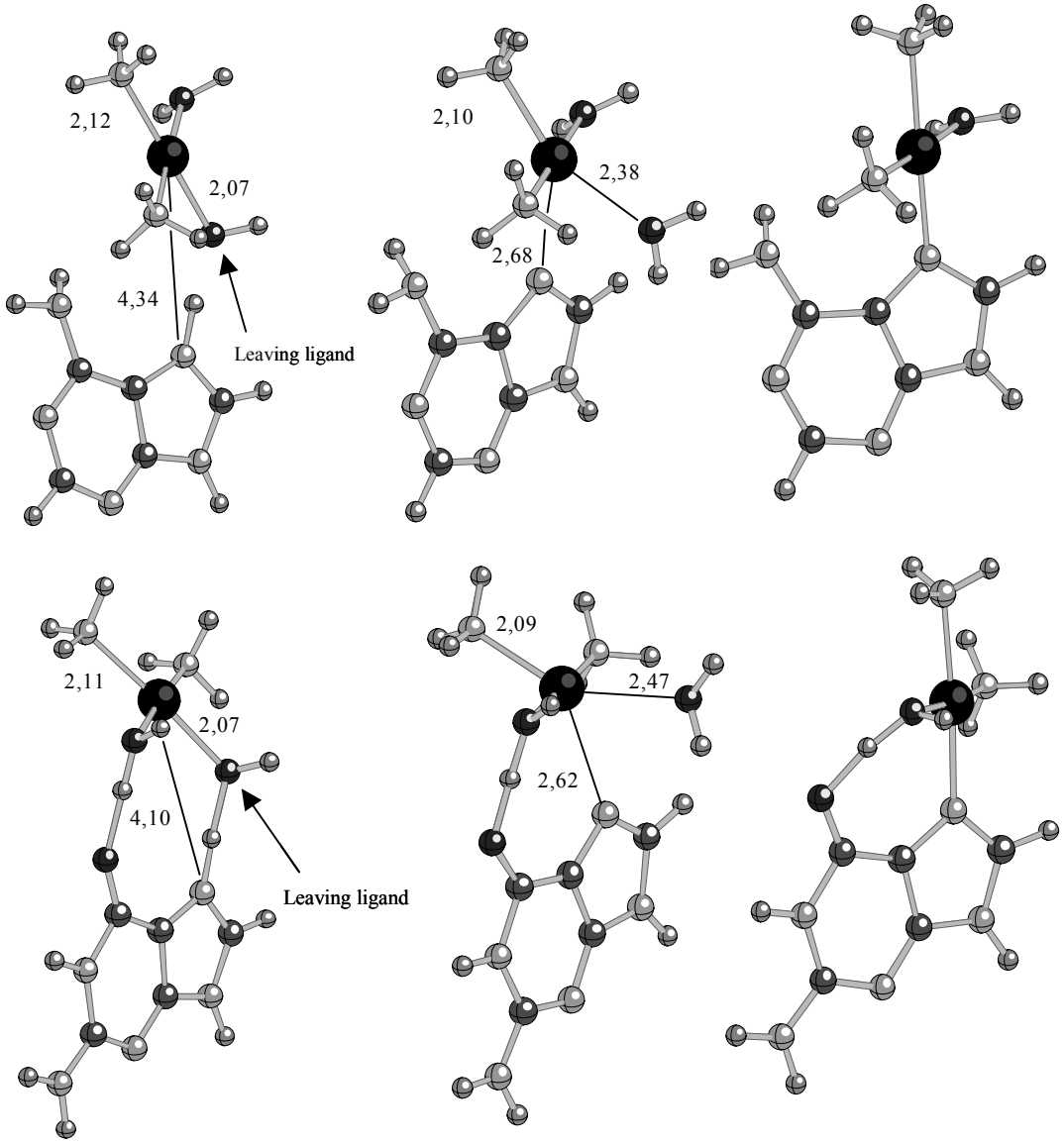


Figure 19. Reactant complexes, transition states and intermediary products for the first substitutions of Adenine (top) and Guanine (bottom).

a)

Angle (deg.) \ State	Reac.	RC1	TS1	IMP	RC2	TS2	PC2	Prod.
L_{ent}-Pt-L_{leav}	-	71.2	66.8	-	57.0	66.7	52.6	-
L_{ent}-Pt-L_{equat}	-	108.9	147.3	-	127.6	158.3	174.5	-
L_{leav}-Pt-L_{equat}	177.7	178.6	145.9	177.1	174.7	135.0	122.1	178.7
L_{ent}-Pt-L_{ax1}	-	84.5	87.7	-	43.6	73.7	83.3	-
L_{ent}-Pt-L_{ax2}	-	92.6	88.3	-	134.6	104.0	93.2	-
L_{leav}-Pt-L_{ax1}	81.0	82.0	73.8	86.7	89.9	102.7	122.1	88.5
L_{leav}-Pt-L_{ax2}	96.8	96.8	104.0	82.7	87.8	70.6	52.6	90.5
L_{equat}-Pt-L_{ax1}	101.3	99.4	101.2	96.2	94.4	98.9	99.7	92.7
L_{equat}-Pt-L_{ax2}	81.0	81.8	82.7	94.4	86.9	85.5	83.3	88.4

b)

Distance (Å) \ State	Reac.	RC1	TS1	IMP	RC2	TS2	PC2	Prod.
L_{ent}-Pt	-	3.27	2.50	-	3.73	2.46	2.10	-
L_{leav}-Pt	2.43	2.45	2.85	2.41	2.42	2.83	3.61	2.14
L_{equat}-Pt	2.13	2.12	2.12	2.15	2.13	2.12	2.09	2.08
L_{ax1}-Pt	2.13	2.13	2.11	2.07	2.06	2.07	2.09	2.08
L_{ax2}-Pt	2.43	2.44	2.45	2.11	2.15	2.13	2.10	2.14

c)

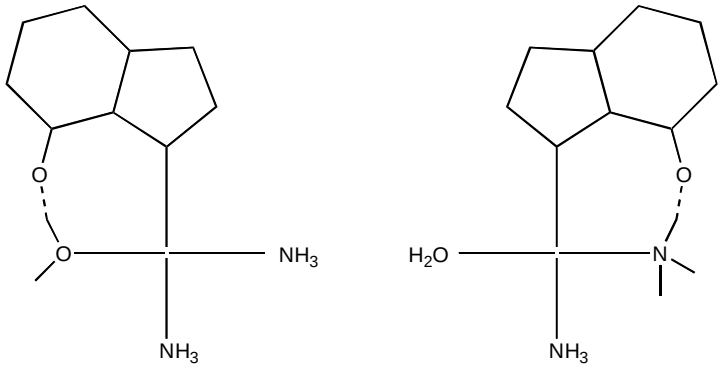
Reactant \ Energy (a.u.)	Solvation model	
	In vacuo	PCM
Cisplatin	-262.080075	-262.128872
Monohydrated cisplatin	-323.330580	-323.450097
Dihydrated cisplatin	-384.586010	-384.759430
Cl^-	-14.947635	-15.075382
H_2O	-76.395255	-76.409383
RC1	-338.494117	-338.541104
TS1	-338.472060	-338.521057
RC2	-399.764520	-399.874250
TS2	-399.726521	-399.822078
PC2	-399.744940	-399.868804

d)

Species	Frequency (cm^{-1})
TS1 (cisplatin)	-166.0
TS2 (cisplatin)	-164.5
TS1 (Adenine)	-146.3
TS1 (Guanine)	-144.8
TS2 (Adenine) head-to-head	-148.4
TS2 (Guanine) head-to-head	-145.3
TS2 (Guanine) head-to-tail	-154.9

Table 2. Important geometrical parameters (2a and 2b), energies (2c) and frequencies (2d) at the different stages in the aquation of cisplatin. L_{ent} denotes the entering ligand (H_2O) as described in the text. Consequently does L_{leav} denote the leaving ligand (Cl^-), L_{equat} the equatorially oriented ligand (NH_3) of a TS (the ligand that will become equatorially oriented in a TS) and $L_{ax1,\ ax2}$ the axially oriented ligands of a TS (the ligands that will become axially oriented in a TS). L_{ax1} is defined as the axial ligand on the right hand side in figures 15 and 16, i.e. NH_3 and thus does L_{ax2} correspond to the ligand in trans position to L_{ax1} (Cl^- or H_2O). Distances and angles are in all cases measured between centers directly involved in bond formation to Platinum, e.g. O-Pt-N for an angle.

At this stage a pattern emerges: All transition states located so far have had the entering purine coordinated via a hydrogen bond to a ligand in cis-position to the leaving ligand. If this pattern is followed, the coordinating ligand for the next substitution should be the ligand in trans-position to the leaving ligand of the first substitution. However, the first substitution can give rise to two stable conformers as schematically represented in fig. 20, thereby leaving four possible transition states to locate for the next substitution, two for Adenine and two for Guanine.



Conformer 1 **Conformer 2**

Figure 20. Two stable conformers of the intermediary Guanine product. The difference is the position of the leaving water ligand in the two conformers.

2.2 Transition states, reactant and product complexes of the second purine substitution

Finding the transition states of the second substitution was accomplished by keeping the structural coordination of ligands around the Platinum centre from the first transition states relatively intact, apart from making the necessary substitutions. Three of the possible four transition states have been found and are shown below in figures 21 and 22. The geometric parameters, energies and negative frequencies of these transition states and their adherent reactant and product complexes are listed in table 3. The vibrational modes of the negative frequencies all display the same characteristics as the previous transition states of the purine substitution. Alas, only one reactant complex have been optimized which is included in figure 22.

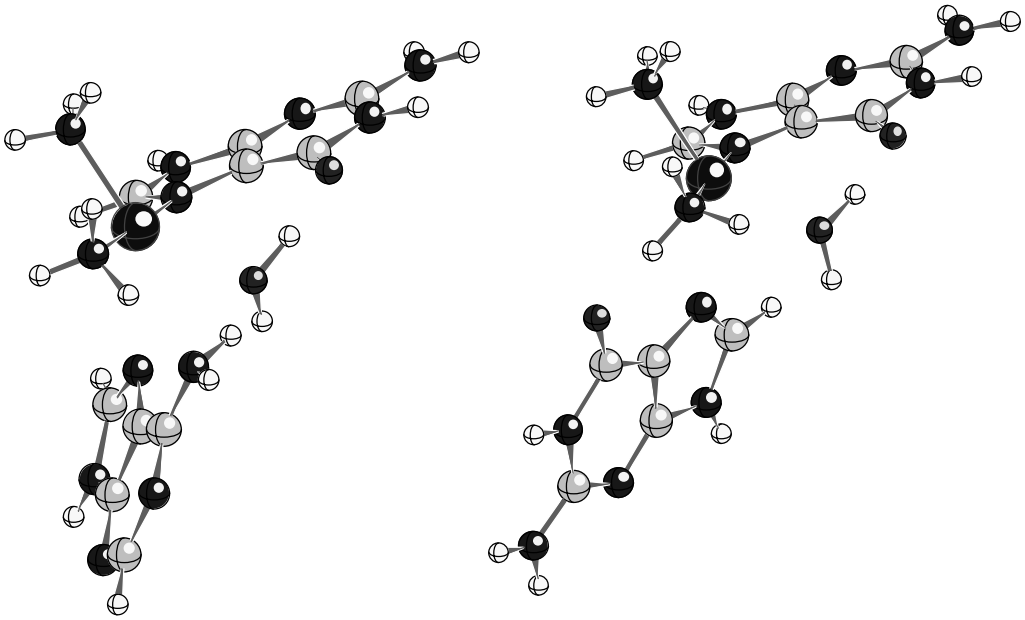


Figure 21. Transition states of the second purine substitution. The intermediary product of Guanine (conformer 1) is in all cases the starting point for these transition states. Adenine is the attacking ligand of the left TS and Guanine is the attacking ligand in the TS on the right.

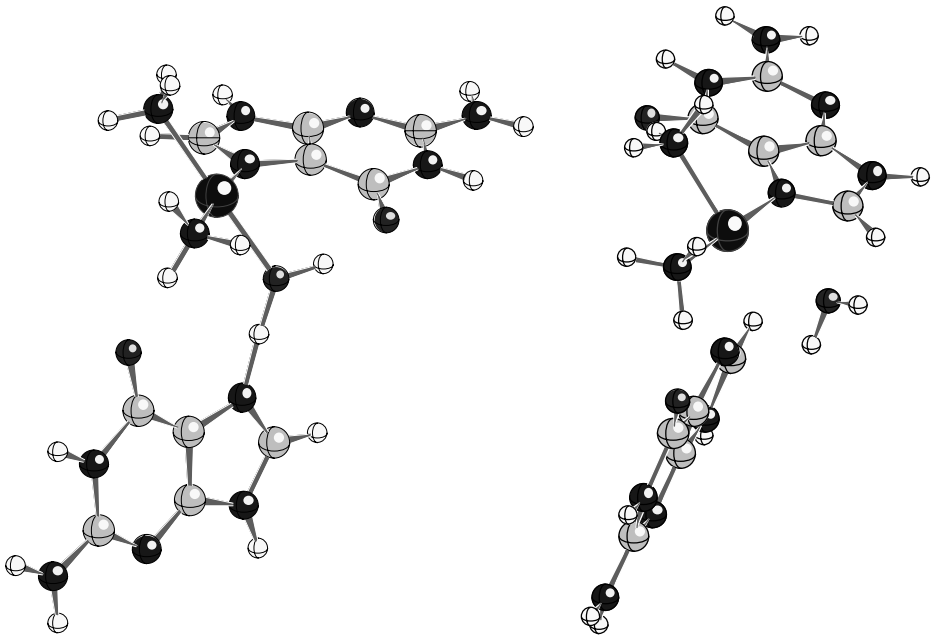


Figure 22. The reactant complex (left) leading to the right TS structure of fig. 21. The structure to the right is the second TS of guanine attacking the Guanine intermediary product (conformer 2).

The not yet determined TS for the second substitution is the Adenine substitution in combination with the conformer 2 of fig. 20. Problems occurred here due to a tendency for N7 of Adenine to hydrogen bond to the leaving water ligand. This phenomenon was also encountered in the other TS optimizations, although not as severe. However, this certainly does not mean that there exist no unambiguous TS for this geometrical arrangement.

2.3 Optimization of the final cis-Pt $[NH_3]_2[G]_2^{2+}$ products.

The final outcomes of the Guanine substitutions presented above are two stable conformers denoted head-to-head and head-to-tail and correspond to the intrastrand and interstrand situation respectively, differing in the relative orientation of the two purines. Since the water molecule present in the Guanine transition state is not included, these structures are not product complexes. Structures are shown in fig. 23 and included in the figure are the corresponding, experimentally determined structures. The head-to-head arrangement is characterized by the oxygen atoms (O6) of Guanine facing the same side of the coordinating platinum plane and the head-to-tail arrangement has the oxygen facing different sides giving it the appearance of a wing-nut. Important geometrical parameters and energies are included in table 3. Frequencies for the transition states of the DNA attack on cisplatin are included in table 2d.

a)

Angle (deg.) \ State	Reac.	RC1	TS1	IMP	RC2	TS2	PC2	Prod.
L_{ent}-Pt-L_{leav}	-	31.2	68.1	-	28.2	74.3	-	-
L_{ent}-Pt-L_{equat}	-	145.7	141.7	-	153.2	142.6	-	-
L_{leav}-Pt-L_{equat}	-	176.9	150.1	174.8	175.3	142.5	-	175.8
L_{ent}-Pt-L_{ax1}	-	99.6	92.3	-	77.0	90.3	-	-
L_{ent}-Pt-L_{ax2}	-	77.3	88.2	-	99.5	90.5	-	-
L_{leav}-Pt-L_{ax1}	-	83.4	79.1	82.6	83.0	78.2	-	87.4
L_{leav}-Pt-L_{ax2}	-	93.5	101.5	91.9	92.6	98.7	-	93.4
L_{equat}-Pt-L_{ax1}	-	98.7	97.5	93.9	93.3	91.8	-	91.4
L_{equat}-Pt-L_{ax2}	-	84.4	81.9	91.6	91.1	89.6	-	87.5

b)

Distance (Å) \ State	Reac.	RC1	TS1	IMP	RC2	TS2	PC2	Prod.
L_{ent}-Pt	-	4.10	2.62	-	4.10	2.54	-	-
L_{leav}-Pt	-	2.07	2.47	2.08	2.05	2.39	-	2.09
L_{equat}-Pt	-	2.11	2.09	2.11	2.12	2.12	-	2.10
L_{ax1}-Pt	-	2.08	2.09	2.11	2.10	2.09	-	2.10
L_{ax2}-Pt	-	2.10	2.08	2.08	2.07	2.08	-	2.09

c)

Relative energies (kcal/mol) \ State	Reac.	RC1	TS1	IMP	RC2	TS2	PC2	Prod.
Guanine subst. of activated cisplatin	0.0	-4.9	29.7	23.9	-36.7	-6.7	-	-13.2
Adenine subst. of activated cisplatin	0.0	35.3	72.6	59.9	-	-	-	-

d)

Reactant	Energy (a.u.)
Adenine	-467.102544
Guanine	-542.328675
RC1 (Adenine)	-851.632327
TS1 (Adenine)	-851.572851
IMP (Adenine)	-775.197884
RC1 (Guanine)	-926.922420
TS1 (Guanine)	-926.867333
IMP (Guanine)	-851.481335
RC2 (Guanine)	-1392.906517
TS2 (Guanine)	-1392.858803
cPt[NH_3][G_2]$^{2+}$ (head-to-head)	-1316.473919

Table 3. Geometrical parameters for the attack of DNA on cisplatin. Angles are presented in 3a and distances in 3b. Definitions follow those of figure 2. Table 3c shows the relative energies along the reaction path and 3d show the calculated individual energies that comprise the relative energies of 3c.

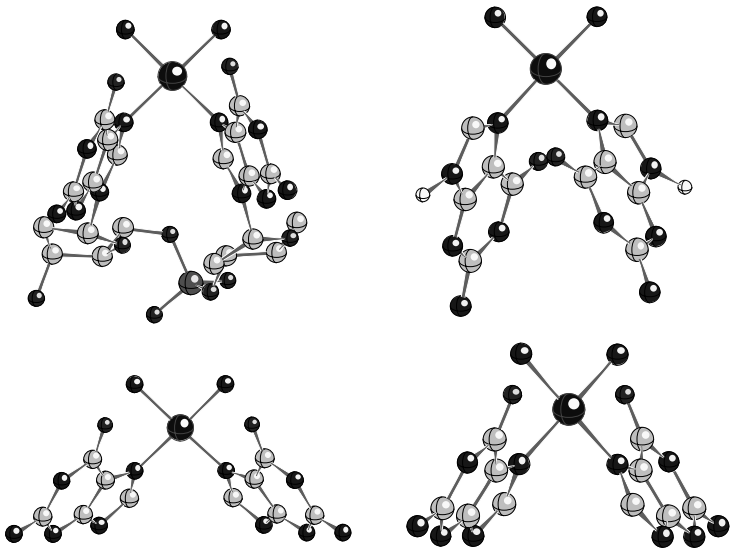


Figure 23. Experimental (top) and calculated (bottom) structures of the head-to-head (left) and head-to-tail (right) arrangement for the final product of two Guanine substitutions.

The reaction energy scheme for the double substitution of purine bases is shown below (fig. 24). The starting point has been arbitrarily set to zero and is the sum of the individual reactant energies, i.e. $E_{start} = E_{purine1} + E_{purine2} + E_{cisplatin(diaquated)}$. The energies of the individual purine molecules are included in table 3.

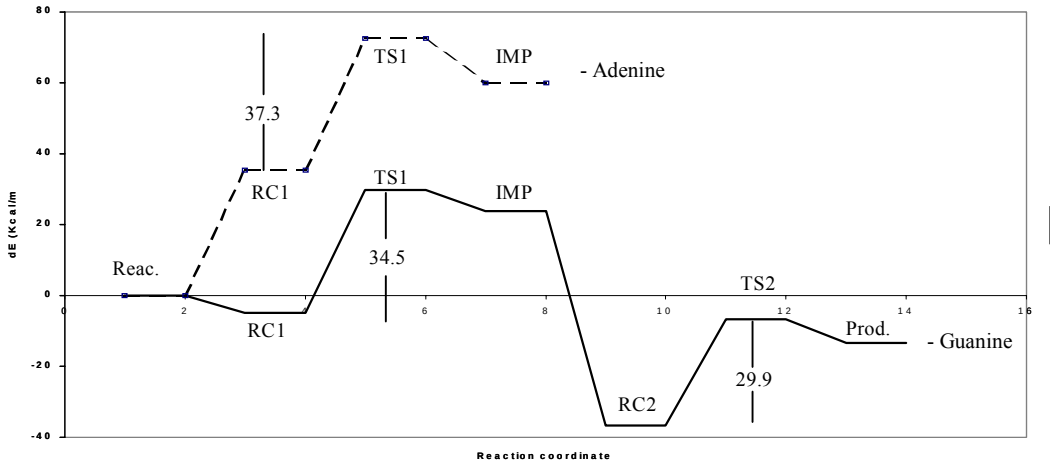


Figure 24. Reaction energy scheme for the purine substitutions.

Discussion

The primary goal of this investigation has been to gain qualitative insights in the reaction mechanism of the purine substitutions and to relate them to experimental observations. An added bonus has been the elucidation of the energetics of cisplatin's aquation. In the introduction some experimental data were presented, but a recapitulation is in order to serve as a basis for the discussion:

- There is evidence that ciplatin does not enter the nucleus in its chlorinated form but rather as the doubly water substituted form Pt $[NH_3]_2[H_2O]_2^{2+}$
- The AG adduct is direction specific, 5'-AG-3'. No exception to this directionality has been found.
- No monofunctional adducts to A has been detected.
- Interstrand adducts are present, albeit in relatively low frequency.
- The 5'-GXG-3' adduct (at ~6% frequency) indicates high flexibility of the DNA molecule, even locally.

The disposition of this section will follow that of results with conclusions included at the end.

Aquation of cisplatin

The main drawback of these calculations is the model system, in that the stoichiometry of the optimizations is not kept constant. This results in a basis superposition error, i.e. the energy of two individually optimized molecules is not the same as the energy of the two molecules if they are optimized simultaneously. This is due to the fact that an increase of system size will in effect have more basis functions ($N_{tot}=N_A+N_B$, N_X denotes the number of basis functions for an individual molecule) available for the description of the total system. A description using more basis functions is in general more accurate which in turn reduces the calculated energy of the system. However, in the present case this error should be quite small since the difference in system size compared to a stoichiometrically correct one is not very large. Another error that should be mentioned is that the optimizations are not performed with the PCM included.

Reasonably, the effects on energies should be rather small, but geometries should be somewhat altered. This is easily seen and understood when looking at the geometry of the second transition state, in which the leaving chloride ion is symmetrically positioned between the two water ligands to achieve maximum stabilization by hydrogen bonding. In an aqueous environment these hydrogen bonds are probably provided by the solvent water molecules, thereby reducing the strain while still providing the same number of hydrogen bonds to chloride. The geometry of the second TS should then more closely resemble that of the idealized TS.

What impact would these errors have on the calculated energies? It is reasonable to assume that, at the different stages of the two substitutions, the basis set superposition error will not be very significant, the reason being that while there is a change in stoichiometry the number of basis functions is kept relatively unchanged, i.e. Cl^- ($N_{b.f.}$=8) is replaced by H_2O ($N_{b.f.}$=13) in the second substitution, thereby not significantly changing the number of basis functions used throughout the two substitutions. The PCM method has been proven very accurate, and errors introduced by this method is generally in the order of ~1 kcal/mol[27] for ions and even less for neutral molecules. It is interesting to note that although the second TS is skewed compared to the first, the difference in bond lengths of entering and leaving ligands is not very large (< 1.6%), perhaps reflecting a phenomena called 'the invariance of the transition state', even though the magnitude of the charge separation is increased by 50 % in the second transition state. In passing, it may be added that when employing the PCM in the optimizations of the second TS, the dissymmetry vanished and the structure closely resembled that of an idealized TS. However, the optimization did not converge due to small oscillations in the optimization procedure (seemingly indefinite), i.e. alternations between two close conformations of the molecule, and the results are therefore excluded in this presentation.

To summarize: The calculations have shown that the aquation of cisplatin is feasible from a thermodynamical point of view, at least within a quantum chemical context. (The rule of thumb is that a reaction is spontaneous if the barrier is ~20 kcal/mol at 25^{o}C. In general, a barrier height of 23 kcal/mol is considered to give three reactions per hour and molecule at room temperature.) Furthermore, when examining the rate law governing this type of substitution, it is evident that the low physiological chloride concentration of the cytosol (~60 nM) enables at least the monoaquated species to form. This conclusion is also supported by recent kinetic experiments[28

where the monoaquated species was prevalent. The rate of formation of the diaquated species was, in that study, half that of the monoaquated species, in line with the calculated energetics of the present study. However, the presented results, especially the first energy barrier, diverge from experimentally determined barriers[39, 40] (~20 kcal/mol) and a previous theoretical study[41] that suggested the first barrier height to be 20.5 kcal/mol. The main difference in that study and the present is the choice of method. In their case they used a Car-Parinello-DFT approach (It can be seen as a quantum mechanical-molecular dynamics hybrid.) employing the same functionals as in this study. They furthermore modeled the solvent using explicit water molecules whereas this study uses a continuum representation. The transition state geometry in that survey slightly diverges from the above presented, most notably the distance of the attacking water molecule was determined to be ~2.7 Å while our study suggest a distance of 2.50 Å. A Car-Parinello type calculation permits no traditional relaxed scans of distances however, and whether their study or this has the most accurate results can only be determined by more calculations. It is certainly so, that the barrier height they suggest is closer to the experimentally determined one but because the methods employed are so different, direct comparisons are difficult to make. An expansion of our model system to explicitly include the first solvation shell (i.e. adding water molecules) could possibly shed some light on this issue. In addition it can be said that BLYP generally gives too low barriers whereas HF gives too high in comparison with experiments.

Interactions between diaquated cisplatin and the DNA bases Guanine and Adenine.

The main error in the calculations of cisplatin-DNA interactions is naturally the small size of the model system compared to the real life situation and the conclusions made in this section must be seen as indications of what would take place in the full size system. When considering the energetics, one should be aware that the basis set superposition error has an impact, the magnitude of it uncertain, especially when the model system is expanded to include the second purine. This means that the energy levels for the first substitution should be lowered in comparison with the second substitution and its reactant and product complexes. Another error is naturally that no solvation model has been included in these calculations. The reasons for this are mentioned in the results section, item 2.0.

As can be seen in the energy diagram (fig. 24), the difference in the barrier height (RC1-TS1) when comparing the first substitution of Adenine to that of Guanine is rather small (~37 kcal/mol for A and ~34 kcal/mol for G) so there being no monofunctional adducts formed to A cannot be explained from the energetics of this substitution. This should perhaps not be a surprise since the two purines are so alike. The answer to why no monofunctional adducts are formed to Adenine is, most likely, to be found in either the reactant complex formation in a model system of larger size or the stabilizing effect a larger piece of the DNA chain could (possibly) have on their respective transition states. When looking at the DNA chain from the major groove (see fig. 4), the main difference between A and G is that Guanine has its hydrogen bond acceptor O6 readily exposed to a putative hydrogen bond donor whereas N6 of Adenine is shielded by the vibrational motions of its non-basepairing hydrogen. This could possibly constitute enough hampering for the Adenine reactant complex formation to direct the first adduct formation to guanine. Also, an examination of their respective reactant complexes indicates that the Adenine RC would distort the tertiary structure of DNA to a larger extent than would the Guanine RC and hence be energetically unfavourable. The reaction energy scheme shows that the formation of the Adenine reactant complex (RC1) is endothermic but due to the small system size no definite conclusions can be made as to if this complex can be formed in a full size system.

The other alternative answer to the discrimination in favor of Guanine could, as mentioned above, be that the surrounding parts of DNA itself differently stabilize their respective transition states. Evidently DNA has enough flexibility to ‘wrap’ around cisplatin, as indicated in the last item of the facts mentioned in the introduction to this section, and has many hydrogen bond donors and acceptors as well as a negatively charged phosphate backbone to facilitate a reduction of the transition state energy. It is however difficult to see why the transition state of Guanine would be more stabilized than its Adenine counterpart. A model system of larger size could probably give the answer to these questions.

Figure 19 shows some features that deserve an explanation. Especially striking is the deprotonation of the leaving water ligand in the reactant complexes. Since the optimizations were performed *in vacuo* the water ligands of the positively charged cisplatin entity are acidified to an extent that makes the waters prone to deprotonation. This, in turn, reduces the energy of the reactant complexes much more than would be the case if the optimizations were done with a

solvation model present, in which case the deprotonation would not be as prominent. If the inclusion of a solvation model would indeed reduce the barrier height is difficult to judge. The barrier height is much too high for the reaction to proceed spontaneously and since the net charge is constant throughout the reaction the inclusion of a solvation model should most likely just shift the energy levels of the different states downwards by the same magnitude. This could imply assistance by the surrounding DNA molecule and, as previously discussed, obviously DNA can 'wrap' itself around an attacking cisplatin molecule.

The second purine substitution is more difficult to discuss, as there is a great lack of data here and most of this section will consist of speculations. One thing can however be said with reasonable certainty: In all of the above substitutions presented the attacking ligand has occupied the position of the leaving ligand. This said, a reasonable guess is that the interstrand adduct (i.e. head-to-tail arrangement) is formed from a monofunctional adduct of the second type in fig. 20 (conformer 2) and consequently the intrastrand adduct (i.e. head-to-head arrangement) is formed from a monofunctional adduct of the first type. Supposing that the first substitution of Guanine is correctly described by these calculations, the most abundant intermediary product should be conformer 1 of fig. 20, which would explain the high frequency of intrastrand adducts. As a consequence the low frequency of interstrand adducts would then be related to the occasions where conformer 2 is formed, presumably by a rotation around the Pt-N7 bond. The energy difference between the two transition states of the second Guanine substitution (Guanine attacking conformer 1 and 2 respectively) is only -0.85 kcal/mol but there would be a barrier height for the rotation around the N7-Pt bond of not less than the energy it takes to break the hydrogen bond to O6 separating the two conformers.

The geometry of the final products of these substitutions diverge somewhat from structures determined in experiments since no strain is put on the model system by the surroundings. In the intrastrand case, not including the phosphate backbone in the calculations results in a wide gap between the two guanines and the two hydrogen bonds between the O6's and the ammine groups of cisplatin are retained, whereas the experimental structure has a rather narrow gap between the Guanines and the hydrogen bonds are missing. Furthermore the displacement of the platinum centre out of the plane of the guanines is not present in the calculated structure. Overall, however, the geometries show close structural similarity (e.g. head-to-head arrangement). The interstrand

adducts, calculated and experimentally determined, are not as similar as in the intrastrand case but the major feature of this adduct, i.e. the head-to-tail arrangement of the bases, is retained. This is also a result of not including the surroundings in the calculations (This is becoming a bit of a mantra by now.). As in the intrastrand case hydrogen bonds are not present between the O6's and ammines in the experimentally determined structure while they are in the calculated structures. The displacement of the Platinum atom out of the Guanine planes evident in the experimental structure is not present in the calculated counterparts since the strain imposed on the adduct by the surroundings is not present.

The direction specificity of the 5'-AG-3' intrastrand adduct cannot, due to the small size of the model system, be addressed by the findings in this study. But some speculations can be interesting. The directionality could be due to a universal reaction order for the intrastrand didentate adduct formation, i.e. all didentate adducts formed could have to proceed via monofunctional adduct formation on the 3' side. This is certainly the case for the 5'-AG-3' adduct, since no monofunctional adduct have been found to Adenine. This hypothesis could probably be tested by experiments. It could also be that reactant complexes or transition states for the 5'-GA-3' adduct cannot form due to the different chemical environment experienced in the 3' direction at the location of Guanine.

Proceedings

In order to draw more definite conclusions concerning the reactions in this study, one would have to expand the model system. However, in an all quantum mechanical description this is not feasible. This problem could be circumvented by means of a hybrid, Quantum Mechanical/ Molecular Mechanics (QM/MM) approach. During the course of this study we have made attempts to utilize this functionality in the Gaussian program package, to very little success (none in fact). It seems, though, that present day codes are not mature enough to handle all types of problems. QM/MM is, in short, a method to connect the forces of a sub-system, accurately described by some Quantum Mechanical method, to a surrounding environment described by Molecular Mechanics. There are some successful applications of this method but there seems to be a general consensus that it is not yet a rigorous methodology.

Any rational improvements of the drug are a bit hard to picture. Replacing the ammine groups of cisplatin to more strongly trans-labilizing groups would probably increase the reactivity, however in both directions of the reaction, thus rendering the adducts unstable. A possible route could be to increase cisplatin's affinity for Adenine, in which case a higher degree of platination would result. This could possibly be accomplished by replacing the ammine groups with a hydrogen bond accepting group, e.g. an NO_2^- group. There is an obvious risk, however that this would ruin the delicate balance of cisplatin's geometry in relation to the purine bases that has become apparent during the course of this investigation. The delicacy of this balance is best understood by considering the fact that in the 35 years that has passed since its discovery, no improvements of the efficacy of cisplatin has been made.

Acknowledgements

I would like to thank the people in Leifs group for their helping hands and uplifting conversations (and sometimes a cigarette, thanks Maria). A big thanks to Martin Agback for helping me with the computers and an occasional gaming tip. Thank you Leif for taking time, even though you sometimes had little, to give advices and especially for your patience with this report. A very special thank you goes to Bo and Llano for valuable advice and company (plus an occasional song). I would also very much like to thank my parents for being there for me.

References

1) Comess, K.M.; Lippard S.J.; In *Molecular Aspects of Anticancer Drug-DNA Interactions*; Neidle, S., Waring, M., Eds.; MacMillan: London, **1993**; p 134.
2) Lippard, S.J.; *Science* **1982**, 218, 1075.
3) Lippard, S.J.; *et al. Nature* **1999**, 399, 708.
4) Wang, A.H.J.; Yang D.; *Prog. Biophys. molec. Biol.* **1996**, 66, p 81-111. And references therein.
5) Brabec, V.; Sip M.; Leng, M.; *Biochemistry* **1993**, 32, p 11676-11681.
6) Douglas, B.; McDaniel, D.H.; Alexander, J.J.; In *Concepts and Models of INORGANIC CHEMISTRY 2nd ed*.; Wiley & Sons: USA, **1983**; p 368-375.
7) Hall, A.J.; Satchell, D.P.N.; *J. Chem. Soc., Dalton Trans.* **1977**, p 1403.
8) Fichtinger-Schepman, A.M.J. *et al. Biochemistry* **1985**, 24, p 707-713.
9) Anderson, K.S.; *MutationRes.* **1979**, 67, p 209-214.
10) Corda, Y.; Anin, M.-F.; Leng, M.; Job, D.; *Biochemistry* **1992**, 31, p 1904-1908.
11) Bradley, *et al.*; *Biochemistry* **1993**, 32, p 982-988.
12) Wang, A.H.J.; Yang D.; *Prog. Biophys. molec. Biol.* **1996**, 66, p 104
13) Coste, F. *et al.*; *Nucl. Acids Res.* **1999**, 27, p 1837-1846.
14) Legendre, F.; Bas, V.; Kozelka, J.; Chottard, J.-C. *Chem.-Eur.J.* **2000**, 6, p 2002-2010.
15) Lippard, S.J.; Gelasco, A.; *Biochemistry* **1998**, 37, p 9230-9239. And references therein.
16) Takahara *et al.*; *Nature* **1995**, 377, p 649-652.
17) Huang, H. *et al*; *Science* **1995**, 270, p 1842-1845.
18) Paquet, F.; Perez, C.; Leng, M.; *J.Biomol.Struct.Dyn.* **1996**, 14, p 67-77.
19) Born, M.; Oppenheimer J.R.; *Ann.Phys*. **1927**, 84, p 457.
20) Hartree, D.R.; *Cambridge Phil. Soc.* **1928**, 24, p 89.
21) Fock, V.; *Physik* **1930**, 61, p 126.
22) Szabo, A.; Ostlund, N.S.; *Modern Quantum Chemistry, 1st ed, Revised.* McGraw-Hill Publishing Company. **1989**.
23) McWeeny, R.; Sutcliffe, B.T.; *Methods of Molecular Quantum Mechanics.* Academic Press, London and New York, **1969**.
24) Parr, R.G.; Yang, W.; *Density-Functional Theory of Atoms and Molecules.* Oxford University Press New York Clarendon Press Oxford, **1989**. And references therein.
25) Tomasi, J.; Cammi, R.; Mennucci, B.; *Int.J.of Quant. Chem.* **1999**, 75, p 783-803.
26) Pacios, L.F.; Christiansen, P.A.; *J.Chem.Phys.* **1985**, 82, p 2664-2671.
27) Barone, V.; Cossi, M.; Tomasi, J.; *J. Chem. Phys.* **1997**, 107, p 3210.
28) Davies, M.S.; Berners-Price, S.J.; Hambley, T.W.; *Inorg. Chem.* **2000**, 39 (25), p 5603-5613.
29) Becke, A.D.; *Phys. Rev*. **1988**, 38, p 3098.
30) Lee, C.; Yang, W.; Parr, R.G.; In *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*; *Phys.Rev. B*, **1988**, 37, p 785.
31) Michlich, B.; Savin, A.; Stoll, H.; Preuss, H.; *Chem. Phys. Lett.*, **1989**, 157, p 200.
32) Dunning Jr, T.H.; Hay, P.J.; In *Modern Theoretical Chemistry*, Schaeffer, H.F. Ed. (Plenum, New York), **1976**, vol. 3, p 1.
33) Hay, P.J.; Wadt, W.R.; *J. Chem. Phys.* **1985**, 82, p 270.
34) Wadt, W.R.; Hay, P.J.; *J. Chem. Phys.* **1985**, 82, p 284.
35) Hay, P.J.; Wadt, W.R.; *J. Chem. Phys.* **1985**, 82, p 299.
36) Miertus, S.; Scrocco, E.; Tomasi, J.; In *Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of ab initio Molecular Potentials for the Prevision of Solvent Effects, Chem. Phys.* **1981**, 55, p 117.
37) Miertus, S.; Tomasi, J.; In *Approximate Evaluations of the Electrostatic Free Energy and Internal Energy Changes in Solution Processes, Chem. Phys.*, **1982**, 65, p 239.
38) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J.; *Chem. Phys. Lett.*, **1996**, 255, p 327.
39) Reishus, J.W.; Martin, D.S.; *J. Am. Chem. Soc.*, **1961**, 83, p 2457.
40) Perumareddi, J.R.; Adamson, A.W., *J. Phys. Chem*, **1978**, 72, p 414.
41) Carloni, P.; Sprik, M.; Andreoni, W.; *J. Phys. Chem. B*, **2000**, 104, p 823-835.