

Theoretical Aspects of Palladium-Catalyzed Direct Arylations of Heterocycles



Johnson Matthey
Catalysts



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Jacques Blignaut

Supervisor (Academic): Prof. H. Marques

Co-supervisors (Industrial): Dr G. Jones and Dr C. Perry

Declaration

I declare that this Dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

9th of October 2016 in Johannesburg

Abstract

In this work, commercially relevant palladium catalysed direct arylation reactions, in which the molecule oxazole is coupled to toluene were investigated computationally. These involve the Pd(II)/Pd(0) manifold. Each of the three hydrogen atoms of oxazole can be replaced by a toluene molecule, and as both mono- and di-substitution occurs there are 6 possible products. Experimental results, using the catalyst Pd(OAc)₂ in DMA solvent, show that both the presence and quantity of tri-tertiary butyl phosphine (P^tBu₃) ligand has a significant impact on the observed product distributions. The aims of this study were to discover why and to ascertain the rate determining step (RDS) of the catalytic cycle under various conditions.

All geometry optimisation and frequency calculations were conducted using the TPSSh hybrid DFT functional in conjunction with the Def2-SVP basis set. Single point energy benchmarks at the DLPNO-CCSD(T)/Def2-TZVP level of theory revealed that the double hybrid (DH) functional DSD-PBEP86 in conjunction with the Def2-TZVP basis set was the best compromise between accuracy and efficiency for the systems at hand.

The generally accepted concerted metallation deprotonation (CMD) mechanism was investigated under phosphine free conditions, where a DMA solvent molecule was modelled bound to the palladium centre. The results correctly predicted the experimental product distributions for both mono- and di-substitution.

The CMD mechanism was also investigated under conditions of 1 equivalent of P^tBu₃ ligand. In this case a P^tBu₃ molecule was modelled bound to the palladium centre. The results correctly predicted the major mono-substituted experimental product and from the data formation of the major di-substituted product was readily rationalized.

Finally, oxidative addition (OA) was investigated under both conditions. It was found that the COSMO solvation model had a dramatic impact on OA reaction Gibbs energies. It was discovered that the presence of P^tBu₃ ligand significantly reduced the total Gibbs energy required for OA. In the absence of P^tBu₃, no DMA is coordinated to the palladium centre during OA. Regeneration of the active Pd(II) catalyst via deprotonation was the modelled completing the entire catalytic cycle.

The RDS under phosphine free conditions was calculated to be the CMD transition state (TS). Under condition of 1 equivalent of P^tBu₃ ligand the RDS was a mixture of both the CMD TS as well as coordination of oxazole to the active Pd(II) catalyst.

Acknowledgements

I would like to express my gratitude to my supervisors, Dr C. Perry and Professor H. Marques for their patience, insight and wealth of technical expertise, which they shared with me freely. In particular, I am indebted to Dr C. Perry for taking time out of his busy schedule and welcoming me to the Johnson Matthey technology centre for many fruitful discussions. It was a pleasure meeting all of the scientists there. To Professor H. Marques, for securing funding, especially for ESQC 2015, I am eternally grateful.

I thank Johnson Matthey as well as the University of Witwatersrand for financial support, without which this work would not be possible.

List of abbreviations:

ACM	Adiabatic Connection Model
AMBER	Assisted Model Building with Energy Refinement
AO	Atomic Orbitals
BOA	Born Oppenheimer Approximation
BSSE	Basis Set Superposition Error
CASPT2	Complete Active Space Self-consistent field with second-order Perturbation Theory
CC	Cooperative Catalysis
CC	Coupled Cluster
CCSD	Coupled Cluster Single Double
CCSD(T)	Coupled Cluster Single Double (perturbative Triple)
CCSDT	Coupled Cluster Single Double Triple
CHPC	Centre for High Performance Computing
CI	Configuration Interaction
CMD	Concerted Metallation Deprotonation
CNDO	Complete Neglect of Differential Overlap
COS	Chain Of Spheres
COSMO	Conductor-like Screening Model
CSD	Cambridge Structural Database
DCACP	Dispersion-Corrected Atom-Centred Potentials
DFT	Density Functional Theory
DFT-D	Density Functional Theory-Dispersion
DFT-D3BJ	DFT-D with Becke and Johnson damping
DH	Double Hydrid
DIIS	Direct Inversion of the Iterative Subspace
DLPNO	Domain-Based Local Pair Natural Orbital
DMA	Dimethylacetamide
DZ	Double Zeta
EC	Electron Correlation
ECP	Effective Core Potential
ESI	Electronic Supplementary Information
GB	Gigabyte

GC-MS	Gas Chromatography - Mass spectrometry
GGA	Generalised Gradient Approximation
GTO	Gaussian Type Orbital
HF	Hartree-Fock
INDO	Intermediate Neglect of Differential Overlap
K_B	Boltzmann's constant
KB	Kilobyte
KOH	Potassium Hydroxide
KS	Kohn Sham
LBFGS	Limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm
LDA	Local Density Approximation
LSDA	Local Spin Density Approximation
MAD	Mean Absolute Deviation
MBPT	Many-Body Perturbation Theory
MEP	Minimum Energy Path
MM	Molecular Mechanics
MO	Molecular Orbital
MP2	Second order Møller–Plesset perturbation theory
MPI	Message Passing Interface
MRCI	Multi-Reference Configuration Interaction
MRSA	Methicillin-Resistant Staphylococcus Aureus
MS	Mass spectrometry
MUE	Mean Unassigned Error
NDDO	Neglect of Diatomic Differential Overlap
NEB	Nudged Elastic Band
NMR	Nuclear Magnetic Resonance
NSAID	Non-Steroidal Anti-Inflammatory Drug
OA	Oxidative Addition
OS	Operating Systems
PES	Potential Energy Surface
PP	Pseudopotential
PPAR	Peroxisome Proliferator-Activated Receptor
RAM	Random Access Memory
RDS	Rate Determining Step
RE	Reductive Elimination

RI	Resolution of the Identity
RMSD	Root Mean Squared Standard Deviation
SAS	Solvent Accessible Surface
SCF	Self-Consistent Field
SCS-MP2	Spin-Component-Scaled Second order Møller–Plesset perturbation theory
SD	Slater Determinant
SIE	Self-Interaction Error
STO	Slater Type Orbital
SV	Split Valence
SVP	Split Valence Polarisation
T	Temperature
TD-DFT	Time-Dependent Density Functional Theory
THF	Tetrahydrofuran
TS	Transition State
TZ	Triple Zeta
TZV	Triple Zeta Valance
TZVP	Triple Zeta Valance Polarisation
vdW-DF	van der Waals Density Functional
XC	Exchange Correlation
ZPE	Zero-Point vibrational Energy

Table of Contents

1. Introduction.....	1
2. Methodology.....	48
3. The CMD Mechanism: Ligand-less Conditions	83
4. The CMD Mechanism: Tri-tert-butylphosphine ligand.....	211
5. Oxidative Addition	270
6. Conclusions.....	319
7. Future work	322
8. Appendix A	325
9. Appendix B	356

1. Introduction

Contents

1.1 The Significance and Synthetic Preparation of Bi-aryl Compounds.....	3
1.2 Aspects of Computational Chemistry	8
1.2.1 Force Field Methods.....	8
1.2.2 <i>Ab initio</i> methods	10
1.2.2.1 The Born–Oppenheimer Approximation	10
1.2.2.2 The Hartree- Fock Independent Particle Method	12
1.2.2.3 The Basis Set Approximation	15
1.2.2.4 Electron Correlation and Relativistic Approaches	20
1.2.2.5 Density Functional Theory	23
1.2.2.6 Solvation Models.....	26
1.3 The Direct Arylation of Oxazole	27
1.3.1 Industrial Applications.....	27
1.3.2 Experimental Results Courtesy of Johnson Matthey	31
1.3.2 The Concerted Metallation Deprotonation Mechanism.....	38
1.3.3 The Cooperative Catalysis Mechanism	41
1.3.4 The Catalytic Cycle	43
1.4 Aims.....	45
References	46

This project concerns the theoretical modelling of reactions in which direct arylation of heterocyclic aromatic compounds takes place and a bi-aryl bond is constructed. In particular, the direct arylation of the molecule oxazole was studied using theoretical methods. This is in the hope that such calculations can provide valuable insight into atomic processes and offer ways to rationalise experimental results and predict the outcome of experiments yet to be performed. In this way, a computer can be used as an experimental “technique” per se to explore reality and generate useful data without physically having to induce chemical transformations and expend the associated material resources.

1.1 The Significance and Synthetic Preparation of Bi-aryl Compounds

The bi-aryl structural motif is a predominant feature in many biologically active molecules and is found in numerous industrially relevant pharmaceuticals and agrochemicals. Examples include the nontoxic agricultural fungicide Boscalid¹, the hypertension management drug Telmisartan², Botulinum neurotoxin inhibitor F4H³, the broad spectrum herbicides Paraquat and Difenzoquat⁴ as well as the non-steroidal anti-inflammatory drug Diflunisal.⁵ The structures of these molecules are shown in Figure 1-1, where the bi-aryl bonds are highlighted in red.

Palladium-catalysed direct arylation is emerging as a valuable and effective alternative to traditional metal-catalysed cross couplings in the construction of bi-aryl compounds. This is due to a number of factors. These include the requirement for only one reactant to be functionalised, avoiding the use of stoichiometric amounts of expensive organometallic reagents and the subsequent toxic metal waste. Additionally, direct arylation has the advantage of reducing the number of synthetic steps necessary and offering the possibility of higher yields when compared to the more understood and developed traditional cross coupling methods, which typically employ organometallic reagents in addition to the catalyst.

In order to introduce the scope of this work in an historical context, a brief survey of the development of transition metal catalysis, in particular the use of palladium in the synthesis of bi-aryl compounds, is necessary. The first known reported use of a transition metal catalyst in the construction of a bond between two aromatic nuclei was reported in 1901 and is now famously known as the Ullmann reaction.⁶ In this reaction, two equivalents of aryl halide are reacted with one equivalent of finely divided copper at temperatures above 200 °C to form a bi-aryl molecule and a copper halide. It took 71 years before a transformation with similar efficiency was reported using nickel, which involved the coupling of Grignard reagents with

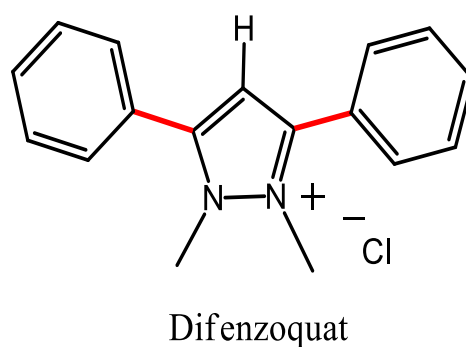
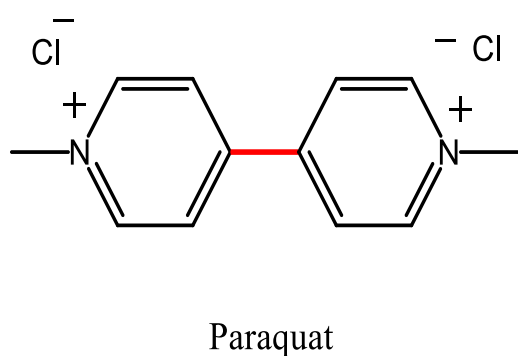
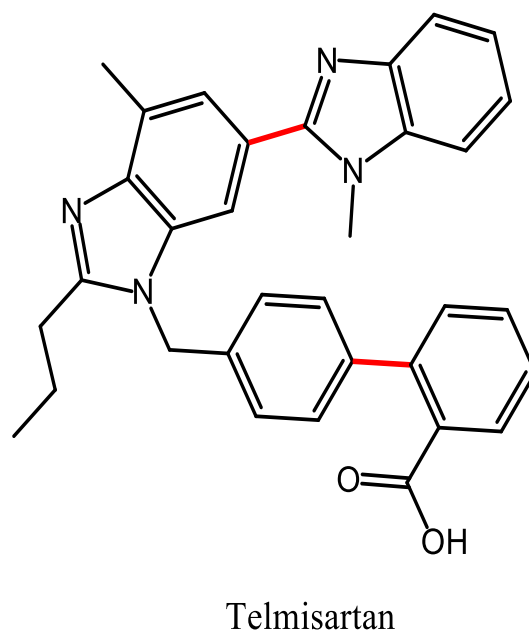
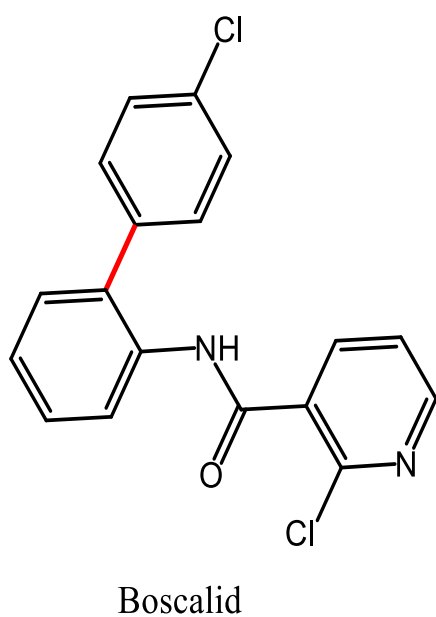
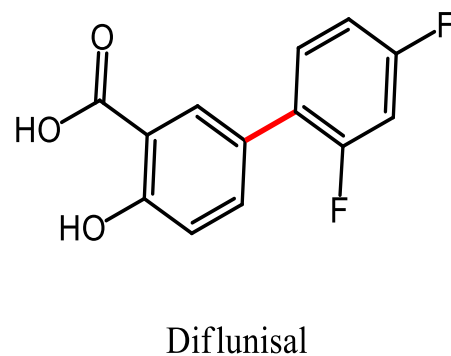
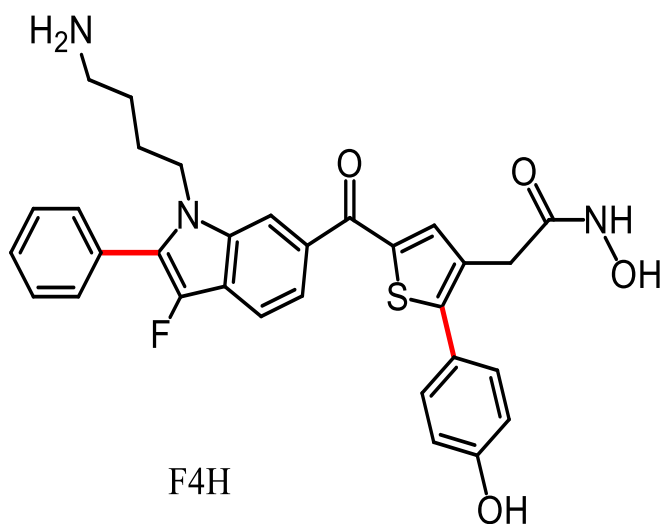
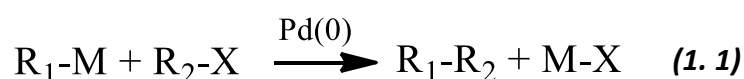


Figure 1-1: Industrially Relevant Examples of Bi-aryl Molecules (bi-ary bond highlighted in red)

organohalides. This discovery was made independently in 1972 by Robert Corriu⁷ and Makoto Kumada.⁸ The transformations were more selective and run under milder conditions than had been achieved using copper. However, the most significant advancement that resulted from

the use of nickel was the subsequent broadening of the scope of the substituents on the aromatic moieties, allowing for far greater diversity of target molecules attainable via these methods.^{9, 10} A few years later, the first reports of similar reactions catalysed by palladium were disclosed.^{11, 12} The use of palladium would develop to make it overwhelmingly the most popular transition metal catalyst. This is due in part to the incredibly low catalyst loadings required when palladium is used, in some cases as low as 0.1 mol %¹³, making these processes industrially very valuable. Furthermore, palladium complexes generally exhibit high stability, allowing for indefinite storage and facile handling. The most developed and popular method employing palladium is the cross coupling of organometallics and halides. This is referred to in the literature as “traditional” cross coupling methods. In these methods the palladium catalyst is in the 0 oxidation state, although in many cases this Pd(0) species may be generated *in situ* from Pd(II) complexes. The general scheme is outlined in Equation (1. 1).



Here R_1 and R_2 are the aromatic moieties while M is a metallic functional group and X is a halide or triflate. A vast variety of metals have been used in the reactants including but not limited to MgX, ZnX, Cu, SnR₃, SiR₃, Zr, Al as well as B(OH)₂.⁹ However, the most famous varieties deserve special attention. They make use of tin and boron organometallic compounds and are known as the Stille¹⁴ and Suzuki¹⁵ coupling reactions, respectively. These relatively mild, versatile reactions are tolerant of a wide variety of functional groups on either coupling partner, and are additionally stereospecific and regio-selective as well as being high yielding. They have thus been adopted widely for the synthesis of elaborate organic molecules. These two reactions have opened up synthetic routes to many important compounds and have had a massive impact on organic chemistry. In particular, the use of boronic acids as the organometallic reagent, introduced by Suzuki and colleagues, represented a major step forward. This discovery was honoured with the 2010 Nobel Prize in chemistry. Boronic acid derivatives present several advantages over other organometallic reagents including their aforementioned ability to tolerate a bewildering range of functional groups as well as the low toxicity of the reagents and by-products, especially when compared to tin-containing compounds. Toxic tin by-products can be difficult to remove completely from the final product, which can be prohibitive for the synthesis of pharmaceuticals as the tolerance for such impurities is often in the parts per million (ppm) range.¹⁶

For example, one of the most common by-products of Stille coupling reactions are tributyltin halide compounds, especially tributyltin chloride. These compounds, before being banned, were used globally in anti-fouling paint on the hulls of ocean going vessels to prevent the growth of aquatic organisms that attach to the hull and affect a vessel's performance and durability. Subsequently they have been shown to have immunotoxic, neurotoxic, mutagenic as well as endocrine disrupting effects in organisms ranging from invertebrates such as marine snails to mammals such as whales and humans.¹⁷

The next major development in the construction of bi-aryl compounds has been direct arylation. In the most advanced form of direct arylation, which represents the pinnacle of bi-aryl methodology reached so far, no organometallic reagent (apart from the catalyst) is required. The reaction can be performed simply with an aryl halide and an unfunctionalised aromatic compound. The scheme in Figure 1-2 contrasts traditional cross coupling with direct arylation.

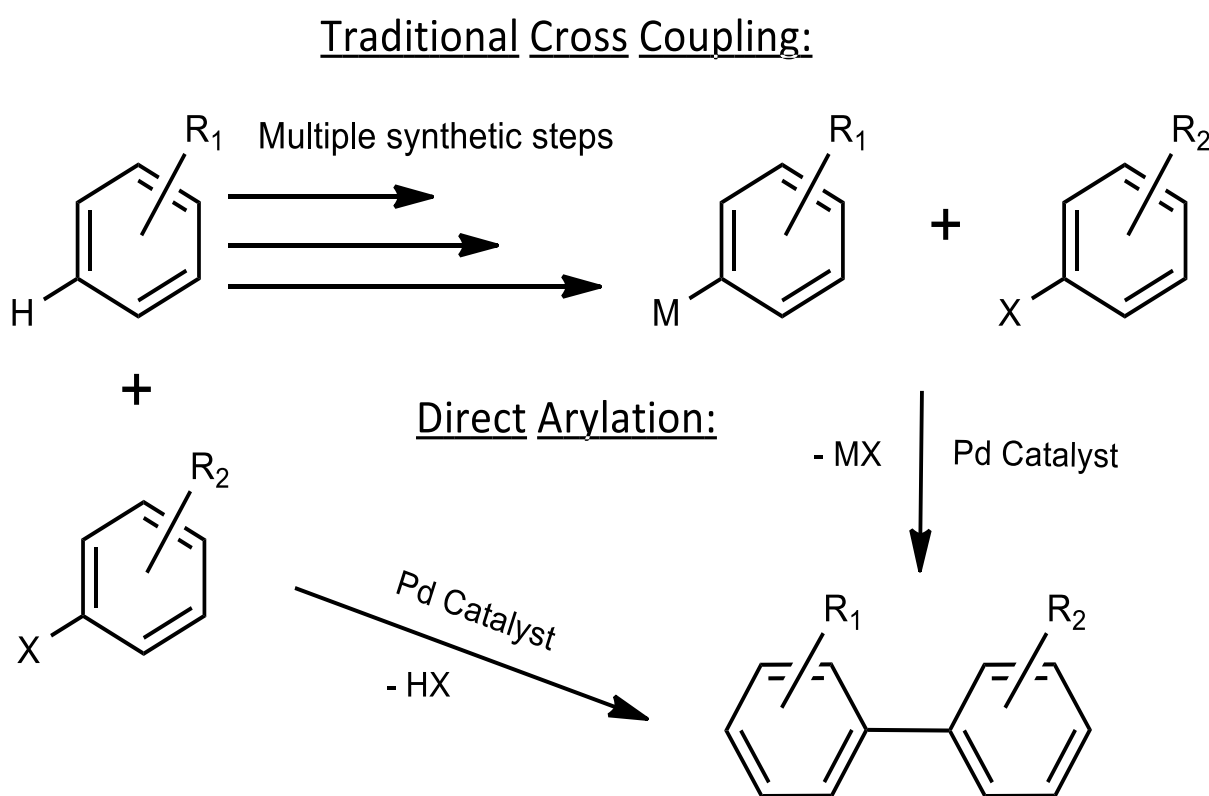


Figure 1-2: Traditional Cross Coupling vs Direct Arylation

Direct arylation reactions themselves can be differentiated based on the nature of the coupling partners involved. The two major divisions are oxidative direct arylations and direct arylations involving halides or pseudo halides. Figure 1-2 shows only the latter, which for reasons discussed below, are generally preferred. Oxidative direct arylations themselves appear in two distinct forms; both inevitably require the use of sacrificial oxidants. The first

involves the coupling of an organometallic reagent (M) and an unfunctionalised arene. The second subcategory involves the coupling of two unfunctionalised arenes and is also known as dehydrogenative arylation. These are contrasted in Figure 1-3 . In dehydrogenative oxidative direct arylation the hydrogen atoms become incorporated into the sacrificial oxidant.

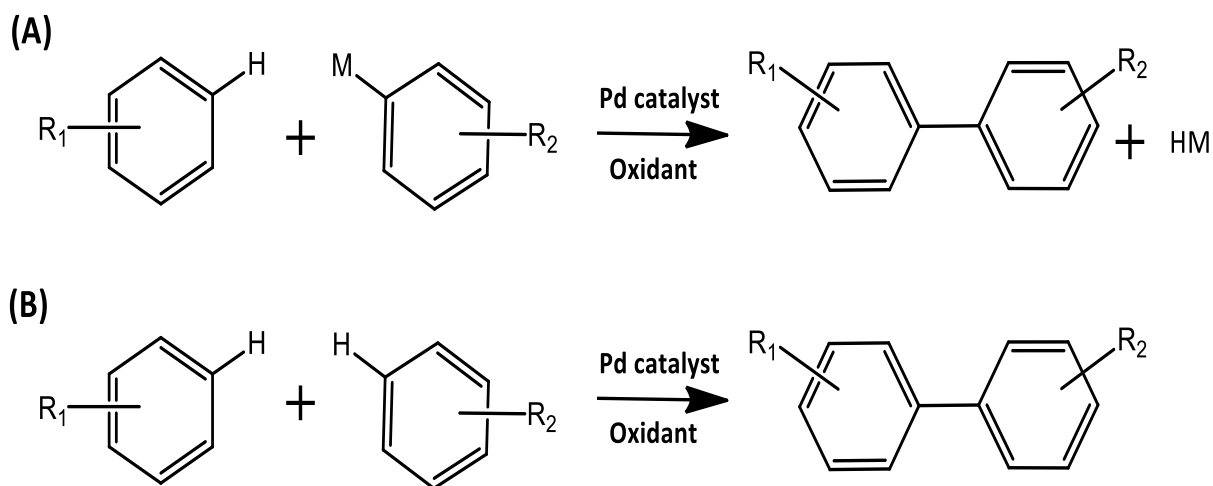


Figure 1-3: (A): Organometallic Oxidative Direct Arylation (B): Dehydrogenative Oxidative Direct Arylation

The sacrificial oxidants are often costly metal compounds, such as AgOAc, that are required in stoichiometric concentrations. Although some reactions have been reported in which the oxidant is atmospheric oxygen, this requires the use of expensive catalysts such as $\text{H}_4\text{PMo}_{11}\text{VO}_{40}$.¹⁸ Additionally, over-oxidation of the product molecules complicates these methods, making some industrially important substrates, such as indoles, incompatible with this methodology.¹⁸ The use of an organometallic reagent gives rise to the same complications as traditional coupling, such as the extra synthetic methodology required to functionalise one coupling partner with an organometallic group as well as the problematic metallic waste.

Additionally, the synthesis of the organometallic reagent generally involves a number of synthetic steps, during which undesired by products are formed.¹⁹

The coupling of two unfunctionalised arenes in dehydrogenative oxidative direct arylation may at first glance seem immensely promising. Producing higher value chemicals from unfunctionalised molecules, available directly from petrochemical feedstocks without the generation of harmful by products is surely the zenith of modern catalysis. However, besides the disadvantages of using oxidants mentioned above, there is the massive problem of regioselectivity, in terms of which hydrogen atoms in the molecule are replaced by the bi-aryl

bond. The ubiquitous nature of hydrogen in organic matter ensures that achieving selectivity in such transformations remains a huge hindrance to the further development of this approach. Admittedly, if this lack of selectivity could be overcome, the advantages would be substantial. Presently, however oxidative direct arylations hardly represent an improvement over traditional cross coupling methods.

In contrast, direct arylations involving one functionalised halide/pseudo halide and an arene do not suffer from the disadvantages mentioned above. No oxidant is required, the synthesis of organometallic reagents is not necessary and no organometallic waste is produced. Furthermore, a high selectivity is often possible. Additionally, the by-products are often commercially valuable acids such as HCl. Such methods thus rank among the most advanced and efficient ways yet developed to construct bi-aryl bonds and will be the central focus of this project.

1.2 Aspects of Computational Chemistry

A huge number of methods exist for the atomistic modelling of systems. The spectrum of techniques and approximations spans vastly different accuracies and computational costs. The approach taken is often a consequence of considering the available computational resources, the type of information one ultimately requires and the nature of the system at hand. Due to approximations made in the formalism of the theory, some methods are inherently better suited for certain systems where such approximations hold true, rather than others where they are relaxed. What follows, is a brief mention of the various methods that are of relevance to this project, including their strengths and weaknesses.

1.2.1 Force Field Methods

On the less computationally expensive side of the spectrum lie the force field methods, also known as molecular mechanics (MM) methods. These are empirical techniques that require the input of experimental data. Bonding information must thus be provided externally, rather than as a result of solving the electronic Schrödinger equation. In addition to bypassing the solution of the electronic Schrödinger equation, the quantum aspects of the nuclear motion are also neglected. The motion of the atoms is treated completely by classical mechanics, i.e. Newton's second law. In this approach, the electronic energy is expressed as a parametric function of the nuclear coordinates. These parameters are obtained by fitting the equations involved to experimental thermochemical data, and more recently also to data obtained from computational calculations performed at higher levels of theory, often coupled cluster

methods. Molecules are described by a ball and spring model, with the parameters mentioned above generally being the stiffness of the bonds to different atoms, expressed by a force constant, and the length of the bond. Additionally, there are also parameters to account for the polarizability and size of the atoms. Thus, the building blocks in MM are atoms; electrons and nuclei are not considered as individual particles. Depending on the atomic number and the type of chemical bonding involved, a given atom has tabulated parameters that are stored on disk and used during calculations. Each different atom is identified as a type. There are many different types depending on the exact implementation of MM; as an example a few common atoms types from the MM2(91) implementation²⁰ are shown below in Table 1-1. The difference in the various implementations of MM is usually a matter of the functional form of the energy term and the data used to fit the parameters. Some methods have been developed especially for large biological systems, such as AMBER.²¹ In these cases the functional forms have been kept as simple as possible and hydrogen atoms are not modelled explicitly. In other implementations the goal has been to reproduce the properties of small molecules with a high degree of accuracy. In these cases the functional forms are more complex, hydrogen atoms are modelled explicitly and they will generally be more computationally expensive for a system of a given size.

Because large amounts of experimental data are available for determining the value of the parameters, the results of a MM calculation for simple hydrocarbons may be as good as that obtained from higher levels of theory, but at a fraction of the cost.²² However, today such semi empirical methods are rarely used in isolation other than for describing ensembles of particles or large macromolecules, and have largely been superseded by *ab initio* methods. This is predominantly as a consequence of the rapid increase in computational power available to the average scientist over the last few decades.

Table 1-1: Selected MM2(91) atom types

Type	Element Symbol	Atoms chemical description
1	C	sp ³ carbon
50	C	sp ² aromatic carbon
56	C	sp ² carbon, cyclobutane
46	N	nitro group nitrogen (R-NO ₂)

1.2.2 *Ab initio* methods

In order to obtain the properties of a molecule with greater accuracy, one must attempt to describe the system of interest at a more fundamental level than that in which the basic units are atoms. A molecule can be considered to be comprised of interacting nuclei and electrons. If we wish to describe the electrons, even if only qualitatively, then we must invoke quantum mechanics. However, the equations involved can only be solved analytically for one electron systems (such as H_2^+) and thus for all chemically relevant systems one must turn to a number of computationally intense methods that are based on approximations. These will be reviewed here briefly, in a mathematically light handed manner simply to emphasise their importance.

1.2.2.1 The Born–Oppenheimer Approximation

One of the fundamental approximations on which quantum chemistry is based is the Born–Oppenheimer approximation (BOA) first proposed in 1927.²³ This is an assumption that allows for simplification of the Schrödinger equation for a molecule. The total wavefunction for a molecule Ψ_{Total} is a function of the nuclear coordinates ($\mathbf{R}$) as well as the coordinates of the electrons ($\mathbf{r}$), Equation (1. 2).

$$\Psi_{\text{Total}} = \Psi(\mathbf{R}_1, \mathbf{R}_2 \dots \mathbf{R}_N, \mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_n) \quad (1. 2)$$

The complete non-relativistic Hamiltonian for a molecule consisting of n electrons and N nuclei can be expressed as:

$$\hat{H} = -\frac{\hbar^2}{2} \sum_I \frac{\nabla_I^2}{m_I} - \frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 - \sum_i \sum_I \frac{Z_I e'^2}{r_{Ii}} + \sum_{I>J} \frac{Z_I Z_J e'^2}{r_{IJ}} + \sum_{i<j} \frac{e'^2}{r_{ij}} \quad (1. 3)$$

Where:

$$e'^2 = \frac{e^2}{4\pi\epsilon_0}$$

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

$\hbar$ is Plank's constant, m_e is the mass of an electron, m_I and Z_I refer to the mass and atomic number of nucleus I respectively. Finally, r represents the distance between the particles. Lower case indices refer to the electrons and upper case indices refer to the nuclei.

The terms from left to right in Equation (1. 3) refer to the kinetic energy of the nuclei, the kinetic energy of the electrons, the attractive potential between nuclei and electrons, the repulsive nuclear-nuclear potential, and finally the electron-electron repulsion. The Hamiltonian in Equation (1. 3) contains pairwise attraction and repulsion terms, implying that no particle is moving independently of all of the others. The term “correlation” is used to describe this interdependency of motion. A proton and electron both possess the same magnitude of electrical charge, namely the fundamental unit of charge. The forces on both electrons and nuclei due to this electric charge are thus of the same order of magnitude. However, this is not so for the changes which occur in the momentum of the particles as a result of these forces. Even the lightest nucleus, in a hydrogen atom, is approximately 1800 times heavier than an electron. For a uranium nucleus this ratio increases to 430 000. This immense difference in mass indicates that the nuclear velocities will be much smaller than the velocities of the electrons as the same force applied to a much heavier object will cause a smaller change in momentum as when applied to a much lighter object. The Born–Oppenheimer approximation assumes that the nuclei are infinitely heavier than the electrons and thus the nuclear kinetic energy term is taken to be independent of the electronic motion. The electrons can thus adjust rapidly to changes in the nuclear geometry, essentially instantaneously at the limit that they are infinitely lighter than the nuclei. This allows for the separation of the nuclear coordinates from the electron coordinates in the wavefunction:

$$\Psi(\mathbf{R}_1, \mathbf{R}_2 \dots \mathbf{R}_N, \mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_n) = \Phi(\mathbf{R}_1, \mathbf{R}_2 \dots \mathbf{R}_N) \psi(\mathbf{R}_1, \mathbf{R}_2 \dots \mathbf{R}_N, \mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_n) \quad (1. 4)$$

where $\Phi(\mathbf{R}_1, \mathbf{R}_2 \dots \mathbf{R}_N)$ represents the nuclear wavefunction while $\psi(\mathbf{R}_1, \mathbf{R}_2 \dots \mathbf{R}_N, \mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_n)$ represents the electronic wavefunction. Physically the above separation corresponds to neglecting the coupling between the nuclear and electron velocities, i.e. neglecting their correlated motion. Thus the nuclei are stationary in the electron’s frame of reference. As a consequence of this separation of the nuclear and electronic motion, the electronic wavefunction depends on the position of the nuclei ($\mathbf{R}_1 \dots \mathbf{R}_N$) but not on their momentum. One can thus fix the nuclear coordinates and use this as input to solve for the energy of the electrons.

The non-relativistic Hamiltonian for a molecule consisting of n electrons and N nuclei after applying the Born–Oppenheimer approximation, expressed in atomic units, simplifies to:

$$\hat{H} = -\frac{1}{2} \sum_i^n \nabla_i^2 - \sum_i^n \sum_I^N \frac{Z_I}{r_{Ii}} + \sum_{i<j}^n \frac{1}{r_{ij}} + V^{nuc} \quad (1.5)$$

where the nuclear-nuclear repulsion energy (V^{nuc}) becomes a constant for a given geometry.

For two nuclei A and B, V^{nuc} has the form:

$$V^{nuc} = \frac{e^2}{4\pi\epsilon_0} \frac{Z_A Z_B}{R} \quad (1.6)$$

It is the above Hamiltonian in Equation (1.5) that is generally used in computational chemistry.

Solving the electronic Schrödinger equation this way for a large number of nuclear geometries generates what is referred to as a multidimensional potentially energy surface (PES). The nuclei move upon this potential energy surface and once the PES is known, their motion can then be solved to generate, for example, a vibrational spectrum. Critical points on this PES represent local minima as well as transition states. The concept of a PES relies on the validity of the Born-Oppenheimer approximation; without it we would have to resort to abstract discussions of probability densities of the nuclear-electron wavefunction rather than stable molecular configurations and transition state structures. Thus, without this approximation the concept of molecular geometry itself becomes blurred.

1.2.2.2 The Hartree- Fock Independent Particle Method

A significant simplification of the many bodied Schrödinger equation can be achieved if we consider the system of electrons and nuclei to be an independent particle problem. In such a system the motion of one electron is independent of the dynamics of all other electrons. The interaction between the electrons is then approximated by taking all repulsion into account in an average fashion. This is the essence of Hartree-Fock (HF) theory. Thus, the fundamental assumption made is to ignore the true correlated motion the electrons and the resulting energetic consequences. In HF theory each electron is described by an independent particle, one electron orbital (ϕ) and the total electronic wavefunction (ψ) is given as a product of these orbitals:

$$\psi(\mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_n) = \phi(\mathbf{r}_1) \phi(\mathbf{r}_2) \dots \phi(\mathbf{r}_n) \quad (1.7)$$

The Hamiltonian in Equation (1. 5) applied to this wavefunction would give the exact solution to the Schrödinger equation, within the limit of the Born-Oppenheimer approximation, if it was not for the electron-electron repulsion, which here is approximated. The exact, instantaneous electron – electron repulsion experienced by each electron is replaced with an effective field produced by the average position of the remaining electrons. This leads to a set of equations for each electron:

$$\left(-\frac{1}{2}\nabla_i^2 - \sum_I \frac{Z_I}{r_{Ii}} + V_i^{eff}\right)\phi_i = E_i\phi_i \quad (1. 8)$$

where ϕ represents the one electron Hartree-Fock orbitals and V_i^{eff} is the effective field felt by electron i as a consequence of the average repulsion of all other electrons. However, the total wavefunction must still obey the Pauli Exclusion Principle as two electrons cannot occupy the same spatial position. This antisymmetry condition is conveniently achieved by building the wavefunction from Slater determinants (Φ), Equation (1. 9), since interchanging any two rows or columns changes the sign of the determinant. This physically represents the fact that electrons have a spin of $\frac{1}{2}$ and are thus indistinguishable fermions and the wavefunction must change sign upon interchanging any two electrons.

$$\Phi_{SD} = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(e_1) & \phi_2(e_1) & \dots & \phi_n(e_1) \\ \phi_1(e_2) & \phi_2(e_2) & \dots & \phi_n(e_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(e_n) & \phi_2(e_n) & \dots & \phi_n(e_n) \end{vmatrix} = |\phi_1, \phi_2, \dots, \phi_n| \quad (1. 9)$$

An assumption is made that the wavefunction consists of only one Slater determinant. From this a trial set of functions ($\phi_1, \phi_2 \dots \phi_n$) is estimated to generate a trial wavefunction from the corresponding Slater determinant. However, solving for these functions is nontrivial as V_i^{eff} depends on all of the functions ($\phi_1, \phi_2 \dots \phi_n$). Thus, as the HF equations depend on their own solutions they must therefore be solved iteratively. This is done via a method known as the Hartree-Fock-Roothaan Procedure.²⁴ After assuming the trial set functions ($\phi_1, \phi_2 \dots \phi_n$), these are then used to produce a set of effective potential operators (V_i^{eff} - V_n^{eff}) and the HF equations are then solved to get the energy. The trial functions ($\phi_1, \phi_2 \dots \phi_n$) are then modified to minimise the energy and these new functions used to calculate an updated set of effective potential operators (V_i^{eff} - V_n^{eff}) which are used to solve the HF equations to get the new

energy. This process continues until the functions ($\phi_1, \phi_2 \dots \phi_n$) no longer change (to within some pre-determined convergence criteria) at which point convergence is said to have been reached. This iterative process is known as a self-consistent field (SCF).

The importance of the Hartree-Fock approach is that it serves as a kind of branching point from which different models have developed. The HF method gives the best possible single determinant wavefunction at the limit of the basis set being composed of an infinite number of functions. This is known as the HF limit, assuming no additional approximations. In spite of its fairly significant fundamental assumption of neglecting instantaneous electron-electron correlation, HF theory provides a well-defined stepping stone on the way to more sophisticated theories. When the HF limit is reached, the difference between the energy determined by this theory and the true energy is referred to as the electron correlation (E_{corr}):

$$E_{\text{corr}} = E - E_{\text{HF}} \quad (1. 10)$$

where E is the “true” energy and E_{HF} the energy at the HF limit. It should be noted that here relativity is neglected and that E is the true energy only if no relativistic corrections are necessary, within the limitations of the BOA. Thus, the aforementioned correlated motion of the electrons gives rise to an energy that is defined relative to the energy calculated at the HF limit for a given system.

In certain situations, such as isodesmic reactions and conformational changes, results from HF can be chemically meaningful. Fortuitously in many cases a favourable cancellation of errors occurs due to the error in energy associated with basis set incompleteness (discussed below) being opposite in sign and equal in magnitude to the error in energy that occurs from neglecting electron correlation. Surprisingly accurate calculations at times result from this technicality. However, the neglect of electron correlation results in significant discrepancies between calculated and experimental properties where such interactions are predominant and no cancellation of errors occurs. This is particularly true for heats of formation, and indeed any process where the nature of the bonding changes significantly. Even as one approaches the practical HF limit the inherent error in an absolute energy calculation for a molecular system can be substantial. For example, studies have shown mean unsigned errors (MUE), Equation (1. 11), of 85, 66, and 62 kcal.mol⁻¹ for the atomization energies of 66 small molecules at the HF level using the aug-ccpVnZ basis sets, with $n = \text{D, T, and Q}$ respectively.²⁵

$$MUE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (1.11)$$

In Equation (1.11) f_i and y_i are the predicted and experimental values, respectively. Thus, in general, HF theory is insufficient for the demands of a modern computational chemistry study.

In the early days of computational chemistry, when access to high performance computing resources was limited, carrying out HF calculations, even for small systems with small basis sets, was quite demanding. Thus, a great amount of effort was expended to develop methods that are less computationally demanding than HF, while implicitly taking into account correlation. This is done by neglecting all integrals involving more than two nuclei in the Hartree-Fock-Roothaan (SCF) procedure. These methods are known as semi empirical *ab initio* methods. As HF theory itself is based on the major simplification of neglecting instantaneous electron-electron correlation any further simplifications will lead to a poor model. However, the success of these semi empirical methods relies on tuning the remaining integrals into parameters that are fitted to experimental data such as molecular energies and geometries; this is the origin of the term *semi-empirical* methods. These methods are therefore limited to systems for which parameters exist. Complete Neglect of Differential Overlap (CNDO), Intermediate Neglect of Differential Overlap (INDO) and Neglect of Diatomic Differential Overlap (NDDO) are just some of the methods that have been developed using this approach. Due to the potential for linear scaling these methods are particularly promising for the modelling of large systems such as biomolecules and materials such as zeolites.

If more Slater determinates are used to construct the trial wavefunction, correlation can be accounted for. These results will eventually converge to an exact solution of the electronic Schrödinger equation. Such methods require much larger computing resources and are the subject of subsequent sections.

1.2.2.3 The Basis Set Approximation

The mathematical functions that make up the trial wavefunction are referred to as basis sets. Essentially the unknown molecular orbital (MO) functions are approximated by a linear combination of known atomic orbitals (AOs). In practice these are not generally solutions to hydrogenic Schrödinger equation but are themselves approximations to the AOs.

In theory, the HF limit is reached by use of an infinite basis set. In practice, this is not possible. A great amount of effort has been expended by the scientific community in order to develop basis functions that approach the HF limit arbitrarily closely, as efficiently as possible. In

electronic structural calculations of molecular species, two fundamental types of functions are used in the construction of basis sets, namely Slater Type Orbitals (STOs) and Gaussian Type Orbitals (GTOs). In spherical polar coordinates STOs have the general form:

$$\chi = Nr^{n-1}e^{-\zeta r}Y_l^m(\theta, \phi) \quad (1.12)$$

while GTOs have the general form:

$$\chi = NY_{l,m}(\theta, \phi)r^{2n-2-l}e^{-\xi r^2} \quad (1.13)$$

In these equations N represents the normalisation constant, r is the distance from the nucleus, and n is the principle quantum number. ζ is an exponent that depends on the atomic number, among other things. Finally, Y_l^m represents the spherical harmonic function from the solution of the Schrödinger equation for the hydrogen atom that depends on the quantum numbers m and l .

STOs have a number of advantageous features over GTOs, including the fact that they closely resemble hydrogenic atomic orbitals and in particular, the exponential dependence on the distance between the nucleus and electron. They thus correctly represent the reduction in amplitude of the electrons wavefunction as the distance from the nucleus increases. The practical use of STOs as basis functions is however, severely limited by the disadvantage that the calculation of three- and four-centre two-electron integrals, inherent to all *ab initio* methods, cannot be performed analytically. Having to solve these integrals by numerical methods ensures that STOs are generally impractical compared to GTOs for all but the smallest of systems, such as atomic and diatomic systems. Accurate and efficient STO basis sets for such calculations have none the less been developed,²⁶ and the well known Amsterdam Density Functional (ADM) suite of programs makes use of STOs as basis sets.²⁷ In semi-empirical methods where such integrals are neglected and replaced by experimentally obtained parameters, STOs commonly find useful implementation for larger, more chemically relevant systems.

In contrast, GTOs have analytical solutions to three- and four-centre two-electron integrals. This ensures that they are computationally much more efficient. However, two features diminish the accuracy of these functions as basis sets. The first is the r^2 dependence in these types of functions, which results in an unrealistically rapid reduction in the wavefunction as one moves away from the nucleus. This is in contrast to real hydrogenic AOs which have a radial decay dependent on r . Consequently, the tail of the wavefunction is represented poorly with GTOs. The second feature is that they fail to describe the cusp at the nucleus present in real hydrogenic AOs, which STOs are able to describe. This is because GTOs are smooth and differentiable at the nucleus, where r is equal to zero, while STOs show a discontinuous derivative in the same region. Consequently, GTOs fail to describe the behaviour of the electronic wavefunction close to the nucleus. Although these are disadvantages, it should be noted that most of the electron density will fall between these two poorly described regions of space.

In order to combine the favourable characteristics of STOs, namely the correct radial decay behaviour and a cusp at the nucleus, with that of GTOs, namely computational efficiency, GTOs are used as building blocks in the construction of STOs. In other words the basis functions used in the SCF process are not individual STOs but a linear combination of GTOs with parameters that are fitted to reproduce an STO as accurately as possible:

$$\varphi = \sum_{a=1}^M C_a \phi \quad (1.14)$$

In Equation (1.14) above φ and ϕ represent the STO approximated by GTOs and the individual GTOs, respectively. M is the number of GTOs used in the linear combination and is referred to as the “degree of contraction”. The coefficients, C_a , are chosen to optimise the shape and guarantee normalisation of the wavefunction. They are generally referred to as contraction coefficients. When a basis function is defined this way, it is referred to as a “contracted” Gaussians function (φ), and the individual GTOs within the summation are referred to as the “primitive” Gaussians (ϕ). A general consensus is that at least 3 GTOs are required to describe a given STO. Taking into account that the scaling behaviour, in terms of

computational resources, of *ab initio* methods is formally (basis functions)⁴ this may seem at first glance counterproductive. However, the increase in the number of basis functions that arises as a consequence of using GTO's is more than compensated for by the reduction in computational expense that occurs as a consequence of evaluating the aforementioned integrals analytically rather than numerically. It is for this reason that GTOs are used almost universally in modern electronic structure packages for the calculation of molecular properties. Additionally, the Gaussian type orbitals are normally centred on the nuclei (i.e. the origin of the coordinate system corresponds to the position of the nuclei). However, this is not always the case; it can be beneficial under certain circumstances to centre the functions in the middle of a bond or between two non-bonded atoms for example, so that dispersion forces can be better accounted for. The contraction scheme in Equation (1. 14) also alleviates a problem inherent to both STOs and GTOs, namely the inability to exhibit radial nodal behaviour. Thus, for example, no choice of parameters exists that will make individual Equations (1. 12) and (1. 13) negative near the nucleus (values close to the origin) and then change sign at particular value of r , such as is necessary to describe a 2s orbital. Using a summation of weighted primitive functions solves this problem, as the sign of the individual contraction coefficients (C_a) in Equation (1. 14) can be chosen to be positive or negative.

The smallest number of basis functions needed to describe the system is referred to as the minimal basis set. This implies that only enough functions are used in order to describe all the electrons of the system; it is the absolute minimum number of functions required and is certainly far from the infinite basis set limit. For example, hydrogen and helium would require a single 1s function. Atoms in the first row would require a 1s, 2s and a set of 2p functions ($2p_x$, $2p_y$ and $2p_z$). Although lithium and beryllium would actually only require the s functions to describe their two electrons, a set of p functions is normally also added to better describe bonding orbitals. These minimal basis sets are also referred to as single zeta basis sets, from the ζ exponent term in the expression for a STO. Doubling the minimum number of basis functions required results in what is referred to as a Double Zeta (DZ) basis set. Tripling them, in a similar fashion, results in a Triple Zeta (TZ) basis set. Doubling the number of basis functions allows for a more accurate description of the reality of different electron distributions in different spatial directions in a molecule. The core orbitals are not, however, significantly perturbed by chemical bonding. Thus, although doubling the 1s functions in a second row element, for example, gives a better description of the 1s electrons, the distribution of these electrons is essentially constant in different chemical environments.

Performing a calculation at a given level of theory, using a given basis set for two different molecules containing a particular atom would result in almost identical core orbitals for that atom in both molecules. For example, the 1s orbital of nitrogen in both nitrogen dioxide (NO₂) and ammonia (NH₃) would be essentially identical. Valence orbitals, in contrast, differ significantly as a function of chemical environment. Thus, in order to correctly predict chemistry, more mathematical flexibility is needed to describe the valence electrons than their core counter-parts. This has led to a variation in the double and triple zeta basis sets in which only the number of functions describing the valence electrons is increased relative to the minimal basis set. These are referred to as split valence (SV) basis sets and are generally employed in modern quantum chemistry packages.

In addition to describing the fundamental electron distribution, polarisation functions are almost always necessary to reach an adequate level of accuracy. These are basis functions corresponding to one quantum number of higher angular momentum (L) than the valence orbitals. Thus, an s function can be polarised by a p function, a p function by a d function and a d function by an f function, etc. Although the atomic electron configuration can be completely described by s and p functions for a second row element for example, the molecular orbitals require more flexibility than a linear combination of the atomic orbitals can provide, and thus a d function is usually also necessary. These *polarisation* functions predominantly help describe charge polarisation effects, for example when an atom is bonded to a much more electropositive element and thus takes on a partial negative charge, and vice versa. For independent particle methods that do not take into account electron correlation the first set of polarisation functions is largely sufficient (i.e. a p function for hydrogen and helium and a d function for second row elements, etc.). This is however not the case for methods that take into account electron correlation.

For many chemically relevant systems, such as anions, loosely bound supramolecular complexes held together by non-covalent bonds and species with highly excited electronic states, an additional type of basis function is required to reproduce reality. As a consequence of weakly bound electrons localising far from the remaining electron density, the highest energy MOs in these types of species are much more spatially diffuse than usual. In to avoid the inaccuracies in energy and other properties so-called *diffuse* basis functions are necessary. These are essentially valence functions, usually of type s and p, with small exponents, i.e. small values of ξ in Equation (1. 13). Generally, the exponents for diffuse functions are 4 times smaller than the lowest valence exponent. Diffuse functions thus reproduce a slow decay of

the electron density with distance from the nucleus. They are particularly important for the calculation of molecular properties such as dipole moments and polarizabilities, as such properties are sensitive to the tail of the wavefunction.

1.2.2.4 Electron Correlation and Relativistic Approaches

On the other side of the spectrum, where the computational costs are often prohibitive and the results converge to that of the exact solution of the electronic Schrödinger equation, lie methods that take into account the electron correlation energy (EC), known as electron correlation methods. As mentioned before, this EC is essentially the instantaneous rather than the average electron-electron repulsion that is calculated by lower level HF theory. The EC energy accounts for only approximately 1% of the total energy. Although seemingly insignificant this contribution is important for describing chemical phenomena as “chemical accuracy” is generally agreed to be $\sim 1 \text{ kcal.mol}^{-1}$, often less than this 1% of the total energy of a molecule.²² More specifically, the difference between the energies of similar molecules is often around 1% of the total energy of the molecules. Thus, in order to gain insight into chemical phenomena and apply the results of calculations to chemical problems, a high degree of accuracy is often advantageous.

Electron correlation methods include Configuration Interaction (CI), Many-Body Perturbation Theory (MBPT) and coupled cluster (CC) methods. The wavefunction here is described by more than one Slater determinant (SD); the Hartree-Fock Slater Determinant as well as “excited” Slater determinants.

$$\Psi = a_0 \Phi_{HF} + \sum_{i=1} a_i \Phi_i \quad (1.15)$$

$\downarrow$
Hartree-Fock SD

$\downarrow$
Excited SDs

These electron correlation methods essentially differ only in the way that they calculate the coefficients of the excited determinants, a_i , which reflect the weight of each determinant in the expansion. The coefficient of the Hartree-Fock determinant (a_0) is determined by the normalization condition. If an infinitely large basis set is used with all possible determinants for this basis set included in the calculation, the method will converge to an “exact” solution

of the Schrödinger equation within the limitations of the Born-Oppenheimer approximation (neglecting relativity).

Of the techniques to calculate electron correlation, coupled cluster methods, particularly CCSD(T), have become the gold standard for single reference calculations. Therefore, CC methods will be briefly expanded upon. CC theory is based upon the idea that the full-configuration interaction wavefunction (i.e. “exact” within the incompleteness of the basis set and neglecting relativity) can be expressed as:

$$\Psi = e^T \Psi_{HF} \quad (1.16)$$

T is referred to as the cluster operator and is defined as:

$$T = T_1 + T_2 + \dots + T_n \quad (1.17)$$

where n is the total number of electrons. The T_i operators generate all possible determinants having i excitations from the reference determinant. For example, for $i = 2$:

$$T_2 = \sum_{i < j}^{occ.} \sum_{a < b}^{vir.} t_{ij}^{ab} \Psi_{ij}^{ab} \quad (1.18)$$

where the notation used in Equation (1.18) is that the occupied molecular orbitals, from where the electrons are removed from are designated by a subscript (ij) while the virtual orbitals that the electrons enter when excited are designated as a superscript (ab).

In coupled cluster theory, it is customary to use the term amplitudes for the expansion coefficients t , which are equivalent to the a_i coefficients in Equation (1.15). The amplitudes are then determined by the constraint that Equation (1.16) be satisfied. It should be noted that the expansion ends at T_n because no more than n electrons can be excited. Making the approximation that $T = T_1 + T_2$, and thus including only single and double excitations, corresponds to the CCSD method i.e. Coupled Cluster Singles and Doubles. Generally speaking, the computational cost of including single excitations in addition to doubles is worth the increase in accuracy, being only slightly more taxing than just including doubles. CCSD scales as B^6 (Where B is the number of basis functions used). The inclusion of triple excitations, corresponding to $T = T_1 + T_2 + T_3$, is immensely demanding and scales as B^8 . This makes CCSDT impractical for all but the smallest of systems. Various methods of estimating the

contributions of the triple excitations from the link between CC theory and many bodied perturbation (MBP) theory have been proposed, to surmount this. The most widely used is referred to as CCSD(T)²⁸ to distinguish it from CCSDT methods. The (T) term indicates that the energy of the triplet states has been obtained from invoking a perturbative treatment. CCSD(T) additionally includes a singles-triples coupling term. Fortuitously, the (T) term actually slightly overestimates the triple excitation correction by an amount closely equal to the ignored quadruples correction. This favourable cancellation of errors makes the CCSD(T) method essentially the most accurate approach that is practically feasible for the computation of chemically relevant systems, as long as they show no strong multireference character.

In order to take into account relativistic effects, one must incur further computational expense and solve for the Dirac equation. This equation is significantly more complex to solve than the Schrödinger equation. It takes into account, among other things, the dependence of the electron's mass on its velocity as well as the interaction of the electron's spin with the magnetic field generated by having this velocity (the *spin-orbit* term). However, the incorporation of such relativistic effects is only necessary for electrons that will achieve velocities comparable to *c*. This is only applicable to core electrons that reside in the proximity of heavy nuclei. For example, relativistic effects for a hydrogen 1s electron account for only 0.4% of the total wavefunction and 10⁻³% of the electron density. As one moves further down the periodic table however, this discrepancy becomes larger. For example, for a uranium 1s electron such effects account for a third of the wavefunction and 11% of the density²². Additionally, it is necessary to use a large number of basis functions to describe the core electrons as the electron-electron repulsion they generate has a pronounced effect on the energies of the valence electrons. These problems, namely the inclusion of relativistic effects and the large number of basis functions needed, can be solved simultaneously by considering the invariant nature of the core electrons. By using a suitable function to model their effects they are not treated explicitly and this results in a significant reduction in computational expense with only a small reduction in accuracy. Although such functions are complex they need only be created once and are valid for a diverse array of systems. Such functions that model the core electrons are known in the chemistry community as Effective Core Potentials (ECP) while physicists generally refer to them as Pseudopotentials (PP).

1.2.2.5 Density Functional Theory

For the majority of the work in this project, the method that was employed is Density Functional Theory (DFT). This is based on the demonstration by Hohenberg and Kohn²⁹ that the ground state electronic energy is determined completely by the electron density. Therefore, there exists a one-to-one relationship between the electron density of a system and the energy. More specifically, there exists a unique functional of the ground state density, $E[\rho(\mathbf{r})]$, that is equal to the exact electronic energy. This can be expressed as:

$$E[\rho(\mathbf{r})] = E_{elec} \quad (1. 19)$$

In the above equation $\rho(\mathbf{r})$ is the ground state electron density.

The significance of the Hohenberg-Kohn theorem is the reduction in the variables needed to describe the system. For a system of N electrons the wavefunction depends on the three spatial coordinates (in Cartesian coordinates x, y and z) as well as the spin of each electron. Thus $4N$ variables are needed. The electron density is the square of the wavefunction, integrated over $N - 1$ electron coordinates. If one considers the spin density for spin $+1/2$ and spin $-1/2$ separately then each spin density only depends on three spatial coordinates and is independent of the number of electrons. Thus while the complexity of a wavefunction increases exponentially with the number of electrons, the electron density has the same number of variables regardless of the size of the system. However, although the Hohenberg-Kohn theorem proves the existence of a functional that relates the electron density to the energy, it offers no guidance as to the form of this functional. The success of modern DFT methods is based on the introduction of auxiliary sets of orbitals used to represent the electron density which leads to a functional with the following form:

$$E[\rho(\mathbf{r})] = T_e[\rho(\mathbf{r})] + V_{ne}[\rho(\mathbf{r})] + V_{ee}[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] \quad (1. 20)$$

$T_e[\rho(\mathbf{r})]$ is the kinetic energy of a fictitious non-interacting system and V_{ne} is the nuclear - electron attraction functional with the form:

$$V_{ne}[\rho(\mathbf{r})] = \sum_j^{nuclei} \int \frac{Z_j}{|\mathbf{r} - \mathbf{r}_k|} \rho(\mathbf{r}) d\mathbf{r} \quad (1. 21)$$

V_{ee} is the classical electron-electron repulsion term:

$$V_{ee}[\rho(\mathbf{r})] = \frac{1}{2} \iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (1. 22)$$

$E_{xc}[\rho]$ is the exchange correlation functional which can be written as a sum of the separate exchange, $E_x[\rho]$, and correlation, $E_c[\rho]$, functionals.

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho] \quad (1. 23)$$

This is referred to as Kohn and Sham (KS) DFT, named after its developers.³⁰ In this theory it is the fictitious non-interacting system, which has the same density and thus energy as the true system. The exchange correlation functional is the only unknown functional in Equation (1. 20). This term acts as a kind of catch-all term to account for aspects of the true system that the theory neglects. Fortunately, it represents a rather small fraction of the total energy and relatively crude approximations for the exchange correlation functional provide quite accurate results. There is no way of theoretically deriving this term and different DFT functionals essentially only differ in the way they estimate the exchange-correlation. The choice of the functional form of $E_{xc}[\rho(\mathbf{r})]$ and the approximations made in estimating this term have developed over time. This has given rise to the concept of Jacob's ladder of DFT functionals³¹, where each subsequent generation of functionals represents a rung on the ladder to the ideal case of an analytical, universally accurate $E_{xc}[\rho(\mathbf{r})]$ term colloquially referred to as "DFT heaven".

The Local Density Approximation (LDA) assumes that $E_{xc}[\rho(\mathbf{r})]$ can be determined just from the density; this is the most basic approach. The Local Spin Density Approximation (LSDA) is an extension where the spin +1/2 and spin -1/2 densities are assumed not to be equal. It should be noted that for closed shell systems LDA is equivalent to LSDA. Together LDA and LSDA represent the first rung on this ladder. Supposing that the density is not uniform and thus that $E_{xc}[\rho(\mathbf{r})]$ is also dependent on the gradient of the density, gives rise to the generalised gradient approximation (GGA). Most GGA functionals are constructed with a correction term, being a term added to the LDA functional. This is the second rung of Jacobs ladder. Meta GGA functionals include an orbital kinetic energy density functional or a second derivative of the density. Since the orbital kinetic energy density and the Laplacian of the density essentially carry the same information, these methods are equivalent and both grouped under the meta

GGA classification on the third rung. Hybrid functionals incorporate some mixture of HF exchange, again increasing accuracy. These hybrids are also referred to as the Adiabatic Connection Model (ACM) as the adiabatic link between the non-interacting system and the true system is exploited in their construction. As exchange in HF theory is exact, its inclusion is often found to improve the calculated results, although the optimum fraction to include depends on the specific property of interest. Hybrids represent the fourth rung. On the fifth and currently final rung, of the still-incomplete Jacobs ladder of DFT functionals lie the double hybrid functionals. The composition of a double hybrid functional means that it is essentially a hybrid-DFT calculation plus an MP2 calculation, using the DFT orbitals. Not only the occupied but also the virtual orbitals are included in this calculation, which significantly improves dispersion interactions. The perturbational correlation from the MP2 calculation can correct for traditional correlation deficiencies in DFT while simultaneously improving the self-interaction error. However, double hybrids have basis set requirements similar to correlated wavefunction methods and thus cannot benefit from the faster basis set convergence of traditional DFT. The accuracy of double hybrid methods is slightly better than MP2 quality although, in favourable cases, the performance may approach that of coupled cluster calculations.

The strength of DFT is that it can produce results that are reasonably accurate with the same computational expense as HF. It is this efficiency, exemplified by small calculations being able to run on desktop PC's, that have made this method so widely popular. However, DFT is not without its innate disadvantages. One of the problems with DFT is that if a chosen functional produces poor results there is no way to systematically correct its performance other than turning to a different functional. Weak interactions due to dispersion forces (van der Waals type interactions) arise from electron correlation, but this is poorly described by current DFT methods. Thus, such corrections must be added on in an *ad hoc* fashion. Although many programs offer modern empirical dispersion correction methods (for example those developed by Grimme et al³²) this results in decreased accuracy when describing such interactions when compared to wavefunction methods. This inherent lack of dispersion in the formalism of DFT is exemplified by the inability to describe the slight attraction between rare gas atoms. Most functionals calculate a purely repulsive energy, while those that do predict an attraction underestimate the effect and the variation between systems.³³ DFT fails systematically when attempting to describe charge transfer systems. This arises as a consequence of the exchange–correlation functional being inherently local, depending only

on the density (LSDA) and possibly its derivatives (GGA) and second derivative (meta-GGA) at a given point. These models thus fail to describe processes where electrons are transferred over large distances, typified by charge transfer systems. A source of error in DFT that is not present in wavefunction methods is the self-interaction error (SIE). This error arises when a single electron interacts with itself, erroneously giving rise to a non-zero electron–electron Coulomb repulsion, even for a one-electron system. This error is a consequence of the fact that the exchange and correlation functionals are approximations and are not known exactly. Theoretically, if the exchange-correlation term was known and able to be derived analytically then the self-interaction in the individual correlation and exchange functionals would cancel exactly. Unfortunately, this is not the case and the spurious interaction of an electron with itself contributes to the total energy. Many attempts have been proposed to correct for this unrealistic behaviour, with particular success involving transition state structures.³⁴ Unfortunately however, these corrections add an additional level of self-consistency into the SCF process as they depend on the KS orbitals. This results in accompanying convergence issues and large increases in calculation times. As HF theory is completely self-interaction-free, hybrid functionals show much less susceptibility to SIE as a consequence of the added HF exchange.

1.2.2.6 Solvation Models

Thus far the focus has been on determining the properties of single molecules, in the gas phase. However, a fundamental part of chemistry consists of solvent effects. Different solvents can dramatically alter a reaction's profile and the extent to which it occurs. In fact, it is often only possible to induce certain chemical transformations by carefully screening for compatible solvents experimentally. These effects are however usually not a consequence of a single solvent molecule interacting with the reactants, but many hundreds (and possibly thousands in biological systems) of solvent molecules that together with the reactants constitute the reacting system. Due to these persistent interactions, the PES of a solvated system can be significantly different from its gas phase counterpart. Various physical effects contribute to the overall solvation process such as electrostatic interactions, the formation of a cavity in the solvent, changes in dispersion as well as changes in bulk solvent structure. Electrostatic interactions are however considered to be the dominant contribution to the free energy of solvation (ΔG_{sol}). Even if one only explicitly includes the first solvation shell, the size of the system rises rapidly and a geometry optimisation quickly becomes intractable. Thus, the most practical way to model solvent effects are implicit solvation models, also known as

continuum approaches. Here individual solvent molecules are not considered but are replaced by a dielectric continuum that represents a statistical average of their electrostatic interactions with the solute at thermal equilibrium. These models only consider the electrostatic interactions that contribute to ΔG_{sol} . The particular implicit solvation model used in this project was the Conductor-like Screening Model (COSMO).³⁵ Here the solute is embedded in a cavity within the dielectric continuum. Initially it is assumed that the permittivity, ε , of the dielectric continuum is infinite i.e. that the solvent is a perfect conductor. Under this assumption no electrostatic potential exists in the bulk continuum, and instead charge develops on the continuum surface in contact with the solute. This assumption leads to a significant simplification of the relevant Poisson classical electrostatic equations that need to be solved. The interface between the cavity and the dielectric continuum is commonly referred to as the solvent accessible surface (SAS). The response of this homogeneous continuum to any charge distribution of the solute (the molecular charge density) consists of a charge distribution at the SAS (referred to as the screening charge). Both the molecular charge density and the screening charge need to be minimised in an iterative process until self-consistency is reached. This results in a double iterative procedure within a single SCF-cycle. Once self-consistency has been reached the electrostatic interaction between the solvent and the solute is scaled to a finite permittivity by applying the factor:

$$\frac{\varepsilon - 1}{\varepsilon + 0.5} \quad (1.24)$$

where ε is the dielectric constant of the solvent.

The use of COSMO typically adds an additional 30% to the timing of each SCF cycle. Fortunately, analytical gradients are available making geometry optimisations in conjunction with COSMO efficient.

Having briefly outlined the fundamental principles underlying the computational methods used in this work, attention is now focussed on the chemistry of the problem addressed.

1.3 The Direct Arylation of Oxazole

1.3.1 Industrial Applications

The small heterocyclic organic molecule oxazole is shown in Figure 1-4. This project centres on the arylation of oxazole. Each of the three unique hydrogen atoms in oxazole can be replaced

by a biaryl bond during a direct arylation reaction. Industrially, such reactions could be used in the synthesis of modern medicines and other commercially valuable substances. In all of the figures shown, oxazole itself, as well as any biaryl substituents are highlighted in red.

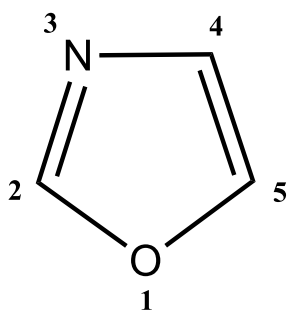


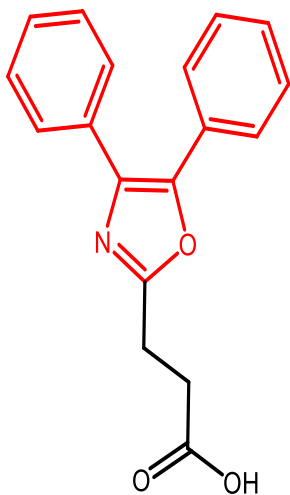
Figure 1-4: Oxazole

Oxaprozin (Figure 1-5) is an example of a possible oxazole arylation product found in a commercially available drug. Sold under the trade names Daypro, Dayrun and Duraprox it is a non-steroidal anti-inflammatory drug (NSAID) that is used to treat symptoms of arthritis such as inflammation, stiffness and pain. Additionally, oxazole with a biaryl bond can be found in a number of peroxisome proliferator-activated receptor (PPAR) agonists that are currently being developed to treat metabolic disorders. These include preclinical drugs that have not yet reached the market such as Farglitazar, which is currently in phase II clinical trials, being developed by GlaxoSmithKline for the treatment of hepatic fibrosis. Additionally, a large number of PPAR agonists that were sought to relieve insulin resistance in type 2 diabetes mellitus, but whose development has been discontinued contain the products of possible oxazole arylation reactions. These include Darglitazone³⁶ developed by Pfizer, Muraglitazar³⁷, discontinued at phase III clinical trials by Bristol-Myers Squibb, Imiglitazar³⁸ placed on hold by Takeda Pharmaceuticals at the Phase III clinical trials stage as well as Reglitazar³⁸, discontinued at phase III trials by Japan Tobaccos Pharmaceutical Division. The structures of all these molecules are shown in Figure 1-5. The fact that the development of this afore-mentioned group of molecules has been discontinued implies that further research into fine tuning the structure-activity relationships is necessary. A better understanding of the underlying factors governing the direct arylation of oxazole would assist such processes into the future.

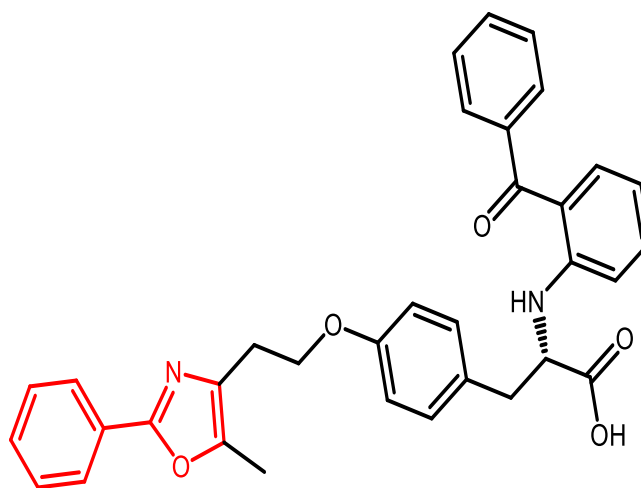
The oxazole moiety is also found in the sulphonamide antibiotic, sulfamoxole³⁹ as well in a number of peptide-based natural products that have been shown to exhibit significant antifungal, antibiotic, antitumor and antiviral biological activities. These include the HIV-1 protease inhibitor Thiagazole⁴⁰ as well as the macrolide antibiotic Pristinamycin IIA⁴¹.

Pristinamycin IIA forms one half of the synergetic combination used in the commercial antibiotic Pyostacine, which is highly effective against Methicillin-Resistant *Staphylococcus Aureus* (MRSA). Additional analogues containing the oxazole moiety display commercial promise as lead compounds in the development of pancreatic cancer drugs.⁴²

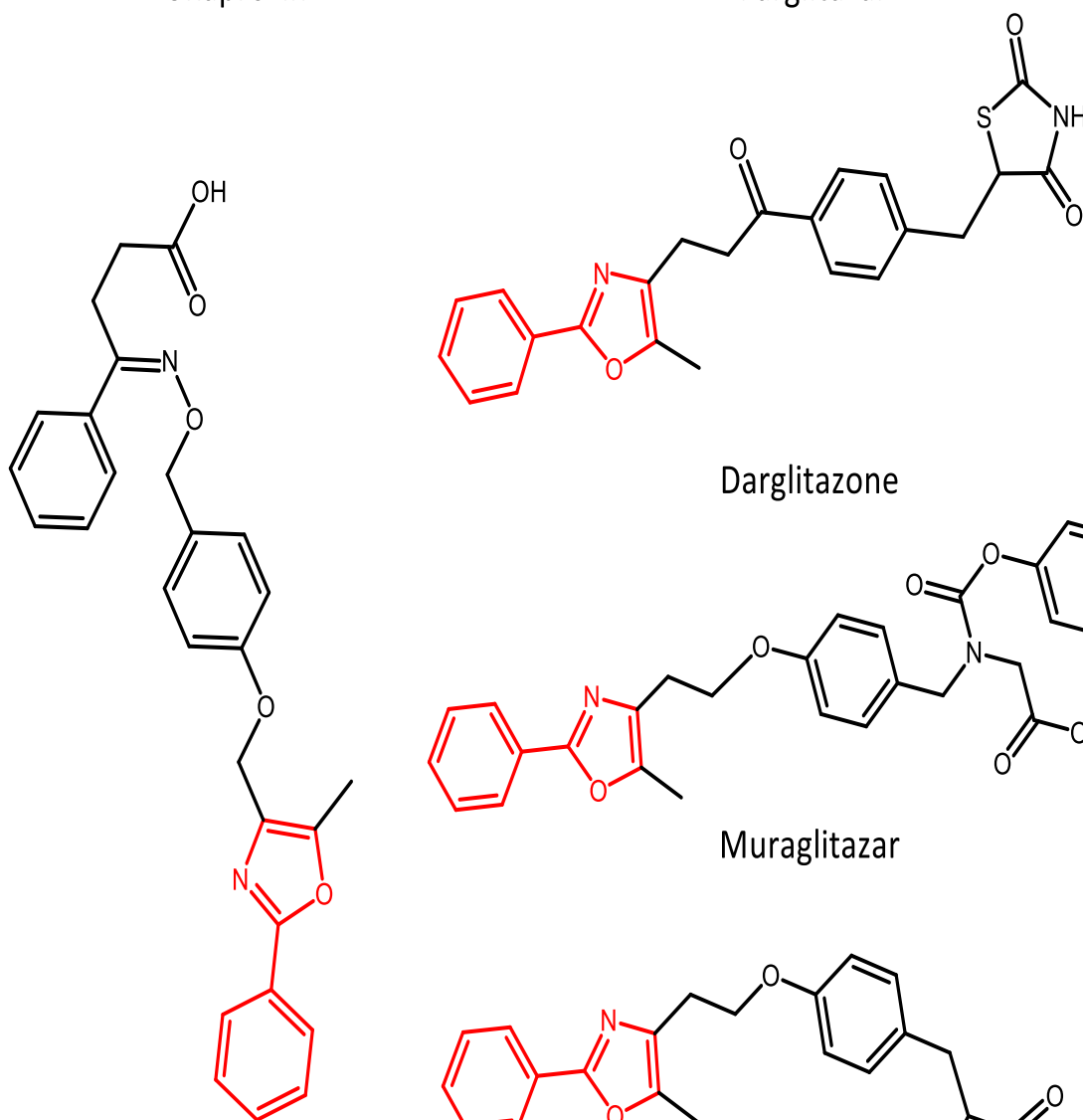
An example outside of pharmaceutical industry is Naphtho[1,2-d][1,3]oxazole, which is used as a decolorizable dye in the photography and imaging industries, currently patented by the Fuji Photo Film Company.⁴³ The structures of the aforementioned molecules are shown in Figure 1-6. Lastly, polymers consisting of oxazole biaryl analogues have been proposed as functional materials for the detection of fast neutrons in the presence of background gamma radiation.⁴⁴



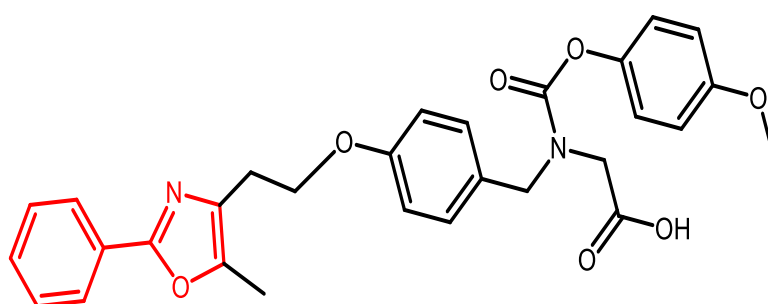
Oxaprozin



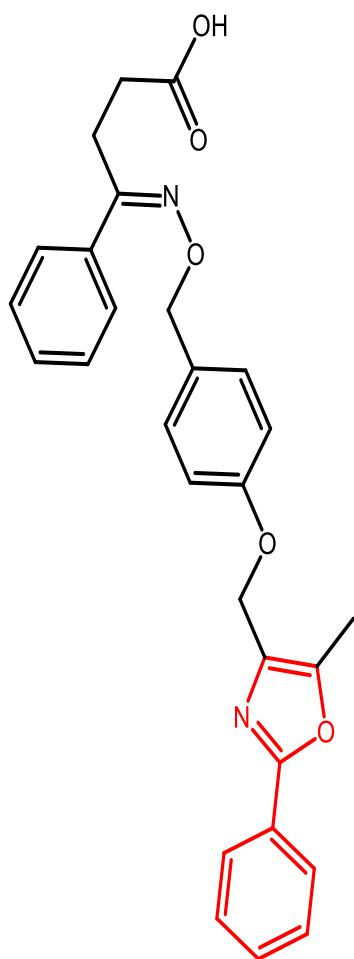
Farglitazar



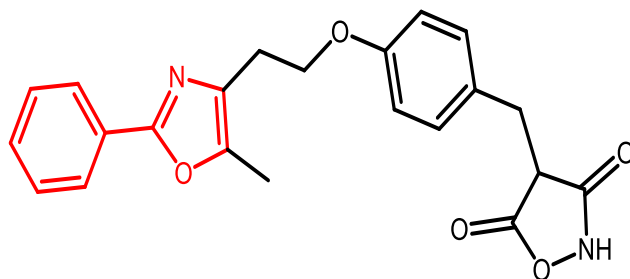
Darglitazone



Muraglitazar

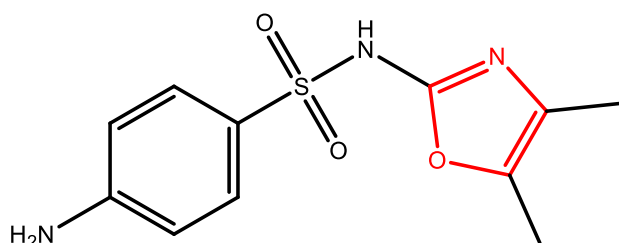


Imiglitazar

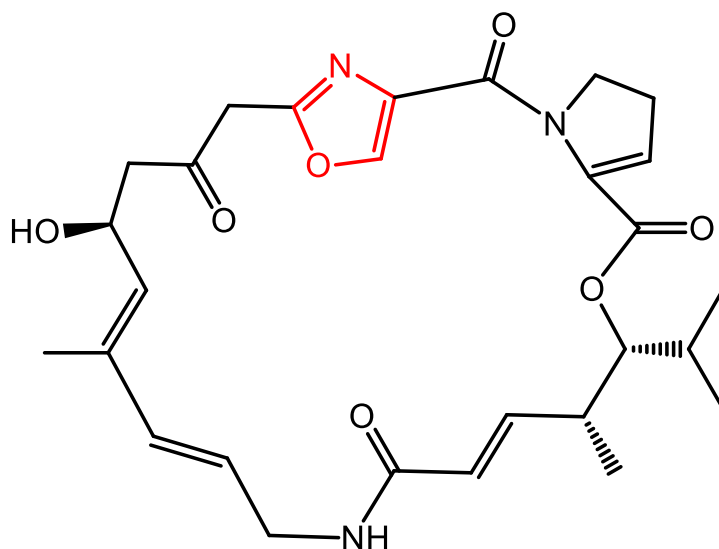


Reglitazar

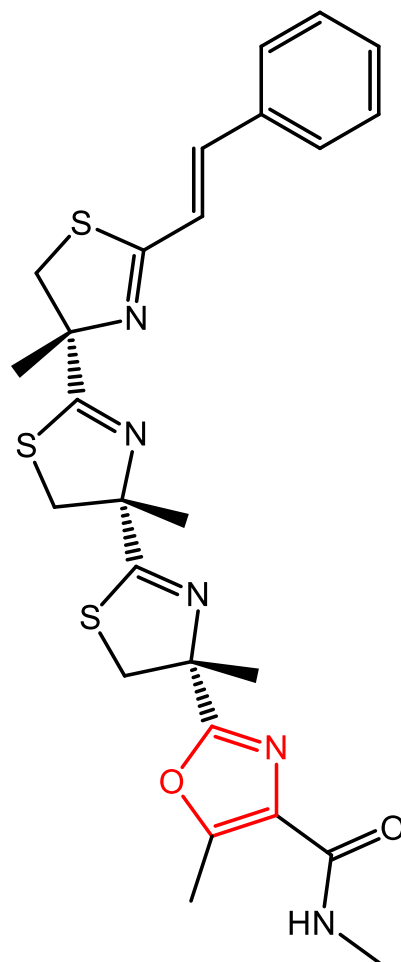
Figure 1-5: Oxazole Arylation Products



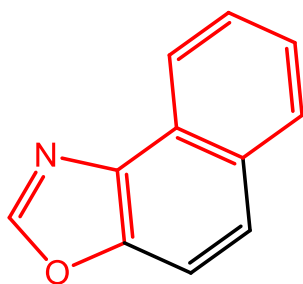
Sulfamoxole



Pristinamycin IIA



Thiagazole



Naphtho[1,2-d][1,3]oxazole

Figure 1-6: Industrially Useful Molecules Containing the Oxazole Moiety

1.3.2 Experimental Results Courtesy of Johnson Matthey

The arylation of oxazole is thus an interesting topic for industrialists. Experiments have been performed in the research laboratories of Johnson Matthey. These experiments involved the direct arylation of oxazole with 4-Bromotoluene (1:1 mole ratio). The catalytic system

consisted of palladium acetate (PdOAc_2), while the solvent system used was dimethylacetamide, DMA (Figure 1-7), containing pivalic acid (Figure 1-8) and KOH base. These reactions were conducted at a temperature of 100°C . The crucial variable in these experiments was the presence, identity and concentration of various ligands. This project will focus on a small subset of these experiments involving the bulky phosphine ligand tri-tertiarybutylphosphine, P^tBu_3 , (Figure 1-9).

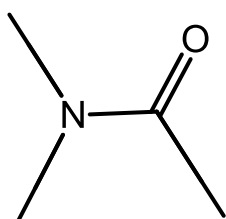


Figure 1-7: Dimethylacetamide (DMA)

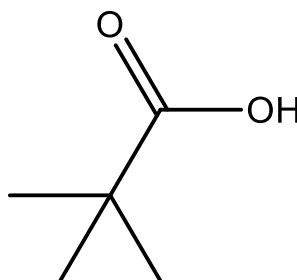


Figure 1-8: Pivalic Acid

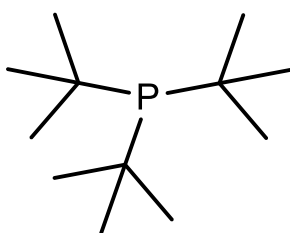


Figure 1-9: Tri-tertiarybutylphosphine

It should be noted that the actual form of the phosphine ligand used was the phosphonium salt $[\text{P}^t\text{Bu}_3\text{H}]^+[\text{BF}_4]^-$, namely tri-tert-butylphosphonium tetrafluoroborate (Figure 1-10). Apart from being malodorous and pyrophoric tri-tertiary phosphines are also notoriously air-sensitive. This is due to oxidation by atmospheric oxygen which leads to the formation of by products such as tri-tertiary-butylphosphine oxide in the case of P^tBu_3 and well as the products of free radical chain oxidation.⁴⁵ These by-products are unsuitable as ligands in transition metal catalysis. This sensitivity requires careful handling of the ligands prior to reaction. However, the oxidation-stable, easily handled phosphonium salt has been shown to be stable in air for longer than several months. Most importantly, it has been confirmed that simply substituting the phosphonium salt for the original ligand in diverse array of palladium-catalyzed reactions results in equivalent (within 5%) catalytic activity.⁴⁶ This is expected, as in direct arylation at-least, nearly all experimental protocols make use of a Brønsted base

additive and thus the active ligand will simply be generated by *in-situ* deprotonation of the phosphonium species. The corresponding fluoroborate counter-ion is non-coordinating and serves as a spectator species. The phosphonium salt will thus not be considered during modelling in subsequent chapters of this work.

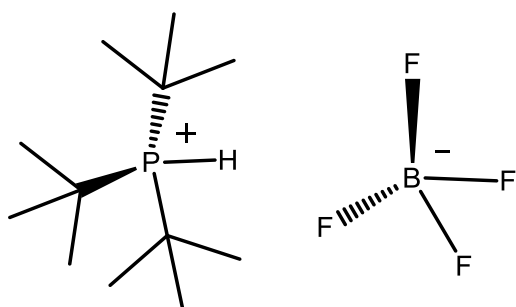


Figure 1-10: Tri-tert-butylphosphonium tetrafluoroborate

Before discussing the details and results of these experiments, it is necessary to introduce the possible products of direct arylation of oxazole with toluene. Initial arylation of oxazole can lead to three possible mono-substituted products, each corresponding to arylation at one of the 3 unique hydrogen atoms in the molecule (Figure 1-11).

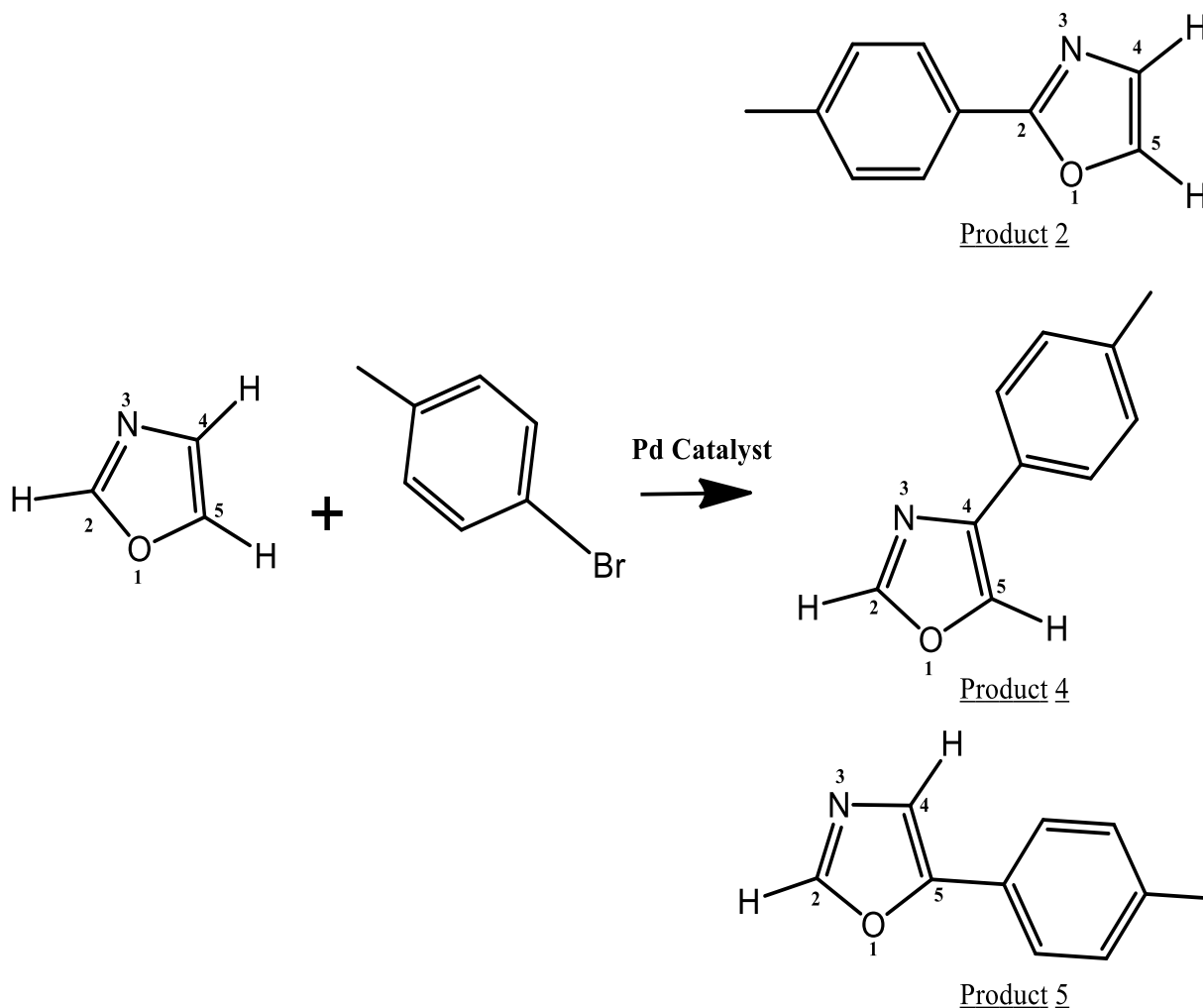


Figure 1-11: Direct arylation of oxazole with bromotoluene highlighting the three mono-substituted products than can arise. The name of the product corresponds to the position of the oxazole substitution.

However, each of these mono-substituted products can enter the catalytic cycle, to be discussed later, for a second time. This leads to the formation of a series of disubstituted products (Figure 1-12).

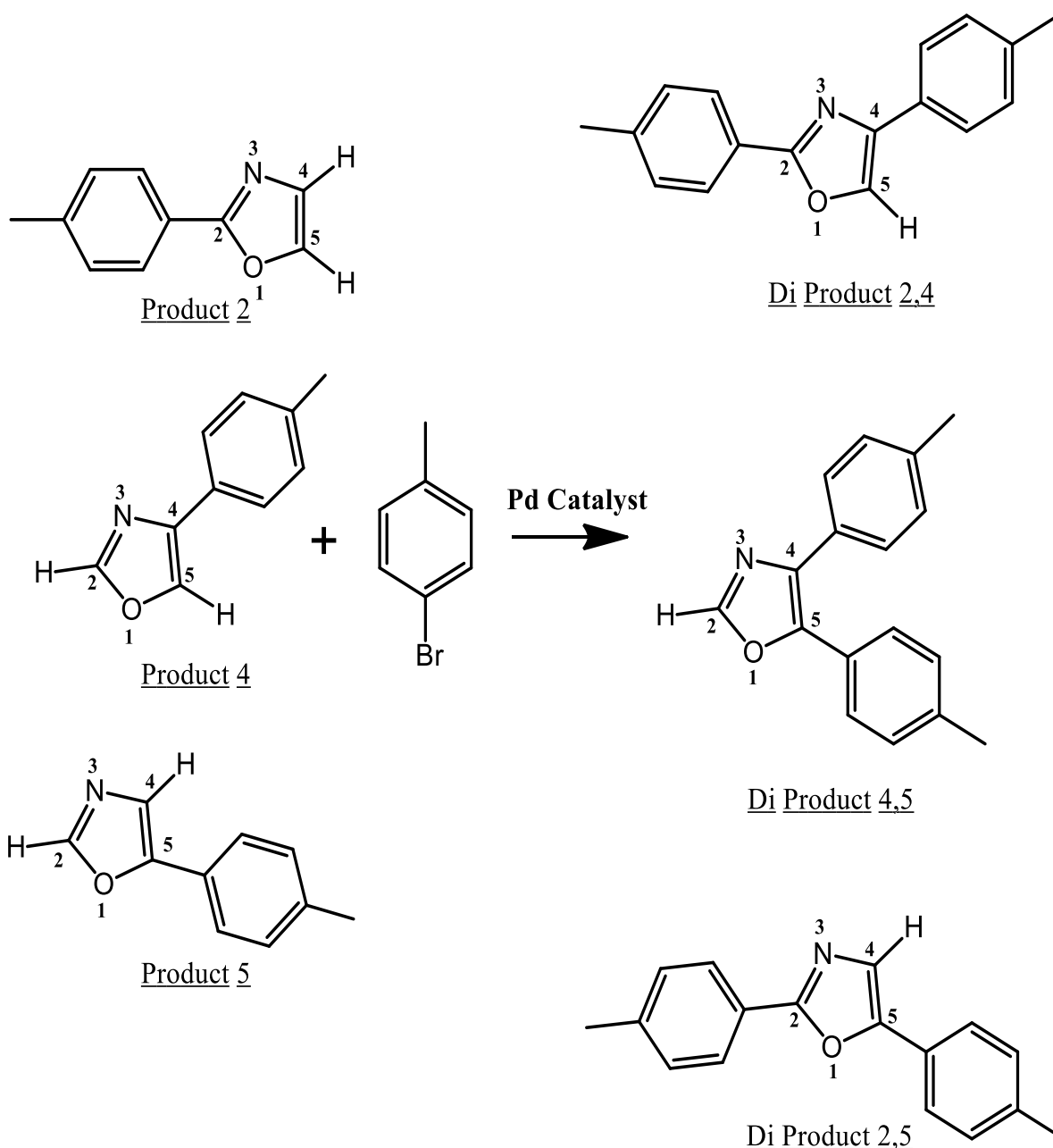


Figure 1-12: Direct arylation of oxazole with bromotoluene highlighting the three di-substituted products than can arise. The name of the product corresponds to the positions of the oxazole substitution.

Lastly, each of the disubstituted products can theoretically enter the catalytic cycle for a third time leading to the formation of one possible tri-substitution product (Figure 1-13).

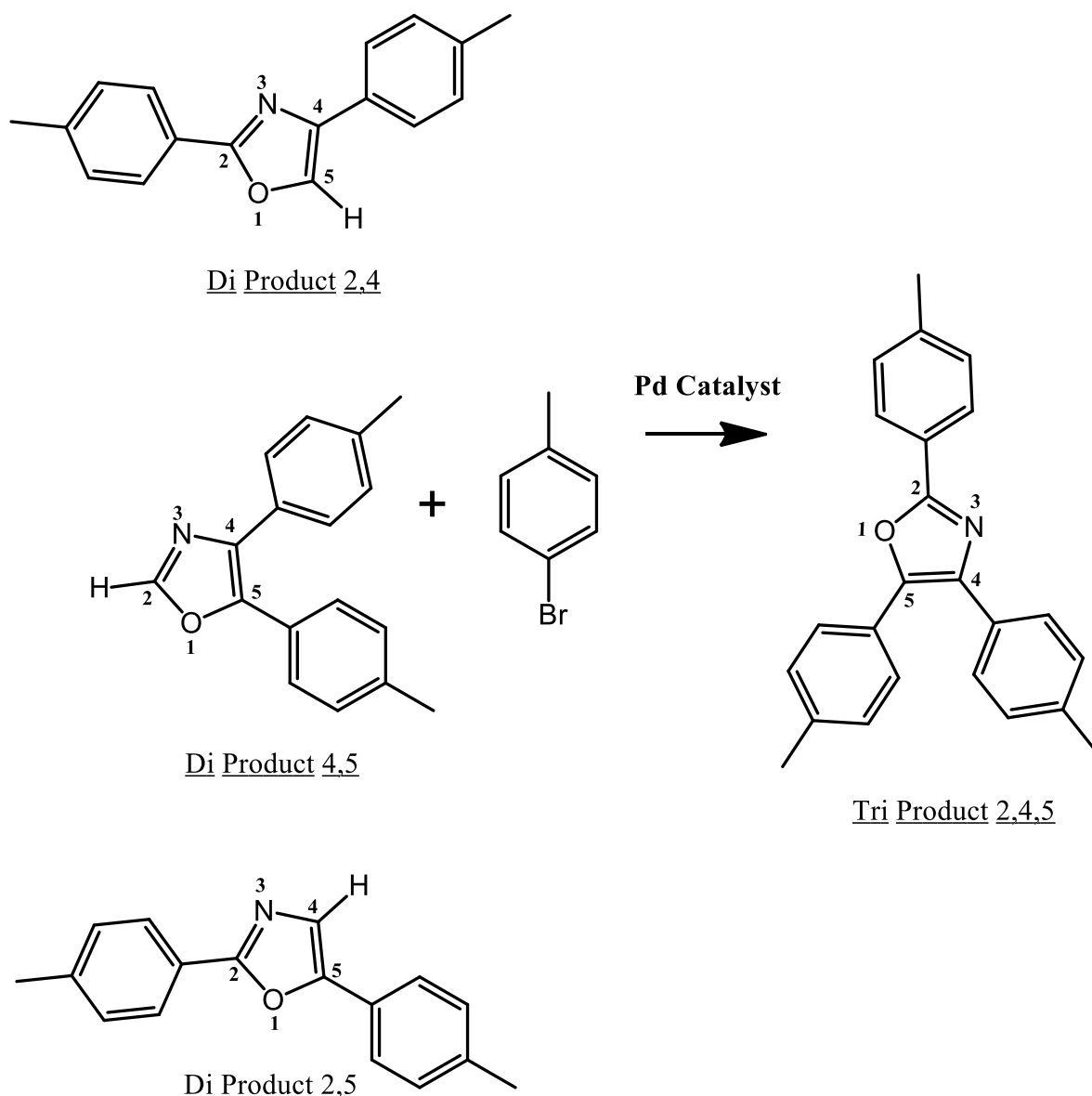


Figure 1-13: Third arylation reaction – Tri substitution

Now that various possible products of the arylation of oxazole with bromotoluene have been introduced, the results of the experiments that are of concern to this project can be presented. Table 1-2 below, shows the product distributions for three experiments in which the mole ratio of P^tBu_3 ligand was varied ranging from 0, 1 and finally to 2 equivalents with respect to the catalyst.

Table 1-2: Experimental Product Distributions

Entry	Catalyst	Ligand	Product Distributions (%)			
			Product 5 ^a	Product 2 ^b	Product 2,5 ^c	Product 4,5 ^d
1	$\text{Pd}(\text{OAc})_2$	-	65.8	0.3	28.1	3.0
2	$\text{Pd}(\text{OAc})_2$	1 x P^tBu_3	30.6	3.7	59.3	0.9
3	$\text{Pd}(\text{OAc})_2$	2 x P^tBu_3	6.3	35.3	49.7	0.1

^a refers to the percentage of mono-substituted product at the 5-position of oxazole that was formed.

^b refers to the percentage of mono-substituted product at the 2-position of oxazole that was formed.

^c refers to the percentage of di-substituted product at the 2 and 5 position of oxazole that was formed.

^d refers to the percentage of di-substituted product at the 4 and 5 position of oxazole that was formed.

It should be noted that product distributions do not add up to 100% in Table 1-2 as small quantities of unreacted oxazole were still present. The structure and relative concentrations of the products were confirmed by GC-MS and a series of NMR experiments. It should be noted that the percentages in Table 1-2 take no account of the relative response at the MS detector and so are only semi-quantitative. However, for the intents and purposes of this study, it will be assumed that the relative response at the MS detector is constant, in the absence of contrary evidence. We will consider the results as quantitative and use them to guide our modelling of the system, simply because no other pertinent data are available.

The presence as well as the concentration of the phosphine ligand has a profound effect on the product distributions that are observed, as manifested in Table 1-2. Specifically, when no phosphine ligand is present Product 5 is favoured, making up approximately two thirds of the distribution with no Product 2 being detected, while the other third is made up by Di Product 2,5 and a small amount of Di Product 4,5. This is completely reversed when one equivalent of phosphine ligand is present; now Product 5 makes up approximately one third of the total distribution with the remaining two thirds consisting of Di Product 2,5 as well as small amounts of Product 2 and Product 4,5. The situation with the mono- substituted products again changes drastically when two equivalents of phosphine ligand are present. Now, only a small amount of Product 5 is formed, while approximately a third of the total distribution consists of Product 2. However, the di-substituted distribution remains similar to one phosphine equivalent.

Thus, the amount of PtBu_3 appears to affect the regioselectivity of these reactions. The conventional mechanisms used to understand direct arylation, to be discussed below, incorporate only one phosphine ligand into the active catalytic species. Thus, it is not obvious why 2 equivalents of P^tBu_3 should favour arylation at the 2 position over the 5 position. The other changes in product distribution also cannot be readily explained. The question central to this project is, why do these observed changes in the regioselectivity of the reaction occur? By answering this question, it is hoped that insight into the underlying chemistry will be revealed that can subsequently be used to tailor reaction conditions so that the most valuable product distributions can be achieved experimentally. Ultimately, the goal will be to offer

insight to experimentalists as they endeavour to develop new synthetic methodologies that could possibly be applied on an industrial scale in the near future.

In order to begin an investigation into this anomalous regioselectivity, two mechanisms that appear in the literature for the regioselectivity-determining C-H bond cleavage step will be considered. The first of these is the well-known Concerted Metallation Deprotonation (CMD) mechanism, while the second, is a much more recently proposed and less well studied mechanism which we will refer to as Cooperative Catalysis (CC). The reasons for these choices will be discussed below in sections dedicated to the elaboration of each mechanism.

1.3.2 The Concerted Metallation Deprotonation Mechanism

Palladium catalysed direct arylation has been proposed in the literature to take place through a variety of mechanisms. These include a Heck-type arylation followed by hydride elimination⁴⁷, anionic cross coupling⁴⁸, electrophilic aromatic substitution (S_EAr)⁴⁹, oxidative C-H insertion⁵⁰ as well as concerted metallation deprotonation (CMD).⁵¹ However, under the standard conditions of a catalytic system comprising of metal, phosphine ligand as well as carbonate or carboxylate base a large array of mechanistic investigations support the CMD mechanism. These studies involve numerous substrates including pyridine⁵², fluorobenzenes⁵³, pyridine N-oxide⁵⁴, substituted benzenes (PhCF₃, PhAc, PhCO₂Me)⁵⁵, tethered arenes⁵⁶, diarenes⁵⁷, as well as benzene itself.⁵⁸ Perhaps the most compelling evidence from a single source is the work by Gorelsky et al.⁵⁹ Here the authors show, through a computational study at the B3LYP/TZVP level of theory, that the CMD mechanism correctly predicts the reactivity and regioselectivity for the direct arylation of a wide range of arenes with benzene. The molecules studied span the entire spectrum of known direct arylation coupling partners. These differ drastically in the electron density around the arene, ranging from electron deficient in the case of penta-fluorobenzene to electron rich in imidazo[1,2-a]pyrimidine (Figure 1-14 (B) and (A) respectively).

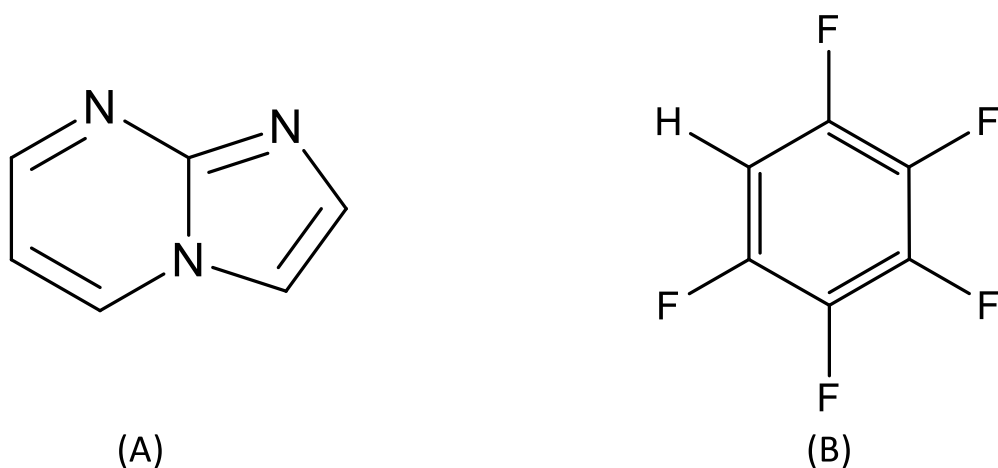


Figure 1-14: (A): Imidazo[1,2-a]pyrimidine (B): Penta-fluorobenzene

In the CMD mechanism (Figure 1-15) a basic ligand attached to the metal centre deprotonates the arene, while the electrons, previously in the C–H bond, simultaneously form a new organometallic bond with the metal atom, giving rise to the transition state. This then allows for the formation of a biaryl bond by subsequent reductive elimination. For the purposes of the current investigation, the basic ligand is an acetate anion; while L in Figure 1-15 is P^tBu_3 or DMA, and $n = 1$. When the basic ligand is a carbonate or carboxylate species (X in Figure 1-15) then associated with this mechanism is a characteristic 7 membered cyclic transition state. Although the CMD mechanism is seemingly ubiquitous in direct arylation reactions, it should be noted that in special cases alternative reaction pathways are favourable. When the properties of the catalytic system are adjusted to lower the activation barriers for the alternative reaction pathways mentioned previously (such as hydride elimination, anionic cross coupling etc.) , they can indeed become favourable. An example is the case of a catalytic system incorporating the highly fluorinated phosphorus alkoxide ligand, $\text{P}(\text{OCH}(\text{CF}_3)_2)_3$, which leads to product distributions that imply a Heck type arylation.⁶⁰ Concerted metallation deprotonation will be the first mechanism considered during the modelling of the reactions in this project and is the subject of Chapter 3.

1.3.3 The Cooperative Catalysis Mechanism

All of the mechanisms mentioned up to this point involve only one phosphine ligand in the active catalytic species. Thus, an increase in phosphine ligand concentration is unlikely to result in a change from CMD to Heck like arylation for example, or to any of the other possible pathways for that matter. Relatively recently, a “cooperative” mechanism involving a cyclometalated complex has been proposed, based on data from the direct arylation of pyridine N-oxide.⁶¹ In this mechanism, henceforth referred to as Cooperative Catalysis (CC), the active catalyst in the CMD pathway (**1** in Figure 1-16) forms a dimer. The higher the concentration of phosphine, the more **1** will be formed and the more likely that two of these complexes will collide, react and subsequently form a dimer. Thus, this mechanism is expected to be favoured under conditions of excess tri-tertiary-butylphosphine ligand and offers potential help to rationalise the changes observed in regioselectivity when the phosphine ligand concentration is doubled.

The origin of the CC mechanism has its roots in an observation by Hartwig et al.⁶¹ that the generally accepted active catalyst in the CMD mechanism is significantly less reactive than expected when isolated from the catalytic cycle. When the equivalent of **1** (containing 3-fluoromethyl benzene as Ar₁) was synthesised separately and combined with pyridine N-oxide in isolation from the catalytic cycle, a yield of 52% was observed, in contrast to the 77% yield observed when starting from the palladium pre-catalyst, phosphine ligand and aryl halide i.e. the catalytic reaction.

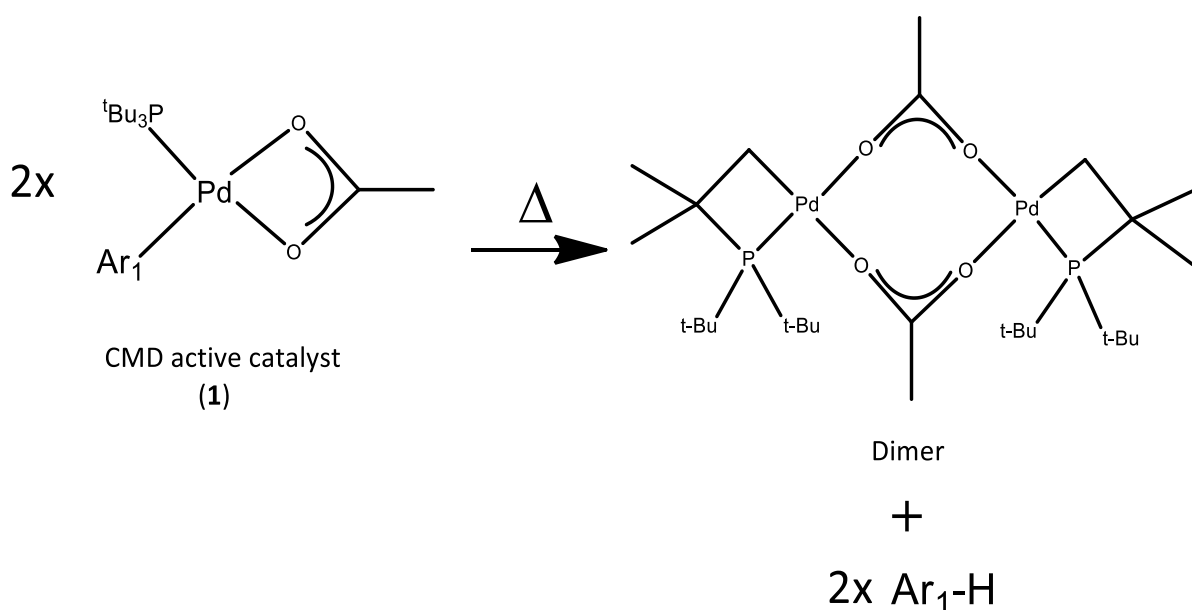


Figure 1-16: Dimer formation in the CC mechanism

A similar decrease in yield from 62% to 48% was observed when the equivalent of **1** (containing *meta*-xylene as Ar₁) was reacted with 4-nitro pyridine N-oxide in isolation from the catalytic cycle in a separate investigation.⁵⁴ An additional example of this phenomenon is the equivalent of **1** (containing 3-fluoromethyl benzene as Ar₁) reacting with benzothiophene and a corresponding decrease in yield from 85% to 65%.⁶¹ Thus, this phenomena is not restricted to pyridine N-oxides.

These discrepancies led Hartwig et al.⁶¹ to investigate the rate of the isolated reaction. They found that it proceeded through an induction period in which almost no bi-aryl product could be detected followed by a linear decay of **1**. During this induction period the formation of the dimer species was detected. Furthermore, when the same reaction was performed with one equivalent of dimer no induction period was observed and the yield of product was 84%, higher than that of the catalytic process. The rate of the reaction with added dimer was zero order with respect to **1**. These data imply that the rate determining C-H cleavage step does not involve **1** but rather the dimer. Furthermore, the concentration of the dimer was constant throughout the reaction implying that it is somehow regenerated and acting as a catalyst. Interestingly, the rate of the reaction with respect to the dimer was half order. This half order dependency has been extensively associated with systems in which the catalyst resting state is a dimer, while the active form of the catalyst is a monomer.⁶²⁻⁶⁶ These data imply that an equilibrium is set up between the observed dimer and a catalytically active monomer before the essential C-H bond cleavage step. These results lead to the conclusion that it is the monomer (Figure 1-17) that is involved in the regio-selective determining step, in which a C-H bond of pyridine N-oxide is cleaved. The next question is how this C-H bond is cleaved by the monomer? The first likely scenario is cleavage through an initial CMD mechanism in which the acetate ligand deprotonates the pyridine N-oxide. A second scenario is through the carbon bonded to palladium in the metallacycle deprotonating the pyridine N-oxide, thereby regenerating P^tBu₃. To elucidate this step, experiments were performed with fully deuterated pyridine N-oxide. No deuterium was observed incorporated into P^tBu₃ in the products, ruling out the second scenario. An additional experiment was performed in which a cyclometalated palladium bromide complex, free from acetate, was used in place of the dimer in the previously described reactions which were isolated from the catalytic cycle. The yield here was only 37%. Thus it can be concluded that C-H cleavage of the arene takes place through an acetate ligand. Finally, DFT calculations at the B3LYP/TZVP level confirmed that C-H cleavage by the monomer is a lower energy pathway than C-H cleavage by **1**.

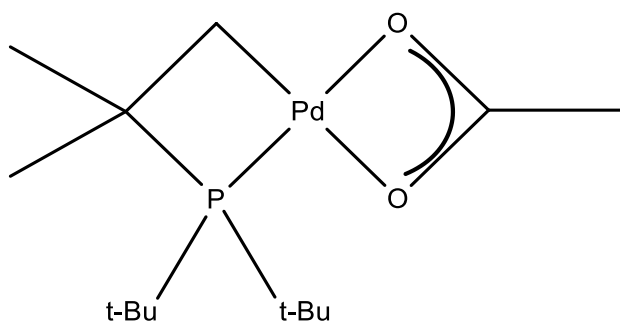


Figure 1-17: Monomer in the CC mechanism

On the basis of the aforementioned evidence, the cooperative catalysis mechanism proposed by Hartwig et al. is shown in Figure 1-18.

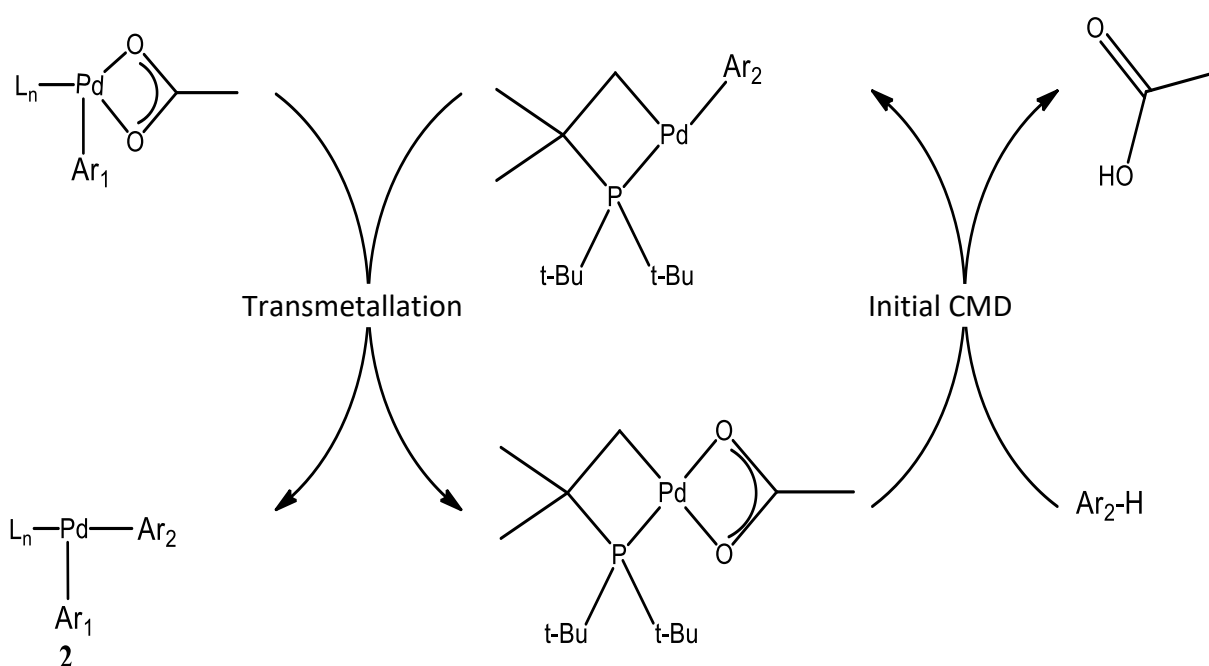


Figure 1-18 : The Cooperative Catalysis Mechanism

The two essential steps in this mechanism are initial deprotonation of Ar_2 via a CMD pathway by the monomer, followed by transmetallation. The species that subsequently undergoes reductive elimination is the same as in the CMD mechanism (See structure **2** in both Figure 1-15 and Figure 1-18).

1.3.4 The Catalytic Cycle

However, the aforementioned mechanisms do not operate in isolation and take place within a well-established catalytic cycle that occurs throughout transition metal catalysis. In the case of the system under consideration we are studying the $\text{Pd}(0)$ to $\text{Pd}(\text{II})$ manifold. This involves oxidative addition, C-H bond cleavage and finally reductive elimination (Figure 1-19). Thus, the

aforementioned CMD and CC mechanisms are but one crucial step in the cycle that determines region-selectivity, namely the C–H bond cleavage step. In order to gain a thorough understanding of the reactions at hand and unambiguously ascertain the rate determining step(s) it is necessary to also consider oxidative addition. This will be the focus of Chapters 5.

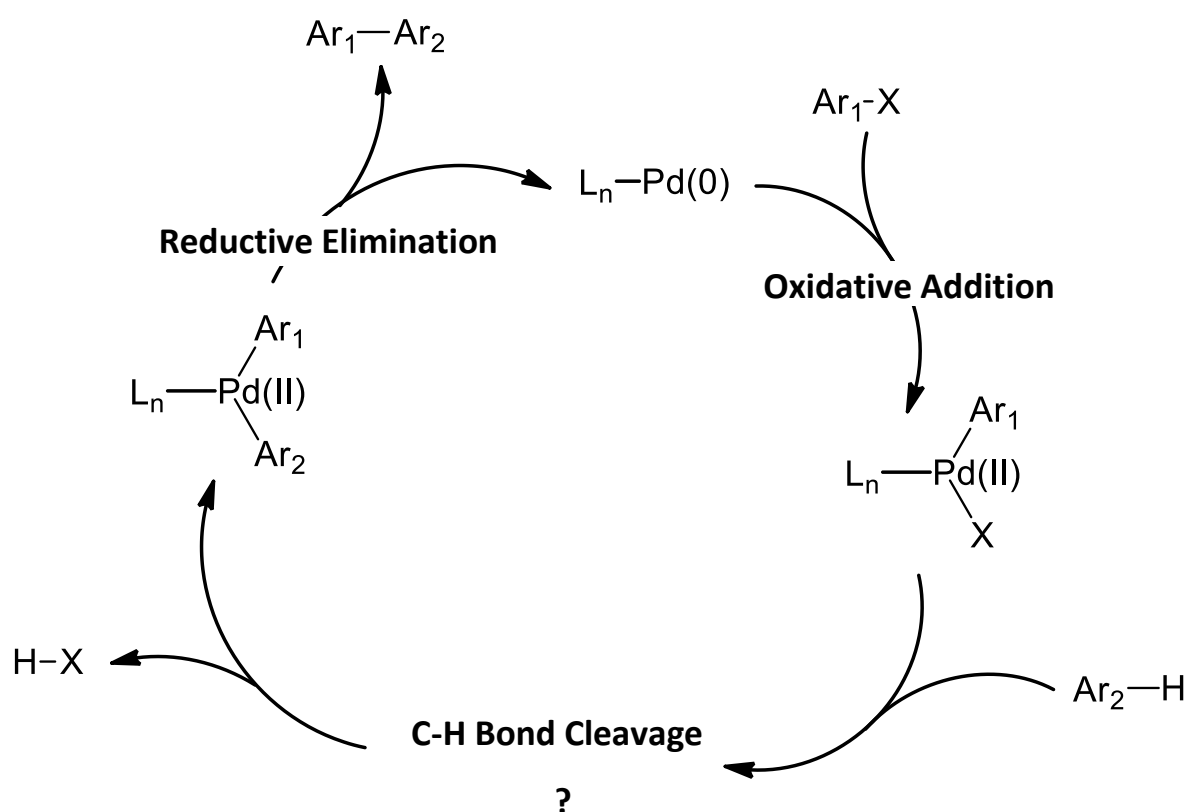


Figure 1-19: The Catalytic Cycle

1.4 Aims

- To investigate the pathway of the CMD mechanism under phosphine free conditions and determine whether such a mechanism can correctly predict the experimental product distributions observed.
- To investigate the pathway of the CMD mechanism under conditions of 1 equivalent of P^tBu_3 ligand, and determine whether such a mechanism can correctly predict the experimental product distributions observed.
- To investigate the pathway of the Cooperative Catalysis mechanism and determine whether such a mechanism can correctly predict the experimental product distributions observed under conditions of 2 equivalents of P^tBu_3 ligand.
- To investigate the mechanism of active catalyst formation via the process of oxidative addition, and thus complete modelling of the catalytic cycle.
- To ascertain the most efficient level of theory for these systems, both in terms of solvation model as well as level of correlation treatment.
- From the above, to elucidate the rate determining step for the entire catalytic cycle, building a concrete preliminary investigation of these systems and providing a target for future research to optimize.

References

1. M. E. Matheron and M. Porchas, *Plant Disease*, 2004, **88**, 665-668.
2. S. Yusuf, *American Journal of Cardiology*, **89**, 18-25.
3. B. Li, N. P. Peet, M. M. Butler, J. C. Burnett, D. T. Moir and T. L. Bowlin, *Molecules*, 2010, **16**, 202-220.
4. O. Núñez, J.-B. Kim, E. Moyano, M. T. Galceran and S. Terabe, *Journal of Chromatography A*, 2002, **961**, 65-75.
5. K. F. Tempero, V. J. Cirillo and S. L. Steelman, *British Journal of Clinical Pharmacology*, 1977, **4**, 315-365.
6. F. Ullmann and J. Bielecki, *Berichte der deutschen chemischen Gesellschaft*, 1901, **34**, 2174-2185.
7. R. J. P. Corriu and J. P. Masse, *Journal of the Chemical Society, Chemical Communications*, 1972, DOI: 10.1039/C3972000144A, 144a-144a.
8. K. Tamao, K. Sumitani and M. Kumada, *Journal of the American Chemical Society*, 1972, **94**, 4374-4376.
9. J. Hassan, M. Sévignon, C. Gozzi, E. Schulz and M. Lemaire, *Chemical Reviews*, 2002, **102**, 1359-1470.
10. H. Felkin and G. Swierczewski, *Tetrahedron*, 1975, **31**, 2735-2748.
11. M. Yamamura, I. Moritani and S.-I. Murahashi, *Journal of Organometallic Chemistry*, 1975, **91**, C39-C42.
12. A. Sekiya and N. Ishikawa, *Journal of Organometallic Chemistry*, 1976, **118**, 349-354.
13. D. Alberico, M. E. Scott and M. Lautens, *Chemical Reviews*, 2007, **107**, 174-238.
14. J. K. Stille, *Angewandte Chemie International Edition in English*, 1986, **25**, 508-524.
15. N. Miyauchi, T. Yanagi and A. Suzuki, *Synthetic Communications*, 1981, **11**, 513-519.
16. V. Farina, V. Krishnamurthy, W. J. Scott and W. J. Scott, *The Stille Reaction*, Wiley, 1998.
17. I. J. Boyer, *Toxicology*, 1989, **55**, 253-298.
18. S. Potavathri, A. S. Dumas, T. A. Dwight, G. R. Naumiec, J. M. Hammann and B. DeBoef, *Tetrahedron Letters*, 2008, **49**, 4050-4053.
19. L. Ackermann, R. Vicente and A. R. Kapdi, *Angewandte Chemie International Edition*, 2009, **48**, 9792-9826.
20. N. L. Allinger, *Journal of the American Chemical Society*, 1977, **99**, 8127-8134.
21. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2013, **3**, 198-210.
22. F. Jensen, *Introduction to Computational Chemistry*, Wiley, 2013.
23. M. Born and R. Oppenheimer, *Annalen der Physik*, 1927, **389**, 457-484.
24. C. C. J. Roothaan, *Reviews of Modern Physics*, 1951, **23**, 69-89.
25. D. Feller and K. A. Peterson, *The Journal of Chemical Physics*, 1998, **108**, 154-176.
26. I. Ema, J. Garcia De La Vega and G. Ramirez, *Journal of Computational Chemistry*, 2003, **24**, 859-868.
27. G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders and T. Ziegler, *Journal of Computational Chemistry*, 2001, **22**, 931-967.
28. K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chemical Physics Letters*, 1989, **157**, 479-483.
29. P. Hohenberg and W. Kohn, *Physical Review*, 1964, **136**, B864-B871.
30. W. Kohn and L. J. Sham, *Physical Review*, 1965, **140**, A1133-A1138.
31. J. P. Perdew and K. Schmidt, *AIP Conference Proceedings*, 2001, **577**, 1-20.
32. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *The Journal of Chemical Physics*, 2010, **132**, 154104.
33. J. Tao and J. P. Perdew, *The Journal of Chemical Physics*, 2005, **122**, 114102.
34. I. Ciofini, C. Adamo and H. Chermette, *Chemical Physics*, 2005, **309**, 67-76.
35. A. Klamt and G. Schüürmann, *Journal of the Chemical Society, Perkin Transactions 2*, 1993, 799-805.
36. M. Li, L. C. Pan, H. A. Simmons, Y. Li, D. R. Healy, B. S. Robinson, H. Z. Ke and T. A. Brown, *Bone*, **39**, 796-806.

37. C. R. Waites, M. A. Dominick, T. P. Sanderson and B. E. Schilling, *Toxicological Sciences*, 2007, **100**, 248-258.
38. E. Adeghate, A. Adem, M. Y. Hasan, K. Tekes and H. Kalasz, *The Open Medicinal Chemistry Journal*, 2011, **5**, 93-98.
39. R. A. Potter, B. G. Burns, J. M. v. d. Riet, D. H. North and R. Darvesh, *Journal of AOAC International*, 2007, **90**, 343-348.
40. R. L. Parsons and C. H. Heathcock, *The Journal of Organic Chemistry*, 1994, **59**, 4733-4734.
41. V. Blanc, D. Lagneaux, P. Didier, P. Gil, P. Lacroix and J. Crouzet, *Journal of Bacteriology*, 1995, **177**, 5206-5214.
42. A. Y. Shaw, M. C. Henderson, G. Flynn, B. Samulitis, H. Han, S. P. Stratton, H.-H. S. Chow, L. H. Hurley and R. T. Dorr, *Journal of Pharmacology and Experimental Therapeutics*, 2009, **331**, 636-647.
43. *United states , Japan Pat.*, 2001.
44. V. N. L. P.N Zhmurin , V.D Titskaya , A.F Adadurov , D.A Elyseev , V.N Pereymak, *Functional Materials -National Academy of Sciences of Ukraine*, 2014, **21**, 8.
45. S. T. Diver, *Encyclopedia of Reagents for Organic Synthesis*, 1995.
46. M. R. Netherton and G. C. Fu, *Organic Letters*, 2001, **3**, 4295-4298.
47. B. Glover, K. A. Harvey, B. Liu, M. J. Sharp and M. F. Tymoschenko, *Organic Letters*, 2003, **5**, 301-304.
48. T. S. S. Pivsa-Art, Y. Kawamura, M. Miura and M. Nomura, *Bulletin of the Chemical Society of Japan* 1998, **71**, 467.
49. B. S. Lane, M. A. Brown and D. Sames, *Journal of the American Chemical Society*, 2005, **127**, 8050-8057.
50. T. Okazawa, T. Satoh, M. Miura and M. Nomura, *Journal of the American Chemical Society*, 2002, **124**, 5286-5287.
51. D. L. Davies, S. M. A. Donald and S. A. Macgregor, *Journal of the American Chemical Society*, 2005, **127**, 13754-13755.
52. M. Ye, G.-L. Gao, A. J. Edmunds, P. Worthington, J. A. Morris and J.-Q. Yu, *Journal of the American Chemical Society*, 2011, **133**, 19090-19093.
53. M. Lafrance, C. N. Rowley, T. K. Woo and K. Fagnou, *Journal of the American Chemical Society*, 2006, **128**, 8754-8756.
54. H.-Y. Sun, S. I. Gorelsky, D. R. Stuart, L.-C. Campeau and K. Fagnou, *The Journal of Organic Chemistry*, 2010, **75**, 8180-8189.
55. S. Zhang, L. Shi and Y. Ding, *Journal of the American Chemical Society*, 2011, **133**, 20218-20229.
56. D. L. Davies, S. M. Donald and S. A. Macgregor, *Journal of the American Chemical Society*, 2005, **127**, 13754-13755.
57. D. García-Cuadrado, P. de Mendoza, A. A. Braga, F. Maseras and A. M. Echavarren, *Journal of the American Chemical Society*, 2007, **129**, 6880-6886.
58. M. Lafrance and K. Fagnou, *Journal of the American Chemical Society*, 2006, **128**, 16496-16497.
59. S. I. Gorelsky, D. Lapointe and K. Fagnou, *Journal of the American Chemical Society*, 2008, **130**, 10848-10849.
60. S. Y. Tang, Q. X. Guo and Y. Fu, *Chemistry-A European Journal*, 2011, **17**, 13866-13876.
61. Y. Tan, F. Barrios-Landeros and J. F. Hartwig, *Journal of the American Chemical Society*, 2012, **134**, 3683-3686.
62. D. P. Fairlie and B. Bosnich, *Organometallics*, 1988, **7**, 946-954.
63. G. P. Van Strijdonck, M. D. Boele, P. C. Kamer, J. G. De Vries and P. Van Leeuwen, *European journal of Inorganic Chemistry*, 1999, 1073-1076.
64. A. Kina, H. Iwamura and T. Hayashi, *Journal of the American Chemical Society*, 2006, **128**, 3904-3905.
65. D. B. Collum, A. J. McNeil and A. Ramirez, *Angewandte Chemie International Edition*, 2007, **46**, 3002-3017.
66. Z. Shen, P. K. Dornan, H. A. Khan, T. K. Woo and V. M. Dong, *Journal of the American Chemical Society*, 2009, **131**, 1077-1091.

2. Methodology

Contents

2.1 General Computational Details.....	50
2.1.1 The ORCA Computational Chemistry Package	50
2.1.2 Hardware Resources	50
2.1.3 SCF Convergence Acceleration.....	51
2.1.4 SCF Convergence Criteria	52
2.1.5 The RIJCOSX approximation	53
2.1.6 The Def2-SVP Basis Set.....	54
2.1.7 The Def2-TZVP Basis Set.....	57
2.1.8 The ma-Def2-SVP Basis Set	59
2.1.9 Visualisation and Molecule Building Software.....	61
2.2 Density Functional Theory	61
2.2.1 The TPSSh Functional	61
2.2.2 Dispersion Correction.....	62
2.2.3 Exchange-Correlation Integration Grids	66
2.2.5 Geometry Optimisations	68
2.2.6 Frequency Calculations	70
2.2.7 Transition state optimisations	71
2.3 Double Hybrid DFT	75
2.4 DLPNO-CCSD(T).....	78
References	81

2.1 General Computational Details

2.1.1 The ORCA Computational Chemistry Package

All the calculations reported in this work were performed using the quantum chemistry package ORCA,¹ unless otherwise stipulated. The particular version used was ORCA 3.0.2. ORCA is a general purpose quantum chemistry package developed by Frank Neese and colleagues at the Max Planck Institute for Chemical Energy Conversion in Germany. The binaries, but not source code, of ORCA are available free of charge for academic users for a variety of operating systems (OSs). ORCA is thus completely free, but not open source. ORCA is flexible, efficient and user-friendly. Developed with a specific emphasis on larger molecules, transition metal complexes and their spectroscopic properties, it is an ideal package to study the palladium catalysis problem that is the focus of this project. ORCA features virtually all modern electronic structure methods ranging from semi-empirical methods to DFT to single- and multi-reference correlated *ab initio* approaches. Additionally, ORCA includes the capabilities to treat environmental and relativistic effects. It is completely parallelized using the MPI interface. Thus, ORCA is ideally suited for situations where license fees for commercial software packages are prohibitively expensive but the latest computational chemistry methodology is required.

2.1.2 Hardware Resources

Calculations were performed by making use of the national Centre for High Performance Computing (CHPC). Specifically, the Dell Westmere and Westmere components of the Tsessebe cluster were used. Here each node was equipped with 2 x 2.93 GHz Xeon processors with 6 individual cores each, or 12 processors per node, with a total of 24 GB local RAM. The cluster makes use of Centos 5.8 as an OS and the Lustre file system. ORCA has not specifically been optimised to run on this file system and thus it was only possible to use one node per calculation. The limited functionality of the local cluster in the School of Chemistry, University of the Witwatersrand, was used as much as possible. Here some nodes were equipped with 2 x E5520 Intel Xeon processors running at 2.27GHz with four cores each, or 8 processors per node and 8192 KB cache memory for each processor. Other nodes were equipped with 2 x E5420 Intel Xeon processors running at 2.50GHz, again with four cores each, or 8 processors per node and 6144 KB cache memory for each processor. Here ORCA could be run across more than one node. This was especially beneficial for frequency calculations, to be discussed below.

2.1.3 SCF Convergence Acceleration

In order to accelerate and stabilize self-consistent field (SCF) convergence, like all modern quantum chemistry packages, ORCA employs the Direct Inversion of the Iterative Subspace (DIIS) method,^{2, 3} also known as Pulay mixing, after its inventor. This is a technique that significantly accelerates the rate of SCF convergence towards the end of the procedure. Without DIIS the energy can fluctuate problematically towards the end of the SCF procedure, leading to unnecessarily long convergence times. In brief, during each SCF step an error vector is constructed from the commutator of the Fock and density matrices, $[\mathbf{F}, \mathbf{P}]$. $[\mathbf{F}, \mathbf{P}]$ is referred to as the DIIS error. This information is then used to extrapolate a new Fock matrix using both present and previous Fock matrices. Thus DIIS is essentially a sophisticated extrapolation technique. In ORCA there is an input variable "DIISStart" that controls the minimum energy difference between two SCF cycles before DIIS is started. Throughout this project the default value of 0.2 Eh was used. In case the SCF iterations proceeded to oscillate drastically without ever reaching this minimum energy difference ORCA sets a limit on the number of iterations that are allowed to proceed before DIIS is activated. This is controlled by the variable "DIISMaxIt". The default value of 12 was always used.

In order to ensure that the SCF process is as efficient as possible before it is appropriate to turn on DIIS, two other SCF convergence acceleration methods are implemented in ORCA. The first of these is static damping, which is mainly used to limit oscillatory energy values. Damping consists of mixing the old density with the new density by replacing the current density matrix with a weighted average of old ($\mathbf{D}_{\text{old}}$) and new ($\mathbf{D}_{\text{new}}$) density matrices in each SCF iteration:

$$\mathbf{D}_{\text{new}}' = w\mathbf{D}_n + (1 - w)\mathbf{D}_{n+1} \quad (2.1)$$

In Equation (2.1) $\mathbf{D}_{\text{new}}'$ is the damped density matrix while w is the weighting factor. Static damping simply refers to the fact that the weighting factor remains constant throughout the entire SCF cycle. Throughout this work the default value of $w = 0.70$ was used, unless convergence issues were encountered. As alluded to earlier, static damping is most important in the first few SCF iterations, before, or just after DIIS is invoked. In ORCA the variable "DampErr" controls exactly when the static damping is turned off. This is necessary as damping used in conjunction with DIIS towards the end of the SCF procedure can cause unnecessarily slow convergence. When the DIIS error, $[\mathbf{F}, \mathbf{P}]$, falls below "DampErr" the damping is no longer deemed necessary and is turned off. The default value of "DampErr" = 0.1000 was used. Therefore when the DIIS error < 0.1000 static damping is terminated by the program. This approach ensures that damping will never be turned off before DIIS is initiated.

The other SCF convergence acceleration method that acts before (and after) DIIS is activated is known as “Level Shifting”.⁴ During the SCF process, linear combinations of occupied and virtual MOs are created. This is referred to as mixing of the occupied and virtual orbitals. If the coefficients of the virtual orbitals become large during this optimisation the total energy will increase rapidly, resulting in oscillations. This mixing of the occupied and virtual orbitals can be reduced if the energy of the virtual orbitals is artificially increased by adding a sufficiently large constant to their energy. It has been shown that this helps to force convergence. Level Shifting also helps prevent flipping between near-degenerate electronic states. Analogous to the “DampErr” variable, the parameter that controls the degree of Level Shifting in ORCA is “ShiftErr”. Level Shifting is thus implemented from the first SCF cycle and when the DIIS error falls below “ShiftErr” the shifting is terminated. The default value of “ShiftErr” = 0.001 was used. The actual value of the constant added to the energy of the virtual orbitals was 0.25 Eh. This is dictated by the “LShift” variable in ORCA. Therefore, an SCF cycle starts with both static damping and Level Shifting. Generally, after a few cycles DIIS is initiated. After some time, static damping will be terminated with Level Shifting usually persisting concurrently with DIIS for a few iterations before it is finally switched off, leaving the last sequence of iterations under the influence of DIIS convergence acceleration alone.

2.1.4 SCF Convergence Criteria

ORCA, by default when using DFT for geometry optimisation and frequency calculations resorts to what its developers refer to as “TightSCF” convergence criteria. Additionally, TightSCF was applied to all single point calculations performed in this work. These “Tight” tolerances ensure that the SCF process goes through a sufficient number of iterations to produce reliable energies that are based on adequately minimised wavefunctions or densities. The convergence criteria refer to the change between two cycles and are shown in Table 2-1.

Table 2-1: Tight SCF Convergence Criteria

Property	Threshold
Total energy	$\leq 1.000 \times 10^{-8} \text{ Eh}$
One electron (i.e. orbital) energy	$\leq 1.000 \times 10^{-5} \text{ Eh}$
DIIS error	$\leq 5.000 \times 10^{-7}$

When all three of these criteria have been met SCF convergence is signalled. Although this is tuneable, ORCA by default also checks the ratio of the total energy and one-electron energy

between two cycles. If this ratio is constant, then the SCF can be considered to no longer fluctuate and convergence is signalled regardless of the DIIS error.

2.1.5 The RIJCOSX approximation

A unique feature of ORCA that helps to speed up the calculation of both the Coulomb (J) energy and exact exchange (X) is the RIJCOSX approximation.⁵ This is actually a combination of two separate approximate algorithms, the RI-J approximation used in conjunction with the COSX approximation. RI refers to “Resolution of the Identity” and J refers to the Coulomb part of the of Fock matrix, the calculation of which this approximation significantly speeds up. In brief, charge distributions arising from products of basis functions are approximated by a linear combination of auxiliary basis functions. This allows for a reduction in the computational effort of two-electron four-centre integrals as these are replaced with two- and three-centre integrals which can be calculated with higher efficiency. RI-J thus requires an additional auxiliary basis set specified by “/J” after the basis set’s keyword in an ORCA input file. The errors introduced by the RI approximation are marginal, usually much smaller than the errors in a standard DFT calculation due to basis set incompleteness. ORCA 3.0.2 actually implements an improved version of the RI method, known as the Split-RI-J approximation that is significantly faster if the basis set contains many high angular momentum functions (d-, f-, g-functions). This could be for example basis sets with many polarisation functions. For small basis sets there is virtually no difference in the results of the two algorithms.

COS refers to “Chain Of Spheres” while X refers to the exact exchange term (HF exchange), the calculation of which is accelerated by this approximation. In brief, when performing two electron integrals, the first integration is done numerically on a grid and the second which involves the Coulomb singularity is done analytically. This introduces a grid on top of the DFT integration grid (Section 2.2.3). The size of this grid is controlled by the “GridXn” keyword, where n ranges from 1 to 9.

When the RIJCOSX approximation is used in conjunction with large basis sets, accelerations in timing up to a factor of 60 have been observed.⁵ The corresponding decrease in accuracy is marginal with the total energy as well as relative energy error introduced being around 1 kcal.mol⁻¹. Bond distances are typically affected by a fraction of a picometre, bond angles by a tenth of a degree and dihedral angles by approximately 1 degree. Lastly, for vibrational frequencies the error introduced ranges from 2-10 cm⁻¹. However, all the aforementioned errors can be reduced by increasing the size of the RIJCOSX grid from the default value. The

RIJCOSX approximation was thus used extensively throughout this project for hybrid DFT, double hybrid DFT as well as Domain-Based Local Pair Natural Orbital-CCSD(T), DLPNO-CCSD(T), to be discussed below. It is invoked by using the “RIJCOSX” keyword in ORCA. However, in the cases where GGA DFT functionals were used, only the Split-RI-J approximation was implemented as no exact exchange is present here for COSX to speed up. This was invoked by using the “RI” keyword in ORCA.

2.1.6 The Def2-SVP Basis Set

For all routine geometry optimisations and frequency calculations the Def2-SVP basis set from the Ahlrichs group was used.⁶ Here, “Def2” refers to “default” as these basis sets were originally developed as the default basis sets in the TURBOMOLE program, the “2” simply indicating a revised version. “SVP” refers to Split Valance Polarisation. As mentioned in section 1.2.2, in a SV basis set, the number of contracted GTOs (Gaussian-type orbitals) used to describe the valence electrons is double the number of valence atomic orbitals; hence “split valance”. At this point it is useful to again refer to the general functional form of a normalized GTO in atom-centred Cartesian coordinates:

$$\phi(x, y, z; \alpha, i, j, k) = \left(\frac{2\alpha}{\pi} \right)^{3/4} \left[\frac{(8\alpha)^{i+j+k} i! j! k!}{(2i)!(2j)!(2k)!} \right]^{1/2} x^i y^j z^k e^{-\alpha(x^2+y^2+z^2)} \quad (2. 2)$$

In Equation (2. 2) α is an exponent controlling the size of the GTO (ϕ); i , j , and k are non-negative integers that determine the shape of the orbital. If i , j , and k are all 0 then the orbital has spherical symmetry i.e. an s orbital. When only one of the indices is 1 then the function has axial symmetry about a single Cartesian axis i.e. a p orbital. When the sum of the indices is equal to two the function represents a d orbital, and so on. GTOs are then linearly combined in a contraction scheme (section 1.2.2) to form contracted GTOs (φ):

$$\varphi = \sum_{i=1}^M c_a \phi(x, y, z, \alpha, i, j, k) \quad (2. 3)$$

For a hydrogen atom with an electron configuration $1s^1$, two contracted s type GTOs will be used to describe the single valence atomic s orbital in a SV basis set. However, SVP also includes a single polarisation function (hence “Polarisation”), and thus a p type contracted GTO is also present to increase the flexibility of the s functions, better accounting for a variety of bonding situations. Thus overall $2 \times s$ and $1 \times p$ type (collectively referring to a set of three

possible p orbitals) contracted GTOs would be employed for an H atom. In basis set nomenclature this would be referred to as 2s1p, not to be confused with the standard way for writing atomic electronic structure.

In basis set nomenclature the $2 \times s$ and $1 \times p$ functions for hydrogen are enclosed in square brackets to indicate that they refer to the number of contracted GTOs:

[2s1p].

The number and type of primitive GTOs used to construct these contracted GTOs in a similar fashion are enclosed in round brackets; for H in Def2-SVP:

(4s1p).

A forward slash separates the primitive and contracted brackets, starting with the primitives: (4s1p) / [2s1p].

Thus the $2 \times s$ contracted GTOs are constructed from a total of 4 primitive GTOs, while the $1 \times p$ function is itself a primitive GTO. This notation however provides no information on the specific contraction scheme. Specifically, how many primitives are used to construct each of the s type contracted GTOs? A number of possibilities exists, such as using two primitives in each or three primitives for the first and the remaining orbital consisting of a single primitive. This information is normally stipulated in curly brackets with a forward slash separating the different types of orbitals starting from s, for H in Def2-SVP:

{31/1}

Thus the first contracted s type GTO is constructed from 3 primitive GTOs ($M = 3$ in Equation (2. 3)), with the remaining s orbital being the normalised 4th primitive. The “/1” indicates that the p type GTO is also a single primitive. Each of these primitives is defined by an exponent (α in Equation (2. 2)) and its contribution to the GTO by a contraction coefficient (C_a in Equation (2. 3)). These are shown for all the primitive GTOs used to describe the H atom in Def2-SVP in Table 2-2, (C_a values unnormalised). Similarly, for a carbon atom with an electron configuration of $1s^2 2s^2 2p^2$ the $1 \times s$ and $1 \times p$ valence orbitals are each described by two GTOs (again collectively referring to a set of three possible p orbitals), giving a total of $2 \times s$ and $2 \times p$ contracted valence GTOs. This combined with the $1 \times s$ core orbital and a d polarisation function gives:

[3s2p1d],

which is then constructed of primitive GTOs as:

(7s4p1d),

giving:

(7s4p1d) / [3s2p1d] {511/31/1}.

Thus one of the three contracted s orbitals is made up of 5 primitives, the last two by a single primitive each, the first contracted p set by 3 primitives, the last by 1 and the d orbital by a single primitive. The α and C_a (unnormalised) values for each primitive are shown in Table 2-3.

Table 2-2: H Def2-SVP

(4s1p) / [2s1p]	{31/1}
α	C_a
3 s	
13.010701	0.019682
1.962257	0.137965
0.444538	0.478319
1 s	
0.121950	1.000000
1 p	
0.800000	1.000000

Table 2-3: C: Def2-SVP

(7s4p1d) / [3s2p1d]	{511/31/1}
α	C_a
5 s	
1238.401694	0.005457
186.290050	0.040638
42.251176	0.180256
11.676558	0.463151
3.593051	0.440872
1 s	
0.402451	1.000000
1 s	
0.130902	1.000000
3 p	
9.468097	0.038388
2.010355	0.211170
0.547710	0.513282
1 p	
0.152686	1.000000
1 d	
0.800000	1.000000

An effective core potential (ECP) was applied to palladium. This ECP effectively replaces 28 electrons considered to be the core, the entire 1s to 3d shells. The remaining $4s^2 4p^6 4d^{10}$ electrons in palladium are then described by the Def2-SVP basis set. In this way the required basis set size is reduced substantially and importantly scalar relativistic effects are accounted for. This was stipulated in an ORCA calculation by the “ECP{Def2-SVP=Pd,Def2-SVP/J}” keyword.

The contracted GTOs, primitive GTOs, contraction scheme as well as the α and C_a values for all primitives of each atom considered in this study are shown in Appendix A for the Def2-SVP

basis set. The corresponding Def2-SVP/J auxiliary Coulomb fitting basis set⁷ was used whenever the RI-J approximation was invoked in conjunction with Def2-SVP. This auxiliary basis set is approximately three times larger than Def2-SVP. For example, Def2-SVP/J uses the following contraction scheme for an H atom, which even includes d polarisation functions:

(5s2p1d) / [3s1p1d] {311/2/1},

while for carbon the contraction scheme is:

(12s5p4d2f1g) [6s4p3d1f1g] {711111/2111/211/2/1}.

Appendix A also contains the contracted GTOs, primitive GTOs, contraction scheme as well as the α and C_a values for all primitives for each atom considered in this study for the Def2-SVP/J auxiliary Coulomb fitting basis set.

2.1.7 The Def2-TZVP Basis Set

The Def2-TZVP basis set, again from the Ahlrichs group⁶, was used for Double Hybrid DFT and DLPNO-CCSD(T) calculations discussed in sections 2.3 and 2.4, respectively. As mentioned in section 1.2.2, in a TZV basis set the number of contracted GTOs used to describe the valence electrons is triple the number of valence atomic orbitals. This larger basis set was necessary as wavefunction based methods (and thus DH DFT) show a much slower convergence of the energy with respect to the number of basis functions compared to DFT. For a hydrogen atom with an electron configuration of $1s^1$, only a single valence orbital is present, the 1s orbital. The Def2-TZVP basis set uses three contracted s type GTOs to describe this 1s orbital. Analogous to Def2-SVP, a single polarisation function is also added giving: [3s1p]. Things however get more complex beyond hydrogen with the valence s orbitals of 1st row elements being described by 4 rather than 3 GTOs, the valence p orbital by the expected 3 GTOs and $2 \times d$ as well as $1 \times f$ polarisation functions. This gives [5s3p2d1f] as the contracted GTOs for carbon. The need for extra s functions arises as a consequence of shielding of the nuclear charge by core electrons, allowing valence s orbitals beyond He greater flexibility. These extra s and polarisation functions as well as an additional steep d function on certain elements ensure that the Def2-TZVP basis set is approximately 1.92 times larger than its Def2-SVP counterpart, rather than the expected ratio of 1.5. The number of primitive GTOs as well as contracted GTOs for each type of orbital together with the α and C_a (unnormalised) values for each primitive for hydrogen and carbon in the Def2-TZVP basis set are shown in Table 2-4 and Table 2-5. For the rest of the atoms considered in this study these values are provided in Appendix A.

Table 2-4: H Def2-TZVP

(5s1p) / [3s1p]	{311/1}
α	C_a
3 s	
34.061341	0.006025
5.123575	0.045021
1.164663	0.201897
1 s	
0.327230	1.000000
1 s	
0.103072	1.000000
1 p	
0.800000	1.000000

Table 2-5 : C Def2-TZVP

(11s6p2d1f) / [5s3p2d1f]	{62111/411/11/1}		
α	C_a	α	C_a
6 s		1 d	
13575.349682	0.000222	1.097000	1.000000
2035.233368	0.001723	1 d	
463.225624	0.008926	0.318000	1.000000
131.200196	0.035728	1 f	
42.853016	0.110763	0.761000	1.000000
15.584186	0.242956		
2 s			
6.206714	0.414403		
2.576490	0.237450		
1 s			
0.576963	1.000000		
1 s			
0.229728	1.000000		
1 s			
0.095164	1.000000		
4 p			
34.697232	0.005333		
7.958262	0.035864		
2.378083	0.142159		
0.814332	0.342705		
1 p			
0.288875	0.464458		
1 p			
0.100568	0.249558		

An effective core potential (ECP) was again applied to palladium and replaced the 28 core electrons. This was stipulated in ORCA by the “ECP{Def2-TZVP=Pd, Def2-TZVP /J}” keyword.

The Coulomb fitting basis set used in the RI-J approximation is exactly the same as that for the Def2-SVP, Def2-TZVP and Def2-QZVP canonical basis sets,⁷ although, it was invoked using the keyword “Def2-TZVP/J” for consistency. This universal auxiliary basis set can actually be used for the entire Def2- range of canonical orbital basis sets; see Appendix A, Section 3, for details.

2.1.8 The ma-Def2-SVP Basis Set

As mentioned in Section 1.2.2.3 (Chapter 1), for many chemically relevant systems, such as anions and highly excited states, diffuse functions are required to reproduce reality.⁸ In such systems loosely bound electrons are present, which are localised far from the majority of the electron density. As a consequence, the tail of wavefunction becomes important, an area which is usually neglected, in the sense of being described by much fewer basis functions than the core of an atom. The diffuseness of a basis set is measured by the size of the smallest exponential in the GTO for each type of orbital i.e. the size of α in Equation (2. 2) for each angular momentum (s, p, d orbital etc.). The inclusion of GTOs with smaller values of α means that the basis set is capable of better describing a more diffuse charge distribution.

The ma-Def2-SVP basis set is an extension of the Def2-SVP basis set in which diffuse functions have been added.⁹ This basis set will be applied extensively during the modelling of oxidative addition to anionic palladium complexes in Chapter 5. Specifically, these extra diffuse functions are a set of s and p GTOs that have been added to all atoms across the periodic table, except hydrogen. Hence, they are referred to as minimally augmented (ma). This is an economical way of creating a diffuse basis set that can provide accurate anion calculations. Reliable electron affinities have been calculated using this basis set , despite its small size.⁹

The exponential parameters of the added functions were determined consistently across the periodic table by simply dividing α of the most diffuse s and p GTOs in the corresponding Def2-SVP basis set by a factor of 3. For a carbon atom in the Def2-SVP basis set (Table 2-3), the lowest exponential for a s function corresponds to a 1s orbital and is 0.130902. The lowest exponential for a p function corresponds to a 1p orbital and is 0.100568. Therefore, the ma-Def2-SVP basis set for carbon will contain an extra s function where $\alpha = 0.043634$ as well as an extra p function where $\alpha = 0.033523$. The number of primitive GTOs as well as contracted GTOs for each type of orbital together with the α and C_a (unnormalised) values for each primitive for carbon in the ma-Def2-TZVP basis can be found in Table 2-6. For the rest of the atoms considered in this study these values are provided in Appendix A.

Table 2-6: C ma-Def2-SVP

(8s5p1d) / [4s3p1d]	{5111/311/11}
α	C_a
5 s	
1238.401694	0.005457
186.290050	0.040638
42.251176	0.180256
11.676558	0.463151
3.593051	0.440872
1 s	
0.402451	1.000000
1 s	
0.130902	1.000000
3 p	
9.468097	0.038388
2.010355	0.211170
0.547710	0.513282
1 p	
0.152686	1.000000
1 d	
0.800000	1.000000
1 s	
0.043634	1.000000
1 p	
0.050895	1.000000

An ECP was applied to palladium in exactly the same way as when using a Def2-SVP basis set. I.e. the 28 core electrons were replaced by invoking the “ECP{Def2-SVP=Pd, Def2-SVP /J}” keyphrase. When using the ma-Def2-SVP basis set in conjunction with the RI-J approximation, no appropriate auxiliary Coulomb fitting basis set has been published. This problem was circumvented by invoking the “AutoAux” feature of ORCA. When using this feature the program automatically constructs of a general purpose fitting basis for the system.

2.1.9 Visualisation and Molecule Building Software

In order to visualise the results of calculations as well as to construct input molecules in Cartesian coordinates (.xyz files) the commercial program Chemcraft¹⁰ was extensively used. Chemcraft is a graphical program for working with quantum chemistry data. Amongst others, it can read ORCA output files directly, allowing one to view SCF convergence graphs, geometry optimisation and vibrational frequency animations. Additionally, ORCA can open Cube files as well as render molecular orbitals if the coefficients are present in the output file. Additionally, it allows for the generation of publication-ready images and all three dimensional pictures of molecules in this project were generated using this program. A modified version of the open source molecule editor and visualizer Avogadro¹¹ was used to aid in the construction and modification of structures. This modified version of Avogadro (version 1.1.1) contains extended ORCA support including an ORCA input generator and the ability to view spectra, vibrations and molecular orbitals.

2.2 Density Functional Theory

2.2.1 The TPSSh Functional

Despite their importance in chemistry, particularly in catalysis, transition metal complexes can be notoriously problematic to model due to strong near-degeneracy correlation (non-dynamic/static correlation) which is typical of atoms with partially filled *d* shells. An accurate multi-reference wavefunction treatment is generally required in order to predict the correct ground-state symmetry. These include methods such as complete active space self-consistent field with second-order perturbation theory (CASPT2) or multi-reference configuration interaction (MRCI) methods. Although invaluable for benchmark studies these methods scale unfavourably with system size, precluding a multi-reference treatment for many problems of chemical interest. However, as palladium (4d¹⁰) has a closed shell electron configuration, single-determinant reference methods are fortunately sufficient, allowing us to generate useful information from hybrid DFT calculations.

The workhorse of this project, in the sense of being used for final geometry optimisations and frequency calculations was hybrid DFT. This is a compromise between efficiency and accuracy. As mentioned in section 1.2.2, currently hybrid DFT represents the penultimate rung on Jacob's ladder of functionals. Starting from a (meta) GGA functional a percentage of exact exchange from HF is incorporated in the E_{xc} term:

$$E_{XC}^{Hybrid} = aE_X^{exact} + (1-a)E_X^{GGA} + E_C^{GGA} \quad (2.4)$$

Initially due to its proven track record in a diverse array of systems we considered the three-parameter exchange functional of Becke (B3)¹² and the correlation functional of Lee, Yang, and Parr (LYP)¹³ i.e. B3LYP. This mixes in 20% exact HF exchange, i.e. a in Equation (2.4) is 0.20. However, later in the study we encountered troublesome transition state structures with multiple negative frequencies. The TPSSh^{14, 15} functional (both exchange and correlation) was then implemented and subsequently resolved this issue. TPSSh mixes in 10% exact HF exchange and thus $a = 0.10$. This is the only empirical parameter in the functional. The value of 10% was originally deduced by minimising the mean absolute deviation of the standard enthalpy of formation, ΔH_f^0 , for 223 molecules from the G3/99¹⁶ test set. This choice was made based on numerous studies that indicated the slight superiority of the TPSSh functional when modelling systems containing transition metals to that of B3LYP and other hybrid functions¹⁷⁻¹⁹.

2.2.2 Dispersion Correction

The instantaneous dipole moments on atoms due to the polarizability of their electron distribution creates an attractive force present in all matter, the dispersion force (also known as the London dispersion force). Such interactions are referred to as long range in electronic structure theory as they arise as a consequence of the interplay between electrons belonging to the densities of two otherwise non-interacting molecules or atoms. Dispersion forces are entirely due to dynamic electron correlation. The unknown exact density functional must account for such electron correlation effects. Unfortunately, in the present implementations of DFT the exchange-correlation (XC) terms are approximated by local functionals. At any given point in a molecule the XC potential is calculated from the density (LDA) at that point and depending on the type of functional, possibly the first derivative of the density (GGA), the Laplacian of the density or equivalently the orbital kinetic energy density (meta-GGA), all at that particular point in space. Thus, the XC functional is a local functional and remains unaffected by the density of another, distant system if no overlap in density occurs. Therefore, in principle, a local exchange-correlation functional, as found in all implementations of DFT, is unable to correctly describe the lowering in energy that occurs as a consequence of two unshared electron distributions interacting i.e. dispersion forces.

A number of different methods have been developed in an attempt to improve this inherent lack of dispersion in DFT. This is because dispersion effects play a crucial role in the ability (or

inability) of DFT to reach chemical accuracy, especially for large molecules. Furthermore, dispersion effects underlie many fundamental molecular processes such as the formation of the tertiary and quaternary structures of proteins, the formation of molecular crystals and binding of pharmaceuticals to drug targets within living organisms, to mention but a few. Such methods include nonlocal van der Waals density functionals²⁰, vdW-DF. In this approach, extensions have been developed for a number of common hybrid GGA functionals. These use the standard exchange and correlation terms but include an extra non-local term in the XC functional, which covers mainly long-range dispersive energy. They represent a non-empirical way to compute the dispersion energy for arbitrary systems, solely based on the electron density. Although accurate, they are significantly costlier computationally than the corresponding standard functionals. Additionally, the extra non-local term introduces a high level of numerical complexity, and these methods suffer from the induced instability with respect to convergence during the SCF process. The biggest disadvantage of nonlocal van der Waals density functionals is that currently no analytical gradients for the computation of atomic forces during geometry optimisations are available. These are thus useful for single point reference calculations but too costly for geometry optimisations.

Other approaches such as dispersion-corrected atom-centred potentials (DCACPs)²¹ have also been developed. In this approach, dispersion forces are approximated by an atom-electron interaction mediated by appropriate nonlocal effective core potential projectors. DCACPs have however, not been extensively adapted and tested. Although these methods perform well for equilibrium structures of small organic molecules in standard benchmarks²², little is known about how they reproduce intramolecular dispersion or account for thermochemistry. Additionally, DCACPs contain a high degree of empiricism, making their use unsuitable for systems that have not been well explored. From a conceptual standpoint a major flaw in the DCACP approach is that the atomic parameters are fixed for each element, regardless of the environment the atom finds itself in. Therefore, these parameters do not reflect the changes of dispersion coefficients with the hybridization or oxidation state of an atom.

Grimme's DFT-D method (DFT-Dispersion) is currently the most widely implemented method of dispersion correction for DFT. It has been applied successfully to thousands of different systems in the last decade.²³ These including both intermolecular and intramolecular cases, for both molecular and extended systems ranging from noble gas dimers²⁴, to the dimers of a large variety of organic molecules²⁵ to the physisorption of nucleic acid bases on graphene layers.²⁶ DFT-D is a kind of semi-classical approach in which the dispersion energy is

computed separately and added to the standard DFT energy, in a non-self-consistent fashion (Equation (2. 5)).

$$E = E_{DFT} + E_{Disp} \quad (2. 5)$$

Thus, this dispersion energy does not influence the calculation of the electron structure but is added after the orbitals have already been determined. Although seemingly a significant approximation this approach has proved remarkably useful. In the latest implementation, namely DFT-D3²⁷ both a two body and a three body term are included. These account for the dispersion between each possible pair and triplet of atoms in the system, respectively. The total dispersion is then just the sum of these two terms:

$$E_{Disp} = E^2 + E^3 \quad (2. 6)$$

However, the two bodied term contributes significantly more to dispersion than the three body term. Inclusion of the three bodied term additionally increases the computational cost from $(N_{atoms})^2$ to $(N_{atoms})^3$. Although this increase in cost is somewhat insignificant as the DFT SCF energy will take 2 to 3 orders of magnitude longer to calculate, we chose not to implement this term. This is because only numerical gradients are available in ORCA making geometry optimisations inefficient. Furthermore, for small and medium sized systems including this terms leads to slightly less accurate results.²⁷

The most important two body term, employing an atom pairwise additive treatment, is shown below:

$$E_{Disp} = - \sum_{AB} \sum_{n=6,8,10,...} S_n \frac{C_n^{AB}}{R_{AB}^n} f_{damp}(R_{AB}) \quad (2. 7)$$

The first summation is over all atom pairs (AB) in the system. S_n is a global scaling parameter specific for the exact density functional used. It is tailored to rectify for the incorrect short range repulsive behaviour of the specific functional. C^{AB} refers to the n th-order dispersion coefficient for atom pair AB (orders $n = 6, 8, 10, ...$) while R_{AB} is the interatomic distance between atoms A and B. f_{damp} is a damping function that depends on the interatomic distance. A damping function is necessary in order to avoid near singularities (i.e. E_{Disp} abruptly becoming $\pm\infty$) for small R values as well as double-counting the effects of correlation at

intermediate distances. Damping functions typically have the form:

$$f_{damp}(R_{AB}) = \frac{1}{1 + 6(R_{AB} / (S_{r,n} R_0^{AB}))^{-\gamma}}$$

In Equation (2. 8) above R_0^{AB} is a cut-off radius for atom pair AB and γ is a global parameter that determines the steepness of the damping for low interatomic distances. The dispersion coefficients are calculated in an *ab initio* fashion by employing time-dependent DFT (TD-DFT). Furthermore, based on the concept of a fractional coordination number, the hybridization state of the atom is taken into account when calculating these dispersion coefficients. This effectively accounts for environmental effects, rather than always having the same dispersion coefficient for a given atom, regardless of its hybridization or oxidation state.

In the above equations E_{disp} approaches zero as $R_{AB} \rightarrow 0$. Theoretically, however, it can be shown that according to a convergent multipole expansion the dispersion energy should approach a constant finite value as the interatomic distance approaches 0.²⁸ The most recent advancement in DFT-D is to use a type of damping term developed by Becke and Johnson.²⁹ This is referred to as DFT-D with Becke and Johnson damping (DFT-D3BJ). DFT-D3BJ fulfils the requirement of the dispersion energy converging to a constant value as the united atom limit is approached i.e. as $R \rightarrow 0$. Apart from being theoretically sound, DFT-D3BJ is also currently the most accurate DFT-D implementation. In this case the dispersion energy is given by:

$$E_{Disp} = -\frac{1}{2} \sum_{A \neq B} S_6 \frac{C_6^{AB}}{R_{AB}^6 + [f(R_{AB}^0)]^6} + S_8 \frac{C_8^{AB}}{R_{AB}^8 + [f(R_{AB}^0)]^8} \quad (2. 9)$$

where:

$$f(R_{AB}^0) = a_1 R_{AB}^0 + a_2 \quad (2. 10)$$

and:

$$R_{AB}^0 = \sqrt{\frac{C_8^{AB}}{C_6^{AB}}} \quad (2. 11)$$

Here a_1 and a_2 are free fit parameters used to fine tune the damping. S_6 is set to 1 for GGA and hybrid functionals. In this case S_8 is used to correct the short range repulsive behaviour of the specific functional. A substantial increase in performance for non-covalent interactions and conformational energies arises as a consequence of using DFT-D3BJ. For the TPSS functional the mean absolute deviation (MAD), averaged over 9 reference sets (S22, S22+, PCONF, SCONF, ACONF, CCONF, ADIM6, RG6) decreases from 2.3 kcal.mol⁻¹ to 0.48 kcal.mol⁻¹ using the Def2-QZVP basis set.³⁰

ORCA implements both the non-local van der Waals density functional VV10³¹ as well as D3BJ dispersion correction for a variety of common functionals. Due to the advantages discussed above, D3BJ was implemented throughout this project whenever DFT, hybrid DFT and double hybrid DFT were implemented. This was done by invoking the keyword "D3BJ". For the TPSSh functional; the parameters for Equations (2. 9) and (2. 10) are as follows: $S_6 = 1.0000$; $a_1 = 0.0000$; $S_8 = 0.4243$ and $a_2 = 5.5253$.

2.2.3 Exchange-Correlation Integration Grids

A technical detail regarding DFT calculations is the grid used for numerical integration of the exchange-correlation potential. The complexity of the expressions for these quantities ensures that no analytical solutions are available and that the required integrals must be solved using approximate numerical methods. This is done through three-dimensional numerical integration over a grid in most modern quantum chemistry packages. If the RI-J approximation is used then it is normally this numerical integration of the XC potential that requires the most computing power in a DFT calculation. The larger the grid used, the less numerical noise and the more accurate the result. However, computational cost increases linearly with grid size.

The use of a finite grid unfortunately gives rise to the completely unphysical behaviour of rotational invariance. Thus the total energy of a molecule is not independent of a rotation of the coordinate system. In other words, the exact same molecule in two different orientations will give two (slightly) different energy values. This rotational invariance can in particular adversely affect calculated harmonic frequencies, especially low-lying vibrational modes.³² It is possible to construct the integration grid so large that for all practical purposes the dependence of energy on the rotational orientation of the molecule is negligible. However, this is not always practically feasible. Throughout this study we avoided calculating closely

related molecules, whose energies would later be compared, in grossly different orientations. This was done in order to minimise the impact of rotational invariance.

In general, the integration grid consists of a series of radial shells around each atom and an angular grid for each radial shell. The particular radial integration method used was the default Chebyshev–Gauss quadrature.³³ In ORCA the variable “IntAcc”, Integral Accuracy, determines the number of radial points, from which the number of radial shells for a particular atom is calculated as:

$$n_{rad}(A) = -5(3\log(10^{-IntAcc} - n_A + 8)) \quad (2.12)$$

where n_A is the row in the periodic table of atom A.

For the angular integration a Lebedev type³⁴ grid was used. A weighting scheme was applied in order to decompose three-dimensional space into atomic regions which are then treated separately. Although ORCA does feature the common Becke weighting scheme,³⁵ an alternative in which the weight of a given grid point is determined from a superposition of atomic exchange densities was employed.³⁶ This is the default setting in ORCA as it is more efficient than the Becke scheme, especially for larger systems. Additionally, ORCA employs an angular grid pruning algorithm in order to reduce the number of grid points necessary. This algorithm exploits the fact that regions of space near the nuclei of atoms require less angular grid points to achieve a given accuracy than regions of space considered to be bonding regions. In ORCA the keyword “Gridn” serves as a compound variable that determines the exact nature of the exchange-correlation integration grid. A series of default grids is available in the program. These range in increasing complexity from “Grid0” to “Grid7”. It is well known that an increase in efficiency can be achieved when during a DFT calculation the SCF iterations are performed with a smaller less accurate grid, while the final energies and gradients are performed with a larger grid once the SCF cycle has converged. This results in a negligible loss of accuracy as long as the first grid is not too small. ORCA features this multigrid option. In order to reduce the possibility of rotational invariance errors as well ensure accuracy the largest two default grids were used, namely Grid 6 and Grid 7 as the final grid, unless otherwise explicitly mentioned. Thus, Grid 6 was used for the SCF iterations while the final energy and gradient were evaluated using Grid 7. “IntAcc” as well as the specific Lebedev grid associated with these options are shown in Table 2-7.

Table 2-7: Final DFT XC Grid Details

	Angular Grid	IntAcc
Grid 7	Lebedev590	5.34
Grid 6	Lebedev770	5.67

2.2.5 Geometry Optimisations

When beginning the modelling of the various systems in this project it was necessary to initially use the GGA functional BP86 and the Def2-SVP basis set, in conjunction with the RI approximation. BP86 consists of the 1988 Becke exchange³⁷ and the 1986 Perdew³⁸ correlation functionals. During this stage the default XC integration grids were also used, namely Grid 2 and Grid 4 as the final grid. These are invoked if no grid keywords are specified in the input file. “IntAcc” as well as the specific Lebedev grid associated with these defaults are shown in Table 2-8.

Table 2-8 Preliminary DFT XC Grid Details

	Angular Grid	IntAcc
Grid 2	Lebedev110	4.34
Grid 4	Lebedev302	4.67

When the system became sufficiently large we also resorted to using the Def2-SV(P) basis set, in which there are no polarisation functions on hydrogen. This ensured that the exploration of potential energy surfaces (PES) could take place relatively quickly in order to develop an initial approximation for the structures of the various stable molecules and transition states before using these results as input for hybrid DFT. Additionally, at this stage, the default convergence criteria with respect to the structural optimisation were chosen (for the SCF criteria see section 2.1.3). These refer to the changes between two optimisation cycles and are defined in Table 2-9.

Table 2-9: Default Geometry Optimisation Convergence Criteria

Property	Threshold
Energy	$\leq 5.000 \times 10^{-6} \text{ Eh}$
Energy gradient	$\leq 3.000 \times 10^{-4} \text{ Eh/bohr}$
RMS gradient	$\leq 1.000 \times 10^{-4} \text{ Eh/bohr}$
Maximum displacement	$\leq 4.0000 \times 10^{-3} \text{ bohr}$
RMS displacement	$\leq 2.0000 \times 10^{-3} \text{ bohr}$

Due to the difference in hybrid and GGA timings, initially exploring the system at the GGA level, especially for larger systems, helped avoid excessive computational expense when much was still uncertain. Thus in summary, at the exploratory stage of the various sections in the project the following level of theory was implemented for geometry optimisations, given in ORCA input file format to aid reproducibility:

```
RI-BP86
Def2-SVP or Def2-SV(P) Def2-SVP/J
ECP{Def2-SVP=Pd, Def2-SVP/J}
D3BJ
Grid2 FinalGrid4
Opt
```

After an idea of the initial structures had been established, final geometry optimisations were then performed with TPSSh, again using Def2-SVP. The choice of Def2-SVP was a compromise between cost and efficiency. This ensured that the systems studied could be considered in their full breadth with the necessary accuracy, in the given time frame. It is known that molecular geometries are much less sensitive to the level of theory used than relative energies. As the relative energies would later all be computed by single-point calculations using higher level correlation methods and TZVP basis sets, def2-SVP was deemed sufficient for this purpose. For all final geometry optimisations, tight convergence criteria for the structural optimisation were chosen. Again, these refer to the changes between two optimisation cycles and are defined as:

Table 2-10: Tight Geometry Optimisation Convergence Criteria

Property	Threshold
Energy	$\leq 1.000 \times 10^{-6} \text{ Eh}$
Energy gradient	$\leq 1.000 \times 10^{-4} \text{ Eh/bohr}$
RMS gradient	$\leq 3.000 \times 10^{-5} \text{ Eh/bohr}$
Maximum displacement	$\leq 1.000 \times 10^{-3} \text{ bohr}$
RMS displacement	$\leq 6.000 \times 10^{-4} \text{ bohr}$

These were invoked via the “TightOpt” keyword. Once converged, an optimised molecular geometry is produced, and a single point energy is reported. In summary the following level of theory was used during final geometry optimisations, again in ORCA input file format:

```
RIJCOSX TPSSh
```

Def2-SVP Def2-SVP/J

ECP{Def2-SVP=Pd, Def2-SVP/J}

D3BJ

Grid6 FinalGrid7

TightOpt

2.2.6 Frequency Calculations

In order to account for thermal effects, i.e. conditions of finite temperature, as well as to confirm whether an optimisation resulted in a stationary point or transition state, frequency calculations were performed on all optimised geometries and transition state structures (TSs). Unfortunately, ORCA does not yet feature analytical frequencies for meta-GGA functionals. Numerical frequencies were thus implemented, which are very time consuming calculations. Numerical frequencies were invoked using the keyword “NumFreq”, with all other settings (i.e. SCF convergence, XC grid, etc.) identical to the parent geometry optimisations.

ORCA displaces each atom in both the positive and negative x, y and z directions generating the force between each atom and subsequently a vibrational spectrum. This is then scaled up using standard ideal gas statistical thermodynamics to ultimately produce a Gibbs free energy value. Thus it is assumed that all vibrations are strictly harmonic, that there are no orbitally degenerate electronic states, and that no electronic excited states are thermally accessible.

The inner energy (U), Equation (2. 13) is given by:

$$U = E(\text{el}) + E(\text{ZPE}) + E(\text{vib}) + E(\text{rot}) + E(\text{trans}) \quad (2. 13)$$

where $E(\text{el})$ is the “electronic energy” from the electronic structure calculation:

$$E(\text{el}) = E(\text{kin-el}) + E(\text{nuc-el}) + E(\text{el-el}) + E(\text{nuc-nuc}) \quad (2. 14)$$

In Equation (2. 14) $E(\text{kin-el})$ is the kinetic energy of the electrons, $E(\text{nuc-el})$ is the potential energy as a consequence of the attractive interaction of the nuclei and the electrons, $E(\text{el-el})$ is the potential energy as a consequence of the repulsion between the electrons and lastly $E(\text{nuc-nuc})$ is the potential energy due the repulsion between the electrons. $E(\text{el})$ is what would have been calculated at the end of a single point energy calculation or geometry optimisation, for example. However, the other terms in Equation (2. 13), are computed during a frequency calculation. These include the zero-point vibrational energy, $E(\text{ZPE})$, the finite temperature correction due to population of excited vibrational states, $E(\text{vib})$, as well as the

rotational, $E(\text{rot})$, and translational, $E(\text{trans})$, thermal energies. The enthalpy (H) is then computed as follows:

$$H = U + k_B \cdot T \quad (2.15)$$

In Equation (2.15) k_B is Boltzmann's constant and T is the temperature of the system. The entropy (S) is also calculated during a frequency calculation and is given by Equation (2.16)

$$S = S(\text{el}) + S(\text{vib}) + S(\text{rot}) + S(\text{trans}) \quad (2.16)$$

Here $S(\text{el})$, $S(\text{vib})$, $S(\text{rot})$ and $S(\text{trans})$ represents the electronic, vibrational, rotational and translational entropy respectively. Finally, the Gibbs free energy (G) is calculated as follows:

$$G = H - T \cdot S \quad (2.17)$$

To match the experimental results, the frequency calculations were performed at 373.15 K and 1 atm. For ground state structures, the vibrational spectrum was checked for any negative frequencies. If these appeared, attempts were made to remove them through modification of the structure and subsequent re-optimisation. In the case of TS optimisations, the structure was confirmed to be a TS by the presence of a single large negative frequency along the expected reaction coordinate. At this point it is useful to introduce constrained optimisations. With this feature it is possible to constrain the positions of certain atoms in the molecule while the rest are optimized. Constrained optimisations were used extensively to eliminate unwanted negative vibrational modes during the optimisation of both transition states and stable molecules. It is possible to constrain atoms, bonds, angles, dihedral angles, all bonds involving certain atoms and many other combinations in ORCA. In summary, the level of theory used in frequency calculations was:

```
RIJCOSX TPSSh
Def2-SVP Def2-SVP/J
ECP{Def2-SVP=Pd, Def2-SVP/J}
D3BJ
Grid6 FinalGrid7
NumFreq
%freq 373.15 end
```

2.2.7 Transition state optimisations

To obtain an accurate and reliable transition state structure, it is necessary to begin the geometry optimisation with a set of coordinates that are as close as possible to what the final

transition state is expected to be. Two methods were used to find approximate TS structures in this project. The first of these is the relaxed surface scan feature of ORCA. In a relaxed surface scan one particular structural variable is systematically varied while all others are relaxed, i.e. optimised. For example, a particular bond can be varied through a set of predefined values and the rest of the molecule optimised at each step, essentially generating a series of optimised structures in which the bond was lengthened or shortened. The highest energy structure in this sequence of optimisations is then used as an estimate for the TS. It is possible to stipulate in ORCA the starting value, end value as well as the number of steps the scan should take between these values. It is also possible for the scan variable to be adjusted non-sequentially. Although mostly used for mimicking reactions in which bonds are shortened or lengthened in this work, it is also possible to scan though bond angles and dihedral angles. Due to the large number of geometry optimisations in a surface scan it was necessary to perform them using BP86 with the Def2-SV(P) basis set and the default integration grids. As always, the RI-J approximation was invoked with D3BJ dispersion correction. In certain instances, when experimenting with large systems it was also necessary to specify loose structural optimisation criteria via the “looseopt” keyword, defined as:

Table 2-11: Loose Geometry Optimisation Convergence Criteria

Property	Threshold
Energy	$\leq 3.000 \times 10^{-5} \text{ Eh}$
Energy gradient	$\leq 2.000 \times 10^{-3} \text{ Eh/bohr}$
RMS gradient	$\leq 5.000 \times 10^{-4} \text{ Eh/bohr}$
Maximum displacement	$\leq 1.000 \times 10^{-2} \text{ bohr}$
RMS displacement	$\leq 7.000 \times 10^{-3} \text{ bohr}$

Thus the maximum energy difference between two structures before convergence is signalled is six times larger when using loose optimisation criteria compared to the tight alternative. This of course decreases the certainty of reaching a PES global minimum but also prevents the many slight structural modifications that are energetically insignificant and tend to occur toward the end of an optimisation. In summary the following level of theory was used for surface scans:

RI-BP86

Def2-SV(P) Def2-SVP/J

ECP{Def2-SVP=Pd, Def2-SVP/J}

D3BJ

Grid2 FinalGrid4

Although useful to estimate transition states for relatively simple chemical transformations, such as the formation or breaking of one bond, surface scans cannot readily mimic more complex reactions. For example, transition states for processes that involve simultaneous rotation of multiple parts of a molecule and subsequent concerted breaking and formation of bonds are unlikely to be found by scanning through a single structural variable. For such cases it was necessary to use the Nudged Elastic Band method (NEB).³⁹ In this method the algorithm is provided with the reactant and product molecules and attempts to calculate the Minimum Energy Path (MEP) between the two structures. Every atom in the reactant maps to every atom in the product. The highest energy structure in the MEP is then used as an estimate for the TS. The MEP is found by constructing a set of images between the reactant and product molecules. The number of images to construct is defined by the user. The more images used, the subtler the structural changes between images, and ultimately the better the estimate of the transition state structure. A spring interaction between adjacent images is added to ensure continuity and prevents any of images from simply being optimised into the reactant or product. ORCA does not feature a built in NEB algorithm. NEB calculations were thus performed with a program developed by Dr Christopher Perry.⁴⁰ The program implements the limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm (LBFGS). It interfaces with ORCA, using ORCA to perform the required energy and energy gradient calculations during the NEB. Due to the large number of optimizations involved, NEBs are extremely expensive calculations. The SCF cycle was thus limited to 100 iterations and the maximum number of MEP iterations to 400. In order to ensure an accurate as possible TS estimate in a reasonable time, the energy and energy gradient calculations were performed using RI BP86 with the Def2-SVP basis set and D3BJ dispersion correction.

As mentioned previously, the highest energy structure from either a surface scan or a NEB was then used as input for a TS optimisation in ORCA. ORCA applies a quasi-Newton like Hessian mode following algorithm in the TS optimiser.⁴¹ By varying the structure, this algorithm maximises the energy with respect to the lowest negative eigenmode of the Hessian, i.e. maximises the most negative imaginary vibration, while minimising the energy with respect to all other eigenmodes. Therefore, before a TS optimisation is started an estimate of the Hessian or an exact Hessian is required. After initially using a Hessian computed at the BP86 level it was discovered that a more exact Hessian was required in order to efficiently locate TSs of interest for the remainder of the project. Thus a TPSSh frequency calculation using the

default DFT integration grid was used to perform a “quick” frequency calculation on the TS estimate obtain from the surface scan or NEB:

```
RIJCOSX TPSSh
Def2-SVP Def2-SVP/J
ECP{Def2-SVP=Pd, Def2-SVP/J}
D3BJ
NumFreq
%freq 373.15 end
```

Once this essentially exact Hessian was generated, it was used together with the estimate TS geometry to perform the transition state optimization. It was found that using very tight SCF convergence criteria helped to ensure that only one negative frequency in the vibrational spectrum was present in the optimised TS. These criteria are shown in Table 2-12.

Table 2-12: Very Tight SCF Convergence Criteria

Property	Threshold
Total energy	$\leq 1.000 \times 10^{-9} \text{ Eh}$
One electron (i.e. orbital) energy	$\leq 1.000 \times 10^{-6} \text{ Eh}$
DIIS error	$\leq 5.000 \times 10^{-8}$

Thus all criteria are required to be 10 times smaller before convergence is signalled when compared to tight convergence criteria. They were invoked by the “VeryTightSCF” keyword. Apart from this the level of theory was exactly the same as for geometry optimisations, namely:

```
RIJCOSX TPSSh
Def2-SVP Def2-SVP/J
ECP{Def2-SVP=Pd, Def2-SVP/J}
D3BJ
Grid6 FinalGrid7
VeryTightSCF
OptTS
```

2.3 Double Hybrid DFT

In order to have confidence in the results of a quantum chemical calculation, dynamic electron correlation should be accounted for. As mentioned in section 1.2.2, accurate electron correlation methods do exist and ideally all calculations would be performed at the CCSD(T) level. However, such equations are currently impractical for all but the smallest systems and again a compromise must be found. Currently one of the most efficient ways to account for correlation is Double Hybrid DFT (DH-DFT). DH-DFT represents the 5th rung on Jacob's ladder of DFT functionals. Here virtual orbitals are introduced by means of MP2 perturbation theory. Not only is the exchange energy a hybrid of DFT and HF exchange (Hybrid DFT), but the correlation energy is now also a hybrid of DFT and MP2 correlation within the Kohn–Sham orbitals. A DH functional consists of two adjustable parameters and the exchange–correlation energy can be expressed as:

$$E_{xc}^{DH} = a_x E_x^{HF} + (1 - a_x) E_x[n] + (1 - a_c) E_c[n] + a_c E_c^{MP2} \quad (2.18)$$

In the above equation $E_x[n]$ and $E_c[n]$ are, as usual, the DFT exchange and correlation functionals that depend on the density n . E_x^{HF} is the HF exact exchange energy and E_c^{MP2} refers to the MP2 correlation energy carried out within the DFT orbitals. The parameters a_x and a_c adjust the percentage of HF and MP2 correlation respectively.

The latest variations of DH-DFT implement spin-component-scaled MP2 (SCS-MP2). SCS-MP2 allows for a dramatic increase in the accuracy of MP2 at no additional cost. It is based on a separate scaling of the correlation energy contributions from antiparallel/opposite spin (singlet) and parallel /same spin (triplet) pairs of electrons. This method was developed from a realisation that the MP2 correlation energy from the interaction of antiparallel spin (E_T) and parallel spin (E_S) electrons has a ratio (E_T/E_S) that is typically in the range 3 – 4. Thus, MP2 shows a systematic energy bias toward unpaired electrons while the contributions from spin-paired electrons are underestimated. To rectify this, two separate scaling factors are applied to E_S and E_T while keeping the total correlation energy roughly constant:

$$E_{SCS-MP2} = p_s E_s + p_T E_T \quad (2.19)$$

In the original SCS-MP2 work,⁴² it was proposed that $p_s = 1.2$ while $p_T = 0.33$. This relatively simply modification results in a reduction of the RMSD (Root Mean Squared Standard Deviation) from 9.5 kcal.mol⁻¹ for standard MP2 to 4.8 kcal.mol⁻¹ for SCS-MP2 when compared

to the W4-11 thermochemical benchmark.⁴³ The benchmarks used Weizmann-4 (W4) theory, which is essentially a layered extrapolation to the all-electron, relativistic CCSDTQ5 infinite basis set limit. Thus, in this limited data set SCS increases MP2 accuracy almost twofold, but introduces additionally two more adjustable parameters into Equation (2. 18). If it is assumed that a_c is factored into the optimised MP2 spin scaling factors (p_s and p_T), this gives:

$$E_{xc}^{DH} = a_x E_x^{HF} + (1 - a_x) E_x[n] + (1 - a_c) E_c[n] + p_s E_s^{MP2} + p_T E_T^{MP2} \quad (2. 20)$$

The inclusion of D3BJ dispersion correction (Section 2.2.2) has been shown to further improve the results of such functionals as long as the dispersion parameters are re-optimised, leading to what is referred to as Dispersion Corrected Spin Component Scaled Double hybrid DFT (DSD-DFT). The particular DH functional used in this work is DSD-PBEP86.⁴⁴ The underlying exchange functional is the PBE⁴⁵ from 1996 while the correlation functional is that of Perdew³⁸ from 1986. The exact parameters for DSD-PBEP86, as implemented in ORCA, corresponding to Equation (2. 20) are as follows: $a_x = 0.69$; $a_c = 0.56$; $p_s = 0.52$ and $p_T = 0.22$.

For the dispersion parameters in Equations (2. 9) and (2. 10) from Section 2.2.2, the requirements for DH functionals are somewhat different than their hybrid counter-parts. Formally, S_6 should be set equal to unity for any method that lacks the inherent ability to describe dispersion interactions. However, with double hybrids S_6 was optimised to complement the same spin fraction of the MP2 correlation energy i.e. p_T . It was discovered that these two parameters are strongly coupled during optimisation of the functional.⁴⁴ This is because correlation between electrons of the same spin is inherently a long range effect, and thus dispersion is partially taken into account by the MP2 calculation anyway, requiring $S_6 < 1$. It was found that the S_8 parameter, which accounts for medium range dispersion, was extremely small during the fitting procedure. This implies that the DFT and MP2 conglomerate effectively takes into account medium range dispersion effects and S_8 was set to zero. For similar reasons a_1 was also deemed sufficiently small to set to zero. Thus for DHs, D3BJ only requires two parameters. The values of these parameters for DSD-PBEP86 are $S_6 = 0.48$ and $a_2 = 5.6$.

It has previously been shown that correlating all electrons in the MP2 part of a DH calculation is equivalent to only calculating the correlation among the valence electrons.⁴⁶ This is largely a consequence of core correlation remaining constant throughout various transformations

and simply cancelling when relative energies are computed. Thus, throughout this work the frozen core approximation was made by invoking the “FrozenCore” ORCA option in the method block. By default, ORCA employs the “FC_ELECTRONS” method of determining the core electrons to exclude when calculating the correlation energy. When this option is employed the program simply counts the number of chemical core electrons as one would do when looking at the periodic table i.e. the electrons of the atom up to group 18 (Noble gases) of the previous period are considered core electrons. For example, chlorine which has an electronic configuration of $1s^2 2s^2 2p^6 3s^2 3p^5$ has a neon core and thus the $1s^2 2s^2 2p^6$ electrons are considered core electrons. An alternative option is to choose the core electrons by stipulating an energy window within which correlation must be calculated. All orbitals having an energy that falls within this window are considered valence orbitals. This is done in ORCA by invoking the “FC_EWIN” option along with “EWin EMin,EMax” where “Emin” and “EMax” represent the minimum and maximum energy of the orbitals in which correlation must be calculated, respectively, in units of Hartree. By default, orbitals with an energy below -6.0 Eh are considered core orbitals and those above 10.0 Eh are the virtual orbitals in which correlation is neglected i.e. “EWin -6.0, 10.0”. In certain instances, such as when DH calculations were performed on the K^+ cation (Chapter 5), it was not possible to use the “FC_ELECTRONS” method as this would result in an error as potassium, which has an electronic configuration of $Ar4s^1$, has no valence electrons once the single 4s electron is removed from the atom. In such instances the “FC_EWIN” option was invoked with the default energy window. It should be noted that when computing ΔE values it is not possible to mix an energy from a DH calculation performed using the “FC_ELECTRONS” method of determining the core electrons with that of a DH energy which made use of the “FC_EWIN” method.

DH-DFT has basis set requirements intermediate between DFT and correlation methods. Thus a much slower convergence of the energy with respect to basis set size is observed when compared to DFT. For this reason throughout this work DHs were run in conjunction with the Def2-TZVP basis set, which was described in Section 2.1.7. As before, a 28 electron ECP was applied to palladium.

Additionally, the RIJCOSX approximation was used for the Coulomb integrals as well the HF exchange in the Hybrid DFT component of the calculation. This thus required the Coulomb fitting auxiliary basis set, Def2-TZVP/J. The RI approximation was also used for the MP2 component of the calculation. In this case the RI algorithm was not approximating the Coulomb integrals involved in DFT but the MP2 correlation integrals. A different correlation fitting auxiliary basis set was thus required, namely Def2-TZVP/C.⁴⁷ This correlation fitting

auxiliary basis set contains approximately 3.3 times as many primitives as the orbital basis set. For the DFT XC integration grid, Grid 4 was used throughout the SCF iterations and the Final grid used was Grid 6. As always, tight SCF convergence criteria were enforced.

Throughout this work Gibbs free energies are reported. However, it is not practically feasible to perform a geometry optimisation at the DH level of theory, never-mind a frequency calculation. Thus, as is generally acceptable, the thermodynamic corrects were imported from the corresponding TPSSh/SVP calculation. Therefore, $E(\text{el})$ in Equation (2. 13) ,was taken from the DH single point energy calculation, while all other terms were imported from the corresponding TPSSh/SVP frequency calculation. Additionally the entire entropy, Equation (2. 16), was also imported from the TPSSh/SVP frequency calculation.

2.4 DLPNO-CCSD(T)

All methods based on DFT, including composite approaches such as DH-DFT, have the potential to provide varying degrees of accuracy depending on the chemical system under consideration. This is because the performance of the underlying XC functionals is not uniform. In contrast, wavefunction methods, ideally CCSD(T), should provide uniform accuracy regardless of the chemical system. CCSD(T) is the *de facto* gold standard wavefunction method as long as a single determinant wavefunction is appropriate i.e. as long as there is not a significant degree of static correlation in the system. It was thus necessary to benchmark DHs against CCSD(T) to confirm that these composite functionals are indeed reliable for the system at hand. However, CCSD(T) scales as N^7 , where as usual N is the number of basis functions. This method is thus unsuitable for conveniently benchmarking the palladium complexes considered in this work. There are continuing efforts by various groups to improve the efficiency of coupled cluster calculations, one such approach, called Domain-Based Local Pair Natural Orbital-CCSD(T) or DLPNO-CCSD(T),⁴⁸ is a linear scaling CCSD(T) approximation. Here the locality of electron correlation is exploited by making use of Pair Natural Orbitals (PNOs) rather than canonical delocalised orbitals. Approximate natural orbitals are constructed that are specific for each electron pair. The electron correlation for the whole system is then given by the sum of these electron pair correlation energies. This method has been shown to faithfully reproduce energy differences delivered by the canonical CCSD(T) method by on average 0.5–1 kcal.mol⁻¹, possibly in severe cases reaching 2 kcal.mol⁻¹. The efficiency of DLPNO-CCSD(T) has recently enabled a CCSD(T) level calculation to be performed, for the first time, on an entire protein. The seed storage protein Crambin, which contains over 650 atoms,

was treated at the DLPNO-CCSD(T)/SVP level.⁴⁸ Currently only closed shell variants are available in ORCA, making DLPNO-CCSD(T) ideal to efficiently benchmark the performance of DHs for the system under consideration. Convergence with respect to basis set size of DLPNO-CCSD(T) is practically identical to CCSD(T). This is also true for basis set superposition errors (BSSE). As these results would ultimately be compared to DHs the same basis set was used, namely Def2-TZVP. This method has been previously used to tackle a transition metal catalysis problem very similar to the focus of this project, namely a change in a reactions enantioselectivity at a metal centre in homogenous catalysis as a consequence of different ligands.⁴⁹ It was found that the computational prediction was in agreement with the experimental results, reinforcing the suitability of DLPNO-CCSD(T) as a benchmark.

As this is a wavefunction method no dispersion correction or DFT integration grid is necessary. Tight SCF optimisation criteria were enforced. The RIJCOSX approximation was implemented. Thus, in a similar manner to DHs three basis sets were required, both an auxiliary Coulomb fitting (Def2-TZVP/J) and an auxiliary correlation fitting basis set (Def2-TZVP/C). In this way the RI algorithm was used to approximate both the Coulomb integrals and the DLPNO-CCSD(T) correlation integrals while the COSX algorithm dealt with HF exchange.

As before Gibbs free energies are reported, with the thermodynamic parameters being imported from the corresponding TPSSh/SVP calculation.

In a similar fashion to DH calculations the frozen core approximation was made. This is actually default ORCA behaviour with DLPNO-CCSD(T) calculations. Again, in situations where the default "FC_ELECTRONS" method of determining the core was unsuitable, the "FC_EWIN" option was invoked with the default energy window.

In conclusion, to illustrate why DLPNO-CCSD(T) was only used as a benchmark rather than routinely throughout the project, a typical DH job run on one node (12 processors) at the CHPC took 8 minutes, while the equivalent DLPNO-CCSD(T) calculation took 5 hours.

2.5 COSMO Solvation Model

All calculations reported thus far are based on the approximation that the system is in the gas phase. Solvents effects are however immensely important in chemistry. It was thus necessary to at least apply an implicit solvent model to determine if the conclusions from gas and solution phase calculations differ qualitatively.

Initially the SMD solvation model of Cramer, Marenich and Truhlar⁵⁰ was chosen. However, this strangely resulted in persistent negative frequency methyl group rotations when used in combination with geometry optimisations and frequency calculations. It thus seems that there are compatibility issues with this method and the ORCA 3.0.2 structural optimizer i.e. the optimizer consistently gets stuck in saddle points on the PES.

We turned to the Conductor-like Screening Model (COSMO)⁵¹ solvation method in order to overcome this. COSMO is fully incorporated into all modules of ORCA, including analytical gradients. Thus full geometry optimisations (including TS optimisation) and frequency calculations were performed using COSMO on certain subsystems in this work. All details of the calculations were identical to their gas phase DFT counterparts, apart from invoking the COSMO module via the “COSMO” keyword. All that is necessary is to specify the dielectric constant and refractive index (actually only used in excited state calculations) of the particular solvent. However ORCA contains a library of 179 solvents including DMA (N,N-dimethylacetamide). Thus it was only necessary to specify “N,N-DIMETHYLACETAMIDE” in the COSMO block of the input file. For the sake of reproducibility with other programs the values of these properties for DMA are $\epsilon = 37.7810$ and refractive index = 1.3300.

The size and shape of the solute cavity (mentioned in section 1.2.3) is determined by overlapping spheres with a radii approximately 1.2 times the Van der Waals radii of the corresponding atoms. These values have been carefully optimised by the COSMO developers and are listed in appendix A for all atoms used in this work.

References

1. F. Neese, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2012, 2, 73-78.
2. P. Pulay, *Chemical Physics Letters*, 1980, 73, 393-398.
3. P. Pulay, *Journal of Computational Chemistry*, 1982, 3, 556-560.
4. V. R. Saunders and I. H. Hillier, *International Journal of Quantum Chemistry*, 1973, 7, 699-705.
5. F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chemical Physics*, 2009, 356, 98-109.
6. F. Weigend and R. Ahlrichs, *Physical Chemistry Chemical Physics*, 2005, 7, 3297-3305.
7. F. Weigend, *Physical Chemistry Chemical Physics*, 2006, 8, 1057-1065.
8. E. Papajak and D. G. Truhlar, *Journal of chemical theory and computation*, 2010, 6, 597-601.
9. J. Zheng, X. Xu and D. G. Truhlar, *Theoretical Chemistry Accounts*, 2011, 128, 295-305.
10. G. A. Andrienko., <http://www.chemcraftprog.com>.
11. M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *Journal of Cheminformatics*, 2012, 4, 1-17.
12. A. D. Becke, *The Journal of Chemical Physics*, 1993, 98, 5648-5652.
13. C. Lee, W. Yang and R. G. Parr, *Physical review B*, 1988, 37, 785.
14. J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Physical Review Letters*, 2003, 91, 146401.
15. J. P. Perdew, J. Tao, V. N. Staroverov and G. E. Scuseria, *The Journal of Chemical Physics*, 2004, 120, 6898-6911.
16. L. A. Curtiss, K. Raghavachari, P. C. Redfern and J. A. Pople, *The Journal of Chemical Physics*, 2000, 112, 7374-7383.
17. F. Furche and J. P. Perdew, *The Journal of Chemical Physics*, 2006, 124, 044103.
18. K. P. Jensen, *The Journal of Physical Chemistry A*, 2009, 113, 10133-10141.
19. K. P. Jensen, *Inorganic Chemistry*, 2008, 47, 10357-10365.
20. K. Lee, É. D. Murray, L. Kong, B. I. Lundqvist and D. C. Langreth, *Physical Review B*, 2010, 82, 081101.
21. O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger and D. Sebastiani, *Physical Review Letters*, 2004, 93, 153004.
22. Y. Sun, Y.-H. Kim, K. Lee and S. Zhang, *The Journal of Chemical Physics*, 2008, 129, 154102.
23. S. Grimme, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2011, 1, 211-228.
24. K. E. Yousaf and E. N. Brothers, *Journal of Chemical Theory and Computation*, 2010, 6, 864-872.
25. L. Goerigk, H. Kruse and S. Grimme, *ChemPhysChem*, 2011, 12, 3421-3433.
26. L. Duy, K. Abdelkader, S. Elsebeth, H. Per and S. R. Talat, *Journal of Physics: Condensed Matter*, 2012, 24, 424210.
27. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *The Journal of Chemical Physics*, 2010, 132, 154104.
28. A. Koide, *Journal of Physics B: Atomic and Molecular Physics*, 1976, 9, 3173.
29. E. R. Johnson and A. D. Becke, *The Journal of Chemical Physics*, 2006, 124, 174104-174104.
30. S. Grimme, S. Ehrlich and L. Goerigk, *Journal of Computational Chemistry*, 2011, 32, 1456-1465.
31. O. A. Vydrov and T. Van Voorhis, *The Journal of Chemical Physics*, 2010, 133, 244103.
32. B. G. Johnson and M. J. Frisch, *Chemical physics letters*, 1993, 216, 133-140.
33. M. Krack and A. M. Köster, *The Journal of Chemical Physics*, 1998, 108, 3226-3234.
34. I. Lebedev and D. N. Laikov, *Doklady Mathematics*, 1999, 59, 5.
35. A. D. Becke, *The Journal of Chemical Physics*, 1988, 88, 2547-2553.
36. R. E. Stratmann, G. E. Scuseria and M. J. Frisch, *Chemical Physics Letters*, 1996, 257, 213-223.
37. A. D. Becke, *Physical Review A*, 1988, 38, 3098-3100.
38. J. P. Perdew, *Physical Review B*, 1986, 33, 8822-8824.
39. G. Henkelman, B. P. Uberuaga and H. Jónsson, *The Journal of Chemical Physics*, 2000, 113, 9901-9904.
40. C. Perry

Senior Scientist
Johnson Matthey Technology Centre
S. A. Pretoria

41. J. Baker, *Journal of Computational Chemistry*, 1986, 7, 385-395.
42. S. Grimme, *The Journal of Chemical Physics*, 2003, 118, 9095-9102.
43. A. Karton, S. Daon and J. M. L. Martin, *Chemical Physics Letters*, 2011, 510, 165-178.
44. S. Kozuch and J. M. L. Martin, *Journal of Computational Chemistry*, 2013, 34, 2327-2344.
45. J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, 1996, 77, 3865-3868.
46. S. Kozuch, D. Gruzman and J. M. Martin, *The Journal of Physical Chemistry C*, 2010, 114, 20801-20808.
47. A. Hellweg, C. Hättig, S. Höfener and W. Klopper, *Theoretical Chemistry Accounts*, 2007, 117, 587-597.
48. C. Riplinger, B. Sandhoefer, A. Hansen and F. Neese, *The Journal of Chemical Physics*, 2013, 139, 134101.
49. A. Anoop, W. Thiel and F. Neese, *Journal of Chemical Theory and Computation*, 2010, 6, 3137-3144.
50. A. V. Marenich, C. J. Cramer and D. G. Truhlar, *The Journal of Physical Chemistry B*, 2009, 113, 6378-6396.
51. A. Klamt and G. Schüürmann, *Journal of the Chemical Society, Perkin Transactions 2*, 1993, 799-805.

3. The CMD

Mechanism: Ligand-less

Conditions

Preface to Chapter 3:

Throughout this chapter, reference will be made to optimised structures of theoretical molecules. As chemistry is inherently a three-dimensional subject, representation of molecules via 2D images in a typed document, although necessary, inherently results in a loss of detail and perhaps clarity. It is for this reason that the coordinates of all optimised structures referred to in this work, as well as all relevant output files, are provided in the electronic supplementary information (ESI). Coordinates are provided in the simple xyz file format that can be read by any computational chemistry visualisation package. This document has been constructed in such a way that by simply clicking the caption of a figure in the text, the relevant location in the ESI will open automatically in Windows explorer, allowing the reader effortless access to all original output files pertaining to the molecule. This functionality however, will be lost if the reader changes the relative location of files inside the “Thesis” folder. The folder itself can be download and moved anywhere on the readers’ PC but the contents should be kept as is and files not moved around within the folder or to other locations. It has been assumed that the examiner will be using a Windows operating system. However, if the examiner is using a Linux or Macintosh operating system a version of this document with compatible paths can be arranged timeously.

This is in the hope that the reader will open these files themselves, examine the various molecules and visualise the transition state vibrational frequency animations so that the contents of this chapter may become more vivid and the arguments presented hopefully more logical.

Contents

3. The CMD Mechanism: Ligand-less Conditions	83
3.1 Introduction	87
3.2 Mono-substitution	90
The Active catalyst:	90
Possibilities for Initial Coordination of Oxazole (Intermediate 1).....	94
Intermediate 1 N2:	99
Intermediate 1 N4:	105
Intermediate 1 =4:.....	109
Intermediate 1 =5:.....	114
Intermediate 1 N=2:	116
Intermediate 1 O2 and Intermediate 1 O5:	118
The Reaction Pathways:	121
An overview of the reductive elimination step:.....	121
Pathway N2:	123
An overview of the CMD step:	130
Validation of the Model System:.....	132
Pathway N4:	137
Pathway =4:.....	140
Pathway =5:.....	143
TSSSh/SVP Results:	146
COSMO Results:.....	153
DLPNO-CCSD(T) Results:.....	160
DSD-PBEP86 Results:	166
3.3 Di-substitution	172
The Reaction Pathways:	174
Pathway 2 → 2,4:	174

Pathway 4 → 2,4:	178
Pathway 2 → 2,5:	183
Pathway 5 → 2,5:	188
Pathway 4 → 4,5:	192
Pathway 5 → 4,5:	197
DSD-PBEP86 Results:	201
References	210

3.1 Introduction

The first mechanism used to model the regio-selective determining step in the direct arylation of oxazole is the concerted metallation deprotonation (CMD) mechanism, the process depicted in blue in Figure 3-1. As mentioned previously, Section 1.3.2, this is the most widely accepted mechanism for direct arylation under the standard conditions in which the catalytic system comprises of palladium, phosphine ligand and a carbonate base. The CMD pathway can theoretically take place in the presence or absence of phosphine ligand. It is thus an ideal place to start a theoretical investigation of the experimental results, reproduced again for convenience in Table 3-1.

Table 3-1: Experimental Results for the Direct Arylation of Oxazole with Toluene – Various Mole Equivalents of P^tBu₃ Ligand

Entry	Catalyst	Ligand	Product Distributions (%)			
			Product 5	Product 2	Product 2,5	Product 4,5
1	Pd(OAc) ₂	-	65.8	0.3	28.1	3.0
2	Pd(OAc) ₂	1 x P ^t Bu ₃	30.6	3.7	59.3	0.9
3	Pd(OAc) ₂	2 x P ^t Bu ₃	6.3	35.3	49.7	0.1

This chapter attempts to elucidate the mechanism of the reactions that correspond to entry 1 in Table 3-1, .i.e. direct arylation of oxazole under ligand free conditions. Numerous examples in the literature of phosphine free direct arylation reactions have been reported.¹⁻³ These include examples in which a change in regioselectivity is observed when moving from phosphine to phosphine-free conditions concerning molecules somewhat similar to oxazole.⁴ However, the most relevant work to the system under consideration is that performed by Hartwig et al.⁵ Here experiments were performed under conditions strikingly similar to those at hand. Hartwig et al.⁵ reacted 4-bromotoluene with benzene using a Pd(OAc)₂ catalyst in DMA solvent with pivalic acid and potassium carbonate as an additive at 110 °C. There are only minor differences between these reactions and those that concern this work, namely the presence of the additional additive K₂CO₃, benzene rather than oxazole as the arylation partner, as well as the additional 10 °C reaction temperature. Their work concluded with the proposal that in the absence of phosphine ligand, a DMA solvent molecule is coordinated to the palladium complex that acts as the active catalyst. Additionally, this DMA coordinated species was optimised at the B3LYP/TZVP level of theory and subsequently a CMD transition state (TS) determined. This pathway was lower in energy than the corresponding structures in which a phosphine ligand was bound to palladium. In this chapter we built upon the work of

Hartwig et al.⁵ and modelled the phosphine free reactions with a DMA solvent molecule coordinated to the palladium centre, i.e. L in Figure 3-1 was a DMA molecule.

Chapter 4 involves modelling the CMD mechanism in the presence of Tri-tert-butylphosphine (P^tBu_3) ligand. These could possibly be reactions which correspond to entry 2 and entry 3 in Table 3-1. In this case, as has generally been observed throughout the literature⁶⁻⁸, a P^tBu_3 ligand molecule was modelled bound to the active catalyst rather than a DMA solvent molecule, i.e. L in Figure 3-1 was a P^tBu_3 molecule.

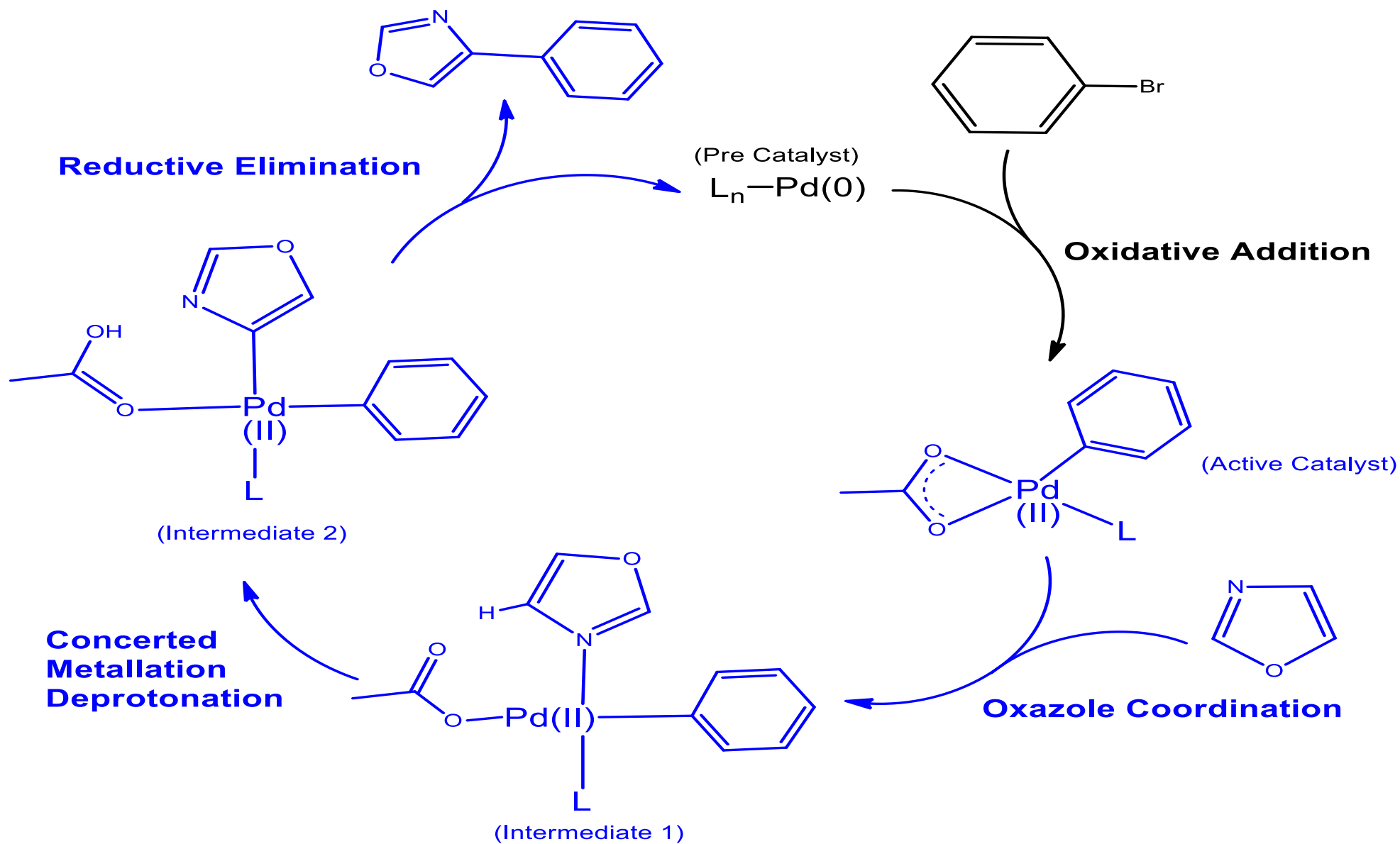


Figure 3-1: Arylation of Oxazole with Benzene at Position 2 via the CMD Mechanism and Subsequent Reductive Elimination

3.2 Mono-substitution

The Active catalyst:

To begin the modelling of the CMD mechanism, under ligand free conditions, we started with the proposed active catalytic Pd(II) species as reported by Hartwig et al.⁵ The formation of this species will be considered in detail in Chapter 5, where an investigation of oxidative addition is reported. For now it is suffice to say that it is generally accepted that palladium acetate, Pd(OAc)₂, serves as a precursor to Pd(0) which is then most likely ligated by a pivalate anion, due to the high concentration of this additive species. This anionic species then undergoes OA. After OA the bromine atom is displaced and a DMA solvent molecule coordinates to palladium (Figure 3-2).

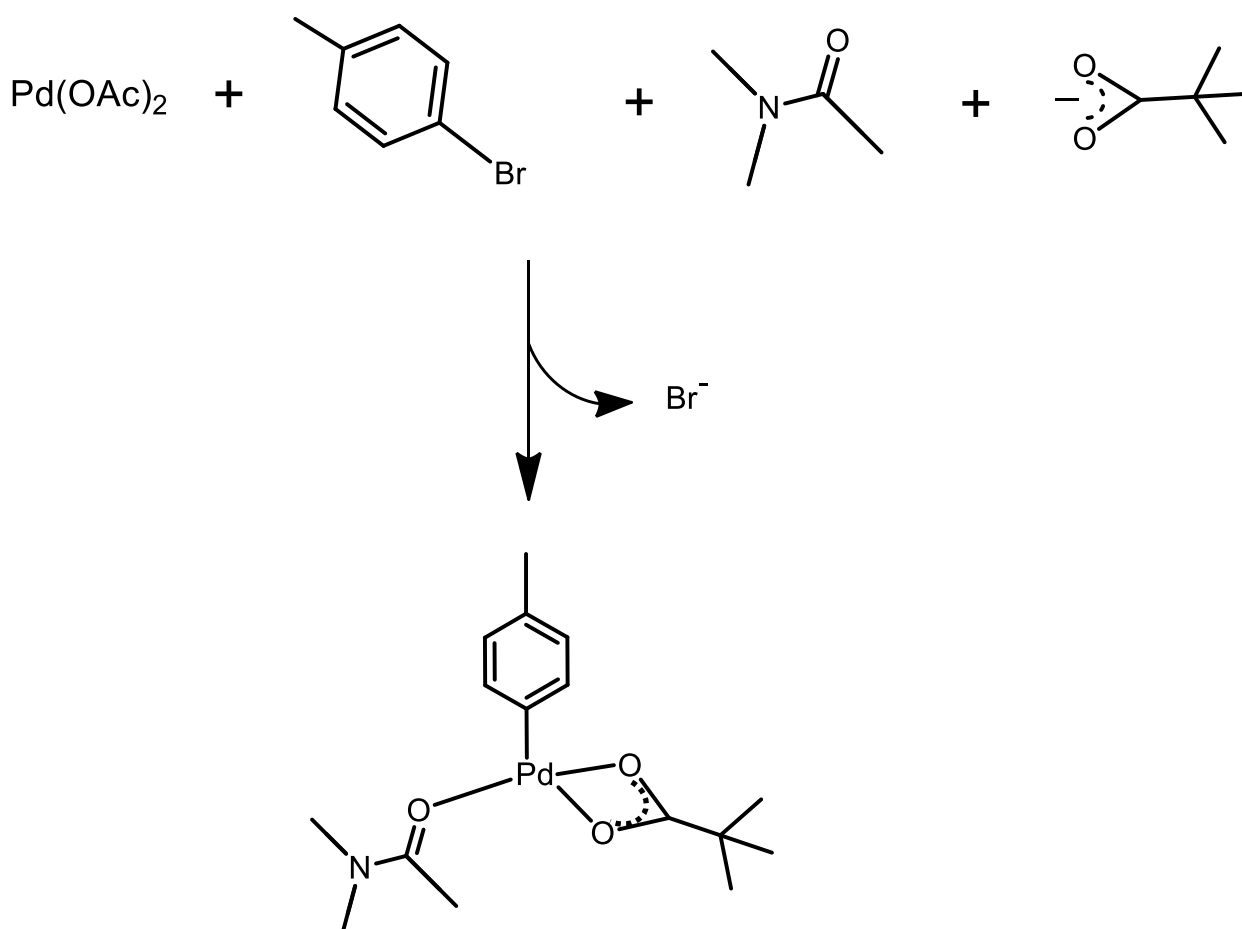


Figure 3-2: Overview of the Formation of the Pd(II) Active Catalyst – Ligand Free Conditions

In order to reduce the computational cost of the study the methyl group of toluene and the three methyl groups of the pivalic acid ligand were replaced with hydrogen atoms. In this way, the model system could be reduced by 12 atoms without significantly affecting relative energy

values or the ultimate conclusions. This assumption was later justified, when the relative energies and conclusions drawn from the true and model systems were compared. The model active catalyst, from henceforth simply referred to as the active catalyst, is shown in Figure 3-3.

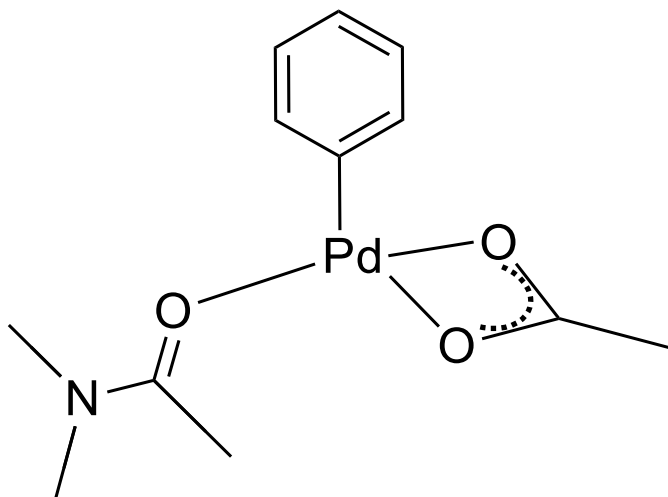


Figure 3-3 : The Active Pd(II) Catalyst – Phosphine Free Conditions

Initially as an estimate of the active catalyst we turned to the optimised structure (B3LYP/TZVP) reported by Hartwig et al⁵ (Figure 3-4) . This optimisation itself was based on a modified crystal structure.

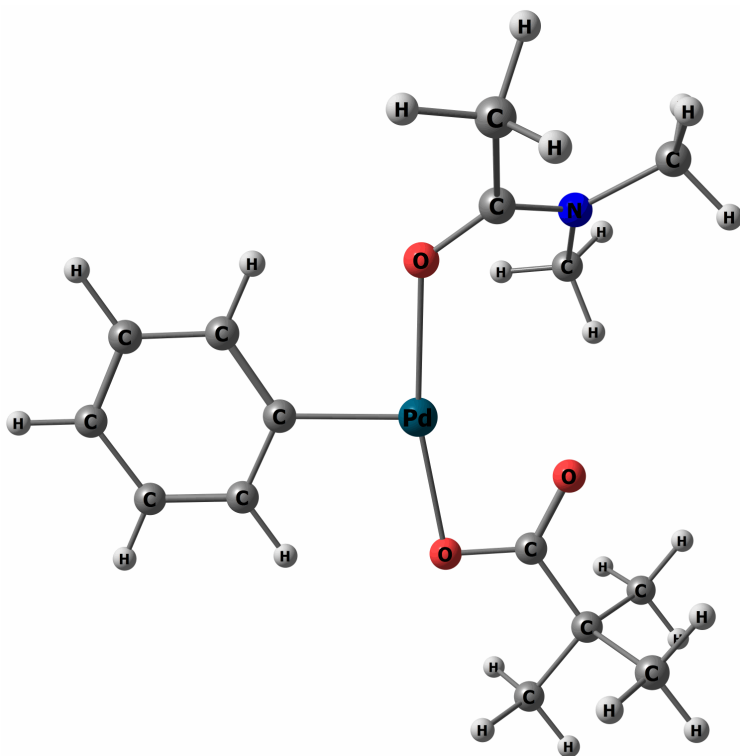


Figure 3-4: Active catalyst from the literature reported by Hartwig et al. (B3LYP/TZVP)

Here the system already had benzene coordinated to Pd, so there was no need to replace the methyl group of toluene with hydrogen. The three methyl groups of pivalic acid were however, replaced with hydrogens and this new structure optimised. This optimised structure was used initially, however during the modelling of the formation of the active catalyst in Chapter 5, a lower energy structure was revealed. Thus, the previous structure was a local minimum on the potential energy surface (PES). This is likely a consequence of the fact that the initially employed catalyst was based on a solid-state crystal structure. In general, the solid-state crystal structure and the lowest energy gas-phase structure of a molecule do not necessarily correspond. Differences arise due to crystal packing forces, which are neglected by gas phase calculations, where a vacuum is assumed. Subsequently all relative energy values were changed and the lower energy version (Figure 3-5) used throughout the project.

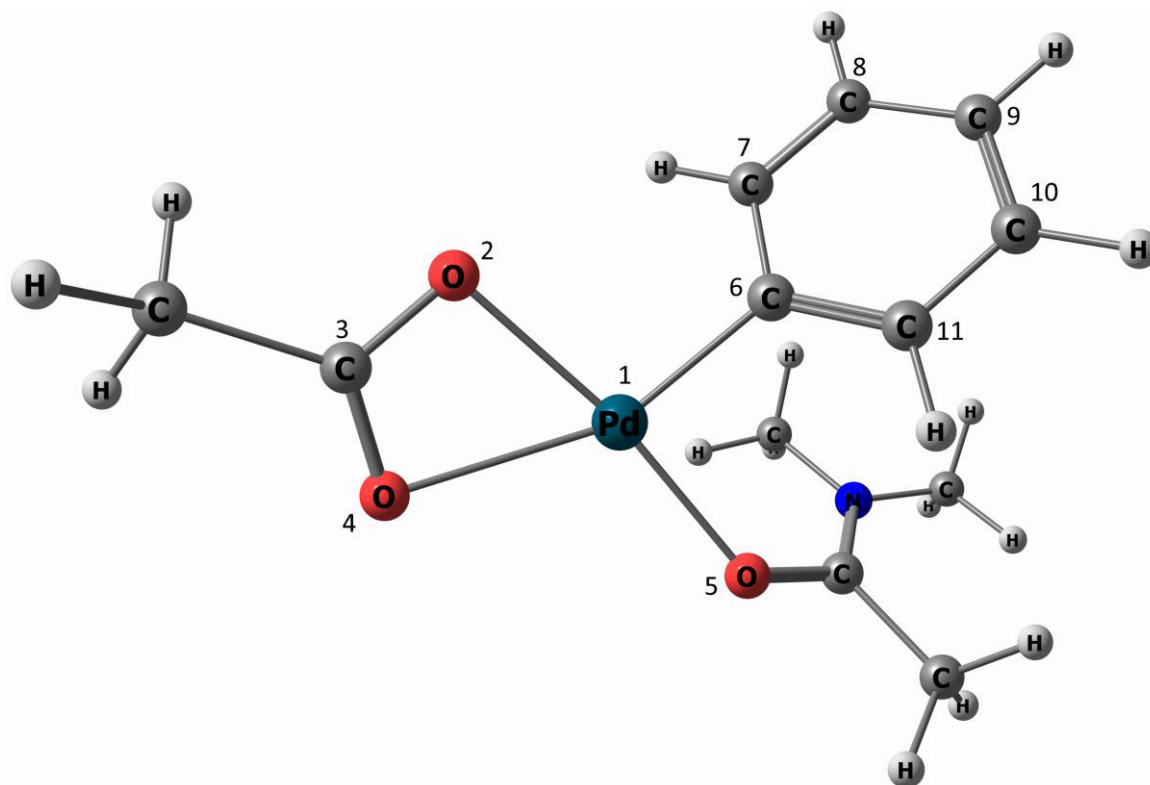


Figure 3-5: Optimised Active Pd(II) Catalyst - Phosphine Free Conditions

Regardless of the fact that the crystal and lowest energy gas-phase structures of a molecule do not necessarily correspond exactly, crystallographic data is commonly used to assess the accuracy of computational methods.

Analysing various bond lengths, with respect to crystallographic data, confirms the reliability of the chosen model and level of theory employed. For example, the distance between the acetate carboxylate carbon and palladium, $D(1-3)$, is 2.521 Å. In this notation, the numbers in brackets refer to the atom numbers in Figure 3-5 and D refers to a distance. A Cambridge Structural Database⁹ (CSD) search of acetate coordinated via both oxygen atoms to palladium generated 13 hits, with an average $D(1-3)$ of 2.495 Å. The O–C–O angle of the acetate group, $\angle(2-3-4)$, is reproduced within 0.68° of the experimental average in the same set of samples. In this notation, $\angle$ refers to an angle and the numbers in brackets to the atom numbers in Figure 3-5. The Pd–C bond length of the benzene group, $D(1-6)$, is also fairly well reproduced when considering 8 experimental crystal structures which contain an organometallic Pd–C benzene bond as well as the additional criterion of acetate coordination (within 0.031 Å). Within the benzene ring itself the average carbon–carbon bond length of the delocalised π system is 1.401 Å, compared to an average 1.387 Å over the 8 experimental samples. For the remaining ligand, DMA, an inquiry of the CSD generated two hits in which DMA is coordinated

to Palladium. In both of these the molecule is coordinated to the metal via the carbonyl oxygen, just as in the optimised structure. The average experimental DMA Pd–O bond length, D(1-5), is 2.080 Å with the two values being 2.004 and 2.155 Å. D(1-5) in the model catalyst is 2.063 Å. The various model vs experimental structural data are summarised in Table 3-2.

Table 3-2: Structural Analysis of Active Catalyst (Refer to Figure 3-5 for atom numbering)

Parameter	Model (Å/°)	Mean Experiment (Å/°)	Difference (Å /°)	Max -Min Experiment (Å /°)	Sample Size
D(1-3)	2.521	2.495	0.026	2.434-2.543	14
D(1-6)	1.951	1.982	-0.031	1.955 - 2.005	8
D(1-5)	2.070	2.080	-0.010	2.004 - 2.155	2
Average C–C Benzene	1.401	1.387	0.014	1.349 - 1.419	8
		Mean absolute bond difference (Å):	0.020		
∠(2-3-4)	119.71	119.03	0.68	117.300 -121.400	14

Possibilities for Initial Coordination of Oxazole (Intermediate 1)

For the reaction to proceed oxazole must coordinate to the palladium atom in the catalyst and subsequently be deprotonated. Coordination of oxazole results in the first, as well as regio-selective determining, intermediate in the reaction pathway, Intermediate 1, abbreviated where appropriate as Int 1 (Figure 3-6).

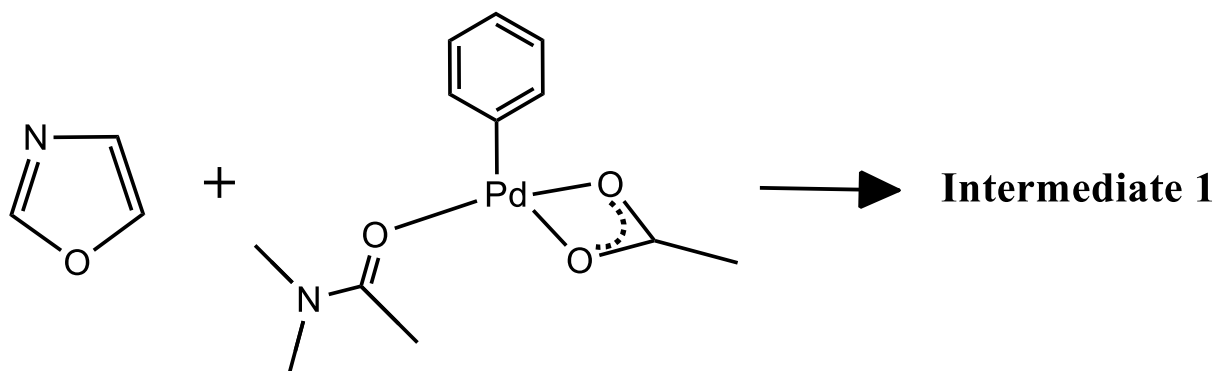


Figure 3-6: Overview of the Formation of Intermediate 1

Regarding the coordination of oxazole, a number of possibilities exist. Chemical intuition suggests that coordination through the nitrogen atom, the double bond between carbons 4 and 5 (Figure 3-7) as well as the double bond between nitrogen and carbon 2 are all possible. Coordination via oxygen is another prospect.

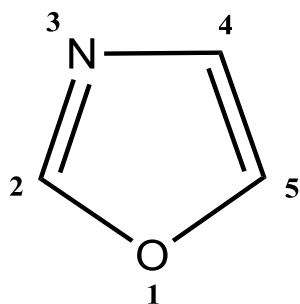


Figure 3-7: Standard Numbering Scheme for Oxazole

Traditional organic chemistry arguments would suggest that both the nitrogen and oxygen atoms in oxazole are sp^2 hybridised. Each of the three sp^2 orbitals would be trigonal planar with the remaining un-hybridized p orbital perpendicular to the plane of the molecule. Thus, the lone pair of the nitrogen atom occupies a sp^2 orbital and via a bonding p orbital nitrogen contributes one electron to the aromatic π system. One of the oxygen lone pairs also occupies a sp^2 orbital while the second lone pair occupies a p orbital. Oxygen thus contributes two electrons to the aromatic π system. Each carbon atom, via a p orbital, also contributes 1 π electron. Therefore, oxazole is an aromatic molecule according to Hückel's rule as it contains $(4n + 2)$ π electrons, with $n = 1$ i.e. 6 π electrons. The lone pair of oxygen in the p orbital contributes to the aromatic π system and therefore it is not available for coordination. The nitrogen atom is expected to be more basic than its oxygen counterpart. This is due to the fact that both available lone pairs on N and O occupy sp^2 orbitals, but the larger nuclear charge of the more electronegative oxygen atom would ensure that this lone pair is less available for coordination.

This is manifested in a variety of crystal structures of Pd(II) complexes in which oxazole and related analogues are coordinated via the nitrogen atom, when oxygen coordination is seemingly also possible. For example the crystal structure of trans-dichloro-bis(oxazole)-palladium(II) is shown in Figure 3-8.¹⁰

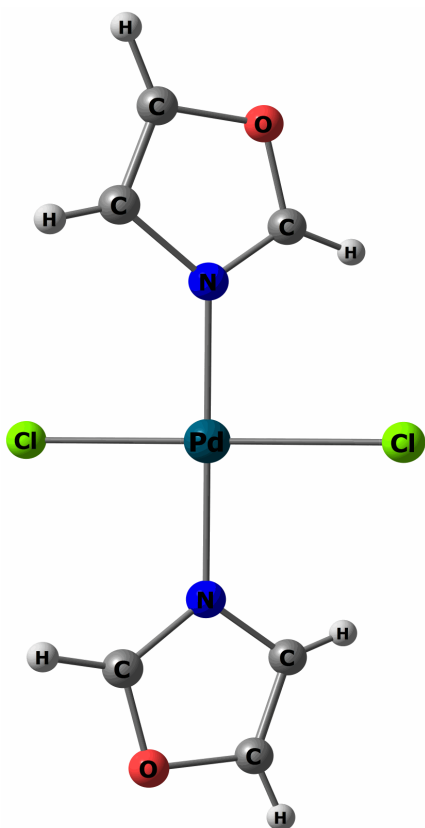
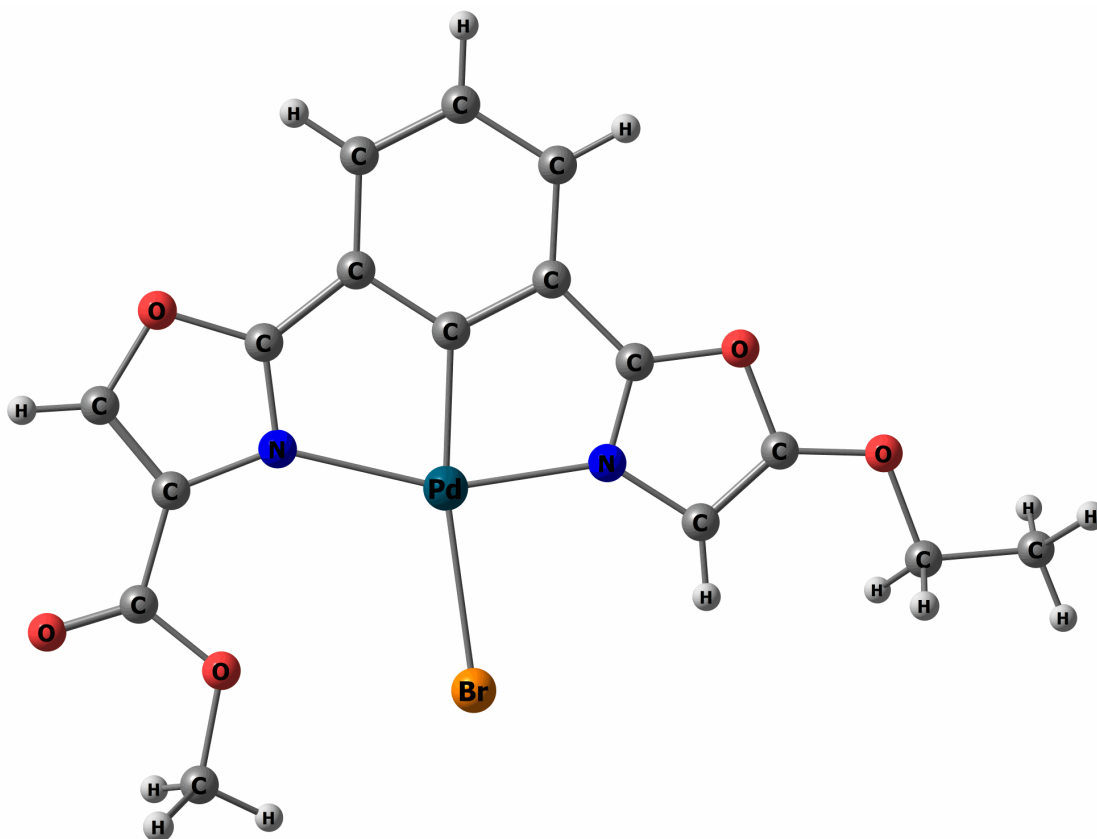
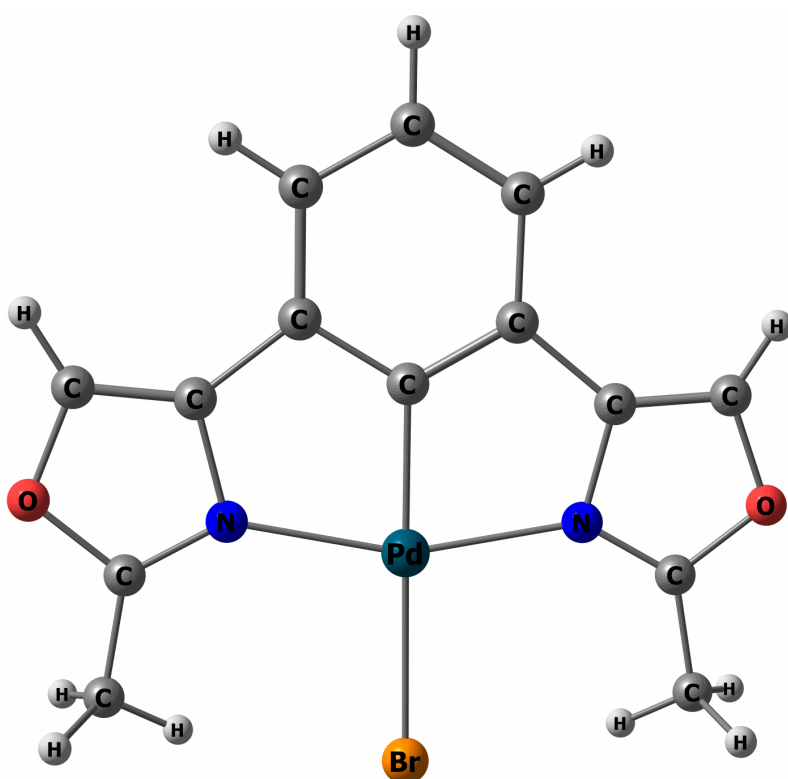


Figure 3-8 : Crystal structure of trans-dichloro-bis(oxazole)-palladium(II)
Showing Preferred Coordination via N

Both oxazole molecules are coordinated to palladium via the nitrogen atom, showing a preference over oxygen as well as any of the double bonds. Additional examples include the crystal structure of the palladium complex shown in Figure 3-9.¹¹ Here both oxazole moieties are also coordinated via the nitrogen atom, even though nitrogen coordination of the substituted oxazole on the left of the complex results in an unfavourable steric interaction between the methyl ester substituent and the bromine ligand. Numerous other crystal structures containing both palladium and other metals reinforce the favourability of nitrogen coordination, some of which are shown in Figure 3-10¹² and Figure 3-11.¹³



[Figure 3-9 : Crystal structure of arylated oxazole palladium complex](#)



[Figure 3-10 : Crystal structure of \(2,6-bis\(2-methyl-1,3-oxazol-4-yl\)phenyl\)-bromo-palladium](#)

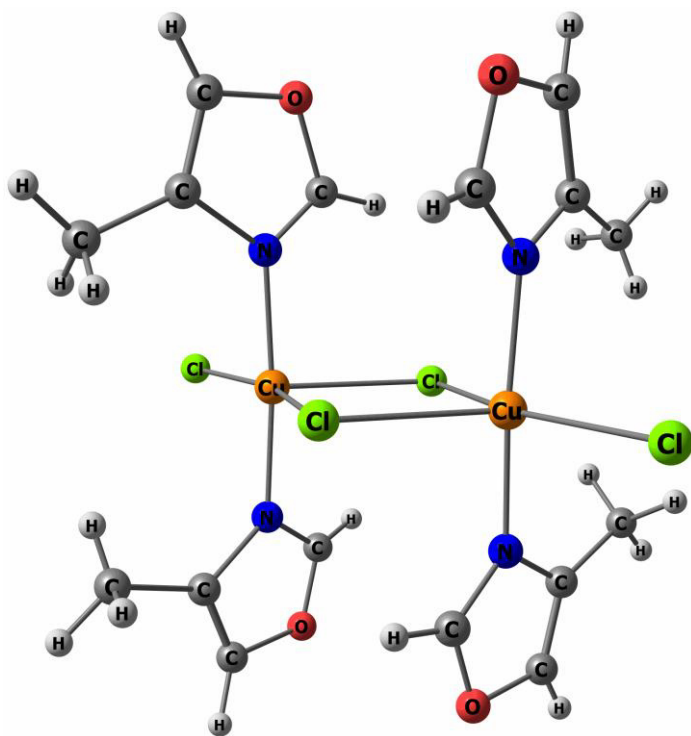


Figure 3-11: Crystal structure of bis(μ_2 -Chloro)-bis(chloro-bis(4-methyloxazole)-copper(II))

For an extensive review of the coordination of oxazole like compounds see reference 14. Both this review and the CSD contain no structures in which the oxygen atom of oxazole is coordinated to a metal centre. Furthermore, they contain no structures in which either of the double bonds of oxazole is coordinated to a metal atom. Although a high quality crystal structure is definite proof of the existence of a particular arrangement of atoms in the solid state, this is not always indicative of the situation in solution. Double/aromatic bond coordination to a metal centre is ubiquitous in organometallic chemistry. As of writing, the 2016 implementation of the CSD contains over 78 000 entries of such molecules, over 1900 of which contain palladium. Thus coordination to palladium through one of the double bonds of oxazole cannot justifiably be excluded if a thorough modelling of the system is to be undertaken. In fact, all of the above mentioned possibilities (i.e. N, O and double bond coordination) were modelled in order to ensure that the project was initiated with a rigorous conformational analysis and to illustrate how calculations can explain experimental findings, verifying their accuracy.

The structure of Intermediate 1 is essential because it dictates the product that will ultimately be formed. There are a number of unique possibilities for the structure of this species, depending both on the mode of coordination as well as the orientation of oxazole. Each of

these possibilities leads to the formation of one of the possible products. Specifically, the hydrogen atom of oxazole closest to the carbonyl oxygen of the acetate ligand will be deprotonated and subsequent formation of a biaryl bond will take place at this position (E.g. H2 in Figure 3-12). The assignment of this “nearest hydrogen” is unambiguous, as in the optimised structures of Intermediate 1 there is a strong interaction between the hydrogen to be deprotonated and the acetate ligand. This is manifested both in the rotation of oxazole as well as in the distance between the carbonyl oxygen and the hydrogen in question.

It should be noted that no transition state was considered when modelling the formation of Intermediate 1, in the various pathways considered below. This is because the geometry optimisations, starting from oxazole in close proximity to the Pd(II) catalyst, rapidly converged to the relevant Intermediate 1 structure, indicating that the formation of this species is highly favourable with no significant energy barrier. Furthermore, to the best of my knowledge, there are no suggestions in the relevant literature regarding such a TS.

Intermediate 1 N2:

Coordination via nitrogen will be considered first as this is expected to be the most stable species. Depending on the orientation of oxazole, coordination via the nitrogen atom can lead to two possible products, Product 2 and Product 4. When the hydrogen atom bonded to carbon 2, H2, is positioned to be deprotonated by the free oxygen of the acetate ligand, Product 2 will be formed. The optimised structure corresponding to this intermediate is shown in Figure 3-12. This possibility will henceforth be referred to as Intermediate 1 N2, i.e. Intermediate 1 Nitrogen coordinated with position 2 aligned to be substituted. The coordination of oxazole to the active catalyst displaces one oxygen atom of the acetate ligand and results in a four-coordinate palladium complex with an approximate square planar geometry. A detailed structural analysis in terms of bond lengths, dihedral angles, etc. is not in line with the aims of this work. However, it is worthwhile to at least examine the four metal–ligand bond lengths and compare them to experimental data to determine if this theoretical molecule’s structure is physically plausible.

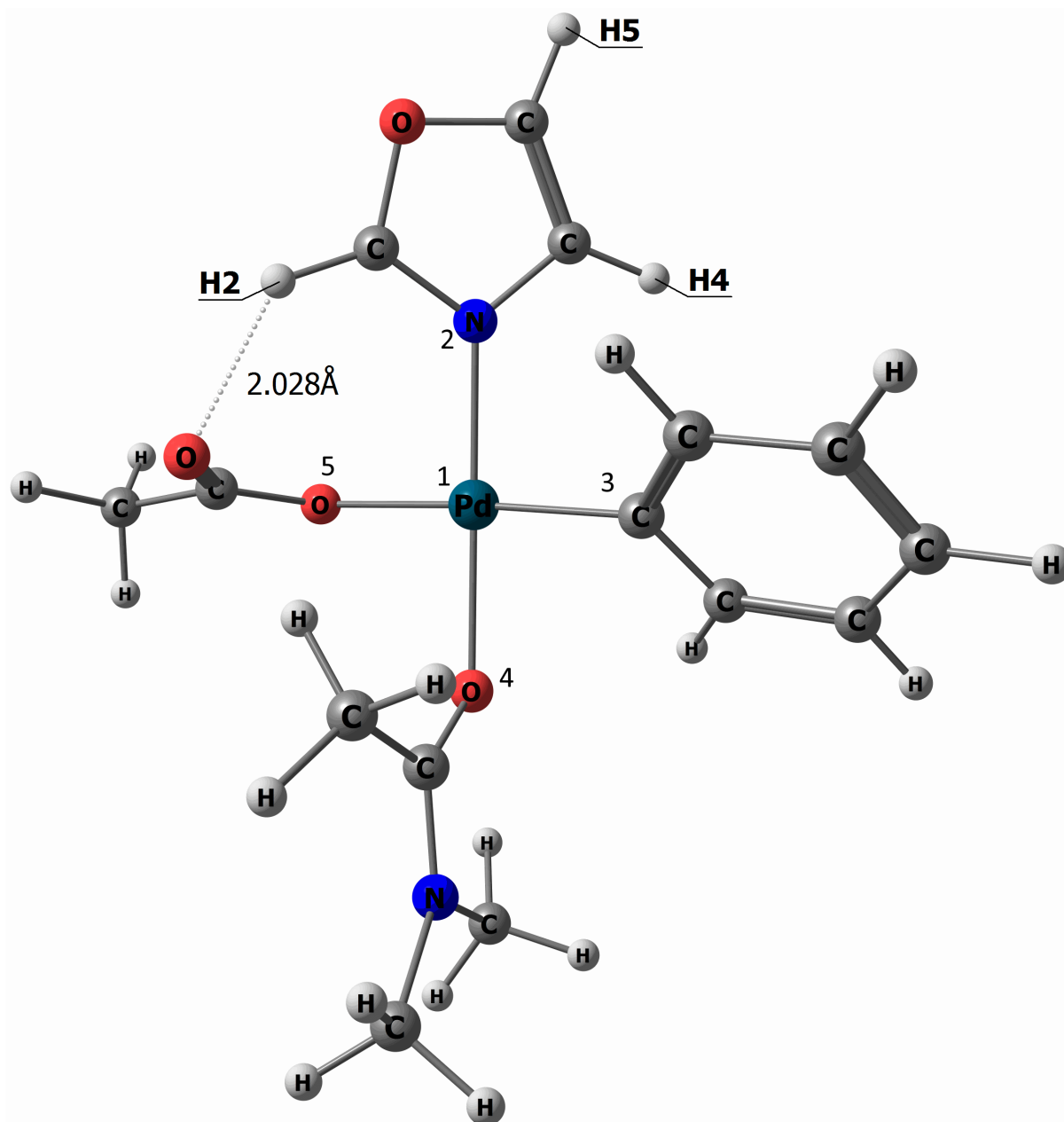


Figure 3-12: Intermediate 1 N2 – Showing H2 Orientated for Deprotonation by the Acetate Ligand

The Pd–N bond length, D(1-2), in Intermediate 1 N2 is 2.001 Å. The two crystal structures in the CSD that contain Oxazole coordinated to palladium are trans-dichloro-bis(oxazole)-palladium(II) (Figure 3-8) as well as the cis alternative of this complex. Here the D(1-2) lengths are 2.015 and 2.016 respectively. The Pd–C bond length of benzene, D(1-3), in the optimised structure is 1.975 Å. A CSD search reveals 8 compounds with Pd–Ph bonds in which the palladium atom was also ligated by a carboxylate group. The average D(1-3) distance here was

1.982 Å with values ranging from 1.955 - 2.005 Å. In the same data set, the corresponding Pd–O bond of the acetate ligand, D(1-5), was on average 2.134 Å with values ranging from 2.085 to 2.181 Å. The optimised structure had an acetate Pd–O bond length of 2.108 Å. For the remaining ligand, DMA, the model Pd–O bond was 2.063 Å and the same data used to assess the active catalyst was employed. These structural parameters and their experimental counterparts are summarised in Table 3-3.

Table 3-3: Selected Bond lengths for Intermediate 1 N2 (Refer to Figure 3-12 for atom numbering)

Bond Type	Model (Å)	Mean Experiment (Å)	Difference (Å)	Max -Min Experiment (Å)	Sample Size
D(1-2)	2.001	2.016	-0.015	2.015 - 2.016	2
D(1-3)	1.975	1.982	-0.007	1.955 - 2.005	8
D(1-5)	2.108	2.134	-0.026	2.085 - 2.181	8
D(1-4)	2.063	2.080	-0.017	2.004 - 2.155	2
		Mean Difference	-0.016		

Considering that the calculations were performed in the gas phase and the experimental values refer to a solid crystal, a mean absolute difference of 0.016 Å indicates that these values are sufficiently close to imbue faith in the model molecule. Although there is a systematic underestimation of the considered bond lengths, which is not expected when moving from gas phase to a solid, this trend is isolated to this particular optimisation and thus assumed to be anomalous. The distance between H2 and the carbonyl oxygen is 2.028 Å. This is well within the range of an intra-molecular hydrogen bond. In fact, the strength of this interaction is responsible for the rotation of oxazole towards the acetate ligand. This is manifested by the dihedral angle between the plane of the four ligand bonds and the plane bisecting oxazole being -48.82° as shown in Figure 3-13 when looking down upon oxazole.

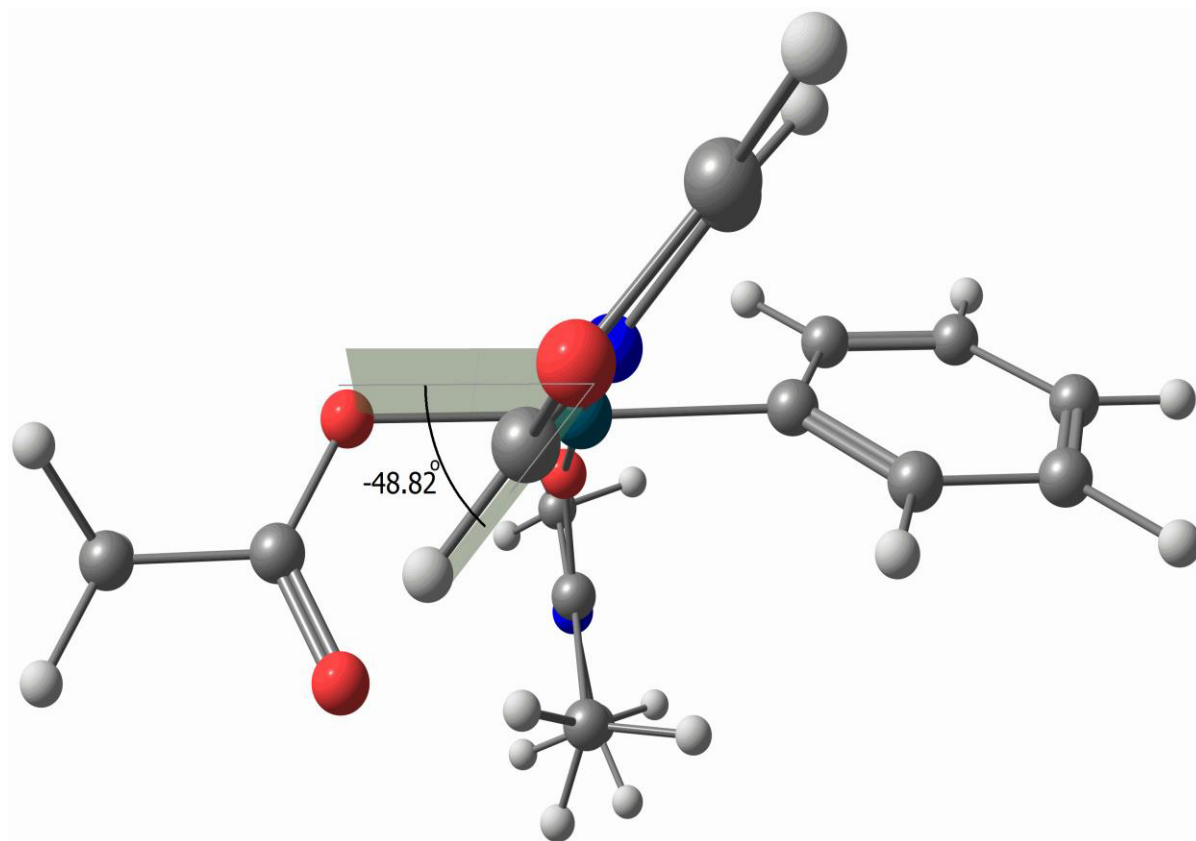


Figure 3-13: The rotation of oxazole in Intermediate 1 N2

A thorough conformational analysis is essential in any computational chemistry study to ensure that the relevant structures are as close as possible to global energy minima. In this way, one can be confident that the reaction pathways modelled mimic the minimum energy pathways followed by nature as closely as possible. There exists another conformation in which oxazole is coordinated to palladium via nitrogen that will also lead to deprotonation at the 2 position. This “enantiomer” of Intermediate 1 N2 has the acetate ligand rotated about the O–Pd bond as well as the oxazole ligand rotated about the N–Pd bond. Again, H2 is exposed to the carbonyl group of the acetate ligand. This optimised alternative to Intermediate 1 N2, Intermediate 1 N2 Alternative, is shown in Figure 3-14. As can be seen by comparing Figure 3-12 and Figure 3-14, in Intermediate 1 N2 Alternative, the acetate ligand is pointing into the page with oxazole also tilted in this direction, while in the original Intermediate 1 N2, the acetate ligand is pointing out of the page with oxazole also tilted out of the plane of the page. However, as the DMA ligand has maintained its orientation in both optimised structures, these are not true enantiomers but simply different conformations.

This relationship is perhaps best emphasised by placing a plane between the two molecules, as shown in Figure 3-15. Here it can be seen that aside from the DMA ligand, the two conformations roughly correspond to reflection in the displayed plane. The Gibbs energies of these two molecules could thus potentially be significantly different.

At the hybrid DFT level of theory (TPSSH/SVP see Section 2.2 for details) Intermediate 1 N2 is $\sim 1.84 \text{ kcal.mol}^{-1}$ more stable than Intermediate 1 N2 Alternative. As always, all reported Gibbs free energy values were calculated at 373.15K and 1 atm, under assumptions of ideal gas behaviour (Section 2.2.6).

It was assumed throughout this project, as a rough estimate, that the absolute error in energy differences is around 2 kcal.mol^{-1} for the system at the TPSSH/SVP level of theory. The energy difference between these two molecules is within the error of method. They could thus possibly be identical in energy.

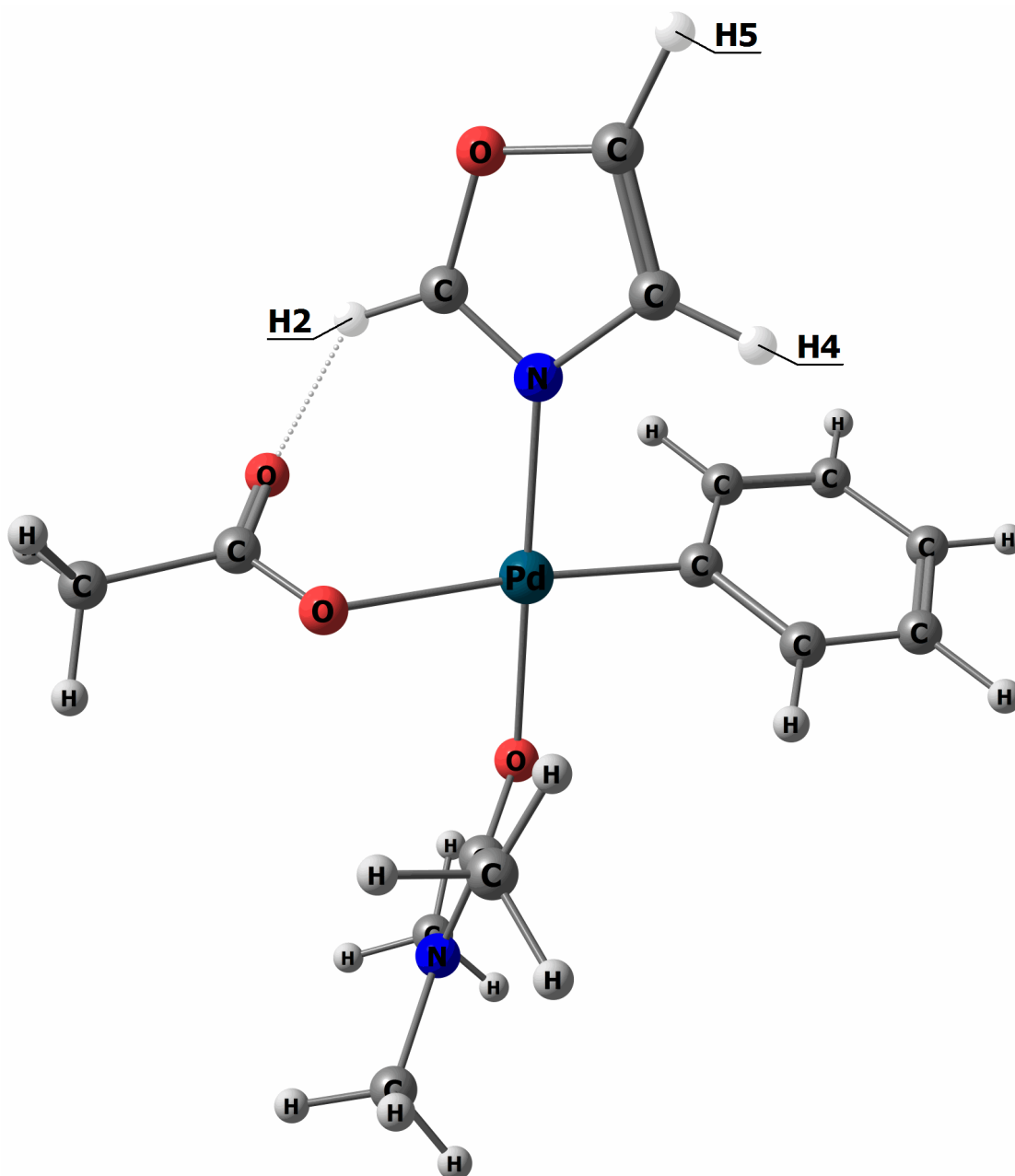


Figure 3-14: Intermediate 1 N2 Alternative - Showing H2 Orientated for Deprotonation by the Acetate Ligand

Additionally, there is likely to be free rotation about both the Pd–N bond of oxazole as well as the Pd–O bond of the acetate ligand at the elevated reaction temperatures making them completely interchangeable. For the sake of consistency, reactions will be modelled starting from the lowest energy intermediate, Intermediate 1 N2, and the alternative will not be further considered.

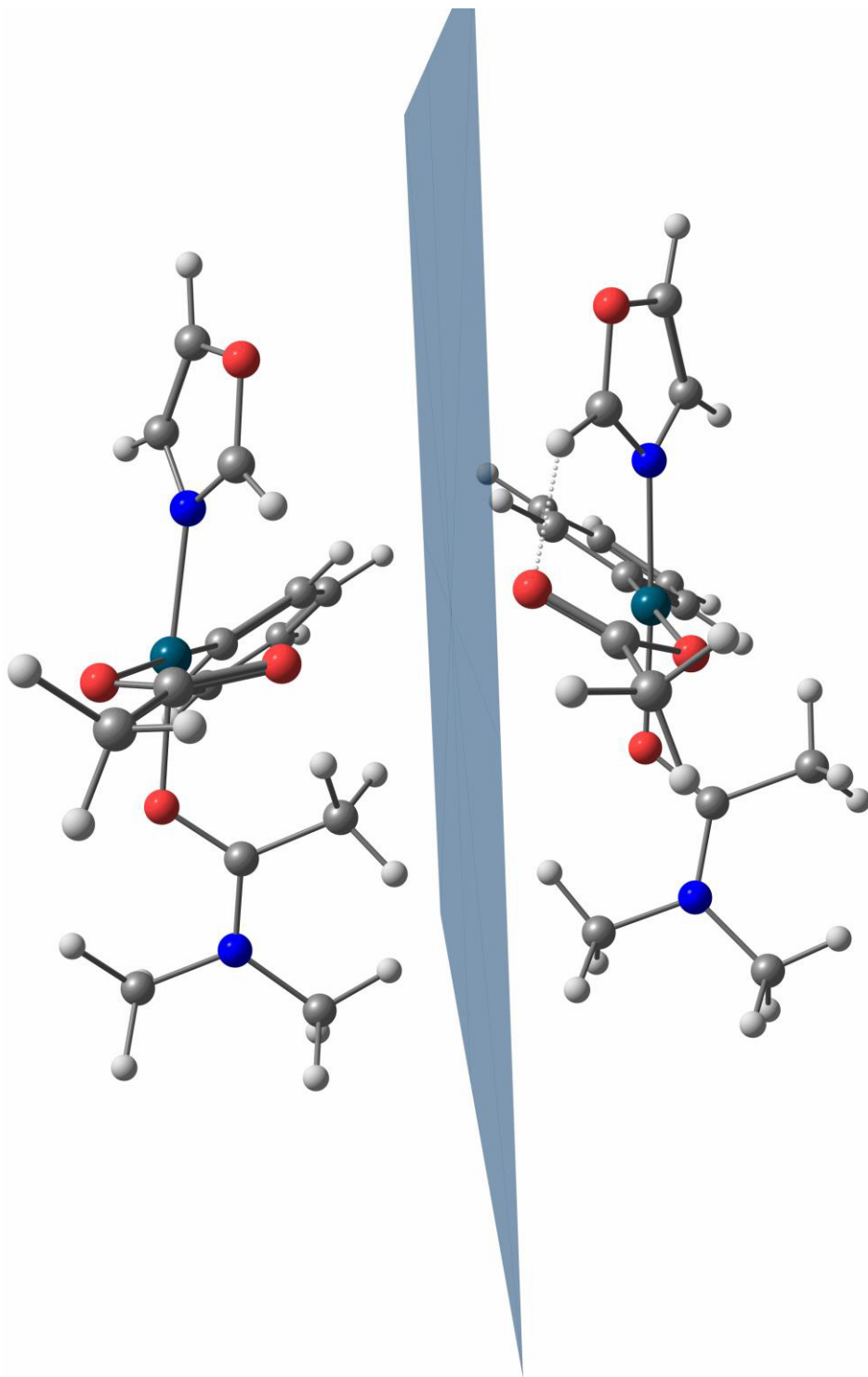


Figure 3-15: Intermediate 1 N2 (left) vs Intermediate 1 N2 Alternative (right)

Intermediate 1 N4:

In a similar fashion, when oxazole is coordinated via nitrogen and the hydrogen atom bonded to carbon 4, H4, is positioned to be deprotonated, Product 4 will be formed. This possibility will henceforth be referred to as Intermediate 1 N4, i.e. Intermediate 1 Nitrogen coordinated with position 4 aligned to be substituted and is shown in Figure 3-16.

An analysis of the ligand-metal bond lengths, shown in Table 3-4, again confirms the physical plausibility of this molecule. The distance between H4 and the carbonyl oxygen is 2.110 Å. This is also within the range of an intra-molecular hydrogen bond and causes oxazole to rotate towards the acetate ligand. In this case the relevant dihedral angle is -51.43° (Figure 3-17).

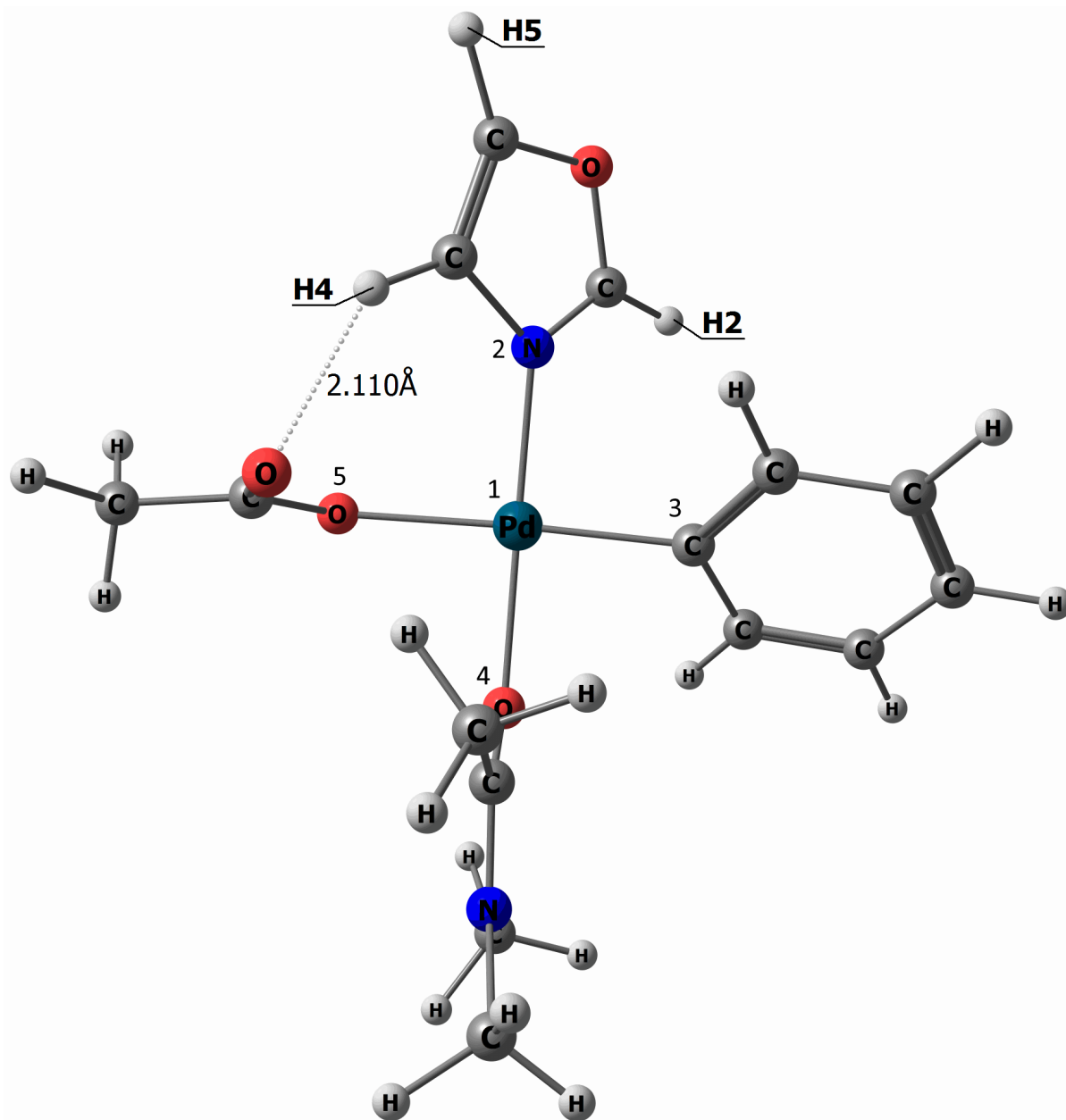


Figure 3-16: Intermediate 1 N4 - Showing H4 Orientated for Deprotonation by the Acetate Ligand

Table 3-4: Selected Bond lengths for Intermediate 1 N4 (Refer to Figure 3-16 for atom numbering)

Parameter	Model (Å)	Mean Experiment (Å)	Difference (Å)	Max -Min Experiment (Å)	Sample Size
D(1-2)	1.994	2.016	-0.021	2.015 - 2.016	2
D(1-3)	1.976	1.982	-0.006	1.955 - 2.005	8
D(1-5)	2.107	2.134	-0.027	2.085 - 2.181	8
D(1-4)	2.061	2.080	-0.019	2.004 - 2.155	2
		Mean Difference	-0.018		

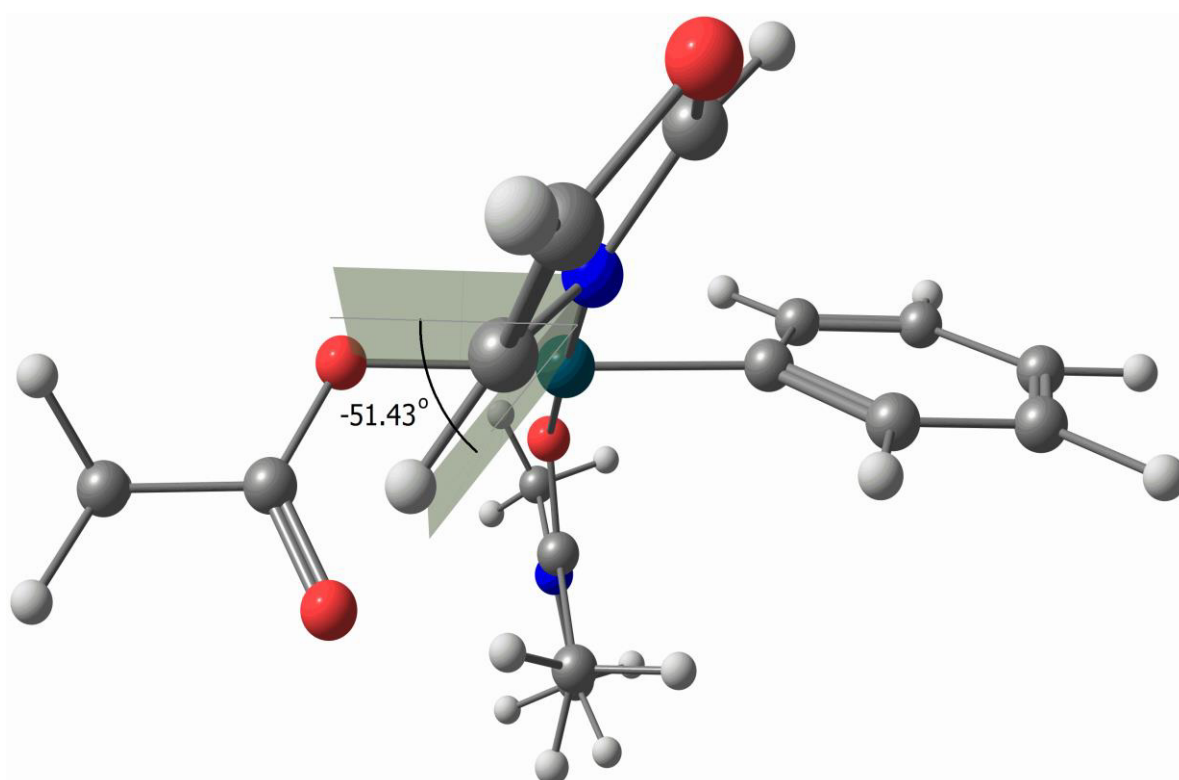


Figure 3-17: Rotation of Oxazole –Intermediate 1 N4

As expected, the intra-molecular hydrogen bond in Intermediate 1 N2 is shorter, by 0.082 Å, than the corresponding interaction in Intermediate 1 N4. This is because the carbon donor atom in Intermediate 1 N2 is bonded to two more electronegative elements, O and N. It is thus partially positively charged and via an inductive effect, electron density is drawn from the C–H bond toward carbon, further polarising this bond and increasing the strength of the O---H interaction. The carbon donor atom in Intermediate 1 N4 is however, not as partially positively charged as it is only bonded to one more electronegative atom, N and the C–H bond is consequently less polarised.

As before, an “enantiomer” like version of Intermediate 1 N4, in which the acetate ligand is rotated about the O–Pd bond and oxazole about the N–Pd bond, was considered. This Intermediate 1 N4 Alternative molecule is shown across a mirror plane from the original in Figure 3-18. In this case, the DMA molecule did not maintain its orientation during the optimisation of the alternative structure and consequently the two molecules are quite close to being true enantiomers. The difference in free energy between these two alternatives is 0.98 kcal.mol⁻¹, with the original being more stable.

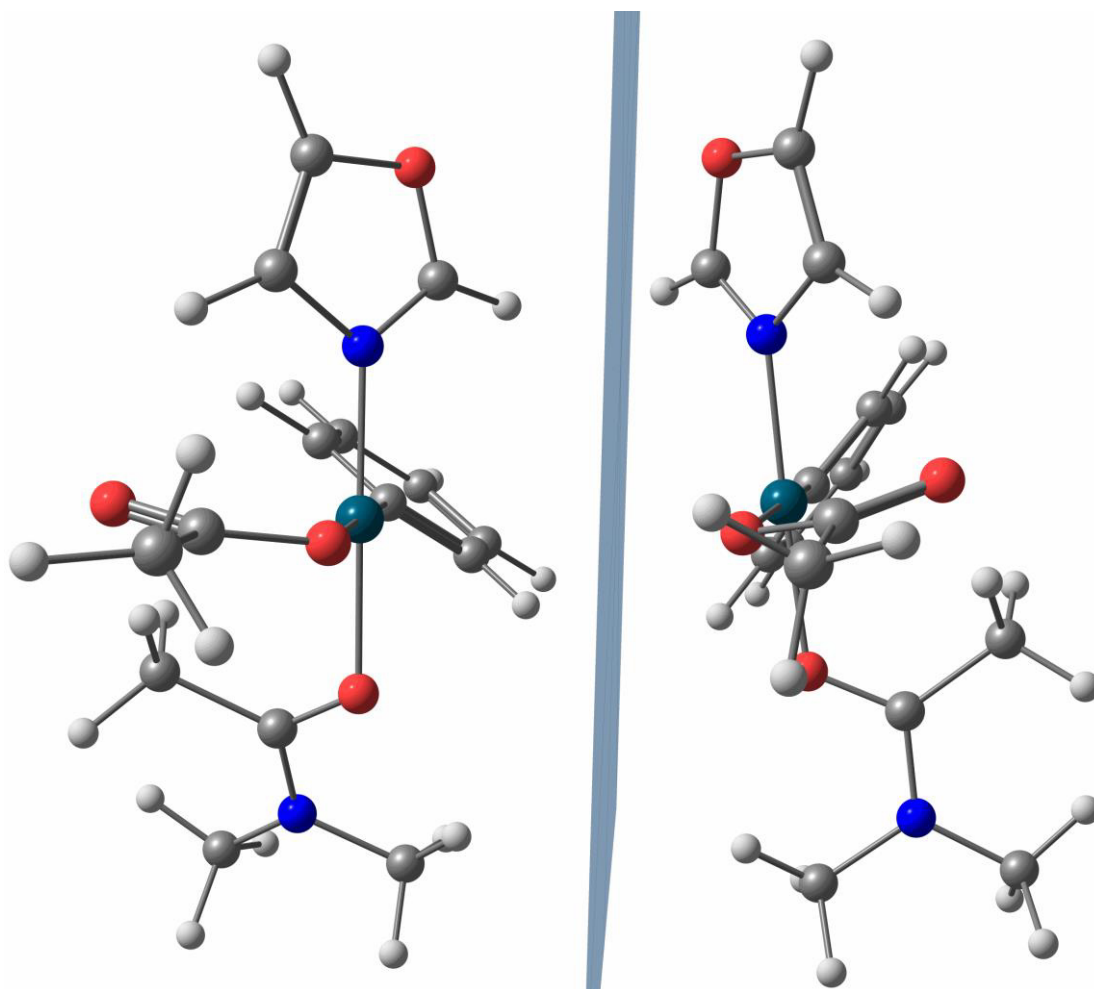


Figure 3-18: Intermediate 1 N4 (right) vs Intermediate 1 N4 Alternative (left)

This Gibbs free energy difference is well within the error of the method. Thus, it is likely that the energies of the two conformations are very similar, although again for consistency subsequent reactions will be modelled using the lowest energy possibility.

Intermediate 1 =4:

Coordination via the double bond between carbons 4 and 5 of oxazole (Figure 3-7) can also lead to two possible products, again depending on the orientation of oxazole in the stable intermediate. The first of these possibilities corresponds to the formation of Product 4 and will henceforth be referred to as Intermediate 1 =4 i.e. Intermediate 1 double bond coordinated, represented as =, with position 4 aligned to be deprotonated. The optimised structure of Intermediate 1 =4 is shown in Figure 3-19. Once more, a clear interaction can be seen between the acetate carbonyl oxygen and H4, with a distance of 2.209 Å separating them.

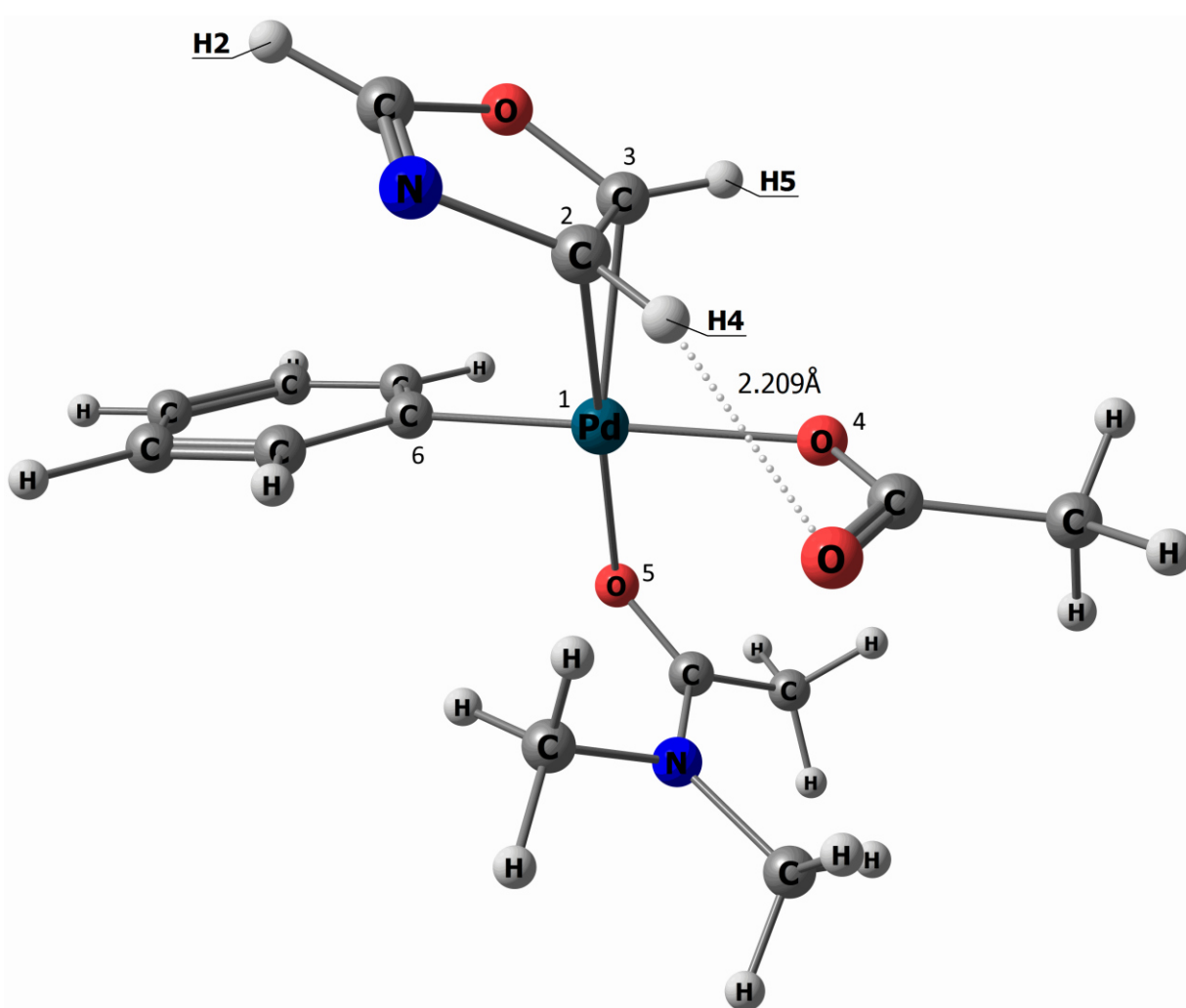


Figure 3-19: Intermediate 1 =4 - Showing H4 Orientated for Deprotonation by the Acetate Ligand

An analysis of various structural parameters is presented in Table 3-5. The length of the coordinated double bond, D(2-3), is 1.414 Å. This compares fairly well with the average value

for 75 crystal structures of 5 member heterocyclic rings coordinated to palladium through a double bond. As expected, the double bond has lengthened somewhat, by 0.051 Å, when compared to an uncoordinated oxazole molecule. This is due to donation of electron density from the double bond onto palladium as well as back donation from the metal atom. Back donation from filled d orbitals on the palladium atom to the empty π^* orbitals of the double bond, weakens the double bond contributing to an increase in its length. However, it seems that back donation does not take place to a significant extent as the bond length before and after coordination, 1.363 Å and 1.414 Å respectively, are both within the range of an aromatic carbon–carbon bond. The distance between the palladium atom and the double bond, as measured by an average of the two Pd–C distances, D(1-2) and D(1-3), is also reasonable. Considering the above as well as the other entries in Table 3-5, it can be concluded that the optimised structure of Intermediate 1 =4 is plausible enough to continue with the study.

As before, there exists an alternative conformation in which oxazole coordination via the double bond between atoms 4 and 5 leads to deprotonation at the 4 position. This Intermediate 1 =4 Alternative structure can be formed by rotation of the Pd–O acetate bond and by rotating oxazole around the axis of the H2–C2 bond and is presented in Figure 3-20. The Gibbs free energy difference ($G_{\text{Original}} - G_{\text{Alternative}}$) of -0.64 kcal.mol⁻¹ is smaller than the assumed error of the method. As per usual, the lower energy structure, in this case Intermediate 1 =4, will be considered exclusively.

Table 3-5: Selected Bond lengths for Intermediate 1 =4 (Refer to Figure 3-19 for atom numbering)

Parameter	Model (Å)	Mean Experiment (Å)	Difference (Å)	Max - Min Experiment (Å)	Sample Size
D(2-3)	1.414	1.397	0.017	1.335 - 1.530	75
Average D(1-2) and D(1-3)	2.154	2.206	-0.052	2.036 - 2.705	75
D(1-6)	1.973	1.982	-0.009	1.955 - 2.005	8
D(1-4)	2.126	2.134	-0.008	2.085 - 2.181	8
D(1-5)	2.115	2.080	0.035	2.004 - 2.155	2
		Mean Absolute Difference	0.024		

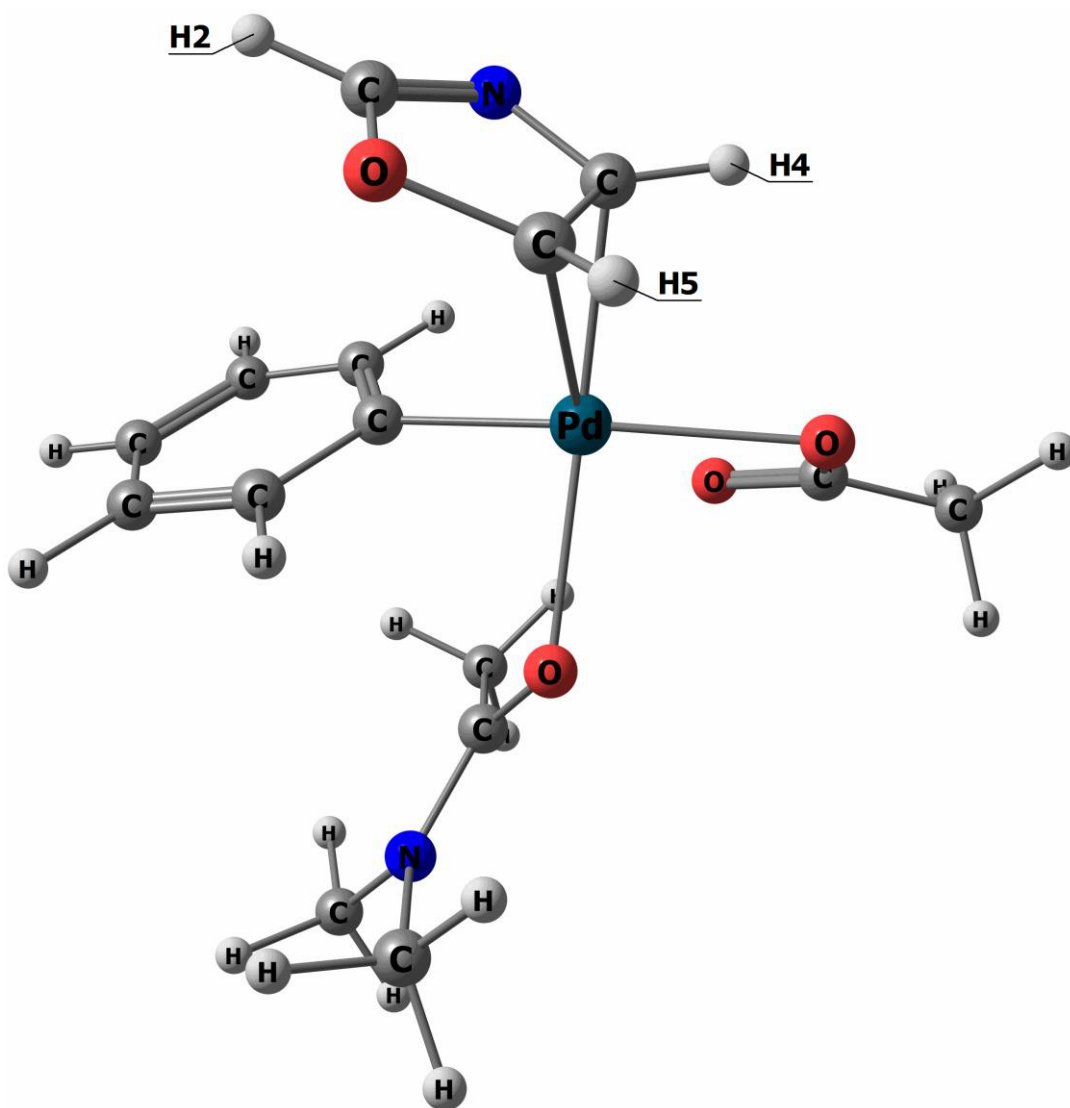
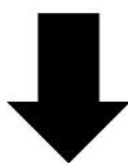
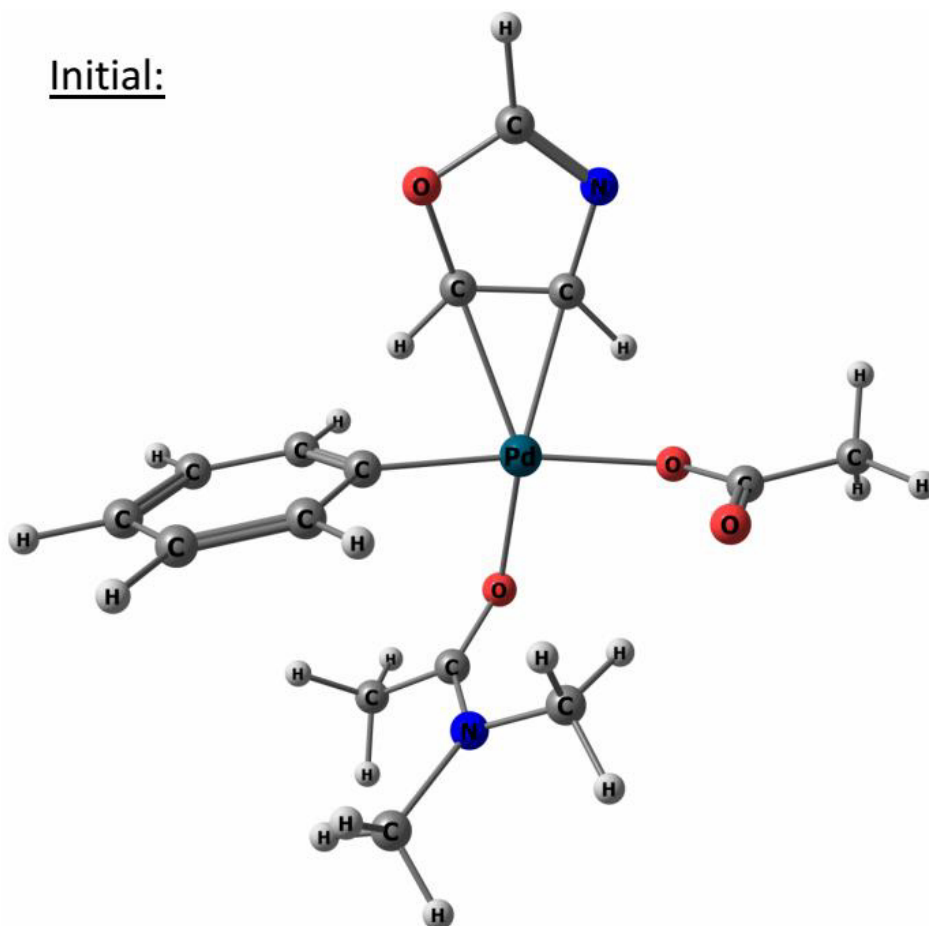


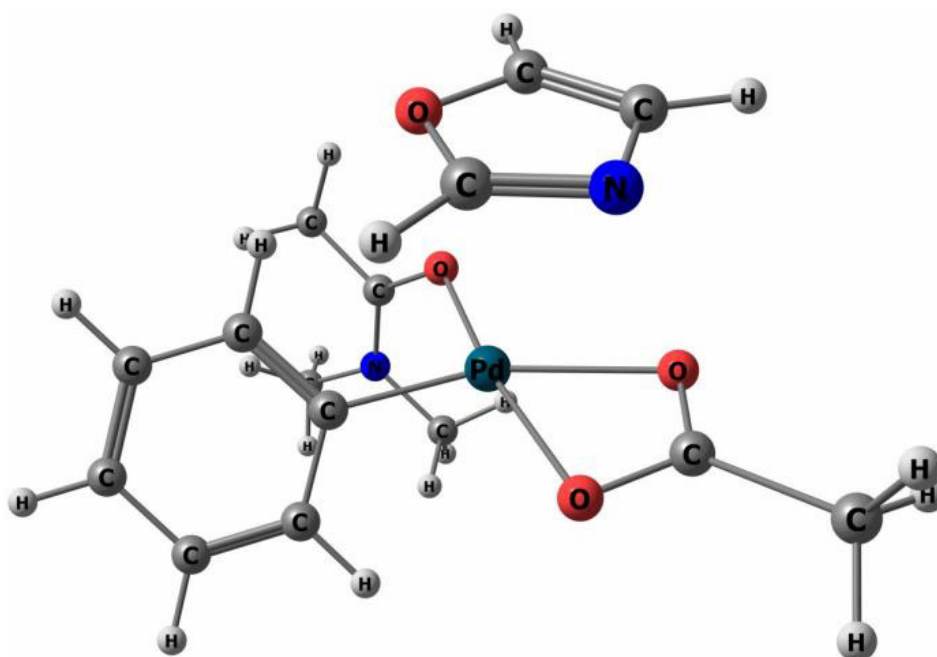
Figure 3-20: Intermediate 1 =4 Alternative

It should be noted that alternative modes of coordination of the double bond were also investigated. For example, starting with the double bond parallel rather than perpendicular to the C (benzene)–Pd–O (acetate) axis. As can be seen from Figure 3-21, an optimization starting with this structure essentially led to the original catalyst and oxazole uncoordinated. Additionally, oxazole could possibly be coordinated with the carbon double bond perpendicular to the C (benzene)–Pd–O (acetate) axis but with the molecule pointing straight upwards (Figure 3-22).

Initial:

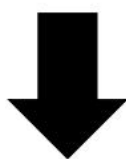
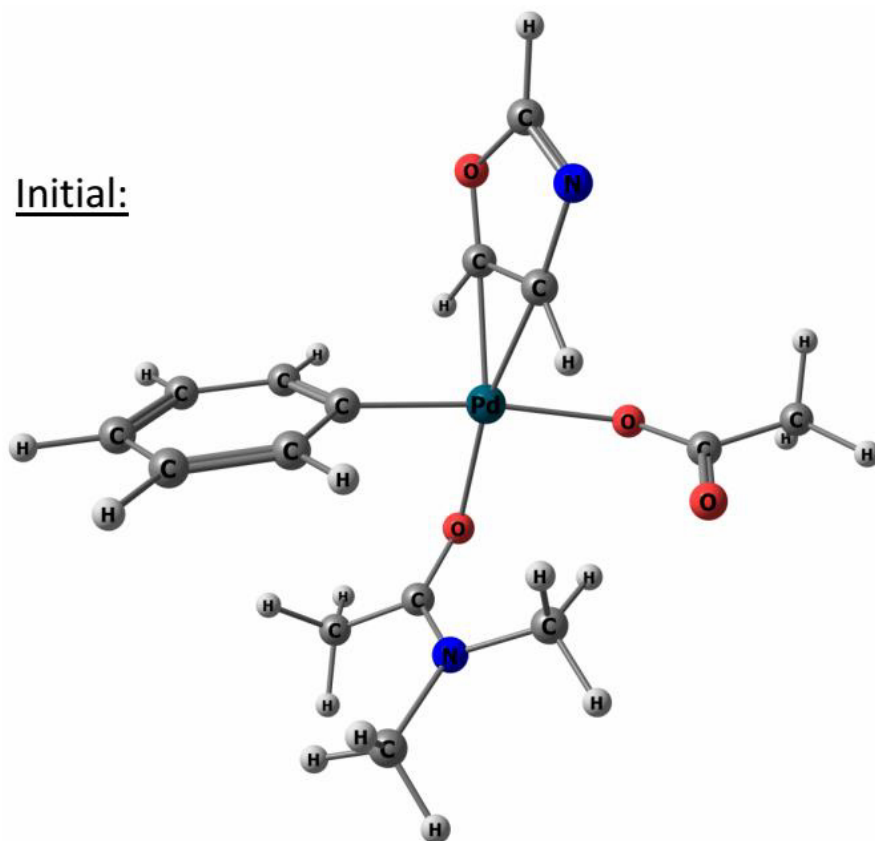


Final:

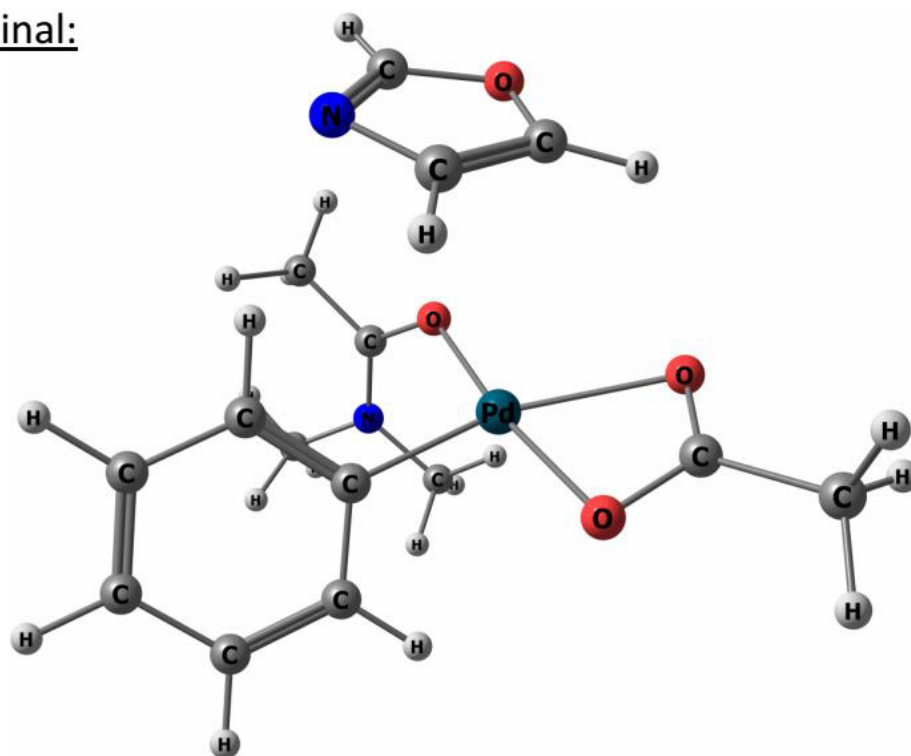


[Figure 3-21: Alternative Mode of Double Bond Coordination 1](#)

Initial:



Final:



[Figure 3-22: Alternative Mode of Double Bond Coordination 2](#)

This optimisation also resulted in the original catalyst and an uncoordinated oxazole. These alternative modes of double bond coordination will thus not be considered again.

Intermediate 1 =5:

Coordination via the double bond between carbons 4 and 5 can also lead to Product 5 if the hydrogen at this position is aligned to be deprotonated. This intermediate, henceforth referred to as Intermediate 1 =5, is shown in Figure 3-23. The distance between the carbonyl oxygen and H5 is 2.223 Å. A bond length analysis confirming the reliability of the optimisation is shown in Table 3-6.

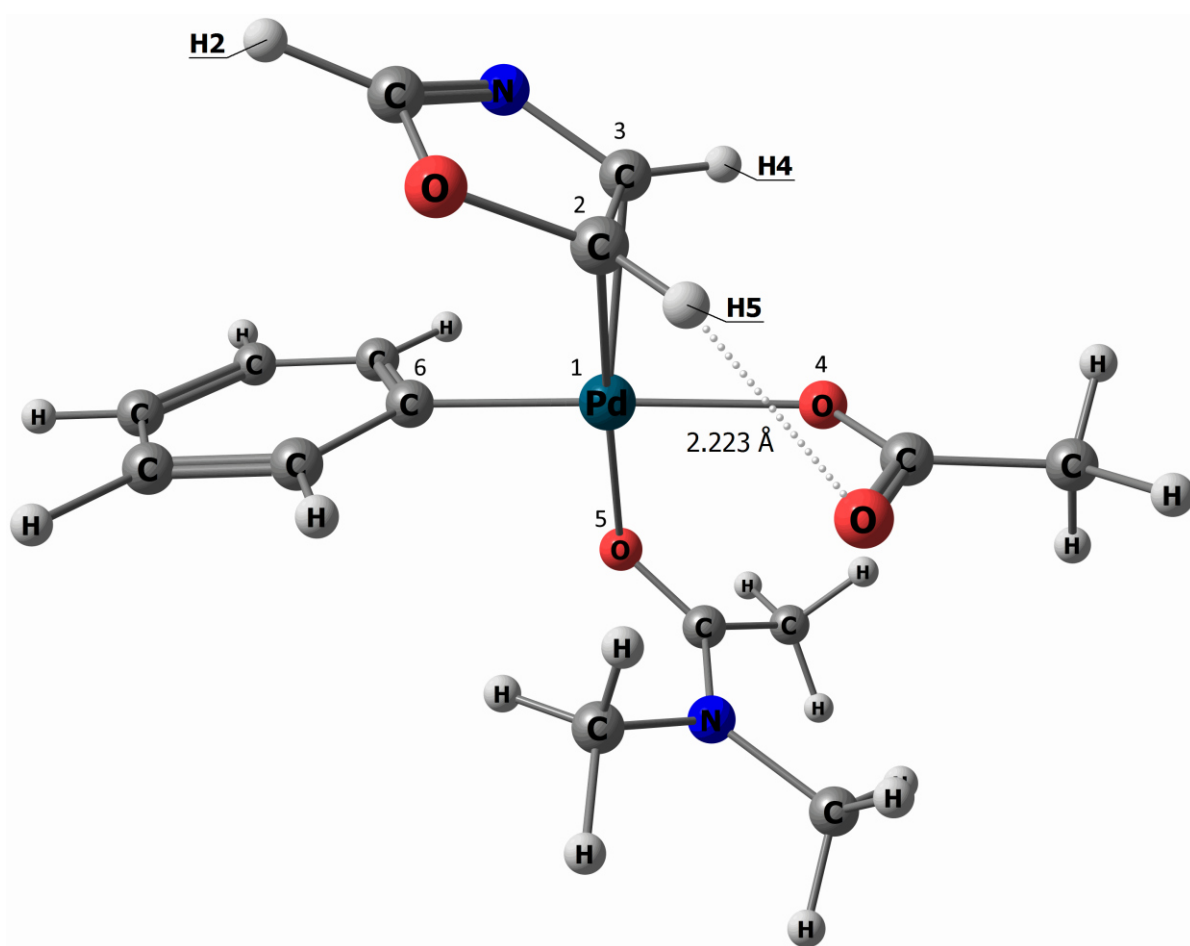


Figure 3-23: Intermediate 1 =5- Showing H5 Orientated for Deprotonation by the Acetate Ligand

Table 3-6: Selected Bond lengths for Intermediate 1 =5 (Refer to Figure 3-23 for atom numbering)

Parameter	Model (Å)	Mean Experiment (Å)	Difference (Å)	Max - Min Experiment (Å)	Sample Size
D(2-3)	1.413	1.397	0.016	1.335 - 1.530	75
Average D(1-2) and D(1-3)	2.154	2.206	-0.052	2.036 - 2.705	75
D(1-6)	1.972	1.982	-0.010	1.955 - 2.005	8
D(1-4)	2.128	2.134	-0.006	2.085 - 2.181	8
D(1-5)	2.116	2.080	0.036	2.004 - 2.155	2
		Mean Absolute Difference	0.024		

As with all the previous intermediates an alternative conformation exists. In an analogous way to Intermediate 1 =4, this can be formed by rotation of the Pd–O acetate bond and by rotating oxazole around the axis of the H2–C2 bond. Intermediate 1 =5 Alternative is depicted in Figure 3-24. The Gibbs free energy of the original intermediate is lower than the alternative intermediate by 2.49 kcal.mol⁻¹. As it will be assumed that nature will predominantly favour the lower energy structure, the alternative will not be further considered.

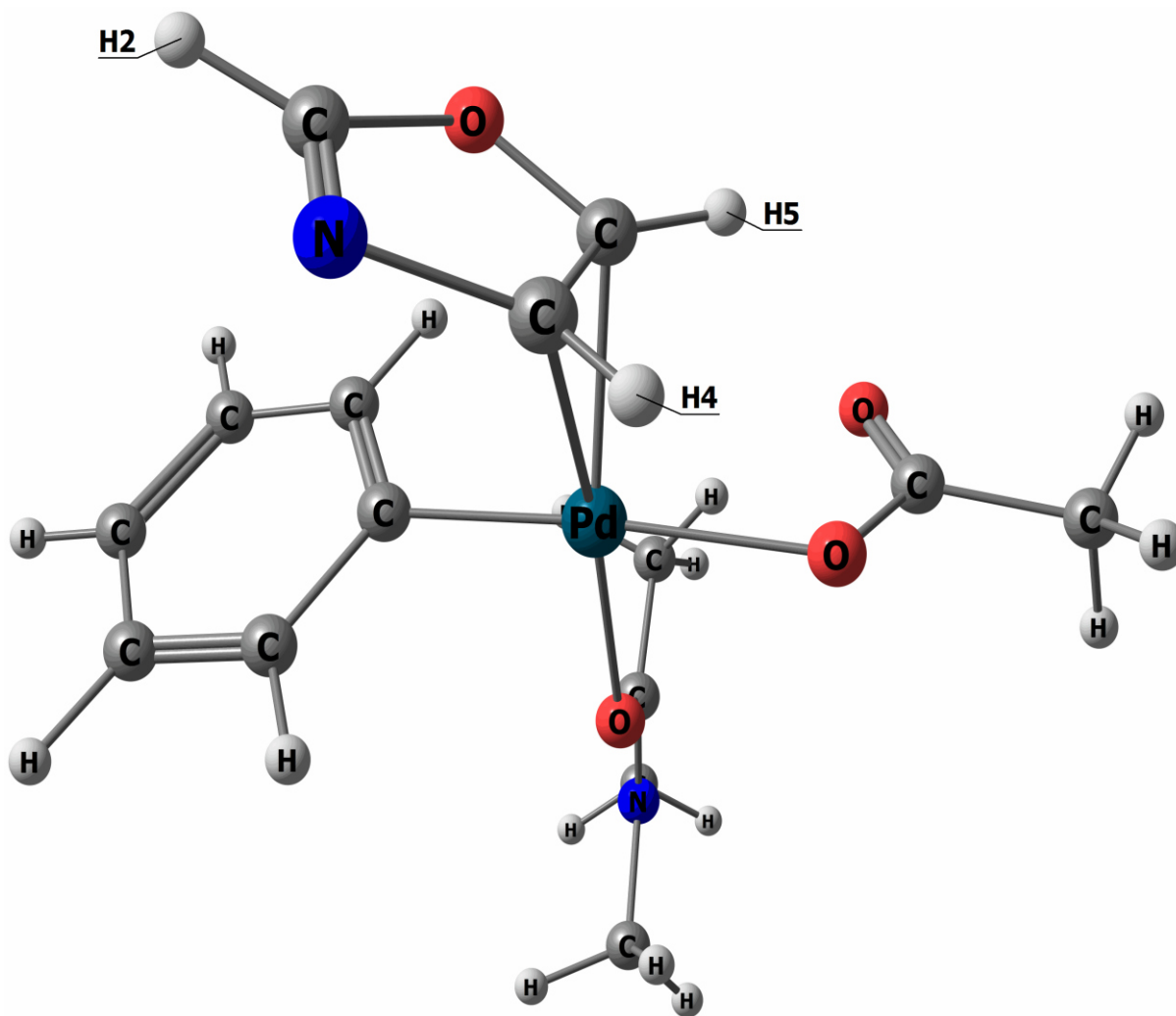
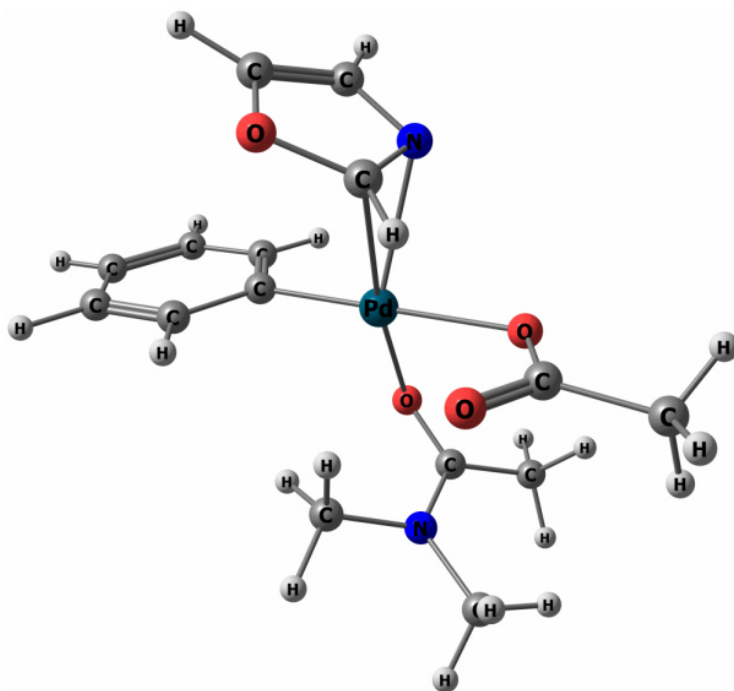


Figure 3-24: Intermediate =5 Alternative

Intermediate 1 N=2:

Coordination via the double bond between nitrogen and carbon 2 of oxazole could lead to deprotonation at the 2 position. However, as shown in Figure 3-25, an optimisation of such an intermediate simply resulted in the formation of Intermediate N2. Thus, it can be concluded that due to the strong preference for coordination via nitrogen, complexation of oxazole to palladium through this particular double bond is unlikely.

Initial:



Final:

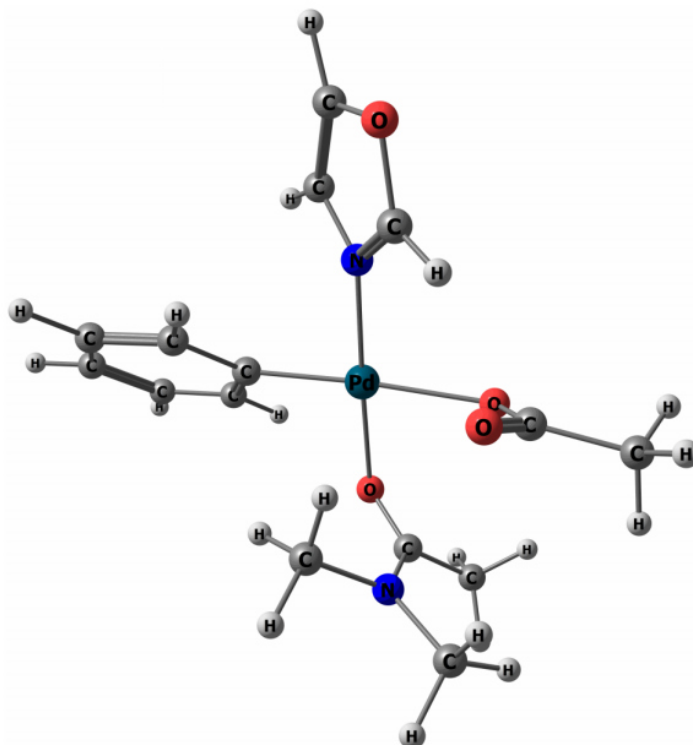
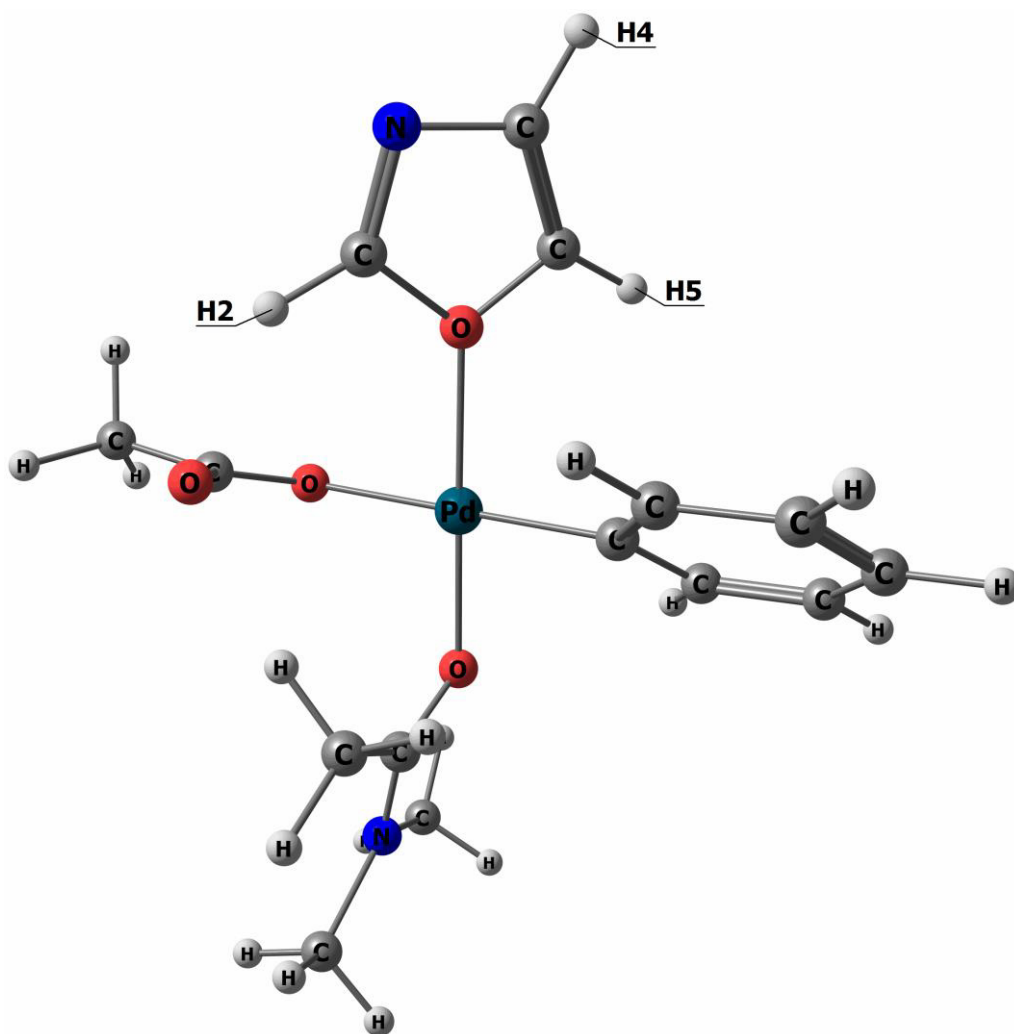


Figure 3-25: Optimization of Intermediate N=2 – Resulting in the formation of Intermediate N2

Intermediate 1 O2 and Intermediate 1 O5:

Coordination via oxygen can lead to deprotonation at the 2 or 5 positions. Optimised structures for these possibilities are shown in Figure 3-26 and Figure 3-27, respectively. However, as will be discussed shortly, the much higher energy of these oxygen coordinated species, compared to the other possibilities, makes their formation unlikely and no structural analysis was deemed necessary.



[Figure 3-26: Intermediate 1 O2- Showing H2 Orientated for Deprotonation by the Acetate Ligand](#)

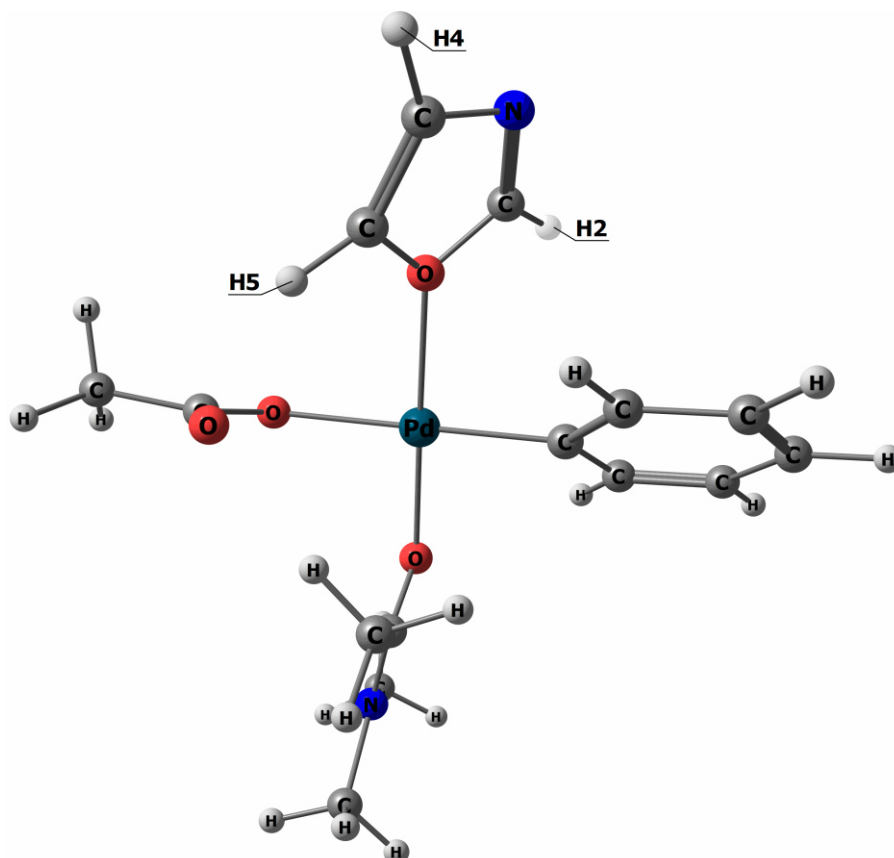


Figure 3-27: Intermediate 1 O5 - Showing H5 Orientated for Deprotonation by the Acetate Ligand

The free energies of formation of the various Intermediate 1 possibilities ($\Delta G_{\text{Int } 1}$) are compared in Figure 3-28. All values are relative to the infinitely separated oxazole and active catalyst molecules, i.e. oxazole + Pd(II) catalyst = 0 kcal.mol⁻¹. The formation of Intermediate 1-N2, N4, =4 and =5 are all negative, indicating that they will form spontaneously when an oxazole molecule and the active catalyst come into contact, assuming no significant transition state (TS). As no covalent bonds are being broken in this process this is a valid assumption. Additionally, for all four of these possibilities the $\Delta G_{\text{Int } 1}$ values are in the relatively narrow range of -2.57 to -7.66 kcal.mol⁻¹. In contrast, the oxygen coordinated intermediates require energy to form and are on average ~20 kcal.mol⁻¹ higher in energy than the other possibilities. It can thus be concluded that in nature an intermediate in which oxazole is coordinated via the oxygen atom will not form to any significant degree and no further reactions incorporating such species will be modelled. This correlates well with the observation that no experimental crystal structures were found with oxazole coordinated to palladium via oxygen.

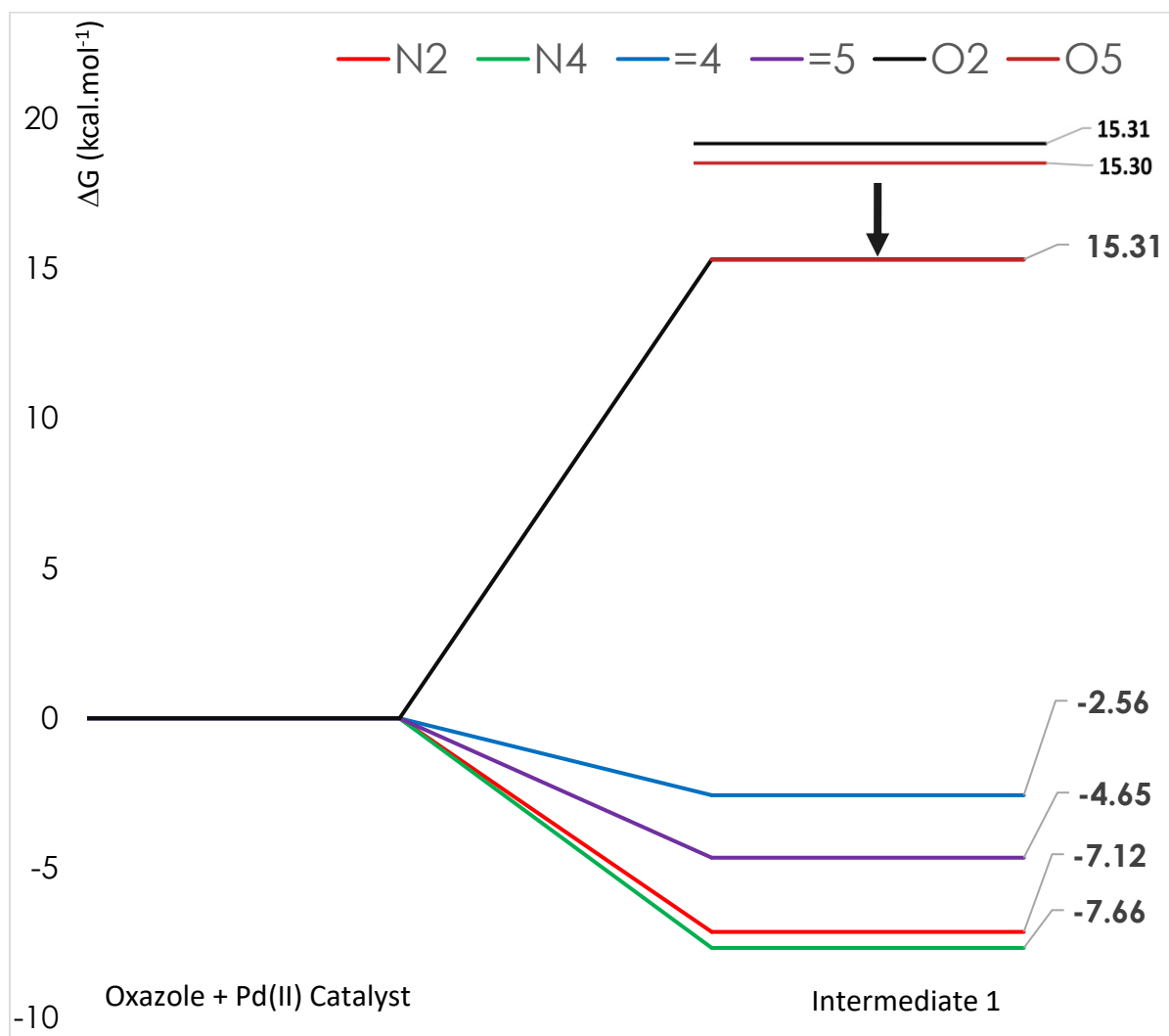


Figure 3-28: Phosphine Free Conditions - Intermediate 1 Gibbs Free Energies of Formation

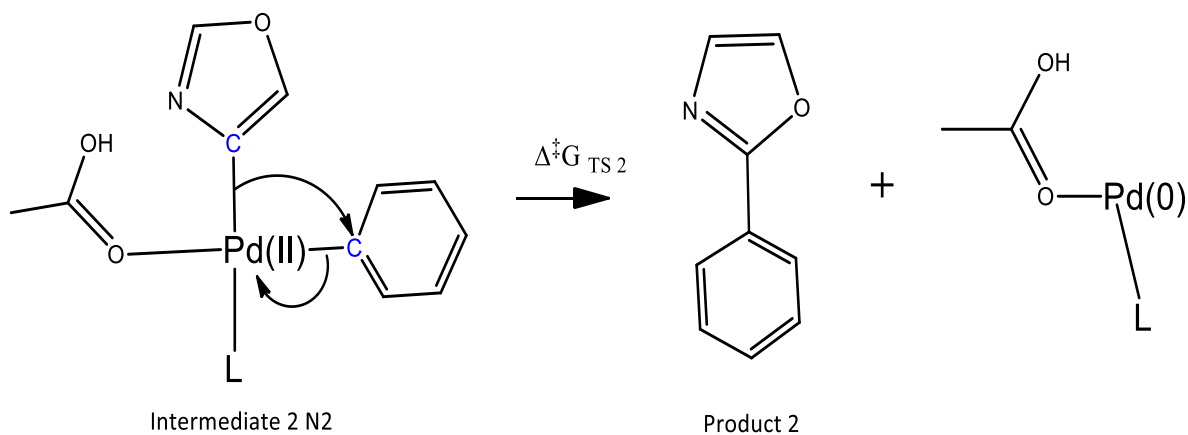
The four remaining plausible Intermediate 1 structures each undergo a sequence of two reactions before for the final mono-substituted products are formed (Figure 3-1). The first of these is the concerted metallation deprotonation (CMD) step and the second a reductive elimination (RE) step. Thus, each Intermediate 1 structure serves as a starting point for a sequence of four reaction pathways. Each of these pathways will be named after the intermediate from whence it originated. The four pathways to be considered are thus Pathway N2, Pathway N4 Pathway =4 and Pathway =5.

The Reaction Pathways:

An overview of the reductive elimination step:

Once Intermediate 2 has been formed the two reactive *cis* Pd–C bonds will undergo reductive elimination (RE). In this process, the oxidation state of palladium returns to 0 from II and a new biaryl bond is formed between the two carbon atoms previously bonded to Pd. The electrons from one of the organometallic bonds localise on palladium while the electrons in the remaining organometallic bond form a biaryl bond which couples the aryl groups together. As mentioned previously, the transition state associated with this reaction will be referred to as Transition State 2 (TS 2).

For each of the reaction pathways to be discussed, the transition state associated with the reductive elimination step was found by performing a relaxed surface scan (see Section 2.2.7). In this surface scan the distance between the two carbon atoms involved in organometallic bonds in Int 2 was systematically shortened (the carbon atoms depicted in blue in Figure 3-29). The highest energy structure from this surface scan was then used as input for a TS optimisation.



Or equivalently:

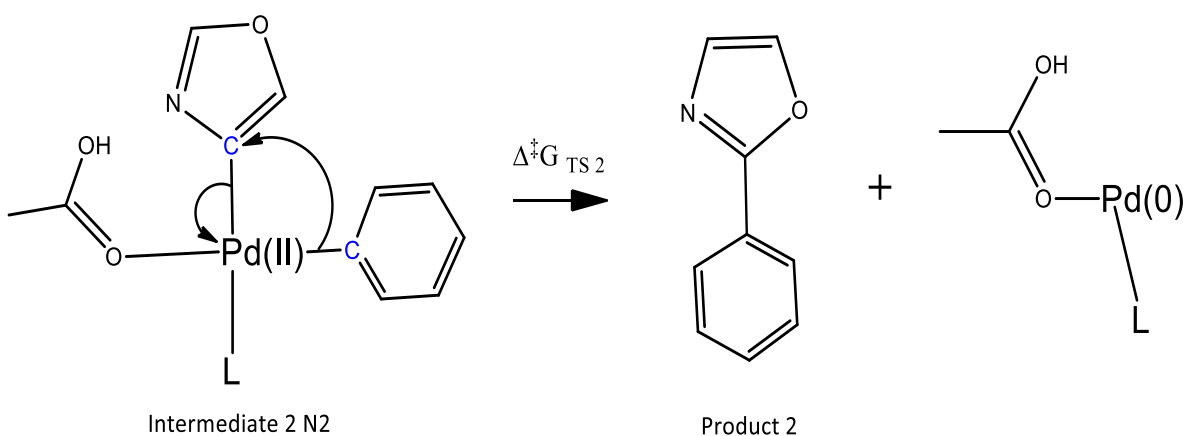


Figure 3-29 Product Formation Via Reductive Elimination

Reductive elimination generally takes place through a two electron three centre transition state.¹⁵ The atomic movements of the imaginary frequency associated with this transition state correspond to the coupling of the two aryl moieties and are depicted by arrows in Figure 3-30.

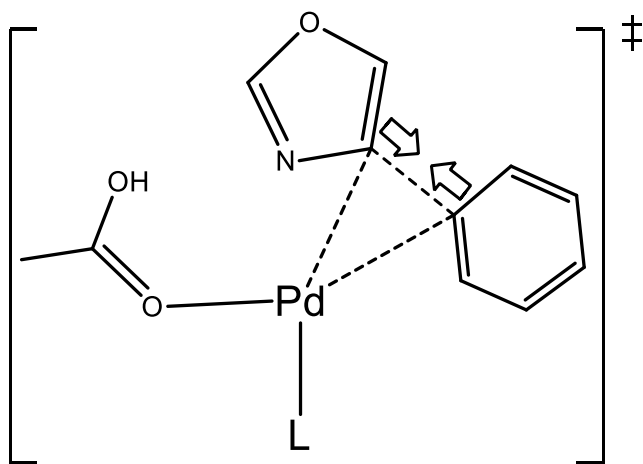
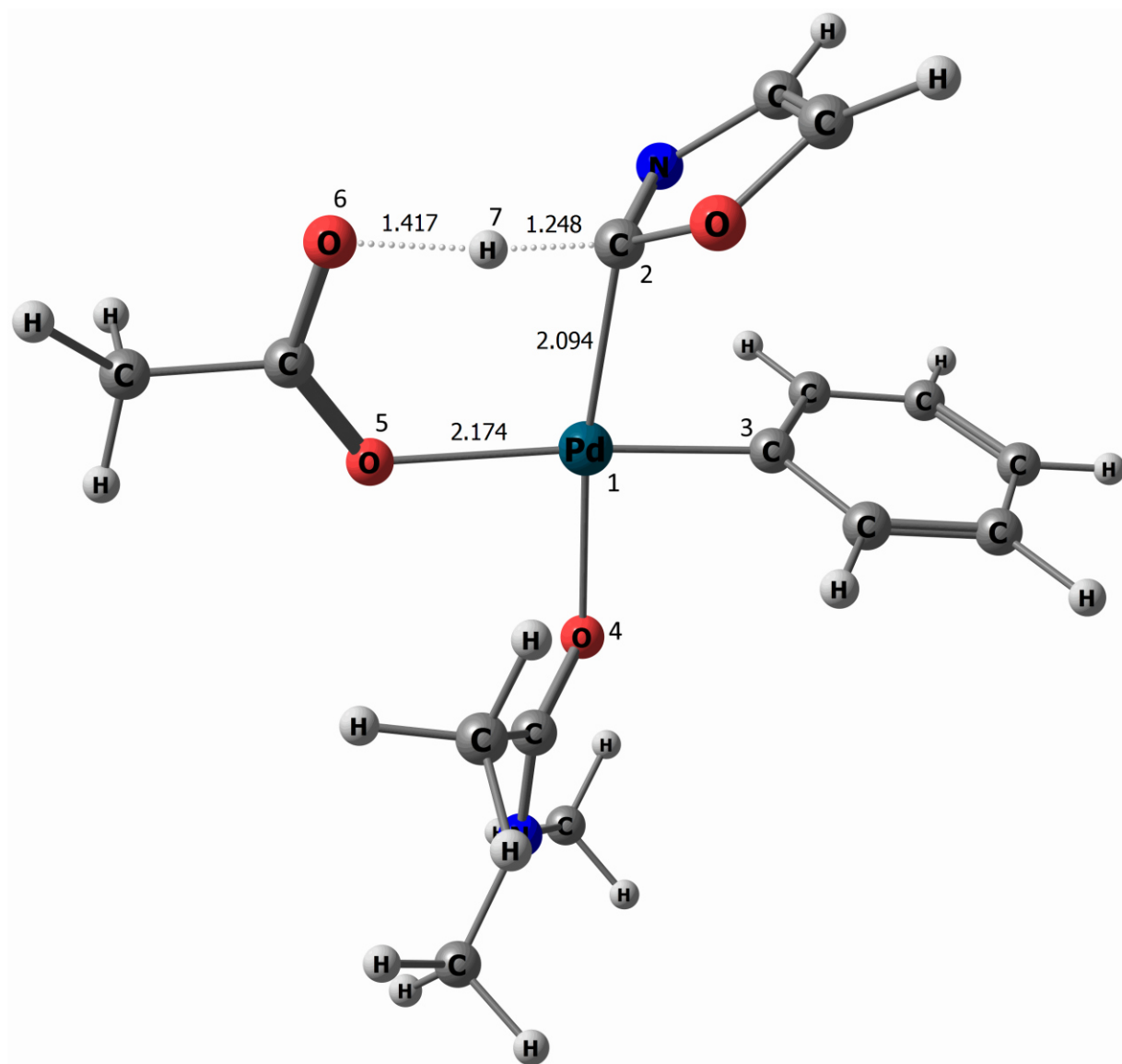


Figure 3-30: Reductive Elimination Transition State (TS 2 Pathway N2)

Pathway N2:

The CMD transition state (TS 1) for the Pathway N2 is illustrated in Figure 3-31, along with selected bond lengths (all in Å). This structure has an imaginary frequency of -145.42 cm^{-1} . As expected the C–H bond being deprotonated, D(2-7), has lengthened significantly from the Int 1 value of 1.094 Å and the molecule corresponds to a structure approximately halfway in the process of O–H bond formation. The new Pd–C bond, D(1-2), is also somewhat longer in the TS than the final value of 1.935 Å in Int 2. The corresponding product, Intermediate 2 N2 is shown in Figure 3-32.



[Figure 3-31: Transition State 1 N2](#)

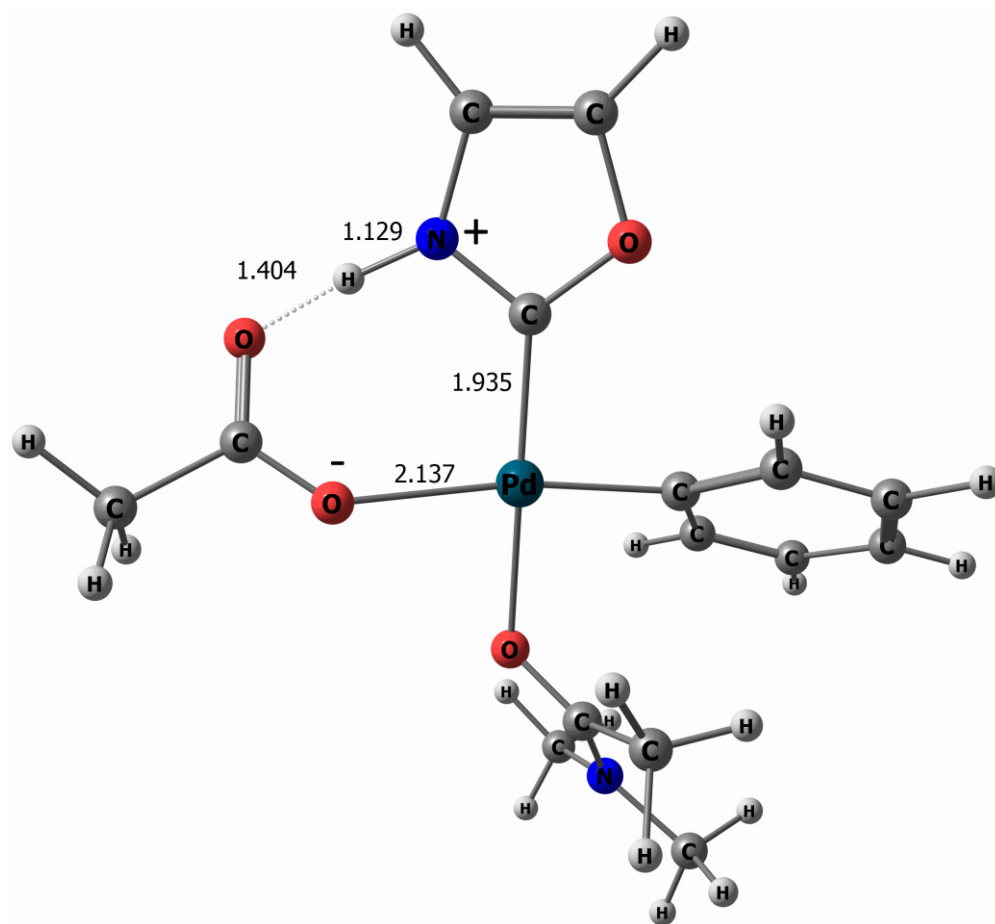


Figure 3-32: Intermediate 2 N2 – Nitrogen of Oxazole Facing the Acetate Ligand

An interesting feature of Intermediate 2 N2 is that oxazole has rotated about the C–Pd bond to enable an acid base reaction to take place between the acetic acid ligand and the basic nitrogen atom of oxazole. The result is a zwitterionic species, as can be seen by considering the relevant bond lengths. The corresponding H bond is strengthened, as the H bond donor is now a positive nitrogen atom. An alternative structure in which oxazole rotates so that the oxygen atom faces the acetate ligand, Intermediate 2 N2 Alternative, was also optimised (Figure 3-33).

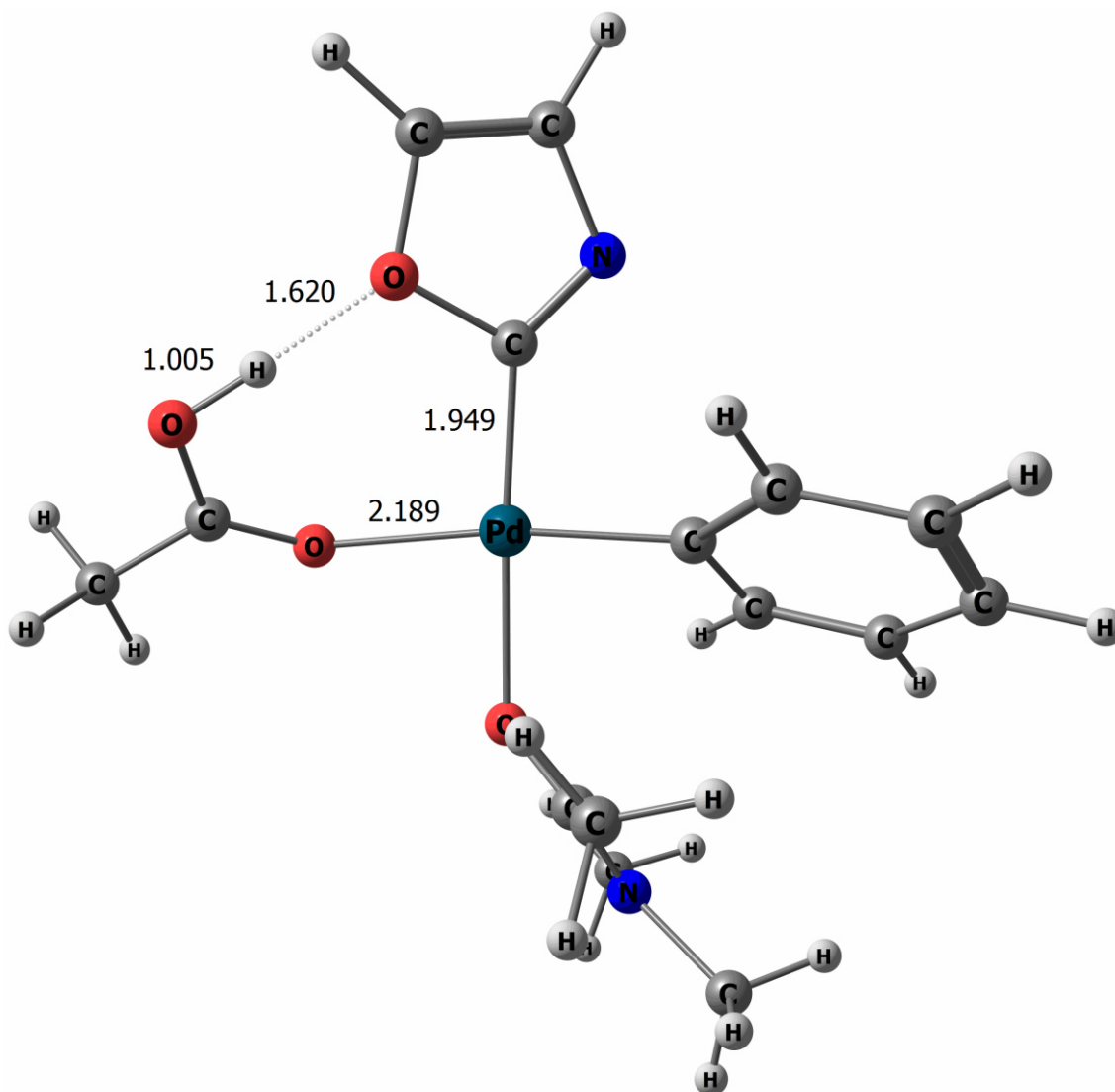


Figure 3-33: Intermediate 2 N2 Alternative – Oxygen of Oxazole facing the Acetate Ligand

No deprotonation takes place in this case and instead a normal hydrogen bond is observed. Thus, the same trends observed during the initial study of oxazole coordination persist. Namely, that the Nitrogen atom is much more basic than its oxygen counterpart. Comparing the energies of these two structures, reveals that Intermediate 2 N2 is 15.05 kcal.mol⁻¹ more stable than the alternative. Therefore, a large stabilization is achieved as a consequence of the deprotonation of the acetic acid ligand, implying that the alternative can be neglected.

The transition state associated with the next reaction along the pathway, reductive elimination (TS 2 Pathway N2), is depicted in Figure 3-34. This structure has an imaginary frequency of -367.09 cm⁻¹. The C–Pd–C angle between the two organometallic carbons is 56.21°. This is significantly smaller than the corresponding value of 91.35° in Intermediate 2. This together with the biaryl C–C distance of 1.872 Å is clear manifestation of a biaryl bond

formation process. The zwitterionic nature of the molecule is maintained in TS 2 and a clear hydrogen bond stabilises the TS with a O---H distance of 1.678 Å.

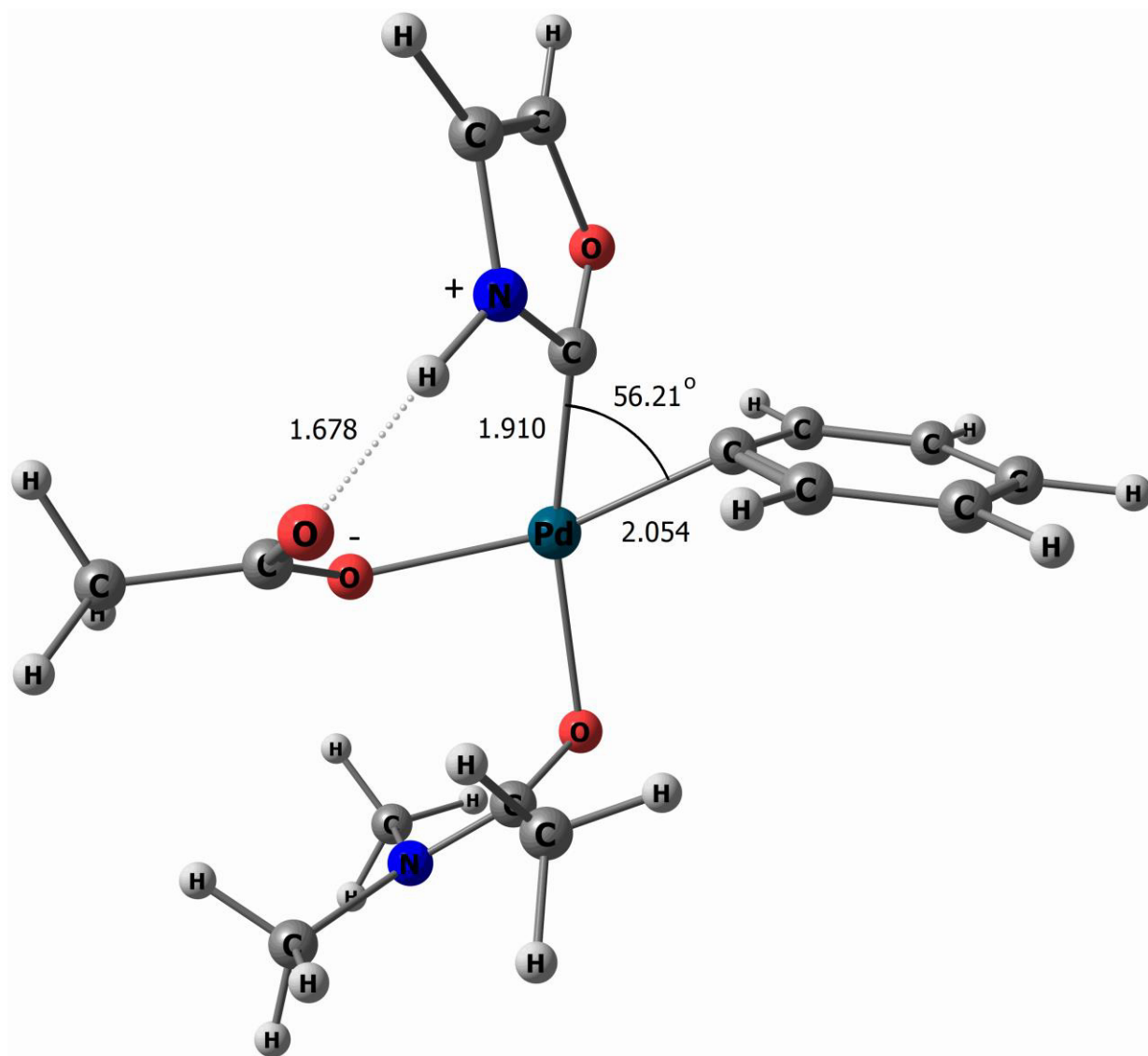


Figure 3-34: Transition State 2 N2

For all the pathways considered, the final structure of the surface scan that revealed TS 2 (in which the distance between the two organometallic carbons was shortened) corresponded to the lowest energy conformation in the scan. When this was optimised (the final structure in the aforementioned surface scan) at the TPSSh/SVP level, the resulting structure consisted of Product 2 still complexed to the now Pd(0) catalyst via a benzene aromatic bond (Figure 3-35). However, the acetic acid ligand is now no longer deprotonated, although there is still a hydrogen bond between the uncharged nitrogen atom and the acetate O-H group.

For the purposes of this investigation it was assumed that in reality the product will easily be displaced by solvent molecules, allowing the freed Pd(0) catalyst to be deprotonated by OH⁻ (see Chapter 5, Section 5.3). Subsequently, oxidative addition will take place when the deprotonated catalyst enters the cycle for a second time. It was assumed that there is no significant energy barrier associated with this dissociation of the product from the Pd(0) catalyst. Justification for this assumption can be drawn from the fact that experimentally the reaction is observed to take place catalytically and thus it is unlikely that the active catalyst ever becomes trapped in a potential energy well upon product formation, but rather is rapidly regenerated.

For all pathways considered, the product and Pd(0) catalyst were manually separated into separate .xyz files and optimised individually. Such a procedure resulted in the structure of Product 2 shown in Figure 3-36 as well as the protonated Pd(0) catalyst shown in Figure 3-37. In Product 2 the biaryl bond length of 1.457 Å matches exactly with the average from 145 CSD hits of oxazole analogues substituted with benzene derivatives at the 2 position.

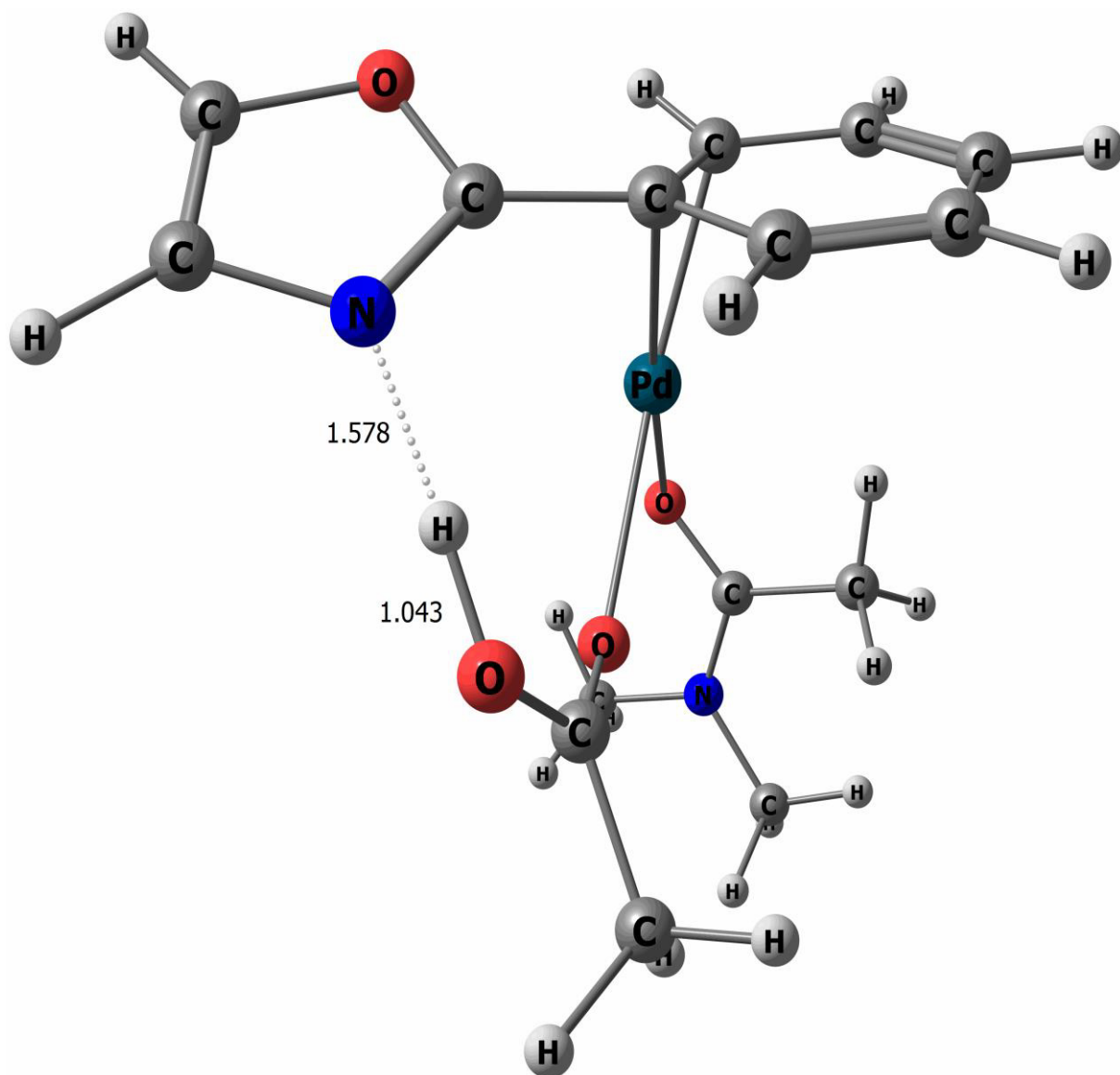


Figure 3-35: Product 2 Complexed to the Pd(0) Catalyst

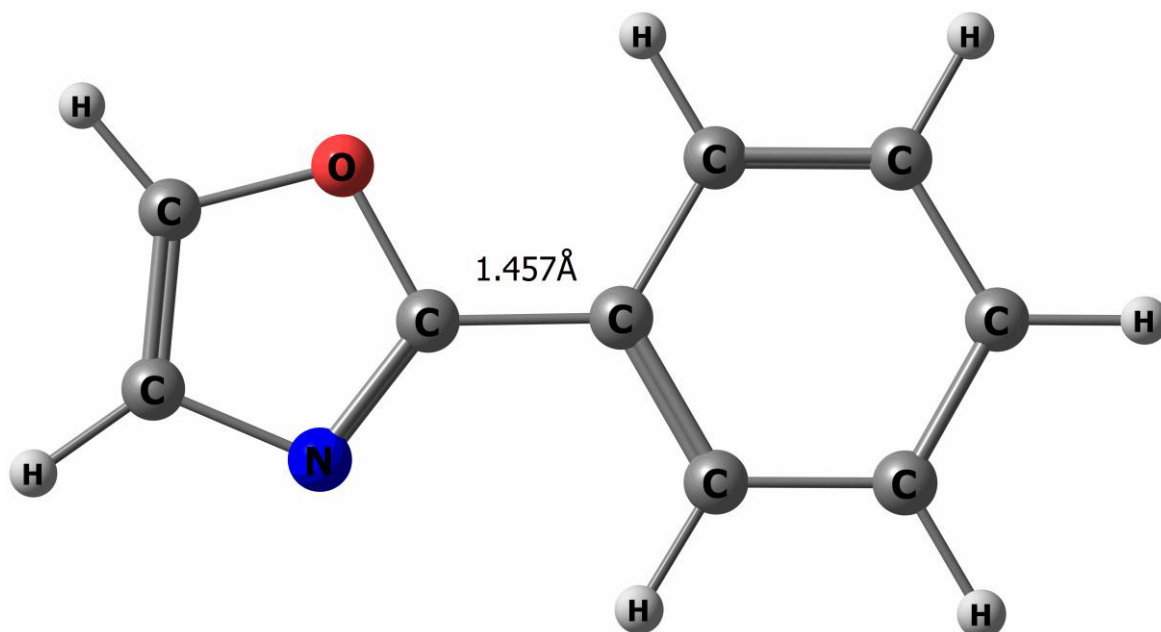


Figure 3-36: Product 2

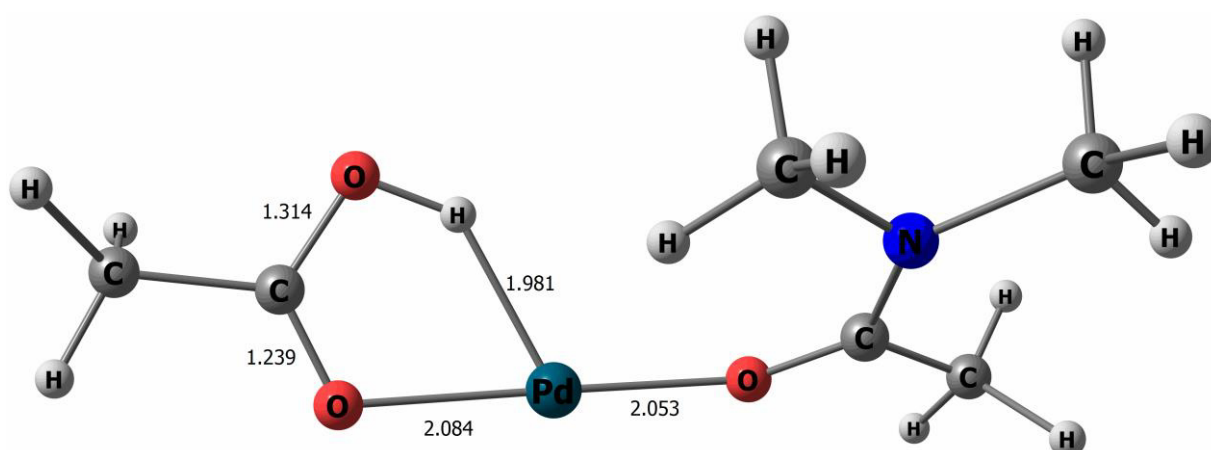


Figure 3-37: Protonated Pd(0) Catalyst

In the protonated Pd(0) catalyst the under-ligated palladium atom forms a bond with the O–H hydrogen atom of the acetate group. In reality, solvent molecules are likely to play a significant role in the structure of this molecule.

An overview of the CMD step:

When the acetate ligand gains sufficient energy to rotate and come close enough to the relevant hydrogen atom for deprotonation to take place, an organo-palladium bond will be formed by the electrons previously in the C–H bond, displacing the coordination bond in the process. This is shown in Figure 3-38 for Pathway N2, as an example. In this reaction, the C–H bond breaks with simultaneous formation of the C–Pd bond, hence *Concerted Metallation*

Deprotonation. The transition state associated with this reaction will be referred to as Transition State 1 (TS 1). The product of the CMD reaction, which contains two organometallic bonds, will henceforth be referred to as Intermediate 2 (Int 2).

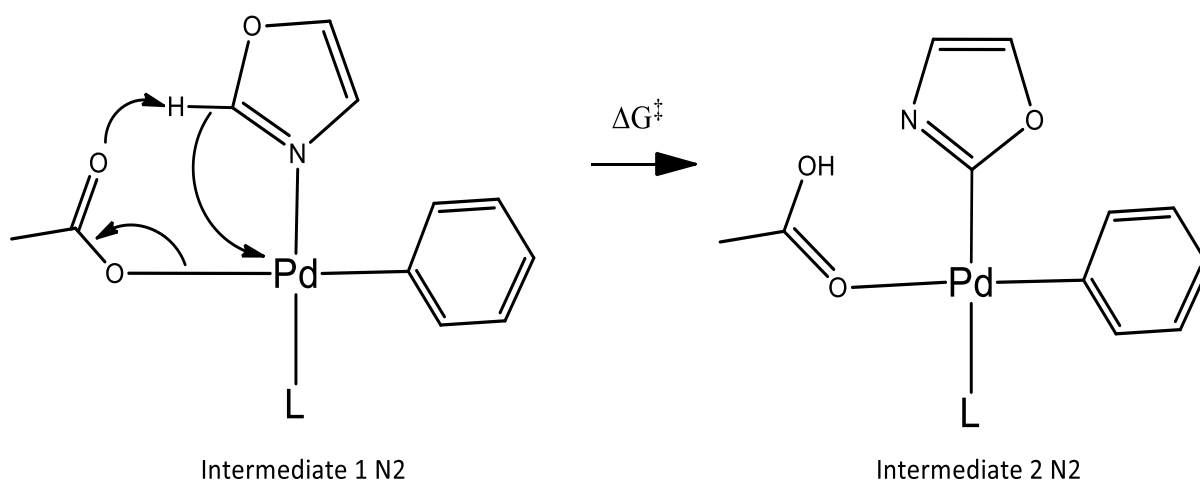


Figure 3-38: The formation of Intermediate 2 via a CMD reaction (Pathway N2)

For each of the reaction pathways to be discussed, the transition state associated with the CMD reaction was found by optimising the structure of the relevant Intermediate 2 molecule and then performing a NEB calculation (see Section 2.2.7) to determine the minimum energy pathway (MEP) between Intermediate 1 and Intermediate 2. The highest energy structure in the MEP pathway was then optimised, using a quasi-Newton like Hessian mode following algorithm, to reveal the transition state. Throughout the literature a characteristic 6 membered CMD transition state is reported when a carboxylate ligand is involved in deprotonation.¹⁶ This transition state possess a single imaginary frequency, the associated atomic movements of which correspond both to breaking of the C–H bond/ formation of the O–H bond as well as formation of the C–Pd bond. These atomic displacements are indicated by arrows in Figure 3-39, which shows the CMD transition state for Pathway N2.

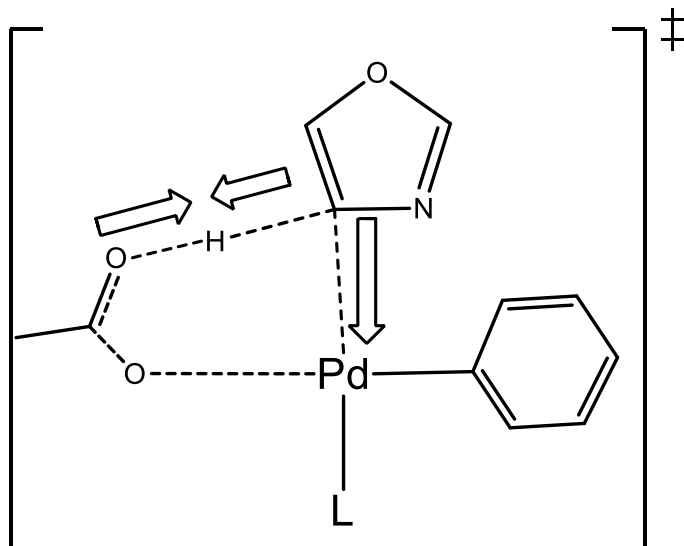


Figure 3-39: Atomic Displacements Associated with the CMD Transition State

Validation of the Model System:

Now that all the structures in this pathway have been presented and described, the energetics can be considered. In particular, it is still necessary to verify if the model system sufficiently reproduces the reaction Gibbs free energies ($\Delta_r G$) of the true system (i.e. toluene and pivalate in place of benzene and acetate, respectively). The $\Delta_r G$ values (TPSSH/SVP) for each step in the complete N₂ pathway for both the true and model systems are shown in Table 3-7. Entries leading to transition states, i.e. where the Gibbs energy of activation ($\Delta^\ddagger G$) is displayed, are highlighted in blue. The reaction zero-kelvin single point electronic energies ($\Delta_r E$) are also displayed. These values were shown simply to highlight the effects of considering Gibbs free energy, as opposed to just the electronic energy of the system, they will not be discussed further.

The $\Delta_r G$ Root Mean Square Deviation (RMSD), Equation (3. 1), between the model and true system is displayed in the same table. Graphically this information is presented in Figure 3-40. As can be seen from the graph, the two systems differ, in absolute terms, significantly only in the Gibbs free energy of TS 1 and Int 2, with all other points along the pathway remaining practically identical.

$$RMSD = \sqrt{\frac{\sum_{i=1}^n (\bar{y}_i - y_i)^2}{n}} \quad (3. 1)$$

Table 3-7: True vs Model System (TPSSh/SVP)

True System:

Chemical Reaction	$\Delta_r E$ (kcal.mol ⁻¹)	$\Delta_r G_{298K}$ (kcal.mol ⁻¹)
Pd(II) Catalyst + Oxazole $\rightarrow$ Int 1	-25.58	-7.02
Int 1 $\rightarrow$ TS 1	25.72	22.73
TS 1 $\rightarrow$ Int 2	-28.55	-24.12
Int 2 $\rightarrow$ TS 2	22.27	19.83
TS 2 $\rightarrow$ Product 2 + Pd(0) Catalyst	10.43	-7.28

Model System:

Chemical Reaction	$\Delta_r E$ (kcal.mol ⁻¹)	$\Delta_r G_{298K}$ (kcal.mol ⁻¹)
Pd(II) Catalyst + Oxazole $\rightarrow$ Int 1	-24.79	-7.12
Int 1 $\rightarrow$ TS 1	25.63	23.53
TS 1 $\rightarrow$ Int 2	-28.93	-27.66
Int 2 $\rightarrow$ TS 2	23.24	22.70
TS 2 $\rightarrow$ Product 2 + Pd(0) Catalyst	9.56	-7.46

$\Delta_r G_{298K}$ RMSD (kcal.mol ⁻¹)
2.07

The reaction $\Delta_r G$ RMSD value of 2.07 kcal.mol⁻¹ is very close to the inherent error in the method (2 kcal.mol⁻¹). Thus, there exists the fortuitous possibility that the two systems differ on average by only 0.07 kcal.mol⁻¹. To confirm that these results are not a consequence of basis set incompleteness and neglect of correlation, DSD-PBEP86/TZVP calculations were also performed (See Section 2.3). As a reminder, the thermodynamic corrects are taken from the TPSSh/SVP level of theory.

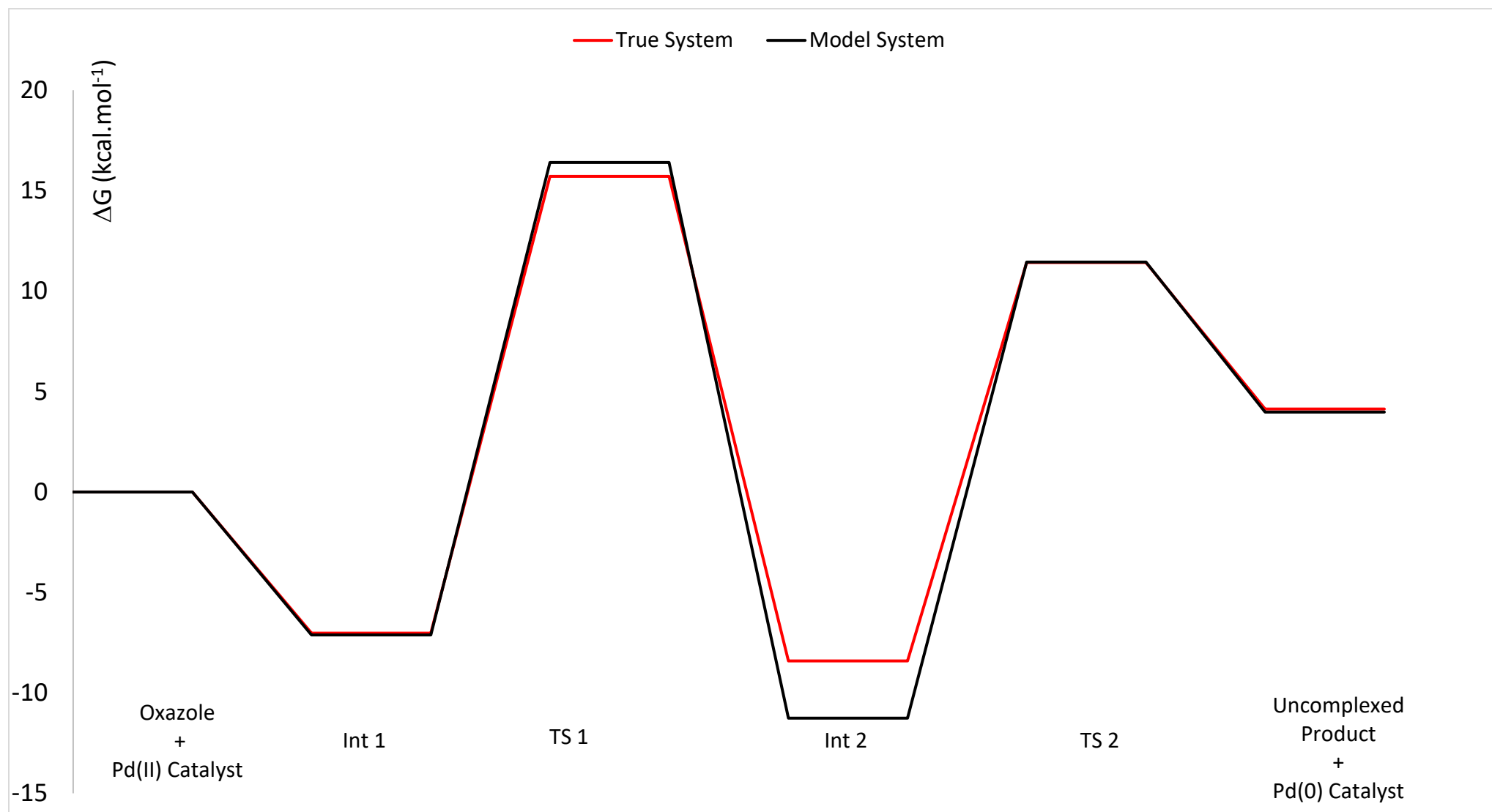


Figure 3-40: Gibbs Free Energies - Pathway N2 Model Vs True System (TPSSh/SVP)

These results are shown in Table 3-8 and represented graphically in Figure 3-41. Again, the model system reproduces the energy profile of the true system well. By observing Figure 3-41 it could be concluded that the absolute values of the two graphs differ more than at the Hybrid DFT level. However, the $\Delta_r G$ values of the two systems are actually closer at the Double Hybrid (DH) level, as can be seen from a RMSD value of 1.57 kcal.mol⁻¹. It was assumed throughout this work, as a rough estimate, that the absolute error in energy differences is around 1 kcal.mol⁻¹ for the system at the DSD-PBEP86/TZVP level of theory. Thus, the two systems possibly only differ on average by 0.57 kcal.mol⁻¹. The relative heights of the two transition states as well as the qualitative features of reaction profile are faithfully reproduced in the model system. Therefore, the model system will be employed throughout the remainder of the project.

Table 3-8: True vs Model System (DSD-PBEP86/TZVP)

True System:

Chemical Reaction	$\Delta_r E$ (kcal.mol ⁻¹)	$\Delta_r G_{298K}$ (kcal.mol ⁻¹)
Catalyst + Oxazole → Int 1	-21.80	-3.24
Int 1 → TS 1	28.75	25.76
TS 1 → Int 2	-32.82	-28.39
Int 2 → TS 2	20.95	18.51
TS 2 → Product 2 + Pd(0) Catalyst	7.02	-10.70

Model System:

Chemical Reaction	$\Delta_r E$ (kcal.mol ⁻¹)	$\Delta_r G_{298K}$ (kcal.mol ⁻¹)
Catalyst + Oxazole → Int 1	-21.56	-3.90
Int 1 → TS 1	28.01	25.90
TS 1 → Int 2	-32.36	-31.09
Int 2 → TS 2	20.70	20.16
TS 2 → Product 2 + Pd(0) Catalyst	7.68	-9.34

RMSD
1.57

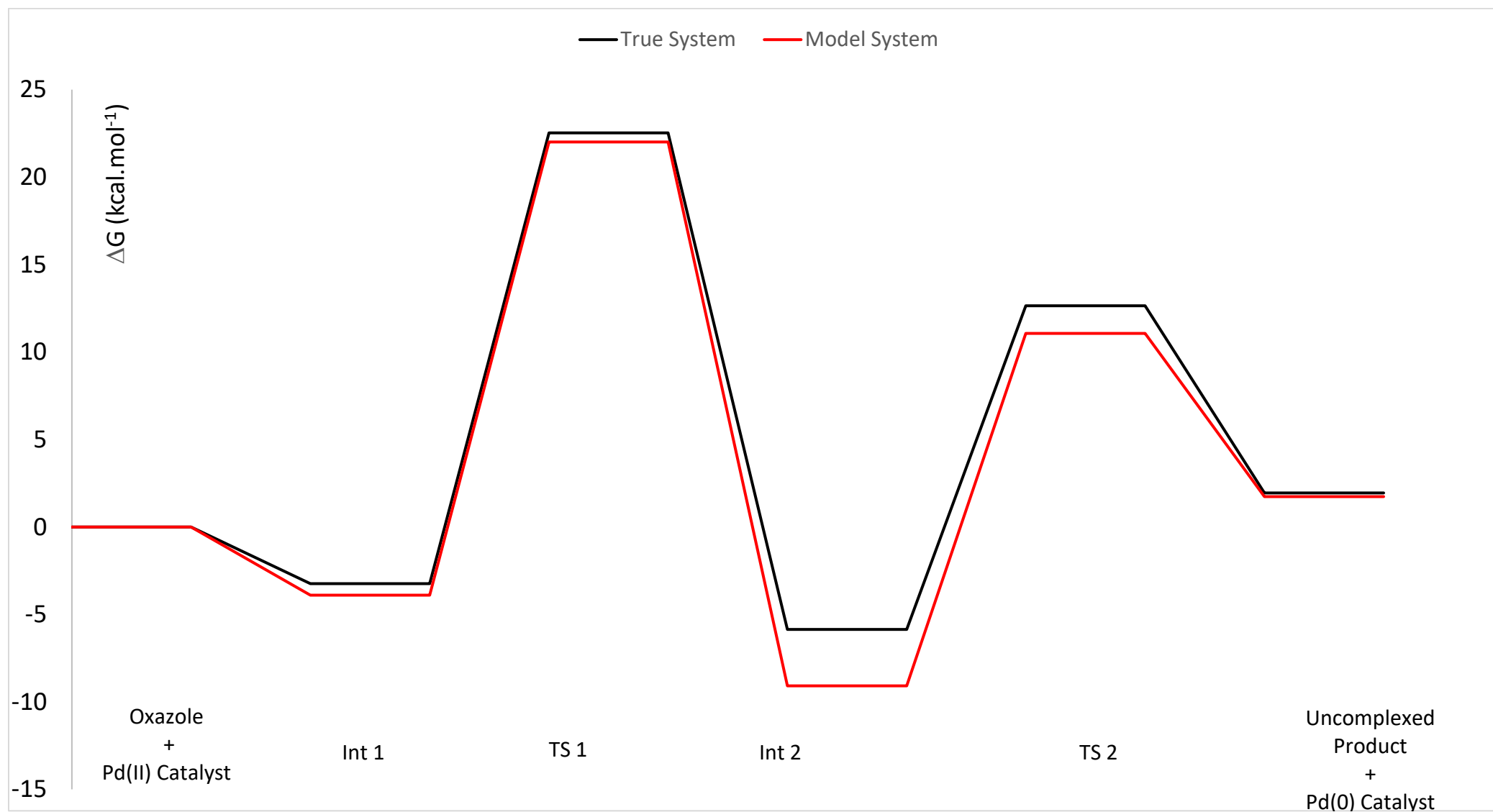


Figure 3-41: Gibbs Free Energies -Pathway N2 Model Vs True System (DSD-PBEP86/TZVP)

Pathway N4:

In a similar fashion to Pathway N2, Intermediate 1 N4 can undergo the same sequence of transformations to produce an associated TS 1 (Figure 3-42), Int 2 (Figure 3-43), TS 2 (Figure 3-44) and ultimately Product 4 (Figure 3-45). In this case TS 1 has a negative frequency of -428.21 cm^{-1} . Intermediate 2 N4 shows a strong intramolecular hydrogen bond, between the hydroxyl hydrogen of the acetate and the nitrogen of oxazole. However, unlike in Intermediate 2 N2, no deprotonation of the acetate ligand takes place. This hydrogen bond persists in TS 2, which has a negative frequency of -340.36 cm^{-1} .

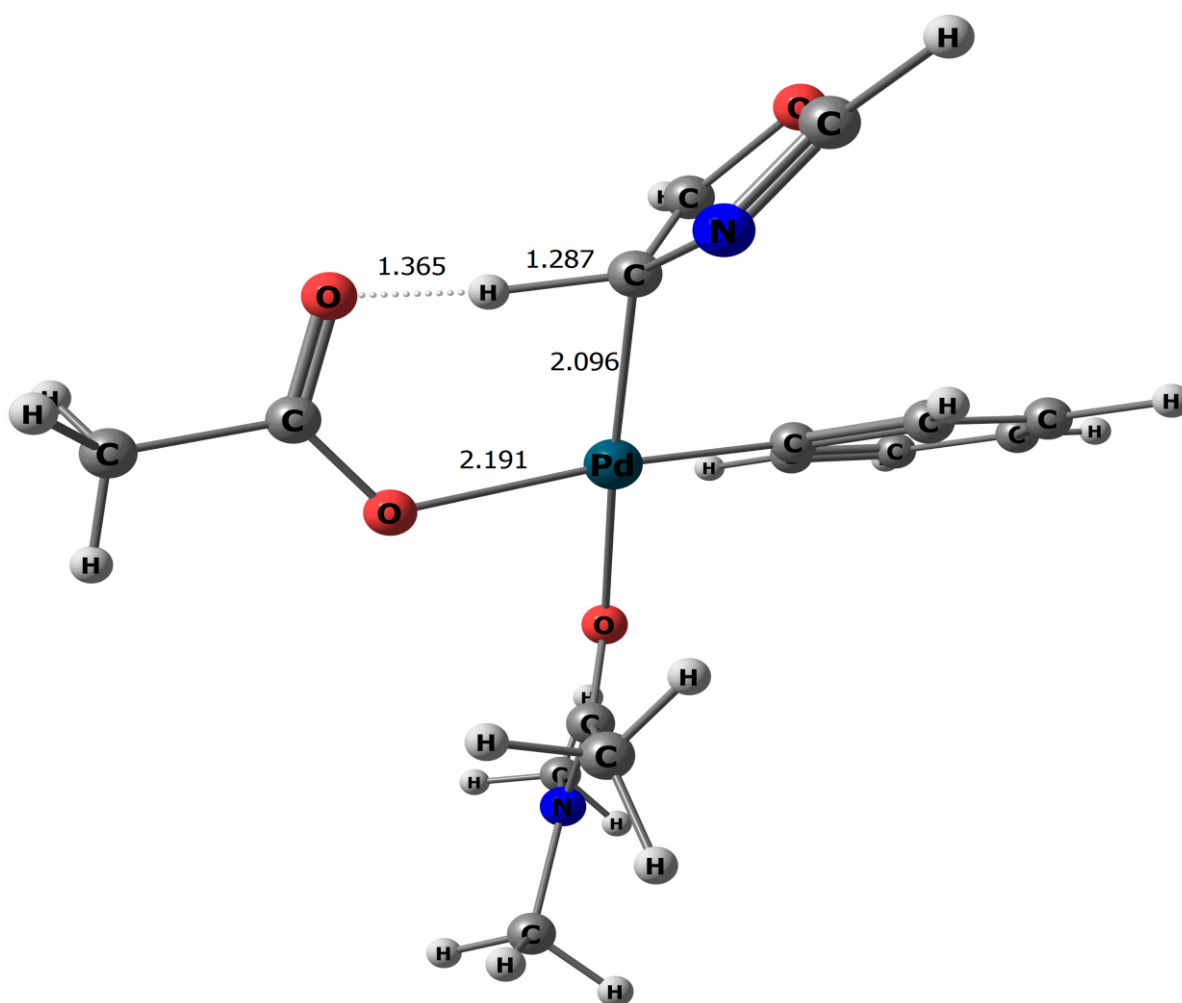
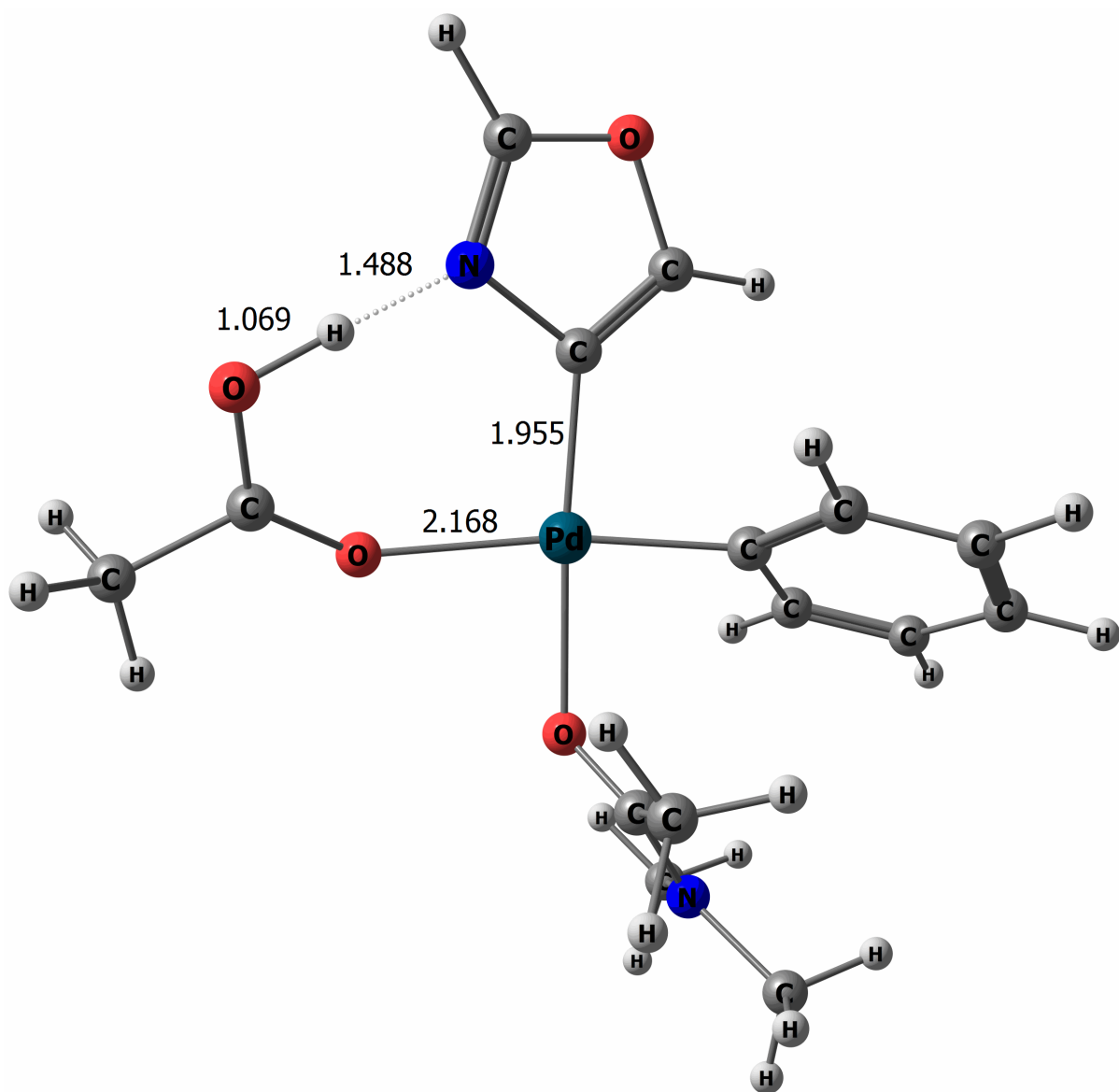
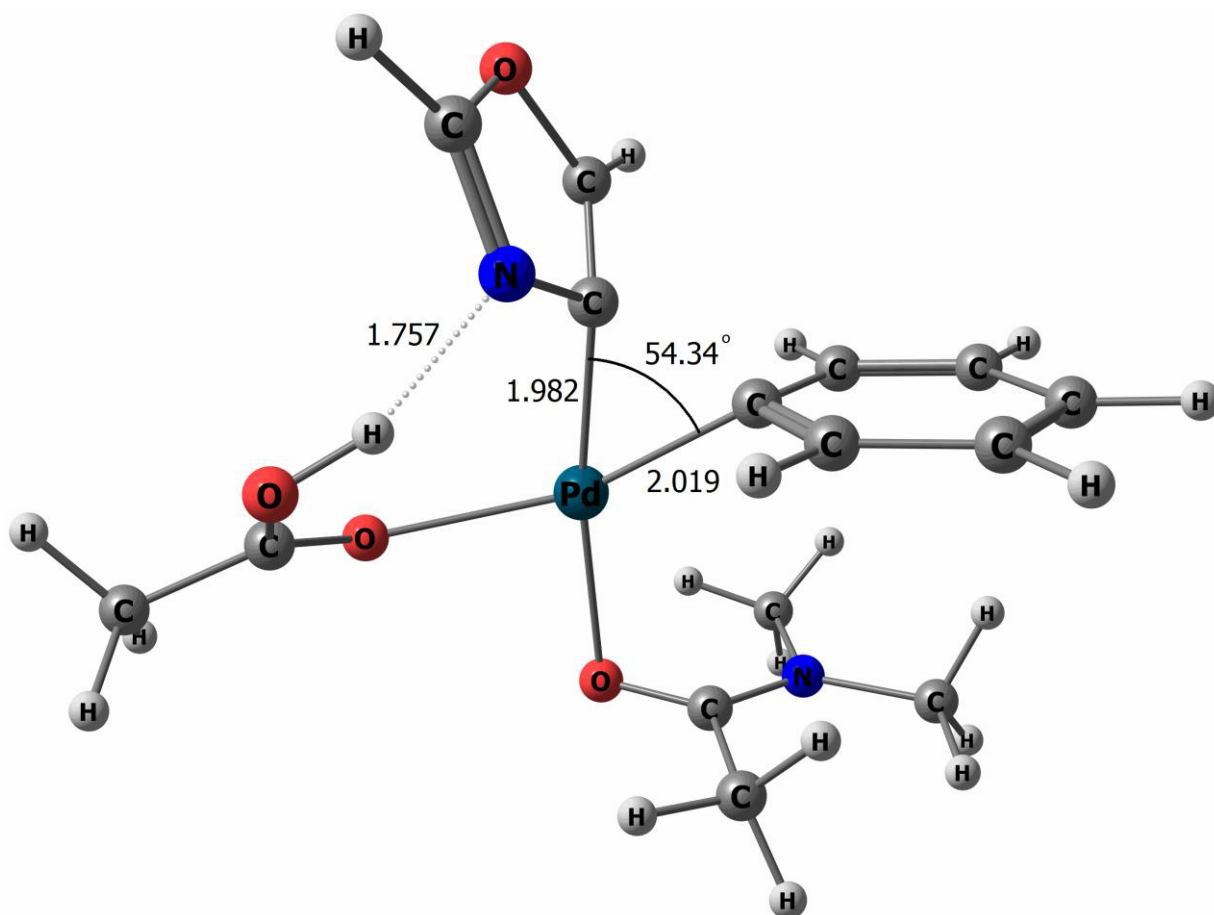


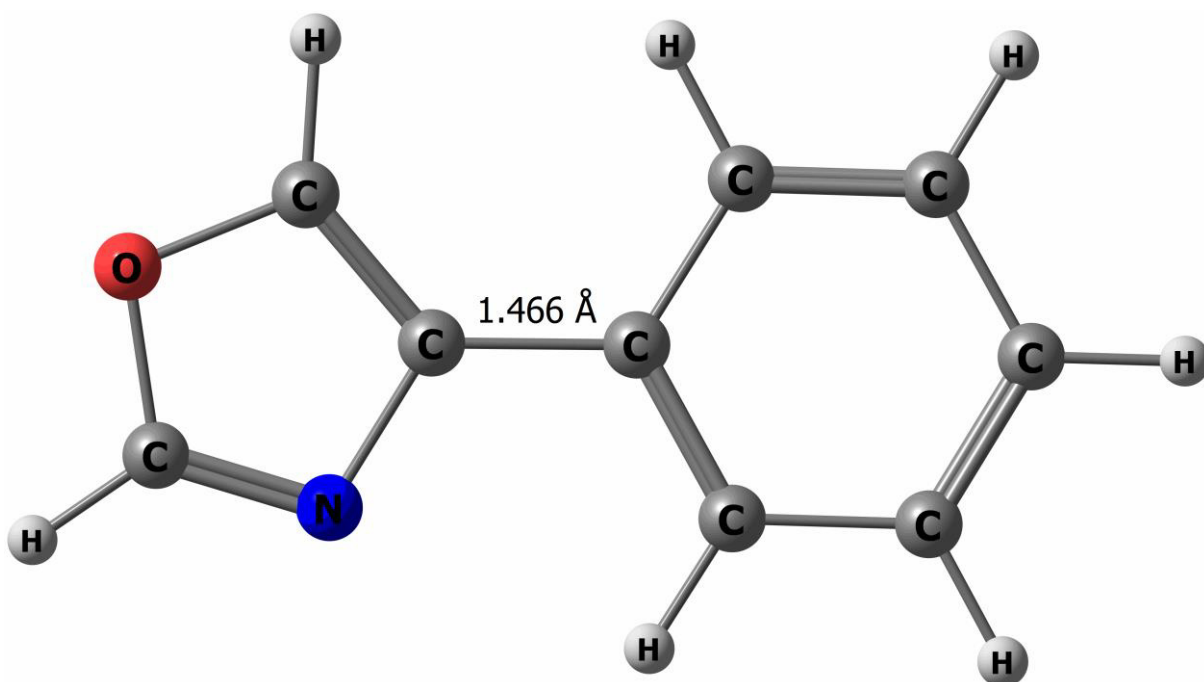
Figure 3-42: Transition State 1 N4



[Figure 3-43: Intermediate 2 N4](#)



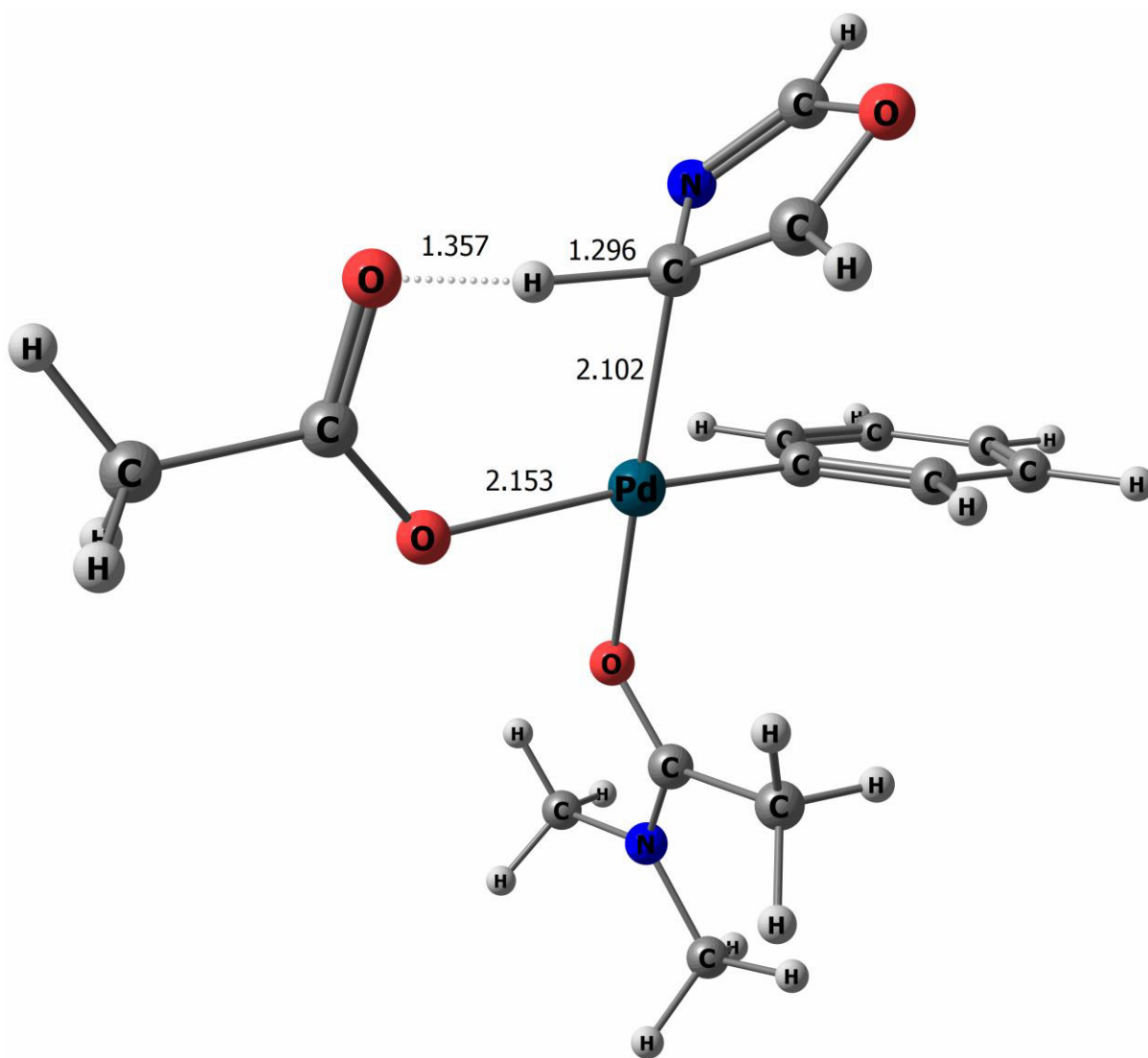
[Figure 3-44: Transition State 2 N4](#)



[Figure 3-45: Product 4](#)

Pathway =4:

TS 1 for the pathway starting at Intermediate =4 is illustrated in Figure 3-46 and has a negative frequency of -539.59 cm^{-1} . At this point, Pathways N4 and =4 actually merge. As can be seen by a comparison of Figure 3-42 and Figure 3-46, TS 1 N4 and TS 1 =4 differ significantly only in the orientation of oxazole. A rotation of the deprotonated oxazole molecule by 180° about an axis bisecting carbon 4 and the O–C2 bond converts one into the other. After deprotonation, oxazole may assume an arbitrary conformation depending on whether rotation about the newly formed C4–Pd bond takes place clockwise or anticlockwise. Starting from TS 1 =4, rotation in the direction of an arrow pointing from the acetate to the benzene ligand (Figure 3-47) leads to what will be referred to as Int 2 =4 (Figure 3-48). Rotation in the opposite direction simply results in Intermediate 2 N4 (Figure 3-43). In fact, a geometry optimisation of Int 2 =4 results in oxazole automatically rotating about the C4–Pd bond and the final optimised structure corresponds to Intermediate 2 N4. This optimisation can be found in the ESI. The driving force behind this rotation is the formation of the aforementioned hydrogen bond. Thus Pathway =4 shares the same Intermediate 2, Transition State 2 and product as Pathway N4.



[Figure 3-46: Transition State 1=4](#)

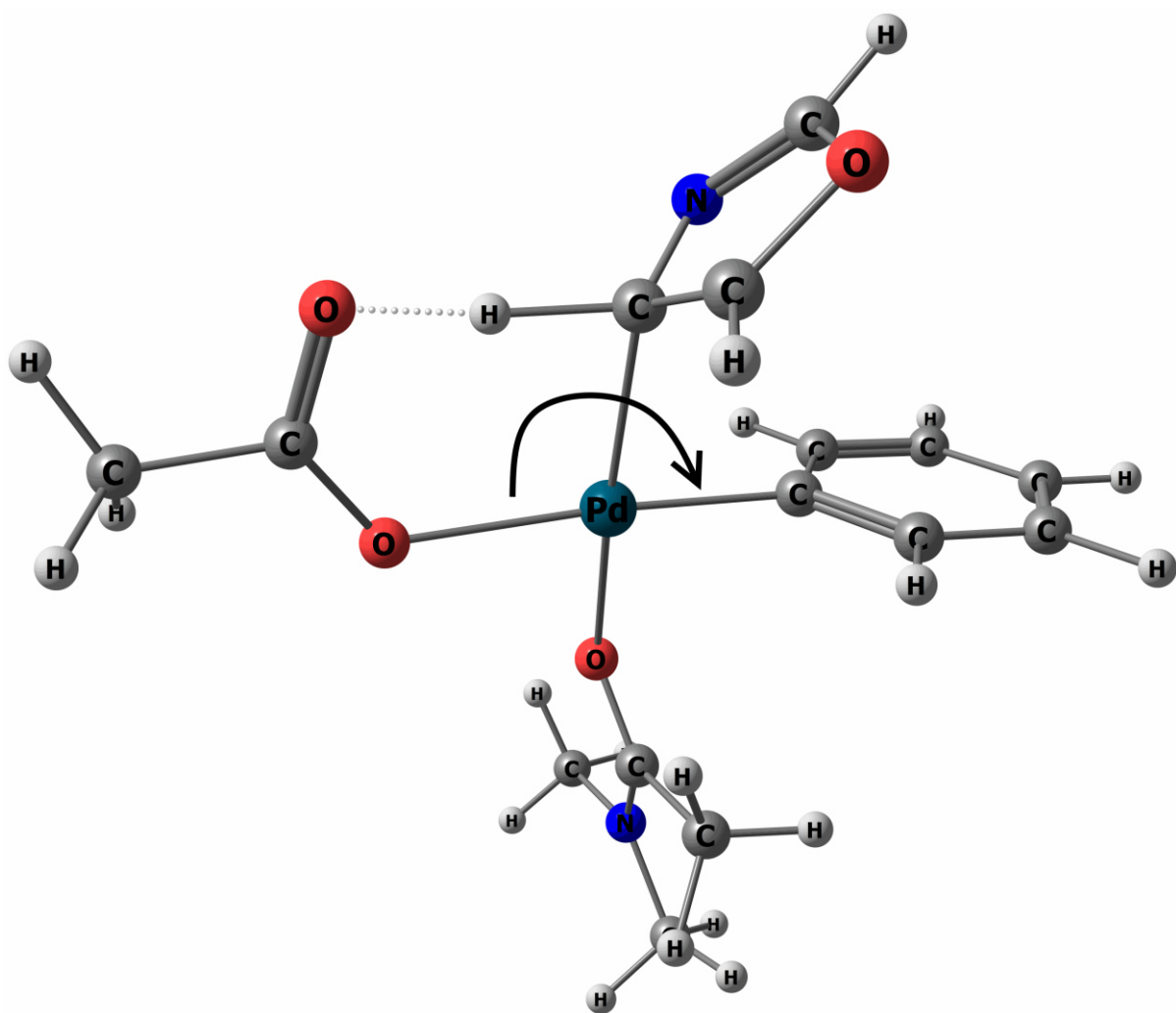


Figure 3-47: Rotation leading to Intermediate 2 =4

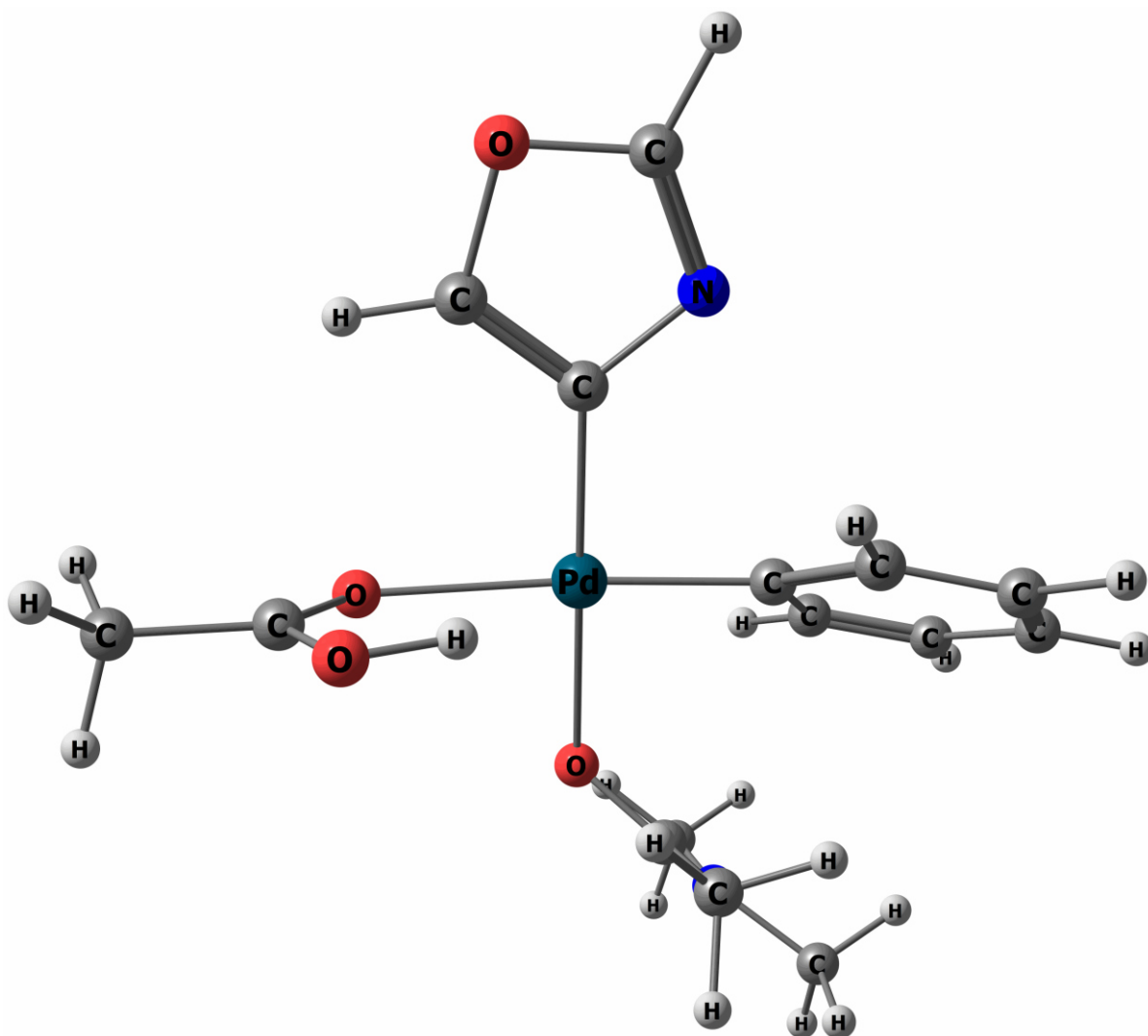
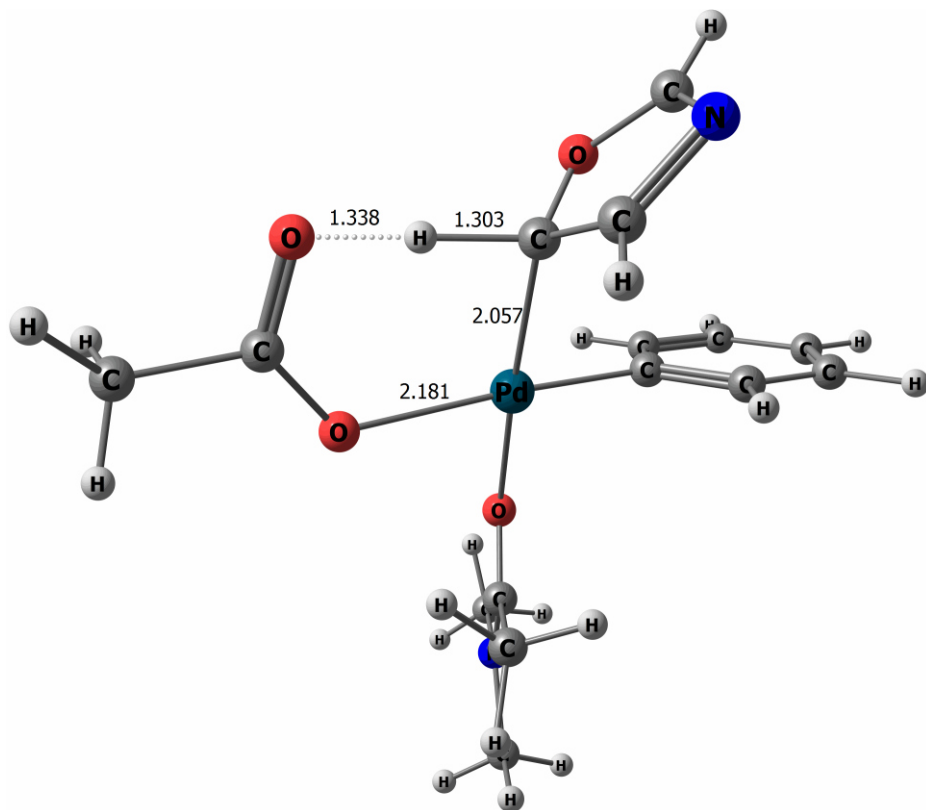


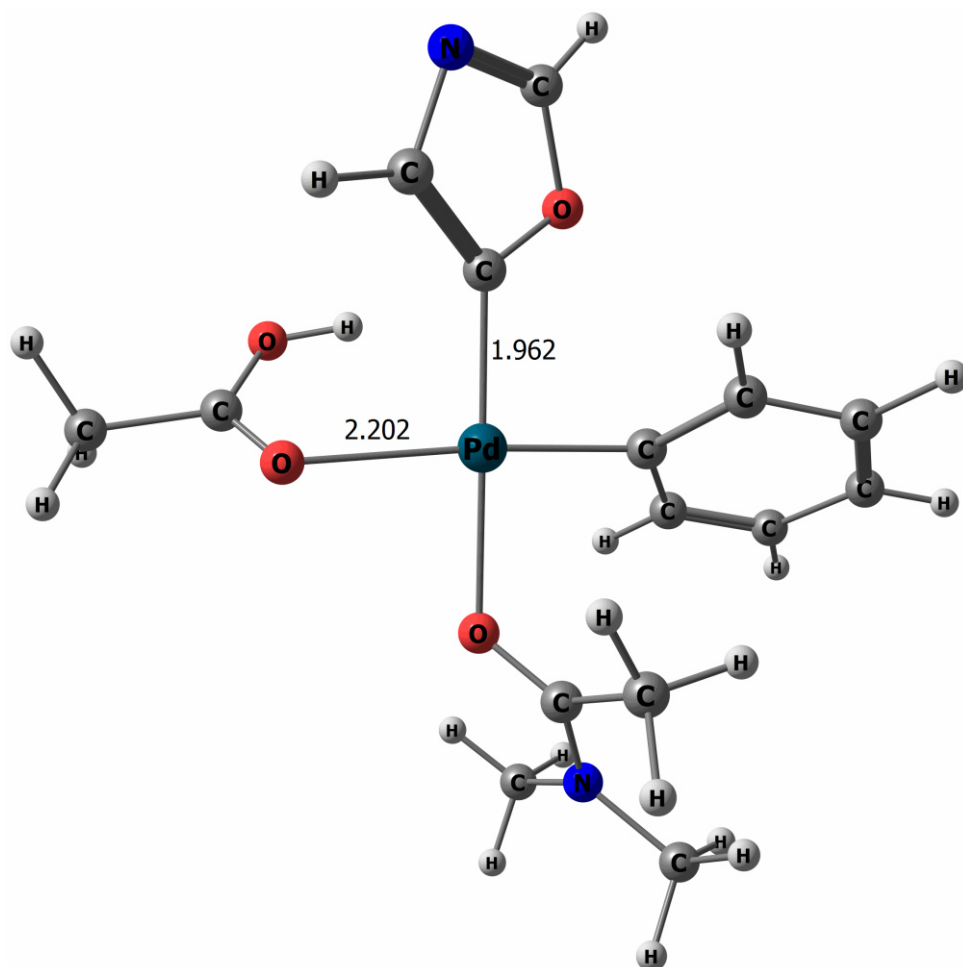
Figure 3-48: Intermediate 2 =4

Pathway =5:

TS 1 (Figure 3-49), Int 2 (Figure 3-50), TS 2 (Figure 3-52) as well as Product 5 (Figure 3-53) are shown for the pathway that initiates from a reaction of Intermediate 1 =5. For this pathway, TS 1 has a negative frequency of -557.076 cm^{-1} . In Intermediate 2, no hydrogen bond is observed between the oxygen atom of oxazole and the hydroxyl group of the acetate ligand. In fact, oxazole is rotated so that the oxygen atom faces away from the acetate ligand. An alternative structure with H-bonding was investigated and the optimised structure is shown in Figure 3-51. However, the repulsion between the two oxygen atoms seems to override any lowering of energy as a consequence of the hydrogen bond, as this alternative structure is slightly higher in energy by $0.33\text{ kcal.mol}^{-1}$. The difference is well within the error of the method, however. TS 2 has a negative frequency of -348.16 cm^{-1} .



[Figure 3-49: Transition State 1 =5](#)



[Figure 3-50: Intermediate 2 =5](#)

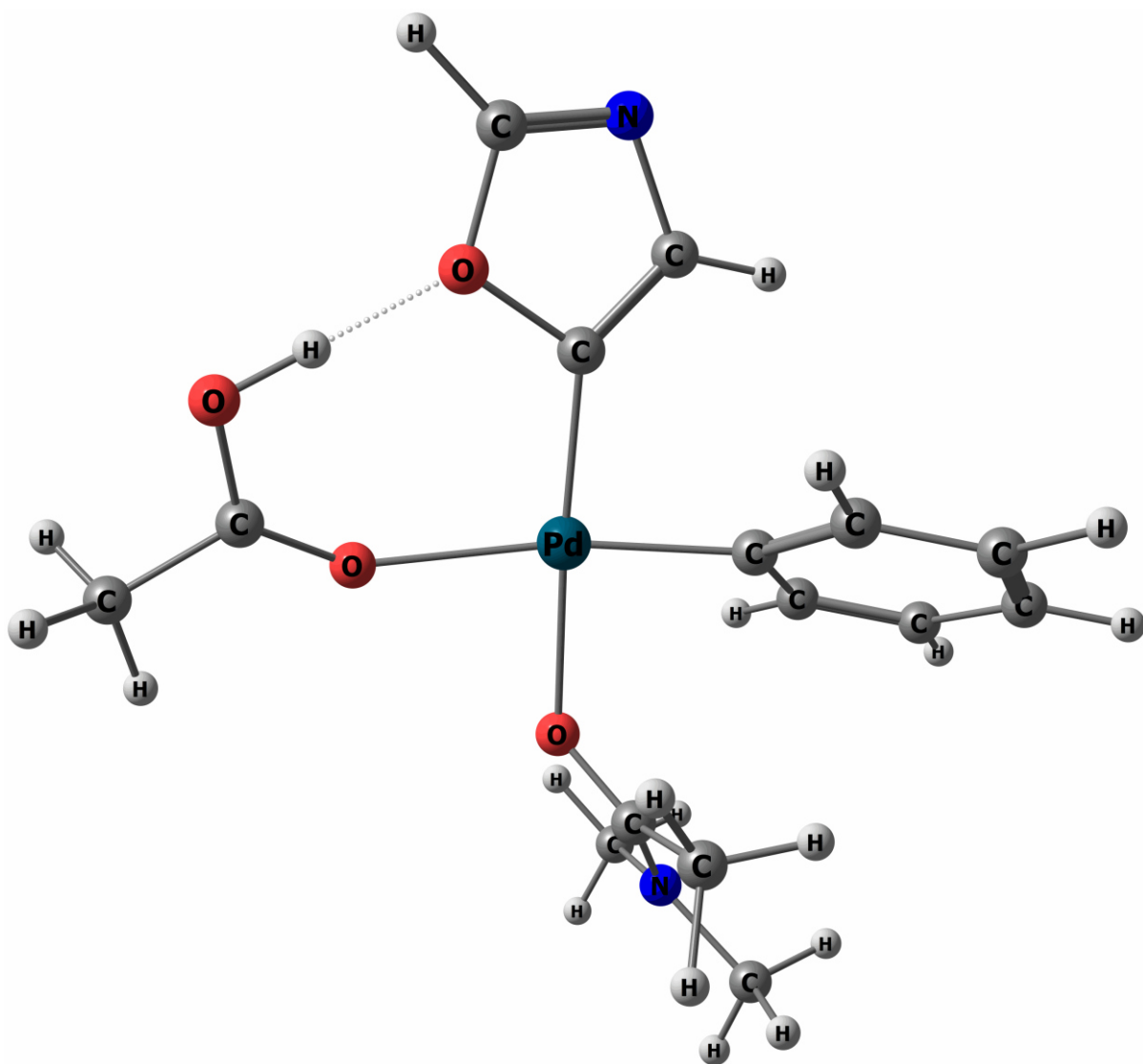


Figure 3-51: Intermediate 2 =5 Alternative

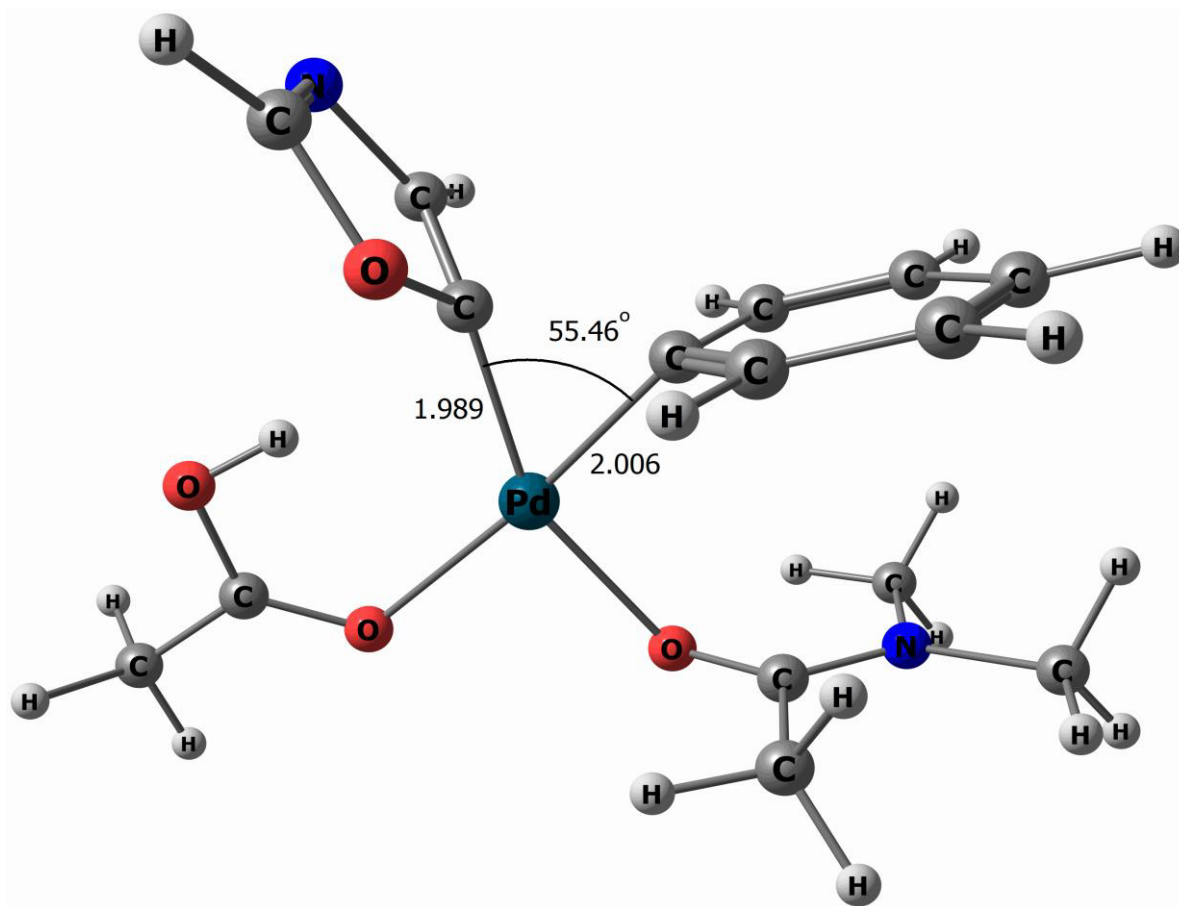


Figure 3-52: Transition State 2 =5

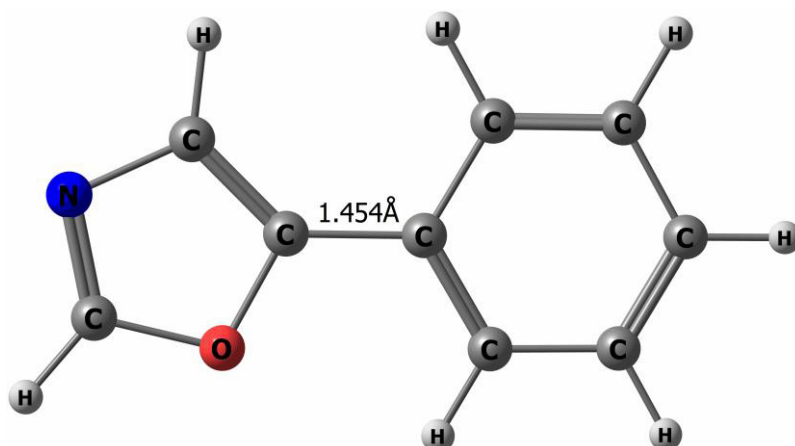


Figure 3-53: Product 5

TSSSh/SVP Results:

It is advantageous for industrial processes to be performed at the lowest possible temperatures. This is in order to minimise energy costs, safety concerns and the associated strain on equipment, among other factors. Faster kinetics are advantageous as they result in the quicker formation of products and thus revenue. It is for these reasons that the experimental results that are the focus of this project were performed under conditions of

kinetic control in which the reaction temperature and time were carefully chosen to allow the oxidative addition of 4-bromotoluene to go to completion. If it is assumed that equilibrium was not established during the remainder of the reactions, then the observed product distributions are a reflection of the kinetics. Therefore, the relative proportion in which two products (P) have been produced at a given stage of the reaction can be expressed by the ratio of the two rates and therefore the two rate constants (k) as shown in Equation (3. 2)

$$\frac{[P_2]}{[P_1]} = \frac{k_2}{k_1} \quad (3. 2)$$

According to the Eyring equation, Equation (3. 3), the rate constant of a chemical reaction depends exponentially on the Gibbs energy required to reach the transition state, i.e. the Gibbs energy of activation, $\Delta^\ddagger G$. In this equation k_r is the rate constant for reaction r , k_B is Boltzmann's constant, T is the reaction temperature, h is Planck's constant and R is the gas constant.

$$k_r = \frac{k_B T}{h} e^{-\Delta^\ddagger G / RT} \quad (3. 3)$$

However, the rate is also exponentially dependent on any errors in calculated $\Delta^\ddagger G$ values. Small errors in $\Delta^\ddagger G$ values will therefore result in large errors in any rate constants calculated in this way. In order to avoid conclusions based on erroneous rates, throughout this project, discussions will be restricted to the Gibbs free energy values themselves.

The Gibbs free energies of the considered structures relative to the infinitely separated Pd(II) catalyst and oxazole (i.e. Pd(II) catalyst + oxazole = 0 kcal.mol⁻¹) are shown in Figure 3-54. The corresponding reaction Gibbs free energies, for each process considered, are shown in Table 3-9. Entries leading to transition states, i.e. where the Gibbs energy of activation ($\Delta^\ddagger G$) is displayed, are highlighted in blue.

Table 3-9 : Mono-substitution Reaction Free Energies (TPSSh/SVP)

Chemical Reaction	ΔE (kcal.mol ⁻¹)	ΔG_{298K} (kcal.mol ⁻¹)
<u>Pathway N2</u>		
Pd(II) Catalyst + Oxazole $\rightarrow$ Int 1	-24.79	-7.12
Int 1 $\rightarrow$ TS 1	25.63	23.53
TS 1 $\rightarrow$ Int 2	-28.93	-27.66
Int 2 $\rightarrow$ TS 2	23.24	22.70
TS 2 $\rightarrow$ Product 2 + Pd(0) Catalyst	9.56	-7.46
<u>Pathway N4</u>		
Pd(II) Catalyst + Oxazole $\rightarrow$ Int 1	-24.33	-7.66
Int 1 $\rightarrow$ TS 1	26.40	24.14
TS 1 $\rightarrow$ Int 2	-24.27	-21.06
Int 2 $\rightarrow$ TS 2	26.73	25.69
TS 2 $\rightarrow$ Product 4 + Pd(0) Catalyst	2.39	-14.99
<u>Pathway =4</u>		
Pd(II) Catalyst + Oxazole $\rightarrow$ Int 1	-20.57	-2.57
Int 1 $\rightarrow$ TS 1	24.92	20.42
TS 1 $\rightarrow$ Int 2	-26.56	-22.44
Int 2 $\rightarrow$ TS 2	26.73	25.69
TS 2 $\rightarrow$ Product 4 + Pd(0) Catalyst	2.39	-14.99
<u>Pathway =5</u>		
Pd(II) Catalyst + Oxazole $\rightarrow$ Int 1	-21.10	-4.65
Int 1 $\rightarrow$ TS 1	19.33	17.38
TS 1 $\rightarrow$ Int 2	-11.84	-9.99
Int 2 $\rightarrow$ TS 2	16.30	16.17
TS 2 $\rightarrow$ Product 5 + Pd(0) Catalyst	3.46	-13.60

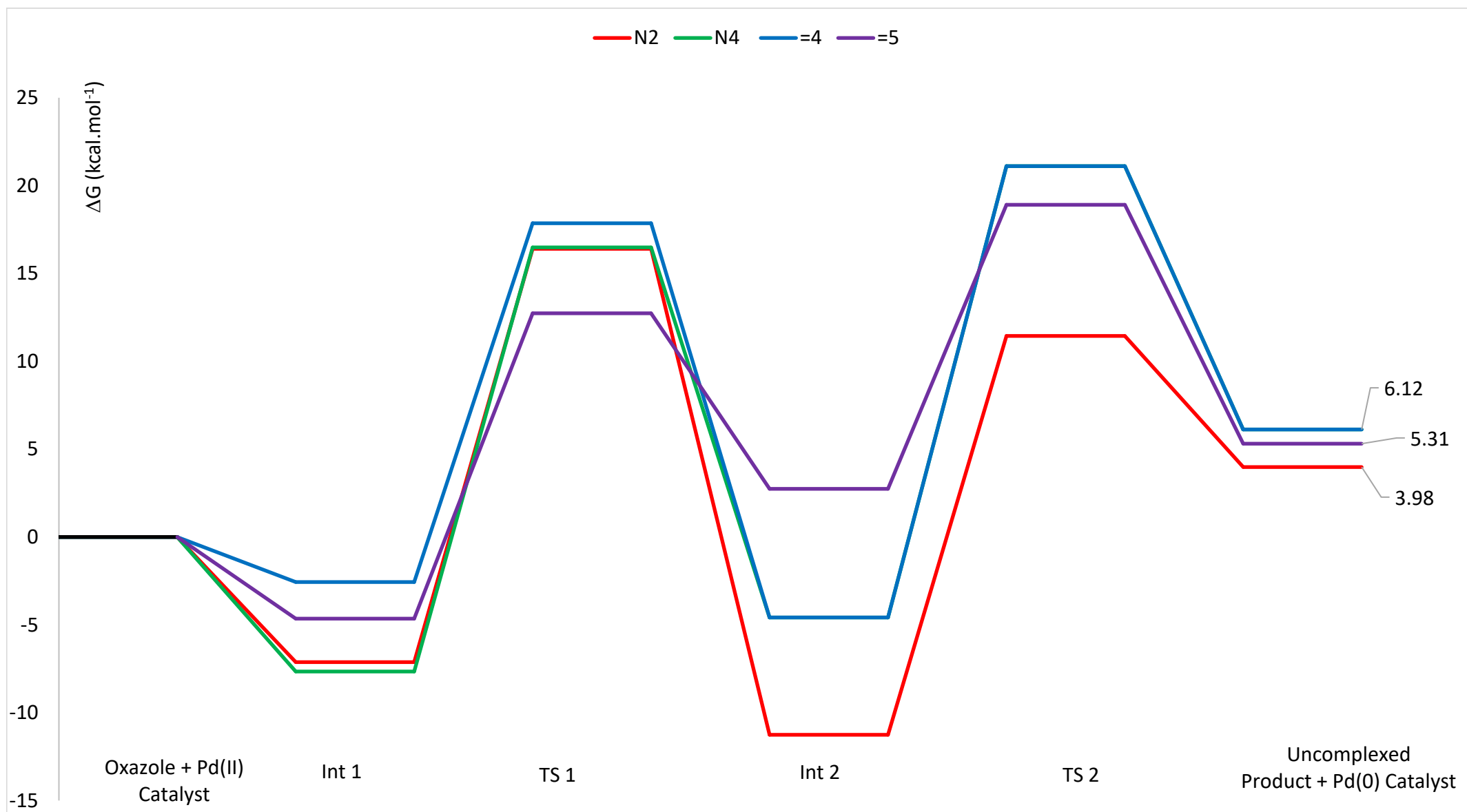


Figure 3-54: Mono-substitution Reaction Profiles Under Phosphine Free Conditions (TPSSh/SVP)

Considering the reaction profiles in general it can be seen that across the pathways there is no clear rate determining step (RDS), but rather that concerted metallation deprotonation and reductive elimination are both energetically relevant processes that need to be considered. This is in contrast to what is reported in literature,⁸ where the CMD step is frequently reported to be rate limiting.

The differences in the Gibbs energy of activation ($\Delta^\ddagger G$), between the two transition states for each pathway are shown in Table 3-10.

Table 3-10: CMD (TS 1) vs Reductive Elimination (TS 2) Gibbs Energies

Pathway	$\Delta^\ddagger G_{TS\ 1} - \Delta^\ddagger G_{TS\ 2}$ (kcal.mol ⁻¹)
N2	0.82
N4	-1.55
=4	-5.27
=5	1.21

The energy required to reach TS 1 (CMD) and TS 2 (RE) is similar in all cases except for Pathway =4. The difference of -5.27 kcal.mol⁻¹ here is significant and it is suffice to say that for Pathway =4, TS 2 is the bottle neck process and RE is largely responsible for determining the speed of product formation, at least within the confines of this level of theory. The differences between TS 1 and TS 2 for Pathways N2 (0.82 kcal.mol⁻¹), N4 (-1.55 kcal.mol⁻¹) and =5 (1.21 kcal.mol⁻¹) are within the error of the level of theory used and thus should be viewed with caution. There is no general trend as to the higher energy transition state across pathways, with Pathways N2 and =5 requiring more energy to reach TS 1 while Pathways N4 and =4 require more energy to reach TS 2.

Analysing Gibbs energy required for the RDS, Table 3-11, reveals that Pathway =5 is the most favourable by a significant margin, requiring 6.15 kcal.mol⁻¹ less Gibbs energy than the next lowest pathway, Pathway N2. Product 5 is thus expected to be the major experimental product under conditions of kinetic control.

Table 3-11: Pathway Favourability in Terms of the RDS (TPSSH/SVP)

Favorability	Pathway	RDS	ΔG RDS (kcal.mol ⁻¹)
1st	=5	CMD (TS 1)	17.38
2nd	N2	CMD (TS 1)	23.53
3 rd /4 th	N4/=4	RE (TS 2)	25.69

Product 5 indeed represents 65.8 % of the total experimental product distribution and is the only significant mono substituted product observed, as seen in Table 3-12.

Table 3-12: Experimentally Observed Mono Substituted Product Distributions - Phosphine Free Conditions

Catalyst	Ligand	Mono-substituted Product Distribution (%)	
		Product 2	Product 5
Pd(OAc) ₂	None	0.3	65.8

The RDS steps for the remaining three pathways are higher in Gibb's energy requirements, and occupy the relatively narrow range of 23.53 - 25.69 kcal.mol⁻¹. Pathways N2 and N4/=4 are separated by only 2.16 kcal.mol⁻¹. At this level of theory, for all practical intents and purposes Pathways =4 and N2 can be considered to require the same Gibbs energy. Thus, the rate of the formation of Products 2 and 4 is expected to be insignificant compared to the rate of formation of Product 5.

If the small amount of Product 2 observed experimentally (0.3%) was due to the rate of the reaction for Pathway N2, a corresponding amount of Product 4 should also have been detected, as the total G requirements for Pathways N4/=4 and N2 are very similar. However, no Product 4 was observed. Additionally, as mentioned above, Pathway =5 is favorable to the extent that the rates of the remaining pathways are expected to be insignificant in comparison, under conditions of kinetic control.

A possible explanation for the small amount of Product 2 can be found by considering the overall Gibbs energy change for the reactions. According to the definition of the thermodynamic equilibrium constant, Equation (3. 4), the negative of the overall reaction Gibbs free energy ($\Delta_r G$) is equal to the product of the gas constant (R), the temperature (T) and the natural logarithm of the equilibrium constant (K).

$$-\Delta_r G = RT \ln K \quad (3. 4)$$

Thus K can be expressed as:

$$K = e^{\frac{-\Delta_r G}{RT}} \quad (3. 5)$$

From Equation (3. 5) it can be seen that the equilibrium constant is inversely proportional to the reaction Gibbs free energy. The more negative $\Delta_r G$, the more the equilibrium will lie to the right and the higher the ratio of products to reactants. As with the rate constant, the

equilibrium constant depends exponentially on the reaction Gibbs free energy, and thus is also exponentially dependent on any errors in calculated $\Delta_r G$ values. The discussion will therefore be restricted to the $\Delta_r G$ values themselves.

The overall reaction for the formation of all three products is endergonic (Figure 3-54). The $\Delta_r G$ values for Product 4, Product 5 and Product 2 are 6.12, 5.31 and 3.98 kcal.mol⁻¹, respectively. However, the bare Pd(II) catalyst that is being considered (Figure 3-37) is under ligated and in reality, will be stabilized by the coordination of solvent molecules to palladium. As no implicit coordination of more than one DMA molecule was considered, the endergonic nature of the overall reactions should be viewed with some caution. Evidence of this can be found when considering explicit solvation of 3 DMA solvent molecules around the Pd(0) catalyst. An optimized geometry of this system is presented in Figure 3-55. The solvate system is 8.61 kcal.mol⁻¹ more stable than the infinitely separated Pd(0) catalyst and three DMA molecules. This stabilization alone would result in all reactions becoming exergonic, even though the effects of solvation on the other entities is neglected.

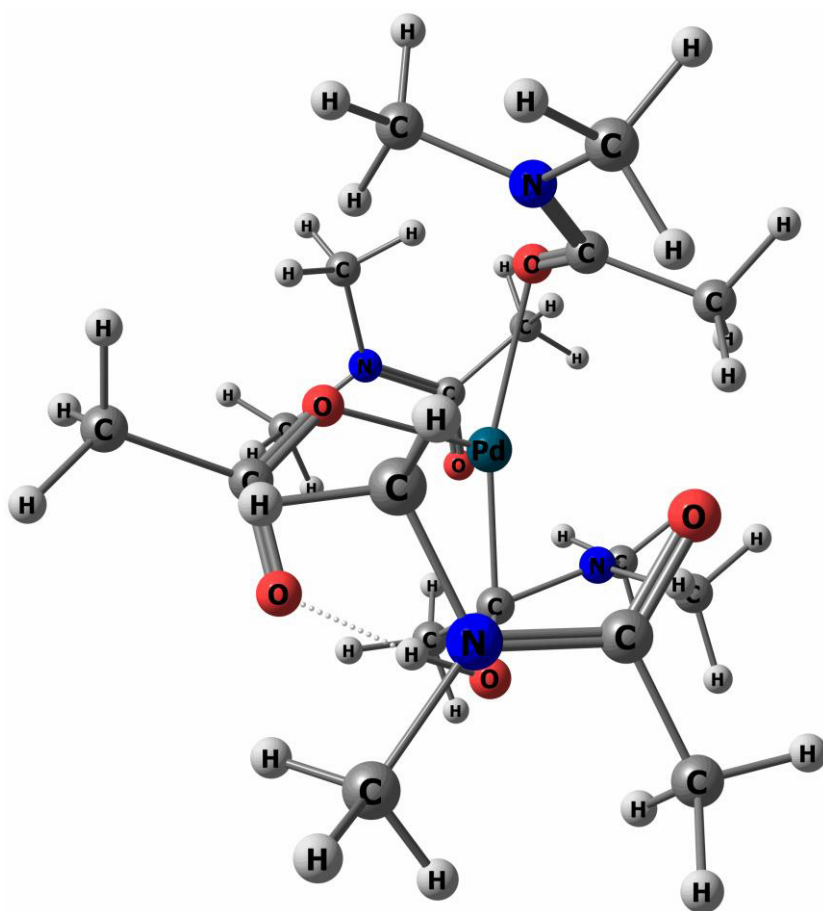


Figure 3-55: Solvated Pd(0) Catalyst – 3 Additional DMA molecules

This lack of implicit coordination with regard to the Pd(0) catalyst, does nothing to change the relative energies of the various products however, and would simply result in a lowering of product Gibbs energies equally across the board.

Product 2 is the most stable molecule and the thermodynamic product. It is possible that the 0.3% of Product 2 observed is because equilibrium was in the process of becoming established towards the end of the reaction time. Assuming all reactions are reversible, it is possible that due to its high concentration, some Product 5 broke down to the starting materials (i.e. the reverse reaction occurred) and subsequently formed Product 2.

As an aside, it should also be noted that the equilibrium constant can also be expressed as the ratio of the rate of the forward (k_f) and reverse (k_r) reactions, Equation (3. 6).

$$K = \frac{k_f}{k_r} \quad (3. 6)$$

On the basis of the TPSSH/SVP data the following preliminary conclusions can be drawn:

- The RDS for Pathway =5 requires the least Gibbs free energy, 6.15 kcal.mol⁻¹ less than any other pathway, and thus Product 5 is thus the kinetic product.
- The rate determining step for each pathway varies as follows:
 Pathway N2: CMD by 0.82 kcal.mol⁻¹
 Pathway N4: RE by 1.55 kcal.mol⁻¹
 Pathway =4: RE by 5.27 kcal.mol⁻¹
 Pathway =5: CMD by 1.21 kcal.mol⁻¹
- Product 2 is the thermodynamic product by 1.33 kcal.mol⁻¹.

COSMO Results:

It is known that applying a solvation model can lead to qualitative changes to the profile of a reaction.¹⁷ A polar solvent such as DMA could possibly stabilise partially charged species more than their neutral counterparts, leading to a change in the relative energies and possibly geometries when compared to gas phase calculations. It was thus necessary to determine if the effects of a solvation model change the conclusions drawn above. For reasons discussed in section 2.5, the COSMO solvation model was implemented. As a reminder, full COSMO geometry optimisations and frequency calculations were performed in DMA solvent, starting from the converged gas phase geometries.

Table 3-13 shows the RMSD values, Equation (3. 1), in terms of the distances between pairs of corresponding atoms in the gas phase and COSMO optimised structures. These RMSD values were calculated using the “Structurers comparer” feature of the ChemCraft programme.¹⁸ The axis and origin of the two structures being compared were manually set to be identical. Before calculating the RMSD, the program rotates and translates one of the structures to minimize the RMSD value. In the current implementation of ChemCraft (Version 1.8, Build 478) the rotation is made iteratively, with a minimal step of 10^{-18} radians. For example, in oxazole the average distance between each corresponding pair of atoms (i.e. between N1 gas phase and N1 COSMO etc.) is 0.004 Å. Larger values in Table 3-13 are mostly a consequence of rotated methyl groups and very little meaningful structural differences are actually observed. Thus, the overall RMSD of 0.280 Å is somewhat inflated.

The COSMO reaction profiles are shown in Figure 3-56. The corresponding reaction Gibbs energies are shown in Table 3-14. The last column in this table shows the difference between the COSMO and gas phase reaction Gibbs energies. The reaction Gibbs energy RMSD for each molecule is also displayed.

The most noticeable change induced by the COSMO model is the increased relative free energy of Intermediate 1 across the board, on average by 3.55 kcal.mol⁻¹. This is particularly severe for the Intermediate 1 structures complexed to palladium via a double bond of oxazole. The formation of both Intermediate 1 =4 and Intermediate 1 =5 is now no longer spontaneous but requires Gibbs energy.

Table 3-13: RMSD of Gas Phase vs COSMO Geometries

Molecule	RMSD (Å)
Pd(II) Catalyst	0.135
Oxazole	0.004
Intermediate 1: N2	0.307
TS 1	0.193
Intermediate 2	0.352
TS 2	0.094
Product 2	0.009
Intermediate 1: N4	0.551
TS 1	0.400
Intermediate 2	0.333
TS 2	0.483
Product 4	0.011
Intermediate 1: =4	0.138
TS 1	0.266
Intermediate 1: =5	0.095
TS 1	0.865
Intermediate 2	0.219
TS 2	1.082
Product 5	0.004
Pd(0) Catalyst	0.052
Average	0.280

Table 3-14: Mono-substitution Reaction Free Energies COSMO (TPSSH/SVP)

Chemical Reaction	COSMO ΔG_{298K} (kcal.mol ⁻¹)	COSMO ΔG - Gas phase ΔG (kcal.mol ⁻¹)
<u>Pathway N2</u>		
Pd(II) Catalyst + Oxazole → Int 1	-4.84	2.28
Int 1 → TS 1	21.33	-2.20
TS 1 → Int 2	-24.12	3.54
Int 2 → TS 2	24.43	1.73
TS 2 → Product 2 + Pd(0) Catalyst	-11.67	-4.21
RMSD:	2.94	
<u>Pathway N4</u>		
Pd(II) Catalyst + Oxazole → Int 1	-4.87	2.79
Int 1 → TS 1	24.92	0.78
TS 1 → Int 2	-18.81	2.25
Int 2 → TS 2	26.19	0.50
TS 2 → Product 4 + Pd(0) Catalyst	-20.65	-5.66
RMSD:	3.02	
<u>Pathway =4</u>		
Pd(II) Catalyst + Oxazole → Int 1	1.34	3.90
Int 1 → TS 1	18.66	-1.76
TS 1 → Int 2	-18.76	3.68
Int 2 → TS 2	26.19	0.50
TS 2 → Product 4 + Pd(0) Catalyst	-20.65	-5.66
RMSD:	3.58	
<u>Pathway =5</u>		
Pd(II) Catalyst + Oxazole → Int 1	0.56	5.21
Int 1 → TS 1	14.96	-2.42
TS 1 → Int 2	-9.18	0.81
Int 2 → TS 2	14.14	-2.03
TS 2 → Product 5 + Pd(0) Catalyst	-14.49	-0.89
RMSD:	2.78	
Total RMSD:	3.08	

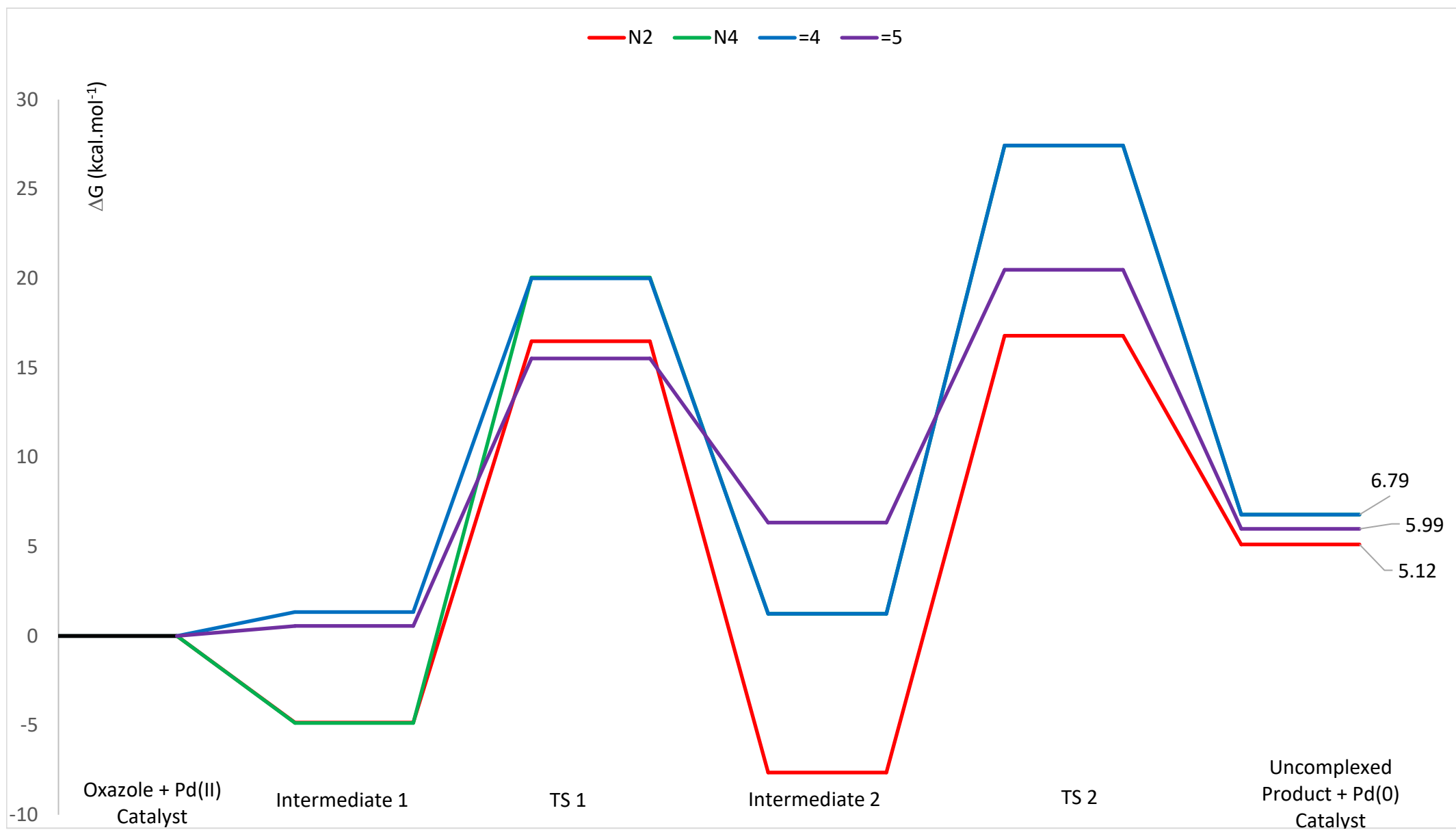


Figure 3-56: Mono-substitution Reaction Profiles for Phosphine Free Conditions COSMO (TPSSH/SVP)

The effect of this increase is that the free energy required to reach TS 1 decreases for all pathways, except Pathway N4. In this case $\Delta^\ddagger G_{TS\ 1}$ increases by 0.78 kcal.mol⁻¹. For the other 3 pathways the average decrease in $\Delta^\ddagger G_{TS\ 1}$ is 2.13 kcal.mol⁻¹. Intermediate 2 is also destabilised by the application of COSMO, this occurs across all pathways, by on average 2.57 kcal.mol⁻¹. However, for all pathways except Pathway =5, TS 2 is destabilised by COSMO to a larger degree than Int 2, resulting in an average $\Delta^\ddagger G_{TS\ 2}$ increase of 1.12 kcal.mol⁻¹ for Pathways N2 and N4. In the case of Pathway =5, TS 2 is stabilised and $\Delta^\ddagger G_{TS\ 2}$ decreases by a significant -2.03 kcal.mol⁻¹.

As mentioned previously, the Pd(0) catalyst is under-ligated and in reality is likely to be stabilised by coordination of solvent molecules. To a certain extent, a continuum solvation model correctly mimics this solvent behaviour, at least in terms of the electrostatics, and the Gibbs energy of the infinitely separated final products and Pd(0) catalyst, relative to TS 2, is lowered across all pathways, on average by 3.59 kcal.mol⁻¹.

In terms of the rate determining step, the difference in $\Delta^\ddagger G$ for the two transition states, $\Delta\Delta^\ddagger G$ (Table 3-15), reveals that TS 2 is now rate determining, by a significant margin (3.10 kcal.mol⁻¹) for Pathway N2, whereas in the gas phase TS 1 was slightly higher than TS 2 (0.82 kcal.mol⁻¹). For the remaining pathways, the magnitude of $\Delta\Delta^\ddagger G$ change slightly, but the RDS remains consistent with gas phase computations.

Table 3-15: CMD vs Reductive Elimination Transition State Gibbs Energies (COSMO TPSSH/SVP)

Pathway	$\Delta^\ddagger G_{TS\ 1} - \Delta^\ddagger G_{TS\ 2}$ (kcal.mol ⁻¹)
N2	-3.10
N4	-1.27
=4	-7.53
=5	0.82

In terms of the Gibbs free energy required for the RDS, Table 3-16, the most favorable pathway remains Pathway =5, requiring 9.47 kcal.mol⁻¹ less Gibbs free energy than the next lowest pathway (N2). The favourability of Product 5 formation has thus increased by 3.32 kcal.mol⁻¹ due to the application of the COSMO model. It can thus be concluded, as before but to an even greater degree, that the major experimental product under conditions of kinetic control will be Product 5. As before, the rate of formation of the remaining products, under conditions of kinetic control, are expected to be insignificant in comparison.

Again, the remaining pathways sit in a narrow range, 24.43 - 26.19 kcal.mol⁻¹. The second lowest pathway is again Pathway N2, followed by Pathway =4 and N2 with 1.76 kcal.mol⁻¹ separating them. The difference between these pathways is slightly smaller than predicted by gas phase calculations and well within the error of the underlying DFT level of theory.

Table 3-16: Pathway Favourability in Terms of the RDS (COSMO TPSSh/SVP)

Favorability	Pathway	RDS	ΔG_{RDS} (kcal.mol ⁻¹)
1st	=5	CMD (TS 1)	14.96
2nd	N2	RE (TS 2)	24.43
3rd	=4/ N4	RE (TS 2)	26.19

Rankings of the relative favourability of the individual transition states is shown for TS 1 in Table 3-17 and for TS 2 in Table 3-18. A similar trend to gas phase computations is revealed.

Table 3-17: Pathway Favourability in Terms of TS 1 (COSMO TPSSh/SVP)

Rank	Pathway	$\Delta^{\ddagger}G_{\text{TS 1}}$ (kcal.mol ⁻¹)
1 st	=5	14.96
2 nd	=4	18.66
3 rd	N2	21.33
4 th	N4	24.92

Table 3-18: Pathway Favourability in Terms of TS 2 (COSMO TPSSh/SVP)

Rank	Pathway	$\Delta G^{\ddagger}_{\text{TS 2}}$ (kcal.mol ⁻¹)
1 st	=5	14.14
2 nd	N2	24.43
3 rd	N4/=4	26.19

Although as previously mentioned, the Pd(0) catalyst is significantly stabilised by the application of COSMO, so too is the Pd(II) catalyst. The consequence is that the overall reaction Gibbs energies remain similar. As can be seen in Figure 3-56, the thermodynamic product is still Product 2 (5.12 kcal.mol⁻¹), followed by Product 5 (5.99 kcal.mol⁻¹) and then Product 4 (6.79 kcal.mol⁻¹). Although the favourability of Product 2 has decreased slightly when compared to gas phase data, the same arguments utilized before can still be used to explain the small amount of Product 2 observed experimentally. Namely, that Product 2 is the thermodynamic product and that the establishment of equilibrium is responsible for the 0.3% detected.

Thus, apart from the minor changes, such as the change in the RDS for Pathway N2, the same major conclusions can be drawn from both gas phase and COSMO calculations. The COSMO

solvation model therefore offers no greater insight into the energetics of these reactions than does gas phase hybrid DFT. The continued use of a solvation model was deemed unnecessary and for the remainder of the project gas phase calculations were utilized. The focus will now shift to how the effects of dynamic electron correlation apply to these reactions.

DLPNO-CCSD(T) Results:

DLPNO-CCSD(T) shows consistently high accuracy across various chemical systems, for which it is appropriate to use single reference wavefunction methods. Additionally, unlike DFT, it is based on a theoretical framework that rigorously describes excited electronic states and thus should provide an improved description of the energetics of transition states. Lastly, as this method is linear scaling, it can be applied to the relatively large systems that concern this work. It was thus used as a benchmark for single point energy calculations. Details of the method can be found in Section 2.4.

Throughout this project, it was assumed, as a rough estimate, that the absolute error in energy differences is around 0.5 kcal.mol⁻¹ for the system at the DLPNO-CCSD(T)/TZVP level of theory. By way of a reminder, the thermodynamic corrections were imported from the corresponding TPSSh/SVP calculations and applied to the single point energy calculations at the DLPNO-CCSD(T)/TZVP level of theory. The DLPNO-CCSD(T) reaction profiles are shown in Figure 3-57. The corresponding reaction Gibbs energies are shown in Table 3-19. The last column of this table shows the difference between the DLPNO-CCSD(T) and TPSSh reaction Gibbs energies. The RMSD for each pathway is also displayed.

Over the entire range of pathways considered, Intermediate 1 was destabilized relative to TPSSh/SVP by on average 5.44 kcal.mol⁻¹. This effect was particularly severe for the intermediates in which oxazole was complexed to palladium via a double bond. It can be seen that at this level of theory the formation of Intermediate 1 =4 and =5 is not spontaneous but requires an input of Gibbs energy, 3.98 and 1.33 kcal.mol⁻¹ respectively. The negative ΔG values previously predicted for these pathways seem to be a relic of TPSSh in conjunction with a relatively small basis set. This increase in the energy of Int 1 has however not resulted in a decrease in $\Delta^\ddagger G_{TS\ 1}$. TS 1 has, across all pathways, also been revealed to be less stable than initially calculated, with an average $\Delta^\ddagger G_{TS\ 1}$ increase of 3.14 kcal.mol⁻¹. $\Delta^\ddagger G_{TS\ 2}$ however changes only marginally from hybrid DFT. It increases by 0.07 kcal.mol⁻¹ for Pathway N2, decreases by -1.07 kcal.mol⁻¹ for Pathways N4 and =4 and increases by 0.53 kcal.mol⁻¹ for Pathway =5. The overall RMSD of 3.88 kcal.mol⁻¹ shows that the Gibbs energies of each reaction have changed significantly.

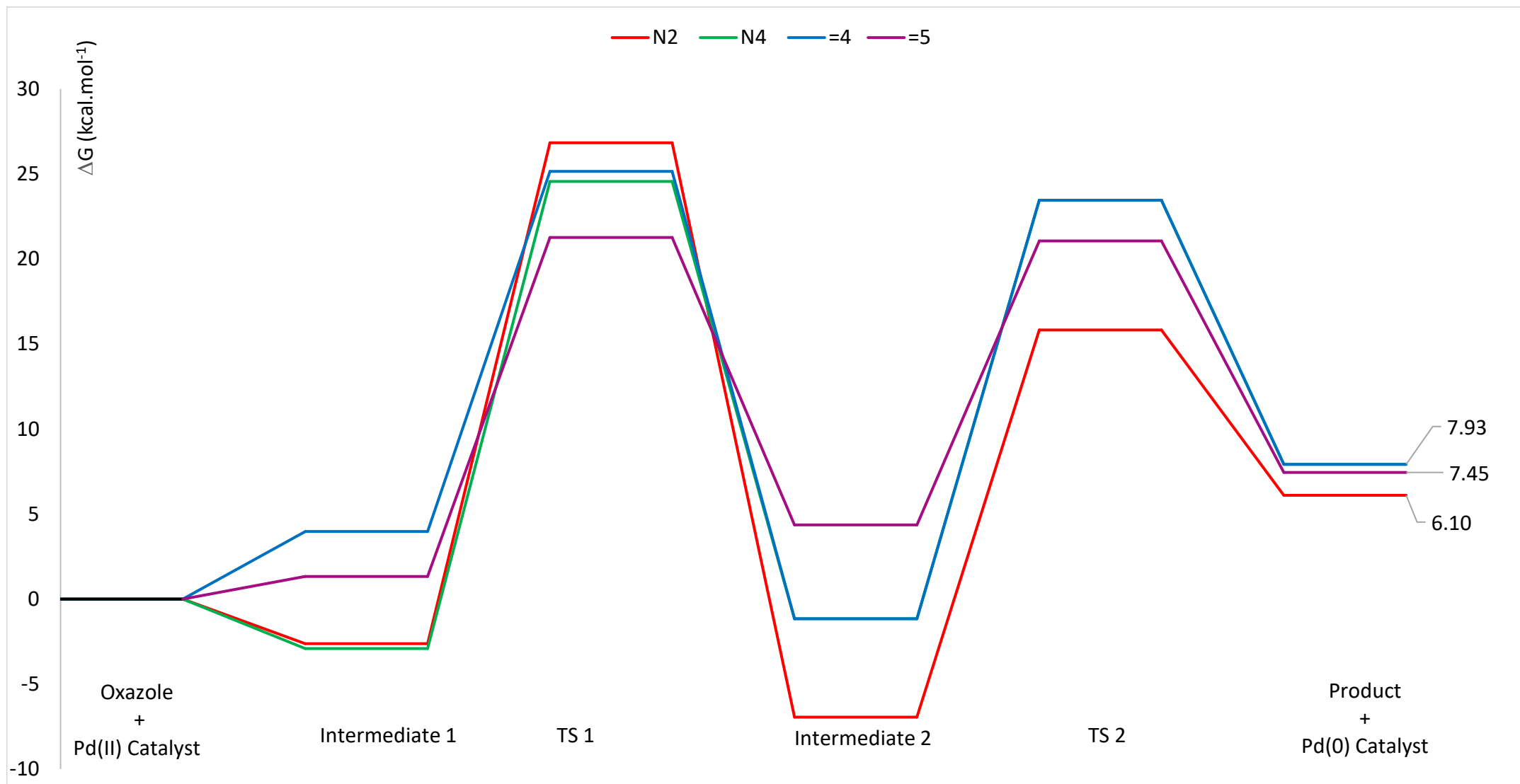


Figure 3-57: Mono-substitution Reaction Profiles (DLPNO-CCSD(T)/TZVP)

Table 3-19: Mono-substitution Reaction Free Energies DLPNO-CCSD(T)/TZVP

Chemical Reaction	DLPNO-CCSD(T) ΔG_{298K} (kcal.mol ⁻¹)	DLPNO-CCSD(T) ΔG - TPSSh ΔG (kcal.mol ⁻¹)
<u>Pathway N2</u>		
Catalyst + Oxazole → Int 1	-2.62	4.50
Int 1 → TS 1	29.46	5.93
TS 1 → Int 2	-33.78	-6.12
Int 2 → TS 2	22.77	0.07
TS 2 → Product 2 + Pd(0) Catalyst	-9.72	-2.26
RMSD:	4.43	
<u>Pathway N4</u>		
Catalyst + Oxazole → Int 1	-2.91	4.75
Int 1 → TS 1	27.47	3.33
TS 1 → Int 2	-25.72	-4.65
Int 2 → TS 2	24.61	-1.07
TS 2 → Product 4 + Pd(0) Catalyst	-15.53	-0.55
RMSD:	3.37	
<u>Pathway =4</u>		
Catalyst + Oxazole → Int 1	3.98	6.54
Int 1 → TS 1	21.18	0.76
TS 1 → Int 2	-26.31	-3.87
Int 2 → TS 2	24.61	-1.07
TS 2 → Product 4 + Pd(0) Catalyst	-15.53	-0.55
RMSD:	3.46	
<u>Pathway =5</u>		
Catalyst + Oxazole → Int 1	1.33	5.98
Int 1 → TS 1	19.94	2.56
TS 1 → Int 2	-16.90	-6.91
Int 2 → TS 2	16.70	0.53
TS 2 → Product 5 + Pd(0) Catalyst	-13.61	-0.02
RMSD:	4.25	
Total RMSD:	3.88	

In terms of the rate determining step for each pathway, things have changed substantially. As can be seen in Table 3-20, for Pathway N2 the RDS is now TS 1 by a large margin of 6.69 kcal.mol⁻¹, whereas at the TPSSh/SVP level the two TSs differed only marginally (0.82 kcal.mol⁻¹). For Pathway N4, TS 1 is now rate determining by 2.86 kcal.mol⁻¹, while TS 2 was computed to be the RDS at the TPSSh/SVP level by 1.55 kcal.mol⁻¹. For Pathway =4, the TPSSh/SVP level resulted in TS 2 being the RDS by the large margin of 5.27 kcal.mol⁻¹. DLPNO- CCSD(T) calculations reveal that TS 2 is still rate determining for this pathway, but now by a smaller margin of 3.44 kcal.mol⁻¹. For Pathway =5, TS 1 is rate determining as before but now significantly so, by an increased factor of almost 3. Thus, this higher level of theory has

revealed that although RE is an energetically relevant process that substantially contributes to the total G input, the CMD reaction is the RDS for all pathways considered, except Pathway =4. Pathway =4 is unique as RE is the RDS by a relatively large margin.

Table 3-20: CMD vs Reductive Elimination Transition State Gibbs Energies (DLPNO-CCSD(T)/TZVP)

Pathway	$\Delta^\ddagger G_{TS\ 1} - \Delta^\ddagger G_{TS\ 2}$ (kcal.mol ⁻¹)
N2	6.69
N4	2.86
=4	-3.44
=5	3.24

In terms of the Gibbs free energy requirements for these reactions, it can be seen in Table 3-19 that TPSSh/SVP actually underestimates the reaction barriers for most pathways, on average by 1.31 kcal.mol⁻¹. However, considering an average value is somewhat deceptive as the free energy barriers of certain transition states increased significantly. For example, TS 1 for Pathway N2 has increased by 5.93 kcal.mol⁻¹ upon the application of DLPNO-CCSD(T)/TZVP.

Ranking of the pathways according to the Gibbs energy required for the RDS, Table 3-21, shows that the most favorable pathway, kinetically, remains Pathway =5. The difference between Pathway =5 and the next most favorable pathway (now Pathway =4) is 4.67 kcal.mol⁻¹, slightly less than that predicted at the TPSSh/SVP level (6.15 kcal.mol⁻¹). The remaining pathways again occupy a relatively narrow range, 24.61 – 29.46 kcal.mol⁻¹. Whereas at the TPSSh/SVP level, Pathway N4/=4 was the highest energy route by 8.31 kcal.mol⁻¹, now Pathway N2 assumes this role. The difference of 1.99 kcal.mol⁻¹ between pathways N4 and N2 is now however, far greater than the inherent error of the method. Thus, Pathway N2 is clearly the most kinetically hindered pathway and very little Product 2 is expected to form under conditions of kinetic control.

Table 3-21: Pathway Favourability in Terms of the RDS (DLPNO-CCSD(T)/TZVP)

Rank	Pathway	RDS	ΔG RDS (kcal.mol ⁻¹)
1st	=5	CMD (TS 1)	19.94
2nd	=4	RE (TS 2)	24.61
3rd	N4	CMD (TS 1)	27.47
4th	N2	CMD (TS 1)	29.46

It can be seen that both TS 1 (Table 3-22) and TS 2 (Table 3-23) are lower in Gibbs energy than the alternative reactions. As Int 1 formation now contributes to the total G input for some pathways, Table 3-24 is also included.

Table 3-22: Pathway Favourability in Terms of TS 1 (DLPNO-CCSD(T)/TZVP)

Rank	Pathway	$\Delta G^{\ddagger}_{TS\ 1}$ (kcal.mol ⁻¹)
1 st	=5	19.94
2 nd	=4	21.18
3 rd	N4	27.47
4 th	N2	29.46

Table 3-23: Pathway Favourability in Terms of TS 2 (DLPNO-CCSD(T)/TZVP)

Rank	Pathway	$\Delta G^{\ddagger}_{TS\ 2}$ (kcal.mol ⁻¹)
1 st	=5	16.70
2 nd	N2	22.77
3 rd	N4/=4	24.61

Table 3-24: Pathway Favourability in Terms of Int 1 (DLPNO-CCSD(T)/TZVP)

Rank	Pathway	$\Delta G_{Int\ 1}$ (kcal.mol ⁻¹)
1 st	N4	-2.91
2 nd	N2	-2.62
3 rd	=5	1.33
4 th	=4	3.98

Returning to Figure 3-57, it can be observed that in absolute Gibbs energy terms TS 2 for Pathway =5 is positioned between the highest energy TS 2 of Pathway =4, and the lowest energy TS 2 of Pathway N2. However, $\Delta^{\ddagger}G_{TS\ 2}$ for Pathway =5 is significantly lower than the other pathways due to the high energy of Intermediate 2. Analysing the structures of Int 2 N2 (Figure 3-32), Int 2 N4 (Figure 3-43) and Int 2 =5 (Figure 3-50) reveals that the nitrogen atom in Int 2 =5 is unable to form a hydrogen bond with the hydroxyl group of the acetate ligand. In constant, deprotonation of the acetate ligand by the nitrogen atom takes place in Int 2 N2, forming a zwitterionic species that sits lowest in energy, while a strong hydrogen bond forms in Int 2 N4. Thus, the low $\Delta^{\ddagger}G_{TS\ 2}$ value for Pathway =5 is because Int 2 for this pathway does not correspond to such a deep well on the potential energy surface, relative to the other pathways.

In terms of the total $\Delta_r G$ values, the reactions remain endergonic, but now by a larger magnitude. The thermodynamic product remains Product 2 (6.10 kcal.mol⁻¹) followed by Product 5 (7.45 kcal.mol⁻¹) and then Product 4 (7.93 kcal.mol⁻¹), the same sequence as predicted at the TPSSh/SVP level. Product 2 is favoured thermodynamically by 1.35 kcal.mol⁻¹, remarkably similar to the 1.33 kcal.mol⁻¹ predicted at the TPSSh/SVP level.

From the results of the DLPNO-CCSD(T)/TZVP single point energy calculations, together with the thermodynamic terms applied from TPSSh/SVP calculations, the following conclusions can be drawn:

- The RDS for Pathway =5 requires the least input of Gibbs energy, 4.67 kcal.mol⁻¹ less than any other pathway, and Product 5 is thus the kinetic product. This is in agreement with experimental results that found Product 5 makes up 99.55% of the mono-substituted product distribution and 65.8% of the total product distribution.
- It is thus highly probable that under phosphine free conditions, these reactions proceed via a CMD mechanism.
- The rate determining step for 3 of the pathways considered is concerted metallation deprotonation, by the following margins:
Pathway N2: 6.69 kcal.mol⁻¹
Pathway N4: 2.86 kcal.mol⁻¹
Pathway =5: 3.24 kcal.mol⁻¹
- Pathway =4 is unique as it is the only pathway where reductive elimination is the rate determining step, by a margin of 3.44 kcal.mol⁻¹.
- Product 2 is the thermodynamic product by 1.35 kcal.mol⁻¹, explaining the trace amounts observed.

When contrasting these conclusions with those derived from TPSSh/SVP data, hybrid DFT does an excellent job of describing the general energetics of the system but fails to consistently determine the nature and magnitude of the RDS. This is to be expected, as DFT is a ground state theory and applying it to calculate TS energies exceeds the assumptions of the underlying equations. These results highlight the importance of performing calculations with methods that include a more accurate description of electron correlation and large basis sets before drawing conclusions on which step in a sequence of reactions is rate determining. However, considering the computational expense of a single point energy DLPNO-CCSD(T) calculation on average for an Intermediate 1 molecule was 7 hours, while the corresponding TPSSh job took a few minutes, DFT is extremely efficient and remarkably accurate for the system at hand.

If this project is to be used as a preliminary investigation to design improved catalytic systems in order to speed up these reactions, the study of mono-substitution has revealed that the CMD step would be the first process to focus on lowering the energy of, other than if Product 4 was desired. In this case, research into lowering the barrier associated with RE is necessary.

If the formation of Product 2 or Product 4 was desired, efforts should be focused on increasing the energy of the relevant Intermediate 2 structures. In particular, if Product 2 formation is desired, research should be invested into preventing the acid base reaction between the acetate ligand and the nitrogen atom of oxazole from taking place. The resultant zwitterionic molecule (Int N2) is significantly stabilized and requires a large input of energy for reductive elimination to occur. Increasing the pH of the solution so that the acetate ligand is deprotonated as soon as CMD takes place is a possibility, for example. Another example could be adding inert cations that complex with nitrogen's lone pair of electrons, after CMD has taken place. Such strategies would also prevent the hydrogen bond between the acetate ligand and the nitrogen atom present in Int 2 N4 from forming, and thus are also expected to favor the formation of Product 4. As CMD is the RDS, RE is expected to occur more slowly and Int 2 could possibly persist for long enough in solution for these suggestions to be implemented. However, it should be noted that any attempt to disrupt the aforementioned interactions would also destabilize TS 2 in these pathways. This is because both the zwitterion in the case of Pathway N2, as well as the hydrogen bond in Pathway =4/N4 persist in TS 2 (see Figure 3-34 and Figure 3-44, respectively) and have the same stabilizing effect as they do in Int 2. Thus, the feasibility of such strategies remains in question.

DSD-PBEP86 Results:

Although used as a benchmark for single point energy calculations, DLPNO-CCSD(T)/TZVP was too computationally costly to be implemented routinely throughout the project. Thus, in order to efficiently account for correlation for the remainder of the work a compromise was necessary. Double hybrid DFT (DH-DFT) is currently one of the most cost-effective ways to account for electron correlation. This method was implemented in the form of the DSD-PBEP86 functional in conjunction with a TZVP basis set (see Section 2.3 for details). It is now necessary to confirm that the same conclusions can be drawn from both DLPNO-CCSD(T)/TZVP and DSD-PBEP86/TZVP calculations.

Throughout this project, it was assumed, as a rough estimate, that the absolute error in energy differences is around 1.0 kcal.mol⁻¹ for the system at the DSD-PBEP86/TZVP level of theory. By way of a reminder, the thermodynamic corrections were imported from the corresponding TPSSH/SVP calculations and applied to DSD-PBEP86/TZVP single point energies. The DSD-PBEP86 reaction profiles are shown in Figure 3-58. The corresponding reaction Gibbs energies are shown in Table 3-25. The last column of this table shows the difference between the DSD-

PBEP86 and DLPNO-CCSD(T) reaction Gibbs energies. The RMSD for each pathway is also displayed.

Using DLPNO-CCSD(T) as the benchmark it can be seen that DSD-PBEP86 overestimates the stability of Intermediate 1, by on average $2.11 \text{ kcal.mol}^{-1}$. This is particularly severe for the intermediates in which oxazole is bound to palladium via a double bond. Overestimation of the Gibbs energy of these molecules seems to be an inherent error in DFT based methods, and also occurred with TPSSh/SVP. The formation of Int 1 =5 is now spontaneous by $1.35 \text{ kcal.mol}^{-1}$ whereas with DLPNO-CCSD(T) it required $1.33 \text{ kcal.mol}^{-1}$ of Gibbs energy to form. This is however the only qualitative failure of DSD-PBEP86 and all other features of the DLPNO-CCSD(T) reaction profiles are reproduced faithfully. The non-spontaneous nature of Int 1 =4 formation is correctly predicted, requiring $0.81 \text{ kcal.mol}^{-1}$, although this is far from

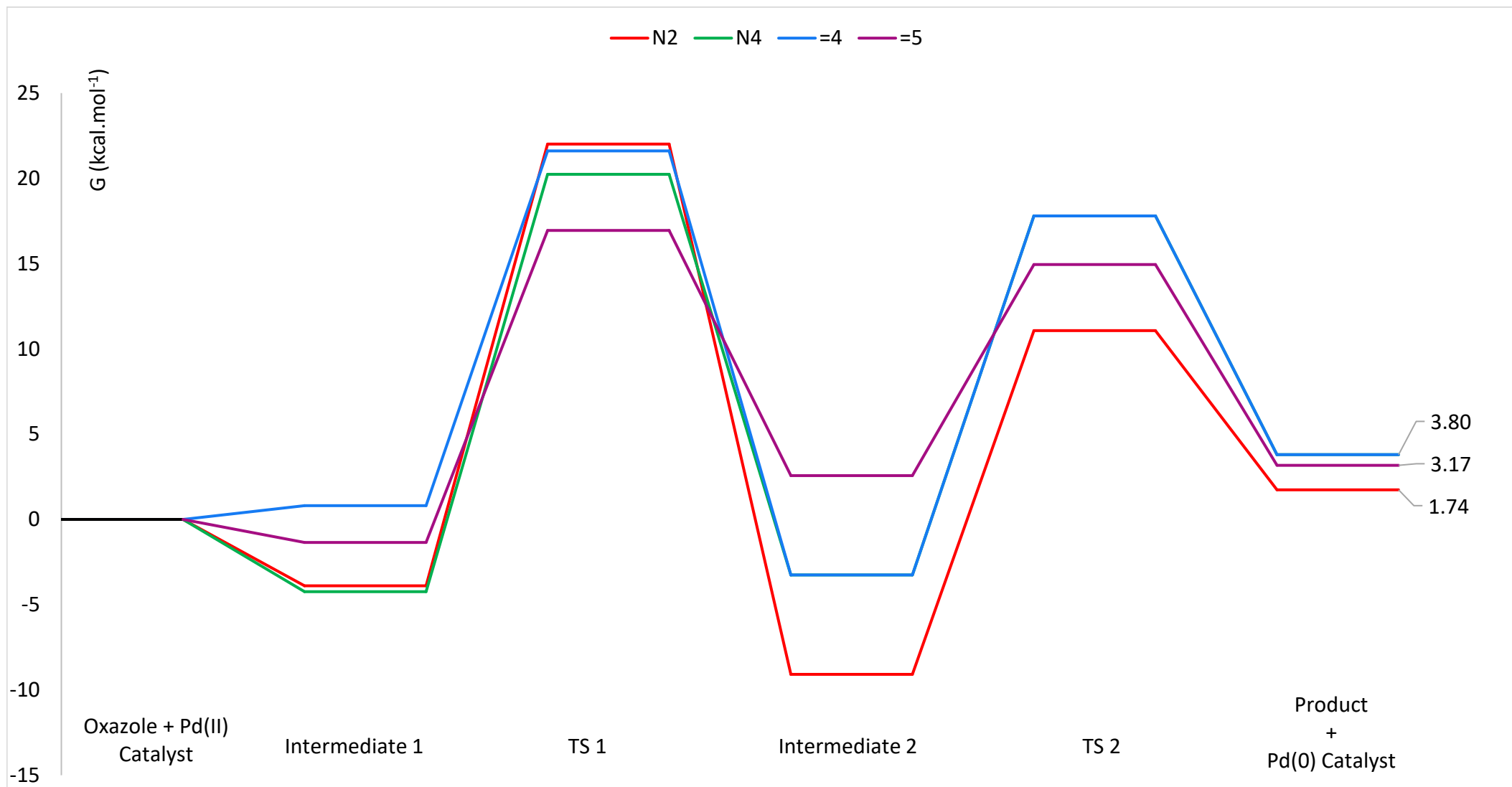


Figure 3-58: Mono-substitution Reaction Profiles (DSD-PBEP86/TZVP)

Table 3-25: Mono-substitution Reaction Gibbs Free Energies (DSD-PBEP86/TZVP)

Chemical Reaction	DSD-PBEP86 ΔG_{298K} (kcal.mol ⁻¹)	DSD-PBEP86 ΔG - DLPNO-CCSD(T) ΔG (kcal.mol ⁻¹)
<u>Pathway N2</u>		
Pd(II) Catalyst + Oxazole $\rightarrow$ Int 1	-3.89	-1.27
Int 1 $\rightarrow$ TS 1	25.90	-3.56
TS 1 $\rightarrow$ Int 2	-31.09	2.69
Int 2 $\rightarrow$ TS 2	20.16	-2.62
TS 2 $\rightarrow$ Product 2 + Pd(0) Catalyst	-9.34	0.39
RMSD	2.39	
<u>Pathway N4</u>		
Pd(II) Catalyst + Oxazole $\rightarrow$ Int 1	-4.24	-1.33
Int 1 $\rightarrow$ TS 1	24.47	-3.00
TS 1 $\rightarrow$ Int 2	-23.49	2.22
Int 2 $\rightarrow$ TS 2	21.05	-3.56
TS 2 $\rightarrow$ Product 4 + Pd(0) Catalyst	-14.00	1.54
RMSD	2.48	
<u>Pathway =4</u>		
Pd(II) Catalyst + Oxazole $\rightarrow$ Int 1	0.81	-3.17
Int 1 $\rightarrow$ TS 1	20.80	-0.38
TS 1 $\rightarrow$ Int 2	-24.86	1.45
Int 2 $\rightarrow$ TS 2	21.05	-3.56
TS 2 $\rightarrow$ Product 4 + Pd(0) Catalyst	-14.00	1.54
RMSD	2.34	
<u>Pathway =5</u>		
Pd(II) Catalyst + Oxazole $\rightarrow$ Int 1	-1.35	-2.68
Int 1 $\rightarrow$ TS 1	18.30	-1.64
TS 1 $\rightarrow$ Int 2	-14.38	2.53
Int 2 $\rightarrow$ TS 2	12.38	-4.32
TS 2 $\rightarrow$ Product 5 + Pd(0) Catalyst	-11.77	1.84
RMSD	2.77	
Total RMSD	2.49	

the 3.98 kcal.mol⁻¹ benchmark. When compared to the benchmark, $\Delta^\ddagger G_{TS\ 1}$ values are underestimated for all pathways by on average 2.14 kcal.mol⁻¹. This decrease in $\Delta^\ddagger G_{TS\ 1}$ is consistent for all pathways and the impact on conclusions drawn from the data is thus minimal. $\Delta^\ddagger G_{TS\ 2}$ is also underestimated across all pathways, with an average decrease of 3.50 kcal.mol⁻¹ when compared to the benchmark. This underestimation is also consistent, with decreases ranging from 2.62 to 4.32 kcal.mol⁻¹, and the impact on the ultimate conclusions is again minimal, as will be shown. In terms of the free energy requirements of the transition states as a whole, DSD-PBEP86/TZVP underestimates this value across the board, by on average 2.83

kcal.mol⁻¹. The overall reaction Gibbs reaction energy RMSD between DSD-PBEP86/TZVP and DLPNO-CCSD(T)/TZVP of 2.49 kcal.mol⁻¹ is significant, but not as large as the 3.88 kcal.mol⁻¹ between TPSSH/SVP and DLPNO-CCSD(T)/TZVP.

In terms of the rate determining step for each pathway, Table 3-26, DSD-PBEP86/TZVP correctly predicts the rate determining step in all cases. For Pathways N2 and N4 the magnitudes of the RDS are very similar to that of the benchmark. However, for Pathway =4, TS 2 is predicted to be rate determining by only 0.25 kcal.mol⁻¹, which is within the assumed error of the method. Without access to DLPNO-CCSD(T) data, it would have been concluded that this pathway has no clear RDS. However, TS 2 is the RDS according to the benchmark data by a clear 3.44 kcal.mol⁻¹. For Pathway =5 TS 1 is predicted to be rate determining although, the magnitude of 5.92 kcal.mol⁻¹ is 1.83 times that of the benchmark and clearly an exaggeration.

Table 3-26: Reductive Elimination vs CMD Transition State Gibbs Energies (DSD-PBEP86/TZVP)

Pathway	$\Delta^\ddagger G_{TS1} - \Delta^\ddagger G_{TS2}$ DSD-PBEP86 (kcal.mol ⁻¹)	$\Delta^\ddagger G_{TS1} - \Delta^\ddagger G_{TS2}$ DLPNO-CCSD(T) (kcal.mol ⁻¹)	Difference (kcal.mol ⁻¹)
N2	5.75	6.69	-0.94
N4	3.42	2.86	0.57
=4	-0.25	-3.44	3.19
=5	5.92	3.24	2.68

In terms of the Gibbs energy required for the RDS, Table 3-27, exactly the same order of favorability is predicted as the benchmark. Pathway =5 is the kinetic product by 2.75 kcal.mol⁻¹, although smaller than the 4.67 kcal.mol⁻¹ benchmark value. At the DH level the remaining pathways occupy a narrow range of 4.85 kcal.mol⁻¹, identical to that predicted by DLPNO-CCSD(T). Pathway =4 and Pathway N2 are separated by 1.43 kcal.mol⁻¹, close to the assumed error of the method, similar to results from the benchmark.

Table 3-27: Pathway Favourability in Terms of the RDS (DSD-PBEP86/TZVP)

Rank	Pathway	RDS	ΔG RDS (kcal.mol ⁻¹)
1 st	=5	CMD (TS 1)	18.30
2 nd	=4	RE (TS 2)	21.05
3 rd	N4	CMD (TS 1)	24.47
4 th	N2	CMD (TS 1)	25.90

Rankings of the relative favourability of the individual transition states are presented for TS 1 in Table 3-28 and for TS 2 in Table 3-29. They reveal an order exactly the same as predicted by DLPNO-CCSD(T). This is also true of the favourability of Int 1, presented in Table 3-30. As

expected, for Pathway =5 both transition states are significantly lower than the alternative reactions.

Table 3-28: Pathway Favourability in Terms of TS 1 (DSD-PBEP86/TZVP)

Rank	Pathway	$\Delta G^{\ddagger}_{TS\ 1}$ (kcal.mol ⁻¹)
1 st	=5	18.30
2 nd	=4	20.80
3 rd	N4	24.47
4 th	N2	25.90

Table 3-29: Pathway Favourability in Terms of TS 2 (DSD-PBEP86/TZVP)

Rank	Pathway	$\Delta G^{\ddagger}_{TS\ 2}$ (kcal.mol ⁻¹)
1 st	=5	12.38
2 nd	N2	20.16
3 rd	N4/=4	21.05

Table 3-30: Pathway Favourability in Terms of Int 1 (DSD-PBEP86/TZVP)

Rank	Pathway	$\Delta G_{Int\ 1}$ (kcal.mol ⁻¹)
1st	N4	-4.24
2nd	N2	-3.89
3rd	=5	-1.35
4th	=4	0.81

In terms of the total $\Delta_r G$ values, the reactions remain endergonic, but by a smaller margin than the benchmark. The thermodynamic product remains Product 2 (1.74 kcal.mol⁻¹) followed by Product 5 (3.17 kcal.mol⁻¹) and then Product 4 (3.80 kcal.mol⁻¹). The correct sequence is thus predicted. Product 2 is favoured thermodynamically by 1.43 kcal.mol⁻¹, similar to the benchmark value of 1.35 kcal.mol⁻¹.

Thus, exactly the same conclusions can be drawn from both DLPNO-CCSD(T)/TZVP and DSD-PBEP86/TZVP calculations. Although, the magnitude of the RDS is inaccurate in some cases. Considering that the average DSD-PBEP86/TZVP Intermediate 1 job took approximately 8 minutes, while the average DLPNO-CCSD(T)/TZVP Intermediate 1 job took 7 hours, DH DFT is both an accurate and extremely efficient way to better account for the effects of electron correlation than standard hybrid DFT theory. DSD-PBEP86/TZVP single point energy calculations will thus be implemented for the remainder of the project and no further use of DLPNO-CCSD(T)/TZVP was deemed necessary.

3.3 Di-substitution

As discussed briefly in Chapter 1, Section 1.3.2, each of the mono substituted products, namely Product 2, Product 4 or Product 5, can enter the catalytic cycle for a second time by reacting with a Pd(II) catalyst molecule. One of the two remaining hydrogen atoms will then be replaced by a toluene moiety in a second direct arylation reaction, made possible by exactly the same catalytic system as before. This was shown in Chapter 1, Figure 1.12. However, each di-substituted product can be formed from one of two mono-substituted products re-entering the catalytic cycle for a second time, thus Figure 1.12 from Chapter 1 is an oversimplification and there are actually two pathways leading to each di-substituted product, each rising from a different monosubstituted product entering the catalytic cycle for a second time. Each of these possible pathways is depicted in Figure 3-59. It should be noted that in this figure the true system is depicted, however, as before benzene will replace toluene in the model systems. Product 2,4 can arise from either Product 2 or Product 4 entering the catalytic cycle for a second time. The reaction pathway starting from Product 2 will be referred to as Pathway $2 \rightarrow 2,4$ while that starting from Product 4 will be referred to as Pathway $4 \rightarrow 2,4$. Exactly the same reactions modelled previously, namely CMD and RE are also responsible for the formation of di-substituted products. Thus, there will again be an Int 1, TS 1, Int 2 and TS 2 for each pathway. A similar naming scheme will be employed for the formation of the remaining two di-substituted products, in which the pathway's name starts with the mono-substituted product that is entering the catalytic cycle for a second time and ends with the resultant di-substituted product. It is essential to model both pathways leading to each di-substituted product, so that the lowest Gibbs energy route followed by nature is discovered.

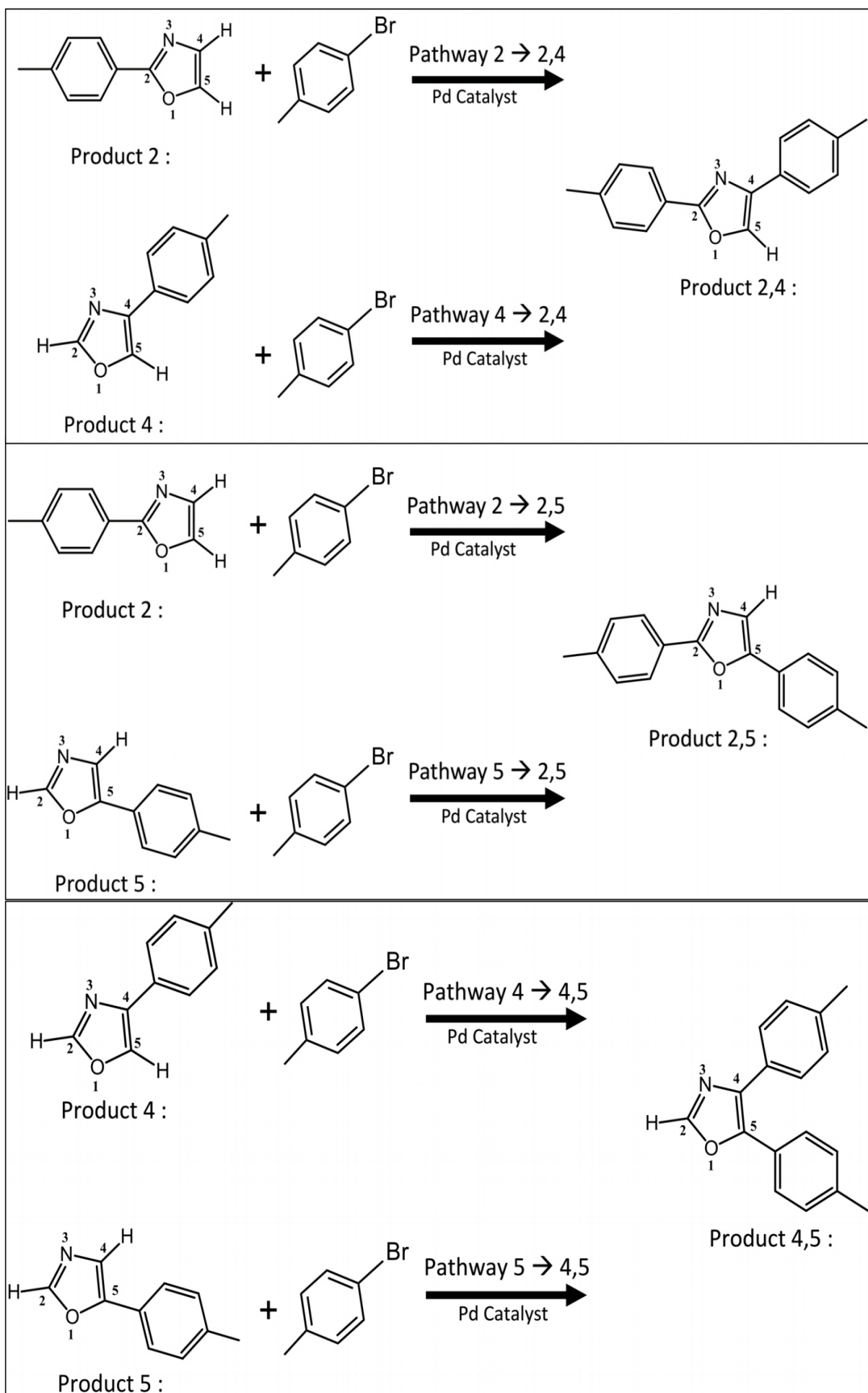


Figure 3-59: Di-substitution Product Formation under Phosphine Free Conditions

The Reaction Pathways:

Pathway 2 $\rightarrow$ 2,4:

Starting from Product 2 entering the catalytic cycle for a second time, there are two ways that this molecule can coordinate to the active catalyst, with position 4 aligned to be deprotonated; via the Nitrogen atom (Figure 3-60) or via the double bond between carbons 4 and 5 (Figure 3-61). Only the lowest “enantiomer” of each intermediate will be presented, with the alternative available in the ESI.

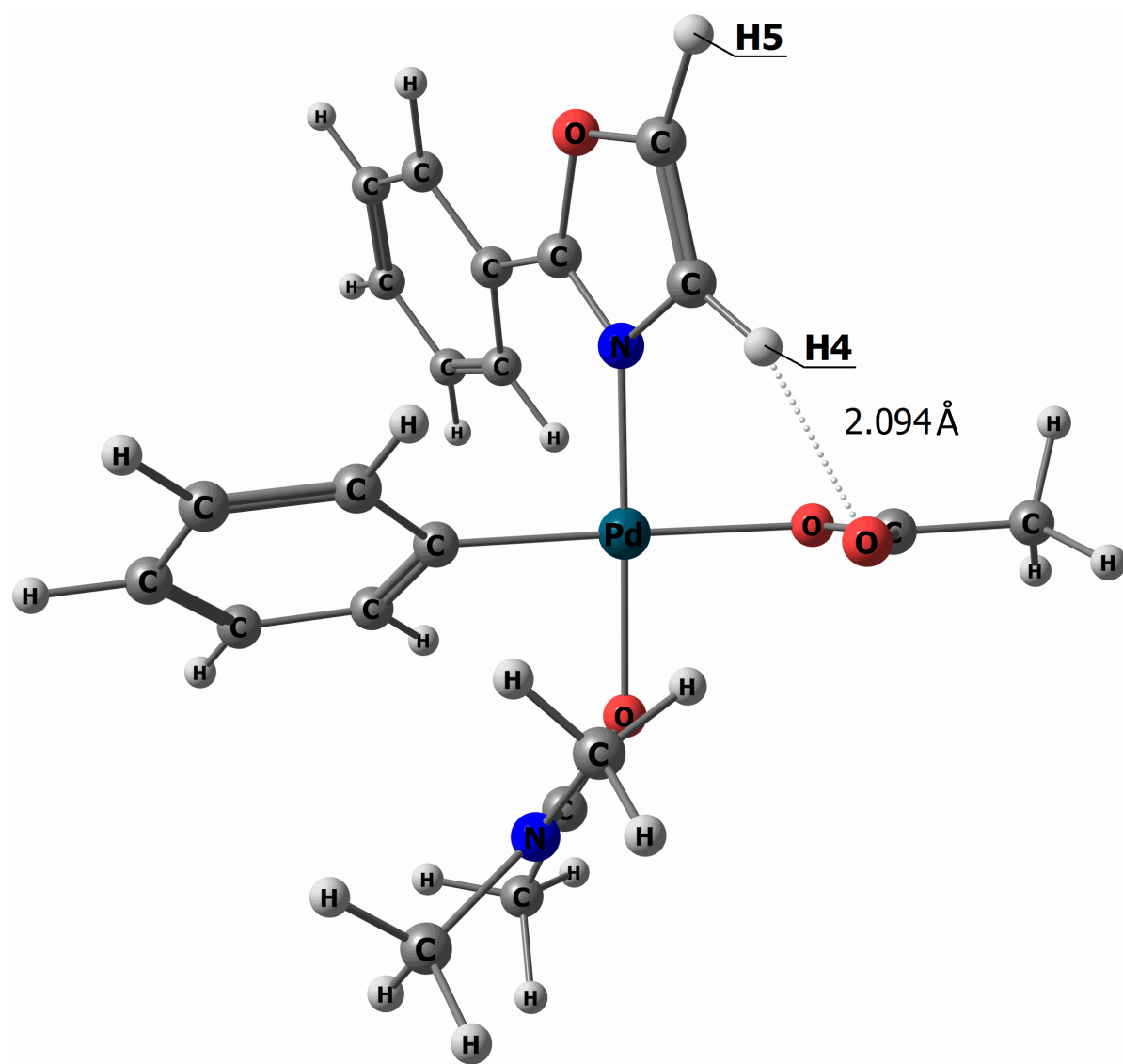


Figure 3-60: Intermediate 1 Pathway 2 $\rightarrow$ 2,4 (Nitrogen Coordinated)

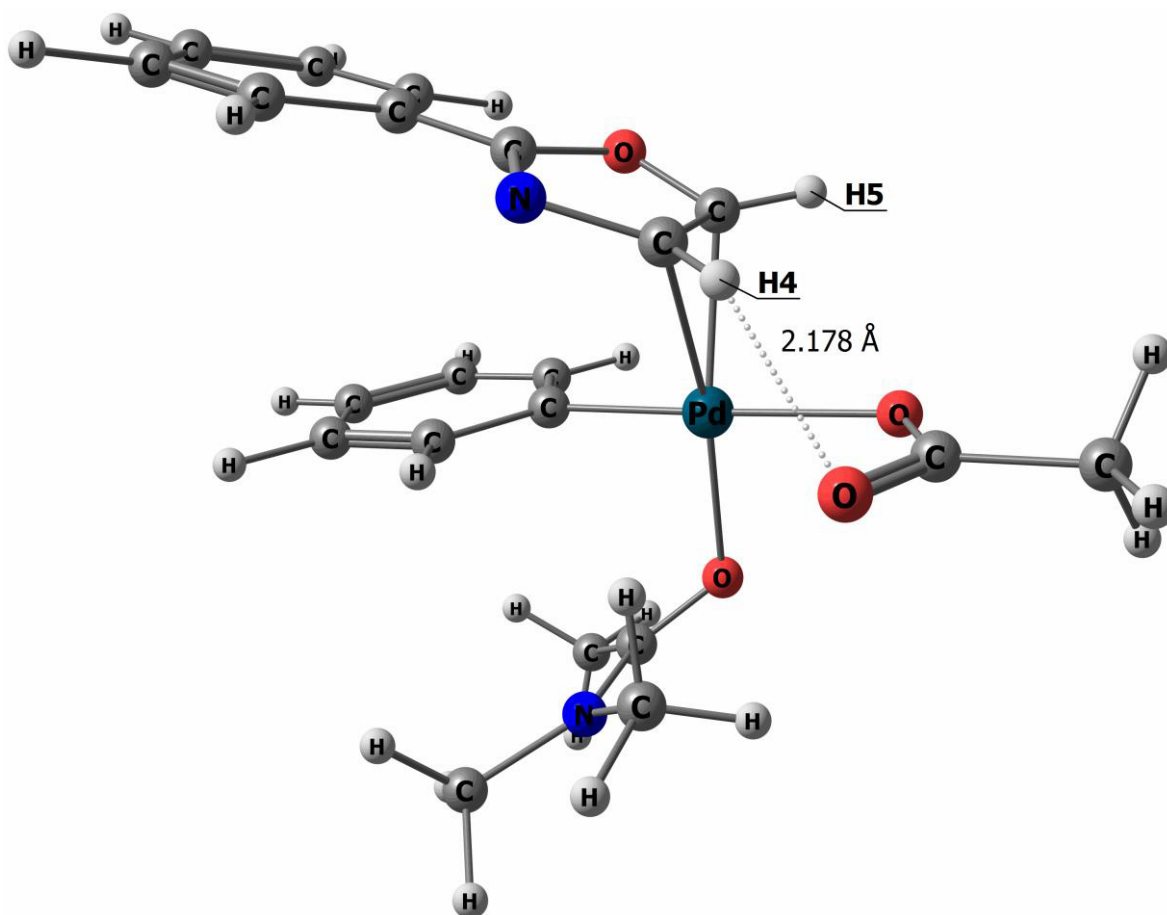
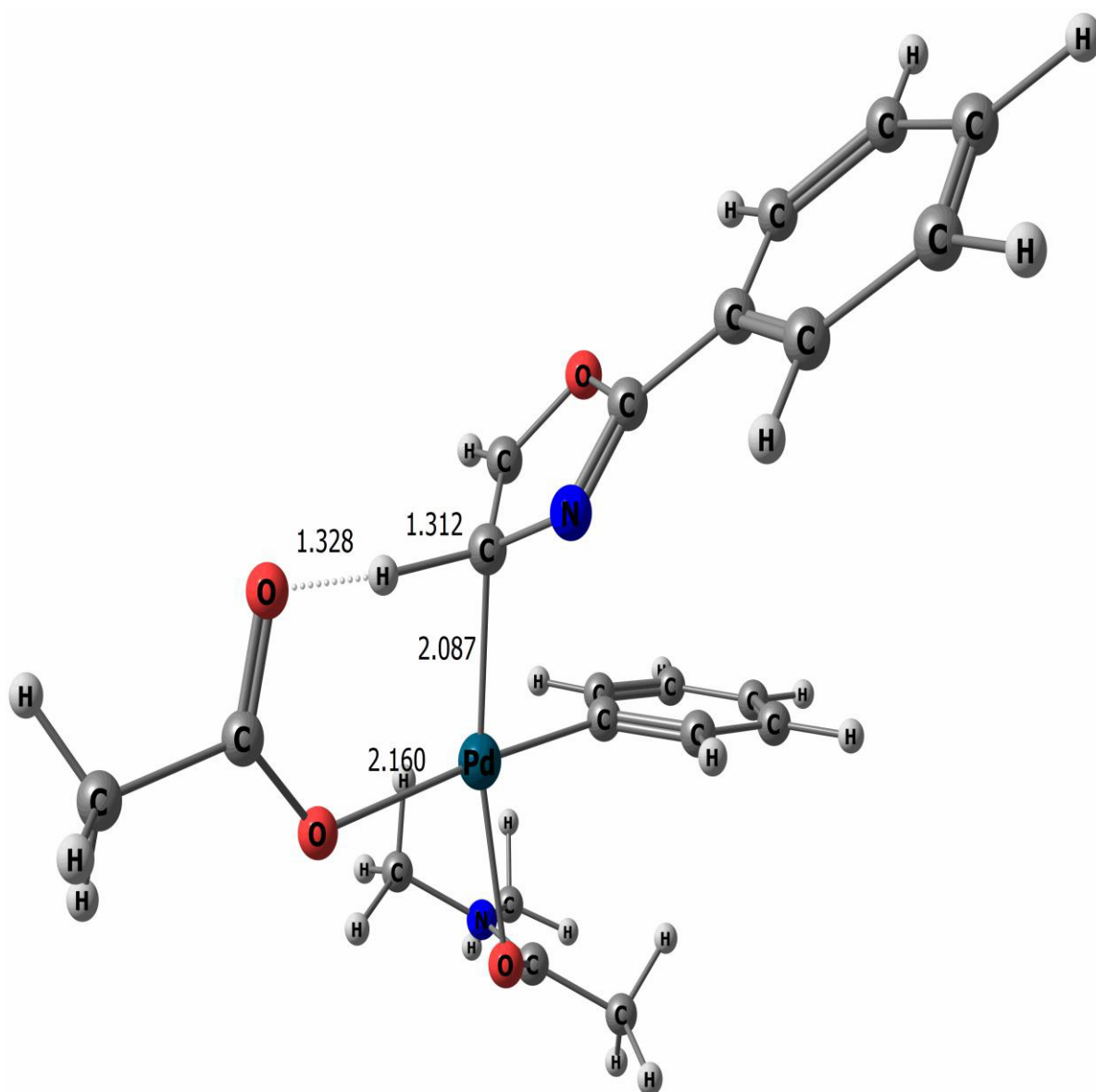


Figure 3-61: Intermediate 1 Pathway 2 $\rightarrow$ 2,4 (= coordinated)

The nitrogen coordinated intermediate is lower in Gibbs energy, by 2.23 kcal.mol⁻¹. We will therefore model the subsequent reactions starting from this species and the alternative will not be further considered.

The CMD transition state (TS 1) for Pathway 2 $\rightarrow$ 2,4 is depicted in Figure 3-62, along with selected bond lengths (all in Å). This structure has an imaginary frequency of -674.51 cm⁻¹. The corresponding Intermediate 2 structure is portrayed in Figure 3-63 and the reductive elimination transition state (TS 2) in Figure 3-64. This structure has an imaginary frequency of -338.75 cm⁻¹. The final di-substituted product, Product 2,4 is pictured in Figure 3-65.



[Figure 3-62: Transition State 1 Pathway 2 → 2,4](#)

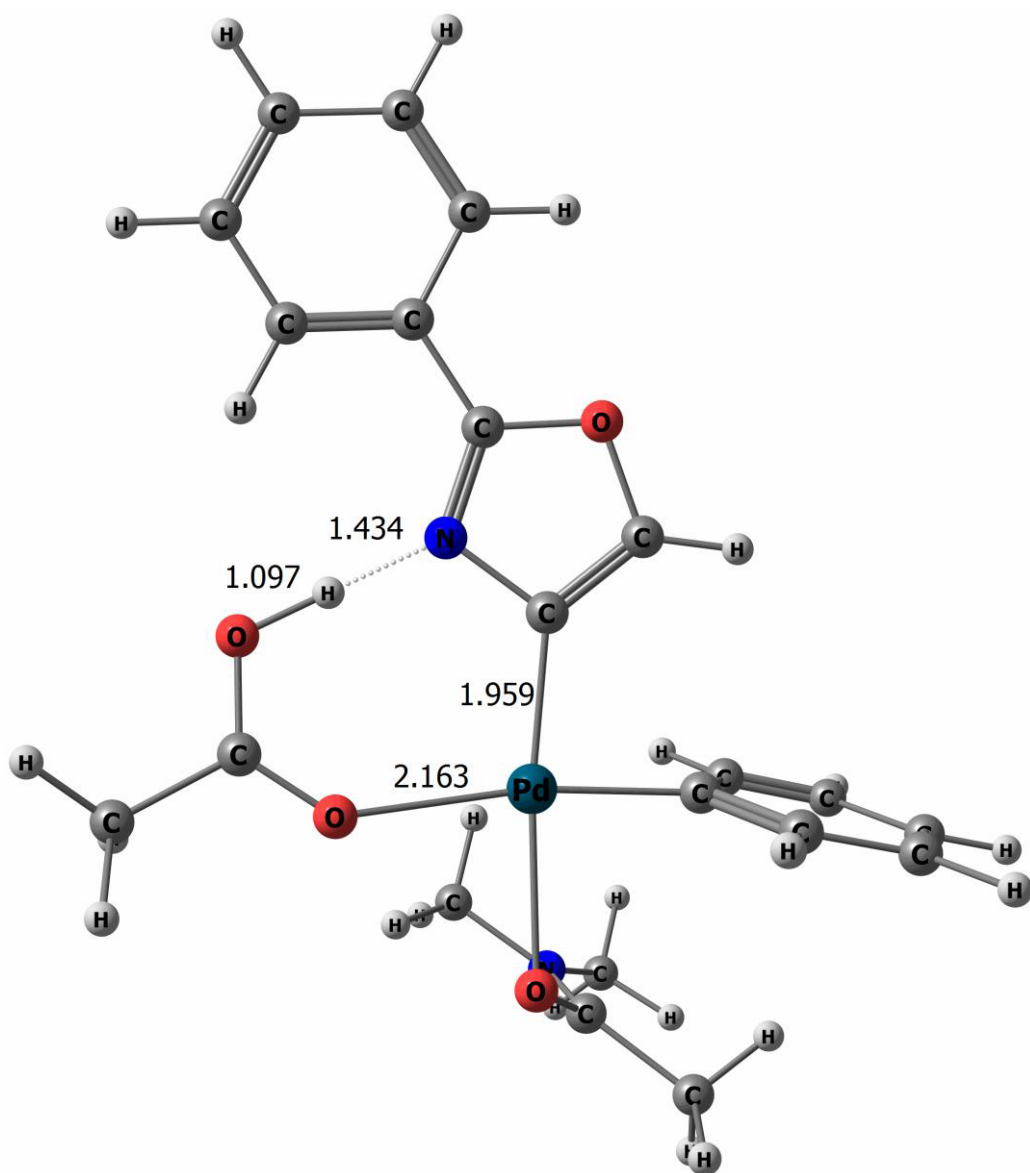
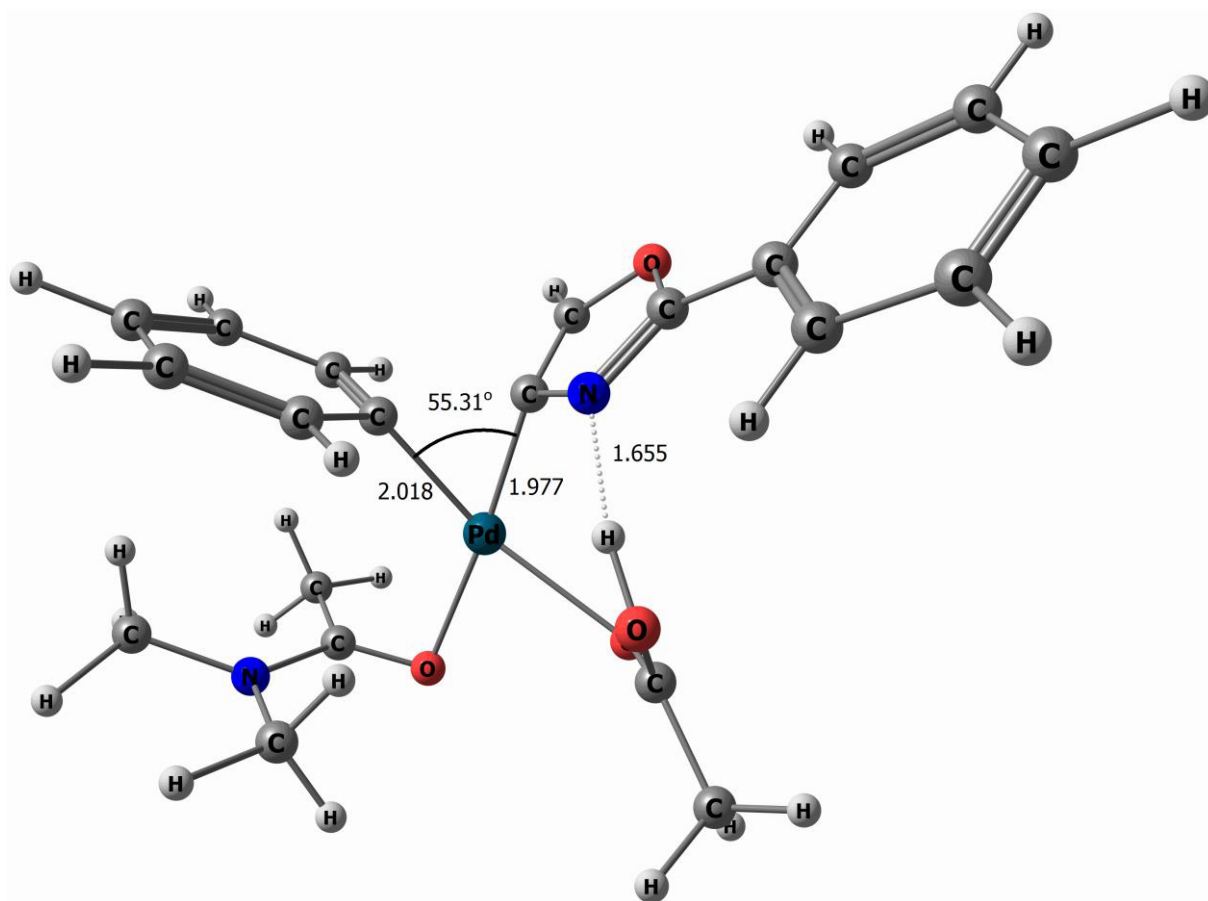
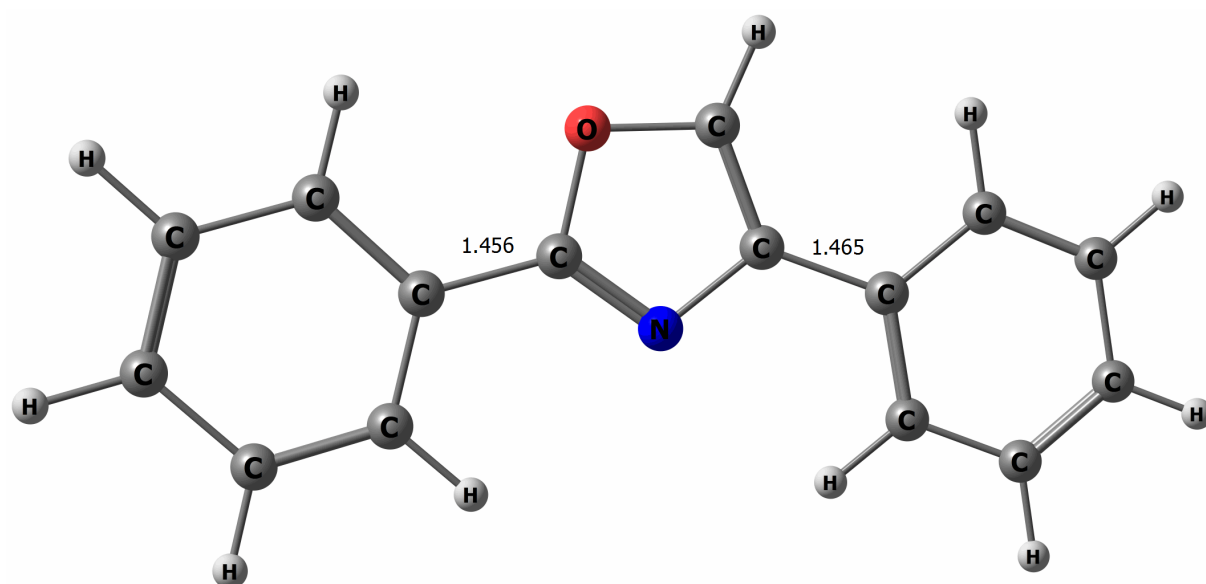


Figure 3-63: Intermediate 2 Pathway 2 → 2,4



[Figure 3-64: Transition State 2 Pathway 2 → 2,4](#)



[Figure 3-65: Product 2,4](#)

Pathway 4 → 2,4:

Starting from Product 4 entering the catalytic cycle for a second time, with the requirement that position 2 is aligned to be deprotonated, leads to the Intermediate 1 structure depicted in Figure 3-66. Another possibility is coordination of Product 4 to palladium via the double

bond between carbons 4 and 5 (Figure 3-67). However, the double bond coordinated intermediate is 7.22 kcal.mol⁻¹ higher in Gibbs energy, and will not be considered further.

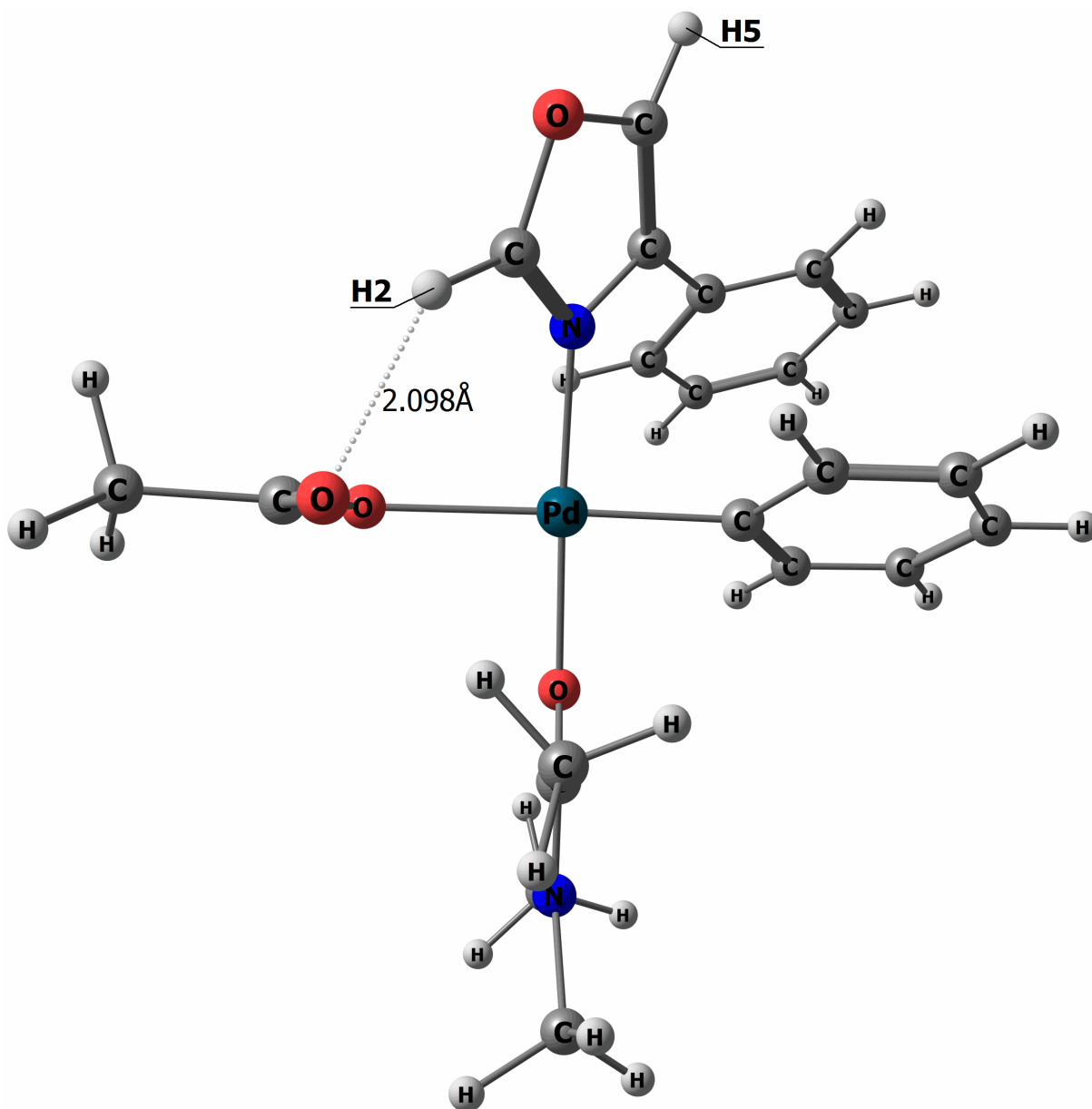


Figure 3-66: Intermediate 1 Pathway 4 → 2,4 (Nitrogen Coordinated)

The CMD transition state (TS 1) for Pathway 4 → 2,4 is portrayed in Figure 3-68. This structure has an imaginary frequency of -441.46 cm⁻¹.

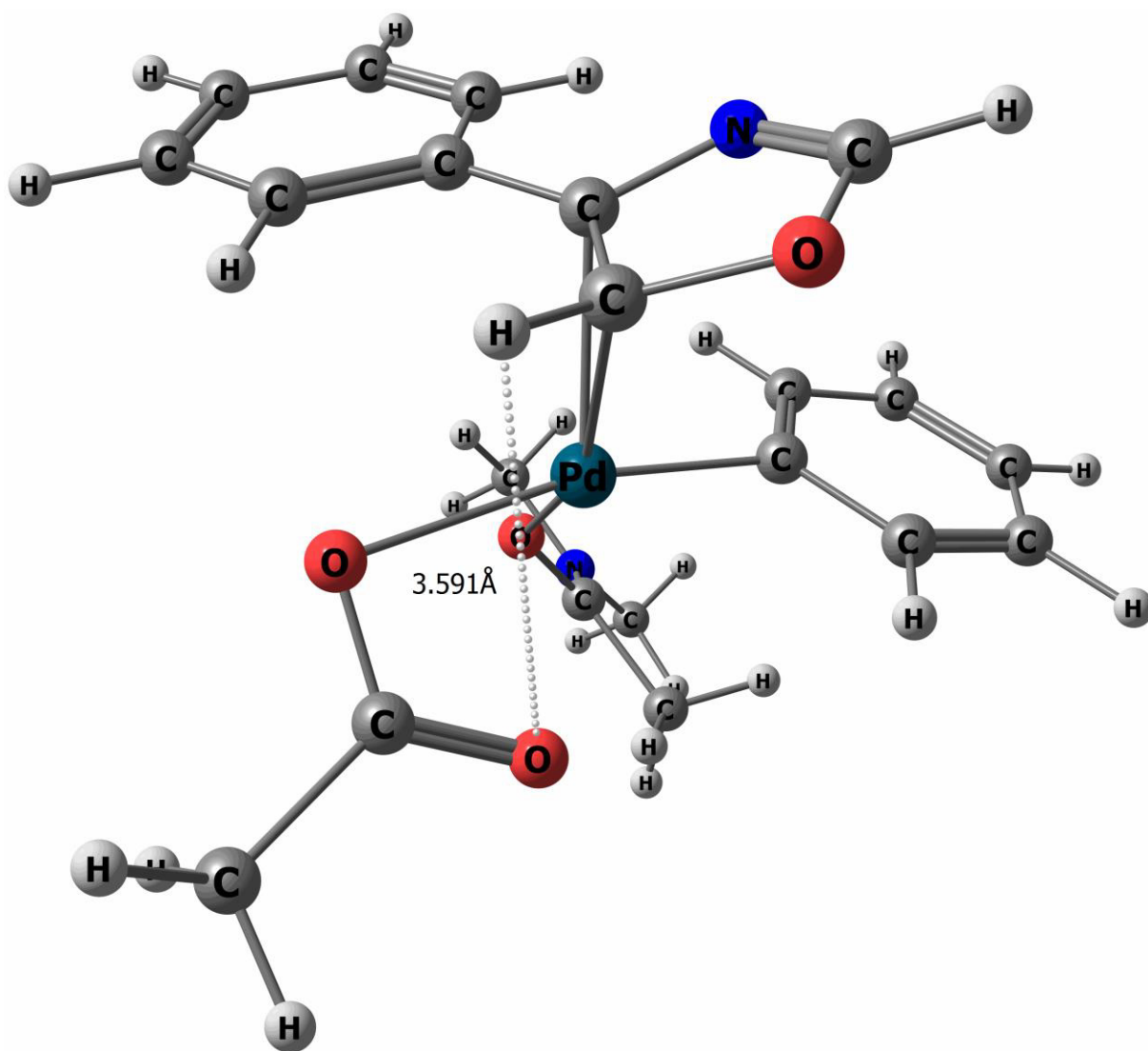
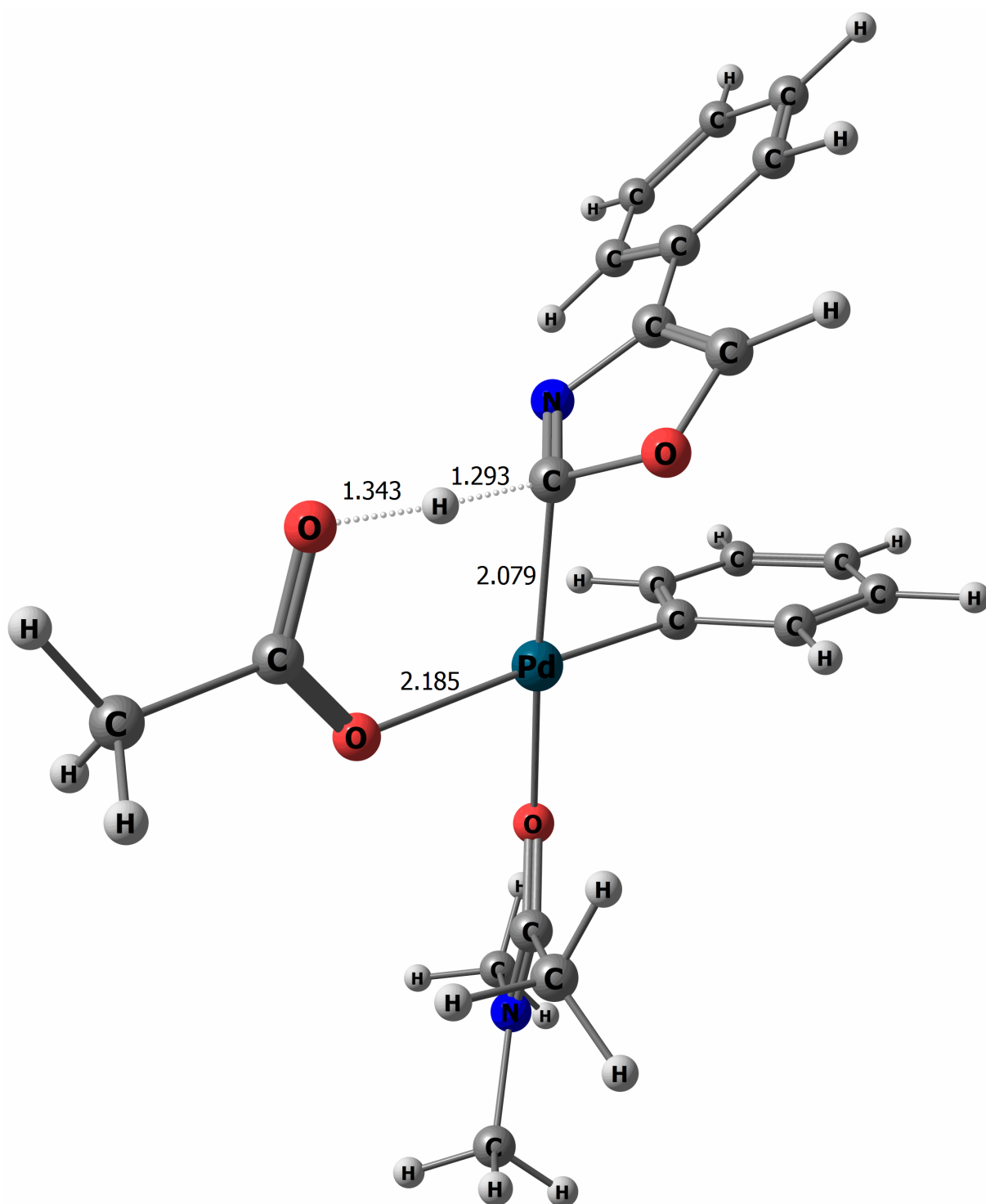
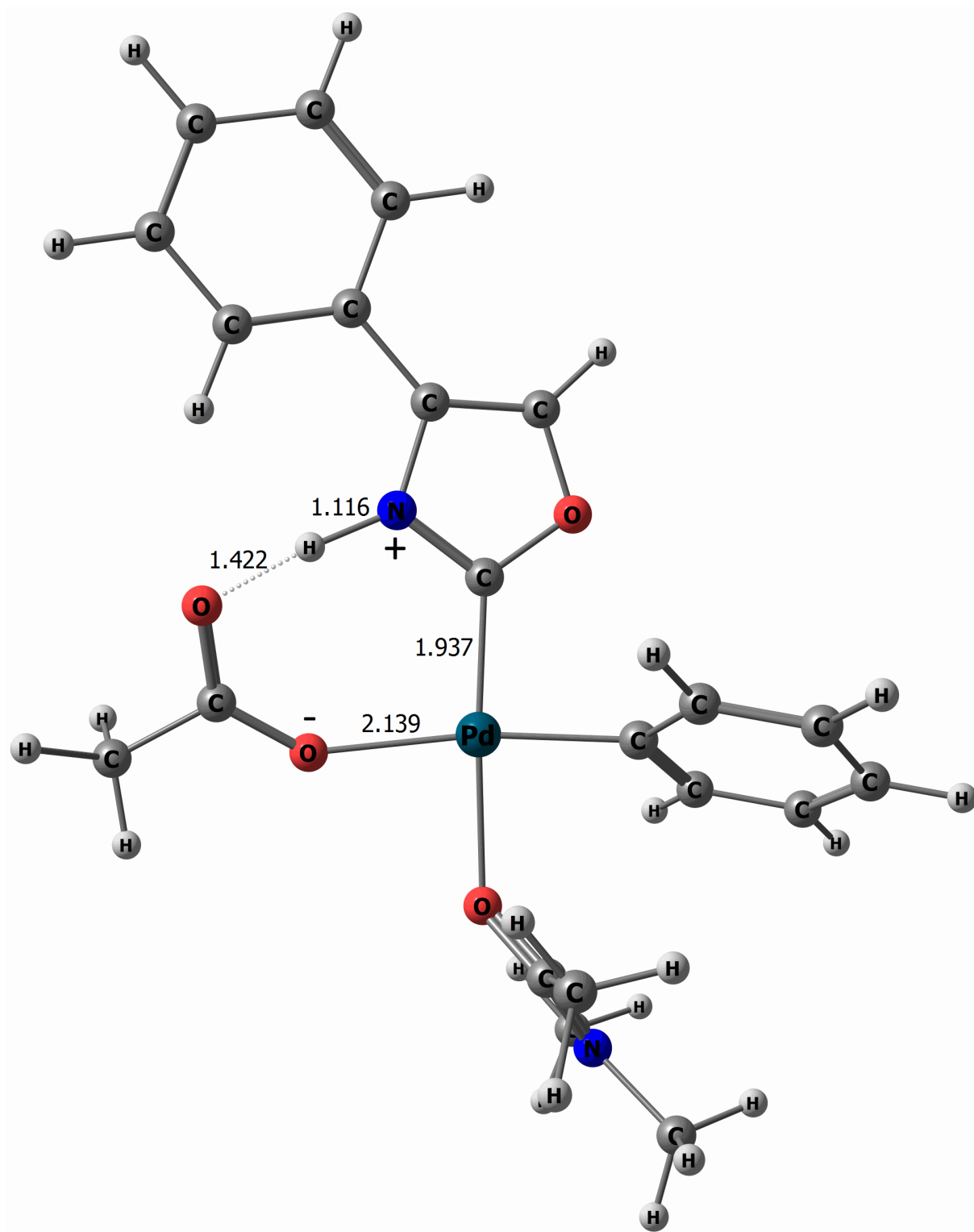


Figure 3-67: Intermediate 1 Pathway 4 → 2,4 (= coordinated)

The corresponding Intermediate 2 structure is shown in Figure 3-69. Just as in Intermediate 2 for Pathway N2 (Figure 3-32), deprotonation of the acetate ligand, by the nitrogen atom of oxazole takes place to generate a zwitterionic species. Therefore, the steric repulsion due to the benzene substituent at position 4 is not large enough to disrupt this interaction or the resultant hydrogen bond, illustrating its strength. This interaction persists in TS2, depicted in Figure 3-70. This structure has an imaginary frequency of -357.72 cm^{-1} .



[Figure 3-68: Transition 1 Pathway 4 → 2,4](#)



[Figure 3-69: Intermediate 2 Pathway 4 → 2,4](#)

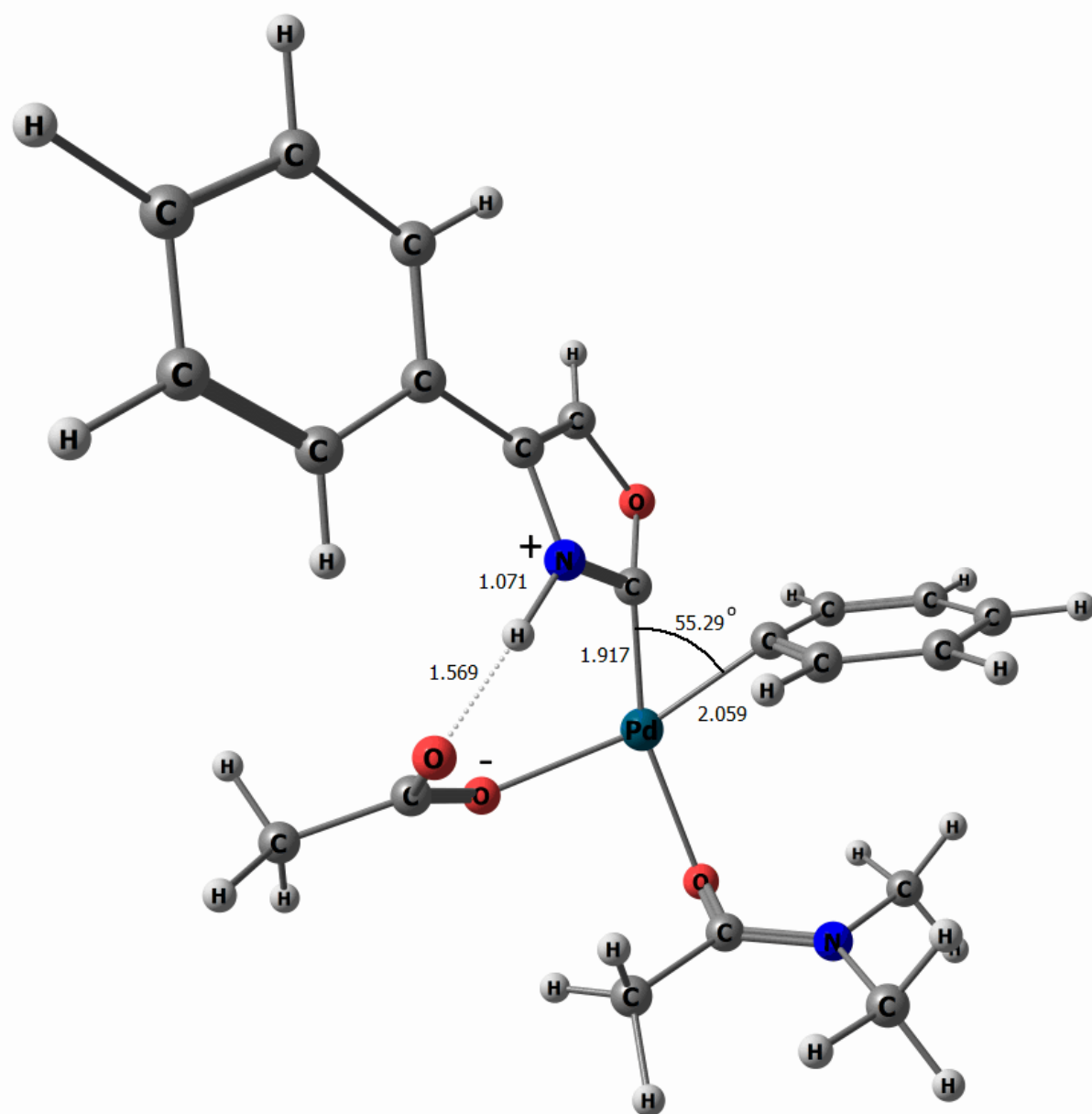
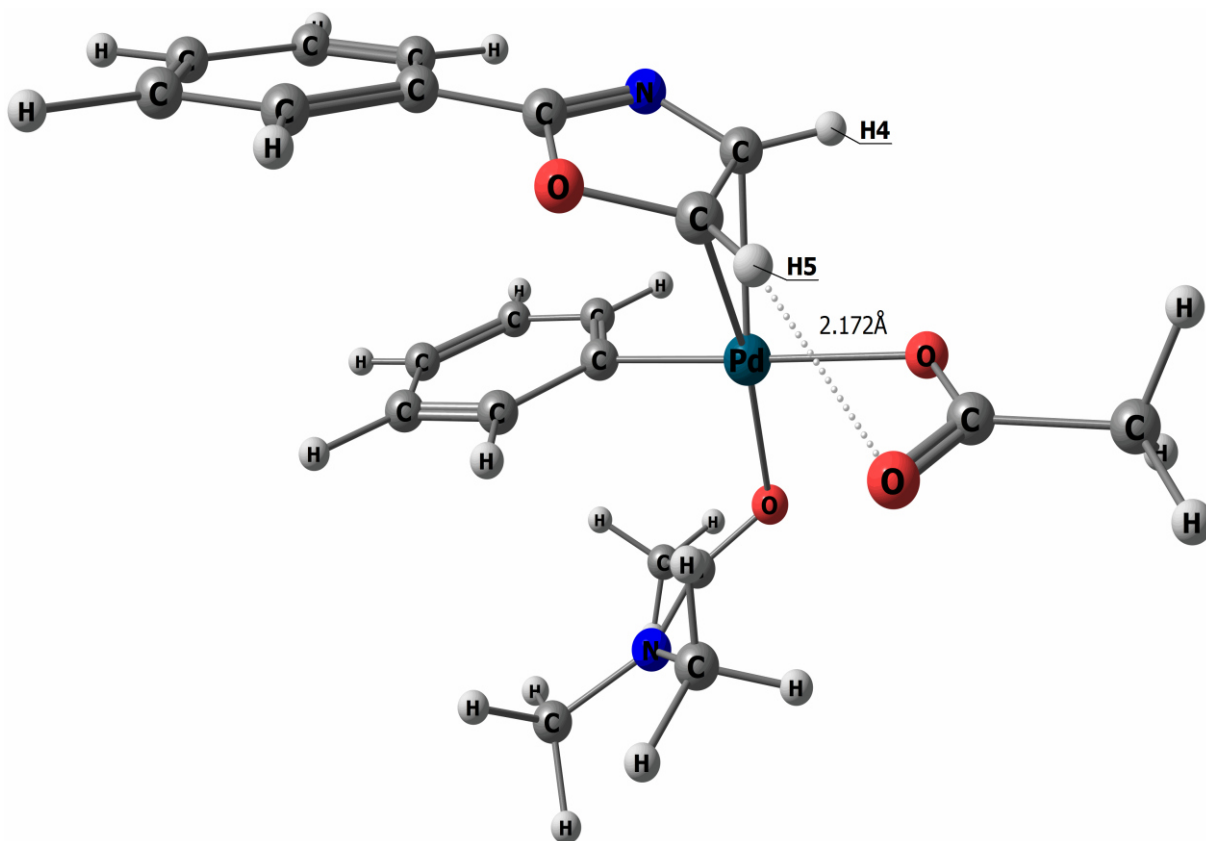


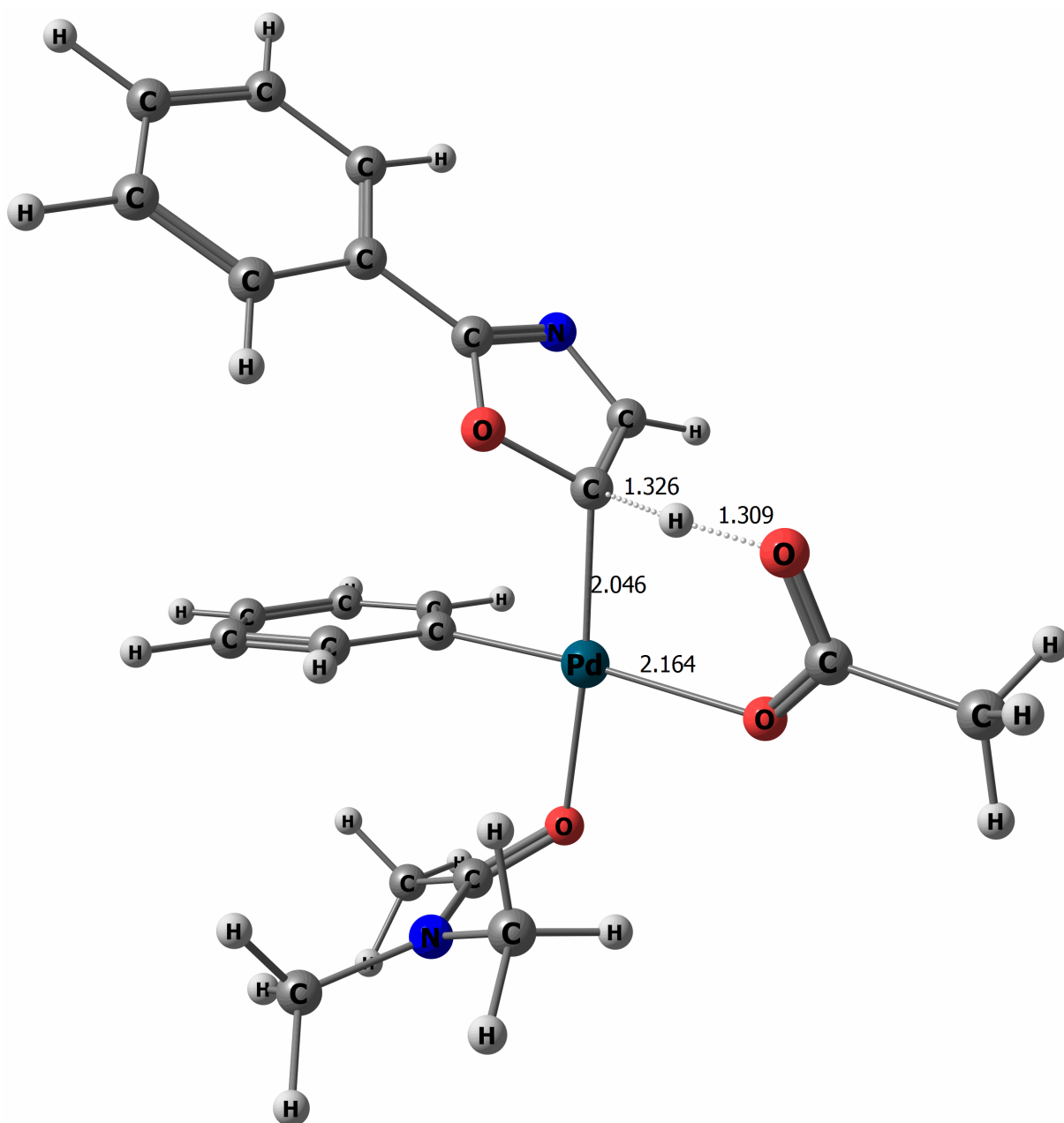
Figure 3-70: Transition State 2 Pathway 4 $\rightarrow$ 2,4

Pathway 2 $\rightarrow$ 2,5:

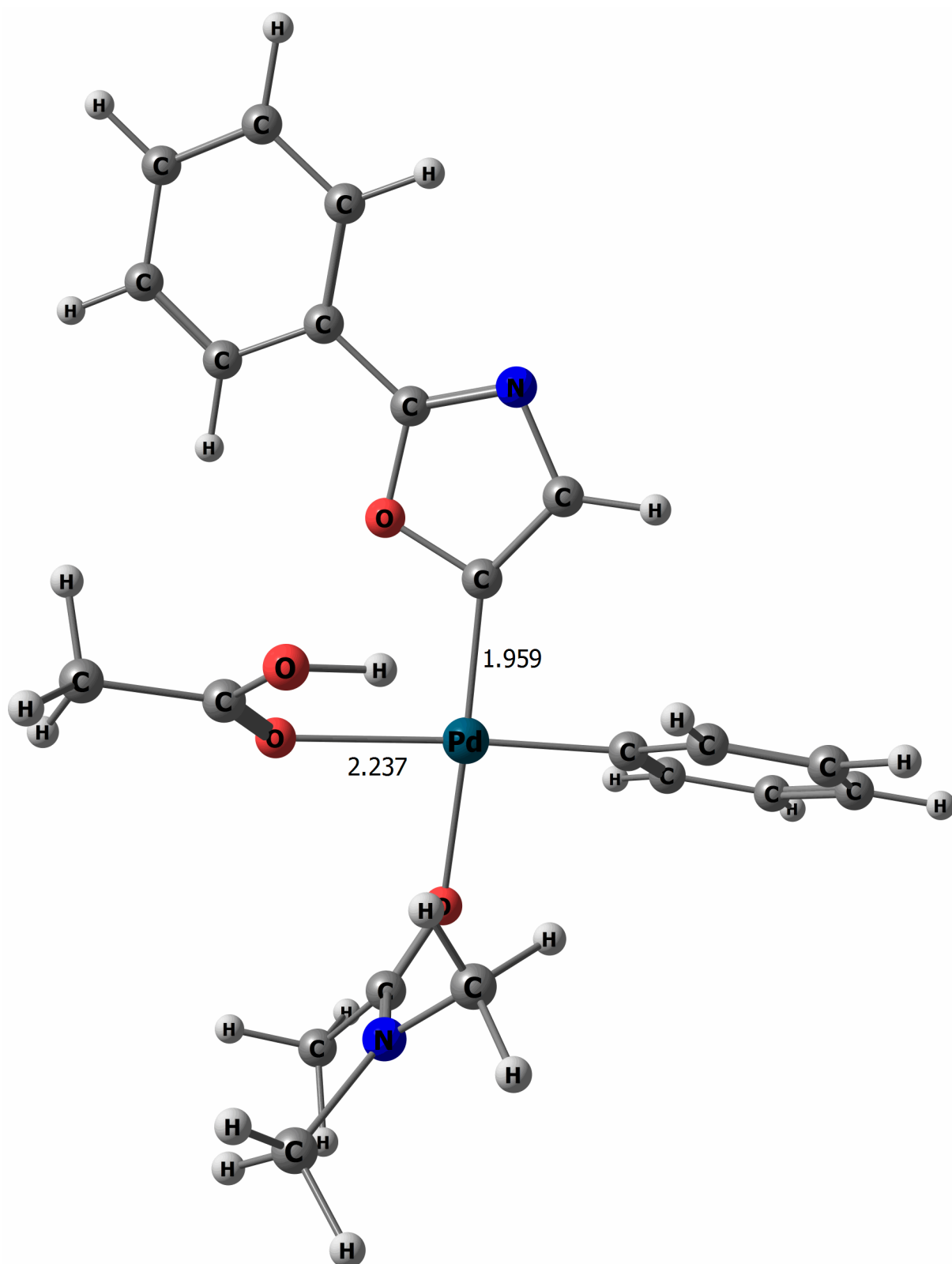
Starting from Product 2 entering the catalytic cycle for a second time, but now with position 5 aligned to be deprotonated, results in the Intermediate 1 structure illustrated in Figure 3-71. Just as with Product 5 formation, the only feasible Intermediate 1 structure has oxazole coordinated to palladium via the double bond between carbons 4 and 5. This is the beginning of Pathway 2 $\rightarrow$ 2,5. The CMD transition state (TS 1) for this pathway is illustrated in Figure 3-72 and has an imaginary frequency of -824.19 cm^{-1} . The corresponding Intermediate 2 structure is pictured in Figure 3-73 and the reductive elimination transition state (TS 2) is shown in Figure 3-74. This structure has an imaginary frequency of -354.09 cm^{-1} . The final di-substituted product, Product 2,5 is portrayed in Figure 3-75.



[Figure 3-71: Intermediate 1 Pathway 2 → 2,5](#)



[Figure 3-72: Transition State 1 Pathway 2 → 2,5](#)



[Figure 3-73: Intermediate 2 Pathway 2 → 2,5](#)

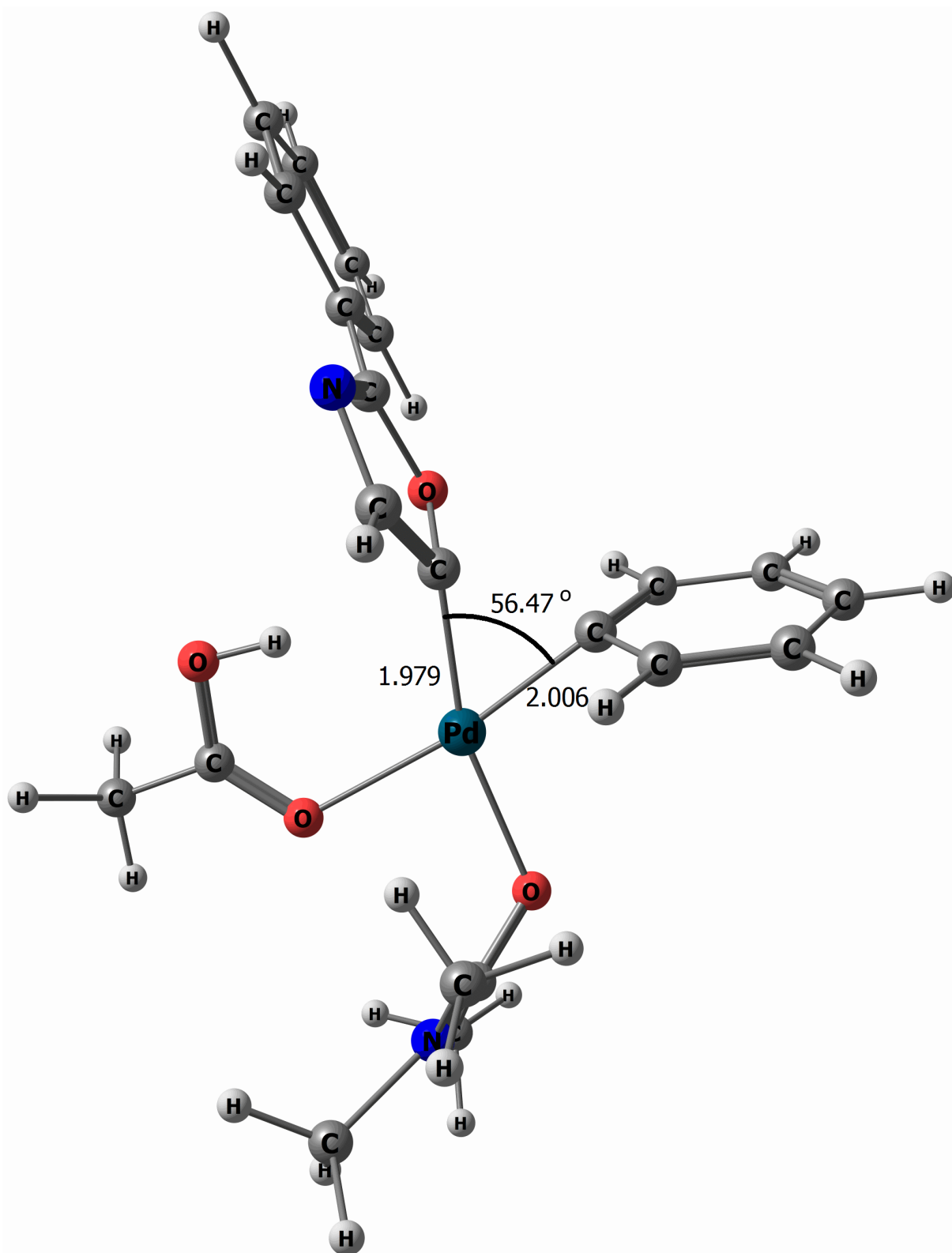


Figure 3-74: Transition State 2 Pathway 2 → 2,5

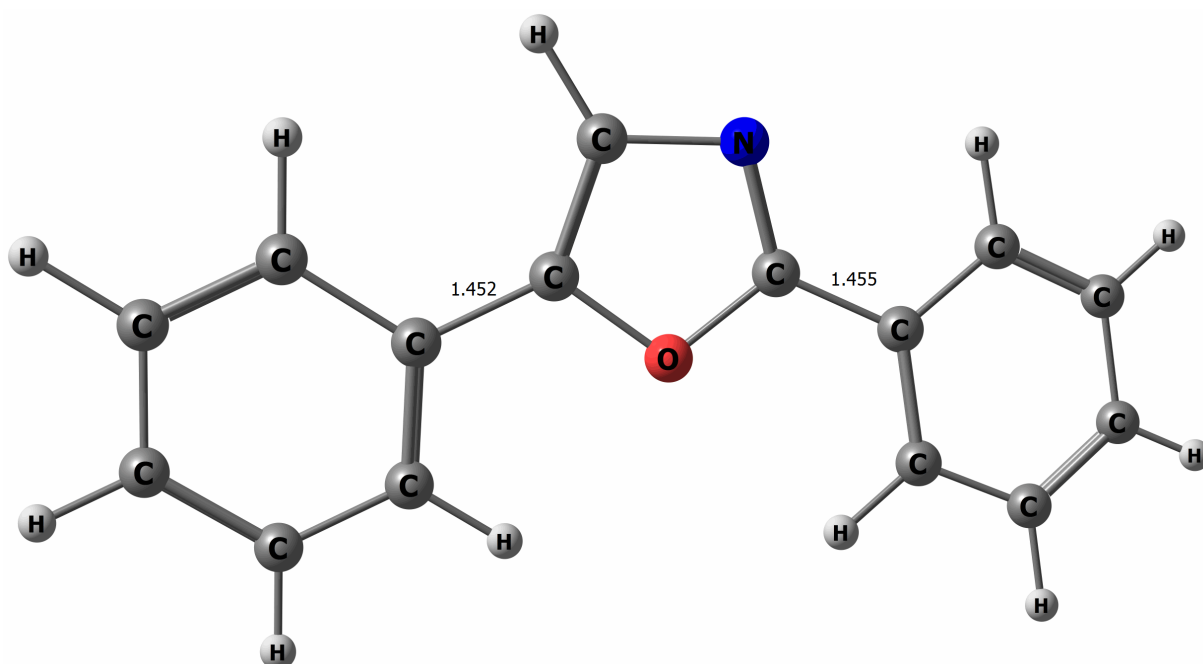
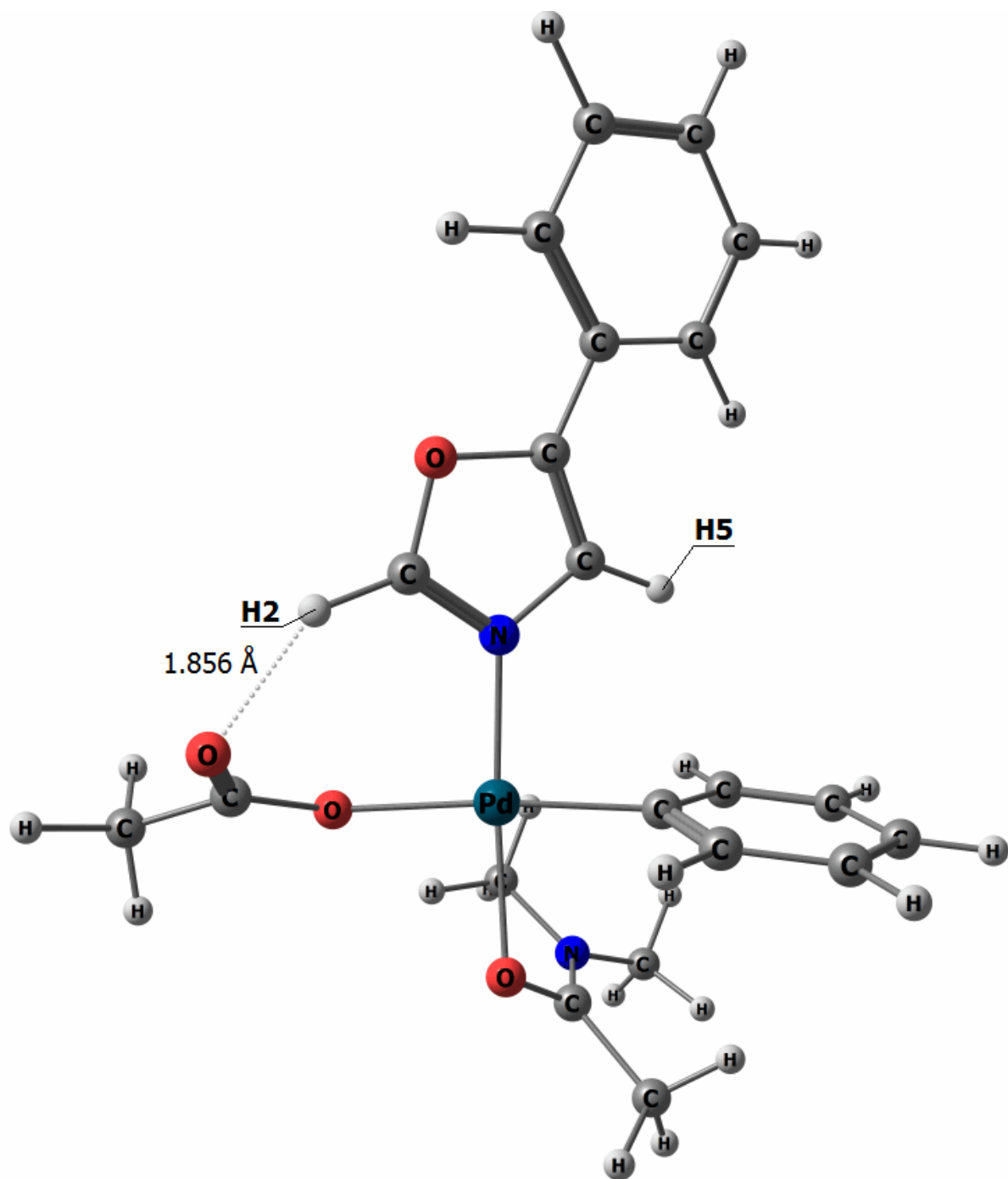


Figure 3-75: Product 2,5

Pathway 5 $\rightarrow$ 2,5:

Starting from Product 5 entering the catalytic cycle for a second time, with position 2 aligned to be deprotonated, results in the Intermediate 1 structure shown in Figure 3-76. This is the beginning of Pathway 5 $\rightarrow$ 2,5. The CMD transition state (TS 1) for this pathway is illustrated in Figure 3-77 and has an imaginary frequency of -644.73 cm^{-1} . The corresponding Intermediate 2 structure is depicted in Figure 3-78. Interesting, although a strong hydrogen bond is present, no deprotonation of the acetate ligand by the nitrogen atom of oxazole to form a zwitterionic species occurs. This is in contrast to the Intermediate 2 structures of other pathways in which substitution at the 2 position of oxazole takes place; see Figure 3-43 and Figure 3-63 for Intermediate 2 structures corresponding to Pathway N2 and Pathway 2 $\rightarrow$ 2,4 respectively. However, TS2 shown in Figure 3-79, is stabilized by deprotonation of the acetate. This reductive elimination transition state has an imaginary frequency of -378.84 cm^{-1} .



[Figure 3-76: Intermediate 1 Pathway 5 → 2,5](#)

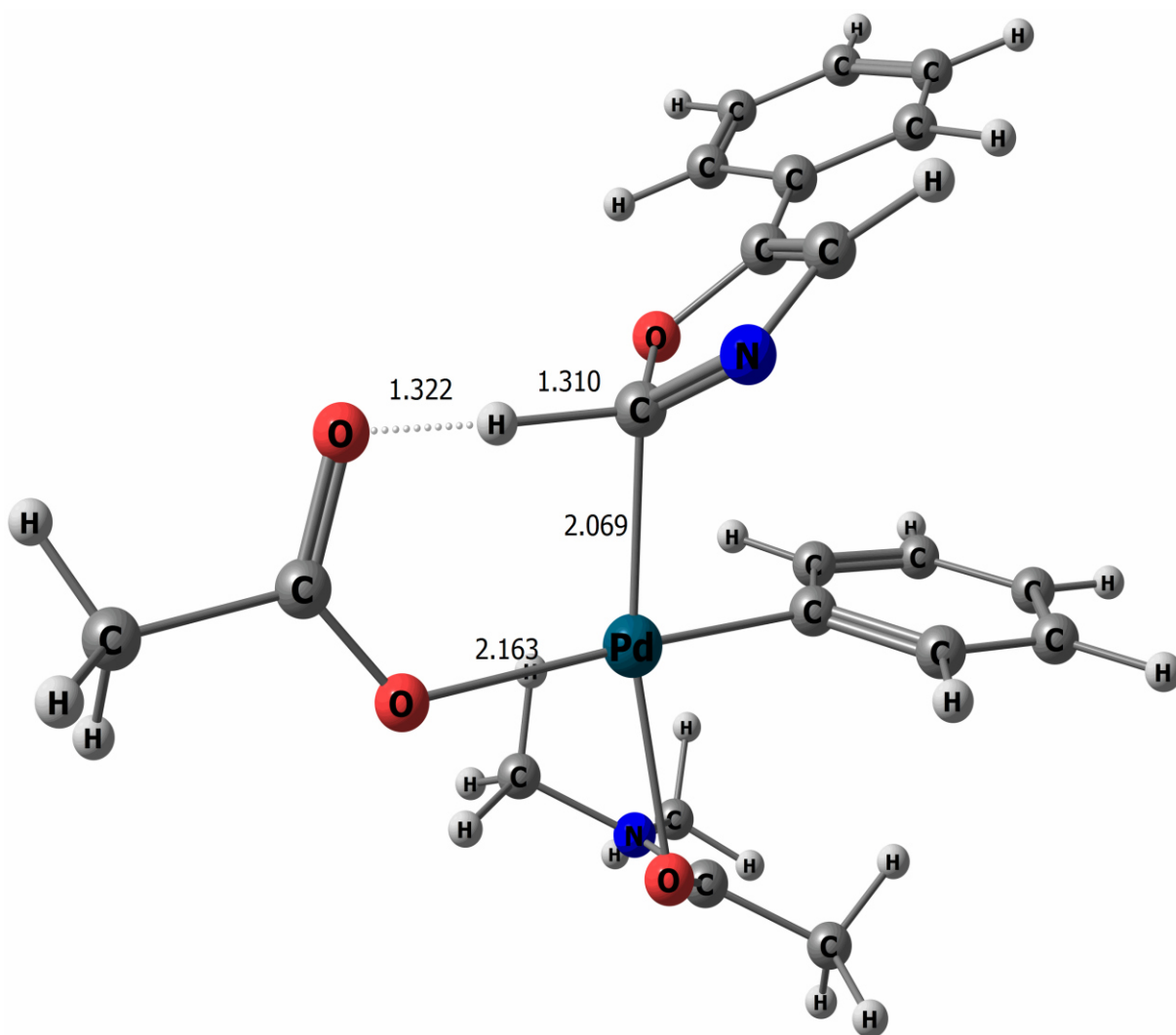
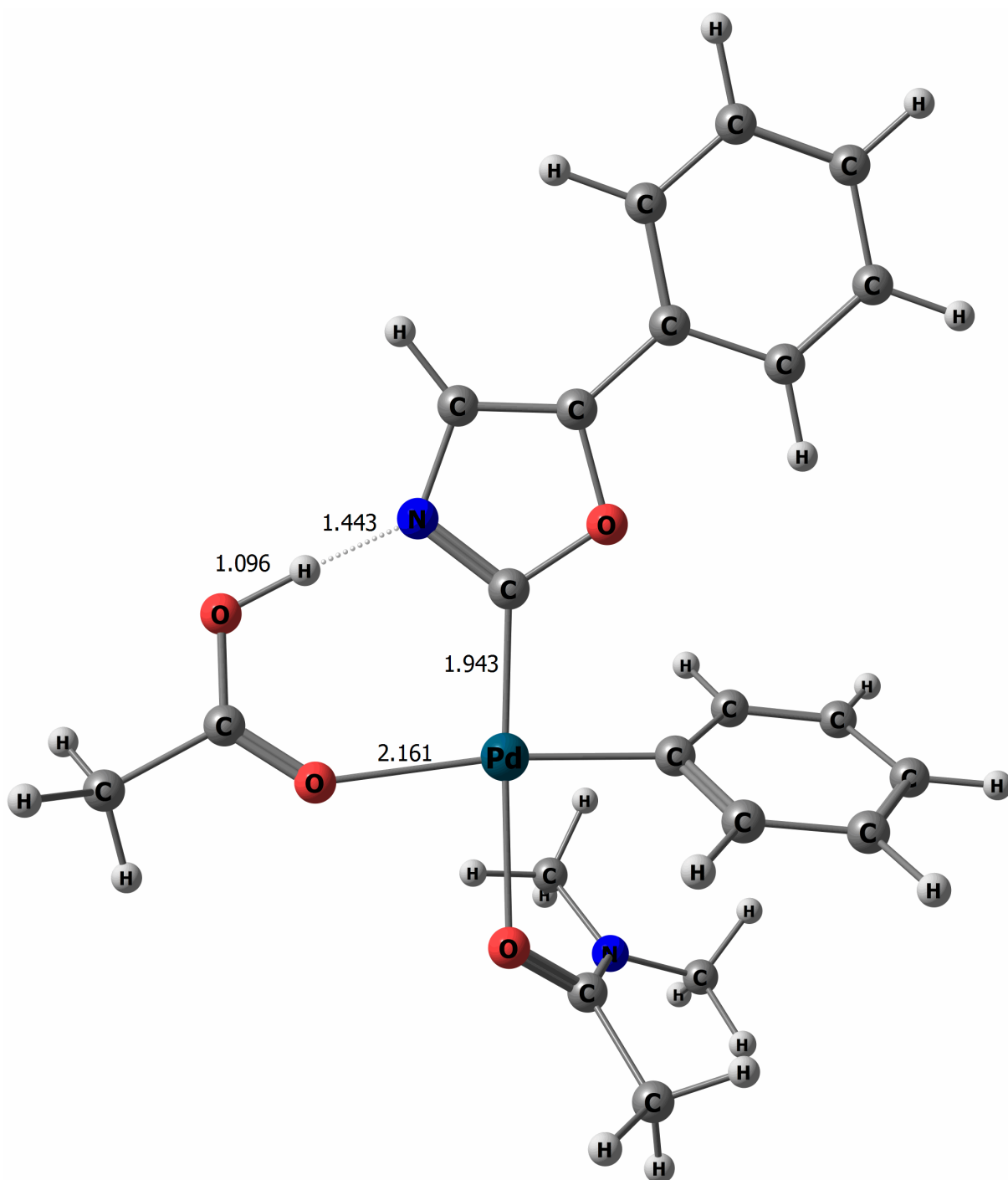


Figure 3-77: Transition State 1 Pathway 5 → 2,5



[Figure 3-78: Intermediate 2 Pathway 5 → 2,5](#)

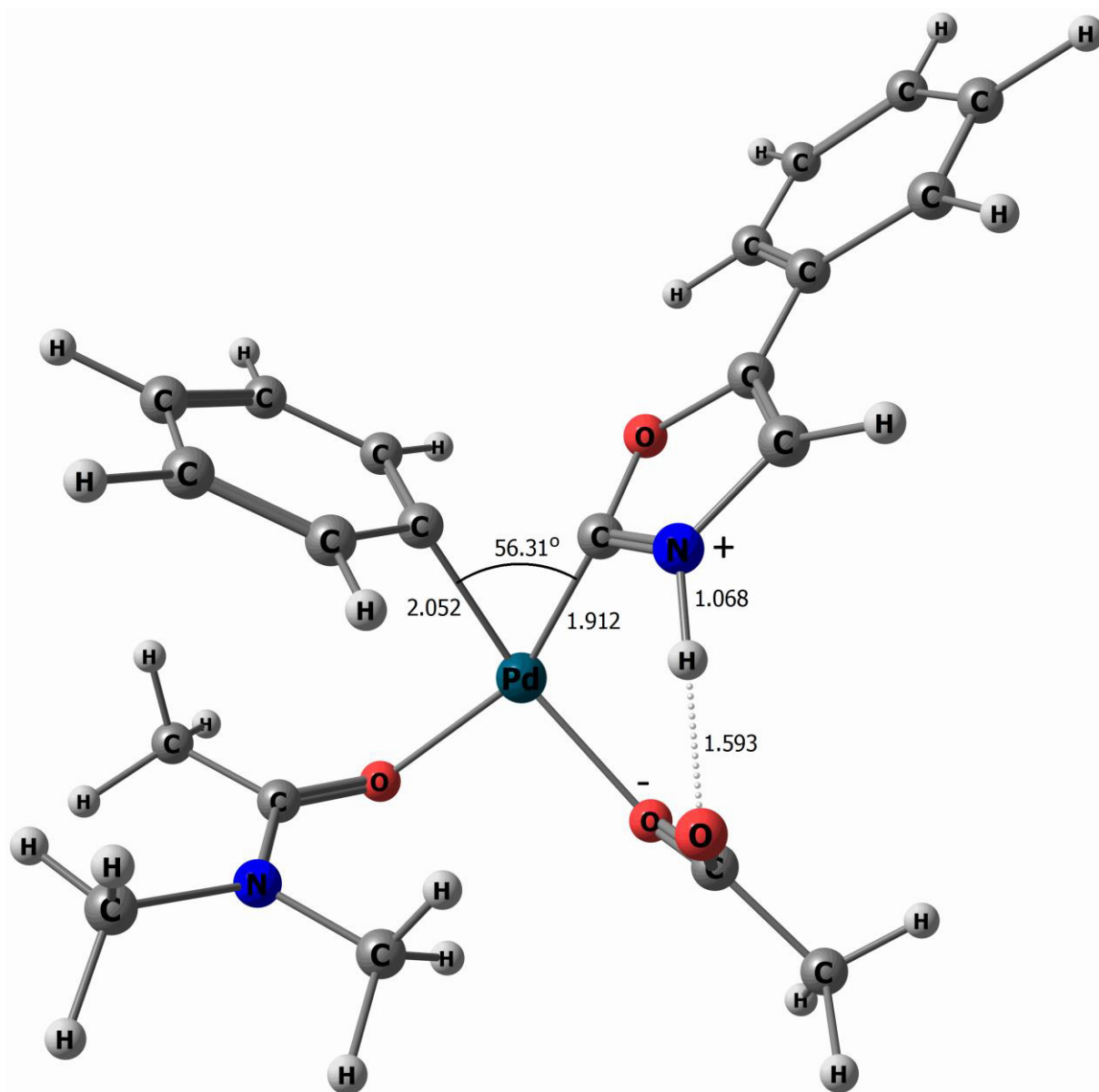


Figure 3-79: Transition State 2 Pathway 5 → 2,5

Pathway 4 → 4,5:

Starting from Product 4 entering the catalytic cycle for a second time, with the requirement that H5 is aligned to be deprotonated, leads to the Intermediate 1 structure shown in Figure 3-80. The CMD transition state (TS 1) for this pathway is portrayed in Figure 3-81 and has an imaginary frequency of -219.79 cm^{-1} . The corresponding Intermediate 2 structure is depicted in Figure 3-82, while the reductive elimination transition state (TS 2) is shown in Figure 3-83.

This structure has an imaginary frequency of -329.29 cm^{-1} . The final di-substituted product, Product 4,5, is illustrated in Figure 3-84.

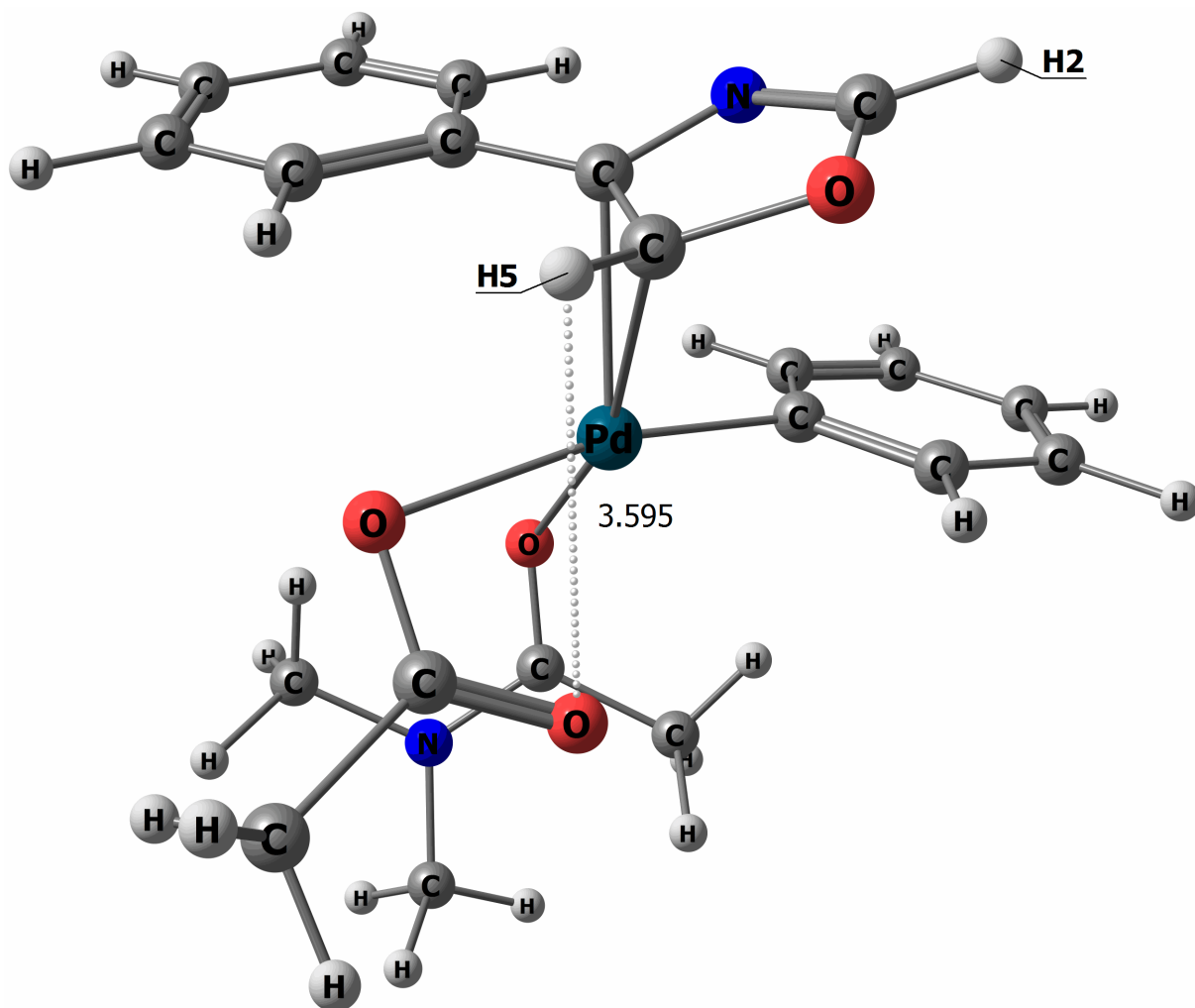
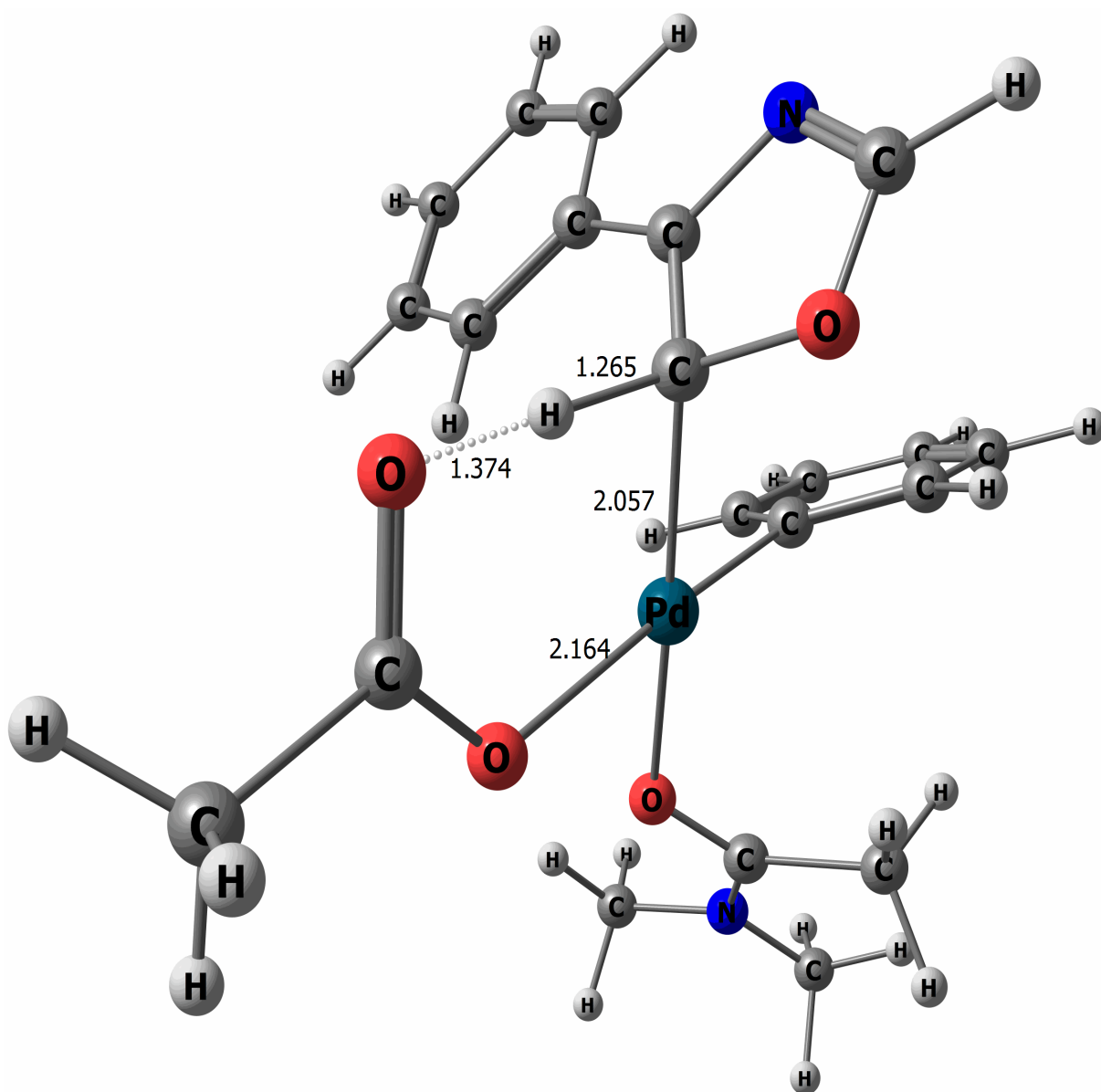


Figure 3-80: Intermediate 1: Pathway 4 → 4,5



[Figure 3-81: Transition State 1 Pathway 4 → 4,5](#)

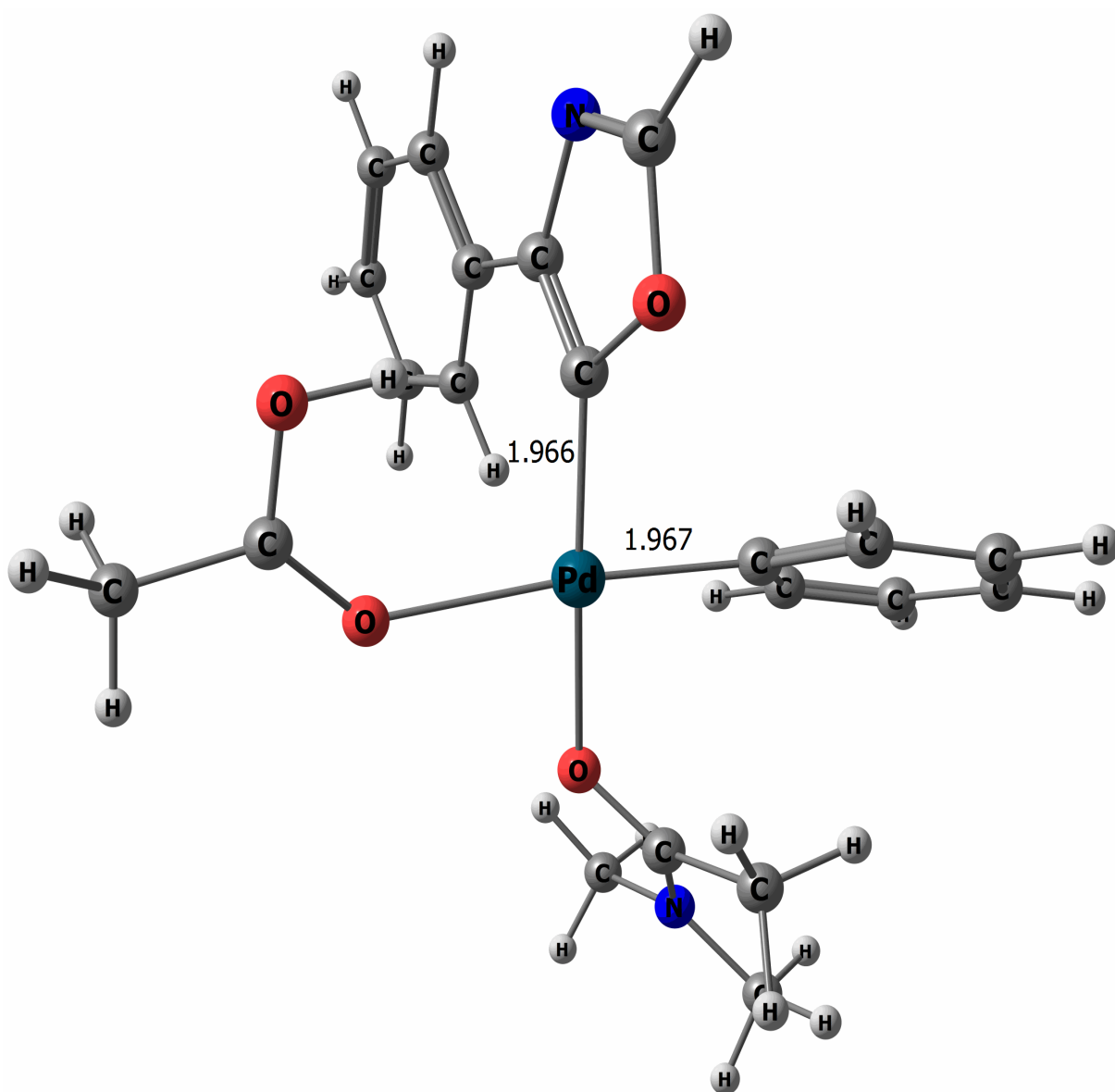
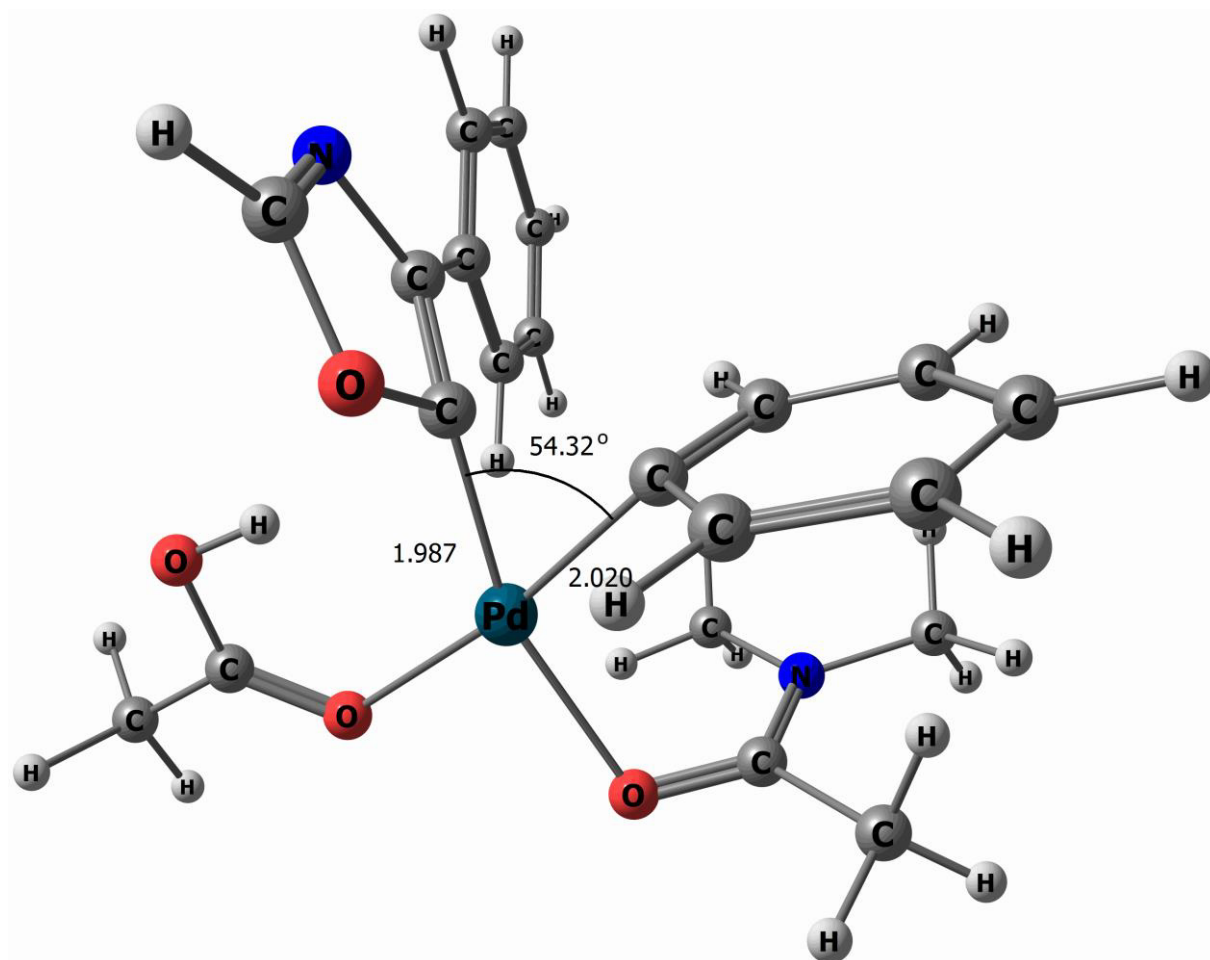
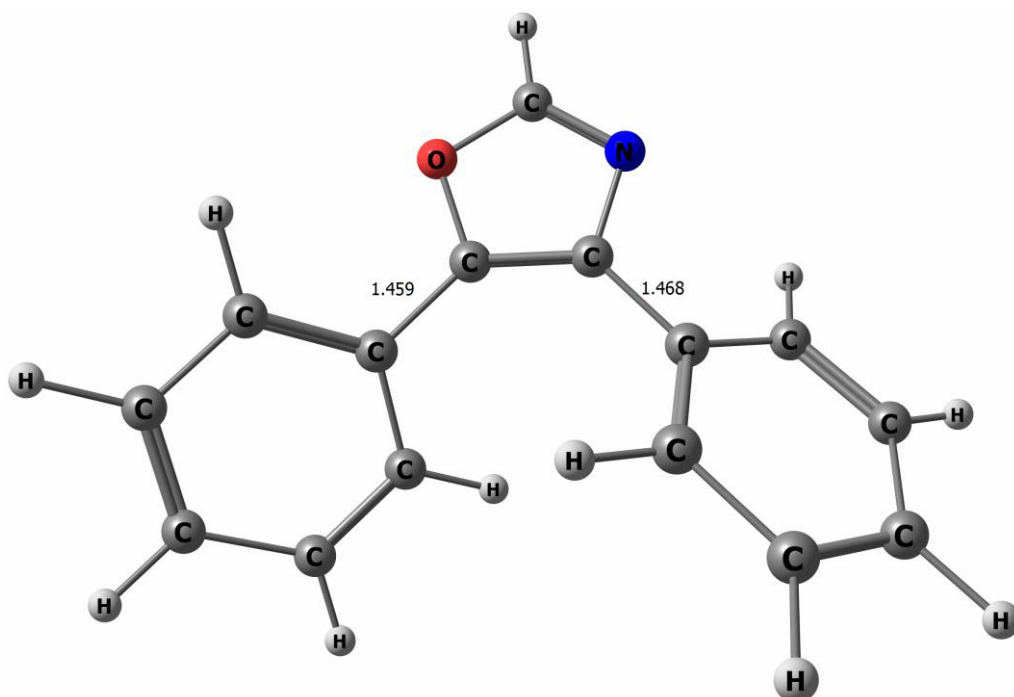


Figure 3-82: Intermediate 2 Pathway 4 → 4,5



[Figure 3-83: Transition State 2 Pathway 4 → 4,5](#)



[Figure 3-84: Product 4,5](#)

Pathway 5 $\rightarrow$ 4,5:

Starting from Product 5 entering the catalytic cycle for a second time, there are two ways that this molecule can coordinate to the active catalyst, with position 4 aligned to be deprotonated; via the Nitrogen atom (Figure 3-85) or via the double bond between carbons 4 and 5 (Figure 3-86).

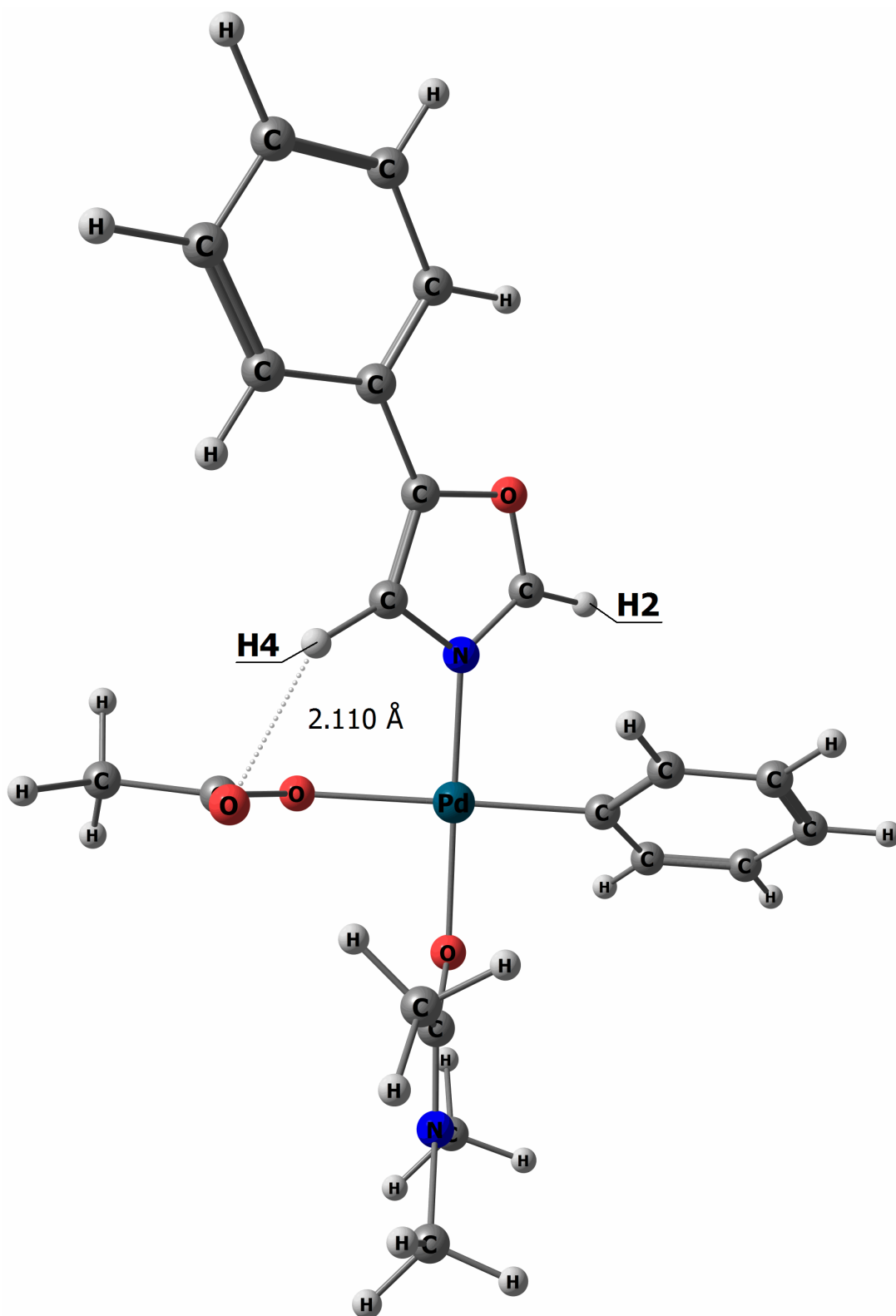


Figure 3-85: Intermediate 1 Pathway 5 $\rightarrow$ 4,5 (Nitrogen Coordinated)

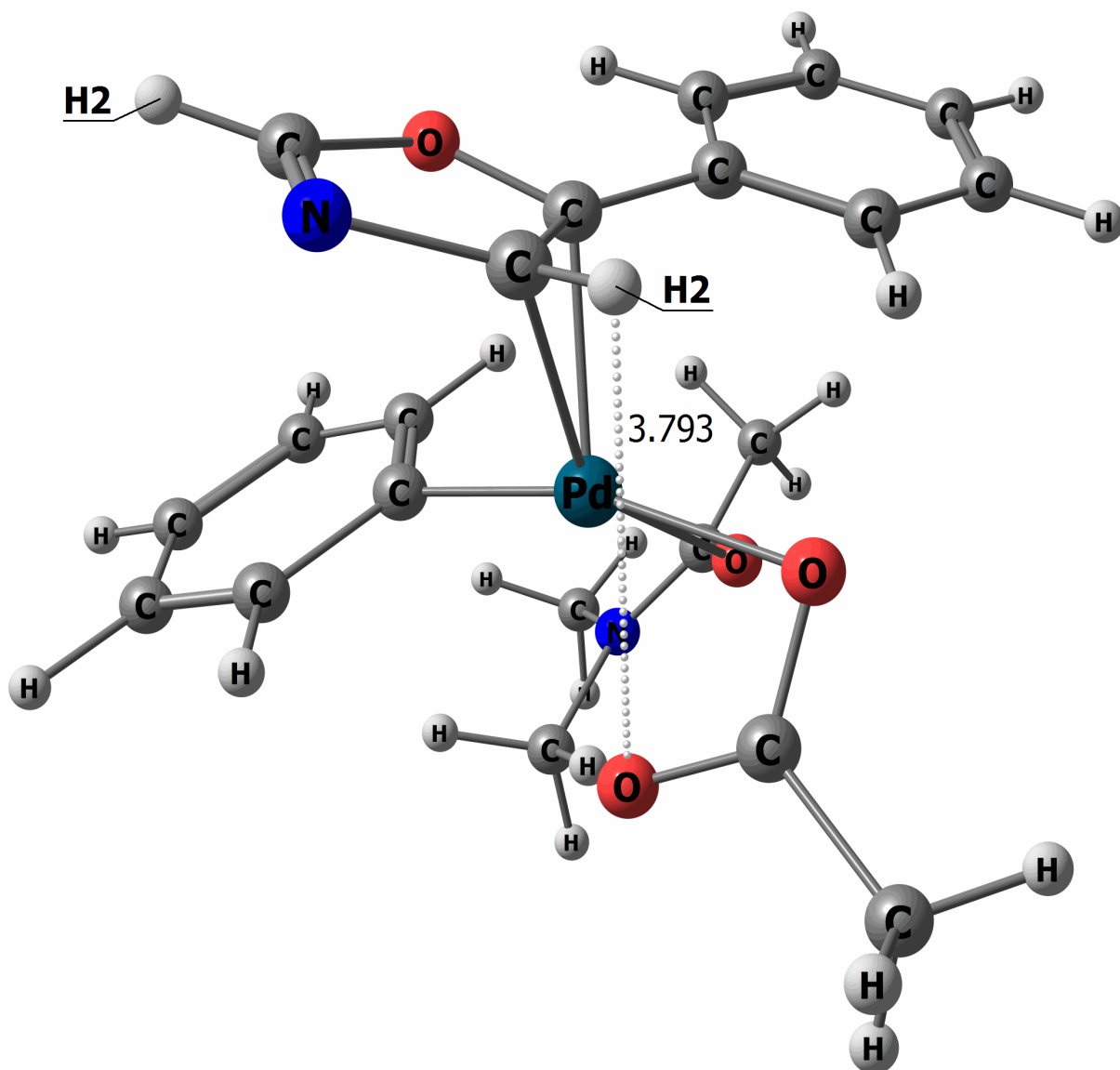
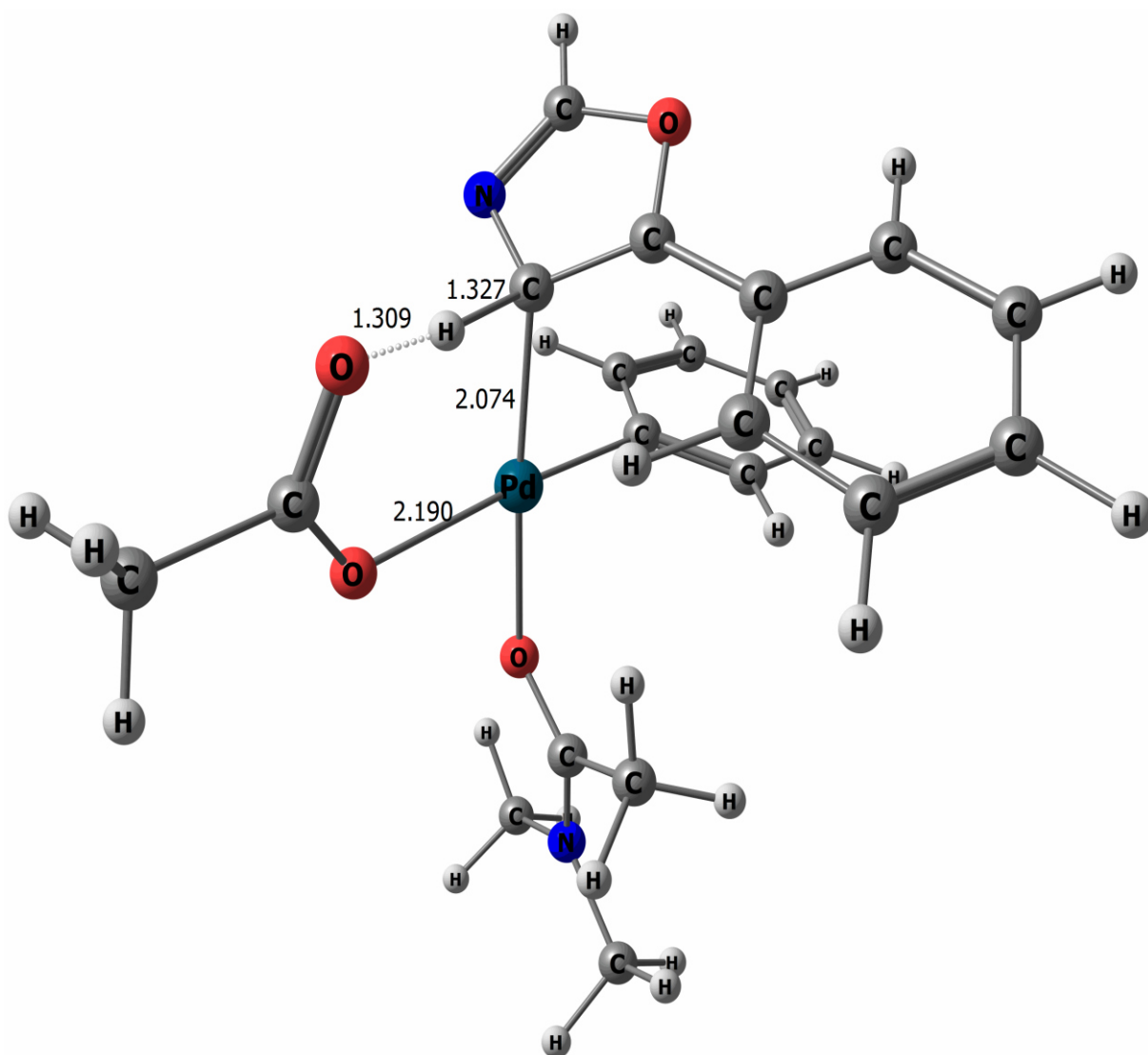
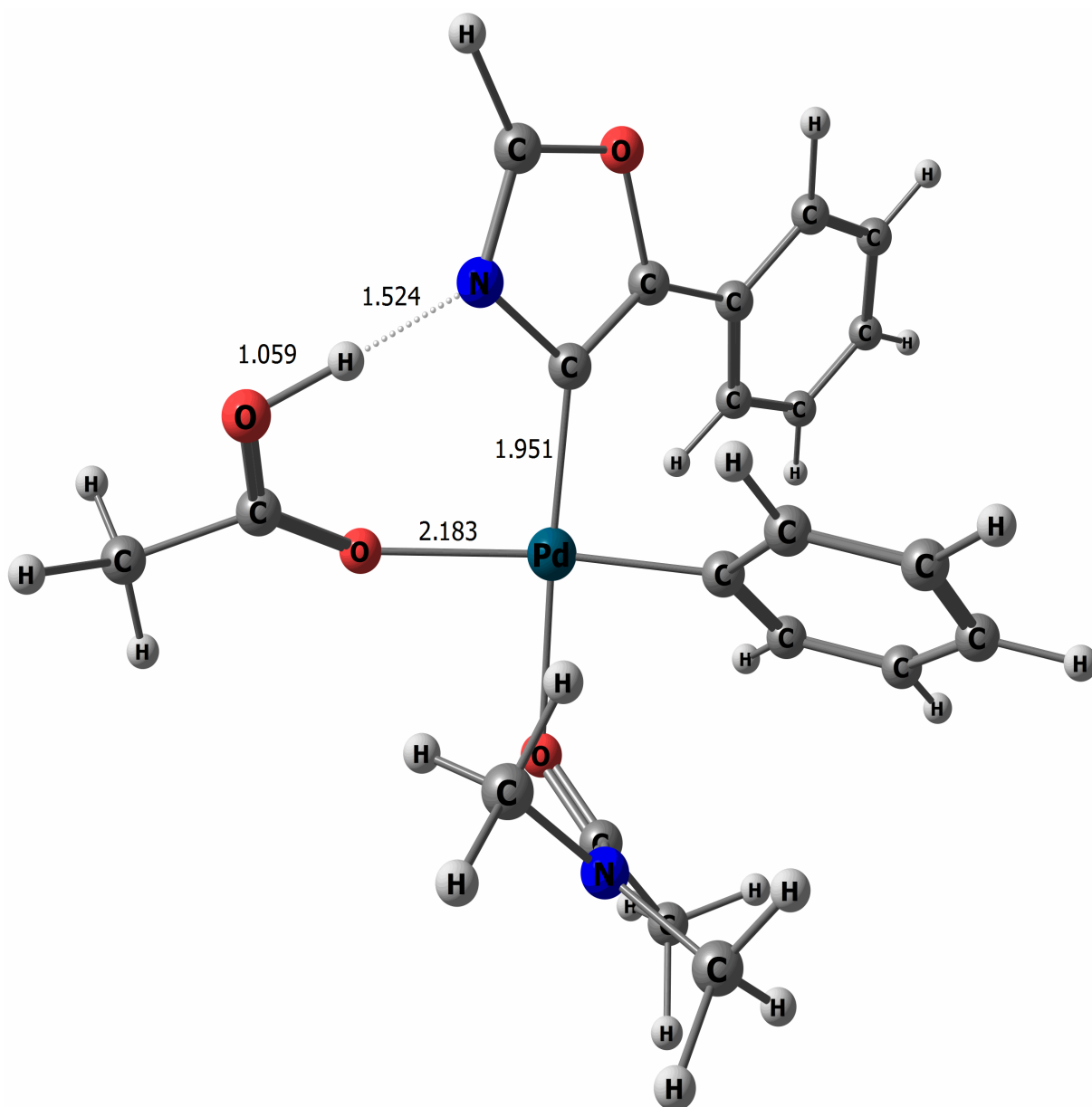


Figure 3-86: Intermediate 1 Pathway 5 → 4,5 (= coordinated)

The nitrogen coordinated intermediate is lower in energy by 4.39 kcal.mol⁻¹, as expected based on previous results. We will therefore model the subsequent reactions starting from this species and the double bond coordinated alternative will not be further considered. The CMD transition state (TS 1) for this pathway is portrayed in Figure 3-87 and has an imaginary frequency of -807.76 cm⁻¹. The corresponding Intermediate 2 structure is shown in Figure 3-88 and the reductive elimination transition state (TS 2) is illustrated in Figure 3-89. This structure has an imaginary frequency of -335.9 cm⁻¹.



[Figure 3-87: Transition State 1 Pathway 5 → 4,5](#)



[Figure 3-88: Intermediate 2 Pathway 5 → 4,5](#)

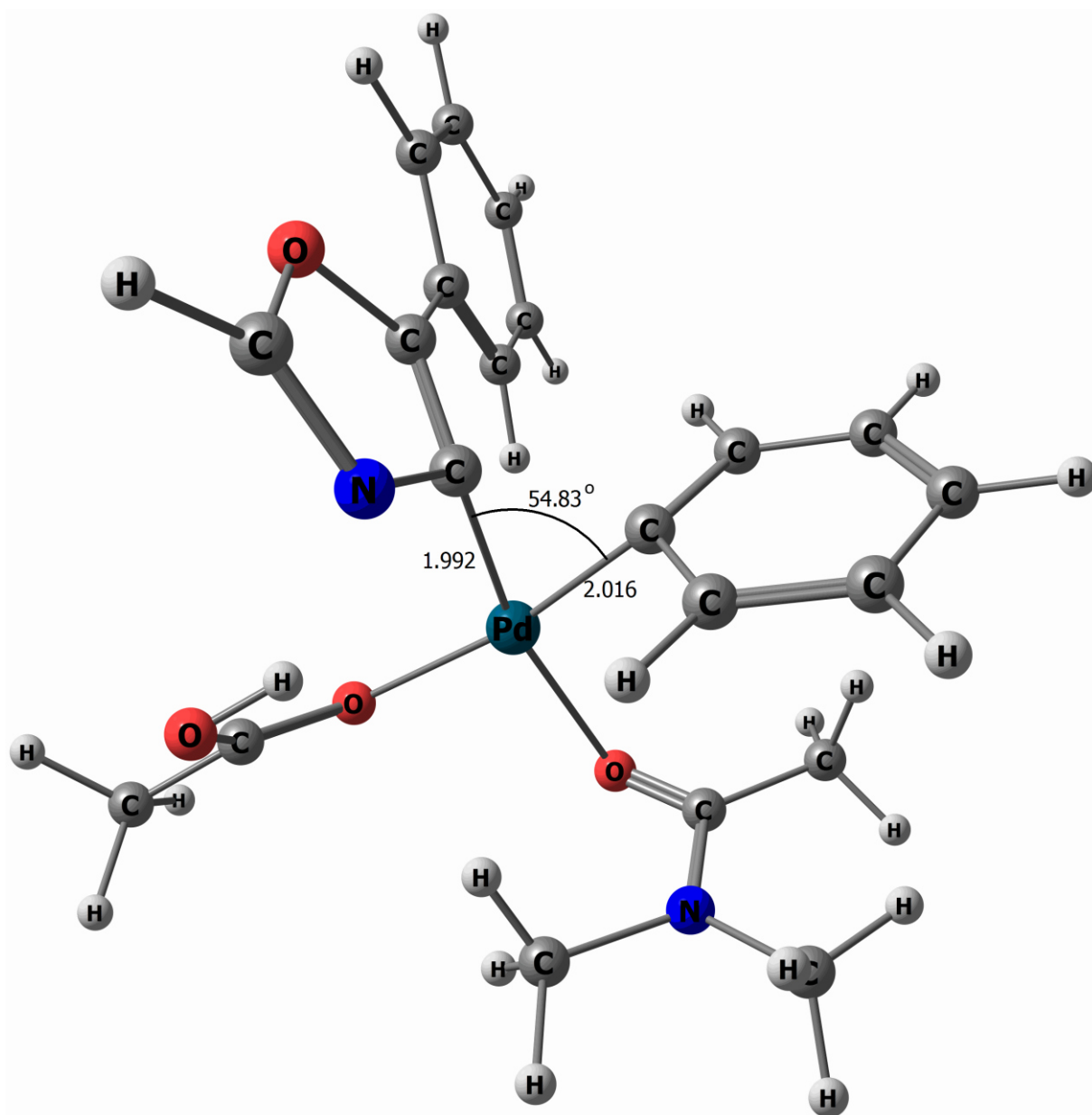


Figure 3-89: Transition State 2 Pathway 5 → 4,5

DSD-PBEP86 Results:

As the accuracy of the DSD-PBEP86/TZVP method has been established for these systems during the study of mono-substitution, we will focus exclusively on the double hybrid results, which themselves incorporate the thermal correction from TPSSh. The less reliable TPSSh energies will not be discussed.

When analysing the energetics of the aforementioned pathways, the reference point, i.e. the zero point of Gibbs energy, was chosen to be the infinity separated oxazole and two Pd(II) catalyst molecules. That is to say, Pd(II) catalyst + 2 x oxazole = 0 kcal.mol⁻¹. This was done so that all the di-substituted pathways could be plotted from the same reference point.

As illustrated in Figure 3-90, which makes use of Pathway N2 and then Pathway 2→2,4 as an example, all of the di-substitution reaction pathways can be considered to arise from the initial reaction of oxazole with two Pd(II) catalyst molecules, ultimately producing a di-substituted product and two Pd(0) catalyst molecules. A consequence of presenting the energetics in this way is that one must consider the Pd(0) catalyst formed along with the monosubstituted product in the 1st arylation, in each subsequent step. This allows for a mass balance. The DSD-PBEP86 reaction Gibbs energies are shown in Table 3-31. The corresponding reaction profiles are shown in Figure 3-91. In this figure, the 1st reaction represents mono-product formation.

Despite the steric hindrance imposed by the now arylated oxazole around the palladium centre, the formation of Intermediate 1 is spontaneous for 3 of the 6 pathways considered. This is especially true for Pathway 4 → 2,4. Here $\Delta G_{\text{Int } 1}$ corresponds to release of 5.76 kcal.mol⁻¹. From observing the structure of the molecule (Figure 3-66), this seems to be a consequence of π - π stacking of the two benzene groups. The average $\Delta G_{\text{Int } 1}$ value across the 6 pathways is -1.19 kcal.mol⁻¹.

Considering the RDS, it can be seen in Table 3-32 that for the majority of pathways the CMD step is rate determining. The exceptions are Pathway 2 → 2,4 and Pathway 5 → 2,5. RE is rate determining by a relatively large margin in both these cases. The extent to which CMD determines the rate of the remaining pathways varies from 2.50 kcal.mol⁻¹ for Pathway 5 → 4,5 to 7.58 kcal.mol⁻¹ for Pathway 2 → 2,5. This information combined with the conclusions regarding the RDS from mono-substitution, leads to the inference that a defining characteristic of these reactions, under ligand free conditions in DMA solvent, is that their rate is predominantly determined by concerted metallation deprotonation, with 7 of the 10 considered pathways from both mono- and di-substitution being CMD limited. However, this is not a blanket statement and RE is energetically relevant and rate determining by a significant margin in certain instances.

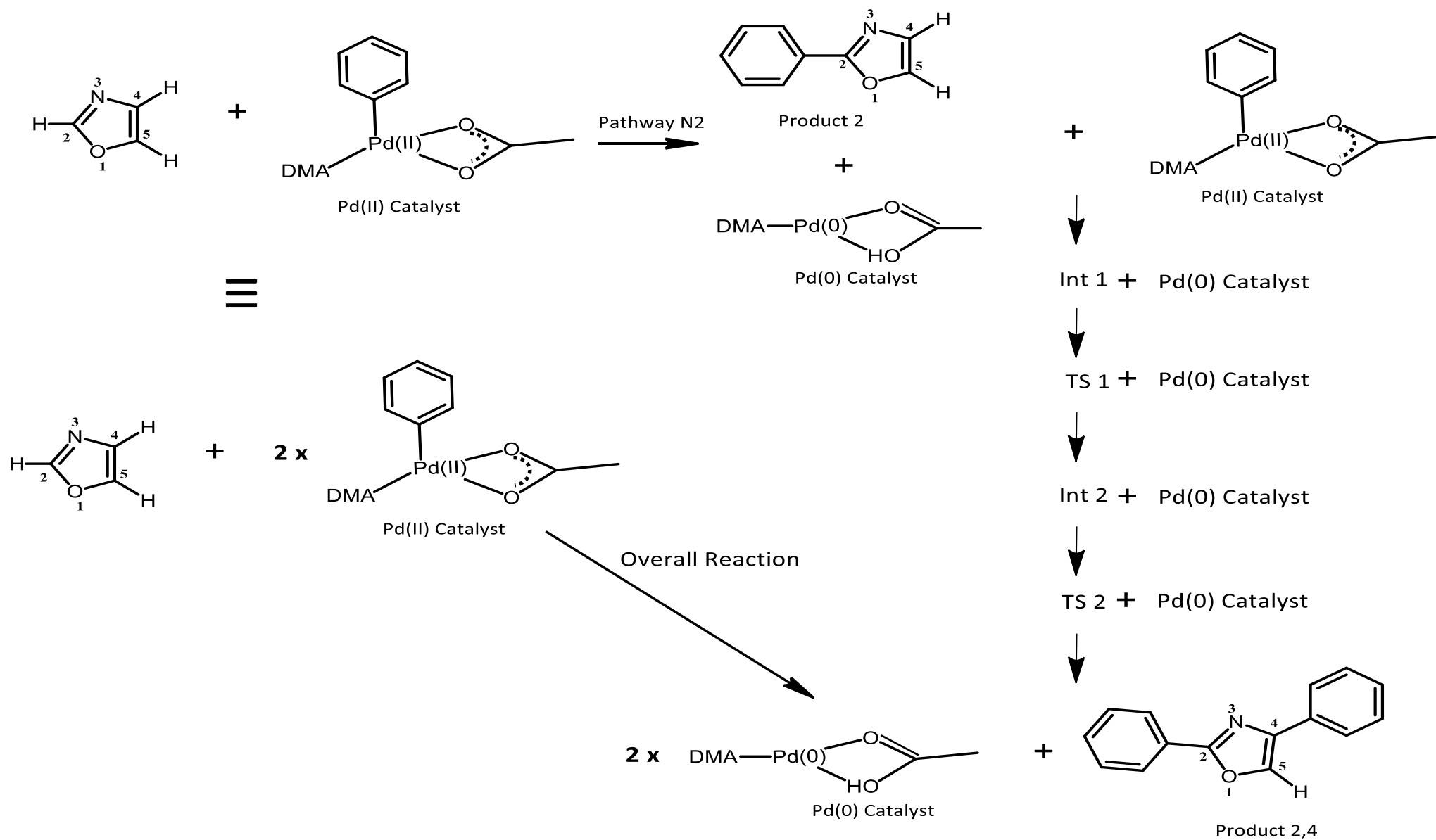


Figure 3-90: Di-substituted Product Formation- Overall Reaction

Table 3-31: Di-substitution Reaction Gibbs Energies (DSD-PBEP86/TZVP)

Product 2,4

Pathway 2 → 2,4	ΔE (kcal.mol ⁻¹)	ΔG_{298K} (kcal.mol ⁻¹)
Product 2 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-21.42	0.57
Int 1 + Pd(0) Catalyst → TS 1 + Pd(0) Catalyst	26.66	21.28
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-26.77	-26.67
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	21.32	24.32
TS 2 + Pd(0) Catalyst → Product 2,4 + 2 x Pd(0) Catalyst	4.44	-15.83
Pathway 4 → 2,4		
Product 4 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-23.95	-5.76
Int 1 + Pd(0) Catalyst → TS 1 + Pd(0) Catalyst	28.12	26.49
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-31.53	-27.58
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	21.32	20.47
TS 2 + Pd(0) Catalyst → Product 2,4 + 2 x Pd(0) Catalyst	8.14	-12.01

Product 2,5

Pathway 2 → 2,5	ΔE (kcal.mol ⁻¹)	ΔG_{298K} (kcal.mol ⁻¹)
Product 2 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-18.94	-0.59
Int 1 + Pd(0) Catalyst → TS 1 + Pd(0) Catalyst	17.59	15.21
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-9.89	-5.46
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	11.75	7.63
TS 2 + Pd(0) Catalyst → Product 2,5 + 2 x Pd(0) Catalyst	2.92	-12.72
Pathway 5 → 2,5		
Product 5 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-19.63	0.19
Int 1 + Pd(0) → TS 1 + Pd(0)	21.91	18.00
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-31.53	-25.42
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	21.54	20.66
TS 2 + Pd(0) Catalyst → Product 2,5 + 2 x Pd(0) Catalyst	8.10	-10.81

Table 3-36 Continued

Product 4,5

Pathway 4 → 4,5	ΔE (kcal.mol ⁻¹)	ΔG_{298K} (kcal.mol ⁻¹)
Product 4 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-18.10	0.89
Int 1 + Pd(0) Catalyst → TS 1 + Pd(0) Catalyst	13.54	14.43
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-13.88	-12.99
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	10.29	8.77
TS 2 + Pd(0) Catalyst → Product 4,5 + 2 x Pd(0) Catalyst	12.35	-7.32
Pathway 5 → 4,5		
Product 5 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-21.63	-2.46
Int 1 + Pd(0) Catalyst → TS 1 + Pd(0) Catalyst	22.00	18.84
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-22.55	-17.24
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	18.63	16.34
TS 2 + Pd(0) Catalyst → Product 4,5 + 2 x Pd(0) Catalyst	8.34	-11.07

Table 3-32: CMD (TS 1) vs Reductive Elimination (TS 2) Transition State Gibbs Energies (DSD-PBEP86/TZVP)

Pathway	$\Delta^\ddagger G_{TS\ 1} - \Delta^\ddagger G_{TS\ 2}$ (kcal.mol ⁻¹)
Pathway 2 → 2,4	-3.04
Pathway 4 → 2,4	6.01
Pathway 2 → 2,5	7.58
Pathway 5 → 2,5	-2.65
Pathway 4 → 4,5	5.67
Pathway 5 → 4,5	2.50

Analysing Gibbs energy required for the RDS, as shown in Table 3-33, reveals Pathway 4 → 4,5 is the kinetic product. This is by a margin of 0.78 kcal.mol⁻¹ when compared to the next lowest pathway, Pathway 2 → 2,5. As stated previously, it was assumed that the absolute error in energy differences is around 1.0 kcal.mol⁻¹ for the system at the DSD-PBEP86/TZVP level of theory. Thus, for all practical intents and purposes the RDS of Pathway 4 → 4,5 and Pathway 2 → 2,5 can be considered to require an equivalent amount of Gibb's free energy. Thus, under conditions of kinetic control and assuming sufficient concentrations of the relevant monsubstituted products, Product 4,5 and Product 2,5 are expected to be the major products.

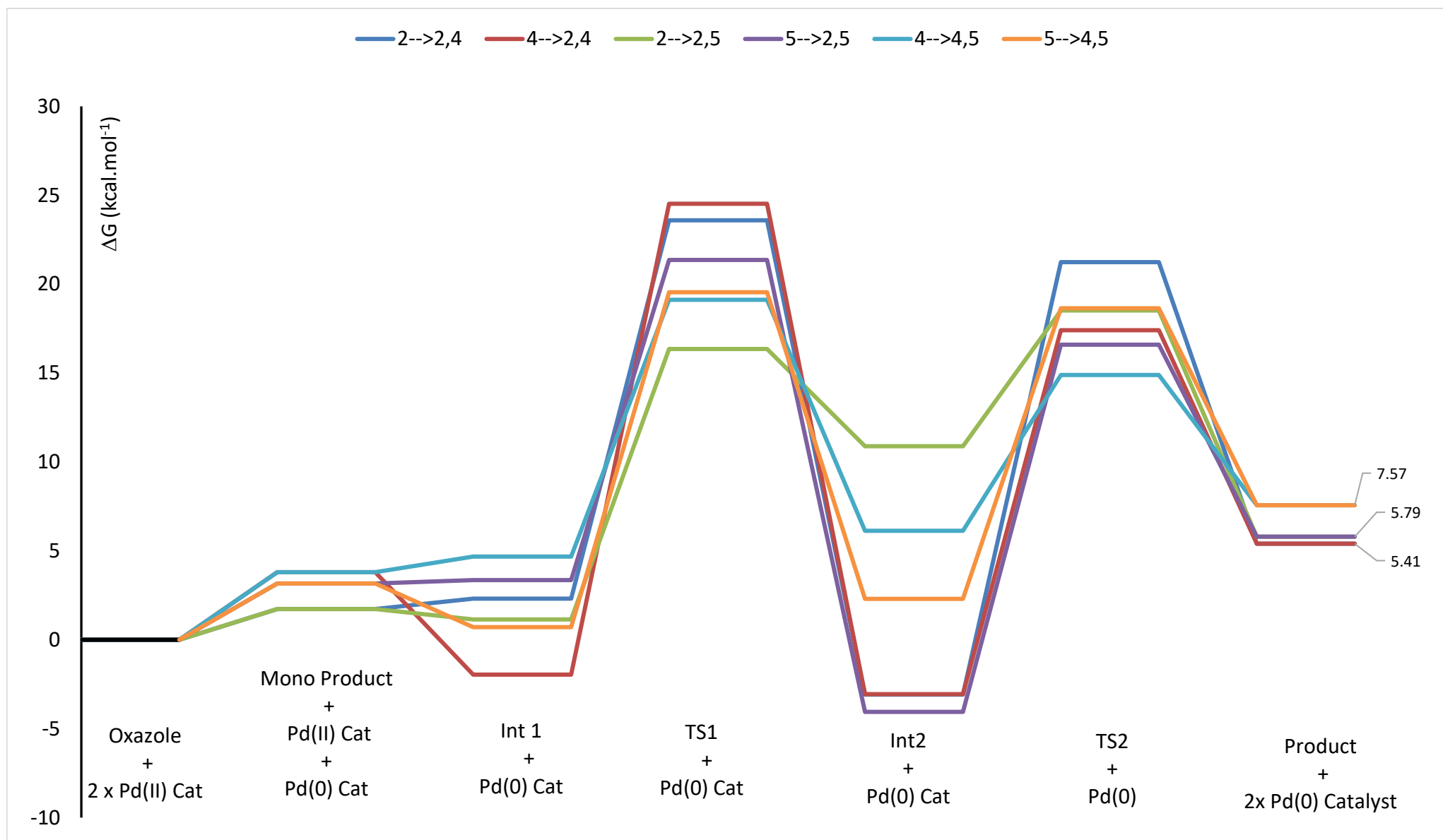


Figure 3-91: Di-substitution Reaction Profiles (DSD-PBEP86/TZVP)

The experimental di-substituted product distributions are shown in Table 3-34. Indeed, the major product is Product 2,5 contributing 28.1% to the total product distribution (this includes both mono and di products, see Entry 1 in Table 3-1). Product 4,5 makes up 3.0%. The two highest energy pathways both correspond to the formation of Product 2,4, which was not observed experimentally. Thus, the experimental results correlate very well with the computational predictions and it is highly probable that the reactions leading to di-substituted products also occur through the CMD mechanism.

Table 3-33 Pathway Favourability in Terms of the RDS - Di-substitution (DSD-PBEP86/TZVP)

Rank	Pathway	RDS	ΔG RDS (kcal.mol ⁻¹)
1 st	Pathway 4 → 4,5	CMD (TS 1)	14.43
2 nd	Pathway 2 → 2,5	CMD (TS 1)	15.21
3 rd	Pathway 5 → 4,5	CMD (TS 1)	18.84
4 th	Pathway 5 → 2,5	RE (TS 2)	20.66
5 th	Pathway 2 → 2,4	RE (TS 2)	24.32
6 th	Pathway 4 → 2,4	CMD (TS 1)	26.49

Table 3-34: Experimental Di-substituted Product Distributions - Phosphine Free Conditions

Catalyst	Ligand	Di-substituted Product Distributions (%)	
Pd(OAc) ₂	None	Product 4,5	Product 2,5
		3.0	28.1

From a consideration of the kinetic products one would expect close to equal amounts of Product 4,5 and Product 2,5 to be observed experimentally. However, this is assuming sufficient concentrations of the relevant monsubstituted products, namely Product 4 and 2, respectively. Although, small amounts of Product 2 were indeed observed experimentally, and this was attributed to its thermodynamic stability, the same is not true for Product 4. Thus, it is possible that although Pathway 4 → 4,5 is accessible kinetically it is severely limited in terms of the concentration of Product 4 available to react with oxazole. The rate of a chemical reaction is dependent on the concentration of reactants, Equation **Error! Reference source not found.** . Therefore, even if the rate constant (k_r) is large, the reaction will not occur to a measurable degree if insufficient concentrations of reactants A or B are available.

$$Rate = k_r[A][B] \quad (3.7)$$

Interpreting the results of di-substitution in a manner consistent with the conclusions from mono-substitution implies that the monosubstituted reactions are governed to a significant degree by the thermodynamics of the system, in order for sufficient amounts of Product 2 (the thermodynamic product) to be available to enter Pathway 2 $\rightarrow$ 2,5 and eventually form the observed 28.1% of Product 2,5. The 0.3% of Product 2 observed experimentally must be what remains after further di-substitution reactions via Pathway 2 $\rightarrow$ 2,5 have occurred to a significant extent.

When comparing mono and di-substitution pathways, it should be noted that these reactions were run with a 1:1 mole ratio of oxazole and 4-bromotoluene. Thus, after a period of initial mono-substitution, significantly less 4-bromotoluene would be left to react via the pathways leading to di-substitution, than was present initially. Therefore, the experimental product distributions are not expected to completely represent the energetics of the various pathways, as di-substitution is concentration limited in terms of 4-bromotoluene.

That being said, the energy barrier of the RDS for Pathway 4 $\rightarrow$ 4,5 is 14.43 kcal.mol⁻¹, 3.87 kcal.mol⁻¹ lower than that of the lowest mono-substituted pathway, namely Pathway =5. Therefore, if these reactions were run under conditions of excess 4-bromotoluene, the relative percentage of di-substituted products, particularly Product 2,5 and Product 4,5 is expected to increase dramatically.

As can be seen in Figure 3-91, in terms of the total Δ_rG values all reactions are endergonic. However, as mentioned previously the bare Pd(II) catalyst that is being considered (Figure 3-37) is under ligated and in reality, will be stabilized by the coordination of solvent molecules to palladium which is likely to change this qualitatively. The thermodynamic product is Product 2,4 (5.41 kcal.mol⁻¹) followed closely by Product 2,5 (5.79 kcal.mol⁻¹) and finally Product 4,5 (7.57 kcal.mol⁻¹). The fact that no Product 2,4 was observed experimentally indicates that an equilibrium among the di-substituted pathways has not been established. Therefore, there must have been insufficient time for an equilibrium to be attained among the di-substituted pathways. Although it is possible to calculate the time required for a system to reach equilibrium, if the rate of the reactions and the initial concentration of reactants are known (see reference 19), this is beyond the scope of this work.

In summary, the following conclusions can be drawn:

- The rate determining step for 4 of the 6 di-substitution pathways considered is concerted metallation deprotonation, by the margins presented in Table 3-32.
- Only for Pathway 2 $\rightarrow$ 2,4 and Pathway 5 $\rightarrow$ 2,5 is reductive elimination rate determining, by 3.04 and 2.65 kcal.mol⁻¹ respectively.
- Pathway 4 $\rightarrow$ 4,5 requires the lowest total Gibbs energy for the RDS, by of 0.78 kcal.mol⁻¹, and thus Product 4,5 is the predicted kinetic di-substituted product. Although, it is only observed experimentally in low concentrations and represents 3.0% of the total product distribution, this can be readily explained by considering the lack of Product 4 available to react further.
- Pathway 5 $\rightarrow$ 2,5 requires the second lowest Gibbs energy for the RDS, the magnitude of which is practically identical to Pathway 4 $\rightarrow$ 4,5. Thus, Product 2,5 is also predicted to form under conditions of kinetic control to a significant extent. This is in agreement with experimental results that found Product 2,5 makes up 90.35% of the di-substituted product distribution and 28.1% of the total product distribution.
- It is thus highly probable that under phosphine free conditions, both mono- and di-substitution reactions proceed via a CMD mechanism.
- The above implies that Product 2 must be produced in sufficient concentrations to react further. The formation of Product 2 is kinetically but not thermodynamically hindered, thus an equilibrium among the mono-substitution pathways must be present.
- Equilibrium was not established among the pathways leading to di-substituted products.

These results imply that if these reactions are run at lower temperatures and for shorter periods i.e. under conditions of further kinetic control then the mono-substituted kinetic product, Product 5, would be formed in even greater concentrations. As an equilibrium among the mono-substitution pathways is necessary in order to produce the reactants that lead to the formation of Products 2,5 and 4,5, higher temperatures and longer reaction times, i.e. conditions of thermodynamic control, would increase the concentration of these products. However, conditions of excessive thermodynamic control will lead to the establishment of an equilibrium among the di-substitution reaction pathways which is expected to result in the formation of Product 2,4. Performing the reactions under conditions of excess 4-bromotoluene would also stimulate a higher percentage of di-substituted products.

References

1. J. Fournier Dit Chabert, L. Joucla, E. David and M. Lemaire, *Tetrahedron*, 2004, **60**, 3221-3230.
2. M. Parisien, D. Valette and K. Fagnou, *The Journal of Organic Chemistry*, 2005, **70**, 7578-7584.
3. X. Wang, D. V. Gribkov and D. Sames, *The Journal of Organic Chemistry*, 2007, **72**, 1476-1479.
4. B. Glover, K. A. Harvey, B. Liu, M. J. Sharp and M. F. Tymoschenko, *Organic Letters*, 2003, **5**, 301-304.
5. Y. Tan and J. F. Hartwig, *Journal of the American Chemical Society*, 2011, **133**, 3308-3311.
6. S. I. Gorelsky, D. Lapointe and K. Fagnou, *Journal of the American Chemical Society*, 2008, **130**, 10848-10849.
7. L. Xue and Z. Lin, *Chemical Society Reviews*, 2010, **39**, 1692-1705.
8. T. Sperger, I. A. Sanhueza, I. Kalvet and F. Schoenebeck, *Chemical Reviews*, 2015, **115**, 9532-9586.
9. F. Allen, *Acta Crystallographica Section B*, 2002, **58**, 380-388.
10. E. Binamira-Soriaga, M. Lundeen and K. Seff, *Journal of Crystal and Molecular Structure*, **9**, 67-75.
11. G. Xu, Q. Luo, S. Eibauer, A. F. Rausch, S. Stempfhuber, M. Zabel, H. Yersin and O. Reiser, *Dalton Transactions*, 2011, **40**, 8800-8806.
12. Q.-L. Luo, J.-P. Tan, Z.-F. Li, Y. Qin, L. Ma and D.-R. Xiao, *Dalton Transactions*, 2011, **40**, 3601-3609.
13. W. E. Marsh, D. S. Eggleston, W. E. Hatfield and D. J. Hodgson, *Inorganica Chimica Acta*, 1983, **70**, 137-142.
14. M. Gómez, G. Muller and M. Rocamora, *Coordination Chemistry Reviews*, 1999, **193–195**, 769-835.
15. P. Anstaett and F. Schoenebeck, *Chemistry – A European Journal*, 2011, **17**, 12340-12346.
16. S. I. Gorelsky, *Coordination Chemistry Reviews*, 2013, **257**, 153-164.
17. Y. Gumrukcu, B. de Bruin and J. Reek, *Catalysts*, 2015, **5**, 349.
18. G. A. Andrienko., <http://www.chemcraftprog.com>.
19. J. P. Marten J. ten Hoor, *Chemistry Education in New Zealand*, 2009, 7-8.

4. The CMD

Mechanism: Tri-tert-butylphosphine ligand

Contents

4.1 Mono-substitution	213
The Active catalyst:	213
The Reaction Pathways:	214
Pathway N2:	215
Pathway N4:	220
Pathway =4:.....	224
Pathway =5:.....	226
DSD-PBEP86 Results:.....	230
4.2 Di-substitution	236
The Reaction Pathways:	236
Pathway 2 $\rightarrow$ 2,4:	236
Pathway 4 $\rightarrow$ 2,4:	240
Pathway 2 $\rightarrow$ 2,5:	244
Pathway 5 $\rightarrow$ 2,5:	248
Pathway 4 $\rightarrow$ 4,5:	252
Pathway 5 $\rightarrow$ 4,5:	257
DSD-PBEP86 Results:	261
References	269

4.1 Mono-substitution

The Active catalyst:

The modelling of the CMD mechanism ,under conditions of one mole equivalent of P^tBu_3 ligand, was investigated starting from the active catalytic species reported by Hartwig et al, ¹ which itself was based on a crystal structure of the isolated catalyst (Figure 4-1).

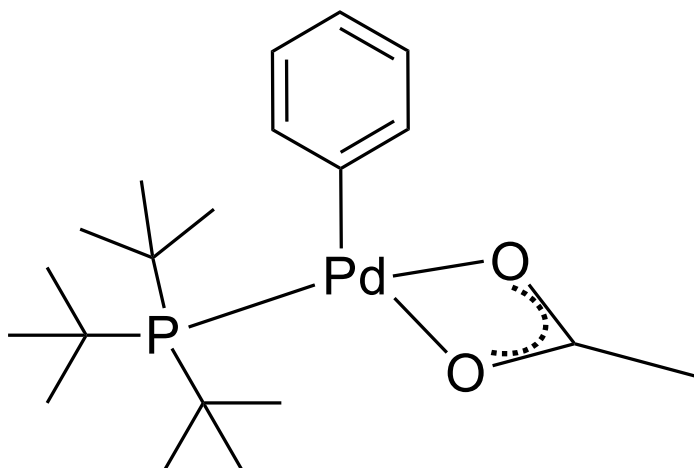


Figure 4-1: Active Catalyst – 1 Mole Equivalent of P^tBu_3 (reported by Hartwig et al.¹)

As an initial estimate of this structure, the B3LYP/TZVP optimized geometry from the same reference¹ was employed. However, this structure contained pivalic rather than acetic acid. The three methyl groups of pivalic acid were replaced by hydrogen atoms and this structure optimized using TPSSH/SVP to generate the model of the active catalyst used in this work, show in Figure 4-2. The full P^tBu_3 ligand was modelled and no methyl groups were replaced with hydrogen atoms. This was done as the size and bulk of this ligand is expected to influence the energetics of the reactions. A substantial speed-up in calculation time could be incurred by replacing P^tBu_3 with PMe_3 , as is often done in literature.² However, this increase in computational efficiency was not deemed worth the resultant loss in accuracy.

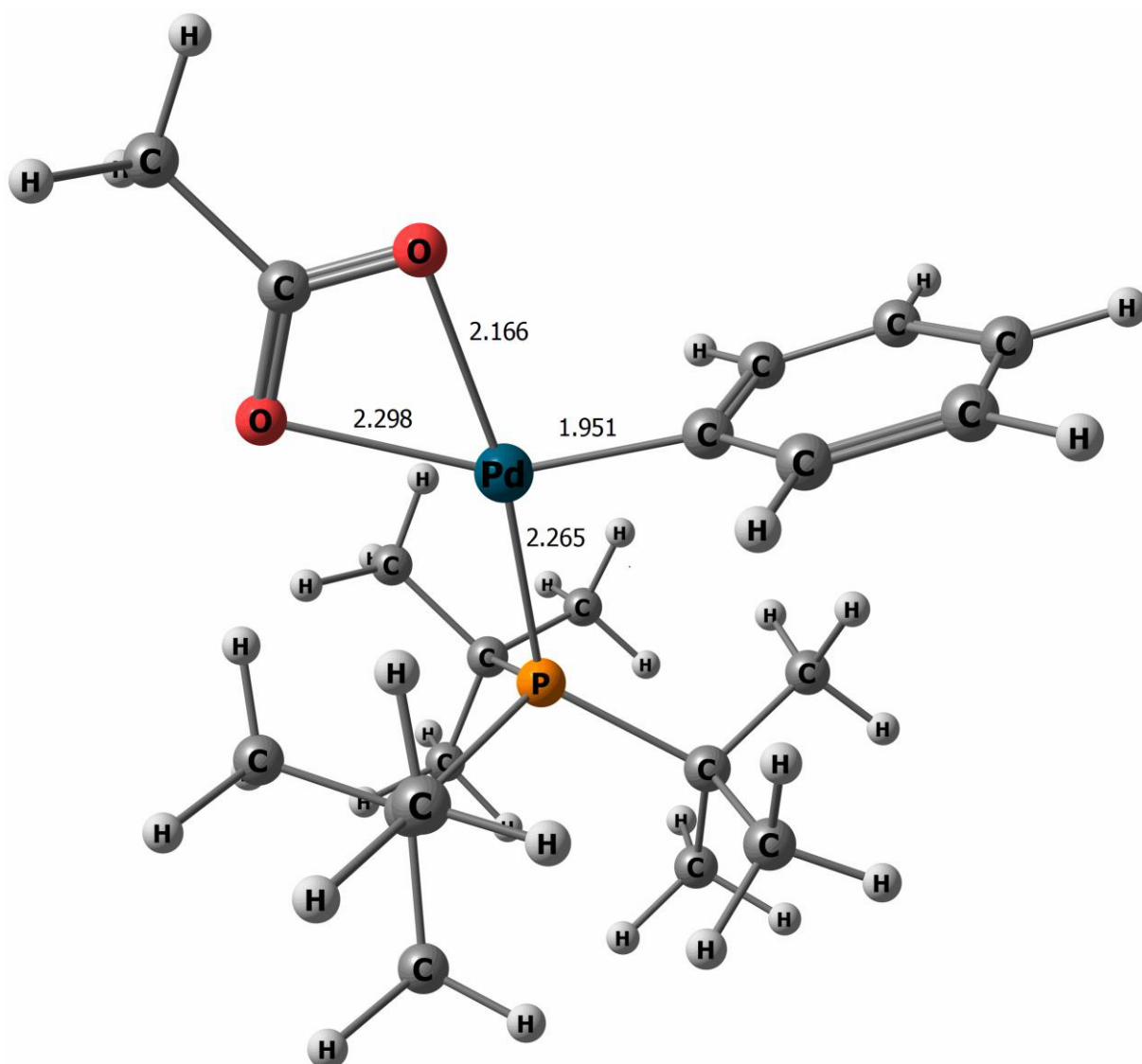


Figure 4-2: The Active Catalyst - 1 Mole Equivalent of P^tBu₃ Ligand

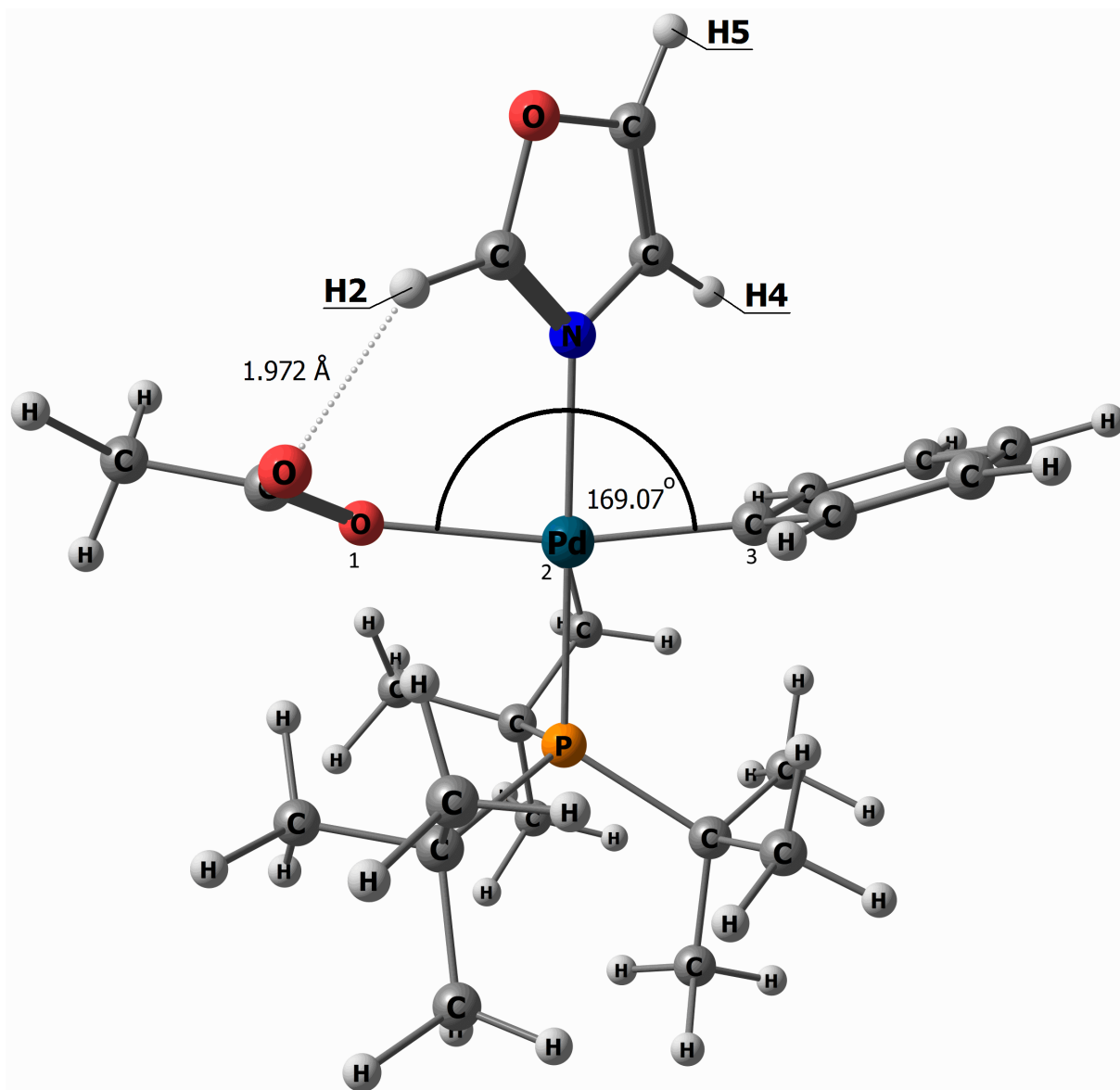
The Reaction Pathways:

Exactly the same mono-substitution pathways were modelled as for ligand free conditions (Chapter 3), namely Pathway N2, Pathway N4, Pathway =4 and finally Pathway =5. It was assumed that the coordination of P^tBu₃ to palladium, rather than DMA, will not dramatically change the overall conformation of the catalyst. Therefore, the Intermediate 1 and Intermediate 2 geometries for all pathways in this chapter were constructed by simply replacing DMA with P^tBu₃ and re-optimising. Alternative conformations of all Intermediate 1 geometries were optimized and can be found in the ESI. Where necessary, different conformations of other molecules were also optimised and compared. This was done in order to ensure that the molecules considered were as close as possible to global minima on the potential energy surface.

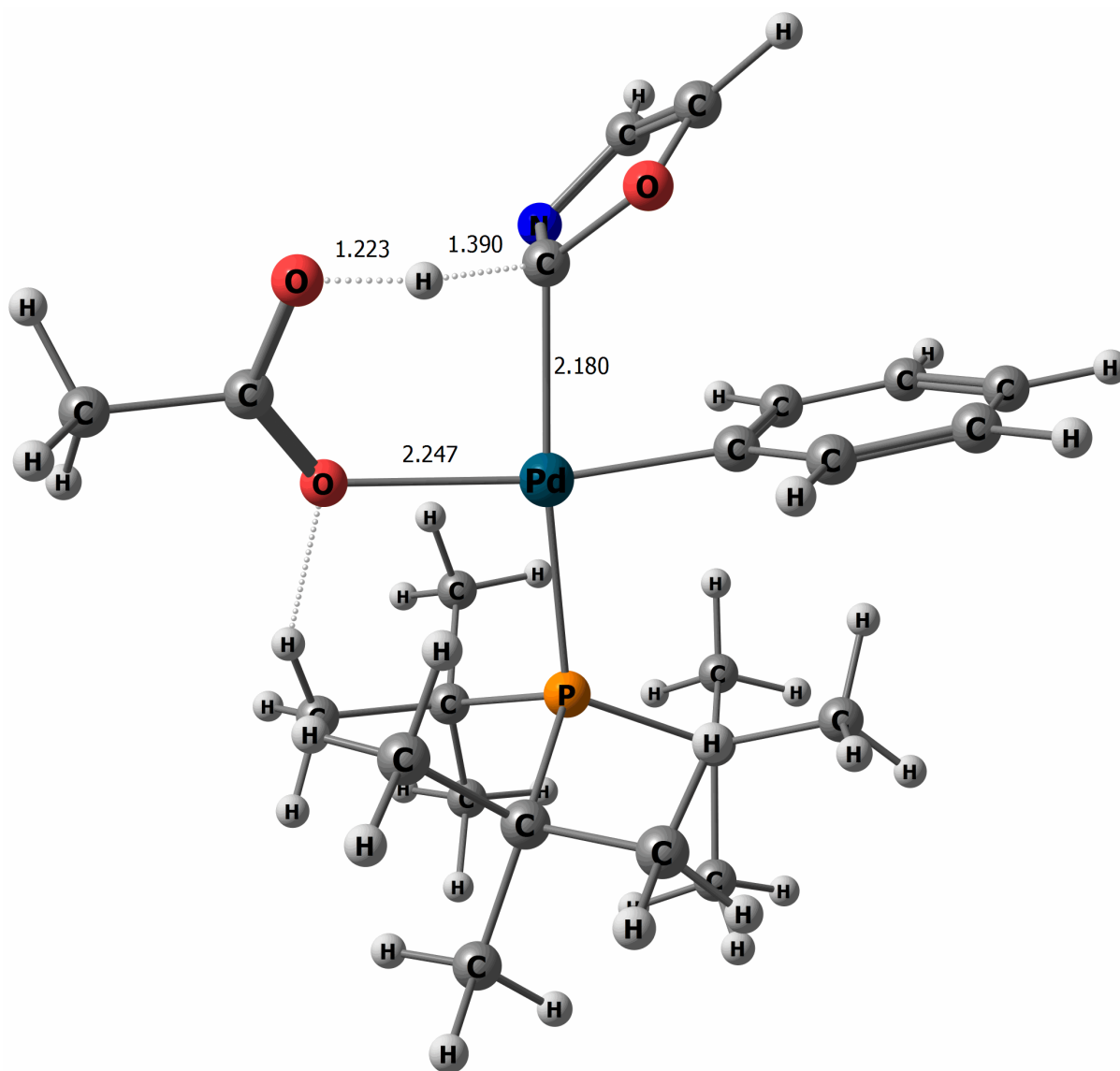
Transition State 1 (CMD) and Transition State 2 (RE) were however not simply obtained from modification of their counterparts in the DMA ligated pathways, but through the use of NEB calculations and surface scans (Section 2.2.7). This was done in order to avoid results biased towards the CMD mechanism, as it was possible that the products form via a different pathway under these conditions. However, as will be shown, the CMD mechanism prevails.

Pathway N2:

Int 1 for Pathway N2 is pictured in Figure 4-3 and it can be seen that the bulk of the P^tBu_3 ligand forces the acetate and benzene ligands to be positioned closer to oxazole. The O(acetate)–Pd–C(benzene) angle, $\angle(1-2-3)$, is 169.07° and is depicted by the black arc in Figure 4-3. The corresponding angle in the DMA coordinated analogue was 176.36° . The acetate oxygen and the hydrogen positioned to be deprotonated, H2, are separated by only $1.972 \text{ \AA}$, the corresponding distance in the DMA analogue is $2.028 \text{ \AA}$. The CMD transition state (TS 1) for this pathway is depicted in Figure 4-4, and has an imaginary frequency of -1130.69 cm^{-1} . Int 2 is shown in Figure 4-5 and is a zwitterionic species, the nitrogen atom of oxazole having deprotonated the acetic acid ligand. This was also the case in the DMA coordinated Pathway N2 from Chapter 3, highlighting the similarity between analogous structures in the same pathway but with different ligands. TS 2 is presented in Figure 4-6 and has an imaginary frequency of -356.44 cm^{-1} . The zwitterionic nature persists in this structure. Just as in Chapter 3, for all pathways considered, the product and Pd(0) catalyst were manually separated into separate .xyz files and optimised individually. This process resulted in the Pd(0) catalyst shown in Figure 4-7. All products are identical to those presented in Chapter 3. The complexed products, i.e. the final products still complexed to the Pd(0) catalyst, can be found in the ESI for this chapter.



[Figure 4-3: Intermediate 1 N2](#)



[Figure 4-4: Transition State 1 N2](#)

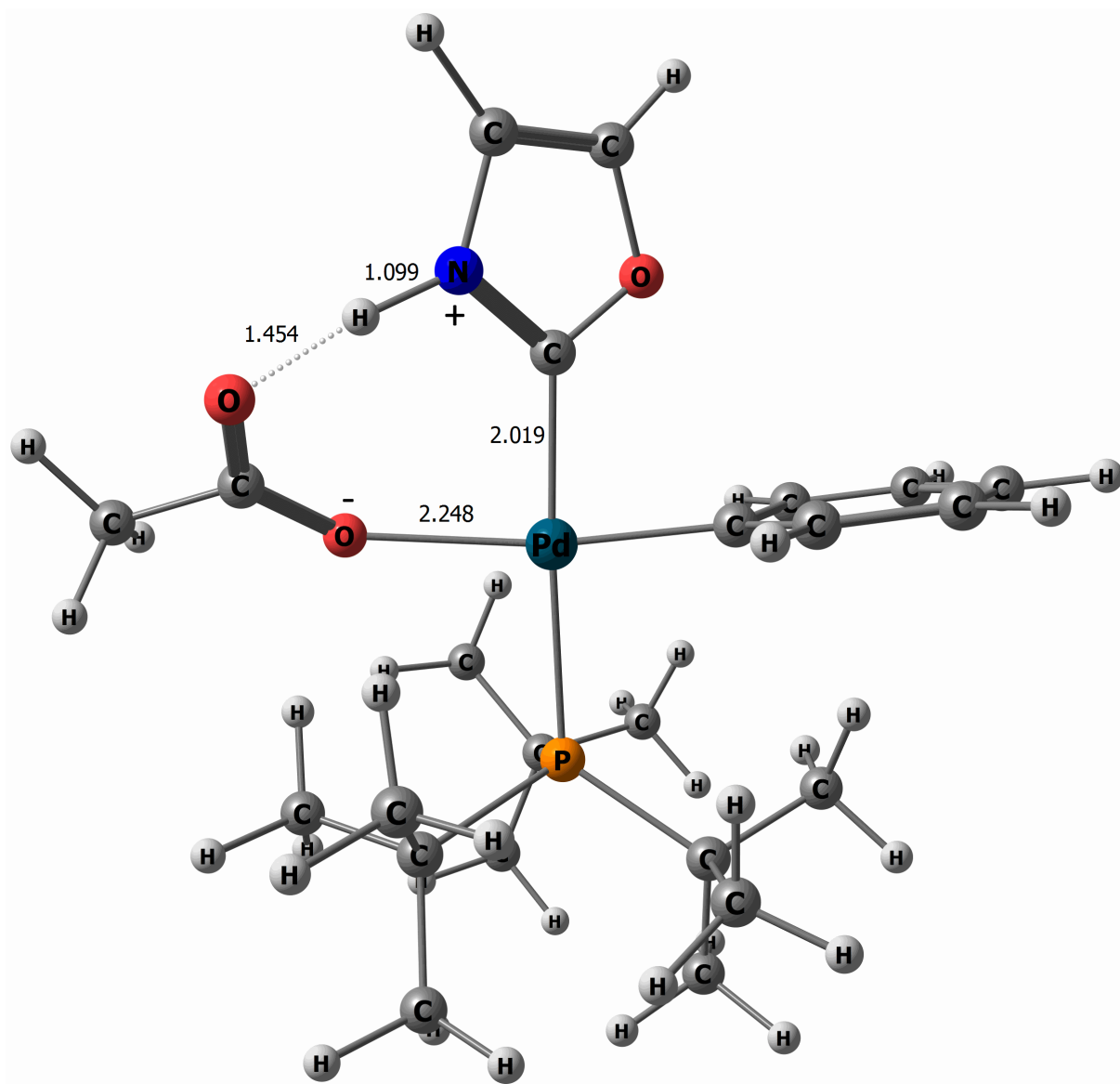


Figure 4-5: Intermediate 2 N2

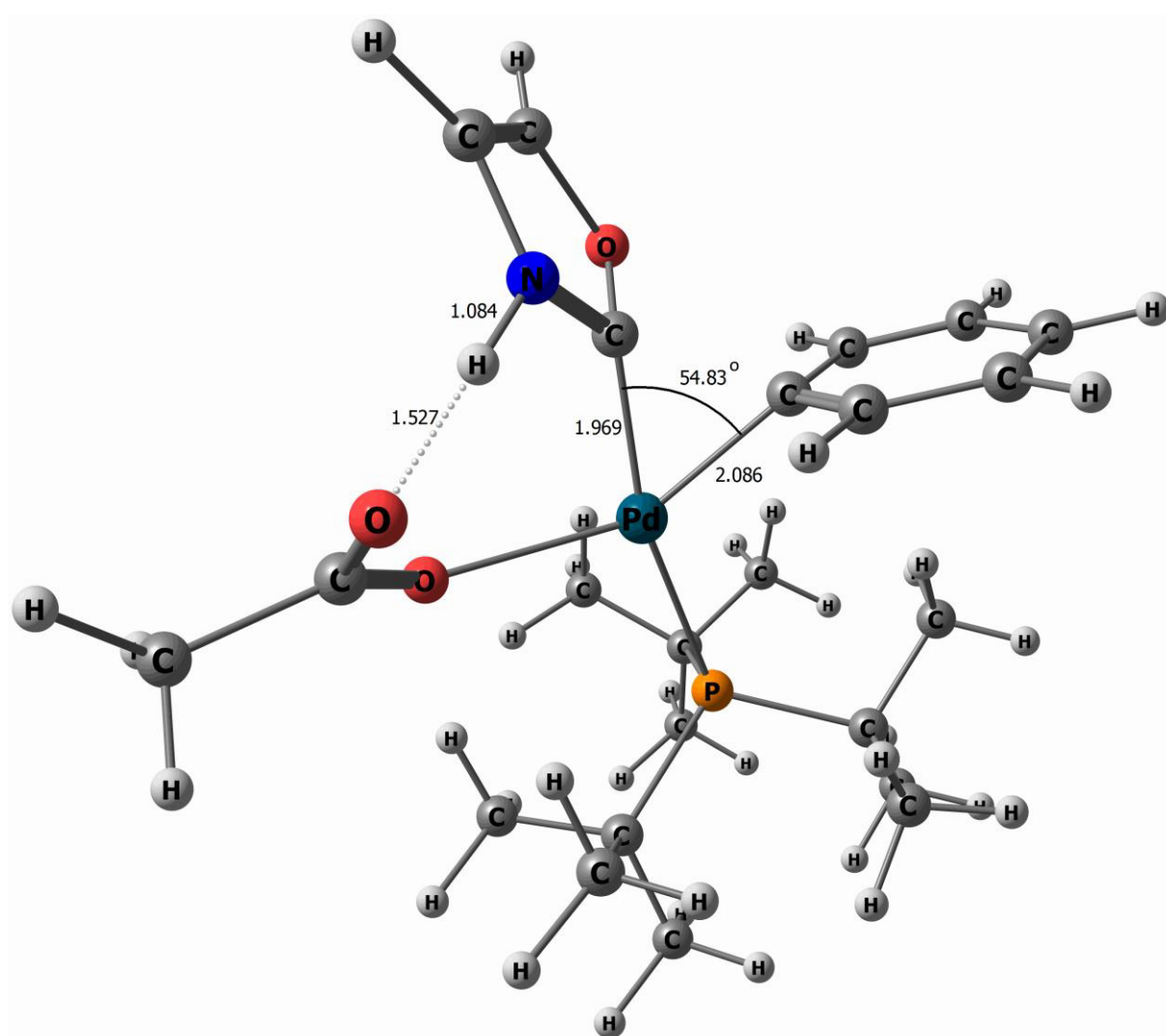


Figure 4-6: Transition State 2 N2

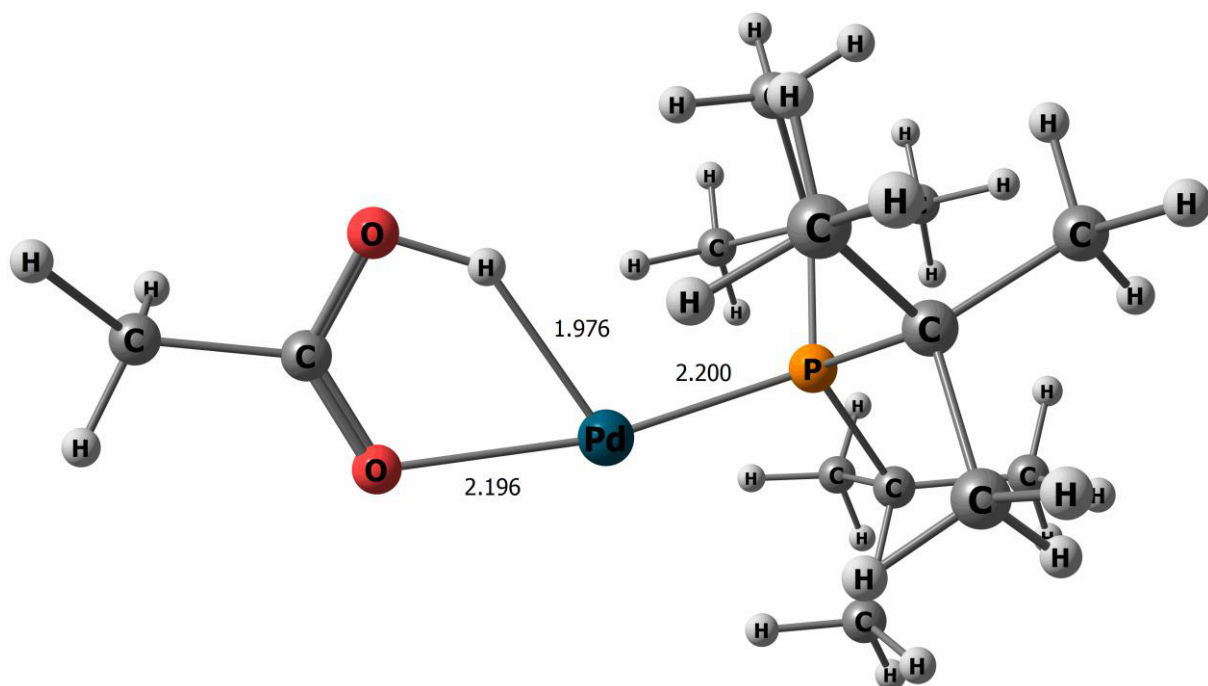
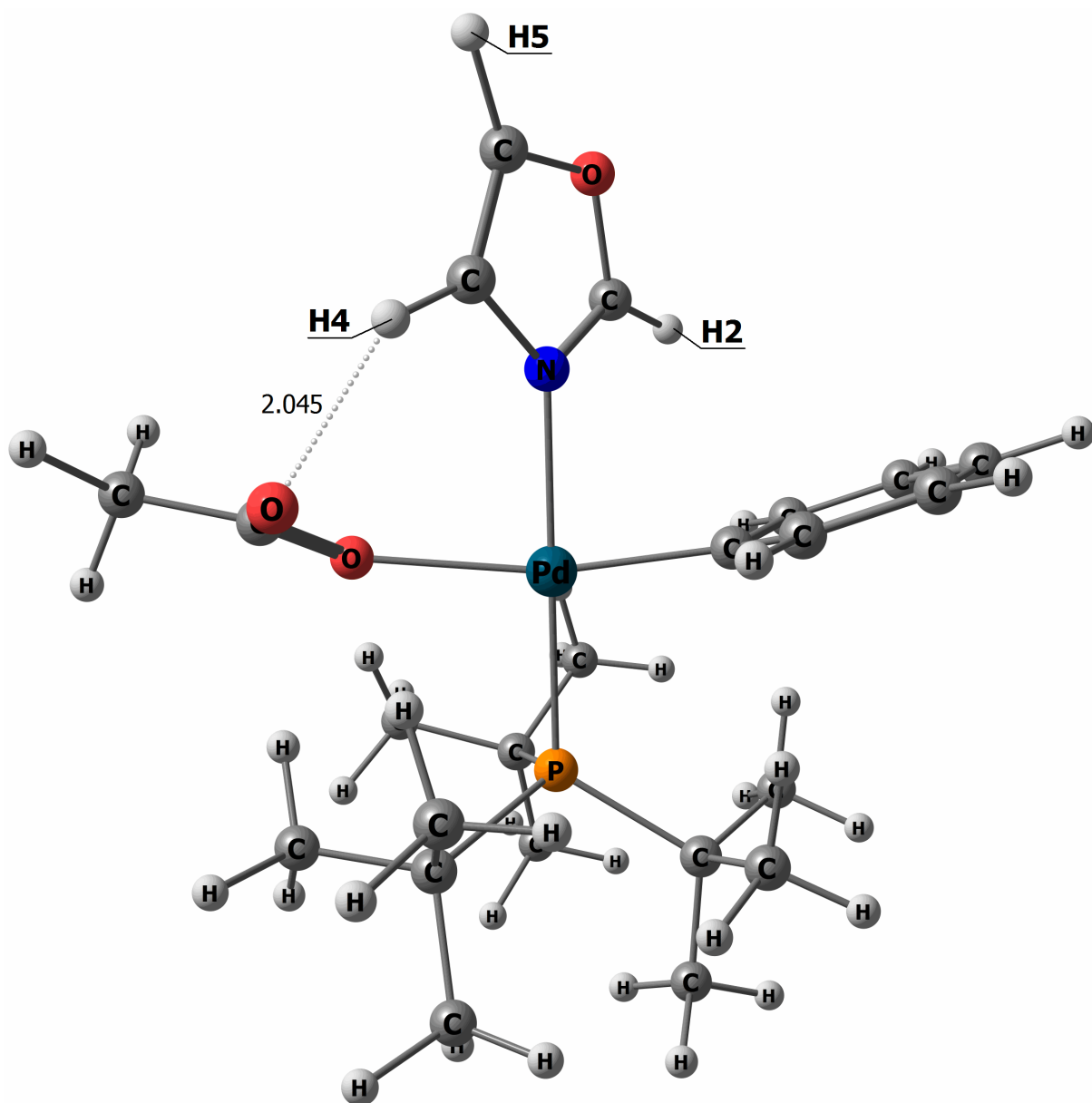


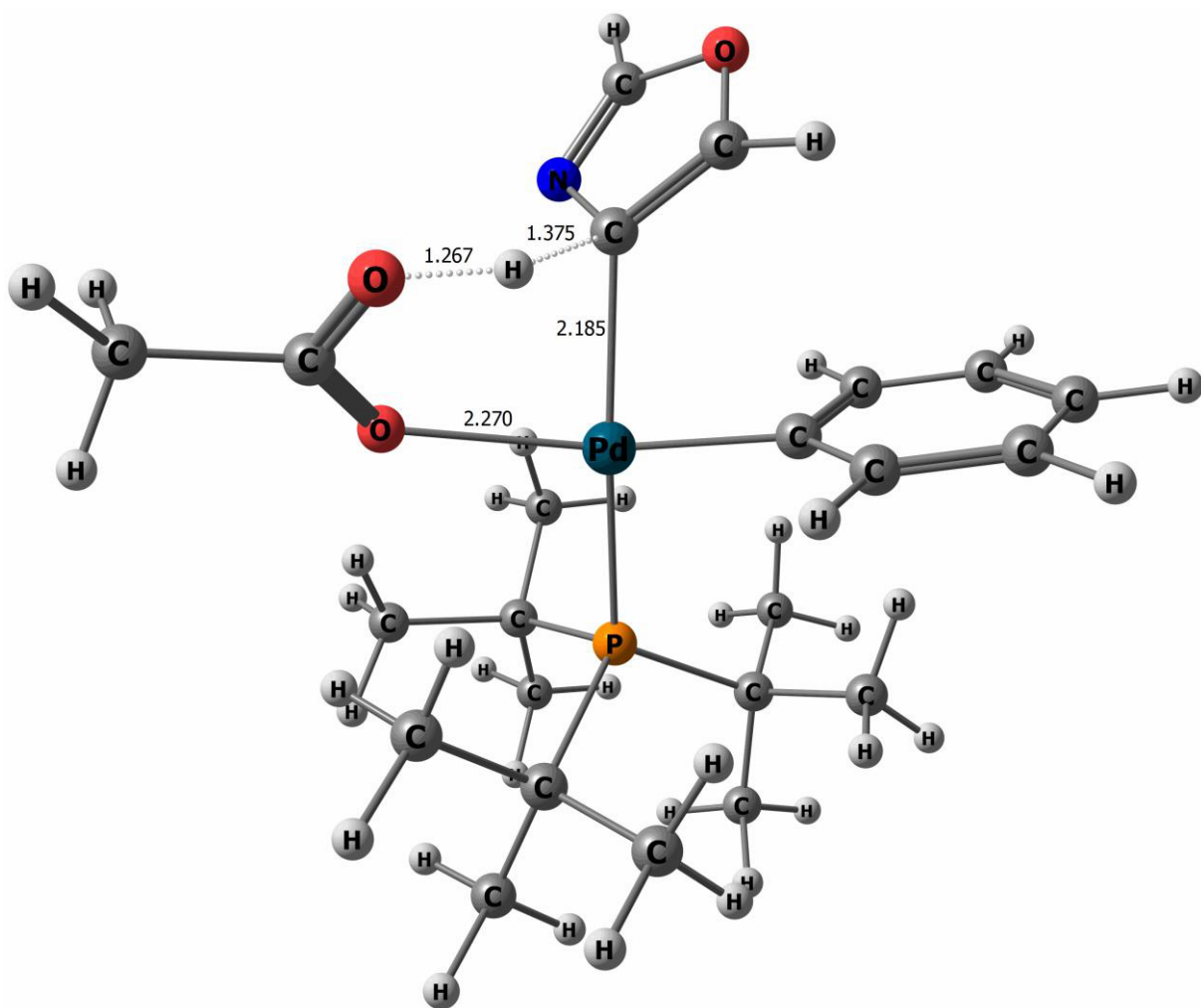
Figure 4-7: Pd(0) Catalyst

Pathway N4:

Int 1 for Pathway N4 is illustrated in Figure 4-8 and TS 1 in Figure 4-9. TS 1 has a negative frequency of -1256.14 cm^{-1} . Int 2 is presented in Figure 4-10, and is a zwitterion. TS 2 is shown in Figure 4-11 and has a negative frequency of -383.09 cm^{-1} .



[Figure 4-8: Intermediate 1 N4](#)



[Figure 4-9: Transition State 1 N4](#)

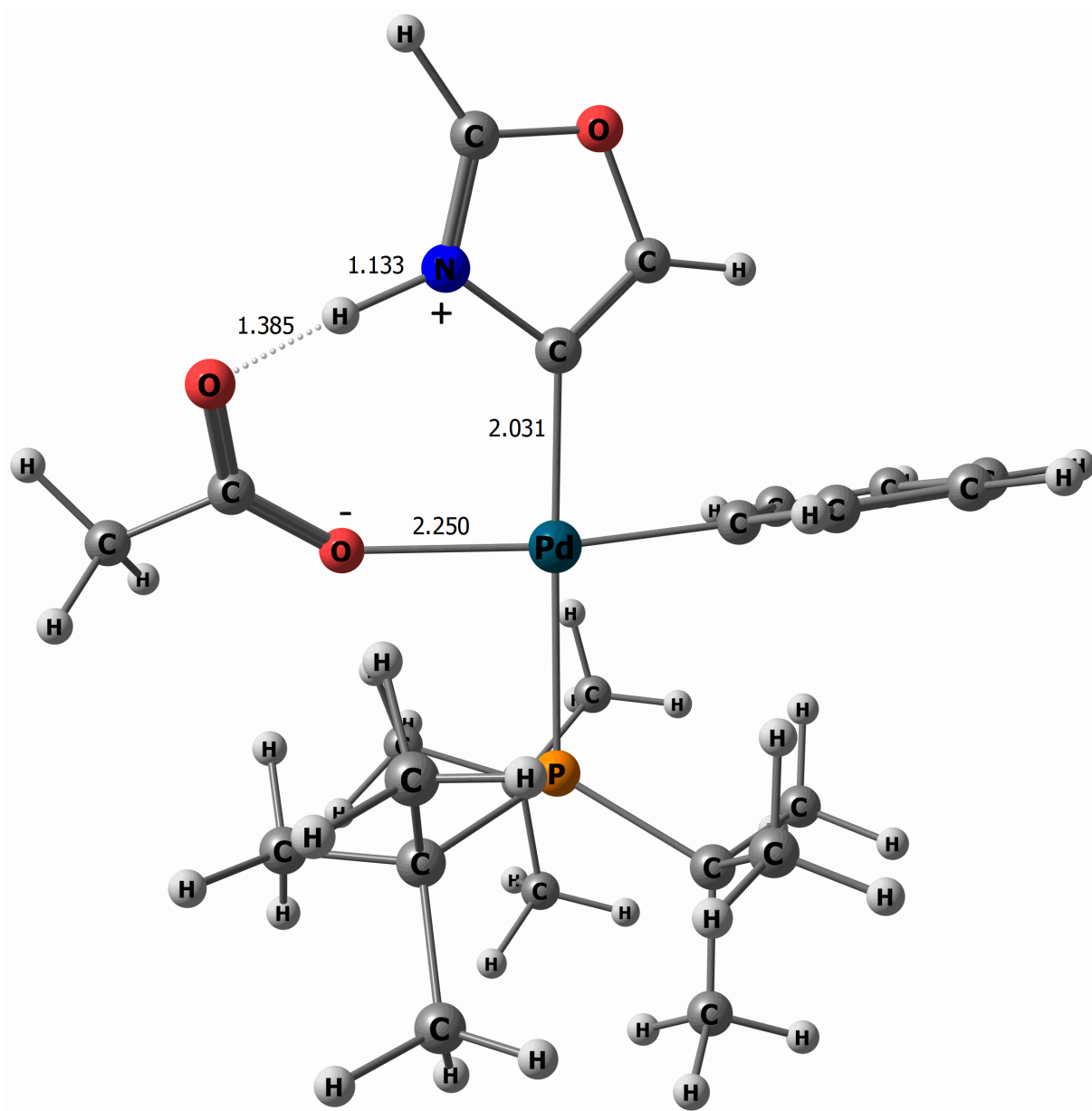


Figure 4-10: Intermediate 2 N4

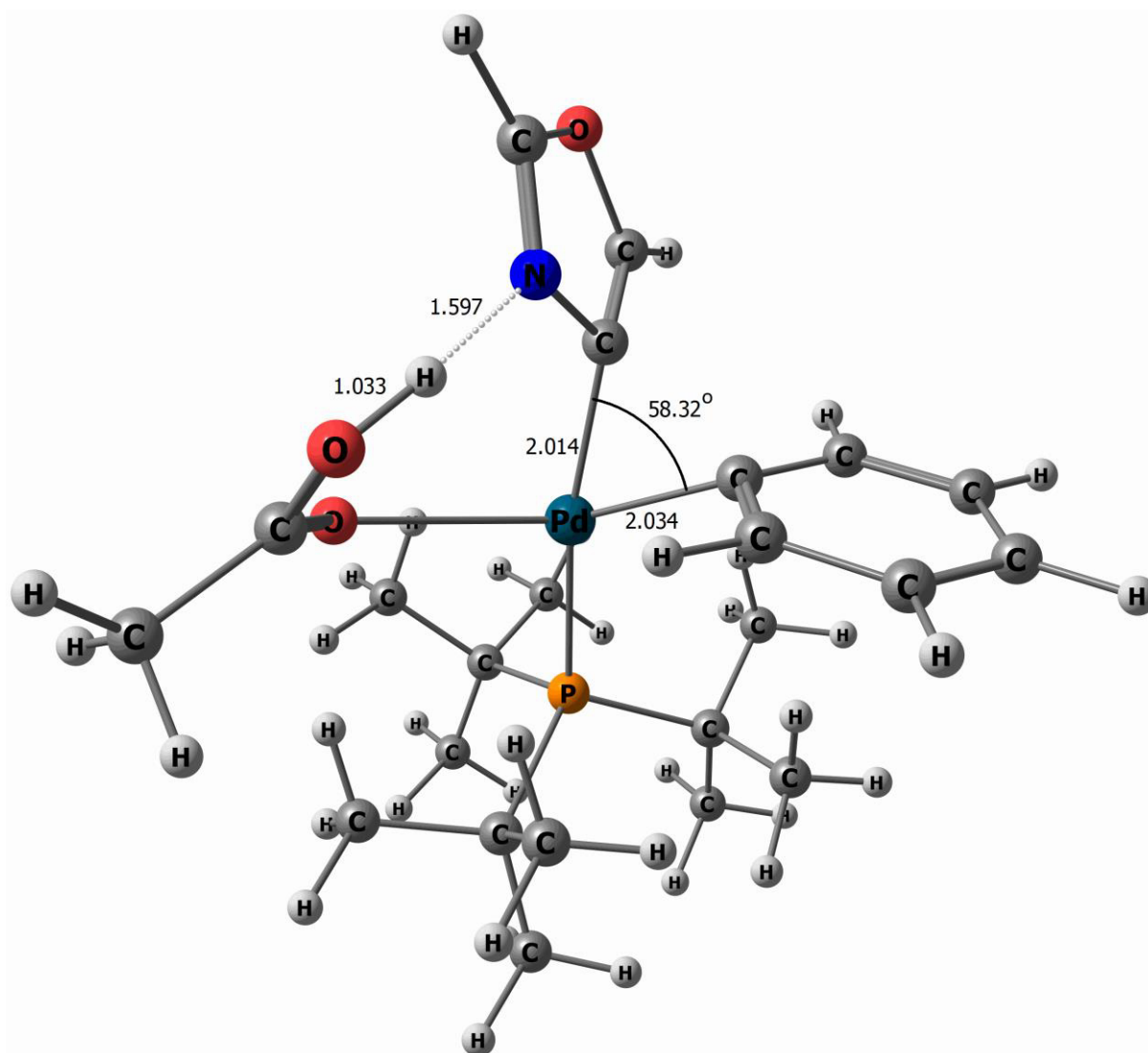
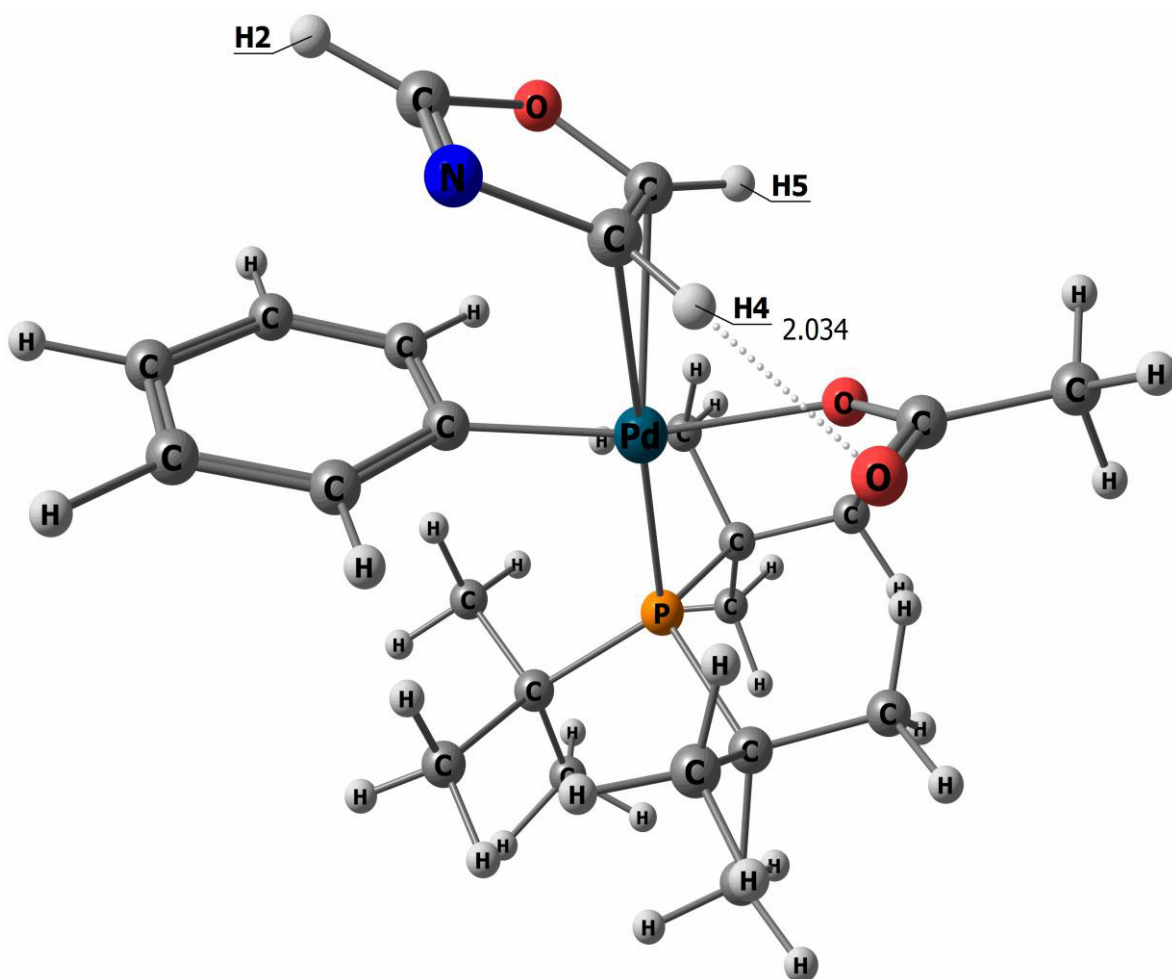


Figure 4-11: Transition State 2 N4

Pathway =4:

Int 1 for Pathway =4 is presented in Figure 4-12 and TS 1 in Figure 4-13. TS 1 has an imaginary frequency of -1267.81 cm^{-1} . Just as in Chapter 3, at this point Pathways N4 and =4 actually merge and Pathway =4 shares the same Int 2 and TS 2 structure as Pathway N4.



[Figure 4-12: Intermediate 1 = 4](#)

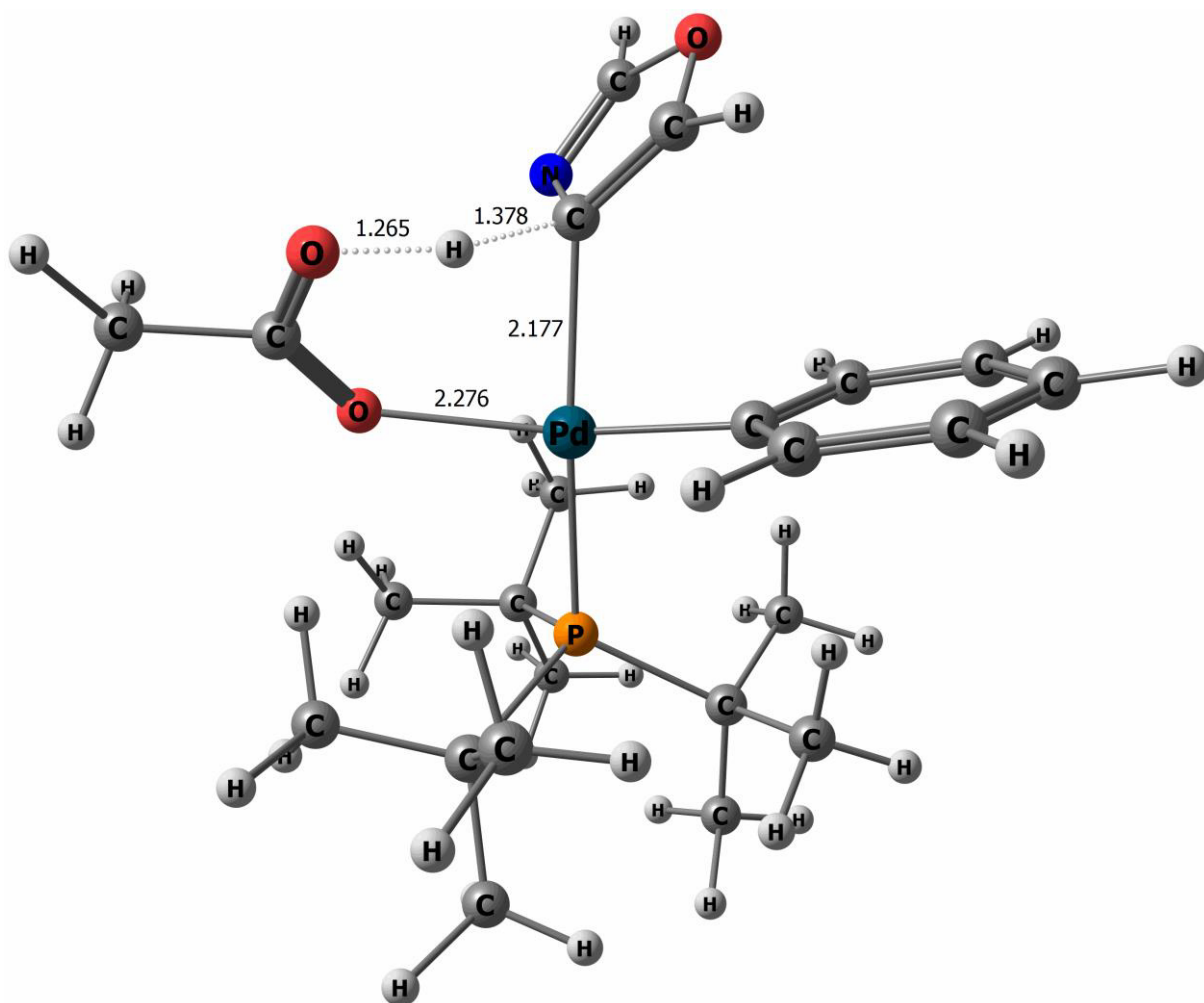
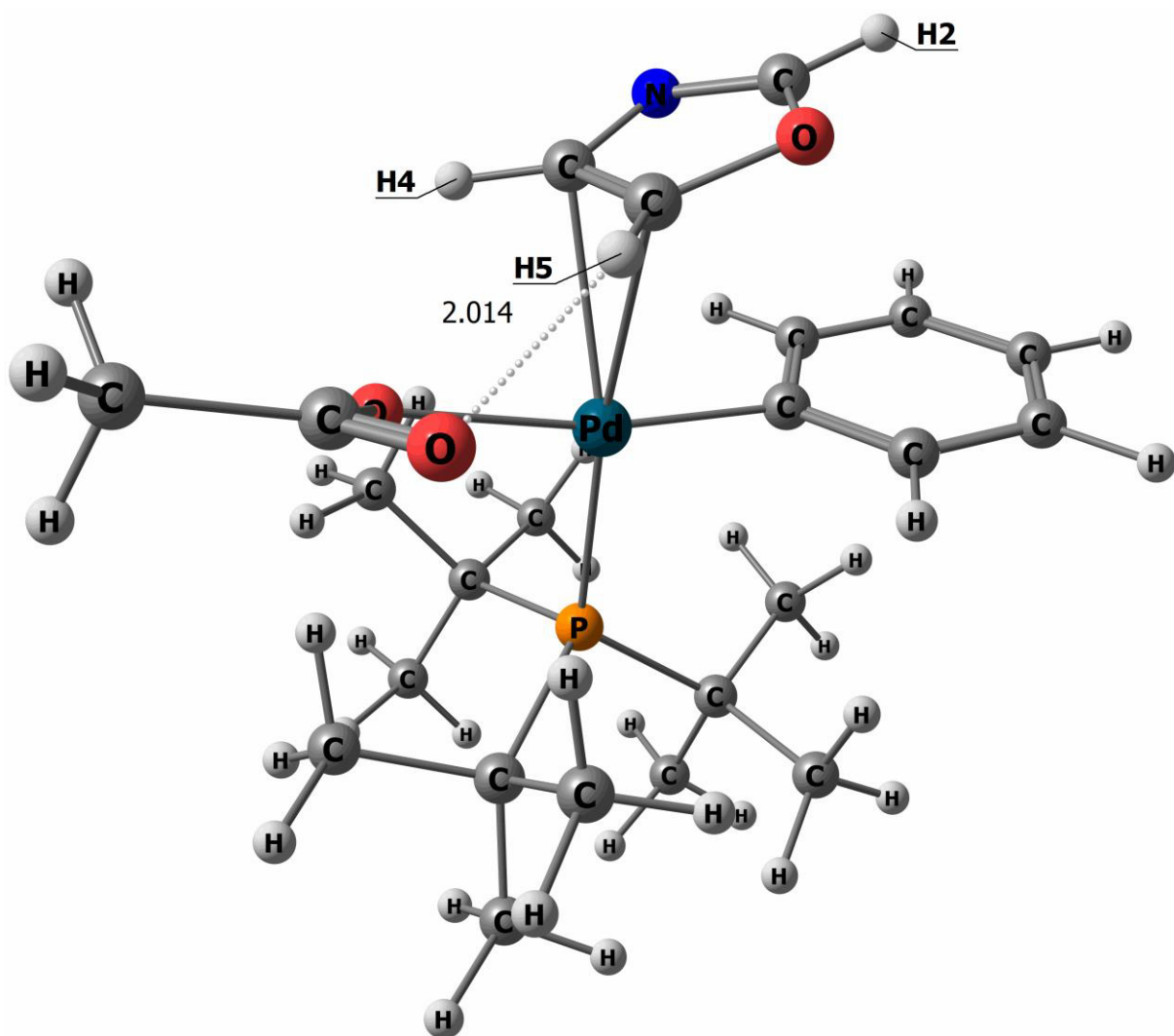


Figure 4-13: Transition State 1 =4

Pathway =5:

Int 1 for Pathway =5 is shown in Figure 4-14 and TS 1 in Figure 4-15. TS 1 has a negative frequency of -1318.83 cm^{-1} . Int 2 is presented in Figure 4-16 and TS 2 in Figure 4-17. TS 2 has a negative frequency of -357.74 cm^{-1}



[Figure 4-14: Intermediate 1=5](#)

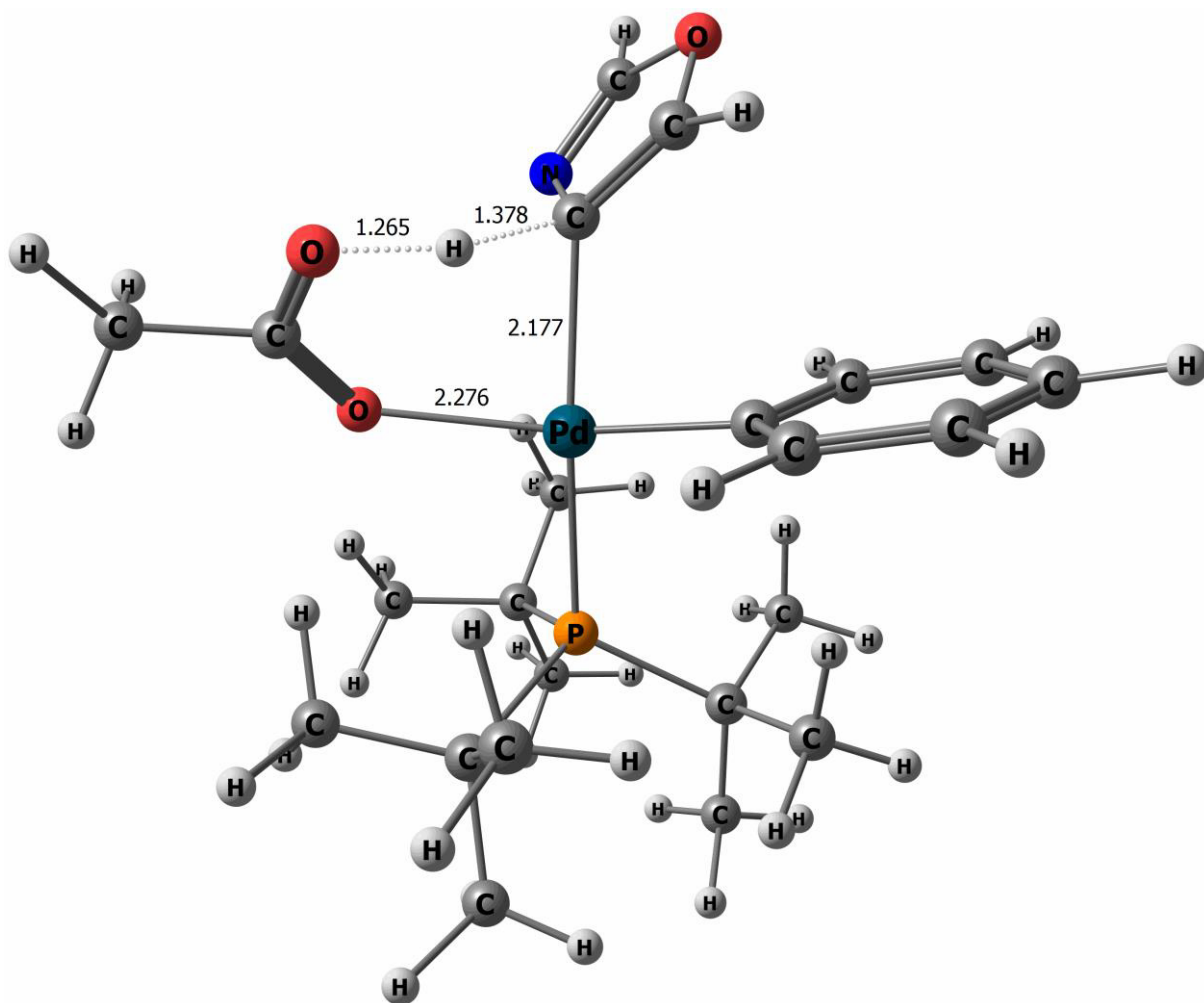


Figure 4-15: Transition State 1 =5

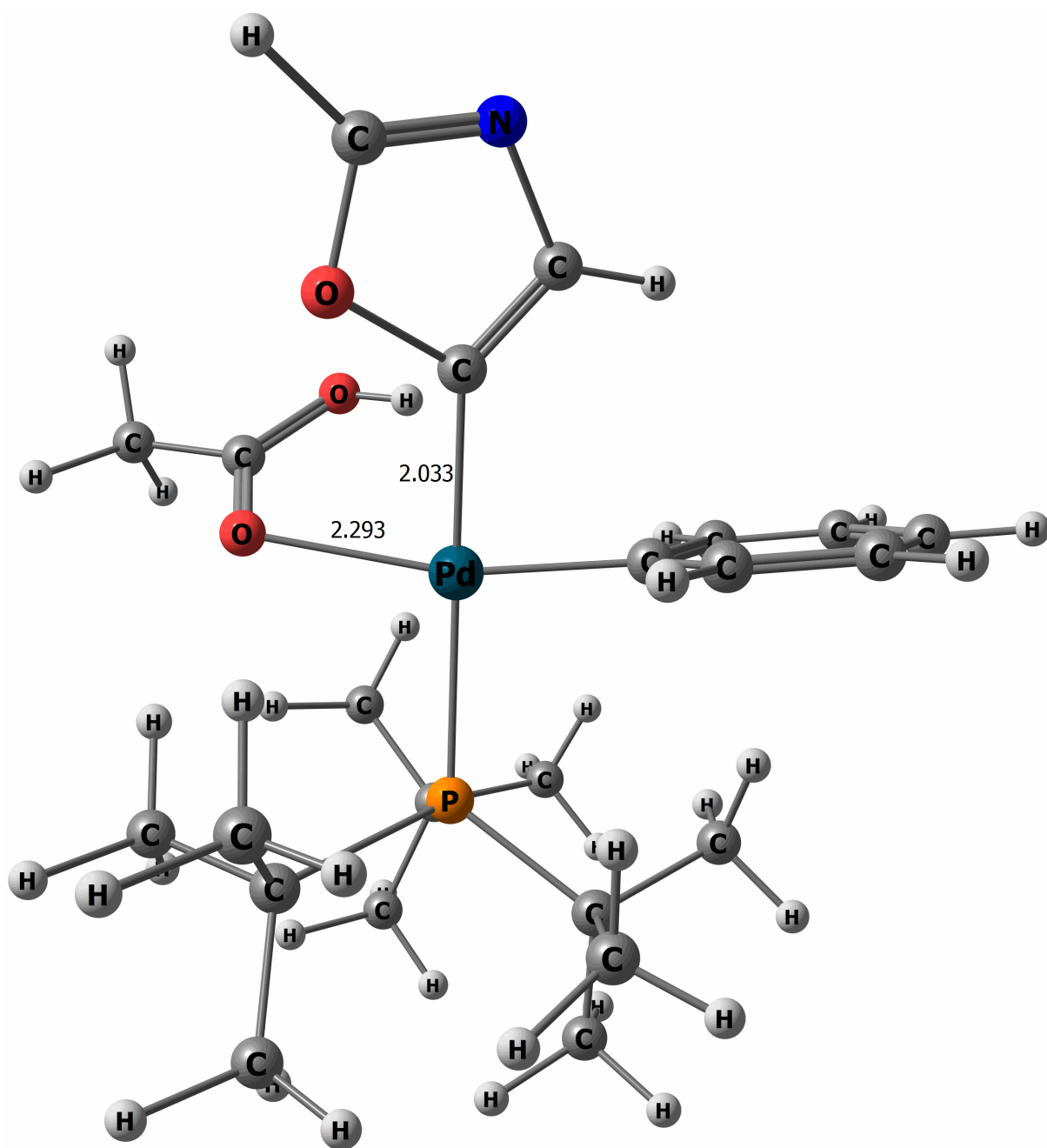


Figure 4-16: Intermediate 2 =5

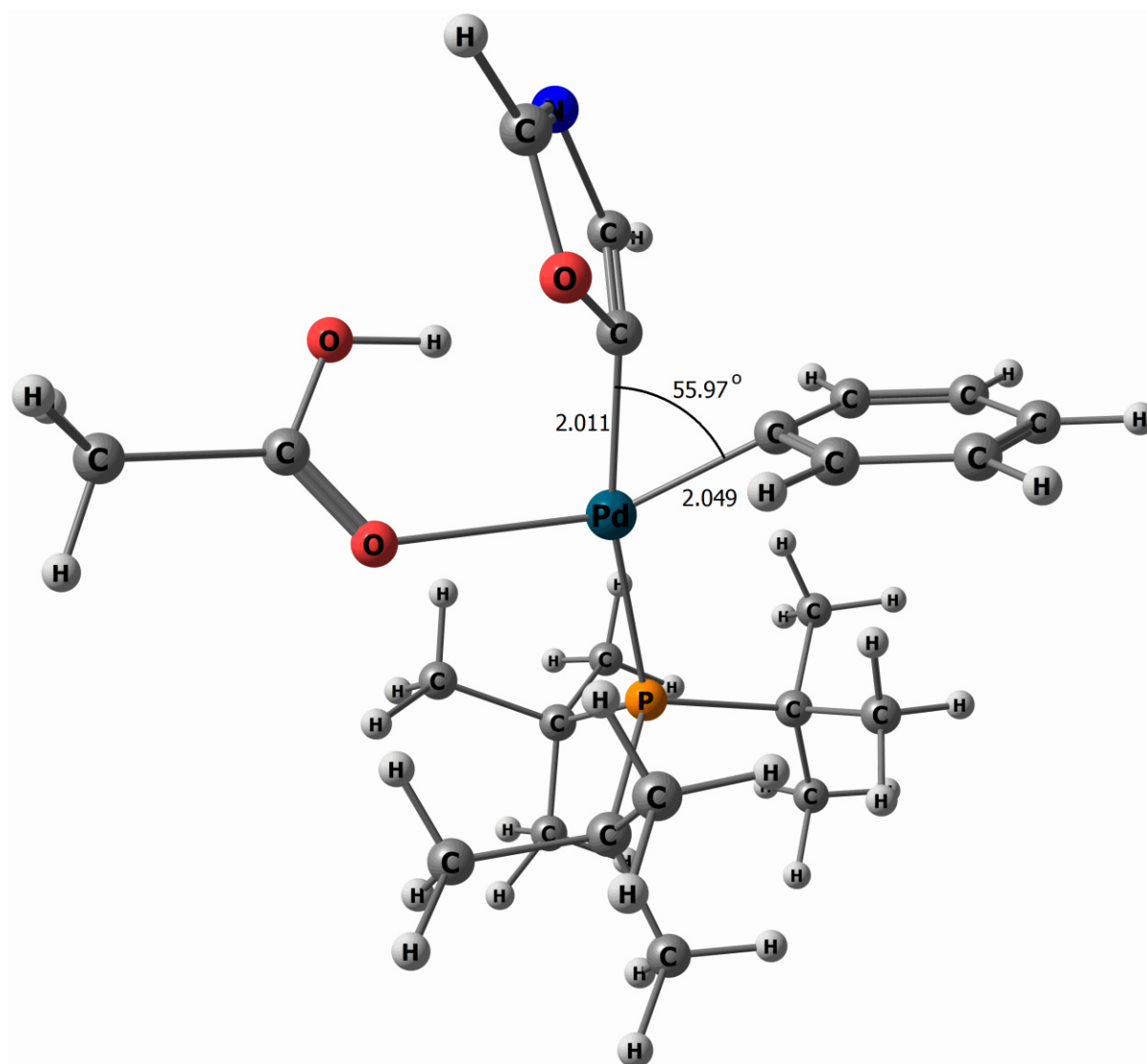


Figure 4-17: Transition State 2

DSD-PBEP86 Results:

The DSD-PBEP86 reaction profiles for mono-substitution are shown in Figure 4-18. The corresponding reaction Gibbs energies are shown in Table 4-1. The reaction profiles differ dramatically from the DMA coordinated analogues. The coordination of oxazole to the active catalyst (i.e. the formation of Int 1) now requires significant inputs of free energy to overcome the steric hindrance that the P^tBu_3 ligand induces. The formation of Int 1 is non-spontaneous across all pathways and requires, on average, $12.08 \text{ kcal.mol}^{-1}$. This initial reaction contributes significantly to the total Gibbs energy requirements of each pathway and in many cases, is equivalent to the barrier faced overcoming a TS.

The rate determining step for each pathway, Table 4-2, is the CMD reaction. For Pathway N4 and Pathway =4 this is by the large margins of 17.31 and 14.08 kcal.mol⁻¹, respectively.

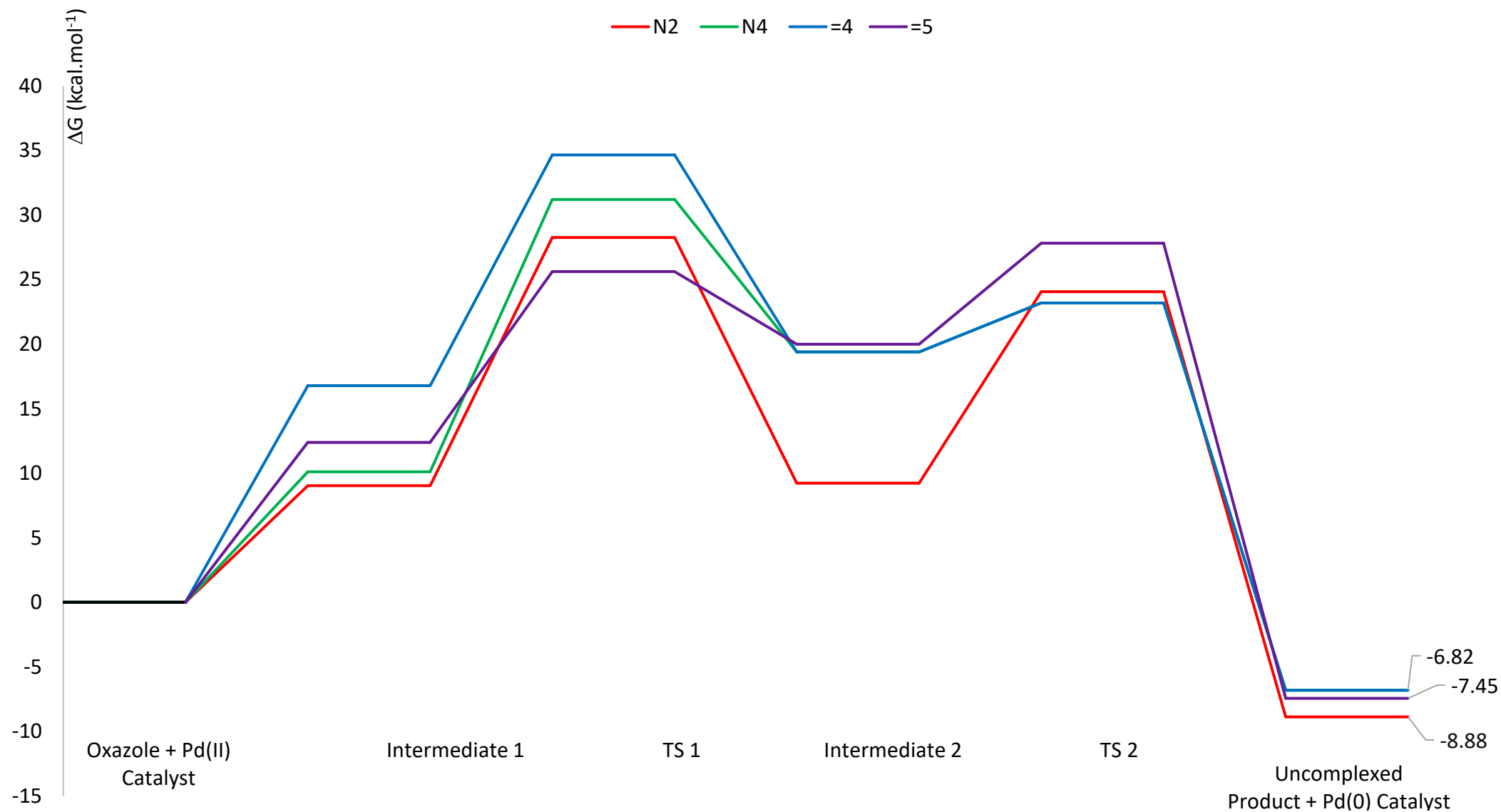


Figure 4-18: Mono-substitution Reaction Profiles (DSD-PBEP86/TZVP)

Table 4-1: Mono-substitution Reaction Free Energies (DSD-PBEP86/TZVP)

Chemical Reaction	ΔE (kcal.mol ⁻¹)	ΔG_{298K} (kcal.mol ⁻¹)
<u>Pathway N2</u>		
Catalyst + Oxazole $\rightarrow$ Int 1	-9.14	9.03
Int 1 $\rightarrow$ TS 1	22.41	19.23
TS 1 $\rightarrow$ Int 2	-22.08	-19.04
Int 2 $\rightarrow$ TS 2	16.31	14.83
TS 2 $\rightarrow$ Product 2 + Pd(0) Catalyst	-12.96	-32.76
<u>Pathway N4</u>		
Catalyst + Oxazole $\rightarrow$ Int 1	-7.80	10.11
Int 1 $\rightarrow$ TS 1	24.10	21.10
TS 1 $\rightarrow$ Int 2	-14.51	-11.81
Int 2 $\rightarrow$ TS 2	5.72	3.79
TS 2 $\rightarrow$ Product 4 + Pd(0) Catalyst	-10.83	-29.81
<u>Pathway =4</u>		
Catalyst + Oxazole $\rightarrow$ Int 1	-1.76	16.78
Int 1 $\rightarrow$ TS 1	21.90	17.88
TS 1 $\rightarrow$ Int 2	-18.35	-15.26
Int 2 $\rightarrow$ TS 2	5.72	3.79
TS 2 $\rightarrow$ Product 4 + Pd(0) Catalyst	-10.83	-29.81
<u>Pathway =5</u>		
Catalyst + Oxazole $\rightarrow$ Int 1	-6.55	12.38
Int 1 $\rightarrow$ TS 1	17.43	13.24
TS 1 $\rightarrow$ Int 2	-8.21	-5.63
Int 2 $\rightarrow$ TS 2	9.24	7.83
TS 2 $\rightarrow$ Product 5 + Pd(0) Catalyst	-15.84	-35.27

Table 4-2: Reductive Elimination vs CMD Transition State Gibbs Energies (DSD-PBEP86/TZVP)

Pathway	$\Delta^\ddagger G_{TS1} - \Delta^\ddagger G_{TS2}$ (kcal.mol ⁻¹)
N2	4.40
N4	17.31
=4	14.08
=5	5.41

It should be mentioned that for all pathways except Pathway N2, $\Delta G_{Int1} > \Delta G^\ddagger_{TS2}$. As can be seen from Table 4-3, ΔG_{Int1} is, in most cases, much closer in magnitude to $\Delta G^\ddagger_{TS1}$ than is $\Delta G^\ddagger_{TS2}$.

Table 4-3 Reductive Elimination vs Intermediate 1 Gibbs Energies (DSD-PBEP86/TZVP)

Pathway	$\Delta^\ddagger G_{TS\ 1} - \Delta G_{Int\ 1}$ (kcal.mol ⁻¹)
N2	10.20
N4	10.99
=4	1.09
=5	0.85

In fact, for Pathway =5 the difference between the two processes is within the assumed error of the method. This implies that research focusing on lowering $\Delta G_{Int\ 1}$ will be equally beneficially in speeding up Product 5 formation as research focusing on lowering the CMD barrier.

Although, RE contributes to the total energy requirement, when these reactions are performed with the P^tBu₃ ligand, their rates are entirely dictated by CMD and possibly in certain cases by Int 1 formation, at least for mono-substitution.

Ranking of the pathways according to the Gibbs energy requirement of the RDS is presented in Table 4-4. It is observed that under these conditions, Pathway =5 requires the least free energy for the RDS and thus Product 5 remains the kinetic product. This is by the margin of 4.64 kcal.mol⁻¹ when compared to the next lowest, Pathway =4. All pathways occupy a range of less than 10 kcal.mol⁻¹.

Table 4-4 Pathway Favourability in Terms of the RDS (DSD-PBEP86/TZVP)

Rank	Pathway	RDS	ΔG_{RDS} (kcal.mol ⁻¹)
1 st	=5	CMD (TS 1)	13.24
2 nd	=4	CMD (TS 1)	17.88
3 rd	N2	CMD (TS 1)	19.23
4 th	N4	CMD (TS 1)	21.10

The experimental results are produced again in Table 4-5 for convenience. In this table, Entry 2 corresponds to the reaction conditions concerning this chapter. Product 5 represents 30.6% of the total product distribution and is the only mono-substituted product formed in significant concentrations. This corresponds well with the predictions from theory.

Table 4-5: Experimental Product Distributions

Entry	Catalyst	Ligand	Product Distributions (%)			
			Product 5	Product 2	Product 2,5	Product 4,5
1	Pd(OAc) ₂	-	65.8	0.3	28.1	3.0
2	Pd(OAc) ₂	1 x P ^t Bu ₃	30.6	3.7	59.3	0.9
3	Pd(OAc) ₂	2 x P ^t Bu ₃	6.3	35.3	49.7	0.1

Comparing the Gibbs free energy of the RDS of each pathway with its DMA coordinated analogue from Chapter 3 (Table 4-6) reveals that all pathways leading to mono-substitution are expected to occur significantly faster under conditions of P^tBu₃ ligand. This is particularly true for Pathway =5 and Pathway N2.

Table 4-6: Pathway Favourability in Terms of the RDS: P^tBu₃ vs DMA

Pathway	RDS (P ^t Bu ₃)	ΔG RDS (P ^t Bu ₃) /kcal.mol ⁻¹	RDS (DMA)	ΔG RDS (DMA) /kcal.mol ⁻¹	Difference /kcal.mol ⁻¹
=5	CMD (TS 1)	13.24	CMD (TS 1)	18.30	-5.06
N4	CMD (TS 1)	21.10	RE (TS 2)	24.47	-3.37
=4	CMD (TS 1)	17.88	CMD (TS 1)	21.05	-3.17
N2	CMD (TS 1)	19.23	CMD (TS 1)	25.90	-6.67

In terms of the total $\Delta_r G$ values the reactions are all exergonic. Exactly the same products are being considered as in Chapter 3 and thus their relative favourability will be identical to before. Product 2 (-8.88 kcal.mol⁻¹) is the thermodynamic product followed by Product 5 (-7.45 kcal.mol⁻¹) and then finally Product 4 (-6.82 kcal.mol⁻¹). The result that these reactions are exergonic is expected to be an accurate description of the situation in solution, as the Pd(0) catalyst is coordinatively saturated and the coordination of solvent molecules (which have not been explicitly modelled) will not change this qualitatively, unlike in Chapter 3 where the Pd(0) catalyst is coordinatively unsaturated.

Pathway N2 is not expected to be kinetically accessible. However, Product 2 is the thermodynamic product. The 3.7% of Product 2 observed experimentally can, as under phosphine free conditions, be explained by the establishment of an equilibrium among the mono-substitution pathways.

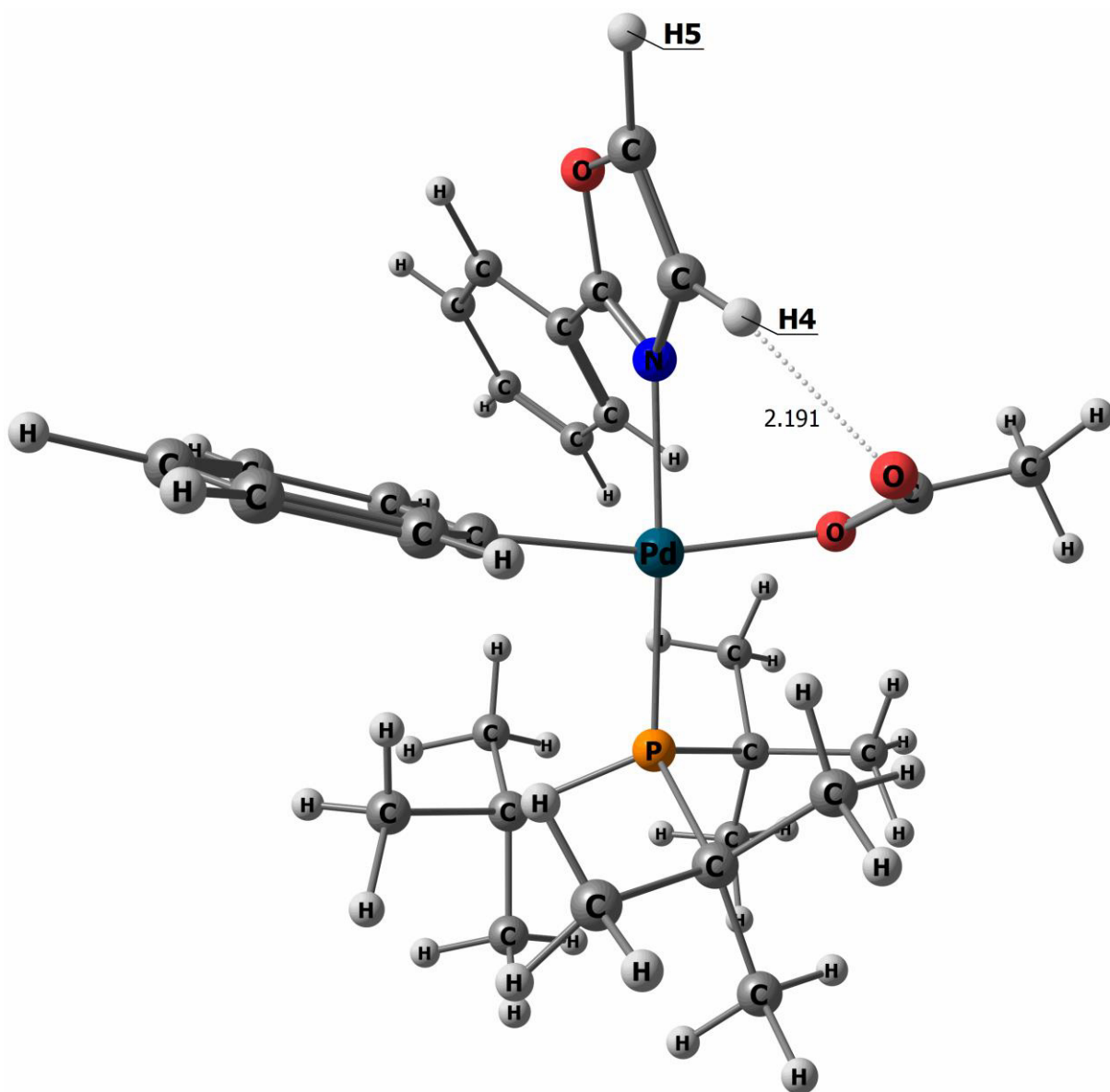
As seen in Chapter 3, the analysis of mono- and di-substitution cannot take place in isolation as the product of one pathway is a reactant in the next and thus further conclusions must be delayed until the di-substituted data is presented.

4.2 Di-substitution

The Reaction Pathways:

Pathway 2 $\rightarrow$ 2,4:

Starting from Product 2 entering the catalytic cycle for a second time, there are two ways that this molecule can coordinate to the active catalyst, with position 4 aligned to be deprotonated; via the Nitrogen atom (Figure 4-19) or via the double bond between carbons 4 and 5. Nitrogen coordination is favoured by 7.05 kcal.mol⁻¹ and will thus be the only option considered. The optimized structure of the double bond coordinated intermediate can however be found in the ESI for this chapter. The CMD transition state (TS 1) for Pathway 2 $\rightarrow$ 2,4 is depicted in Figure 4-20, along with selected bond lengths (all in Å). This structure has an imaginary frequency of -1310.60 cm⁻¹. The corresponding Intermediate 2 structure is shown in Figure 4-21 and is a zwitterionic species. The reductive elimination transition state (TS 2) is shown in Figure 4-22 and has an imaginary frequency of -361.25 cm⁻¹. In this structure, the acetate ligand is no longer deprotonated.



[Figure 4-19: Intermediate 1 Pathway 2 → 2,4](#)

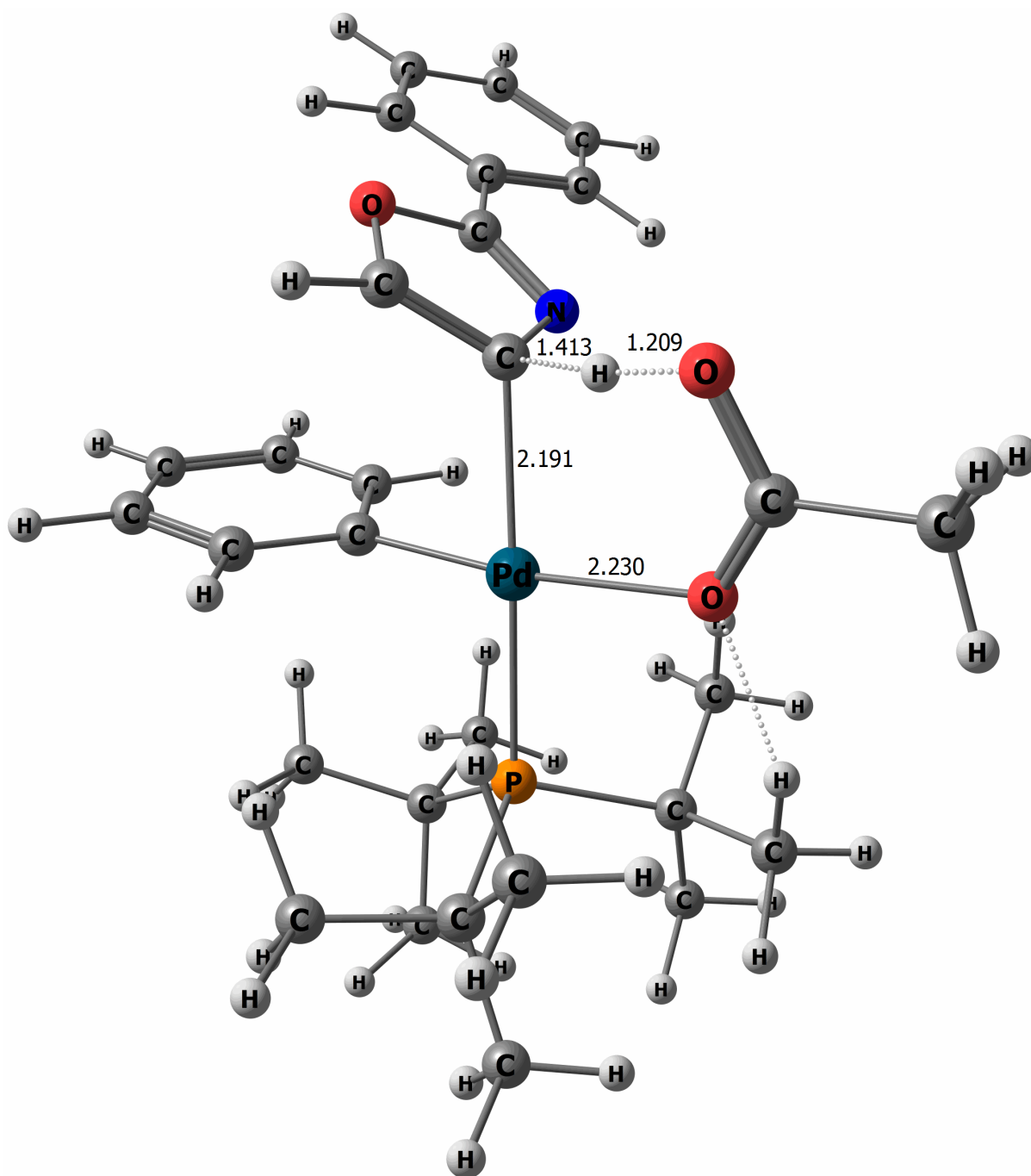
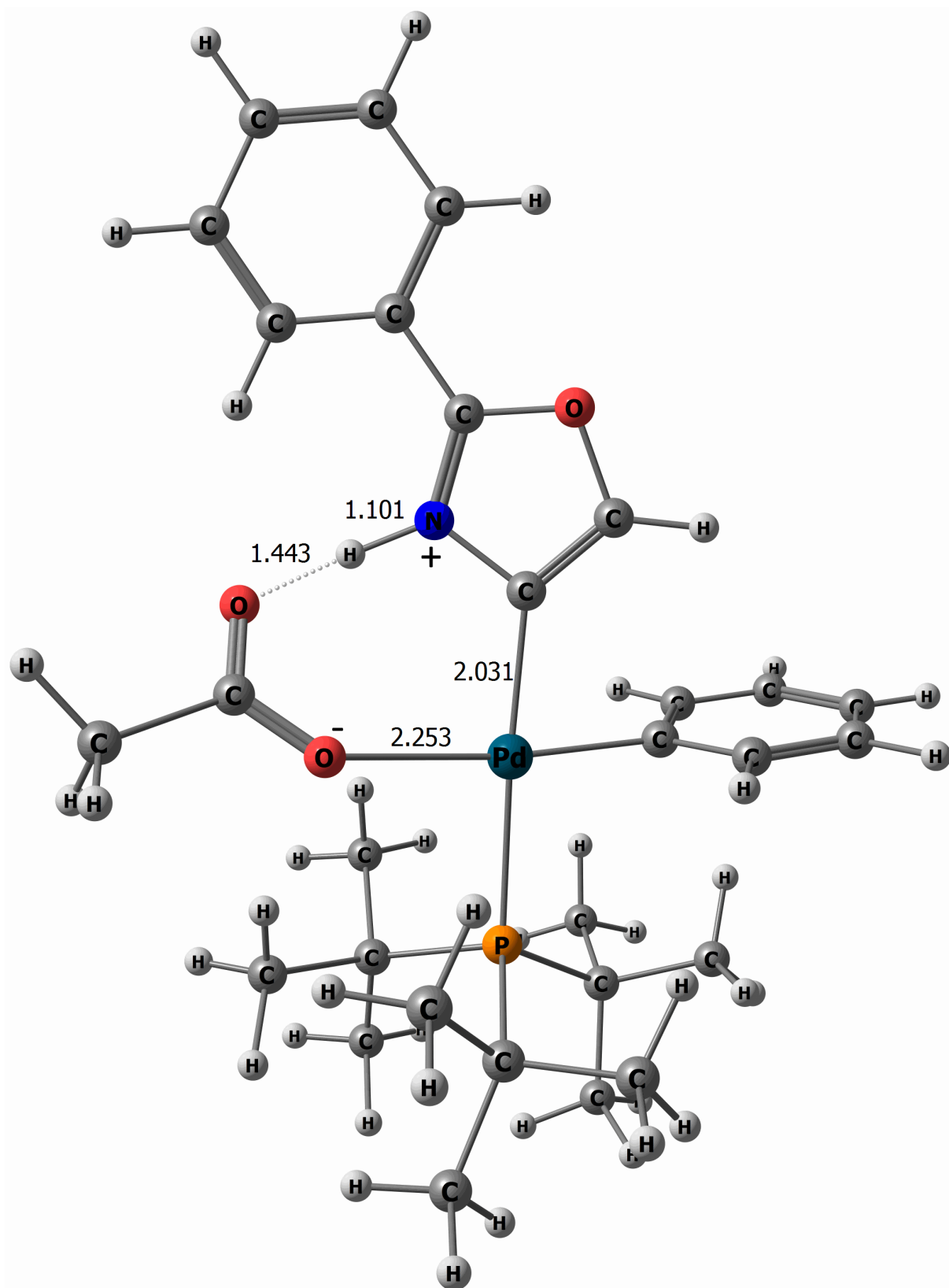


Figure 4-20: Transition State 1 Pathway 2 → 2,4



[Figure 4-21: Intermediate 2 Pathway 2 → 2,4](#)

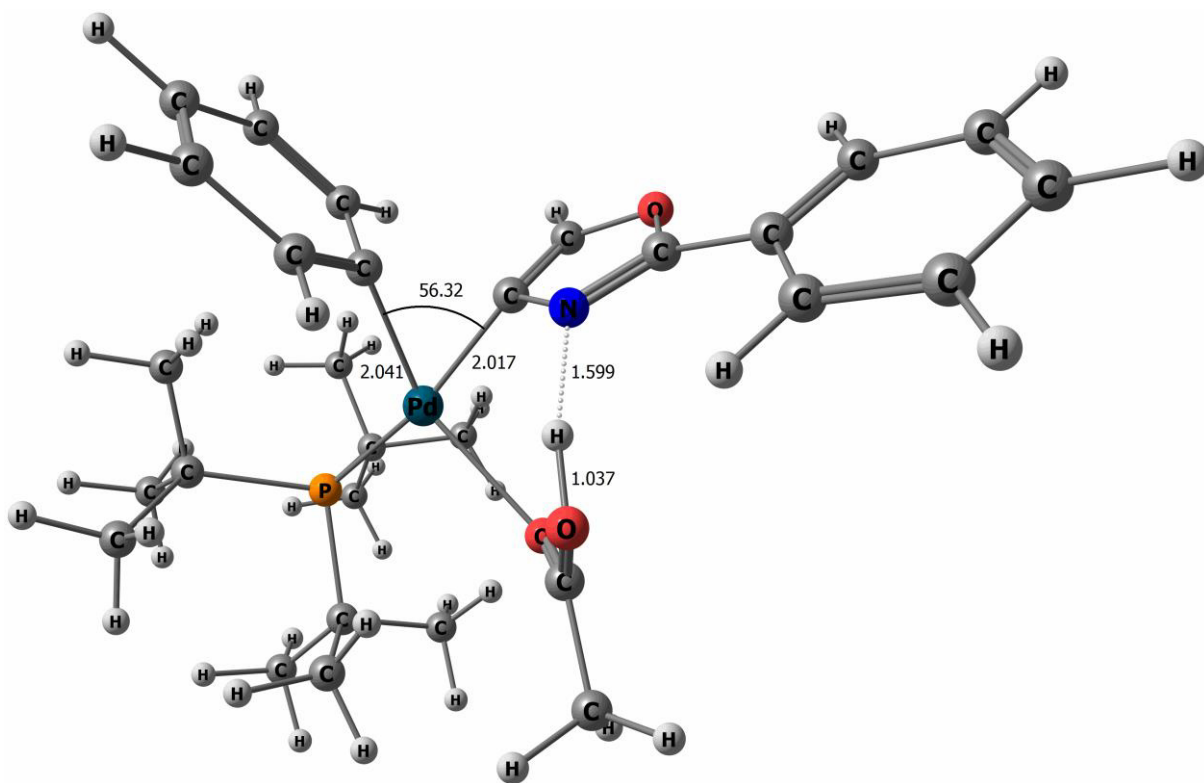
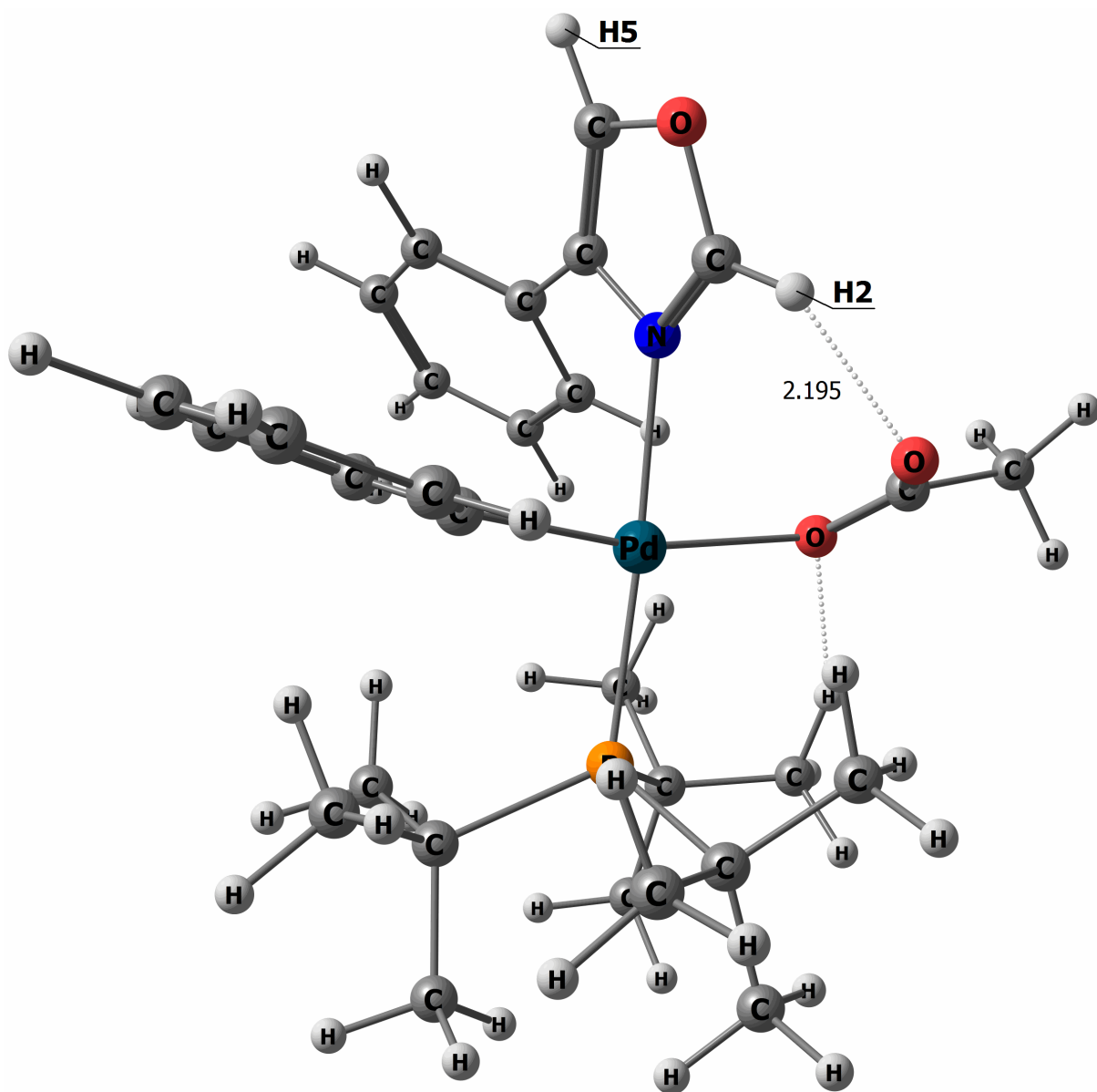


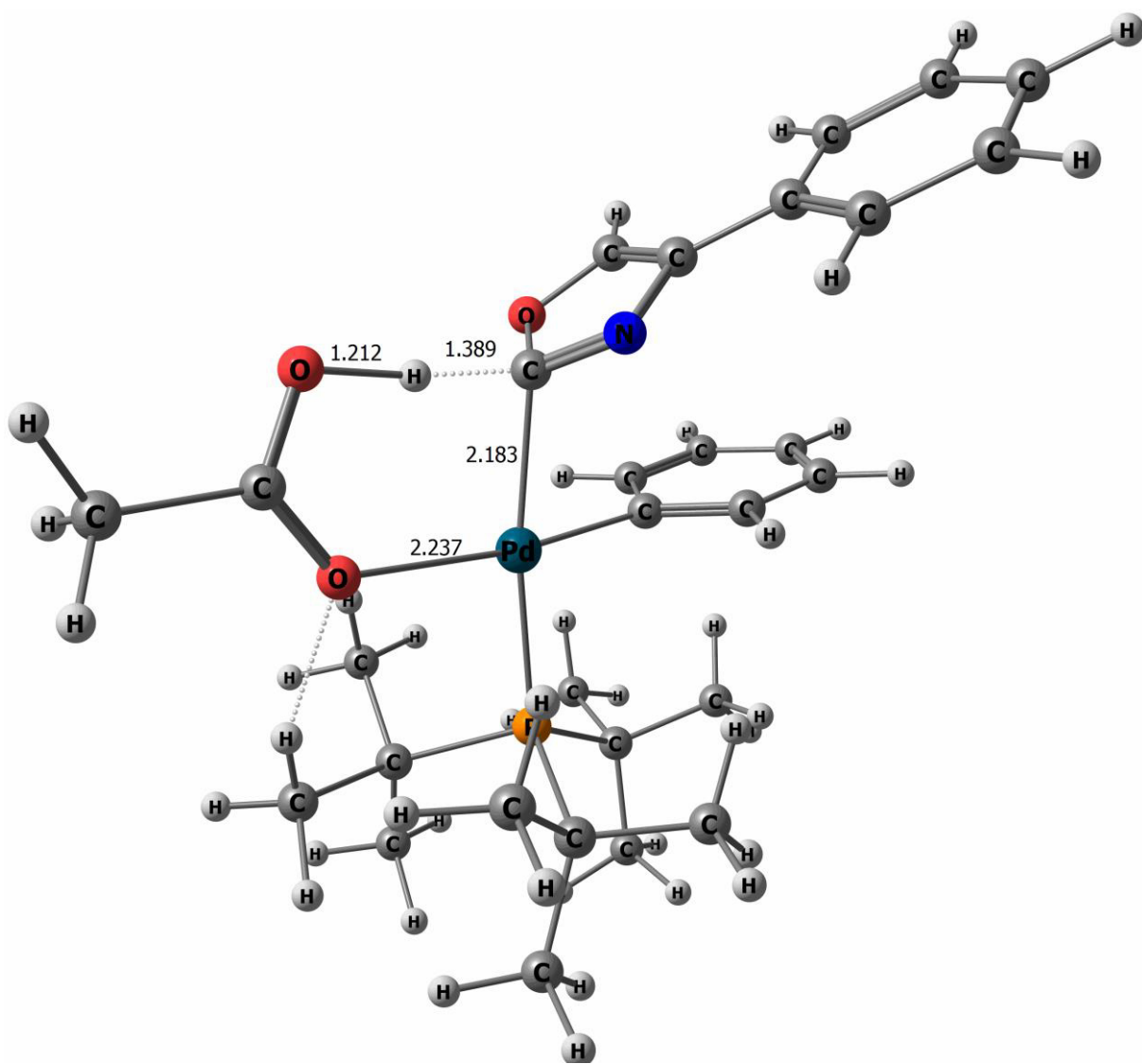
Figure 4-22: Transition State 2 Pathway 2 $\rightarrow$ 2,4

Pathway 4 $\rightarrow$ 2,4:

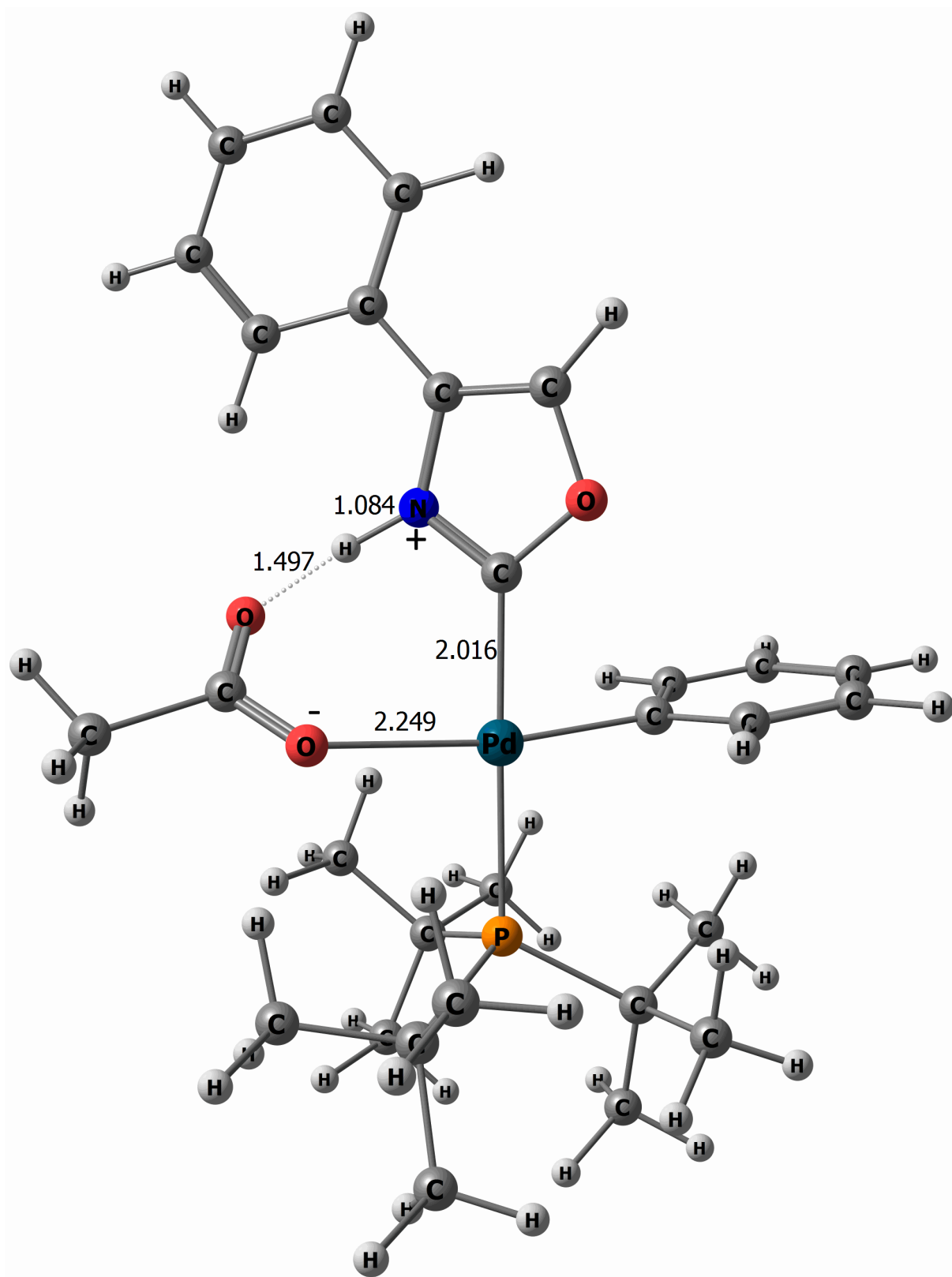
Starting from Product 4 entering the catalytic cycle for a second time, with the requirement that position 2 is aligned to be deprotonated, leads to the Intermediate 1 structure depicted in Figure 4-23. As double bonded coordinated species have consistently proven unfavourable, when compared to Nitrogen coordinated alternatives, none were considered in this case. However, an alternative “enantiomeric” conformation can be found in the ESI. The CMD transition state (TS 1) for Pathway 4 $\rightarrow$ 2,4 is pictured in Figure 4-24. This structure has an imaginary frequency of -1070.02 cm^{-1} . The corresponding Intermediate 2 structure is portrayed in Figure 4-25. Here deprotonation of the acetate ligand by the nitrogen atom of oxazole takes place to form a zwitterion. This interaction persists in TS2, illustrated in Figure 4-26. This species has an imaginary frequency of -355.21 cm^{-1} .



[Figure 4-23: Intermediate 1 Pathway 4 → 2,4](#)



[Figure 4-24: Transition 1 Pathway 4 → 2,4](#)



[Figure 4-25: Intermediate 2 Pathway 4 → 2,4](#)

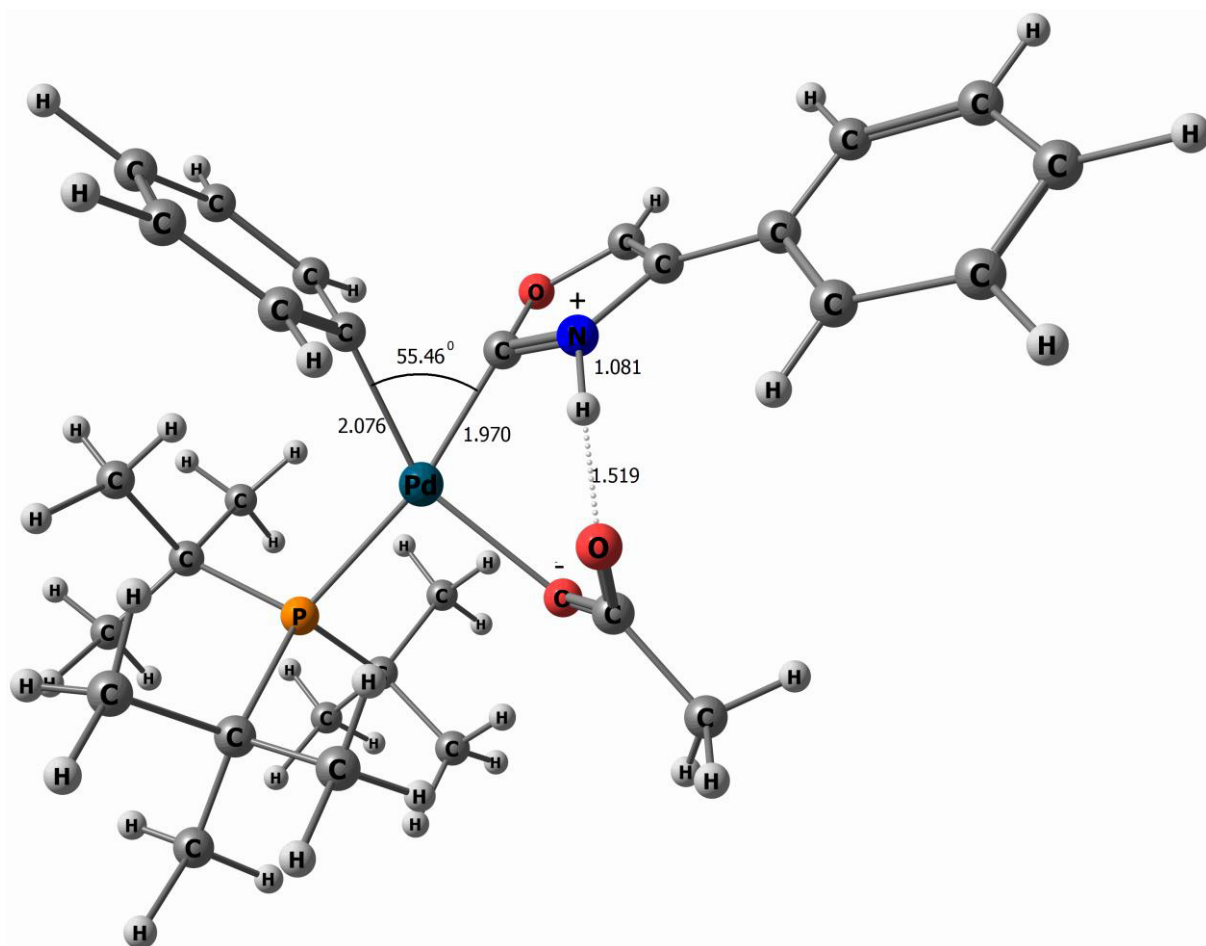
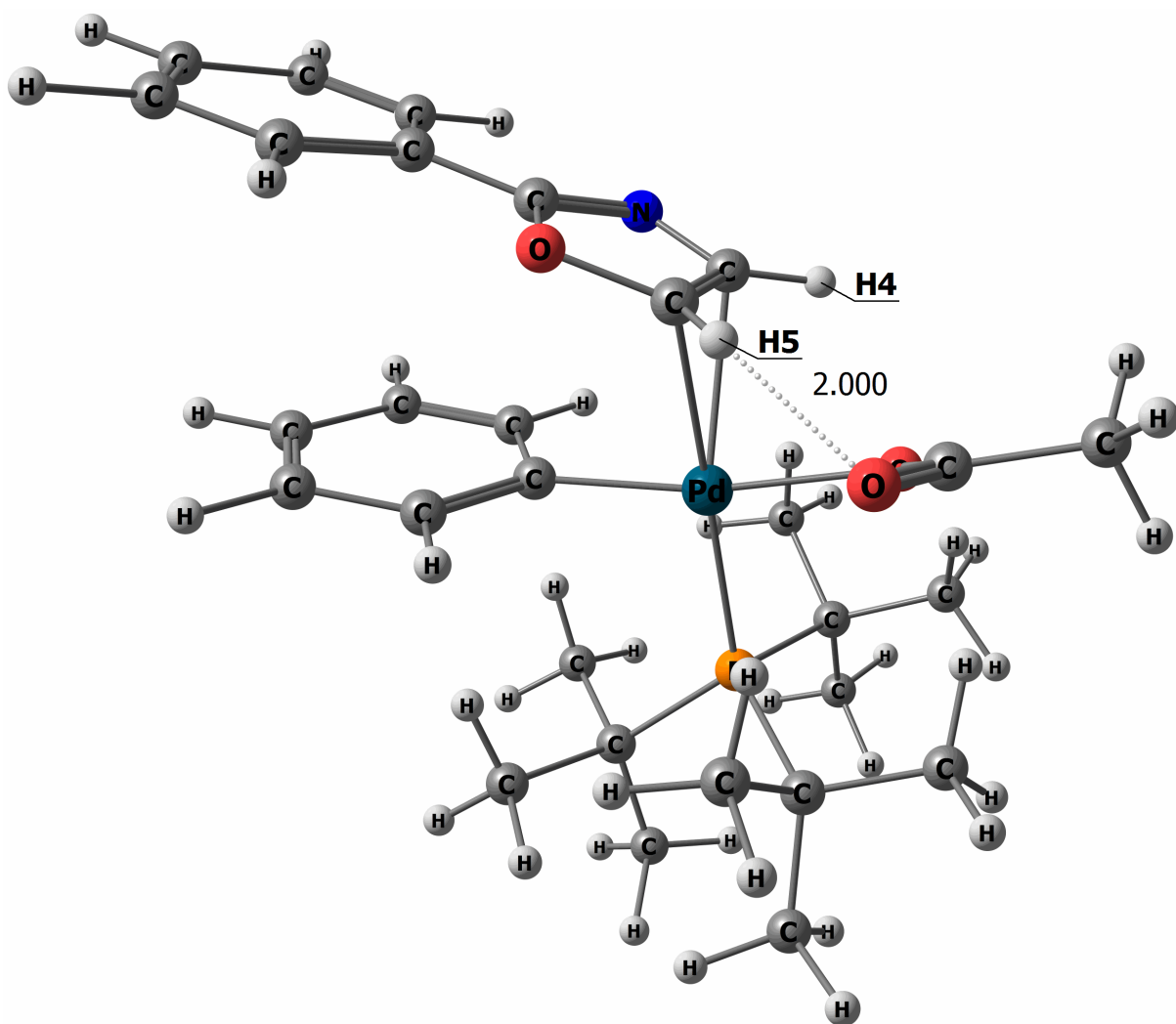


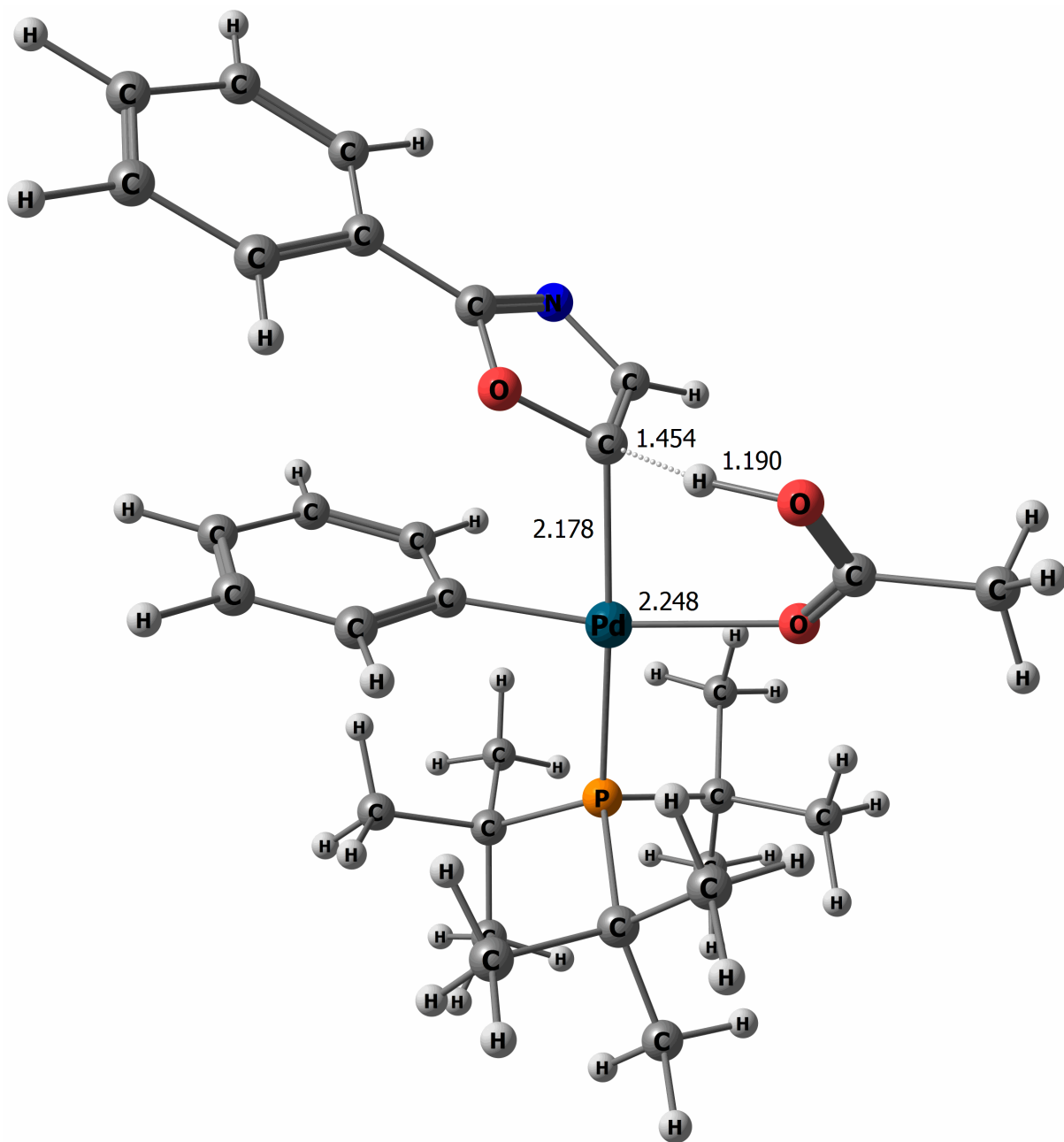
Figure 4-26: Transition State 2 Pathway 4 → 2,4

Pathway 2 → 2,5:

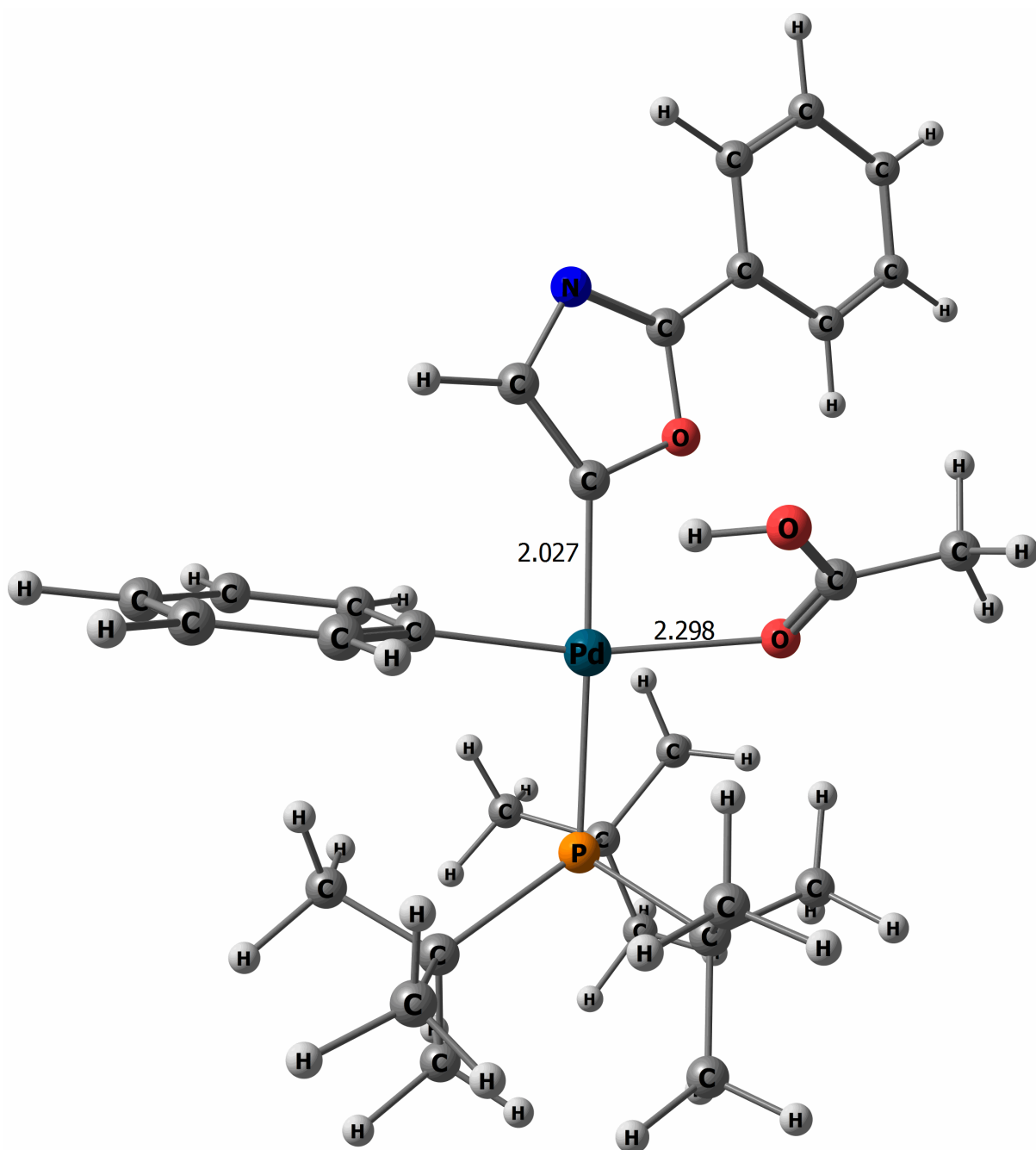
Starting from Product 2 entering the catalytic cycle for a second time, but now with position 5 aligned to be deprotonated, results in the Intermediate 1 structure shown in Figure 4-27. The CMD transition state (TS 1) for this pathway is pictured in Figure 4-28 and has an imaginary frequency of -1166.10 cm^{-1} . The corresponding Intermediate 2 structure is depicted in Figure 4-29 and the reductive elimination transition state (TS 2) in Figure 4-30. This species has an imaginary frequency of -360.50 cm^{-1} .



[Figure 4-27: Intermediate 1 Pathway 2 → 2,5](#)



[Figure 4-28: Transition State 1 Pathway 2 → 2,5](#)



[Figure 4-29: Intermediate 2 Pathway 2 → 2,5](#)

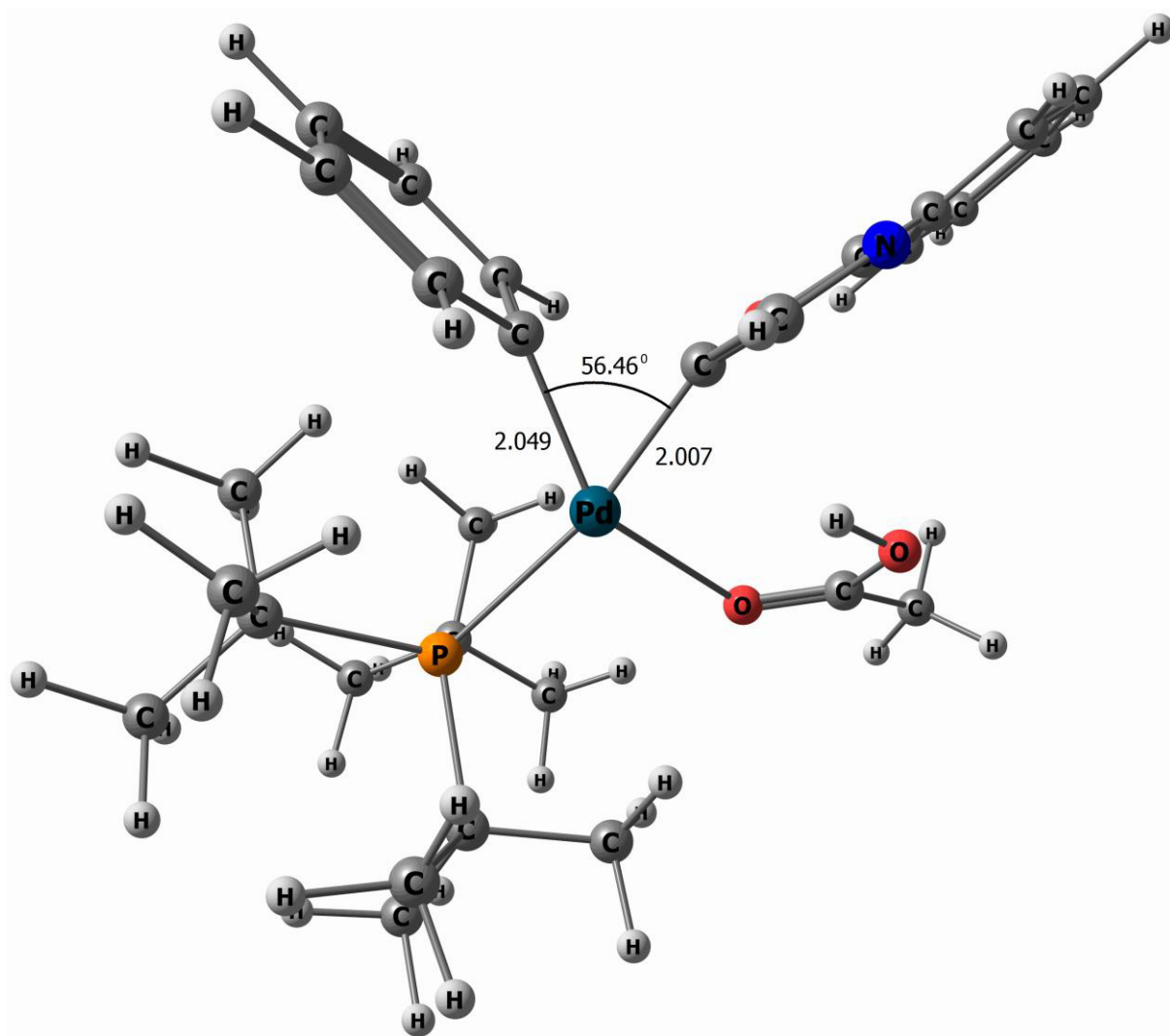
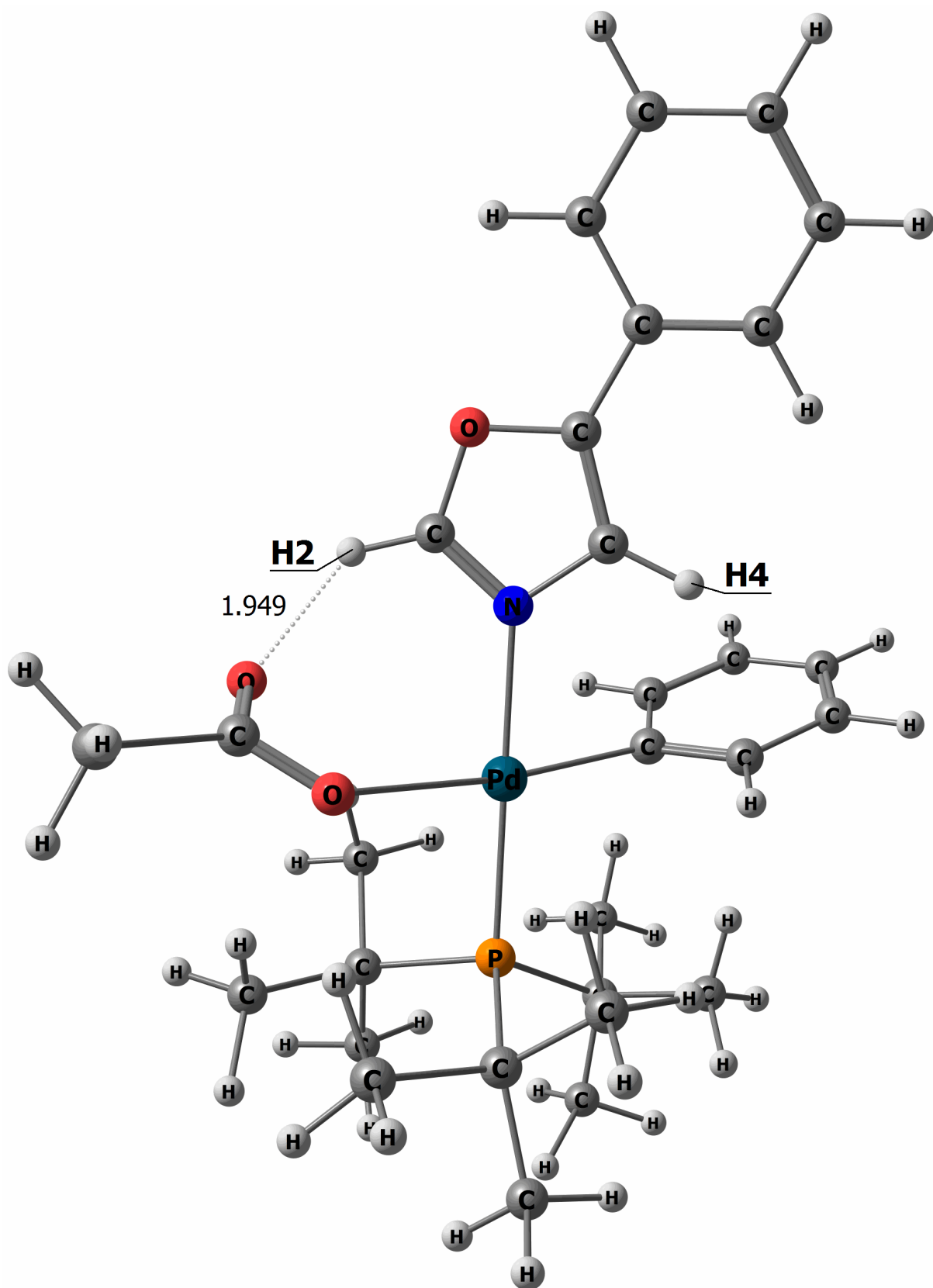


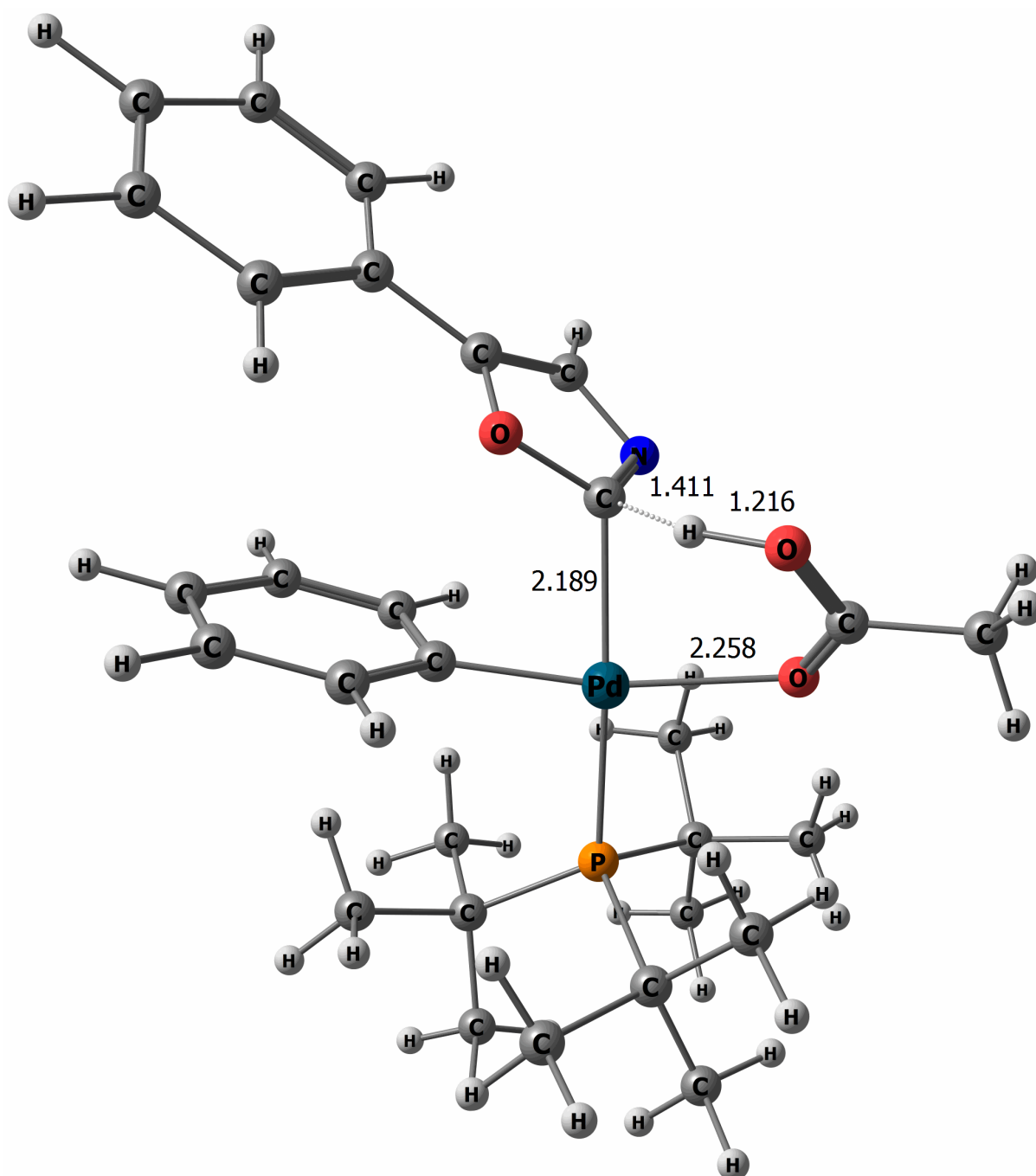
Figure 4-30: Transition State 2 Pathway 2 → 2,5

Pathway 5 → 2,5:

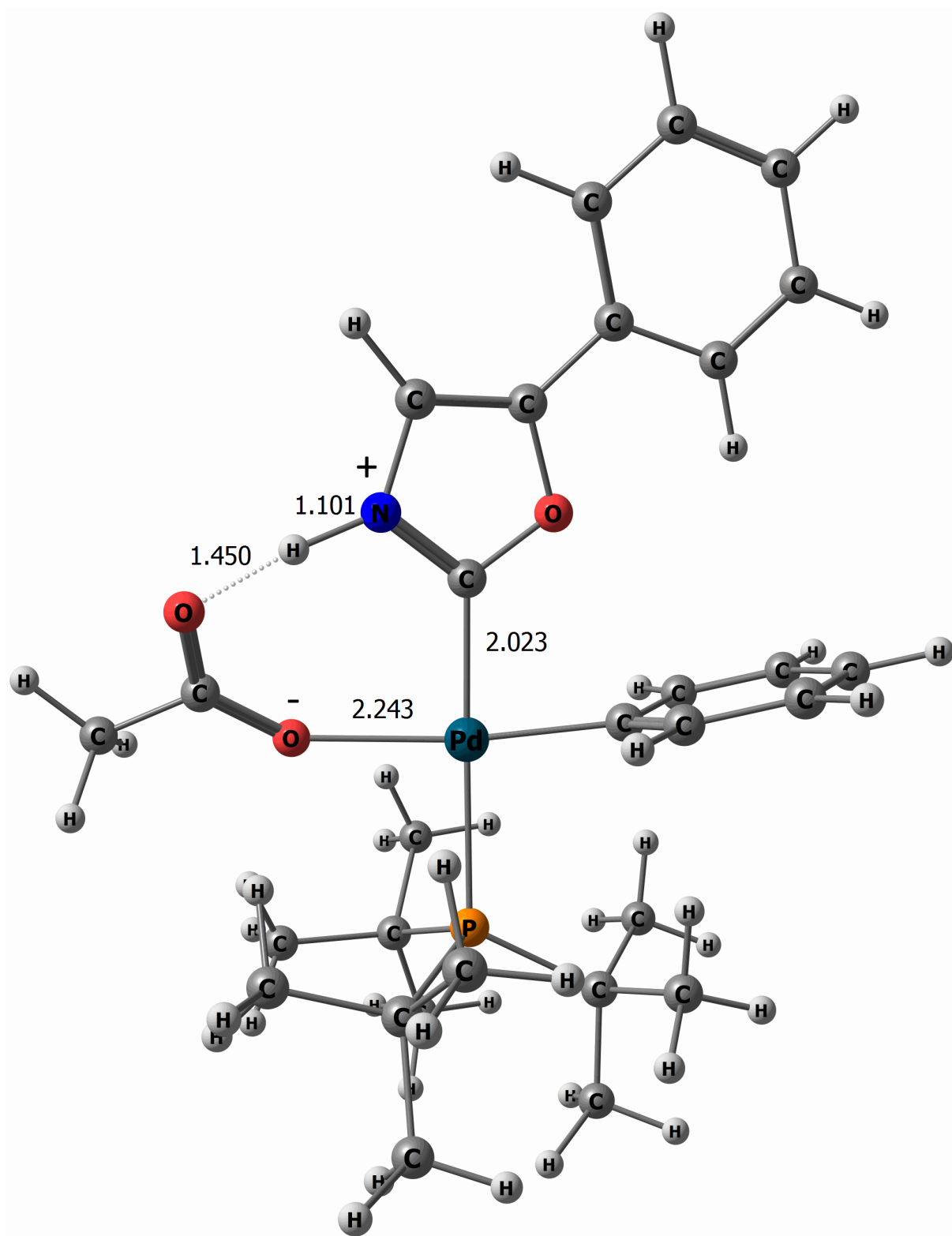
Starting from Product 5 entering the catalytic cycle for a second time, with position 2 aligned to be deprotonated, results in the Intermediate 1 structure shown in Figure 4-31. This is the beginning of Pathway 5 → 2,5. The CMD transition state (TS 1) for this pathway is shown in Figure 4-32 and has an imaginary frequency of -1241.21 cm^{-1} . The corresponding Intermediate 2 structure is shown in Figure 4-33 and is a zwitterion. The zwitterionic nature persists in Intermediate 2, presented in Figure 4-34, which has an imaginary frequency of -357.22 cm^{-1} .



[Figure 4-31: Intermediate 1 Pathway 5 → 2,5](#)



[Figure 4-32: Transition State 1 Pathway 5 → 2,5](#)



[Figure 4-33: Intermediate 2 Pathway 5 → 2,5](#)

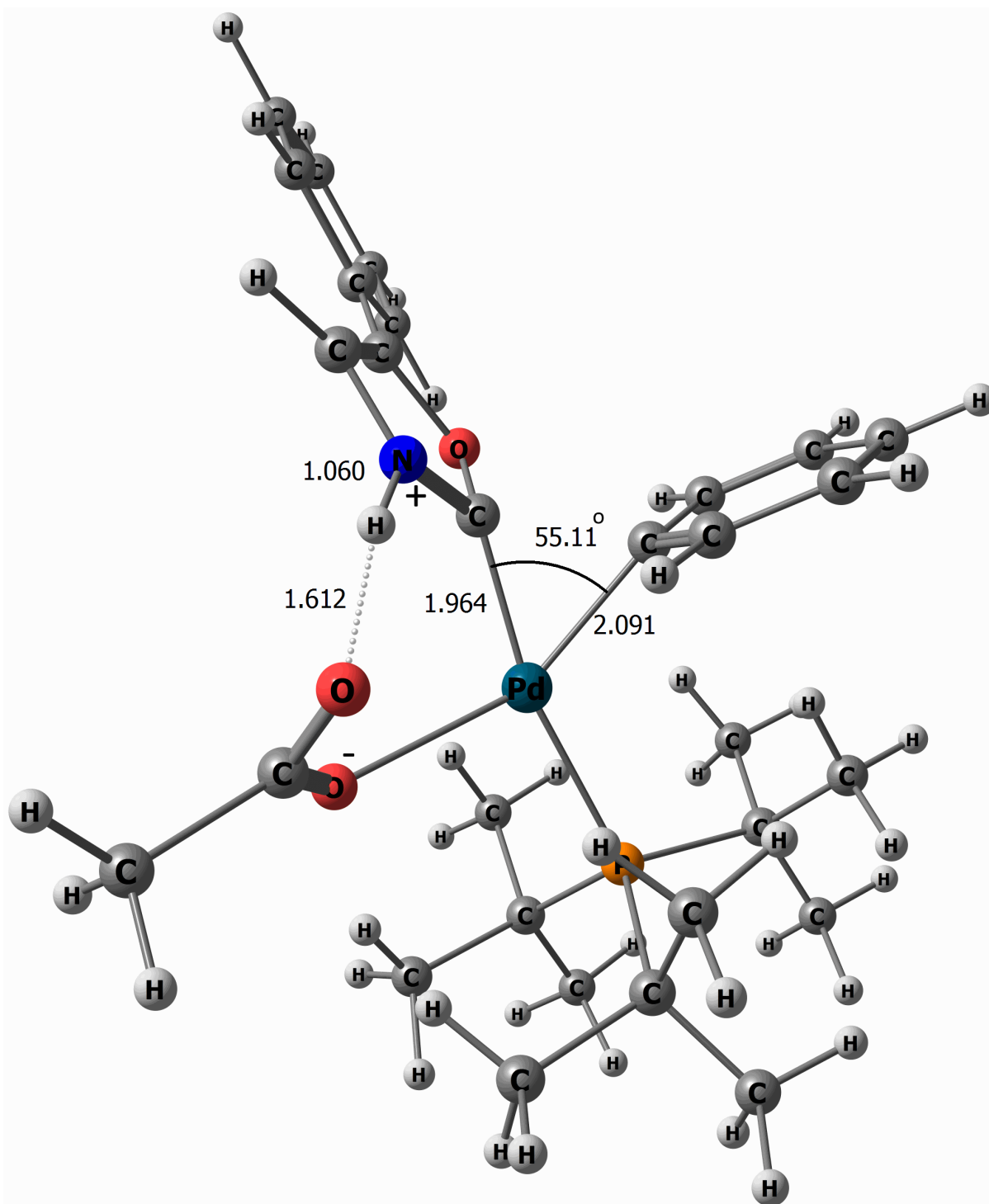
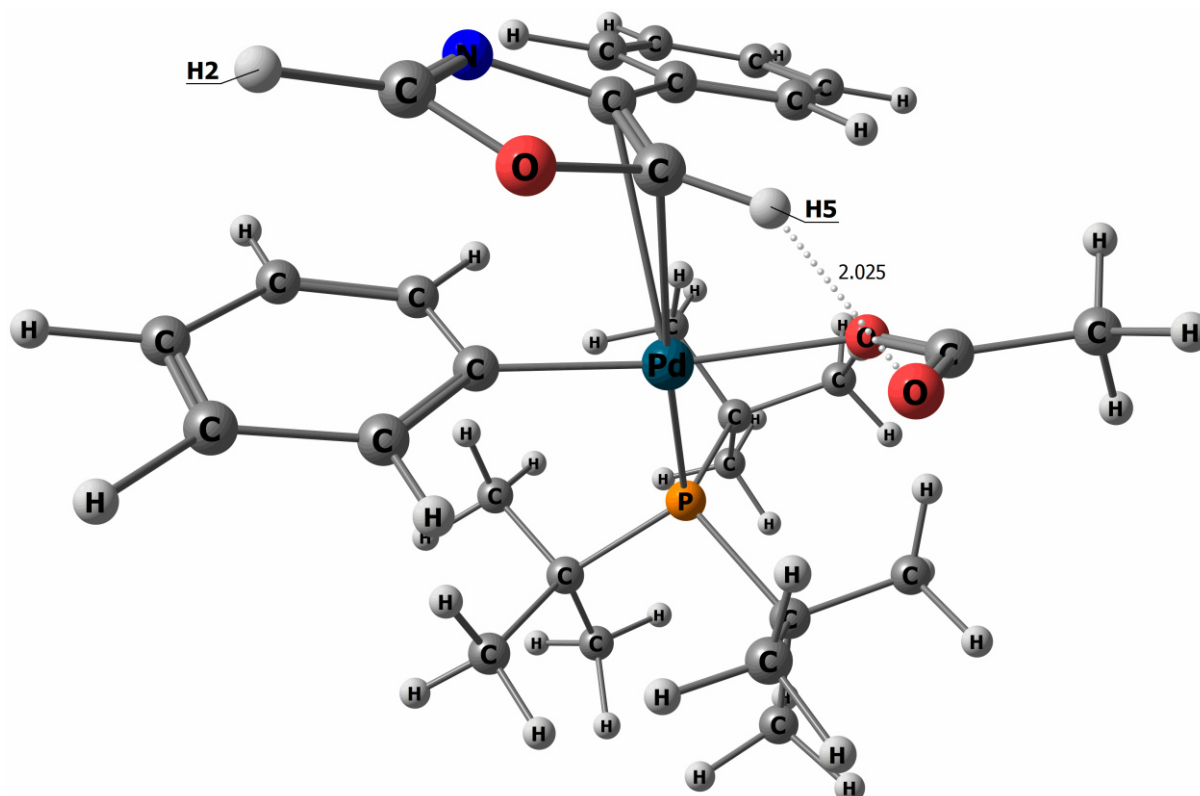


Figure 4-34: Transition State 2 Pathway 5 $\rightarrow$ 2,5

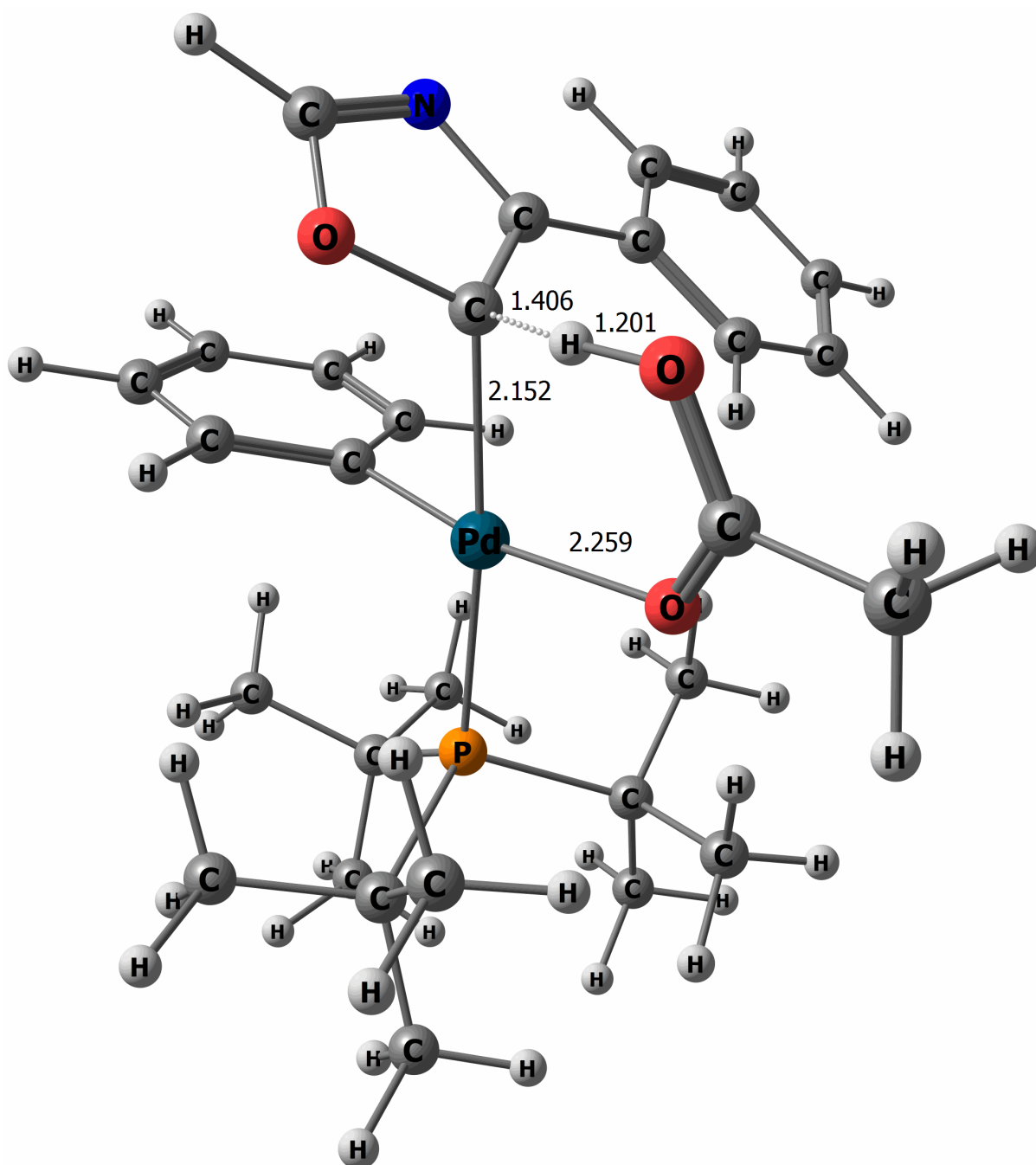
Pathway 4 $\rightarrow$ 4,5:

Starting from Product 4 entering the catalytic cycle for a second time, with the requirement that H5 is aligned to be deprotonated, leads to the Intermediate 1 structure illustrated in Figure 4-35.

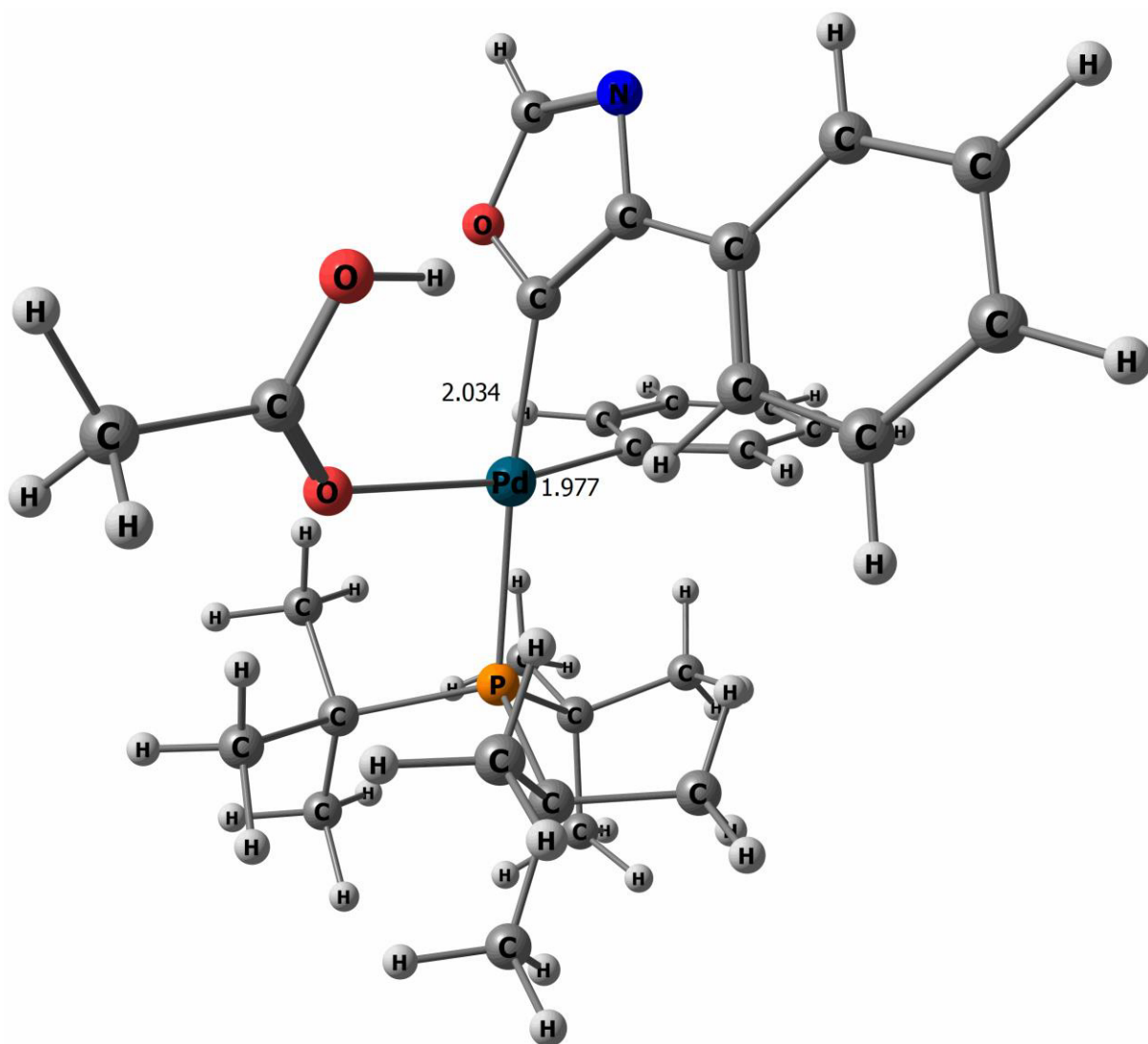
The CMD transition state (TS 1) for this pathway is depicted in Figure 4-36 and has an imaginary frequency of -1219.19 cm^{-1} . The corresponding Intermediate 2 structure is shown in Figure 4-37. As usual for deprotonation at the 5 position, no hydrogen bond between the acetate ligand and the nitrogen atom of oxazole is possible. The reductive elimination transition state (TS 2) is portrayed in Figure 4-38. This structure has an imaginary frequency of -364.78 cm^{-1} .



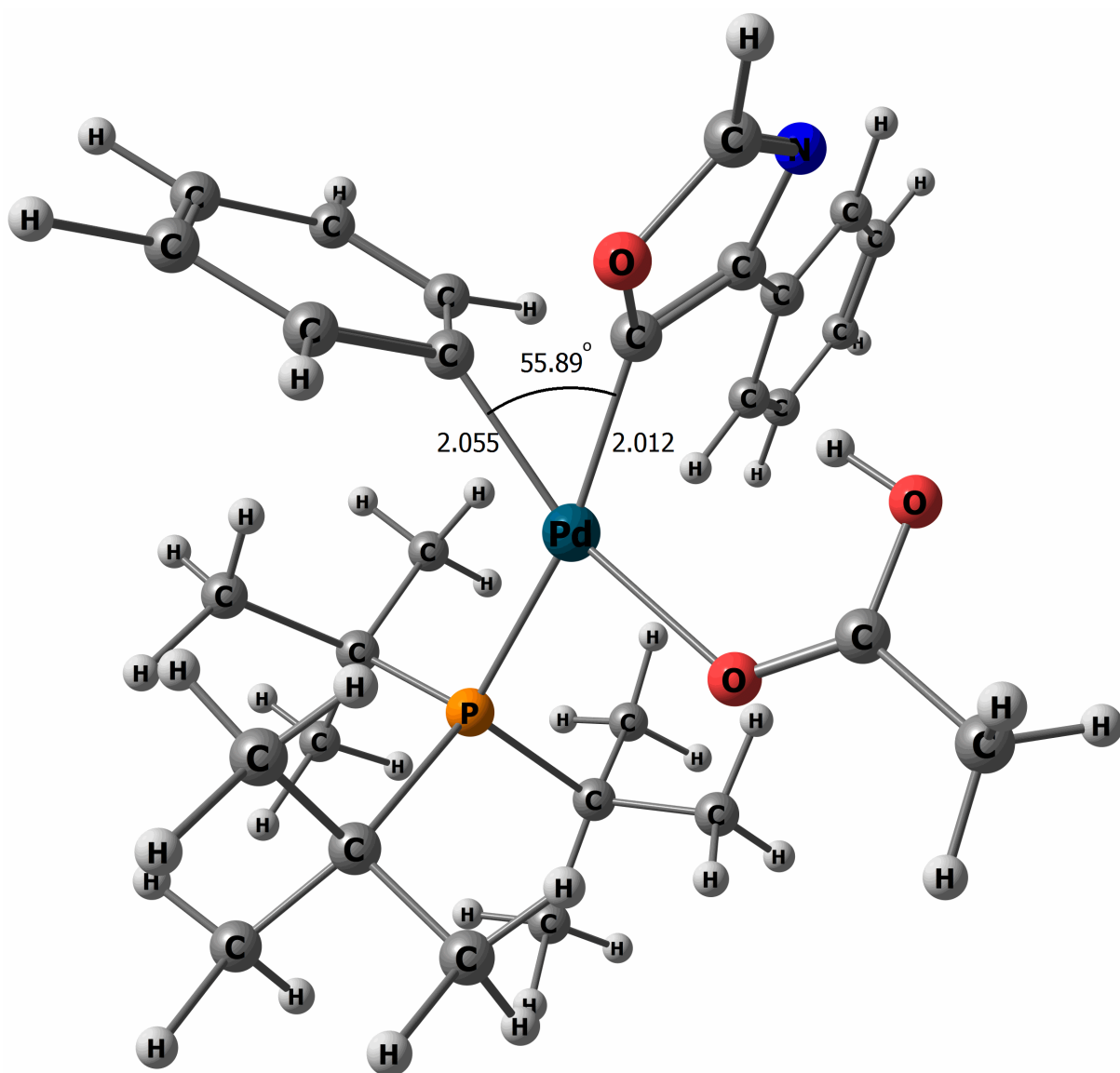
[Figure 4-35: Intermediate 1: Pathway 4 → 4,5](#)



[Figure 4-36: Transition State 1 Pathway 4 → 4,5](#)



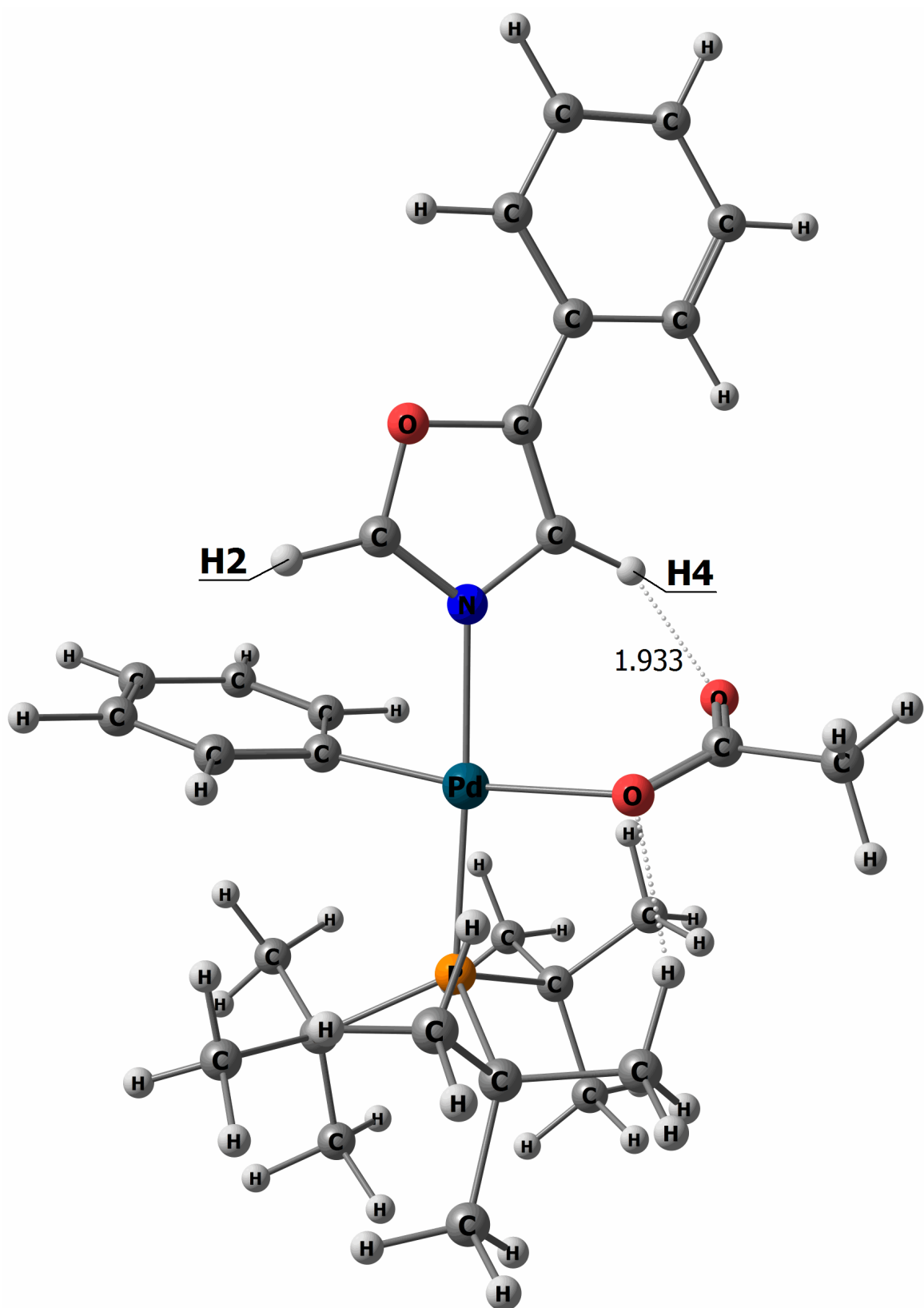
[Figure 4-37: Intermediate 2 Pathway 4 → 4,5](#)



[Figure 4-38: Transition State 2 Pathway 4 → 4,5](#)

Pathway 5 $\rightarrow$ 4,5:

Starting from Product 5 entering the catalytic cycle for a second time, there are two ways that this molecule can coordinate to the active catalyst, with position 4 aligned to be deprotonated; via the Nitrogen atom (Figure 4-39) or via the double bond between carbons 4 and 5. The nitrogen coordinated intermediate is 6.68 kcal.mol⁻¹ lower in Gibbs energy. An optimised structure of the double bond coordinated alternative can be found in the ESI for this chapter. The CMD transition state (TS 1) for this pathway is shown in Figure 4-40 and has an imaginary frequency of -1268.11 cm⁻¹. The corresponding Intermediate 2 structure is depicted in Figure 4-41 and the reductive elimination transition state (TS 2) in Figure 4-42. This structure has an imaginary frequency of -384.56 cm⁻¹.



[Figure 4-39: Intermediate 1 Pathway 5 → 4,5](#)

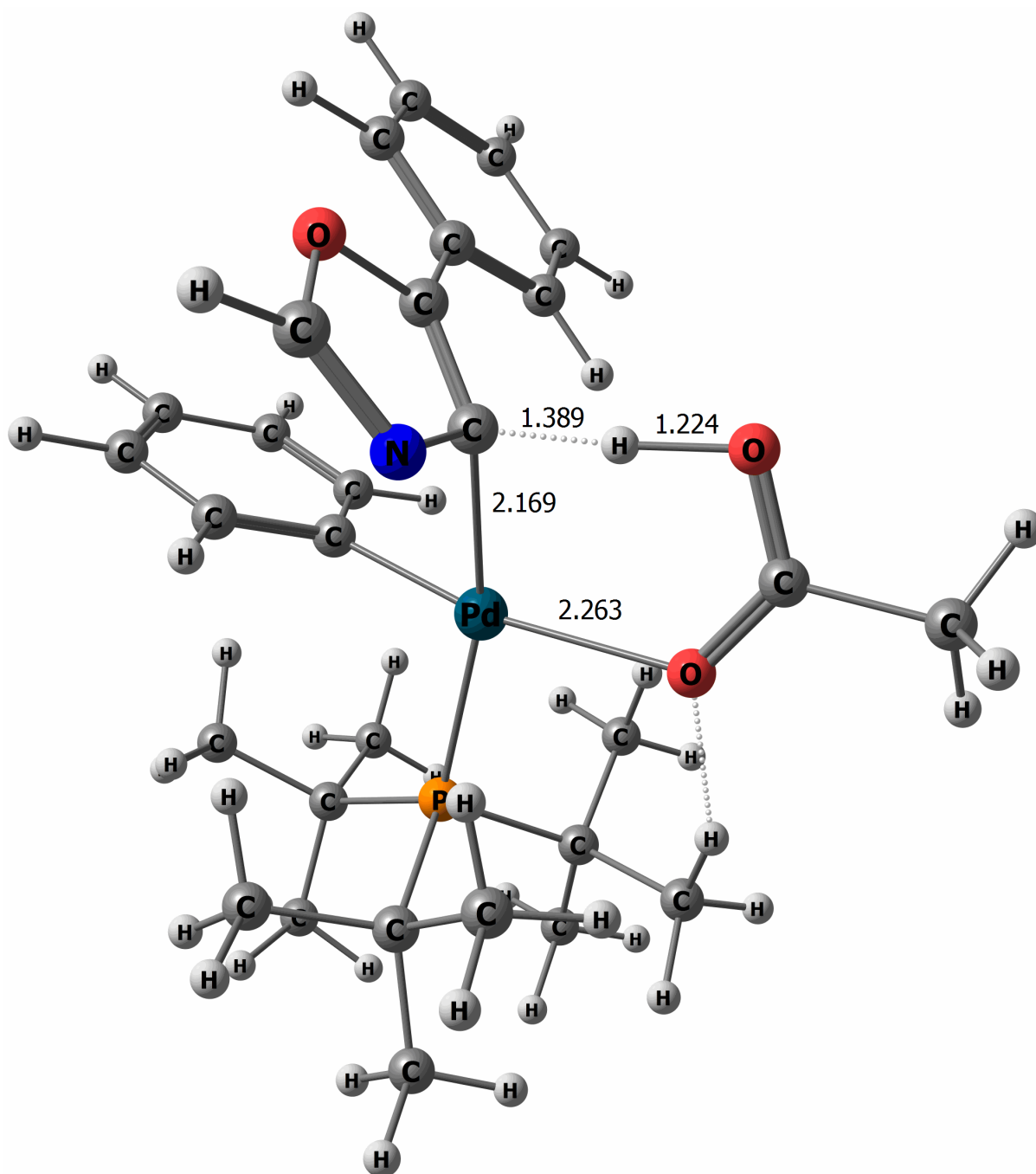


Figure 4-40: Transition State 1 Pathway 5 → 4,5

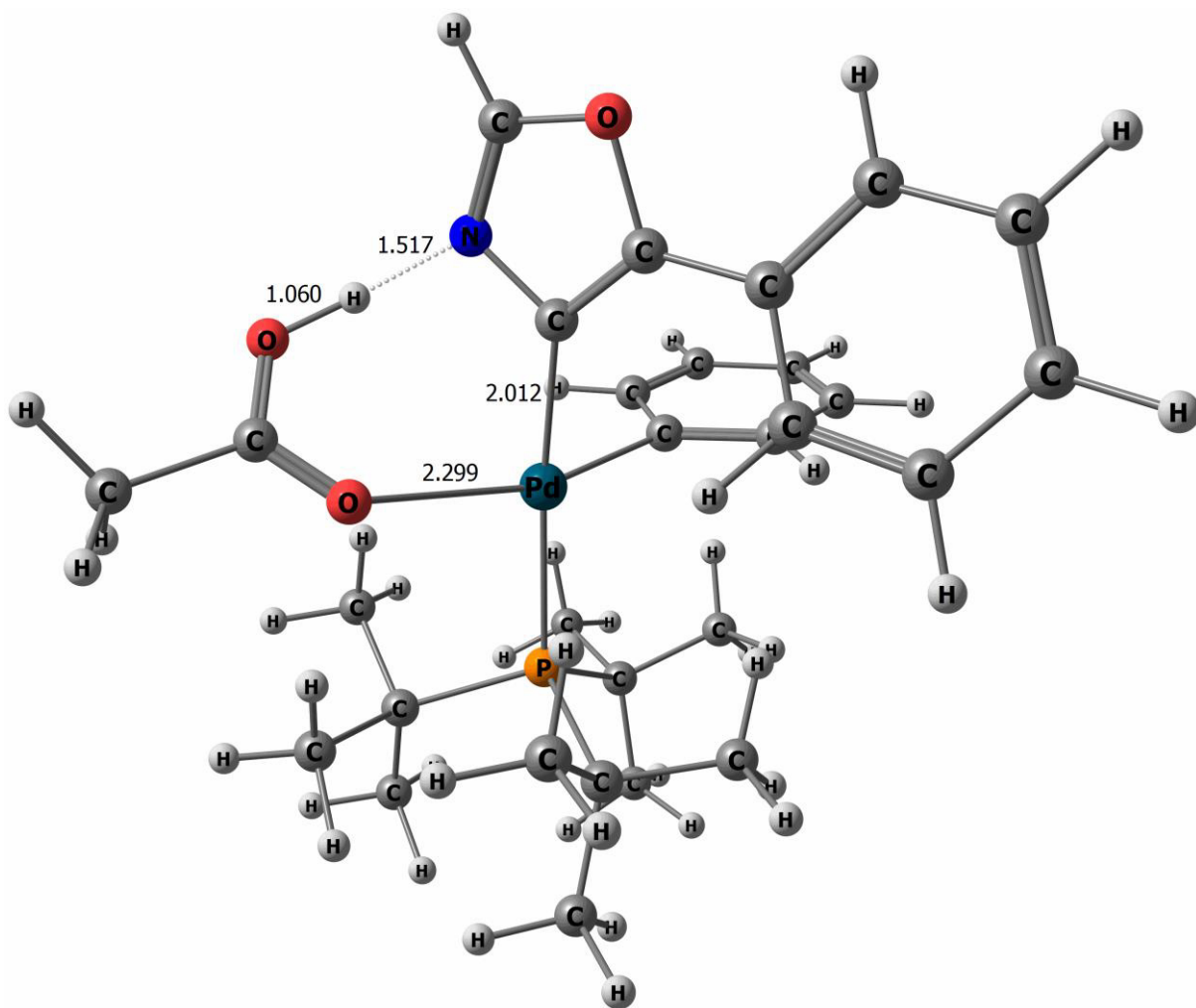


Figure 4-41: Intermediate 2 Pathway 5 → 4,5

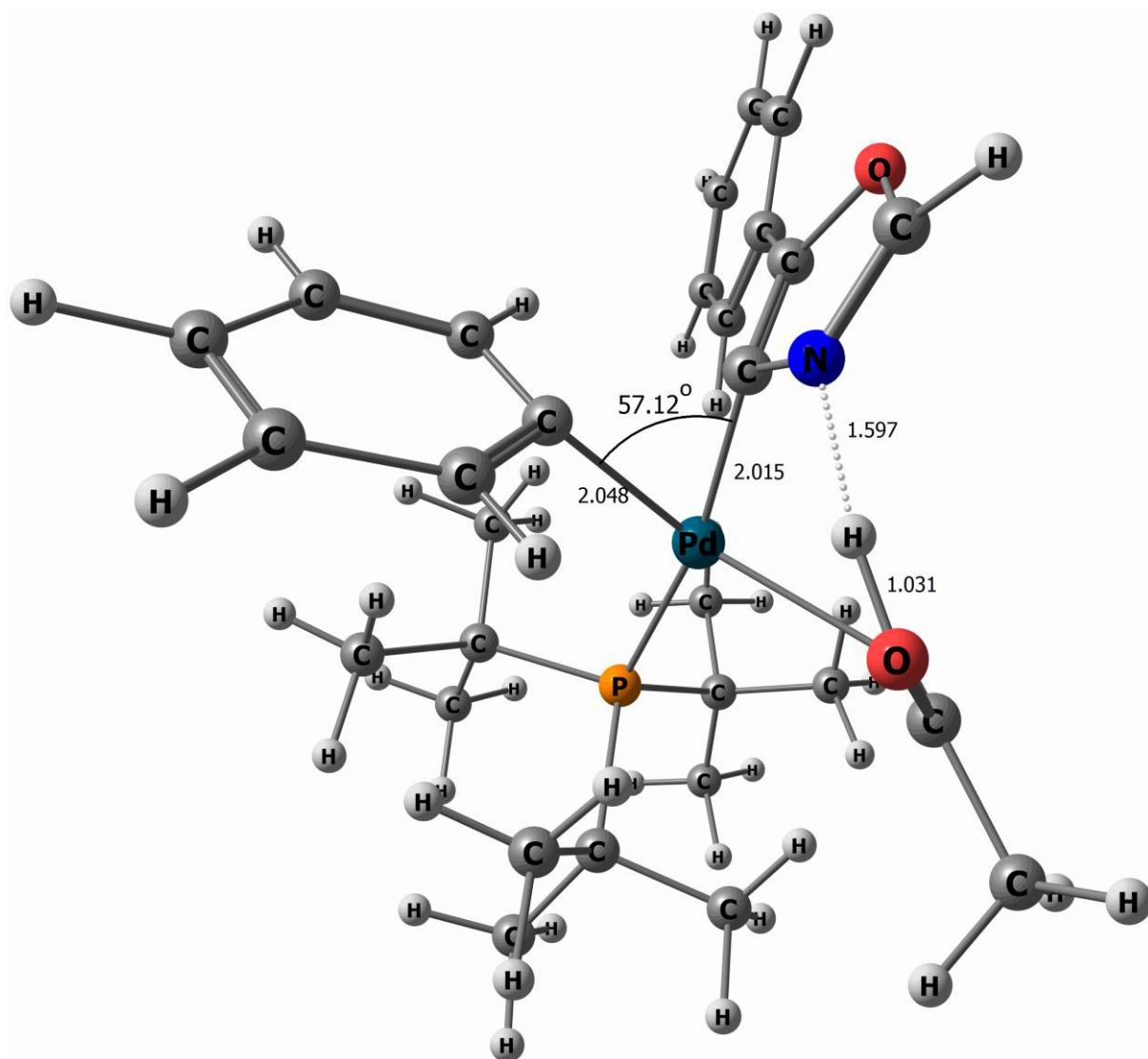


Figure 4-42: Transition State 2 Pathway 5 → 4,5

DSD-PBEP86 Results:

The DSD-PBEP86/TZVP reaction profiles for di-substitution are shown in Figure 4-43 and the reaction Gibbs energies are depicted in Table 4-7. As in Chapter 3, Pd(II) catalyst + 2 x oxazole = 0 kcal.mol⁻¹.

In the case of mono-substitution, the formation of Intermediate 1 across all pathways was endergonic, on average by 12.08 kcal.mol⁻¹. This was attributed to steric hindrance induced by the P^tBu₃ ligand. In the case of di-substitution the formation of Intermediate 1 is also endergonic for all pathways, now on average by a lower 10.35 kcal.mol⁻¹, despite an apparent increase in steric repulsion as the arylated oxazole approaches the catalyst. A similar trend was observed with the DMA ligated pathways, and again this be explained by possible π - π stacking of the two benzene groups, forced closer together than in mono-substitution.

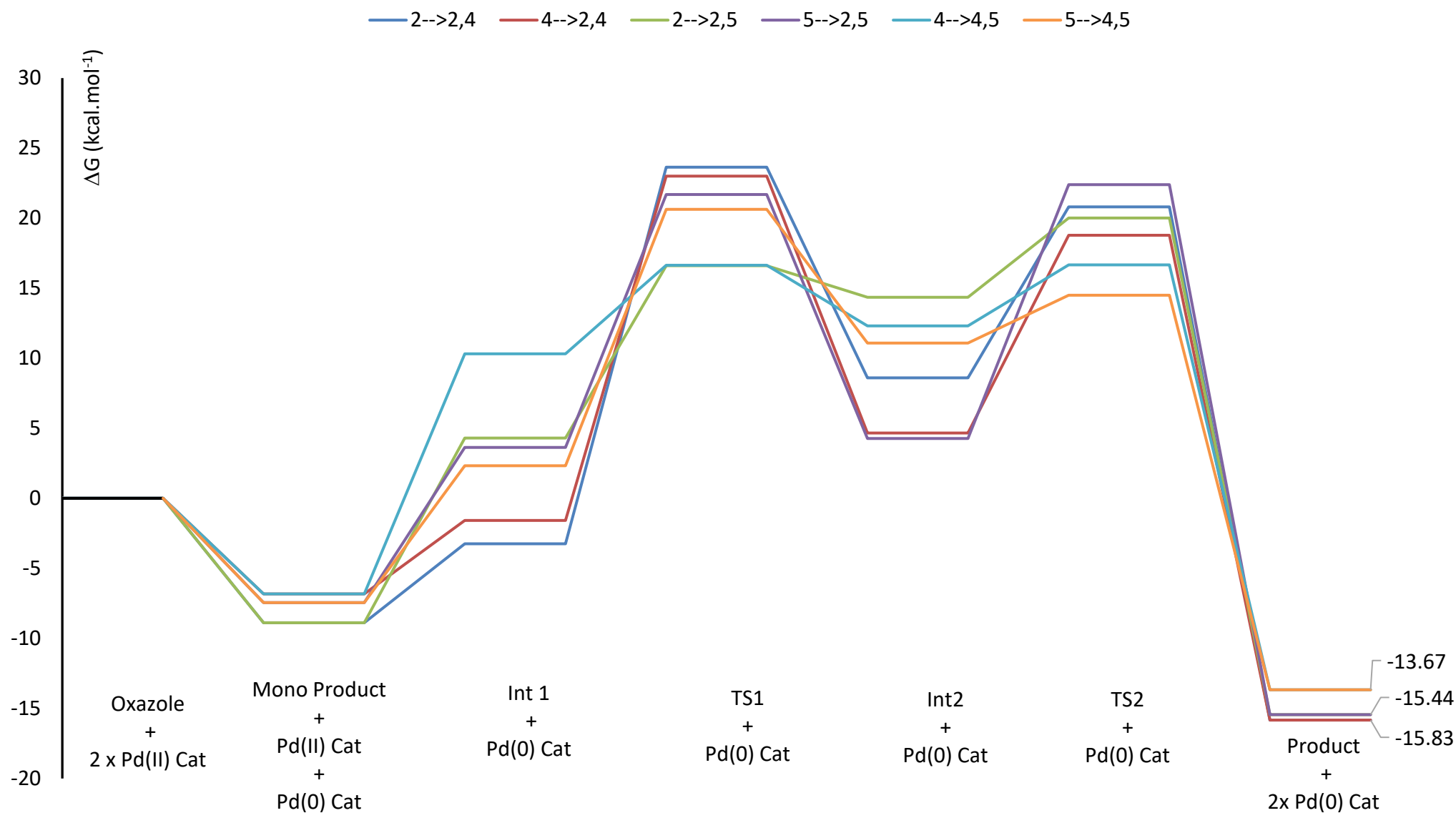


Figure 4-43: Di-substitution Reaction Profiles (DSD-PBEP86/TZVP)

Table 4-7: Di-substitution Reaction Gibbs Energies (DSD-PBEP86/TZVP)

Product 2,4

Pathway 2→2,4	ΔE (kcal.mol ⁻¹)	ΔG_{298K} (kcal.mol ⁻¹)
Product 2 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-14.48	5.65
Int 1 + Pd(0) Catalyst → TS 1 + Pd(0) Catalyst	29.99	26.88
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-19.49	-15.04
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	15.16	12.20
TS 2 + Pd(0) Catalyst → Product 2,4 + 2 x Pd(0) Catalyst	-14.87	-36.63
Pathway 4→2,4		
Product 4 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-16.47	5.25
Int 1 + Pd(0) Catalyst → TS 1 + Pd(0) Catalyst	28.33	24.57
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-22.53	-18.34
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	15.88	14.11
TS 2 + Pd(0) Catalyst → Product 2,4 + 2 x Pd(0) Catalyst	-11.04	-34.60

Product 2,5

Pathway 2→ 2,5	ΔE (kcal.mol ⁻¹)	ΔG_{298K} (kcal.mol ⁻¹)
Product 2 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-6.79	13.18
Int 1 + Pd(0) Catalyst → TS 1 + Pd(0) Catalyst	14.81	12.30
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-6.22	-2.26
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	8.41	5.67
TS 2 + Pd(0) Catalyst → Product 2,5 + 2 x Pd(0) Catalyst	-14.70	-35.45
Pathway 5→ 2,5		
Product 5 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-10.72	11.09
Int 1 + Pd(0) → TS 1 + Pd(0)	20.90	18.04
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-22.53	-17.41
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	17.15	18.13
TS 2 + Pd(0) Catalyst → Product 2,5 + 2 x Pd(0) Catalyst	-15.41	-37.84

Table 4-7 Continued

Product 4,5

Pathway 4 → 4,5	ΔE (kcal.mol ⁻¹)	ΔG_{298K} (kcal.mol ⁻¹)
Product 4 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-4.22	17.13
Int 1 + Pd(0) Catalyst → TS 1 + Pd(0) Catalyst	9.28	6.33
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-6.82	-4.33
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	6.81	4.35
TS 2 + Pd(0) Catalyst → Product 4,5 + 2 x Pd(0) Catalyst	-8.79	-30.33
Total G Input		27.82
Pathway 5 → 4,5		
Product 5 + Pd(0) Cat + Pd(II) Cat → Int 1 + Pd(0) Cat	-12.18	9.78
Int 1 + Pd(0) Catalyst → TS 1 + Pd(0) Catalyst	22.44	18.30
TS 1 + Pd(0) Catalyst → Int 2 + Pd(0) Catalyst	-12.31	-9.55
Int 2 + Pd(0) Catalyst → TS 2 + Pd(0) Catalyst	6.72	3.42
TS 2 + Pd(0) Catalyst → Product 4,5 + 2 x Pd(0) Catalyst	-7.81	-28.17
Total G Input		31.50

Table 4-8 compares the magnitude of CMD vs RE. Table 4-9 compares the magnitude of CMD vs Int 1 formation. It is observed that for the following 3 pathways the RDS is CMD:

Pathway 2 → 2,4: by 14.68 kcal.mol⁻¹

Pathway 4 → 2,4: by 10.46 kcal.mol⁻¹

Pathway 5 → 4,5: by 8.52 kcal.mol⁻¹

For Pathway 5 → 2,5 the difference between CMD and RE is within the assumed error of the method and thus this pathway has no clear RDS and both of these reactions are important. The formation of Int 1 is now an energetically relevant step and is rate determining by the large margin of 10.80 kcal.mol⁻¹ for Pathway 4 → 4,5. The rate of Pathway 2 → 2,5 is governed both by CMD and Int 1 formation as the difference between these two processes is within the assumed error of the method.

Table 4-8: CMD (TS 1) vs Reductive Elimination (TS 2) Transition State Gibbs Energies (DSD-PBEP86/TZVP)

Pathway	$\Delta^\ddagger G_{TS\ 1} - \Delta^\ddagger G_{TS\ 2}$ (kcal.mol ⁻¹)
Pathway 2 → 2,4	14.68
Pathway 4 → 2,4	10.46
Pathway 2 → 2,5	6.64
Pathway 5 → 2,5	-0.08
Pathway 4 → 4,5	1.98
Pathway 5 → 4,5	14.87

Table 4-9: CMD vs Int 1 Gibbs Energies(DSD-PBEP86/TZVP)

Pathway	$\Delta^\ddagger G_{TS\ 1} - \Delta G_{Int\ 1}$ (kcal.mol ⁻¹)
Pathway 2 → 2,4	21.23
Pathway 4 → 2,4	19.33
Pathway 2 → 2,5	-0.88
Pathway 5 → 2,5	6.95
Pathway 4 → 4,5	-10.80
Pathway 5 → 4,5	8.52

Analysing Gibbs energy required for the RDS, as shown in Table 4-10, reveals that Pathway 2 → 2,5 is the most favourable sequence of reactions kinetically. This is by a margin of 3.95 kcal.mol⁻¹ when compared to the next lowest pathway, Pathway 4 → 4,5. Returning again to the experimental product distributions, Table 4-11, it can be seen that Product 2,5 represents 59.3% of the product spectrum and that the only other di-substituted product is Product 4,5; which represents less than 1%. Again, the experimental results are in excellent agreement with the theoretical predictions at the DSD-PBEP86/TZVP level of theory. The two pathways with the highest Gibbs free energy RDS both correspond to the formation of Product 2,4, which was not observed experimentally, further reinforcing the aforementioned point.

Table 4-10: Pathway Favourability in Terms of the RDS: Di-substitution (DSD-PBEP86/TZVP)

Rank	Pathway	RDS	ΔG RDS (kcal.mol ⁻¹)
1 st	Pathway 2 → 2,5	Int 1 formation	13.18
2 nd	Pathway 4 → 4,5	Int 1 formation	17.13
3 rd	Pathway 5 → 2,5	RE (TS 2)	18.13
4 th	Pathway 5 → 4,5	CMD (TS 1)	18.30
5 th	Pathway 4 → 2,4	CMD (TS 1)	24.57
6 th	Pathway 2 → 2,4	CMD (TS 1)	26.88

Table 4-11: Experimental Product Distributions

Entry	Catalyst	Ligand	Product Distributions (%)			
			Product 5	Product 2	Product 2,5	Product 4,5
1	Pd(OAc) ₂	-	65.8	0.3	28.1	3.0
2	Pd(OAc) ₂	1 x P ^t Bu ₃	30.6	3.7	59.3	0.9
3	Pd(OAc) ₂	2 x P ^t Bu ₃	6.3	35.3	49.7	0.1

Pathway 4 → 4,5 is no longer the kinetic product, as it was under phosphine free conditions. This pathway is also expected, as before, to be severely limited in terms of the concentration of Product 4 available to react further. This explains the fact that under conditions of 1 equivalent of P^tBu₃ ligand less than a third of Product 4,5 was produced than under ligand free conditions.

Product 2,5 makes up more than double the product distribution under conditions of 1 equivalent of P^tBu₃ than it does under phosphine free conditions (59.3% vs 28.1%). When contrasting the favourabilities, in terms of the RDS, of the pathways leading to di-substitution under these two conditions (Table 4-12) it can be seen that that Pathway 2 → 2,5 is more favourable under conditions of phosphine ligand, with the RDS being 2.03 kcal.mol⁻¹ lower than under phosphine free conditions.

Table 4-12: Pathway Favourability in Terms of the RDS: P^tBu₃ vs DMA

Pathway	RDS (P ^t Bu ₃)	ΔG RDS (P ^t Bu ₃) /kcal.mol ⁻¹	RDS (DMA)	ΔG RDS (DMA) /kcal.mol ⁻¹	Difference /kcal.mol ⁻¹
Pathway 2 → 2,5	Int 1 formation	13.18	CMD (TS 1)	15.21	-2.03
Pathway 4 → 4,5	Int 1 formation	17.13	CMD (TS 1)	14.43	2.7
Pathway 5 → 2,5	RE (TS 2)	18.13	RE (TS 2)	20.66	-2.53
Pathway 5 → 4,5	CMD (TS 1)	18.30	CMD (TS 1)	18.84	-0.54
Pathway 4 → 2,4	CMD (TS 1)	24.57	CMD (TS 1)	26.49	-1.92
Pathway 2 → 2,4	CMD (TS 1)	26.88	RE (TS 2)	24.32	2.56

Considering the factors responsible for the favorability of Pathway 2 → 2,5, it can be seen in Table 4-13, that it benefits from having an exceptionally low ΔG[‡]_{TS1} value. ΔG[‡]_{TS2} is also lower

than most other pathways. Table 4-14 reveals that this pathway suffers from the highest $\Delta G_{\text{Int } 1}$ requirement. The same trend is observed for Pathway 4 $\rightarrow$ 4,5.

Table 4-13: Pathway Favourability in Terms of TS 1 (DSD-PBEP86/TZVP)

Rank	Pathway	$\Delta G^{\ddagger}_{\text{TS1}}$ (kcal.mol ⁻¹)
1 st	Pathway 4 $\rightarrow$ 4,5	6.33
2 nd	Pathway 2 $\rightarrow$ 2,5	12.30
3 rd	Pathway 5 $\rightarrow$ 2,5	18.04
4 th	Pathway 5 $\rightarrow$ 4,5	18.30
5 th	Pathway 4 $\rightarrow$ 2,4	24.57
6 th	Pathway 2 $\rightarrow$ 2,4	26.88

Table 4-14: Pathway Favourability in Terms of TS 2 (DSD-PBEP86/TZVP)

Rank	Pathway	$\Delta G^{\ddagger}_{\text{TS2}}$ (kcal.mol ⁻¹)
1 st	Pathway 5 $\rightarrow$ 4,5	3.42
2 nd	Pathway 4 $\rightarrow$ 4,5	4.35
3 rd	Pathway 2 $\rightarrow$ 2,5	5.67
4 th	Pathway 2 $\rightarrow$ 2,4	12.20
5 th	Pathway 4 $\rightarrow$ 2,4	14.11
6 th	Pathway 5 $\rightarrow$ 2,5	18.13

Table 4-15: Pathway Favourability in Terms of Int 1 (DSD-PBEP86/TZVP)

Rank	Pathway	$\Delta G_{\text{Int } 1}$ (kcal.mol ⁻¹)
1 st	Pathway 4 $\rightarrow$ 2,4	5.25
2 nd	Pathway 2 $\rightarrow$ 2,4	5.65
3 rd	Pathway 5 $\rightarrow$ 4,5	9.78
4 th	Pathway 5 $\rightarrow$ 2,5	11.09
5 th	Pathway 2 $\rightarrow$ 2,5	13.18
6 th	Pathway 4 $\rightarrow$ 4,5	17.13

In fact, the rate of both of the kinetic products of di-substitution is limited by the formation of Intermediate 1 (with Pathway 4 $\rightarrow$ 4,5 also being limited by CMD). This implies that research into lowering $\Delta G_{\text{Int } 1}$ is essential in speeding up the formation of Product 2,5 as well as Product 4,5.

Product 2,4 is the thermodynamic product. The absence of any Product 2,4 observed experimentally indicates that no equilibrium was established among the di-substituted pathways. This is consistent with the conclusions drawn from di-substitution under phosphine free conditions (Chapter 3).

In summary, the following conclusions can be drawn:

Mono-substitution:

- The RDS for all pathways, except Pathway =5, is CMD.
- For Pathway =5 both Int 1 formation and CMD dictate the rate.
- Pathway =5 requires the least Gibbs energy for the RDS, 4.64 kcal.mol⁻¹ less than any other pathway, and Product 5 is thus the kinetic product. This is in agreement with experimental results that found Product 5 makes up the vast majority of mono-substituted product observed.
- It is thus highly probable that under conditions of 1 equivalent of P^tBu₃, the reactions leading to mono-substitution also proceed via a CMD mechanism.
- Product 2 is the thermodynamic product and, although kinetically hindered, the large negative $\Delta_r G$ value of -8.88 kcal.mol⁻¹ will ensure that significant amounts are produced during equilibrium. This correlates well with the observation that Product 2 constituted 3.7% of the experimental product distribution, and the only other mono-substituted product observed experimentally.
- An equilibrium is thus established among the mono-substituted pathways.

Di-substitution:

- The rate determining step for 3 of the 6 di-substitution pathways is exclusively concerted metallation deprotonation, by the margins presented in Table 4-8 and Table 4-9.
- For Pathway 2 → 2,5 both the formation of Intermediate 1 as well as CMD dictate the rate.
- For Pathway 4 → 4,5 the formation of Intermediate 1 is rate determining by a large margin of 10.80 kcal.mol⁻¹.
- For Pathway 5 → 2,5 both CMD as well as RE dictate the rate.
- Therefore, no blanket statements can be made regarding the RDS of these reactions.
- Pathway 2 → 2,5 requires the lowest Gibbs free energy for the RDS, by 3.95 kcal.mol⁻¹, and Product 2,5 is thus the kinetic di-substituted product. This is in excellent agreement with the experimental results, where Product 2,5 constitutes 98.5% of the di-substituted product portfolio and 59.3% of the overall product portfolio under conditions of phosphine ligand.

- Pathway 4 → 4,5 requires the second lowest Gibbs energy for the RDS, explaining the 0.9% of Product 2,5 observed experimentally.
- It is thus highly probable that under conditions of 1 equivalent of P^tBu_3 ligand, both mono- and di-substitution reactions proceed via a CMD mechanism.
- Equilibrium was not established among the pathways leading to di-substituted.

References

1. Y. Tan and J. F. Hartwig, *Journal of the American Chemical Society*, 2011, 133, 3308-3311.
2. S. I. Gorelsky, D. Lapointe and K. Fagnou, *Journal of the American Chemical Society*, 2008, 130, 10848-10849.

5. Oxidative Addition

Contents

5.1 Introduction	272
5.2 Conditions of One Equivalent of P^tBu_3	274
Formation of the Active Pd(0) Catalyst:	274
The Oxidative Addition Reaction Pathway:	279
TPSSH Results:	283
DLPNO-CCSD(T)/ DSD-PBEP86 Results:	287
5.3 Phosphine Free Conditions	297
Formation of the Active Pd(0) Catalyst:	297
The Reaction Pathway:	299
DSD-PBEP86 Results:	311
References	318

5.1 Introduction

This chapter is concerned with the formation of the active Pd(II) catalysts seen in Chapters 3 and Chapter 4. The formation of the active Pd(II) catalyst does not determine the product distributions, however it is possibly rate limiting and thus an investigation into this step is essential in order to complete the catalytic cycle, shown again in Figure 5-1. This chapter will detail the transformations shown in blue in Figure 5-1. The first part will focus on the formation of the Pd(II) catalyst under conditions of 1 equivalent of P^tBu_3 ligand, to ultimately produce the molecule shown in Figure 5-2 and modelled as the active Pd(II) catalyst in Chapter 4. Due to the popularity of this ligand in direct arylation reactions, oxidative addition (OA) has been extensively studied in the presence of P^tBu_3 and this is thus an appropriate place to initiate modelling. The second part of the chapter will focus on the formation of the Pd(II) catalyst under phosphine free conditions to produce the molecule shown in Figure 5-3 and modelled as the active Pd(II) catalyst in Chapter 3.

Throughout this chapter, reference will be made to two active catalysts, the active Pd(0) catalyst and the active Pd(II) catalyst. This is predominantly to distinguish the Pd(II) pre-catalyst, discussed below, from the active Pd(II) catalyst as well as the protonated Pd(0) catalyst from the active Pd(0) catalyst. It should be emphasised that the active Pd(0) catalyst is the molecule that reacts with the aryl halide and actually undergoes OA to produce the active Pd(II) catalyst, which then catalyses the CMD reaction.

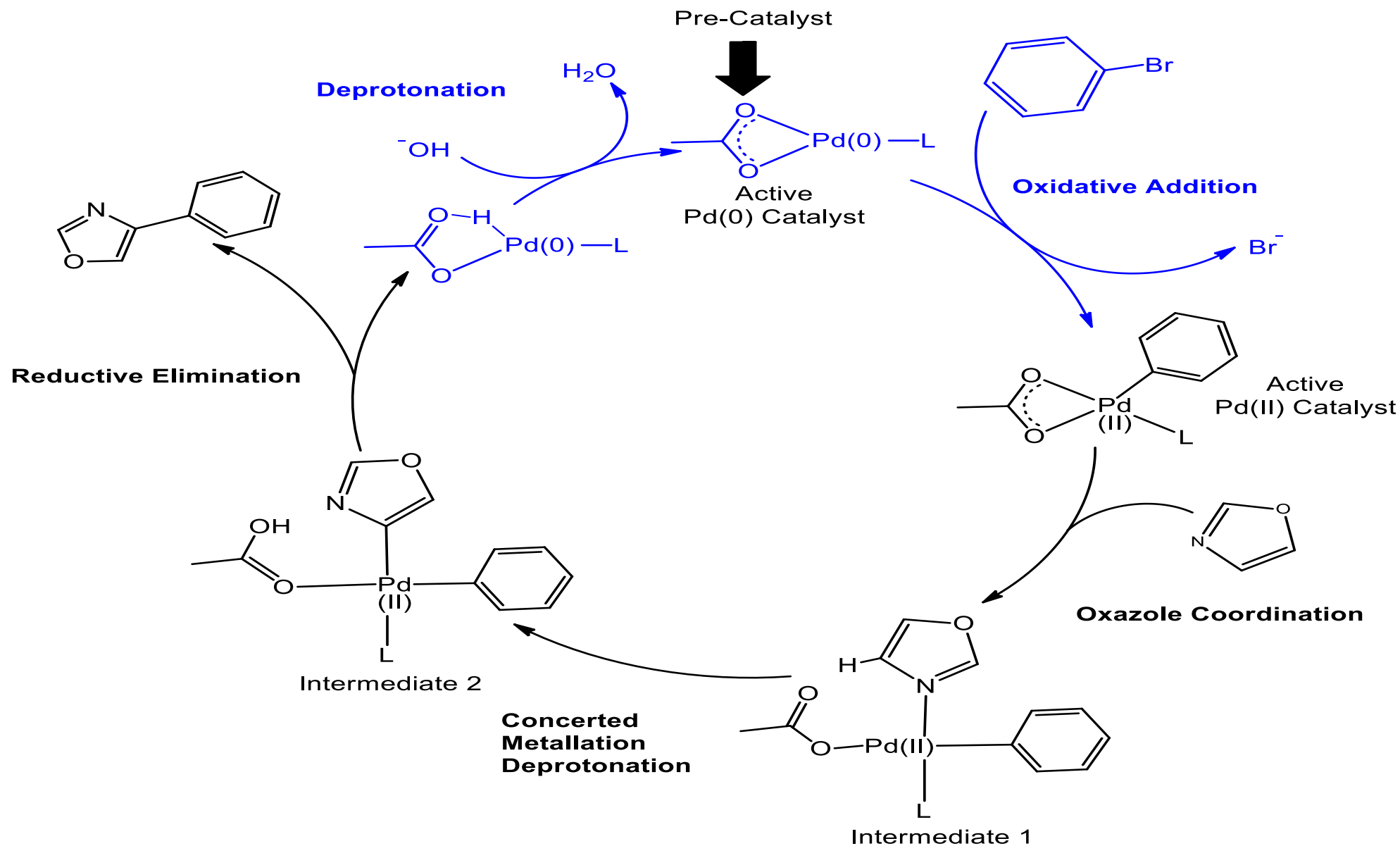


Figure 5-1: The Catalytic Cycle – Oxidative Addition and Related Steps Highlighted in Blue

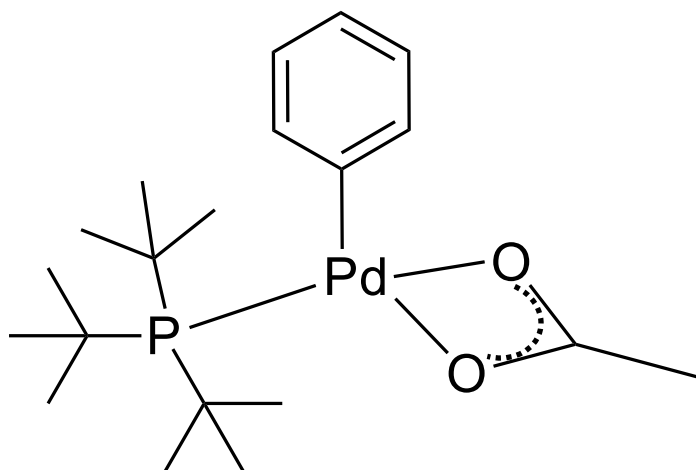


Figure 5-2: Active Pd(II) Catalyst – 1 Mole Equivalent of P^tBu_3

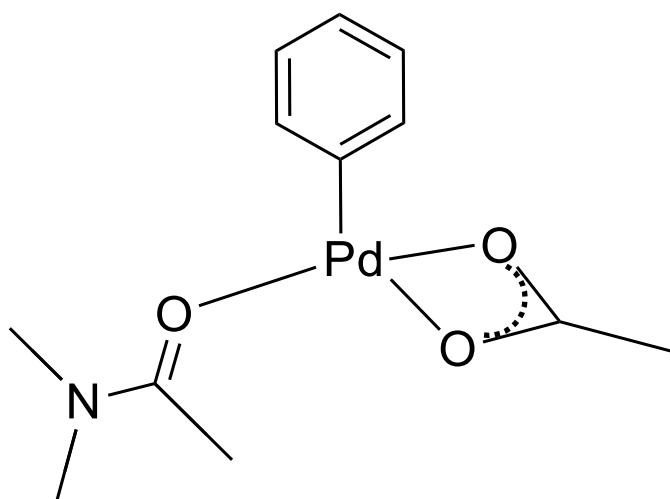


Figure 5-3: Active Pd(II) Catalyst – Phosphine Free Conditions

5.2 Conditions of One Equivalent of P^tBu_3

Formation of the Active Pd(0) Catalyst:

The formation of an active Pd(0) catalyst can be achieved in a number of ways. These include using a source of Pd(0) directly, such as $Pd_2(dba)_3$ ¹ or $Pd(PPh_3)_4$.² The *in-situ* reduction of a Pd(II) pre-catalyst is another possibility. This method avoids problems of Pd(0) instability with respect to air and moisture. A popular approach is to use an allyl chloride palladium dimer, such as $[(allyl)PdCl]_2$.³ Advanced one component Pd(II) pre-catalysts, which already contain ligands bound to the metal atom in the correct stoichiometry, and can operate efficiently at temperatures as low as -40 °C have been reported.⁴ The use of such one component Pd(II) pre-catalysts is however, prohibitively expensive. Probably the most common method, at least in

direct arylation reactions, is an *in-situ* reduction of a Pd(II) salt, normally Pd(OAc)₂. Conveniently, this reduction can take place via a corresponding oxidation of commonly employed tertiary phosphine ligands, in the presence of H₂O or base.⁵ However, tertiary amine ligands have been shown to be equally effective.⁶

The reactions that concern this project made use of a Pd(OAc)₂ pre-catalyst. The mechanism for the formation of the active Pd(0) catalyst from Pd(OAc)₂, when used in conjunction with tri-tertiary phosphine ligands, has been extensively studied.^{5, 7, 8} The generally accepted mechanism in polar solvents, under basic conditions, is shown in Figure 5-4. After initial coordination of two phosphine ligands, an inner-sphere intramolecular reduction of palladium takes place, with a corresponding oxidation of the phosphorous atom of one ligand. In this process, the formal oxidation state of palladium changes from 2+ to 0, while that of the phosphorous atom goes from 3+ to 5+. This is the rate-determining step for the formation of the active Pd(0) catalyst. The result is an anionic Pd(0) complex, [Pd(PR₃)(OAc)]⁻¹, as well as a phosphonium cationic intermediate. This phosphonium cation is then converted to the corresponding phosphine oxide via reaction with a residual water molecule or hydroxide ion, as illustrated in Figure 5-5 for basic conditions. Acetic acid is formed along with the phosphine oxide. Although, not shown in the figure below, under basic conditions this acetic acid will likely be deprotonated rapidly to form an acetate anion and water.

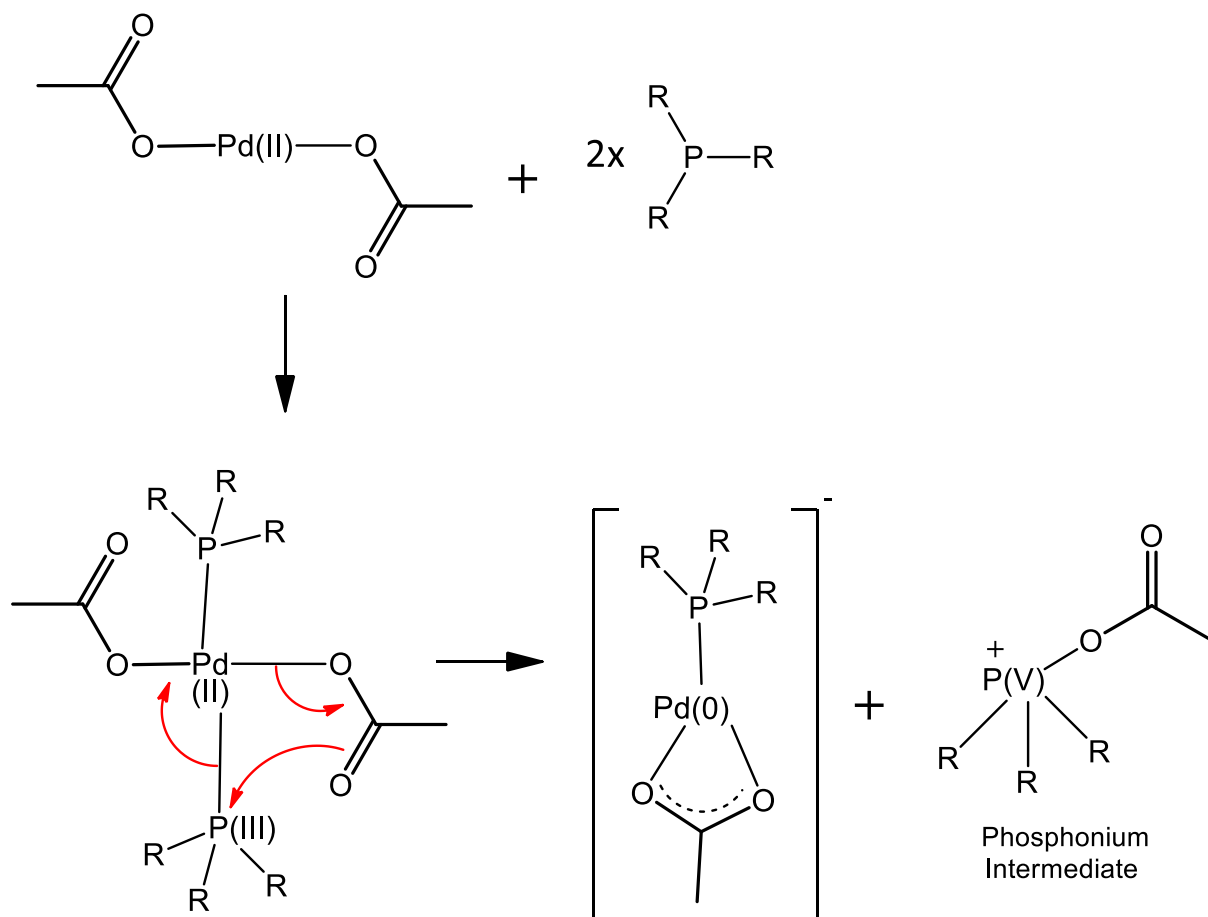


Figure 5-4: Pd(0) Formation from Pd(OAc)₂ Pre-catalyst and Tertiary Phosphine Ligand

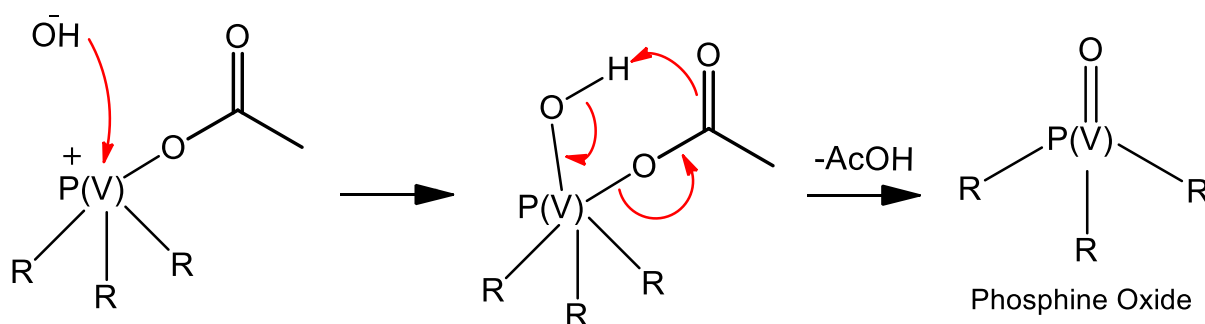


Figure 5-5: The Formation of Phosphine Oxide from Phosphonium Cation via OH⁻

Under conditions of excess phosphine ligand, $[\text{Pd}(\text{PR}_3)(\text{OAc})]^{-1}$ is complexed by an additional two equivalents of ligand to form $[\text{Pd}(\text{PR}_3)_3(\text{OAc})]^{-1}$.⁸ However, the experiments that involve the CMD mechanism in this project used a 1:1 mole ratio of Pd(OAc)₂ and P^tBu₃. As one mole of P^tBu₃ is oxidized per Pd(0) atom formed, there is not expected to be sufficient concentrations of P^tBu₃ in solution for such higher ligated complexes to form. Furthermore, it has been established, from both experimental and computational data, that the species that undergoes OA when the ligand is P^tBu₃ is mono-ligated in terms of phosphine.⁹⁻¹¹ The number of ligands bound to the species that undergoes OA is however dependent, not only on the

nature of the ligand, but also on the aryl halide used.¹² More reactive aryl iodides are able to undergo OA with saturated $\text{Pd}(\text{PR}_3)_2$ complexes. Certain experimental results suggest that OA takes place to a mixture of mono- and bis-ligated phosphine complexes when bromobenzene is the aryl halide.¹² However, computational studies confirm that only OA to a monophosphine complex is energetically feasible.⁹ Therefore, the active $\text{Pd}(0)$ catalyst that will be considered is $[\text{Pd}(\text{OAc})(\text{P}^t\text{Bu}_3)]^{-1}$, illustrated in Figure 5-6. The optimized structure of this molecule is depicted in Figure 5-7.

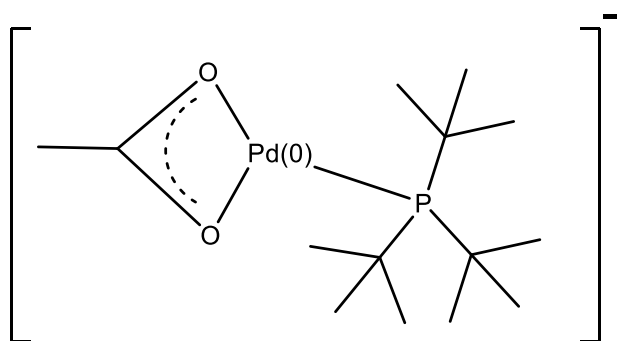


Figure 5-6: Active Pd(0) Catalyst- Phosphine Conditions

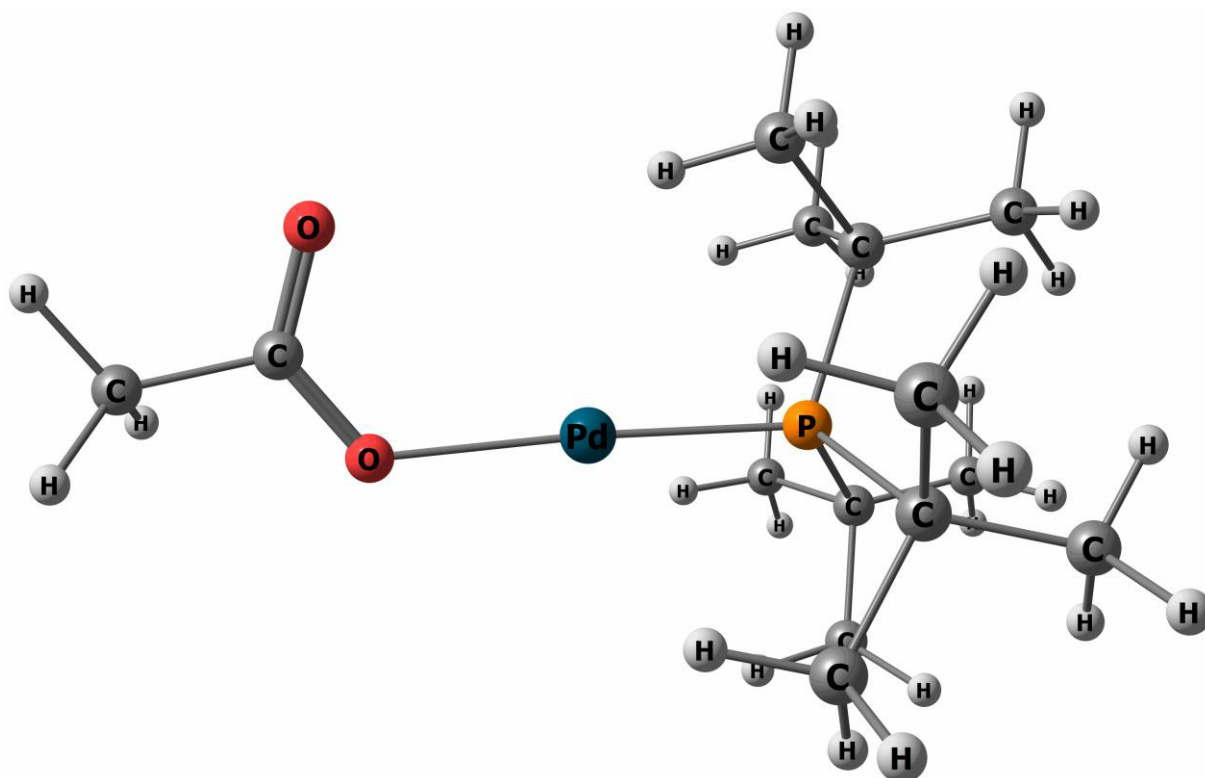


Figure 5-7: Active Pd(0) Catalyst

As ionic species were considered in this chapter, the use of a minimally augmented basis set with added diffuse functions, namely ma-Def2-SVP,¹³ was deemed necessary for all geometry optimisations and frequency calculations. Details of this basis set as well as why its use was

deemed necessary can be found in Chapter 2, Section 2.1.8. For the sake of succinctness, the ma-Def2-SVP basis set will simply be referred to as ma-SVP.

In reality, there is a large negative $\Delta G_{\text{solvation}}$ for ions, especially in polar solvents. Gas phase calculations are particularly unsuited for modelling such species, often severely overestimating their energies.¹⁴ Certain reactions in this chapter consider charged species recombining to form neutral molecules, and thus no favourable cancellation of errors is possible. For this reason, all geometry optimisations, frequency and single point energy calculations were performed in a COSMO continuum solvation model¹⁵ with DMA as the solvent. For details of the implementation of COSMO see Chapter 2- Section 2.5. In summary, all geometry optimizations and frequency calculations in this chapter were performed at the COSMO TPSSH/ma-Def2-SVP level of theory. DH and DLPNO-CCSD(T) calculations were also run in conjunction with the COSMO solvation model. However, minimally augmented basis sets were not deemed necessary for single point energy calculations, for reasons revealed below.

It should be noted that in the literature there is no general consensus as to whether OA takes place to neutral or anionic palladium complexes, particularly when a source of Pd(0) is used directly as the catalyst. Computationally, OA transition states can be located for both pathways, for certain systems.¹⁶ It has been shown that both neutral and anionic pathways correctly predict the experimentally observed reactivities of OA to a range of para-substituted aryl chlorides, with the barrier being lower for anionic species.¹⁷ Furthermore, multiple anionic and neutral pathways were considered in the OA of palladium complexes to acetic anhydride; again the anionic pathway was calculated to be the most favourable alternative.¹⁸ Regarding the relevance of the previous reference, it should be emphasized that the similarities between the OA of palladium complexes to aryl halides and carboxylic anhydrides has been well documented.¹⁹⁻²¹ Although reasonable results have been achieved modelling OA to neutral Pd(P^tBu₃) complexes,²² considering the above it is highly unlikely that such complexes will exist in solution when a Pd(OAc)₂ pre-catalyst is employed. Additionally, even when Pd(0) complexes are used directly as the active catalyst, only by invoking anionic transition states can the correct region-selectivity be predicted in certain systems.²³ Furthermore, it is known that the addition of anionic species accelerates the rate of OA reactions.^{24, 25}

The formation of the active Pd(0) catalyst from the pre-catalyst is assumed to take place immediately when the reacting components are brought together, with no significant reaction barriers. As emphasised in Figure 5-1, this process occurs only once and is not part of the catalytic cycle, with reductive elimination and subsequent deprotonation regenerating the active Pd(0) catalyst. In contrast, the catalytic cycle is expected to occur approximately a hundred times, with typical Pd(OAc)₂ loadings of around 1%. Therefore, it was decided to initiate modelling from the active Pd(0) catalyst and not consider its formation, as this process occurs only once and is assumed to be energetically irrelevant.

The Oxidative Addition Reaction Pathway:

It is generally accepted that when the Pd(0) catalyst and aryl halide come into contact, a reactive “pre-complex” is formed in which the aryl halide is coordinated to palladium via the aromatic bond between the ipso and ortho carbons (Figure 5-8).^{17, 22, 26} The aryl halide considered in this work is bromobenzene. The optimized structure of the pre-complex is depicted in Figure 5-9.

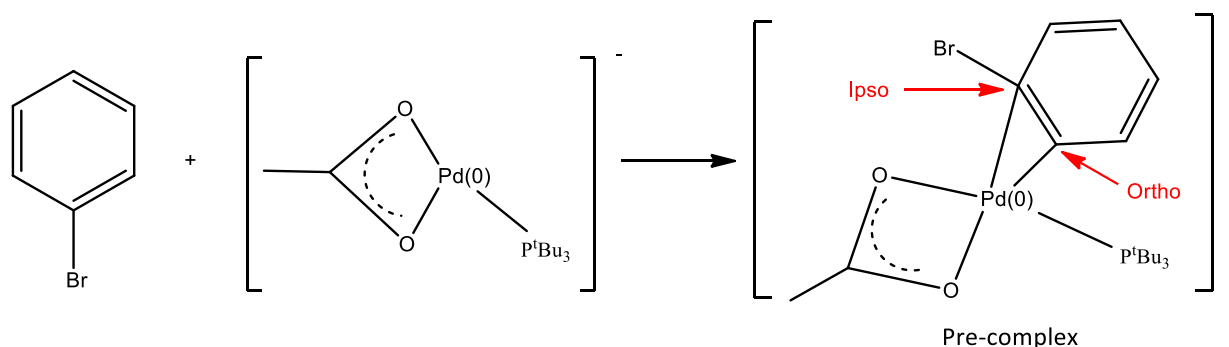


Figure 5-8: Formation of the Reactive Pre-complex

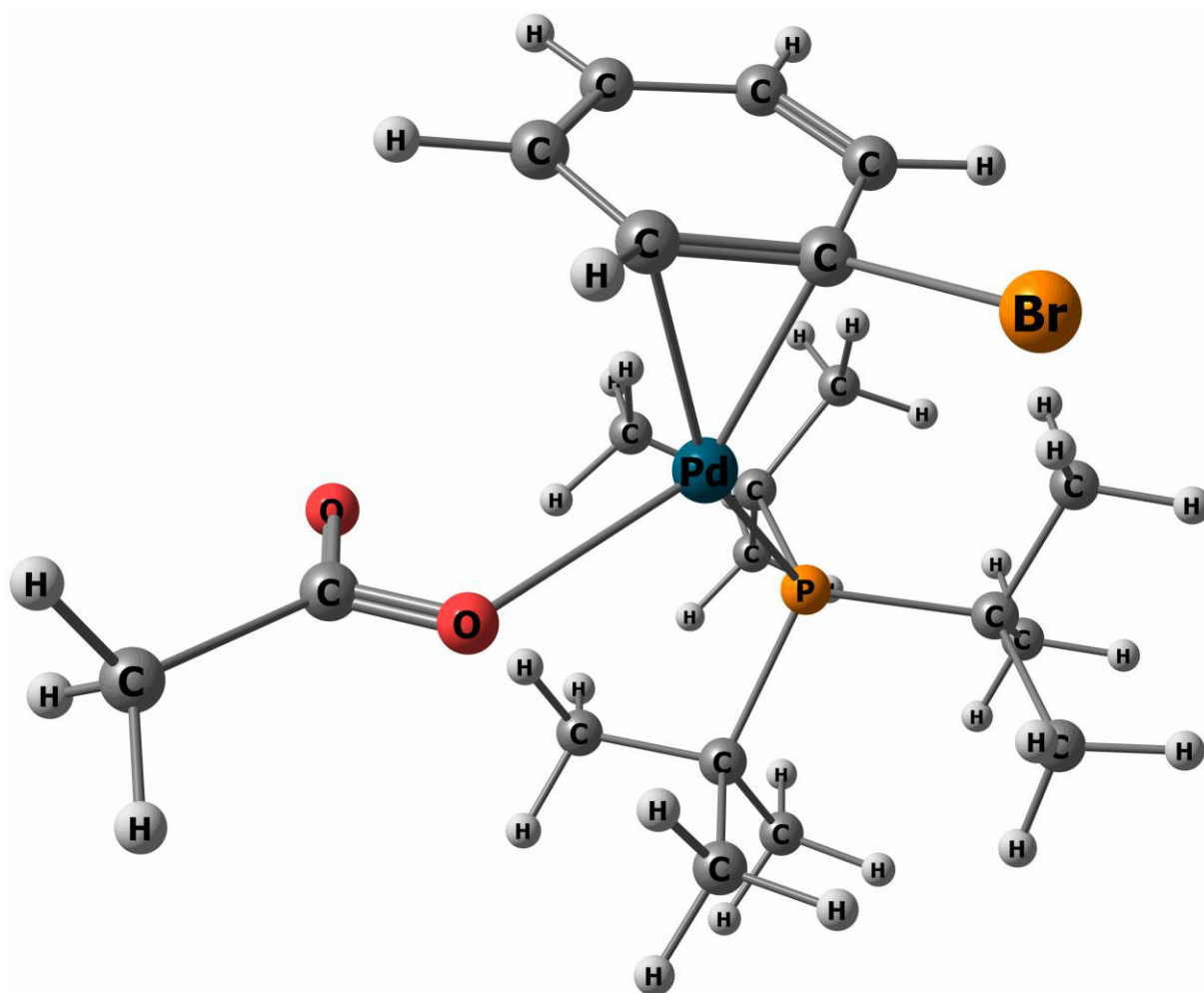


Figure 5-9: Pre-complex

The highly reactive pre-complex then passes through a characteristic oxidative addition transition state in which the imaginary vibrational mode corresponds to the cleaving of the Br–Ar bond. This atomic displacement is indicated by arrows in Figure 5-10. The optimized structure of the TS is portrayed in Figure 5-11. This structure has an imaginary frequency of -138.18 cm^{-1} .

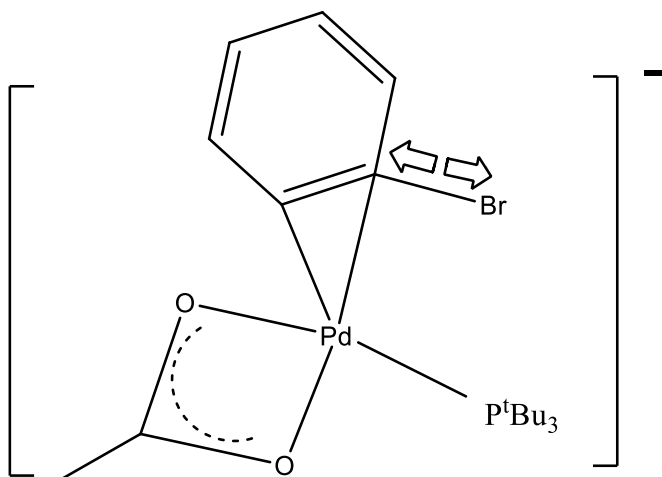


Figure 5-10: Atomic Displacement Associated with the Oxidative Addition TS

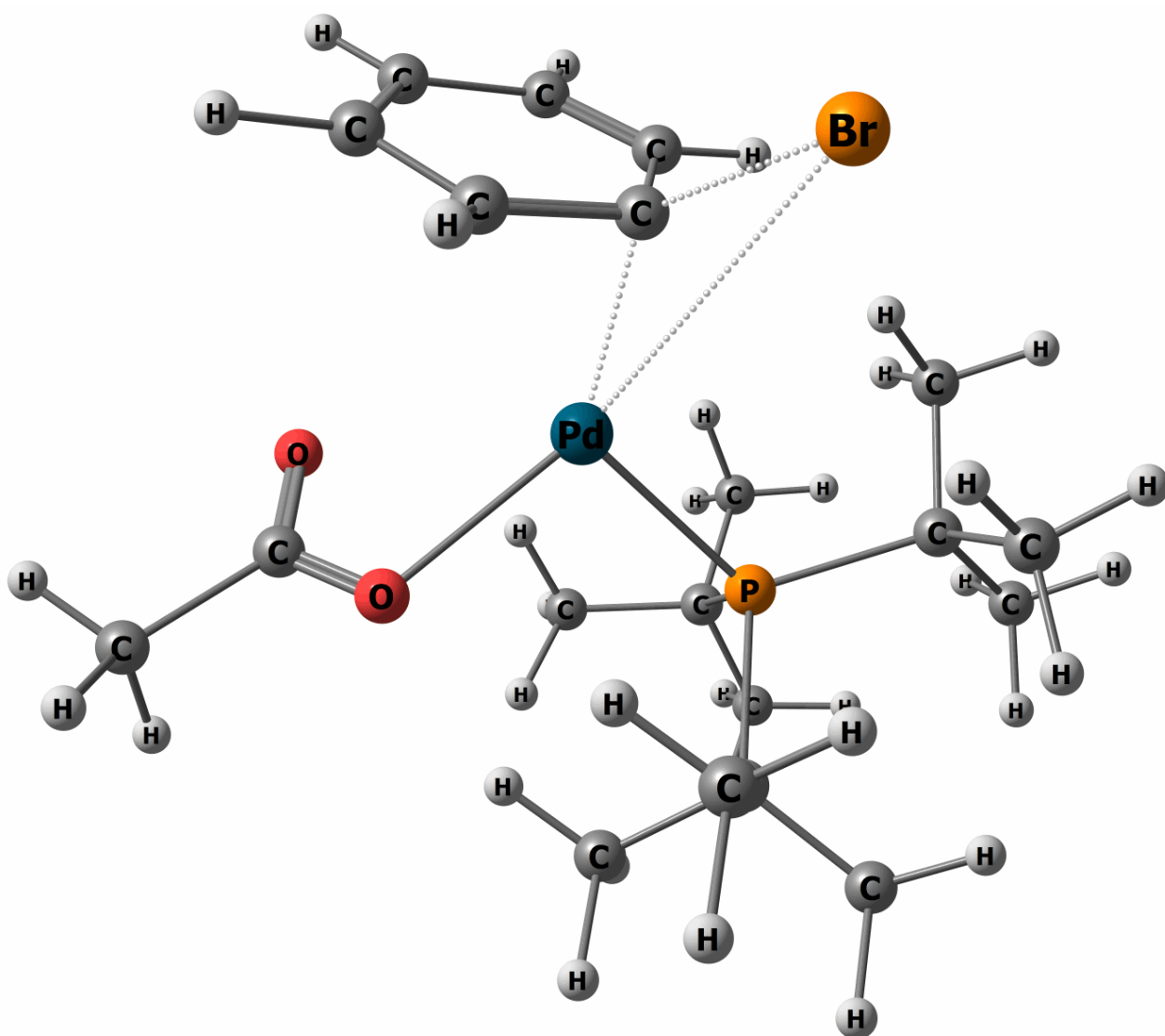
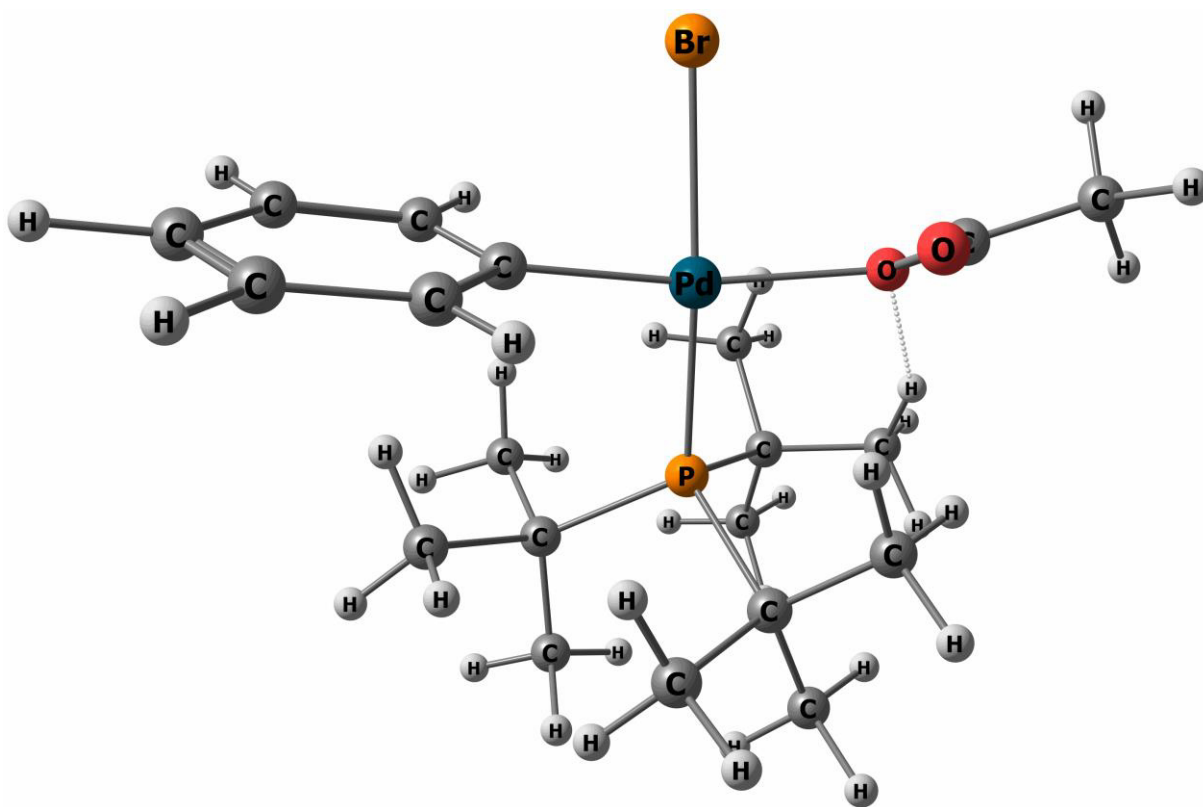


Figure 5-11: Oxidative Addition Transition State

In the process of the aryl halide bond cleaving, the formal oxidation state of palladium changes from 0 to 2+. Thus, palladium is oxidized, hence *oxidative* addition. During this process, there

is a corresponding reduction of the carbon atom bound to bromine as well as the bromine atom, both from a formal oxidation state of 0 to -1. The optimized structure of the resultant product complex is illustrated in Figure 5-12 and has approximate square planar geometry. All that is necessary to form the active Pd(II) catalyst (Figure 5-14), is for a bromine anion to simply dissociate from the product complex. However, as will be elaborated below, other lower energy possibilities exist. Before these are considered, an analysis of the results is necessary.



[Figure 5-12: Product Complex](#)

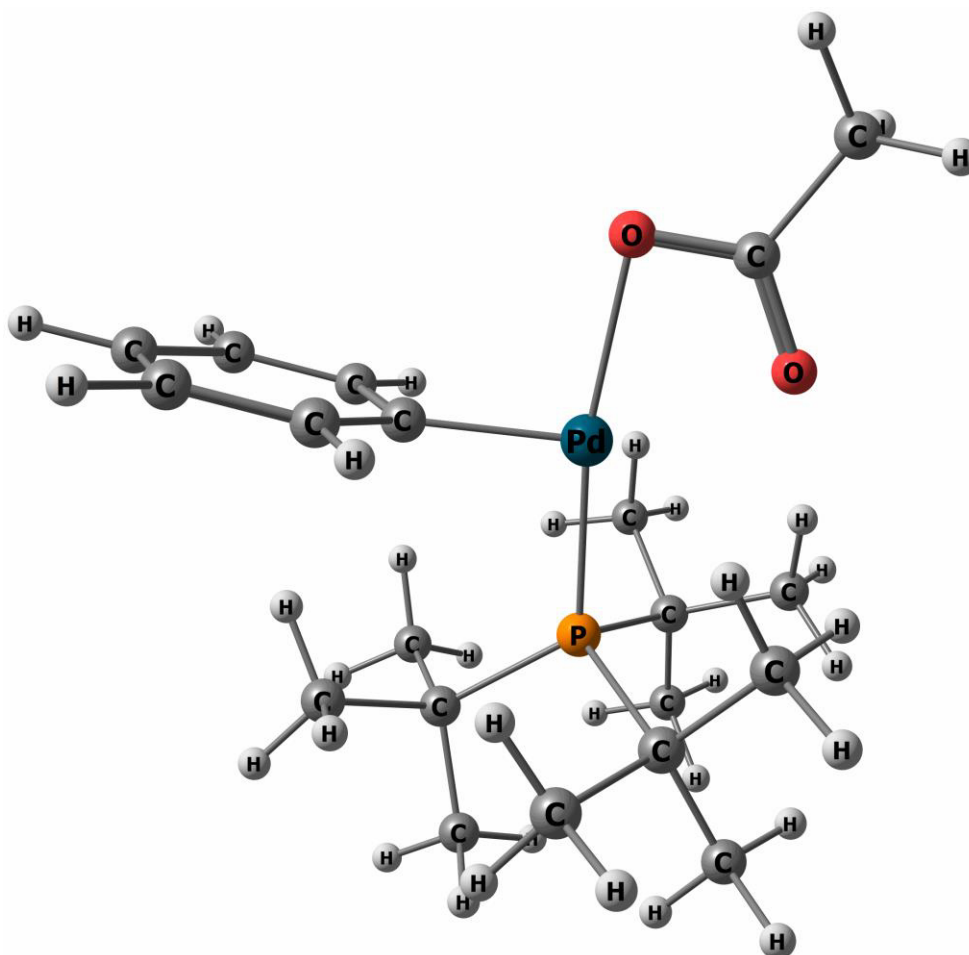


Figure 5-13: Active Pd(II) Catalyst - Phosphine Conditions

TPSSH Results:

Initially, it was imperative to gauge the effects of a minimally augmented basis set as well as a COSMO solvation model, in order to determine if these additional levels of theory were necessary for the system at hand.

Firstly, gas phase geometry optimisations and frequency calculations at the TPSSH/SVP level, as well as single point TPSSH/ma-SVP calculations based on these geometries, were run. As can be seen in Table 5-1, the effect of a ma basis set is negligible for all reactions considered, except for the dissociation of bromine from the product complex to form the active Pd(II) catalyst (Figure 5-14). As in previous chapters, entries corresponding to a TS are highlighted in blue.

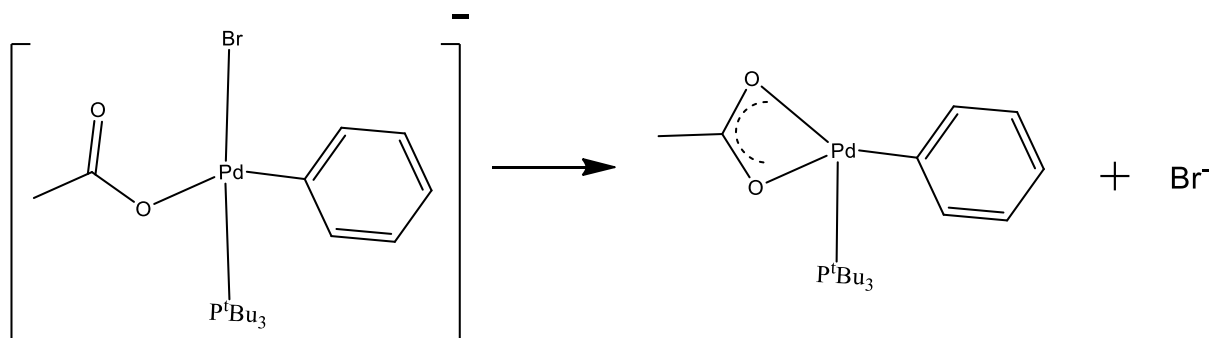


Figure 5-14: Dissociation of Bromine to Form the Active Pd(II) Catalyst

For this reaction, ΔG decreases by a significant $13.14 \text{ kcal.mol}^{-1}$, when moving from SVP to ma-SVP. This illustrates the importance of using a basis set with sufficient diffuse functions when modelling small anions.

However, by far the most dramatic effects are seen when the COSMO solvation model is utilized. Full geometry and frequency calculations at COSMO/ma-SVP level of theory were performed. The formation of the pre-complex now becomes endergonic by a magnitude similar to which it was predicted to be exergonic in the gas phase. The transition state barrier is decreased by $3.50 \text{ kcal.mol}^{-1}$, while the product complex is predicted to be significantly more stable ($7.70 \text{ kcal.mol}^{-1}$). The most striking difference manifests itself in the dissociation of bromine from the product complex to form the active Pd(II) catalyst. Due to the instability of ions in the gas phase, this process was initially predicted to require $14.96 \text{ kcal.mol}^{-1}$ of Gibbs energy. Upon application of a COSMO model, this dissociation is now revealed to be spontaneous by $6.16 \text{ kcal.mol}^{-1}$.

If one considers the reaction conditions (Chapter 1, Section 1.3.2), a KOH base was used and thus there is expected to be a significant concentration of K^+ ions in circulation in solution. It is possible that the dissociation of the bromine anion would be accelerated by the formation of potassium bromide. This reaction is illustrated in Figure 5-15. As can be seen in the last entry of Table 5-1, this reaction corresponds to a release of $17.22 \text{ kcal.mol}^{-1}$ Gibbs free energy and is thus a much more favorable alternative, than simple bromide ion dissociation. Therefore, formation of the active Pd(II) catalyst via a corresponding genesis of potassium bromide was adopted and modelled. The energetics of the final sequence of reactions being considered is presented in Figure 5-16.

Table 5-1: TPSSh Preliminary Oxidative Addition Reaction Free Energies

Chemical Reactions	ΔG_{298K} SVP (kcal.mol ⁻¹)	ΔG_{298K} ma-SVP (kcal.mol ⁻¹)	ΔG_{298K} COSMO/ ma-SVP (kcal.mol ⁻¹)
Ph-Br + Pd(0) Cat → Pre-complex	-5.29	-5.35	6.95
Pre-complex → TS	6.60	6.07	2.57
TS → Product Complex	-33.70	-32.47	-40.17
Product Complex → Active Pd(II) Catalyst + Br ⁻	28.09	14.96	-6.16
Product Complex + K ⁺ → Active Pd(II) Catalyst + KBr	-101.24	-99.75	-17.22

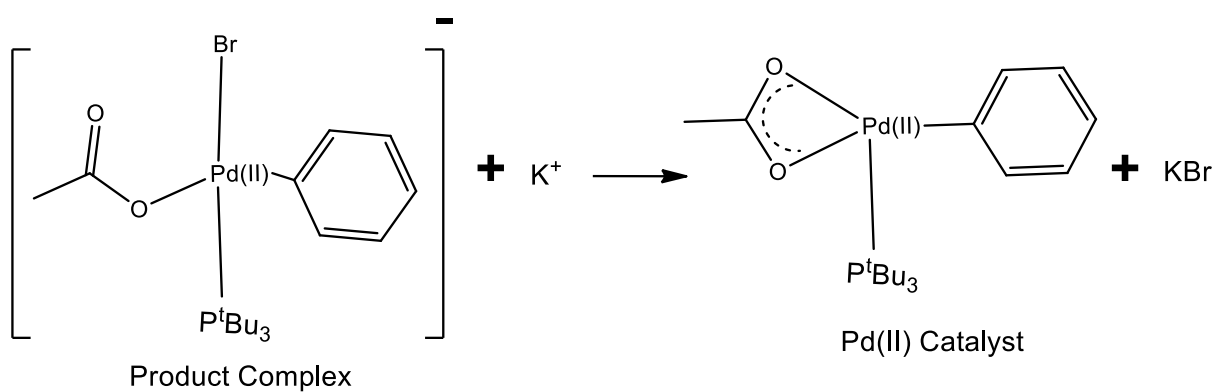


Figure 5-15: Br⁻ Dissociation Accelerated by Potassium Hydroxide Formation

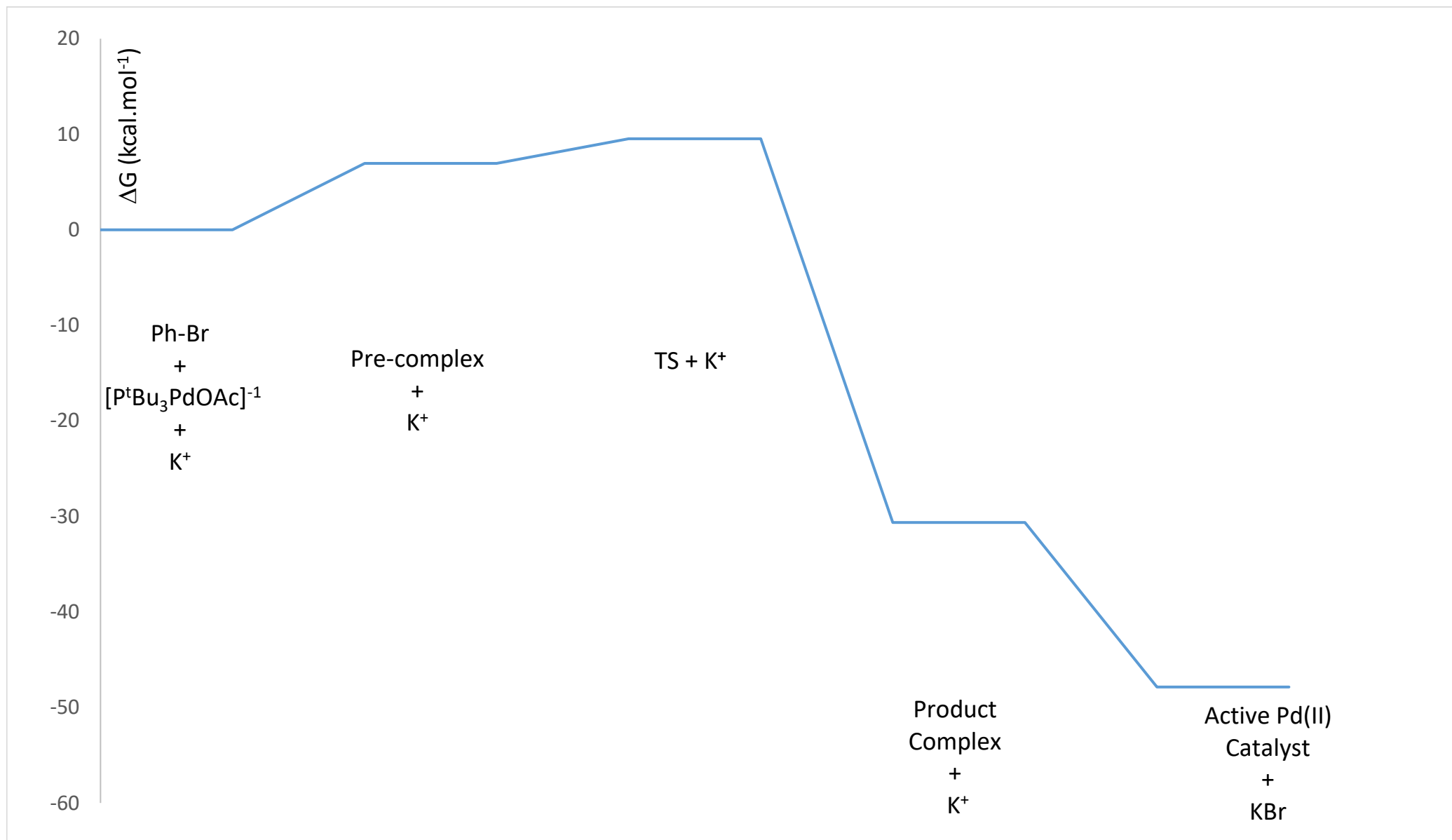


Figure 5-16: Oxidative Addition Reaction Profile (COSMO TPSSh/ma-Def2-SVP)

DLPNO-CCSD(T)/ DSD-PBEP86 Results:

In order to find a reliable way to account for dynamic electron correlation in OA reactions, single point energy benchmark calculations using DLPNO-CCSD(T)/TZVP were performed on the geometries obtained from COSMO TPSSh/ma-SVP. These were then compared to DSD-PBEP86/TZVP single point energies to confirm if the same conclusions can be drawn, in a similar manner to the investigation in Chapter 3. As per usual, thermodynamic and zero-point energy corrections were imported from COSMO TPSSh/ma-SVP frequency calculations and applied to the DLPNO-CCSD(T)/TZVP single point results.

We observed previously at the hybrid DFT level, Table 5-1, that the only reaction in which a minimally augmented basis set changed ΔG was dissociation of the bromide anion from the product complex. This reaction will not be considered, as dissociation via formation of KBr was determined to be a more favourable alternative. This observation, combined with the fact that diffuse functions are much more important for DFT methods than they are for wavefunction based theories, as well as that diffuse functions are most beneficial when added to double-zeta basis sets,²⁷ such as Def2-SVP, lead us to omit the use of ma basis sets during DLPNO-CCSD(T) and DSD-PBEP86 calculations in this chapter. The usual TZVP basis set was employed instead (Section 2.1.7). Due to the dramatic effects of its implementation on ΔG values, the use of the COSMO solvation model was however, deemed essential. It should be noted that as a K^+ ion is being considered in the reaction sequence, all DLPNO-CCSD(T) and DSD-PBEP86 calculations in this chapter made use of an energy window to determine the core electrons; see Chapter 2 Section 2.3 for details.

The results are depicted in Table 5-2 and graphically in Figure 5-17. As in previous chapters, it can be seen that overall, TPSSh gives a reasonable estimate of the reaction Gibbs energies, but severely underestimates the energy required to form the pre-complex as well as the TS barrier. The high degree of electron correlation expected in these species explains why DFT, a ground state theory, is inadequate here. An overall RMSD of $5.09 \text{ kcal.mol}^{-1}$ is observed when comparing TPSSh to the DLPNO-CCSD(T) benchmark.

Although a lower overall RMSD of $3.71 \text{ kcal.mol}^{-1}$ is observed when comparing DSD-PBEP86 reaction Gibbs energies with the DLPNO-CCSD(T) benchmark, the energy required to reach both the pre-complex and TS is again underestimated; by the margins of 6.04 and $1.16 \text{ kcal.mol}^{-1}$, respectively. This is consistent with the results from Chapter 3 that found DH DFT systematically underestimated the TS barriers considered therein, again when compared to a DLPNO-CCSD(T) benchmark. Considering Figure 5-17, it can be seen that DH DFT gives results

intermediate between that of hybrid DFT and DLPNO-CCSD(T). Regarding the rate determining step of oxidative addition as a whole, the same conclusion would be drawn from all three datasets, namely that the formation of the pre-complex is rate limiting. This step requires significantly more Gibbs energy than the TS barrier itself.

Table 5-2: Oxidative Addition Reaction Gibbs Free Energies- Various Levels of Theory

Chemical Reactions	$\Delta G_{298K} / \text{kcal.mol}^{-1}$ (COSMO TPSSH /ma-Def2-SVP)	$\Delta G_{298K} / \text{kcal.mol}^{-1}$ (COSMO DLPNO-CCSD(T)/Def2-TZVP)	$\Delta G_{298K} / \text{kcal.mol}^{-1}$ (COSMO DSD-PBEP86/Def2-TZVP)
Ph-Br + Pd(0) Cat $\rightarrow$ Pre-complex	6.95	16.01	9.97
Pre-complex $\rightarrow$ TS	2.57	7.00	5.85
TS $\rightarrow$ Product	-40.17	-40.57	-36.81
Product + K ⁺ $\rightarrow$ Active Pd(II) Catalyst + KBr	-17.22	-18.61	-16.84

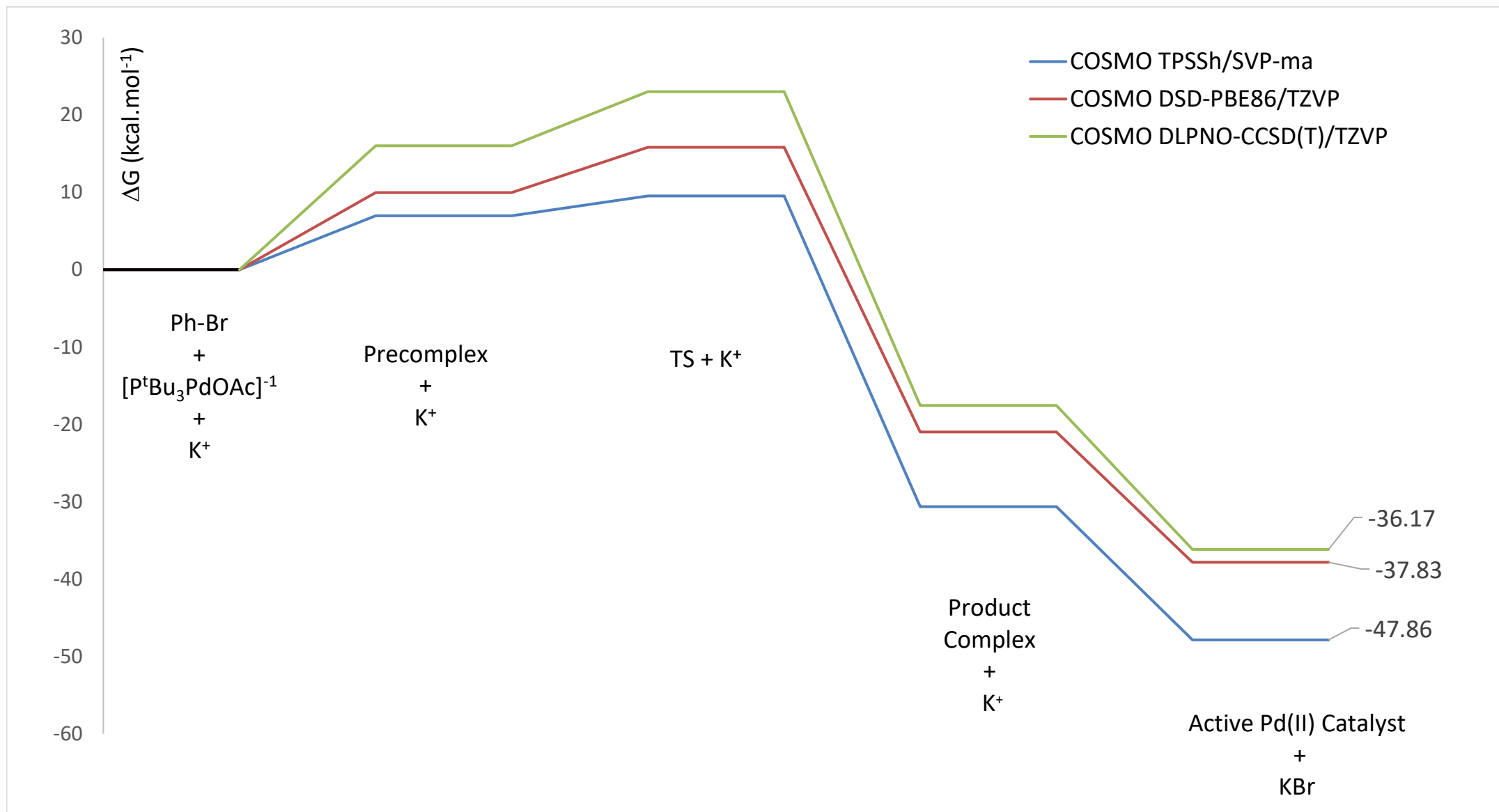


Figure 5-17: Oxidative Addition Reaction Profile – Various Levels of Theory

The overall $\Delta_r G$ of $-37.83 \text{ kcal.mol}^{-1}$ predicted at the DSD-PBE86 level is sufficiently close to the benchmark value of $-36.17 \text{ kcal.mol}^{-1}$. Thus, DH DFT correctly predicts that the overall change in Gibbs energy for the reaction is highly exergonic. In the compromise between efficiency and accuracy, DH DFT strikes an adequate balance when applied to the oxidative addition reactions being considered. No further use of DLPNO-CCSD(T) was deemed necessary during the remainder of the chapter.

The acceleration of the formation of the active Pd(II) catalyst by the K^+ ion has now been confirmed using a higher level of theory. It is possible that this process could become even more favourable by using bases with other cations that bind more strongly to the Br^- anion, or even by the addition of salts containing such cations to the reaction solution. However, further computational as well as experimental studies would be necessary to confirm this suggestion.

In order to determine overall RDS of the catalytic cycle under conditions of 1 equivalent of P^tBu_3 ligand, it was necessary to compare the DH results from this chapter with those from the CMD mechanism under the appropriate conditions (Chapter 4). Unfortunately, the DH calculations in Chapter 4 were run with no solvation model and made use of the "FC_ELECTRONS" method of determining the core electrons, rather than an energy window. See Chapter 2 Section 2.3 for details.

Therefore, the DH results of Chapter 4 needed to be re-run so that a direct comparison would be possible. All geometries in this chapter were optimised in a COSMO solvation model, while those in Chapter 4 were exclusively gas phase. Applying a COSMO solvation model to the CMD mechanism under phosphine free conditions (Chapter 3, Section 3.2) changed the magnitude of the total G input across pathways by an amount, in most cases, within the inherent error of the underlying DFT functional (assumed to be $\sim 2 \text{ kcal.mol}^{-1}$). Furthermore, the major conclusions in that chapter at the TPSSh level, from both gas phase and COSMO data, remained identical. Additionally, no meaningful structural differences between gas and solution phase geometries were observed. Considering the above, as an estimate, it was deemed acceptable to rerun the DH calculations in Chapter 4 using the original gas phase geometries but employing an energy window to determine the core electrons as well as with a COSMO solvation model. To compute G values, the thermodynamic corrections were imported from the original gas phase TPSSh calculations. In this way, the energy barriers of both oxidative addition and concerted metallation deprotonation could be directly compared at approximately the same level of theory.

The results are shown in Table 5-3. It should be noted that as gas phase DH calculations were deemed sufficient for the CMD mechanism in Chapters 3 and 4, no changes in the relative energy as a consequence of applying a COSMO solvation model were considered and the conclusions of that chapter remain valid. The results of Table 5-3 simply serve the purpose of allowing CMD and RE barrier heights to be directly compared to those of OA.

For all of the pathways considered, the CMD step is significantly higher in Gibbs energy requirements than formation of the OA pre-complex; by on average 8.30 kcal.mol⁻¹. Thus OA is in no way rate determining. Interestingly, the OA addition barrier for 3 of the 4 pathways considered is comparable to that of RE.

The average Gibbs energy required to overcome the RDS is very similar, at the gas phase DSD-PBEP86/TZVP level of theory, for both the mono- and di-substitution series; 17.86 vs 19.70 kcal.mol⁻¹ respectively. Considering this, is it thus highly probable that OA will also not be the RDS for di-substitution under the relevant conditions of one equivalent of P^tBu₃. Ideally, rerunning DH calculations for the entire di-substitution series from Chapter 4 but employing an energy window to determine the core electrons as well as with a COSMO solvation model would allow a direct comparison with the results of OA, thus confirming this assumption. Due to the excessive length of the present work as well as time constraints, unfortunately, this was not practical.

Table 5-3: Mono-substitution CMD Reaction Free Energies (COSMO DSD-PBEP86/TZVP)

Chemical Reaction	ΔE (kcal.mol ⁻¹)	ΔG_{298K} (kcal.mol ⁻¹)
<u>Pathway N2</u>		
Catalyst + Oxazole → Int 1	-3.69	15.71
Int 1 → TS 1	22.03	18.85
TS 1 → Int 2	-21.50	-18.45
Int 2 → TS 2	16.36	14.89
TS 2 → Product 2 + Pd(0) Catalyst	-12.28	-33.57
<u>Pathway N4</u>		
Catalyst + Oxazole → Int 1	-2.99	16.15
Int 1 → TS 1	24.25	21.25
TS 1 → Int 2	-13.10	-10.40
Int 2 → TS 2	10.10	8.17
TS 2 → Product 4 + Pd(0) Catalyst	-15.78	-36.18
<u>Pathway =4</u>		
Catalyst + Oxazole → Int 1	2.20	21.98
Int 1 → TS 1	23.19	19.17
TS 1 → Int 2	-17.23	-14.14
Int 2 → TS 2	10.10	8.17
TS 2 → Product 4 + Pd(0) Catalyst	-15.78	-36.18
<u>Pathway =5</u>		
Catalyst + Oxazole → Int 1	-2.39	17.78
Int 1 → TS 1	18.01	13.82
TS 1 → Int 2	-9.20	-6.62
Int 2 → TS 2	11.37	9.96
TS 2 → Product 5 + Pd(0) Catalyst	-16.10	-36.68

Now that OA has been considered, insight into regeneration of the active Pd(0) catalyst, the last step in the cycle, is necessary. As a reminder from Chapter 3, after OA and subsequent CMD, RE generates the relevant product and a protonated Pd(0) catalyst. This is illustrated in Figure 5-18 for Pathway =5.

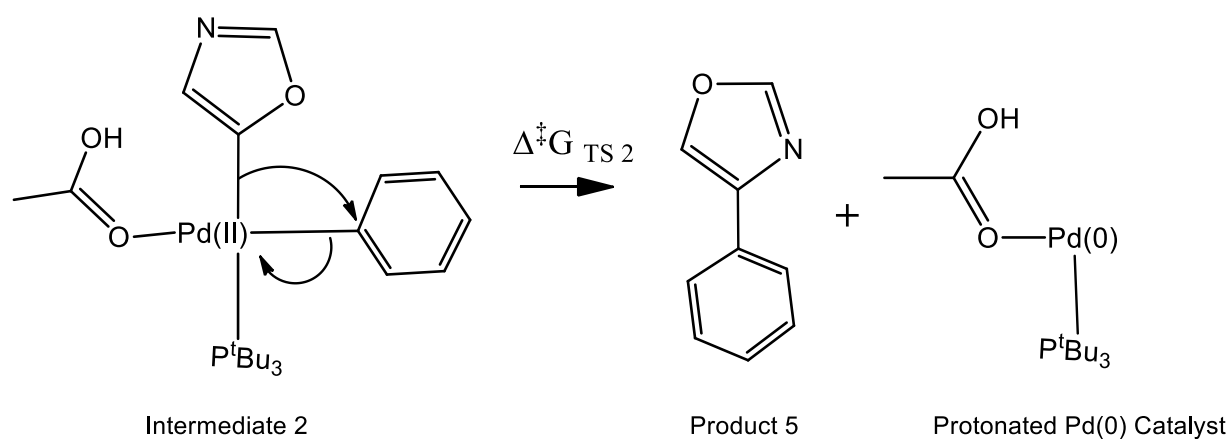


Figure 5-18: Product and Protonated Pd(0) Catalyst formation via Reductive Elimination (Pathway =5)

Deprotonation via a hydroxide ion then regenerates the active Pd(0) catalyst, beginning a new cycle (Figure 5-19).

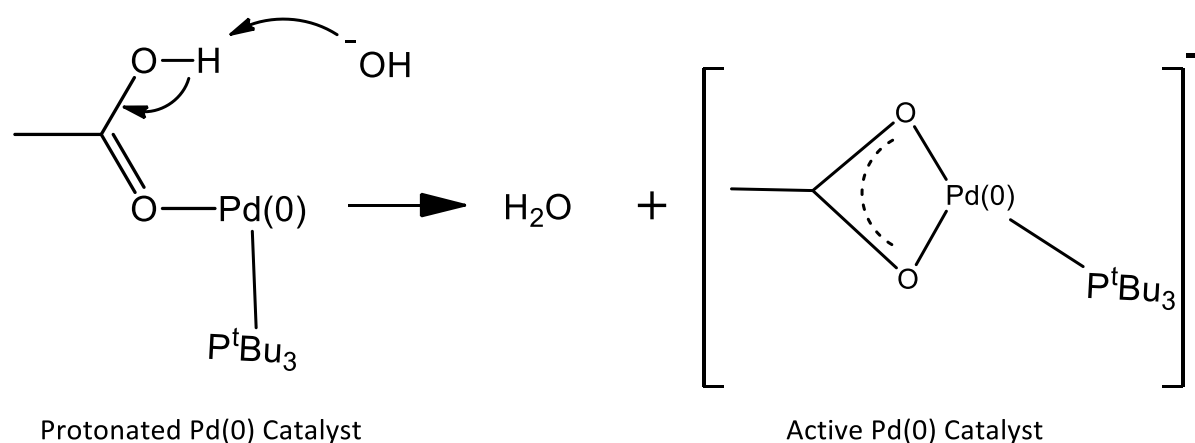


Figure 5-19: Regeneration of the Active Pd(0) Catalyst via Deprotonation

The overall change in Gibbs free energy for this process is -32.17 kcal.mol⁻¹ (COSMO DSD-PBEP86/TZVP). It is thus highly favorable. Additionally, no significant transition state is expected as the O-H bond is acidic; the pK_a of acetic acid being 4.76.²⁸

The entire catalytic cycle (abbreviated in Figure 5-1) has now been modelled. The energetics of the complete cycle, from the formation of the active Pd(0) catalyst to its regeneration, for Pathway =5 is illustrated in Figure 5-20. The molecules corresponding to the reaction coordinate 1-11 in this figure are detailed in Table 5-4. The overall change in Gibbs free energy for one complete catalytic cycle in which Product 5 is generated via Pathway =5 is -72.26 kcal.mol⁻¹.

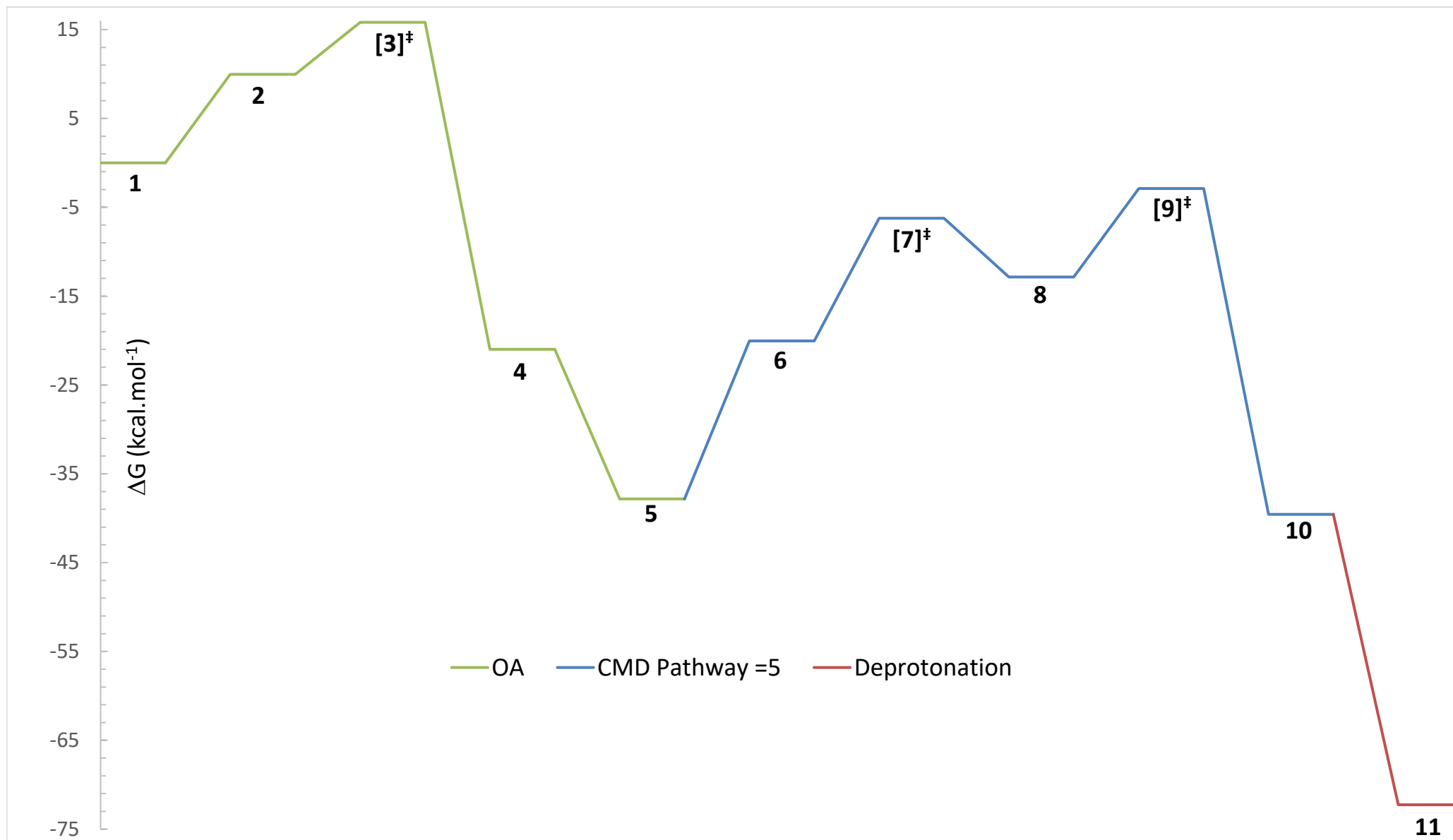


Figure 5-20: Complete Catalytic Cycle - Pathway =5 (COSMO DSD-PBEP86/TZVP). Reaction Coordinate Labels correspond to Table 5-4.

Table 5-4: Molecules Corresponding to the Reaction Coordinate in Figure 5-22

Reaction Category	Molecules	Reaction Coordinate
Oxidative Addition	Ph-Br + Active Pd(0) Catalyst + K ⁺ + Oxazole + OH ⁻	1
	Pre-complex + K ⁺ + Oxazole + OH ⁻	2
	OA TS + K ⁺ + Oxazole + OH ⁻	[3][‡]
	Product Complex + K ⁺ + Oxazole + OH ⁻	4
	Active Pd(II) Catalyst + KBr + Oxazole + OH ⁻	5
Concerted Metallation Deprotonation Pathway =5	Intermediate 1 + KBr + OH ⁻	6
	CMD TS 1 + KBr + OH ⁻	[7][‡]
	Intermediate 2 + KBr + OH ⁻	8
	CMD TS 2 + KBr + OH ⁻	[9][‡]
	Product 5 + Protonated Pd(0) Catalyst + KBr + OH ⁻	10
Deprotonation	Product 5 + Active Pd(0) Catalyst + KBr + H ₂ O	11

In conclusion, modelling suggests that OA is not the RDS for either mono or di-substitution under conditions of one equivalent of P^tBu₃. If this work is to be used as a preliminary investigation regarding optimisation of these reactions, research into lowering the energy required for OA to occur would not yield a significant increase in reaction rate or allow these reactions to be optimised for lower temperature conditions.

5.3 Phosphine Free Conditions

Formation of the Active Pd(0) Catalyst:

Much less is known about the mechanism for Pd(0) formation from the Pd(OAc)₂ pre-catalyst under phosphine free conditions. It is possible that under these conditions the reduction of palladium takes place with a corresponding oxidation of a DMA solvent molecule in a mechanism similar to the oxidation of tertiary amine ligands⁶ (Figure 5-21). It is known that simply warming a Pd(OAc)₂ solution in THF at 66 °C leads to the formation of Pd(0).²⁹ This is supposedly via the electron transfer from two acetates to a Pd(II) unit with formation of gaseous methane, ethane, and CO₂. No mechanism for this transformation has been proposed. Regardless of the mechanism, Pd(0) formation does however occur. Anionic Pd(0) complexes with a solvent molecule coordinated to the palladium centre, rather than a phosphine ligand, have been proposed as lower energy alternative active Pd(0) catalysts.¹⁷ Additionally, the complex [Pd(H₂O)OAc]⁻ has been detected experimentally in Heck reactions that made use of a Pd(OAc)₂ pre-catalyst.³⁰ In this work we replaced the P^tBu₃ ligand of the active Pd(0) catalyst under phosphine conditions, Figure 5-6, with a DMA solvent molecule to generate the complex illustrated in Figure 5-22 and initially employed this molecule, the active Pd(0) catalyst, under phosphine free conditions. As before, modelling was initiated from the active Pd(0) catalyst and its formation was not considered, as this process occurs only once and is assumed to be energetically irrelevant.

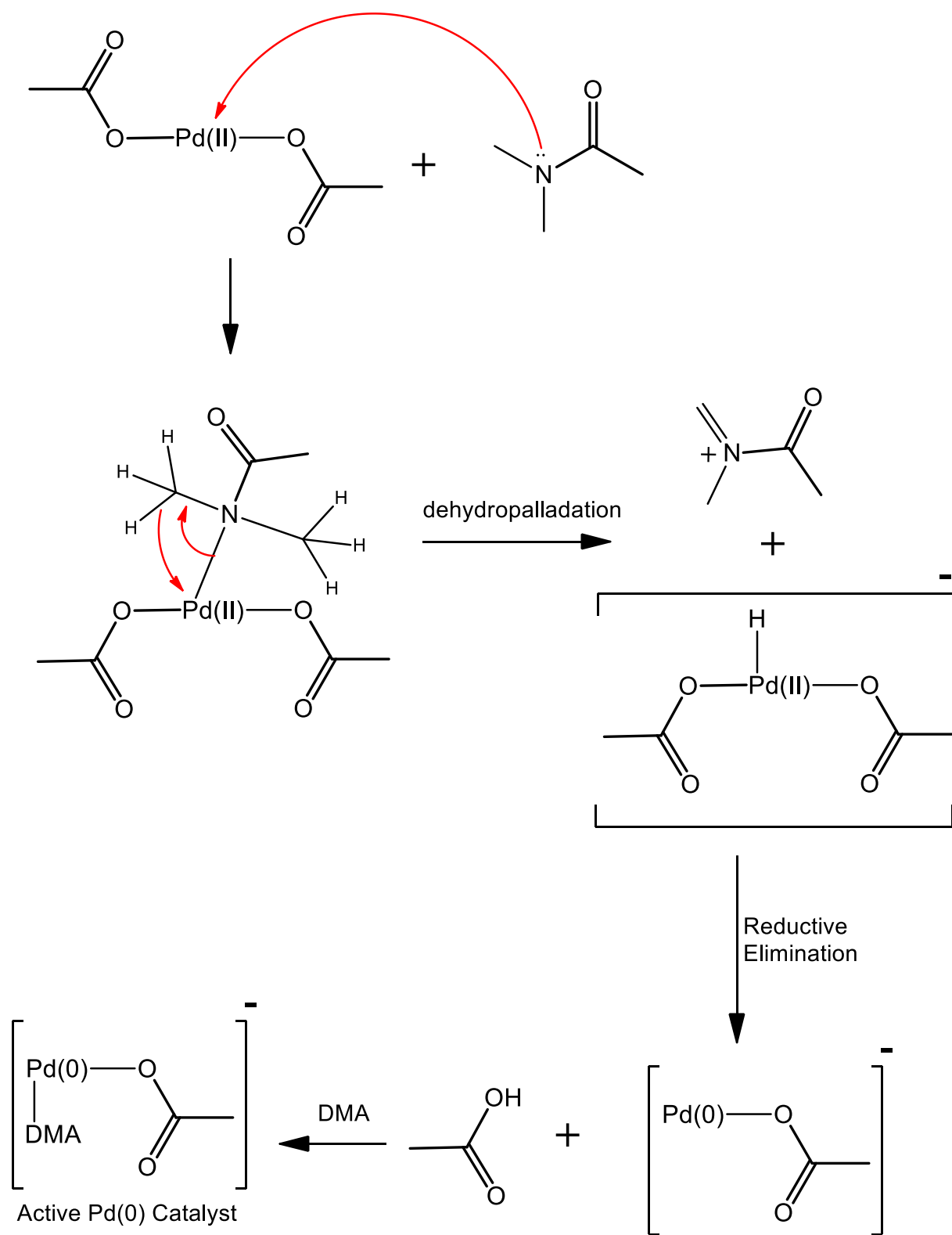


Figure 5-21: Possible Mechanism for the Formation of Pd(0) from Pd(OAc)_2 – Phosphine Free Conditions

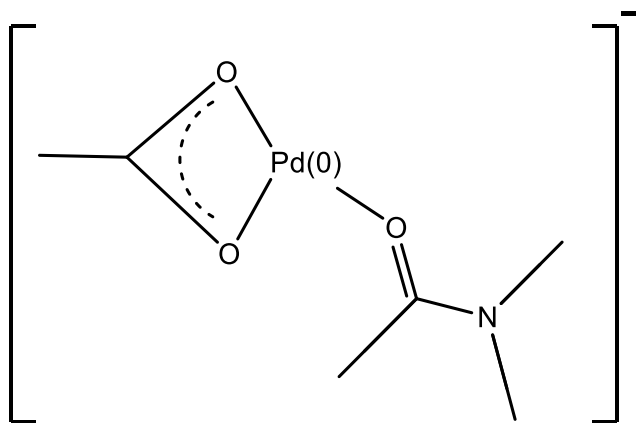


Figure 5-22: Active Pd(0) Catalyst - Phosphine Free Conditions

The Reaction Pathway:

Initially, exactly the same sequence of reactions was considered as for OA under conditions of phosphine ligand, but with P^tBu_3 simply replaced by a DMA molecule (Figure 5-23). For the purpose of consistency with Section 5.1, all geometry optimizations and frequency calculations were performed at the COSMO TPSSh/ma-Def2-SVP level of theory. The optimised structures for this pathway are presented in Figure 5-24 (Active Pd(0) Catalyst), Figure 5-25 (Pre-complex), Figure 5-26 (TS) and finally Figure 5-27 (Product complex). The TS has an imaginary frequency of -132.67 cm^{-1} .

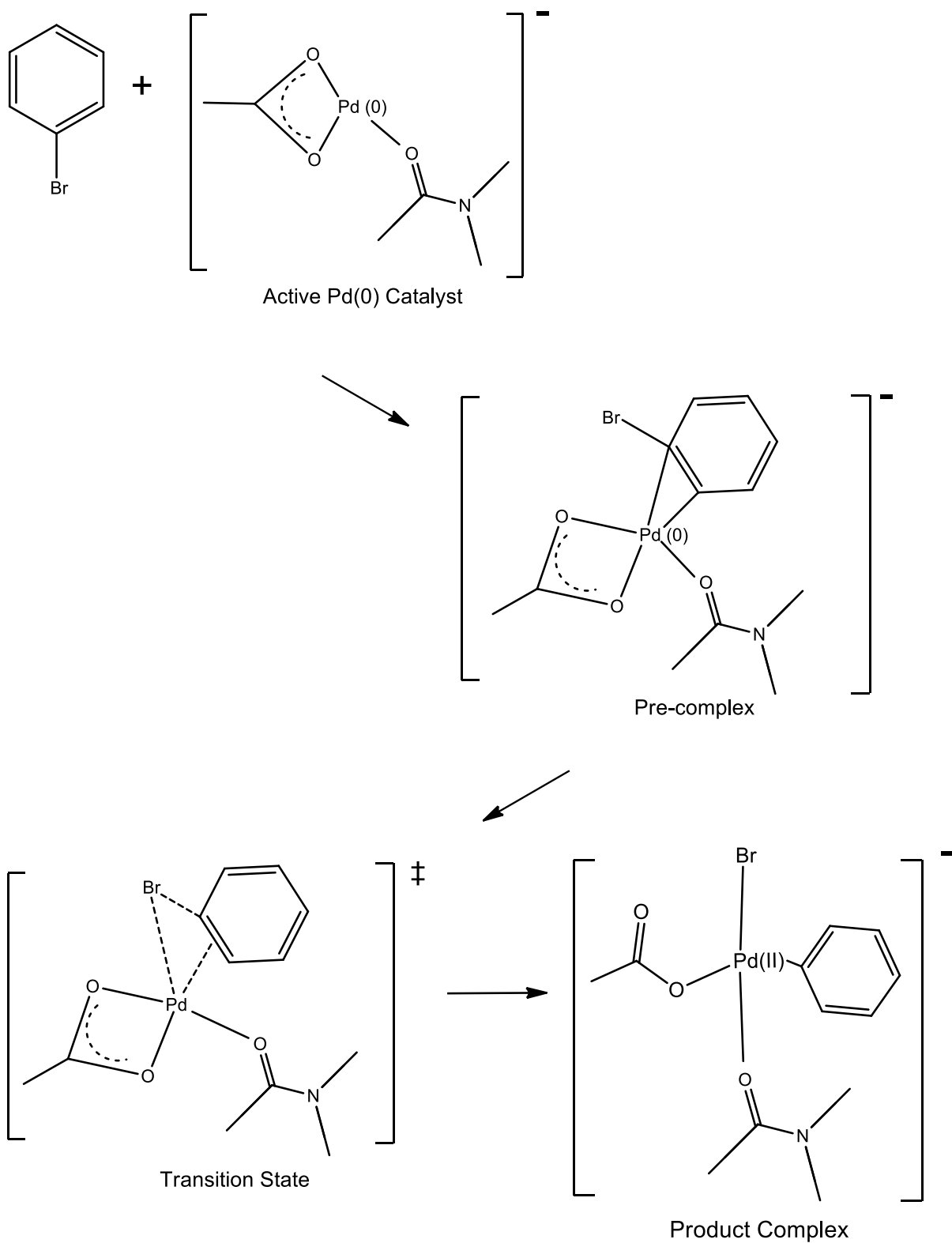
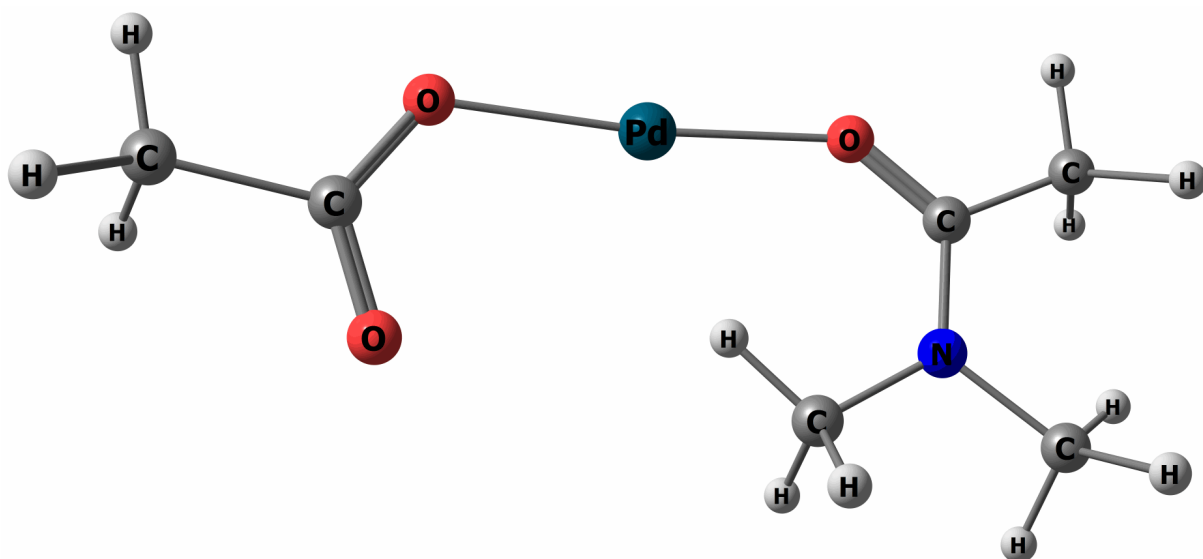
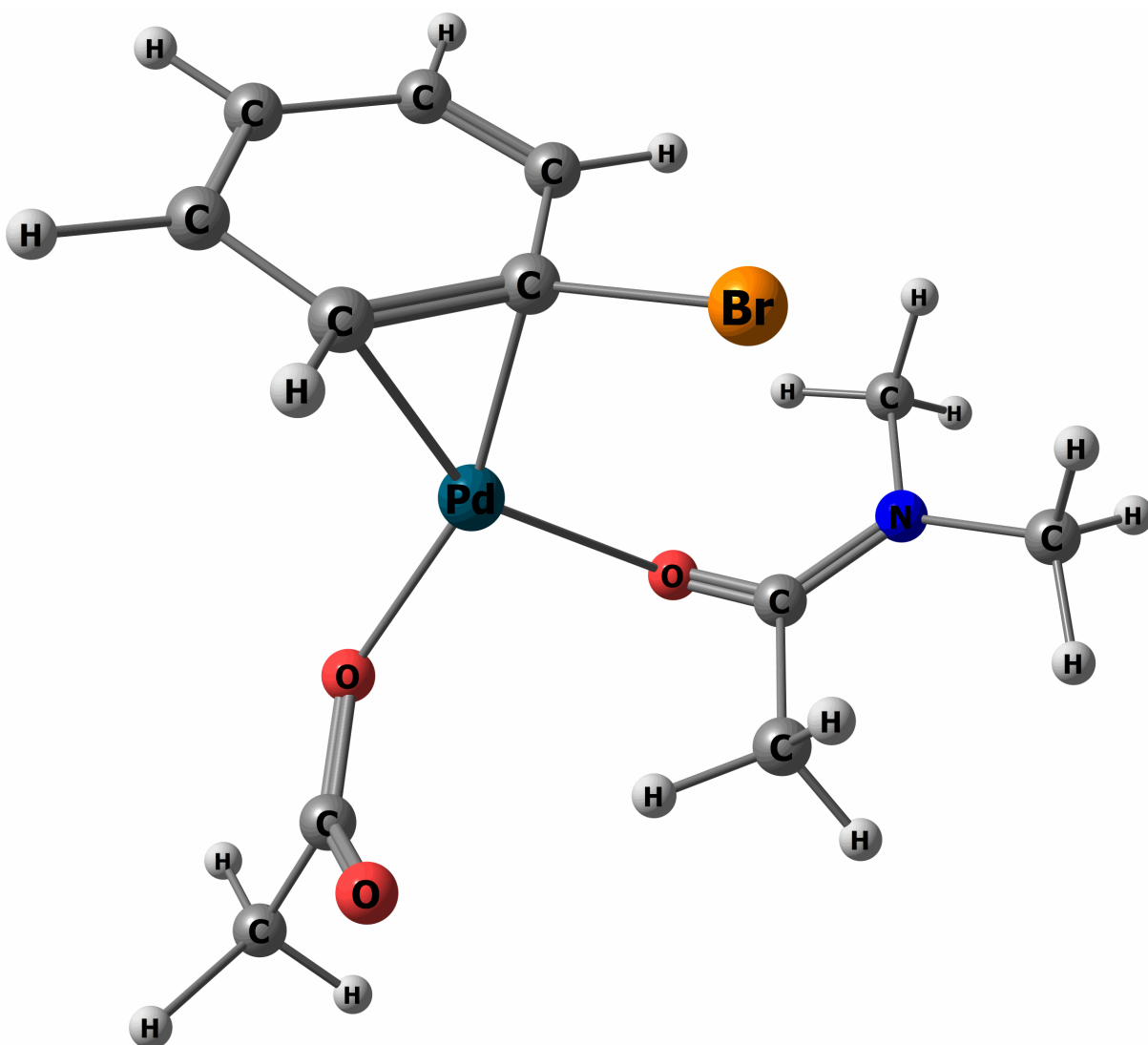


Figure 5-23: Initial Pathway Considered for Oxidative Addition Under Phosphine Free Conditions



[Figure 5-24: Initially Considered Active Pd\(0\) Catalyst](#)



[Figure 5-25: Initially Considered Pre-complex](#)

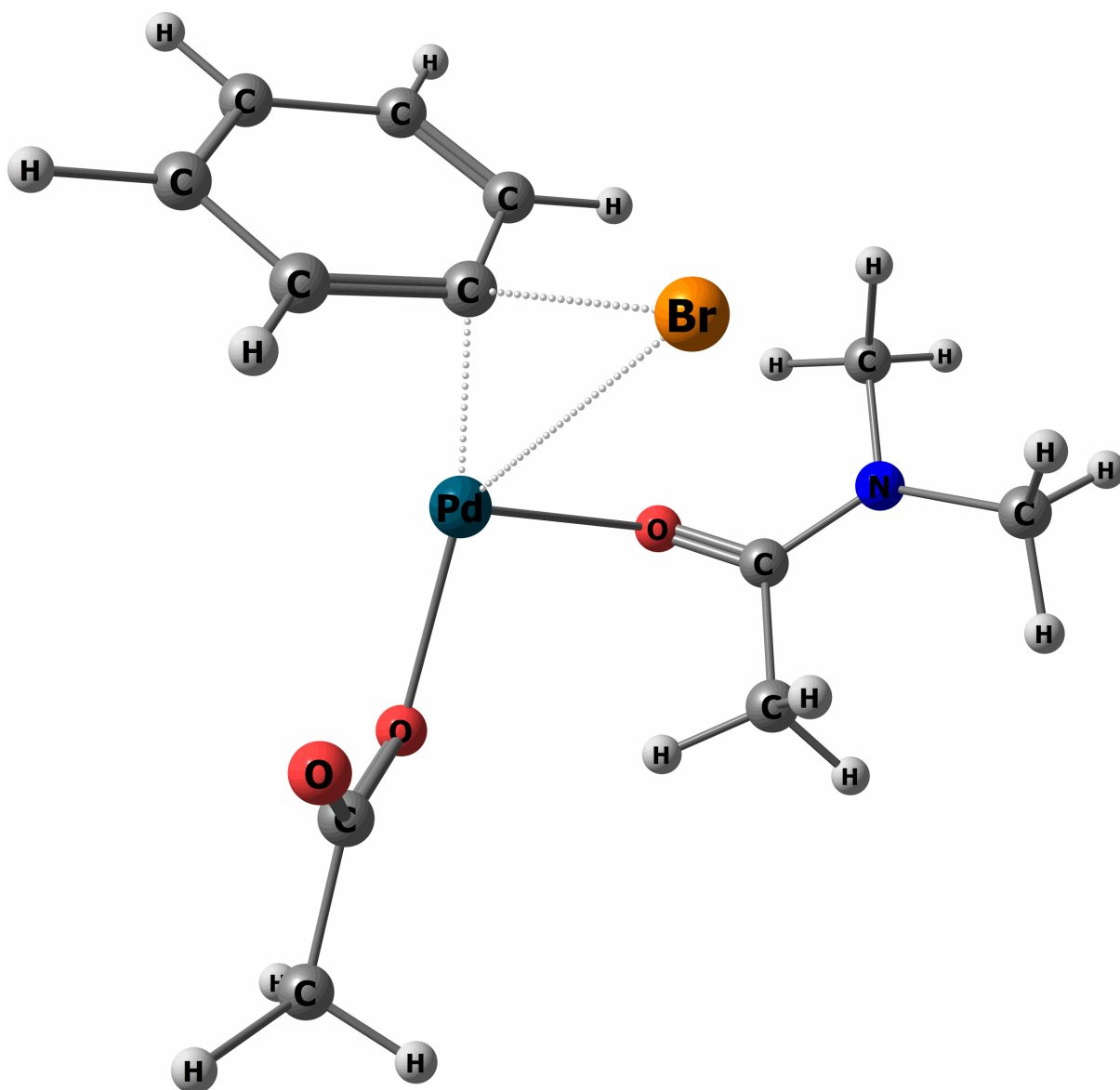


Figure 5-26: Initially Considered Transition State

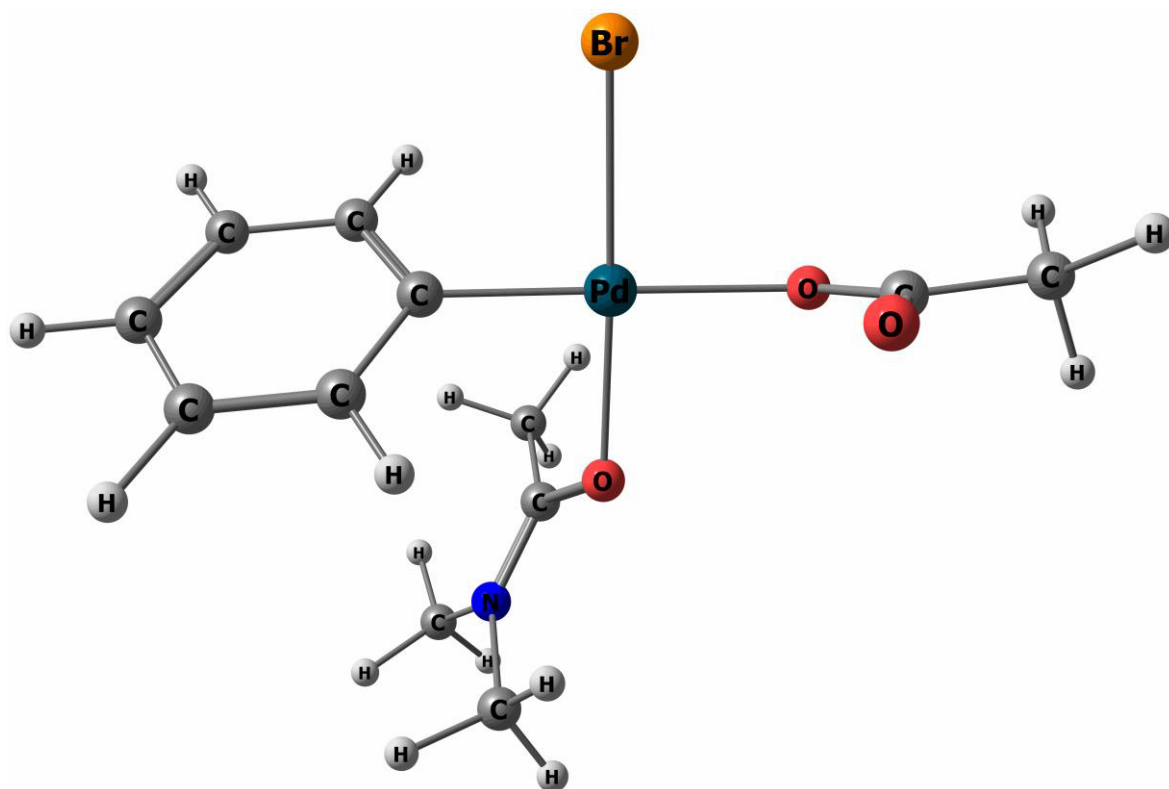


Figure 5-27: Initially Considered Product Complex

While attempting to find an improved TS structure, it was discovered that a lower energy OA pathway exists if one removes the DMA ligand from the system completely. This pathway is presented in Figure 5-28. Optimised structures of the active Pd(0) Catalyst (Figure 5-29), Pre-complex (Figure 5-30), TS (Figure 5-31) and Product Complex (Figure 5-32) are presented below. The TS has a single imaginary frequency of -169.15 cm^{-1} . The energetics of the initial and final pathways are compared in Figure 5-33 at the COSMO TPSSh/ma-Def2-SVP level of theory. It can be seen that removing the DMA molecule during OA results in a pathway that is significantly lower in Gibbs energy, on average by $19.81\text{ kcal.mol}^{-1}$. From now on this lower energy pathway will be considered exclusively.

The question that then arises is, how and when does DMA coordinate and Br^- dissociate to form the active Pd(II) catalyst illustrated in Figure 5-3 and modelled in Chapter 3? After some experimentation, we propose that a potassium cation would coordinate to the Product complex through the Br, producing the neutral complex depicted in Figure 5-34. DMA coordination would then displace KBr, generating the active Pd(II) catalyst by the reaction shown below in Figure 5-35.

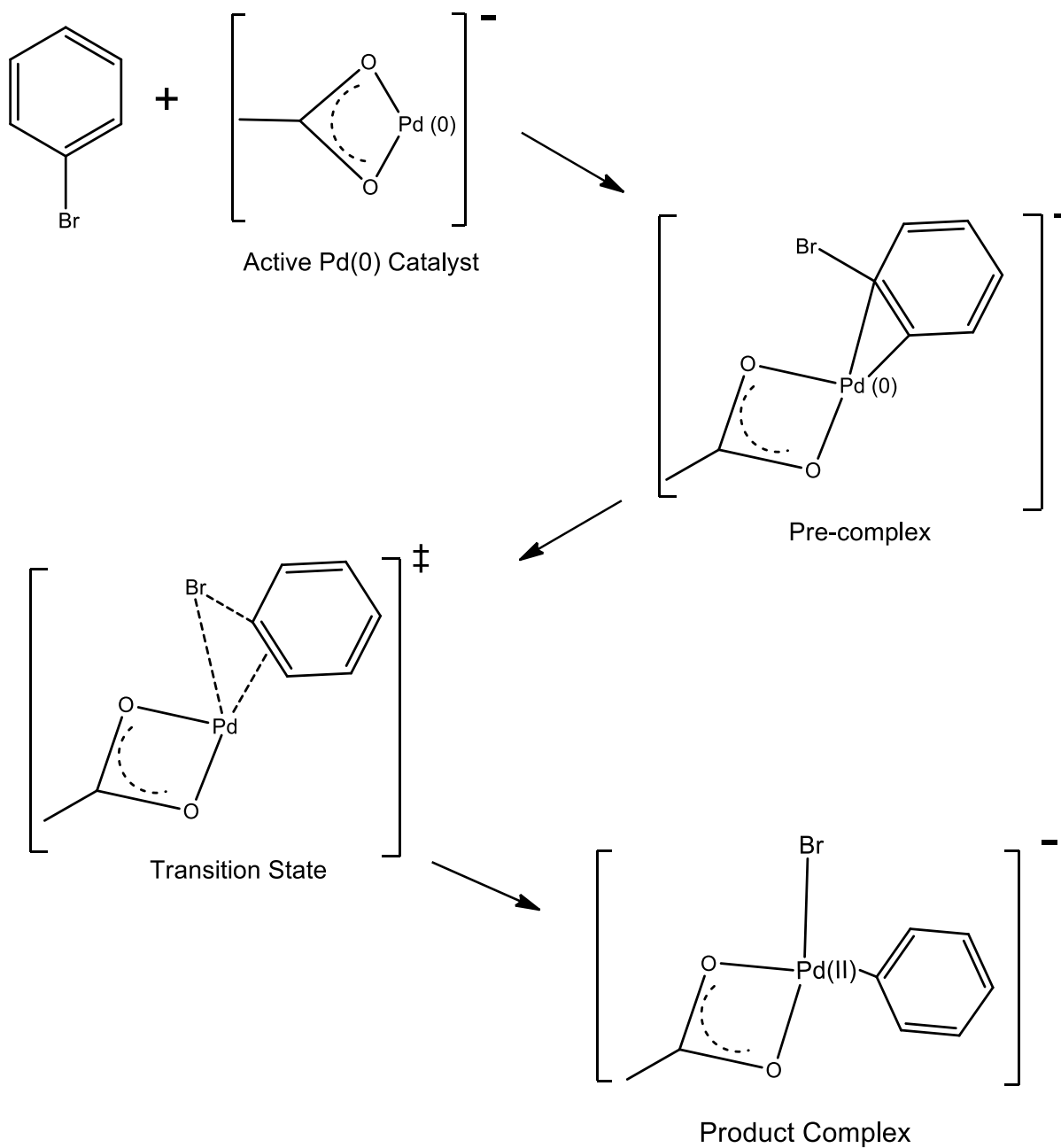
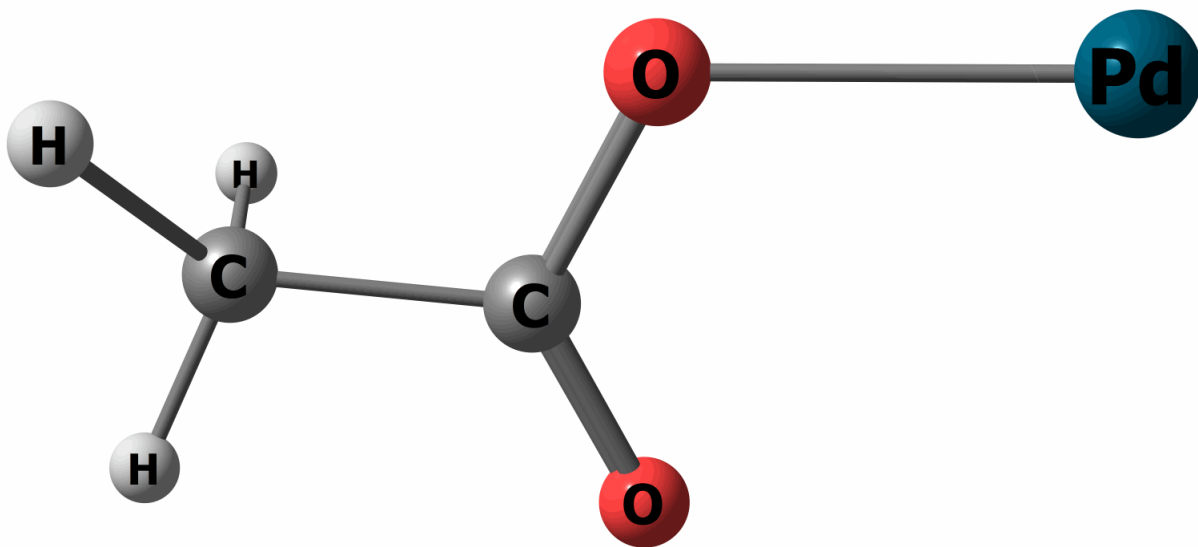


Figure 5-28: Final Pathway Considered for Oxidative Addition Under Phosphine Free Conditions



[Figure 5-29: Active Pd\(0\) Catalyst Under Phosphine Free Conditions](#)

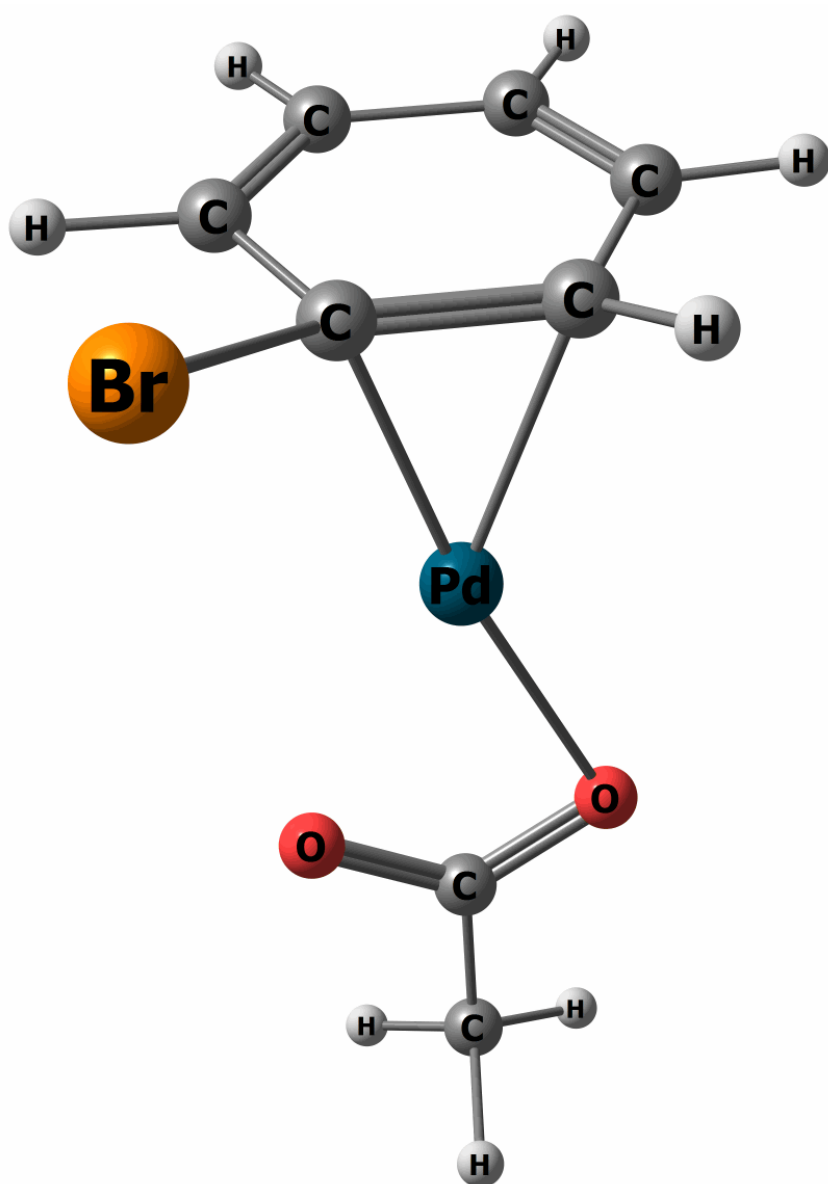


Figure 5-30: Pre-Complex Under Phosphine Free Conditions

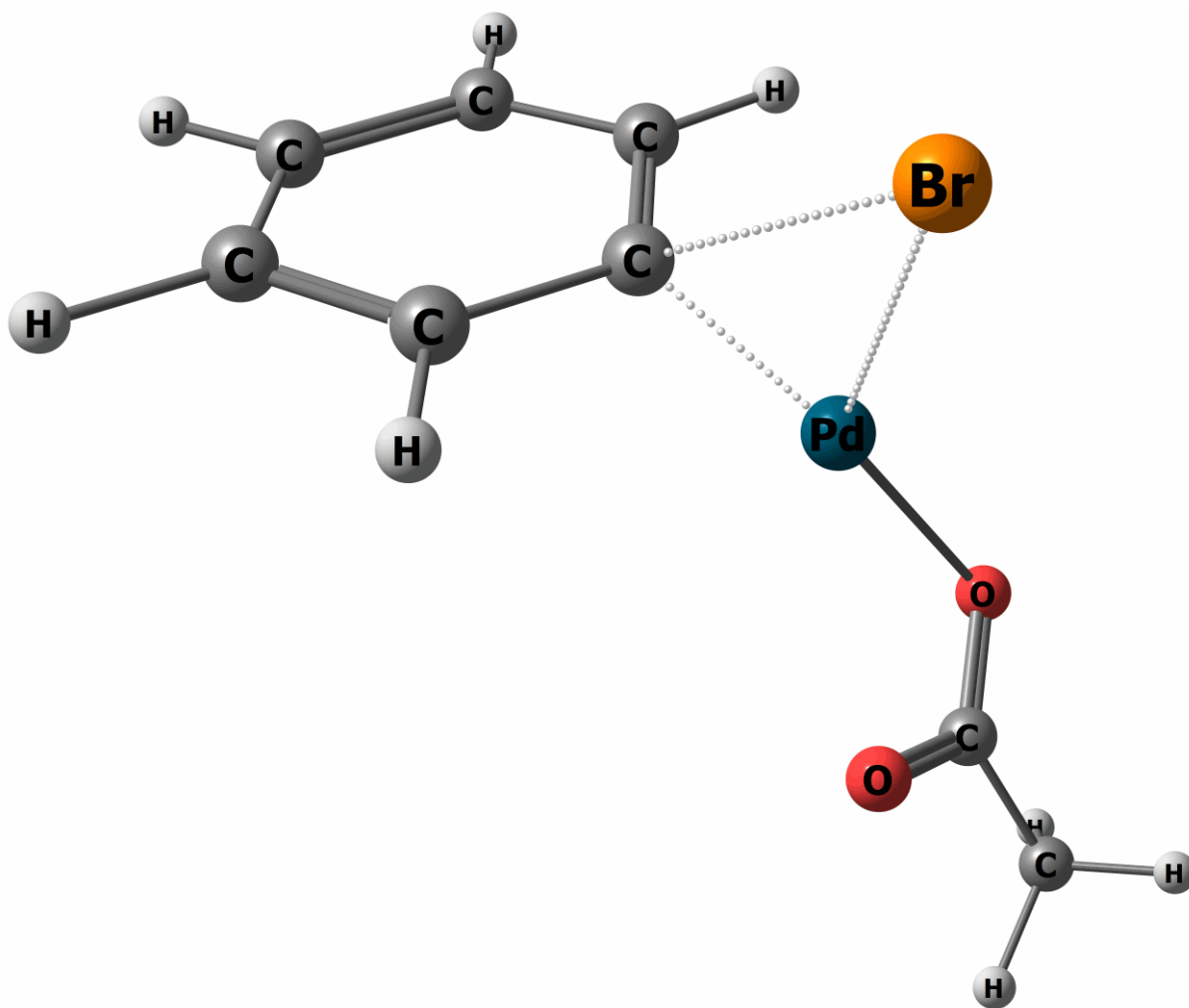


Figure 5-31: TS Under Phosphine Free Conditions

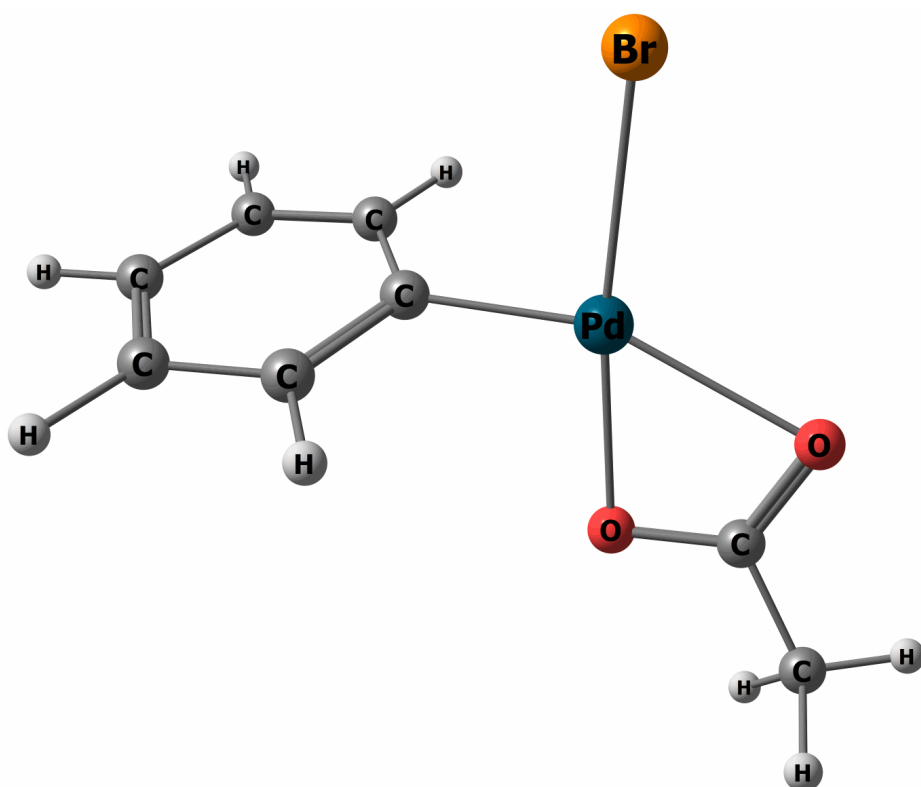


Figure 5-32: Product Complex Under Phosphine Free Conditions

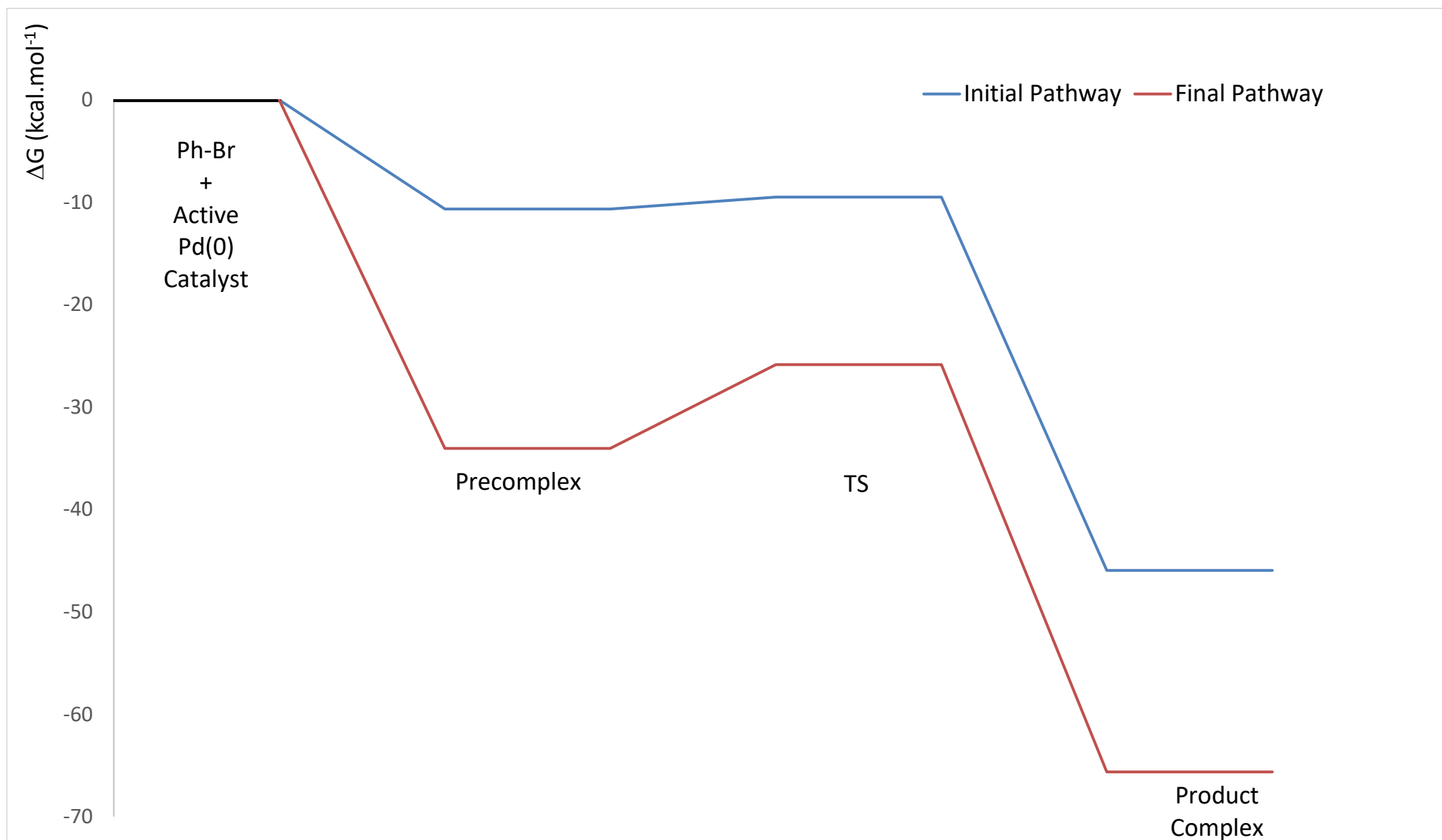


Figure 5-33: Energetics of the Initial and Final Pathways (COSMO TPSSH/ma-Def2-SVP)

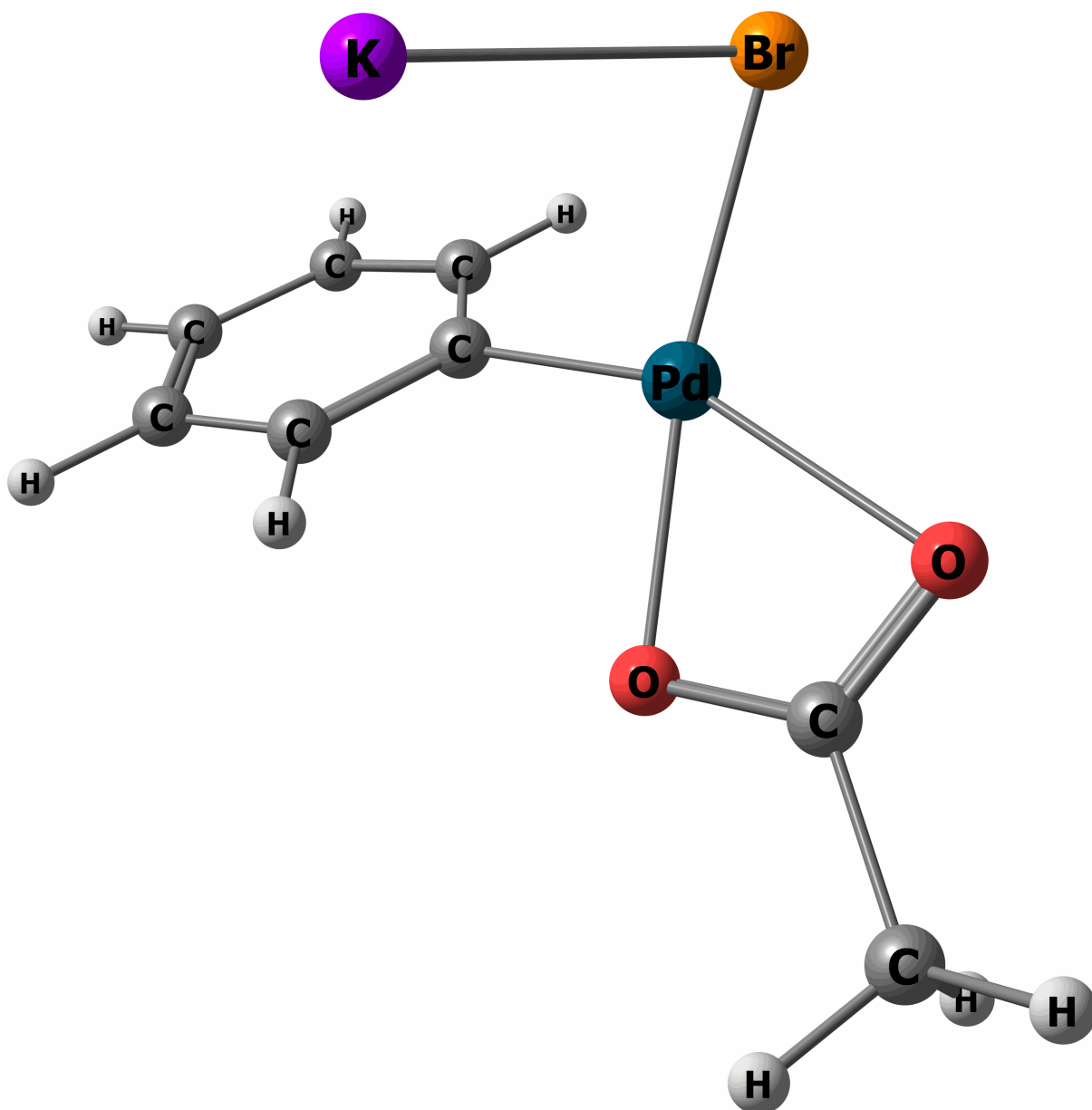


Figure 5-34: Neutral Complex with Coordinated Potassium Cation

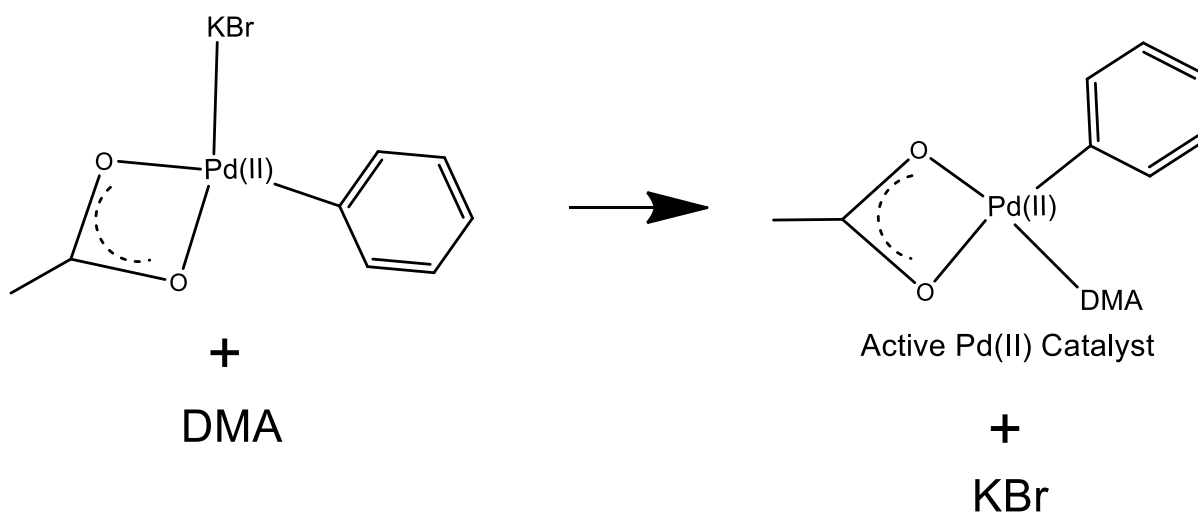


Figure 5-35: Formation of the Active Pd(II) Catalyst

DSD-PBEP86 Results:

DH DFT was shown to be an adequate theoretical treatment of OA reactions in Section 5.2. COSMO DSD-PBEP86/TZVP results for the final reaction pathway considered are presented in the form of reaction Gibbs free energies in Table 5-5 and graphically in Figure 5-36.

Table 5-5: Oxidative Addition Reaction Gibbs Free Energies (COSMO DSD-PBEP86/TZVP)

Chemical Reactions	ΔE /kcal.mol ⁻¹	ΔG_{298K} /kcal.mol ⁻¹
Ph-Br + Active Pd(0) Catalyst → Pre-complex	-48.22	-32.39
Pre-complex → TS	11.30	12.31
TS → Product	-38.27	-38.24
Product + K ⁺ → Pd(Ar)(OAc)(Br)(K)	-8.03	0.82
Pd(Ar)(OAc)(Br)(K) + DMA → Active Pd(II) Catalyst +KBr	4.46	9.01

The reaction profile differs dramatically from OA in the presence of P^tBu₃ in the sense that the formation of the pre-complex is now very favourable and the TS itself is the RDS for the overall OA process. The RDS is somewhat higher than under phosphine conditions (12.31 vs 9.97 kcal.mol⁻¹). It is thus expected the OA will proceed less rapidly and requires higher temperatures when no P^tBu₃ is present in the system.

The overall $\Delta_r G$ of -48.50 kcal.mol⁻¹ reveals that the net OA process, under these conditions is highly exergonic, even more so than under phosphine conditions. Thus, although OA requires more energy to overcome the RDS reaction barrier in the absence of phosphine ligand, once this energy has been provided the reactions are expected to run until completion with a concurrent release of free energy.

In order to determine overall RDS of the catalytic cycle under phosphine free conditions, it was necessary to compare the DH results from this chapter with those from the CMD mechanism under the appropriate conditions (Chapter 3). Unfortunately, the DH calculations in Chapter 3 were run with no solvation model and made use of the “FC_ELECTRONS” method of determining the core electrons, rather than an energy window. See Chapter 2, Section 2.3 for details. Therefore, the DH results of Chapter 3 needed to be re-run so that a direct comparison at the same level of theory was possible. COSMO geometries for mono substitution under phosphine free conditions were fortunately available from Chapter 3 (Section 3.2). DH calculations using these geometries, but employing a COSMO solvation model and an energy window to determine the core electrons, were subsequently performed. The results are shown in Table 5-6.

It can be seen that OA is not rate determining for any of the pathways considered. It was concluded in Chapter 3 that Pathway =5 is the only kinetically accessible mono-substitution pathway, under those conditions. Even here, CMD requires 3.98 kcal.mol⁻¹ more Gibbs energy than OA. For this pathway RE and OA are very similar in magnitude, an observation consistent with the trends in Section 5.2.

Table 5-6: Mono-substitution CMD Reaction Free Energies (COSMO DSD-PBEP86/TZVP)

Chemical Reaction	ΔE (kcal.mol ⁻¹)	ΔG_{298K} (kcal.mol ⁻¹)
<u>Pathway N2</u>		
Catalyst + Oxazole → Int 1	-19.67	-2.39
Int 1 → TS 1	28.38	24.48
TS 1 → Int 2	-31.21	-27.99
Int 2 → TS 2	22.39	22.45
TS 2 → Product 2 + Pd(0) Catalyst	6.16	-11.54
<u>Pathway N4</u>		
Catalyst + Oxazole → Int 1	-19.67	-2.39
Int 1 → TS 1	28.89	26.12
TS 1 → Int 2	-22.97	-20.59
Int 2 → TS 2	22.93	22.72
TS 2 → Product 4 + Pd(0) Catalyst	-1.57	-19.29
<u>Pathway =4</u>		
Catalyst + Oxazole → Int 1	-13.28	3.92
Int 1 → TS 1	22.59	19.40
TS 1 → Int 2	-23.06	-20.17
Int 2 → TS 2	22.93	22.72
TS 2 → Product 4 + Pd(0) Catalyst	-1.57	-19.29
<u>Pathway =5</u>		
Catalyst + Oxazole → Int 1	-13.59	3.18
Int 1 → TS 1	18.29	16.29
TS 1 → Int 2	-15.42	-13.33
Int 2 → TS 2	14.77	11.49
TS 2 → Product 5 + Pd(0) Catalyst	2.77	-11.78

At the gas phase DSD-PBEP86/TZVP level of theory, the average RDS for di-substitution in the CMD mechanism was 19.99 kcal.mol⁻¹ (Chapter 3, Section 3.3). Considering this, it is unlikely that OA will be rate determining for any of the pathways leading to di-substitution either. However, for Pathway 4 → 4,5 and Pathway 2 → 2,5 with rate determining steps of 14.43 and 15.21 kcal.mol⁻¹ respectively, OA might come close to being the bottle neck process. Ideally, rerunning DH calculations for the entire di-substitution series from Chapter 3 but employing an energy window to determine the core electrons as well as with a COSMO solvation model would allow a direct comparison with the results of this chapter, allowing confirmation of

these assumptions. Due to the excessive length of the present work as well as time constraints, this was unfortunately not possible.

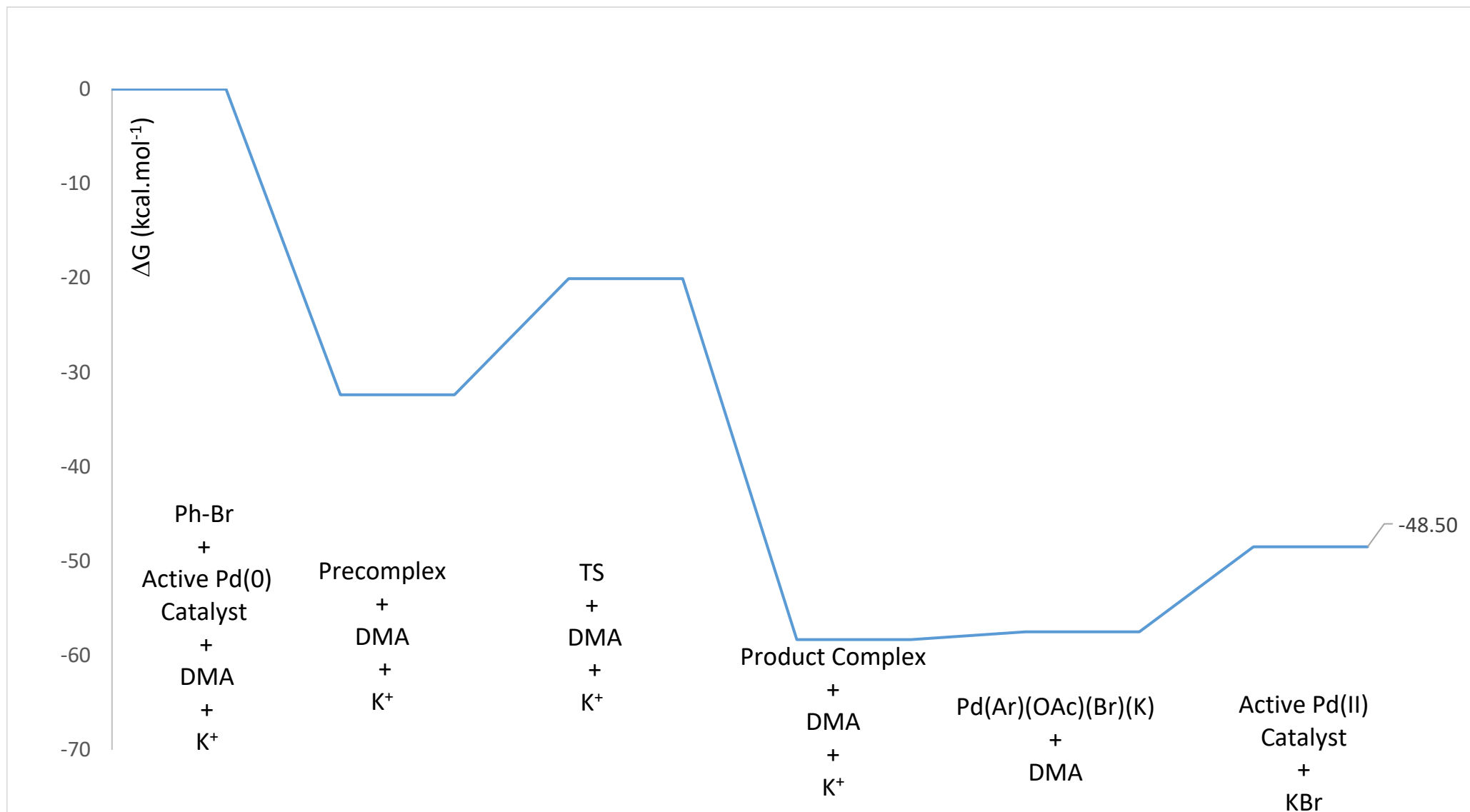


Figure 5-36: Oxidative Addition Reaction Profile – Phosphine Free Conditions (COSMO DSD-PBEP86/TZVP)

Now that OA has been considered, insight into regeneration of the active Pd(0) catalyst, the last step in the cycle, is necessary. As a reminder from Chapter 4, after OA and subsequent CMD, RE generates the relevant product and a protonated Pd(0) catalyst. This is illustrated in Figure 5-37 for Pathway =5.

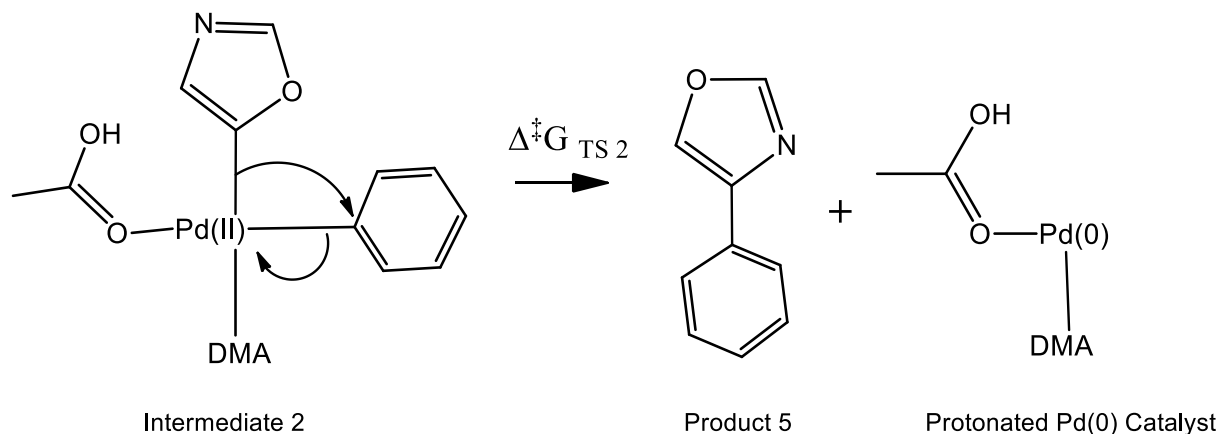


Figure 5-37: Product and Protonated Pd(0) Catalyst formation via Reductive Elimination (Pathway =5)

Under conditions of P^tBu_3 ligand, deprotonation via a hydroxide ion regenerated the active Pd(0) catalyst. However, the situation is slightly more complex under phosphine free conditions. Now deprotonation generates an anionic complex from which DMA then dissociates to regenerate the active Pd(0) catalyst (Figure 5-38). The anionic complex corresponds to the initially considered Pd(0) catalyst. At the COSMO DSD-PBEP86/TZVP level of theory, deprotonation results in an overall change of $-32.49 \text{ kcal.mol}^{-1}$ and is thus highly favorable. As in Section 5.2, no significant TS is expected. Dissociation of DMA to finally regenerate the active Pd(0) catalyst requires a small Gibbs energy input of $3.67 \text{ kcal.mol}^{-1}$, again at the COSMO DSD-PBEP86/TZVP level of theory. Neither of these processes corresponds to a significant energy barrier.

The entire catalytic cycle has now been modelled for phosphine free conditions. In conclusion, OA is not the RDS for either mono or di-substitution under phosphine free conditions. If this work is to be used as a preliminary investigation regarding optimisation of these reactions, research into lowering the energy required for OA to occur would not yield a significant increase in reaction rate or allow these reactions to be optimised for lower temperature conditions.

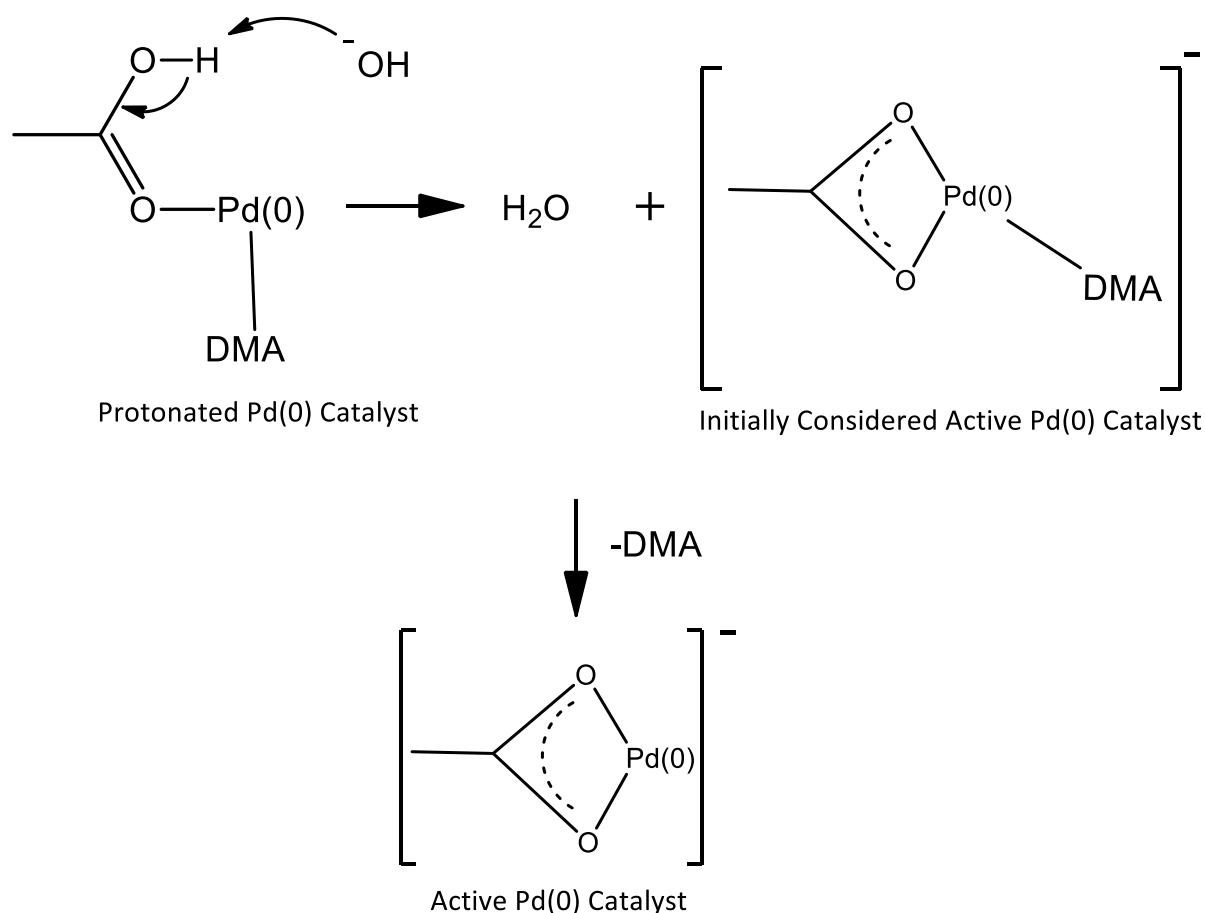


Figure 5-38: Regeneration of the Active Pd(0) Catalyst – Phosphine Free Conditions

In summary, the following conclusions can be drawn:

Conditions of 1 equivalent of P^tBu_3 :

- A COSMO solvation model dramatically changes the energy landscape of oxidative addition reactions, highlighting the sensitivity of these reactions to solvent conditions.
- The RDS of OA is formation of the reactive Pre-complex, rather than the transition state itself (9.97 vs 5.85 kcal.mol⁻¹ at the COSMO DSD-PBEP86/Def2-TZVP level of theory).
- The formation of the active catalyst is accelerated by the presence of the K^+ ion.
- OA is not the rate determining step for either mono- or di-substitution and the conclusions regarding the RDS from Chapter 4 prevail. However, OA is energetically relevant and, in most cases, is comparable in magnitude to the barrier of reductive elimination.

Phosphine free conditions:

- A DMA molecule does not appear to be coordinated to the palladium center during OA.

- Under these conditions, the formation of the pre-complex is highly favorable and the RDS of OA is formation of the TS.
- The total Gibbs free energy of the RDS for OA is higher than under phosphine conditions (12.31 vs 9.97 kcal.mol⁻¹).
- OA is not the rate determining step for either mono- or di-substitution and the conclusions regarding the RDS from Chapter 3 prevail. Again however, OA is energetically relevant and contributes significantly to the total Gibbs free energy required for the catalytic cycle, comparable in magnitude to the barrier of reductive elimination for the only kinetically feasible mono-substitution pathway (Pathway =5).

References

1. T. S. Ito, S. Hasegawa, Y. Takahashi and Y. Ishii, *Journal of Organometallic Chemistry*, 1974, 73, 401-409.
2. L. Malatesia and M. Angoletta, *Journal of the Chemical Society (Resumed)*, 1957, DOI: 10.1039/JR9570001186, 1186-1188.
3. M. S. Viciu, R. F. Germaneau, O. Navarro-Fernandez, E. D. Stevens and S. P. Nolan, *Organometallics*, 2002, 21, 5470-5472.
4. M. R. Biscoe, B. P. Fors and S. L. Buchwald, *Journal of the American Chemical Society*, 2008, 130, 6686-6687.
5. C. Amatore, A. Jutand and M. A. M'Barki, *Organometallics*, 1992, 11, 3009-3013.
6. A. M. Trzeciak and J. J. Ziółkowski, *Organometallics*, 2002, 21, 132-137.
7. F. Ozawa, A. Kubo and T. Hayashi, *Chemistry Letters*, 1992, 21, 2177-2180.
8. C. Amatore, E. Carre, A. Jutand and M. A. M'Barki, *Organometallics*, 1995, 14, 1818-1826.
9. U. Christmann and R. Vilar, *Angewandte Chemie International Edition*, 2005, 44, 366-374.
10. A. F. Littke, C. Dai and G. C. Fu, *Journal of the American Chemical Society*, 2000, 122, 4020-4028.
11. F. Schoenebeck and K. N. Houk, *Journal of the American Chemical Society*, 2010, 132, 2496-2497.
12. F. Barrios-Landeros, B. P. Carrow and J. F. Hartwig, *Journal of the American Chemical Society*, 2009, 131, 8141-8154.
13. J. Zheng, X. Xu and D. G. Truhlar, *Theoretical Chemistry Accounts*, 2011, 128, 295-305.
14. P. Deglmann and S. Schenk, *Journal of Computational Chemistry*, 2012, 33, 1304-1320.
15. A. Klamt and G. Schüürmann, *Journal of the Chemical Society, Perkin Transactions 2*, 1993, 799-805.
16. E. Lyngvi and F. Schoenebeck, *Tetrahedron*, 2013, 69, 5715-5718.
17. M. Ahlquist and P.-O. Norrby, *Organometallics*, 2007, 26, 550-553.
18. L. J. Goossen, D. Koley, H. L. Hermann and W. Thiel, *Organometallics*, 2006, 25, 54-67.
19. K. Nagayama, F. Kawataka, M. Sakamoto, I. Shimizu and A. Yamamoto, *Chemistry Letters*, 1995, 24, 367-368.
20. R. Kakino, H. Narahashi, I. Shimizu and A. Yamamoto, *Bulletin of the Chemical Society of Japan*, 2002, 75, 1333-1345.
21. A. Jutand, S. Négri and J. G. de Vries, *European Journal of Inorganic Chemistry*, 2002, 2002, 1711-1717.
22. Z. Li, Y. Fu, Q.-X. Guo and L. Liu, *Organometallics*, 2008, 27, 4043-4049.
23. F. Proutiere and F. Schoenebeck, *Angewandte Chemie International Edition*, 2011, 50, 8192-8195.
24. S. Verbeeck, C. Meyers, P. Franck, A. Jutand and B. U. Maes, *Chemistry-A European Journal*, 2010, 16, 12831-12837.
25. A. H. Roy and J. F. Hartwig, *Organometallics*, 2004, 23, 194-202.
26. L. J. Goossen, D. Koley, H. L. Hermann and W. Thiel, *Organometallics*, 2005, 24, 2398-2410.
27. B. J. Lynch, Y. Zhao and D. G. Truhlar, *The Journal of Physical Chemistry A*, 2003, 107, 1384-1388.
28. A. Albert and E. Serjeant, *Khimiya, Moscow*, 1964, 126, 145.
29. M. T. Reetz and M. Maase, *Advanced Materials*, 1999, 11, 773-777.
30. A. H. M. de Vries, F. J. Parlevliet, L. Schmieder-van de Vondervoort, J. H. M. Mommers, H. J. W. Henderickx, M. A. M. Walet and J. G. de Vries, *Advanced Synthesis & Catalysis*, 2002, 344, 996-1002.

6. Conclusions

- The proposed CMD mechanism correctly predicted the major products observed experimentally under phosphine free conditions – Product 5 and Product 2,4 for mono-substitution and di-substitution, respectively. These results confirm the operation of this mechanism under these conditions.
- The proposed CMD mechanism correctly predicted Product 5 to be the major mono-substituted product under conditions of 1 equivalent of P^tBu_3 ligand. This result confirms this mechanism prevails under these conditions.
- Although, the major di-substituted product observed experimentally (Product 2,5) was predicted to be the second most favourable product kinetically, its dominance can be readily explained from the data considering the equilibrium established during mono-substitution. Thus, it is highly likely that here too, the CMD mechanism prevails.
- The TPSSh DFT functional, used in conjunction with a SVP basis set, provided adequate geometries and preliminary energies for the CMD reactions considered.
- No additional insight was gain by employing a COSMO solvation model to approximate the effects of DMA solvent for CMD reactions.
- The TPSSh/SVP level of theory, however, does not provide an accurate measure of the RDS for the CMD reactions considered.
- The double hydride functional DSD-PBEP86, used in conjunction with a TZVP basis set, was shown to correctly predict the RDS for the CMD reactions considered. It was demonstrated that this level of theory provides the best compromise between efficiency and accuracy for these systems.
- Although using a DH functional for these systems is essential, greater quantitative insight into the energetics of the CMD mechanism is provided by the DLPNO-CCSD(T)/TZVP level of theory. Should the resources be available, it is recommended over DSD-PBEP86/TZVP.
- Unlike CMD reactions, the use of a COSMO solvation model was shown to be vital when modelling OA reactions. This is primarily due to the inherent inaccuracies introduced by modelling charged species in the gas phase.
- The use of a minimally augmented basis set was shown to have a negligible effect on the modelling of OA reactions.
- OA under conditions of P^tBu_3 ligand was shown to be significantly more favourable than under phosphine free conditions, correlating well with widespread experimental observations.
- COSMO DSD-PBEP86/TZVP was shown to provide an adequate compromise between efficiency and accuracy for OA reactions, however should the resources be available, it

is again recommended to employ DLPNO-CCSD(T)/TZVP calculations in conjunction with a solvation model.

- Under phosphine free conditions, the RDS of the entire catalytic cycle for both mono- and di-substitution, for all kinetically relevant pathways is the CMD TS.
- Under conditions of 1 equivalent of P^tBu_3 ligand the RDS, for both mono- and di-substitution, for all kinetically relevant pathways is a mixture of both the CMD TS as well as the formation of the CMD Intermediate 1 i.e. coordination of oxazole to the active Pd(II) catalyst.

The majority of the original aims of the investigation have thus been met. The CMD step has been identified as a target which future research may optimise in order to develop superior synthetic routes. Unfortunately, due to time constraints and the length of this present work, the cooperative catalysis mechanism was not studied and must be relegated to future work.

7. Future work

- In order to investigate why the experimental product distributions changed when the reactions were conducted with two equivalents of P^tBu_3 , an investigation into the cooperative catalysis mechanism is essential. As the mechanism involves two equivalents of P^tBu_3 ligand, it offers the very real potential of providing mechanistic insight into the observed changes in product distributions.
- This work focused on a small subset of the direct arylation experiments conducted in the research laboratories of Johnson Matthey. In-fact, direct arylation of oxazole with 4-Bromotoluene was conducted with a wide variety of catalytic systems. Future work could include an investigation of catalysts such as $PdCl_2(PPh_3)_2$, $Pd(OAc)_2/PCy_3$ as well $PdCl_2(PCy_3)_2$ to confirm if the results of this study can be generalised.
- A dramatic effect was observed when the COSMO solvation model was applied to OA reactions. This leads to the question of whether newer solvation models could offer greater insight into how the choice of solvent affects OA. Future work could involve re-running calculations with the more advanced SMD solvation model.¹
- Throughout the present work, DLPNO-CCSD(T)/TZVP was used as a single point energy benchmark. Although TZVP is a large basis set, it is not expected to be near the basis set convergence limit for the molecules considered. Future work could include DLPNO-CCSD(T) calculations in conjunction with the much larger Def2-QZVPP basis set, in order to minimise basis set superposition errors when benchmarking. Even more accurate would be to use the automatic extrapolation to the basis set limit feature of ORCA. Although this would be very computationally demanding, the resulting benchmark energy values would be highly accurate.
- During the modelling of OA it was discovered that the K^+ cation could accelerate the formation of the active $Pd(II)$ catalyst by facilitating Br^- dissociation. Future work could involve modelling other cations facilitating this dissociation, in an attempt to further optimise this final step in the formation of the active $Pd(II)$ catalyst.

References

1. A. V. Marenich, C. J. Cramer and D. G. Truhlar, *The Journal of Physical Chemistry B*, 2009, 113, 6378-6396.

Appendix A

1. Details of Selected Atoms from the def2-SVP Basis Set

H

(4s1p) / [2s1p]	{31/1}
α	C_a
3 s	
13.01070	0.01968
1.96226	0.13797
0.44454	0.47832
1 s	
0.12195	1.00000
1 p	
0.80000	1.00000

C

(7s4p1d) / [3s2p1d]	{511/31/1}
α	C_a
5 s	
1238.40169	0.00546
186.29005	0.04064
42.25118	0.18026
11.67656	0.46315
3.59305	0.44087
1 s	
0.40245	1.00000
1 s	
0.13090	1.00000
3 p	
9.46810	0.03839
2.01035	0.21117
0.54771	0.51328
1 p	
0.15269	1.00000
1 d	
0.80000	1.00000

N

(7s4p1d) / [3s2p1d]	{511/31/1}
α	C_a
5 s	
1712.84159	-0.00539
257.64813	-0.04022
58.45825	-0.17931
16.19837	-0.46376
5.00526	-0.44171
1 s	
0.58732	1.00000
1 s	
0.18765	1.00000
3 p	
13.57147	-0.04007
2.92574	-0.21807
0.79928	-0.51294
1 p	
0.21954	1.00000
1 d	
1.00000	1.00000

O

(7s4p1d) / [3s2p1d]	{511/31/1}
α	C_a
5 s	
2266.17678	-0.00534
340.87010	-0.03989
77.36314	-0.17854
21.47964	-0.46428
6.65894	-0.44310
1 s	
0.80976	1.00000
1 s	
0.25531	1.00000
3 p	
17.72150	0.04339
3.86355	0.23094
1.04809	0.51375
1 p	
0.27642	1.00000
1 d	
1.20000	1.00000

<u>P</u>		<u>Br</u>	
(10s7p1d) / [4s3p1d]	{5311/511/1}	(14s10p6d) / [5s4p3d]	{63311/5311/411}
α	C_a	α	C_a
5 s		6 s	
8002.47951	0.00575	113286.3878	1.42830E-03
1203.68136	0.04301	17009.6263	1.09504E-02
273.44227	0.19364	3870.184257	5.44210E-02
76.65554	0.49652	1093.035723	0.190479077
23.51693	0.44983	356.397218	0.390246427
3 s		123.1253964	0.308144325
39.79168	0.09519	3 s	
4.27703	-0.57650	236.7408401	-0.112280657
1.69403	-0.54240	28.46866107	0.647759623
1 s		11.88344372	0.44235576
0.27568	1.00000	3 s	
1 s		21.26963331	-0.226425763
0.10496	1.00000	3.612922684	0.73823712
5 p		1.662664897	0.426838687
219.50756	0.00921	1 s	
51.27416	0.06541	0.348237932	1.00000
15.92160	0.24034	1 s	
5.50699	0.46318	0.130190314	1.00000
1.95377	0.37393	5 p	
1 p		1560.280188	8.71667E-03
0.47803	1.00000	368.4785921	6.62436E-02
1 p		117.2297885	0.264956104
0.13658	1.00000	42.64890925	0.538391606
1 d		16.0872251	0.365793879
0.45000	1.00000	1 p	
		0.530642948	1.00000
		1 p	
		0.15702759	1.00000
		4 d	
		104.8551864	2.26501E-02
		30.28114369	0.134554832
		10.65139427	0.364744545
		3.869945623	0.490445871
		1 d	
		1.324087676	0.27137289
		1 d	
		0.389	1.00000

Pd

(7s6p5d1f) / [5s3p2d1f]	{31111/411/41/1}
α	C_a
3 s	
8.47564	1.23924
7.16572	-1.65631
3.18211	-0.13179
1 s	
1.39736	1.00000
1 s	
0.60533	1.00000
1 s	
0.10230	1.00000
1 s	
0.03734	1.00000
4 p	
4.24610	-0.82324
3.39259	0.87085
1.19759	0.63512
0.52674	0.27225
1 p	
0.20607	1.00000
1 p	
0.03700	1.00000
4 d	
7.36133	-0.01720
2.62910	0.23314
1.12927	0.44536
0.44472	0.39294
1 d	
0.15545	1.00000
1 f	
1.24629	1.00000

2. Details of Selected Atoms from the def2-TZVP Basis Set

H

(5s1p) / [3s1p]	{311/1}
α	C_a
3 s	
34.06134	0.00603
5.12357	0.04502
1.16466	0.20190
1 s	
0.32723	1.00000
1 s	
0.10307	1.00000
1 p	
0.80000	1.00000

C

(11s6p2d1f) / [5s3p2d1f]	{62111/411/11/1}
α	C_a
6 s	
13575.34968	0.00022
2035.23337	0.00172
463.22562	0.00893
131.20020	0.03573
42.85302	0.11076
15.58419	0.24296
2 s	
6.20671	0.41440
2.57649	0.23745
1 s	
0.57696	1.00000
1 s	
0.22973	1.00000
1 s	
0.09516	1.00000
4 p	
34.69723	0.00533
7.95826	0.03586
2.37808	0.14216
0.81433	0.34270
1 p	
0.28888	0.46446
1 p	
0.10057	0.24956
1 d	
1.09700	1.00000
1 d	
0.31800	1.00000
1 f	
0.76100	1.00000

N

(11s6p2d1f) / [5s3p2d1f]	{62111/411/11/1}
α	C_a
6 s	
19730.80065	0.00022
2957.89587	0.00170
673.22134	0.00880
190.68249	0.03536
62.29544	0.11096
22.65416	0.24983
2 s	
8.97915	0.40624
3.68630	0.24338
1 s	
0.84660	1.00000
1 s	
0.33647	1.00000
1 s	
0.13648	1.00000
4 p	
49.20038	0.00556
11.34679	0.03805
3.42740	0.14954
1.17855	0.34949
1 p	
0.41642	0.45843
1 p	
0.14261	0.24429
1 d	
1.65400	1.00000
1 d	
0.46900	1.00000
1 f	
1.09300	1.00000

O

(11s6p2d1f) / [5s3p2d1f]	{62111/411/11/1}
α	C_a
6 s	
27032.38263	0.00022
4052.38714	0.00168
922.32723	0.00874
261.24071	0.03524
85.35464	0.11154
31.03504	0.25589
2 s	
12.26086	0.39769
4.99871	0.24628
1 s	
1.17031	1.00000
1 s	
0.46475	1.00000
1 s	
0.18505	1.00000
4 p	
63.27495	0.00607
14.62705	0.04191
4.45012	0.16154
1.52758	0.35707
1 p	
0.52935	0.44794
1 p	
0.17478	0.24446
1 d	
2.31400	1.00000
1 d	
0.64500	1.00000
1 f	
1.42800	1.00000

<u>P</u>	
(14s9p3d1f) / [5s5p2d1f]	{73211/51111/21/1}
α	C_a
7 s	
52426.99923	0.00055
7863.26606	0.00427
1789.52273	0.02193
506.27300	0.08567
164.60699	0.24841
58.39192	0.46337
21.64366	0.35351
3 s	
99.01384	0.02190
30.55044	0.09565
5.45371	-0.29454
2 s	
2.65034	1.32944
1.27267	0.66109
1 s	
0.31645	1.00000
1 s	
0.11417	1.00000
5 p	
472.27219	0.00257
111.58883	0.02025
35.44594	0.09158
12.99078	0.25749
5.04862	0.42863
1 p	
1.99340	0.34360
1 p	
0.66527	1.00000
1 p	
0.25517	1.00000
1 p	
0.09036	1.00000
2 d	
3.12000	0.20000
0.64800	1.00000
1 d	
0.21800	1.00000
1 f	
0.45200	1.00000

Br

(17s13p8d1f) / [6s5p4d1f]		{842111/64111/5111/1}	
α	C_a	α	C_a
8 s		5 d	
565073.25256	0.00024	168.85370	0.00897
84701.72318	0.00183	49.97795	0.06206
19276.27190	0.00955	18.27491	0.21475
5456.42846	0.03888	7.24557	0.40335
1776.95035	0.12718	2.85623	0.42209
639.19398	0.30438	1 d	
248.78824	0.44491	1.04596	0.17875
98.67831	0.24382	1 d	
4 s		0.56866	1.00000
606.07825	-0.02653	1 d	
188.45598	-0.12485	0.22031	1.00000
31.49714	0.56469	1 f	
13.73601	0.55555	0.57083	1.00000
2 s			
21.20321	-0.24941		
3.76164	0.71213		
1 s			
1.77359	1.00000		
1 s			
0.45197	1.00000		
1 s			
0.16613	1.00000		
6 p			
3019.69557	0.00250		
715.35481	0.02042		
229.98329	0.09690		
86.16784	0.28054		
34.66787	0.44606		
14.11387	0.24410		
4 p			
57.08565	-0.02186		
8.81938	0.32707		
3.93403	0.57855		
1.79988	0.33571		
1 p			
0.66899	1.00000		
1 p			
0.27136	1.00000		
1 p			
0.10084	1.00000		

Pd

(7s7p6d1f) / [6s4p3d1f]	{211111/4111/411/1}
α	C_a
2 s	
18.00000	-0.16605
14.66213	0.34900
1 s	
5.63887	1.00000
1 s	
1.31990	1.00000
1 s	
0.57818	1.00000
1 s	
0.10352	1.00000
1 s	
0.03755	1.00000
4 p	
12.55290	0.06173
7.24445	-0.24179
1.89059	0.49453
0.90737	0.50454
1 p	
0.40877	1.00000
1 p	
0.11500	1.00000
1 p	
0.03700	1.00000
4 d	
22.35746	0.00396
10.68253	-0.01404
2.48582	0.24219
1.07353	0.42580
1 d	
0.42614	1.00000
1 d	
0.15046	1.00000
1 f	
1.24629	1.00000

3. Details of Atoms from the ma-Def2-SVP Basis Set

H

(4s1p) / [2s1p]	{31/1}
α	C_a
3 s	
13.01070	0.01968
1.96226	0.13797
0.44454	0.47832
1 s	
0.12195	1.00000
1 p	
0.80000	1.00000

C

(8s5p1d) / [4s3p1d]	{5111/311/11}
α	C_a
5 s	
1238.40169	0.00546
186.29005	0.04064
42.25118	0.18026
11.67656	0.46315
3.59305	0.44087
1 s	
0.40245	1.00000
1 s	
0.13090	1.00000
3 p	
9.46810	0.03839
2.01035	0.21117
0.54771	0.51328
1 p	
0.15269	1.00000
1 d	
0.80000	1.00000
1 s	
0.04363	1.00000
1 p	
0.05089	1.00000

N

(8s5p1d) / [4s3p1d]	{5111/311/1}
α	C_a
5 s	
1712.84159	-0.00539
257.64813	-0.04022
58.45825	-0.17931
16.19837	-0.46376
5.00526	-0.44171
1 s	
0.58732	1.00000
1 s	
0.18765	1.00000
3 p	
13.57147	-0.04007
2.92574	-0.21807
0.79928	-0.51294
1 p	
0.21954	1.00000
1 d	
1.00000	1.00000
1 s	
0.06255	1.00000
1 p	
0.07318	1.00000

O

(8s5p1d) / [3s2p1d]	{5111/311/1}
α	C_a
5 s	
2266.17678	-0.00534
340.87010	-0.03989
77.36314	-0.17854
21.47964	-0.46428
6.65894	-0.44310
1 s	
0.80976	1.00000
1 s	
0.25531	1.00000
3 p	
17.72150	0.04339
3.86355	0.23094
1.04809	0.51375
1 p	
0.27642	1.00000
1 d	
1.20000	1.00000
1 s	
0.08510	1.00000
1 p	
0.09214	1.00000

P

(11s8p1d) / [5s4p1d]	{53111/5111/1}
α	C_a
5 s	
8002.47951	0.00575
1203.68136	0.04301
273.44227	0.19364
76.65554	0.49652
23.51693	0.44983
3 s	
39.79168	0.09519
4.27703	-0.57650
1.69403	-0.54240
1 s	
0.27568	1.00000
1 s	
0.10496	1.00000
5 p	
219.50756	0.00921
51.27416	0.06541
15.92160	0.24034
5.50699	0.46318
1.95377	0.37393
1 p	
0.47803	1.00000
1 p	
0.13658	1.00000
1 d	
0.45000	1.00000
1 s	
0.03499	1.00000
1 p	
0.04553	1.00000

Br

(15s11p6d) / [6s5p3d]	{633111/53111/411}
α	C_a
6 s	
113286.3878	1.42830E-03
17009.6263	1.09504E-02
3870.184257	5.44210E-02
1093.035723	0.190479077
356.397218	0.390246427
123.1253964	0.308144325
3 s	
236.7408401	-0.112280657
28.46866107	0.647759623
11.88344372	0.44235576
3 s	
21.26963331	-0.226425763
3.612922684	0.73823712
1.662664897	0.426838687
1 s	
0.348237932	1.00000
1 s	
0.130190314	1.00000
5 p	
1560.280188	8.71667E-03
368.4785921	6.62436E-02
117.2297885	0.264956104
42.64890925	0.538391606
16.0872251	0.365793879
1 p	
0.530642948	1.00000
1 p	
0.15702759	1.00000
4 d	
104.8551864	2.26501E-02
30.28114369	0.134554832
10.65139427	0.364744545
3.869945623	0.490445871
1 d	
1.324087676	0.27137289
1 d	
0.389	1.00000
1 s	
0.04340	1.00000
1 p	
0.05234	1.00000

Pd

(7s6p5d1f) / [5s3p2d1f]	{31111/411/41/1}
α	C_a
3 s	
8.47564	1.23924
7.16572	-1.65631
3.18211	-0.13179
1 s	
1.39736	1.00000
1 s	
0.60533	1.00000
1 s	
0.10230	1.00000
1 s	
0.03734	1.00000
4 p	
4.24610	-0.82324
3.39259	0.87085
1.19759	0.63512
0.52674	0.27225
1 p	
0.20607	1.00000
1 p	
0.03700	1.00000
4 d	
7.36133	-0.01720
2.62910	0.23314
1.12927	0.44536
0.44472	0.39294
1 d	
0.15545	1.00000
1 f	
1.24629	1.00000

4. Details of Selected Atoms from def2 Coulomb Fitting Basis Set

Note that this auxiliary Coulomb fitting set is a universal auxiliary that can be used for the following canonical basis sets:

def2-SV(P)

def2-SVP

def2-TZVP

def2-TZVPP

def2-QZVP

def2-QZVPP

H

(5s2p1d) / [3s1p1d]	{311/2/1}
α	C_a
3 s	
15.6752927	0.0186886
3.6063578	0.063167
1.2080016	0.1204609
1 s	
0.4726794	0.0592485
1 s	
0.20181	0.0051272
2 p	
2.0281365	1.000000
0.535873	1.000000
1 d	
2.2165124	0.0033116

C

α	C_a
(12s5p4d2f1g) / [6s4p3d1f1g]	{7111111/2111/211/2/1}
7 s	
1861.091633	0.0744171
642.9939764	0.1653957
235.1105725	0.5576484
90.7028894	1.3108298
36.7794552	2.1694681
15.6046273	1.7668846
6.8907294	0.2930769
1 s	
3.147885	-0.1708702
1 s	
1.4777287	0.1641553
1 s	
0.7076466	0.4149941
1 s	
0.3430122	0.1624366
1 s	
0.1669453	0.0207675
2 p	
13.5472892	-0.0206477
5.4669419	-0.0115282
1 p	
2.1751721	0.0455914
1 p	
0.8582194	0.002836
1 p	
0.337672	0.0181875
2 d	
5.9287253	-0.0225948
1.9809209	-0.0476827
1 d	
0.8055417	-0.0365372
1 d	
0.3531244	-0.0145417
2 f	
1.6755626	0.0088798
0.5997536	0.0069903
1 g	
1.00246	-0.0022192

N	
(12s5p4d2f1g) / [6s4p3d1f1g]	{711111/2111/211/2/1}
α	C_a
7 s	
2542.940179	0.0904668
1029.547295	0.1455675
424.9053546	0.459409
178.4573708	1.1392859
76.1362742	2.1795891
32.9338653	2.7604305
14.4155436	1.2837451
1 s	
6.3718891	-0.1702642
1 s	
2.8381742	0.0504472
1 s	
1.2711793	0.6634588
1 s	
0.5712407	0.4116559
1 s	
0.2569887	0.0489167
2 p	
17.897599	-0.0330656
6.5346319	0.0041582
1 p	
2.3523645	0.0525824
1 p	
0.8397377	0.0017058
1 p	
0.2989335	0.012296
2 d	
7.690987	0.003582
2.3044972	0.0180403
1 d	
0.8403987	0.0194441
1 d	
0.3303797	0.0002361
2 f	
2.3331076	0.0067248
0.8446137	0.0054427
1 g	
1.40377	-0.0023051

Q

(12s5p4d2f1g) / [6s4p3d1f1g]	{7111111/2111/211/2/1}
α	C_a
7 s	
2876.821661	0.1443558
1004.744303	0.2920041
369.7579954	1.0258517
142.9442404	2.2875516
57.8366425	3.6080237
24.3864983	2.3737865
10.6622662	0.0489414
1 s	
4.8070437	-0.1295186
1 s	
2.221077	0.7747158
1 s	
1.0447795	0.7647816
1 s	
0.4968425	0.2369803
1 s	
0.2371384	0.0208099
2 p	
64.2613382	-0.0126659
16.3006076	-0.0378744
1 p	
4.3550542	0.0638078
1 p	
1.2019554	0.0169516
1 p	
0.3354196	0.0065743
2 d	
9.2146611	-0.0597914
2.8435251	-0.0846724
1 d	
0.9955759	-0.0466066
1 d	
0.3649441	-0.0096978
2 f	
2.6420115	1.00000
0.7345613	1.00000
1 g	
1.3931	-0.0016533

P

(14s5p5d2f1g) / [8s4p3d1f1g]		{711111111/2111/311/2/1}	
α	C_a	α	C_a
7 s		1 g	
7402.640563	0.6312551	0.5510844	-0.0032
2444.38063	1.0410015		
1031.048528	3.5177235		
387.5215861	7.8577024		
159.3342518	9.3323916		
72.5438912	3.5808385		
29.2976155	-1.103988		
1 s			
13.1685047	2.8706692		
1 s			
6.4235268	4.9937953		
1 s			
3.1479365	1.200375		
1 s			
1.4941282	-0.4895877		
1 s			
0.6745215	0.3794973		
1 s			
0.3252877	0.362504		
1 s			
0.1527285	0.0512179		
2 p			
10.6100703	0.0333958		
3.3972489	-0.0239301		
1 p			
1.0821336	0.0194214		
1 p			
0.3436265	0.0184052		
1 p			
0.1090044	0.0008768		
3 d			
5.5575593	0.00664		
2.603324	-0.0123559		
1.2378589	0.0087029		
1 d			
0.5939646	0.007218		
1 d			
0.2858732	-0.0025575		
2 f			
1.0262041	0.0104661		
0.2959414	0.0038742		

Br

(19s5p5d3f1g) / [8s4p3d2f1g]		{(12)1111111/2111/311/21/1}	
α	C_a	α	C_a
12 s		1 d	
181929.6907	0.970382	0.6328484	-0.0104
21782.50624	3.1341819	1 d	
7843.239172	11.0443801	0.2711415	-0.0138
2608.027178	24.6916778	2 f	
1564.969647	12.3123948	3.4681858	-0.003
920.5703545	24.6284398	0.8708007	0.01241
552.396338	16.2668646	1 f	
191.1405907	-3.1865943	0.2186427	0.00049
108.0484426	5.1797612	1 g	
64.8354183	25.8419691	1.1078575	-0.0027
38.1384803	15.577103		
22.4343995	-9.1953776		
1 s			
13.1967052	-1.3048405		
1 s			
7.9187991	8.573887		
1 s			
4.658117	9.2928352		
1 s			
2.6331523	0.5617368		
1 s			
0.557718	0.8600521		
1 s			
0.3152683	0.2596004		
1 s			
0.1854519	0.0721463		
2 p			
8.4481854	0.013469		
3.0812881	-0.0389777		
1 p			
1.1536778	0.0500453		
1 p			
0.4389478	-0.008797		
1 p			
0.1679158	0.0062473		
3 d			
8.7933629	-0.0194411		
3.5598266	0.0201204		
1.4866514	-0.0158267		

Pd

(11s5p5d3f3g) / [8s5p5d2f3g]		{41111111/11111/11111/21/111}	
α	C_a	α	C_a
4 s		1 f	
51.0211803	0.0538811	0.3451253	0.0027275
17.6543867	-1.3608599	1 g	
10.384933	5.9322863	7.5375072	0.0047006
6.2315704	-7.8659686	1 g	
1 s		2.7664349	0.0340131
3.6656295	-1.3215872	1 g	
1 s		1.0133462	0.0097463
2.1995933	4.1735311		
1 s			
1.2683839	2.2689133		
1 s			
0.7461081	0.671248		
1 s			
0.4477087	0.3367069		
1 s			
0.2581689	0.0896295		
1 s			
0.151864	0.0380638		
1 p			
5.6999676	0.0141426		
1 p			
2.5155374	-0.0569343		
1 p			
1.1229524	0.0732647		
1 p			
0.5047808	-0.0021359		
1 p			
0.2274325	-0.0233875		
1 d			
3.8006002	0.0012299		
1 d			
1.5882692	-0.0485992		
1 d			
0.6550157	-0.0041355		
1 d			
0.268021	0.003165		
1 d			
0.1093843	-0.0006315		
2 f			
2.0106185	-0.0150205		
0.826846	0.010533		

5. Selected COSMO Cavity Radii

COSMO Cavity Radii

Element	Radii (Bohr)
H	2.4566
C	3.7795
N	3.4582
O	3.2503
Pd	4.2009

Appendix B-ESI

Explained

Electronic Supplementary Information

The Electronic Supplementary Information (ESI) for this thesis is located in:

Thesis\ESI

Each chapter which involved modelling has a dedicated ESI folder:

For Chapter 3: ESI\ESI-Chapter 3

For Chapter 4: ESI\ESI-Chapter 4

For Chapter 5: ESI\ESI-Chapter 5

These folders contain all relevant output files for the chapter. It should be noted that embedded in each output file is also a copy of the input file for each job. This is default ORCA behaviour.

In order to aid in the navigation of the ESI, the Thesis.pdf document has embedded hyperlinks on all figure captions that correspond to a xyz file. By simply clicking on the figure caption, the exact location in the ESI will open automatically in Windows Explorer, allowing the reader instant access to all files concerning the relevant molecule.

This functionality however will be lost if the reader changes the relative location of files inside the “Thesis” folder. The folder itself can be download and moved anywhere on the readers’ PC but the contents should be kept as is and files not moved around within the folder or to other locations. It has been assumed that the examiner will be using a Windows operating system. However, if the examiner is using a Linux operating system a version of this document with compatible paths can be arranged timeously.

It should be noted that as many calculations were not illustrated as figures in the text, but as points in a graph or as values in a table etc., links contained in figure captions do not represent the entire content of the ESI, and there are many more files provided that have no links.

If the figure refers to a structure taken from literature, then the path contains the .xyz file and possibly also a .cif file. If the figure refers to a structure from the project (assuming it is shown in the figure for the first time), then the path will contain a “Gas Phase” folder and possibly also a “COSMO” folder. These contain output files from Gas phase and COSMO calculations, respectively. Depending on the type of calculation performed various files are provided as expounded in Table 1.

Table 1: ESI File Descriptions

File Name	Description
TPSSh_Geo_Opt.out	ORCA geometry optimization output file (TPSSh/SVP).
TPSSh_Optimized_Geo.xyz	xyz file containing the optimized geometry (TPSSh/SVP).
TPSSh_Freq.out	ORCA frequency calculation output file (TPSSh/SVP).
TPSSh_Geo_Opt_Freq.out	ORCA output file for a compound geometry optimization and frequency calculation (TPSSh/SVP).
Corrected_G.txt	A text file containing the number of the point group of the molecule as well as the symmetry corrected Gibbs free energy. This is the final Gibbs free energy reported in the thesis. (TPSSh/SVP).
DH.out	ORCA output file for a single point energy DSD-PBEP86/TZVP calculation.
CC.out	ORCA output file for a single point energy DLPNO-CCSD(T)/TZVP calculation
COSMO_TPSSh_Geo_Opt.out	ORCA geometry optimization output file performed in conjunction with the COSMO solvation model using DMA solvent (COSMO TPSSh/SVP).

COSMO_TPSSh_Optimized_Geo.xyz	xyz file containing the optimized geometry (COSMO TPSSh/SVP).
COSMO_TPSSh_Freq.out	ORCA frequency calculation output file performed in conjunction with the COSMO solvation model using DMA solvent (COSMO TPSSh/SVP).
COSMO_TPSSh_Geo_Opt_Freq.out	ORCA output file for a compound geometry optimization and frequency calculation performed in conjunction with the COSMO solvation model using DMA solvent. (COSMO TPSSh/SVP)
COSMO_Corrected_G.txt	A text file containing the number of the point group of the molecule as well as the symmetry corrected Gibbs free energy. This is the final Gibbs free energy reported in the thesis. (COSMO TPSSh/SVP)
BP86_Geo_Opt.out	ORCA geometry optimization output file (BP86/SVP).
COSMO_TPSSh_ma_Geo_Opt_Freq.out	ORCA output file for a compound geometry optimization and frequency calculation using a minimally augmented basis set. performed in conjunction

	with the COSMO solvation model using DMA solvent. (COSMO TPSSh/ ma-SVP)
COSMO_TPSSh_ma_Optimized_Geo.xyz	xyz file containing the optimized geometry (COSMO TPSSh/ma-SVP).
COSMO_ma_Corrected_G.txt	A text file containing the number of the point group of the molecule as well as the symmetry corrected Gibbs free energy. This is the final Gibbs free energy reported in the thesis. (COSMO TPSSh/ma-SVP)
TPSSh_ma_SP.out	ORCA output file for a single point energy calculation using a minimally augmented basis set. (TPSSh/ma-SVP)
COSMO_DH_EWin.out	ORCA output file for a single point energy DSD-PBEP86/TZVP calculation performed in conjunction with the COSMO solvation model using DMA solvent. The core electrons were determined using an energy window (Ewin).
COSMO_CC_EWin.out	ORCA output file for a single point energy DLPNO-CCSD(T)/TZVP calculation performed in conjunction with the COSMO solvation model using DMA solvent.

	The core electrons were determined using an energy window (Ewin).
--	---

Although any computational chemistry visualization package may be used to examine the .xyz files, it is recommended to use the ChemCraft program¹ to open ORCA output files. In the "Thesis\ESI" folder a Windows installer is provided for the latest version of this program – "Chemcraft_b486_win.exe" as well as a version for Linux – "Chemcraft_b94_lin.tar.gz". Each of these files contains the fully functional version of Chemcraft with a trial period of 150 days. According to the authors, the Linux version of Chemcraft has some significant disadvantages. If using Windows, please note that, again according to the authors of the program, it is not recommended to install Chemcraft into the "C:\Program files\Chemcraft" folder, as this can cause problems.

References:

1. G. A. Andrienko., <http://www.chemcraftprog.com>.