A COMPUTATIONAL INVESTIGATION OF THE REACTION MECHANISM OF ARISTOLOCHENE SYNTHASE

Neil James Young

A thesis submitted to

Cardiff University

in accordance with the requirements of the degree of

Doctor of Philosophy

School of Chemistry

Cardiff University

September 2009

UMI Number: U584434

UMI
Dissertation Publishing

ProQuest

DECLARATION

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

Signed .... (candidate) Date. 29 SEPT 09

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD
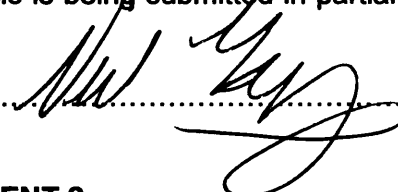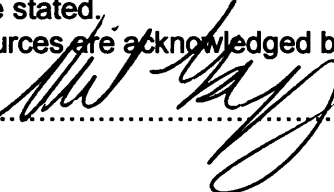
Signed .... (candidate) Date 29 SEPT 04
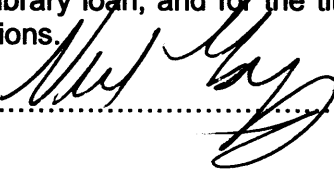
STATEMENT 2

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

Signed .... (candidate) Date 29 SEPT 04

STATEMENT 3

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

Signed .... (candidate) Date 29 SEPT 04

ABSTRACT

A COMPUTATIONAL INVESTIGATION OF THE REACTION MECHANISM OF ARISTOLOCHENE SYNTHASE

Neil James Young

Gas phase *ab initio*, semi-empirical and density functional theory (DFT) quantum mechanical (QM) calculations are used to investigate key steps in the conversion of farnesyl diphosphate to aristolochene by the sesquiterpene synthase, aristolochene synthase (AS-PR), from *P. roqueforti*. Molecular docking and a combined quantum mechanics / molecular dynamics (QM/MM)are then used to simulate the reaction *in silico*. A total of three models of AS-PR are considered, one with two magnesium ions docked and two models containing three magnesiums ions docked in different positions.

Gas phase results indicate that an intramolecular proton transfer, either directly or *via* a water molecule, during a key step in the mechanism is feasible with a maximum barrier height of *ca.* 22 kcal mol^{-1}. Such a mechanism would avoid the generation, quenching and regeneration of a high energy carbocation intermediate. However, experimental evidence obtained contemporaneously with these results suggest that AS does indeed utilise an alternative mechanism involving the formation of a neutral intermediate (germacrene A) but the identity of the acid capable of protonating this intermediate has yet to be

determined. Molecular docking experiments performed in this work reject a previous suggestion that diphosphate ion can perform this role.

A series of QM/MM free energy sampling simulations demonstrate no significant lowering of the energetic barrier to intramolecular proton transfer over the gas phase suggesting that AS-PR does not catalyse this step of the reaction mechanism *in vivo*. Additional data also seemingly rule out Lys 206 as a possible active site acid. The condensed phase data also support gas phase results which suggest the formation of intermediate eudesmane cation occurs *via* a concerted process but that subsequent steps are not concerted and that there are are further true on-path intermediates in the reaction mechanism catalysed by AS-PR.

By considering a number of different models for the AS-PR holoenzyme the role of the magnesium in causing favourable substrate binding and acting as a 'trigger' for the cyclisation cascade are presented and computational results support the tentative evidence obtained by X-ray crystallography.

The results presented in this work add weight to the developing theories in sesquiterpene chemistry in which the biosynthesis of these molecules is achieved by enzymes whose role is the act as chaperones guiding their universal substrate into appropriate conformation, triggering catalysis by the binding of magnesium and preventing premature quenching of reactive carbocations by solvent.

ACKNOWLEDGMENTS

Contents

List of Figures

List of Tables

List of Abbreviations

ABNR Adopted Basis Newton-Raphson

AS-AT Aristolochene synthase from *A. terreus*

AS-PR Aristolochene synthase from *P. roqueforti*

ATP Adenosine triphosphate

CHARMM Chemistry at Harvard Macromolecular Mechanics

DCS δ cadinene synthase

DFT Density Functional Theory

DMAP dimethylallyl pyrophosphate

DOXP 1-deoxy-D-xylulose 5-phosphate

EAS 5-*epi*-aristolochene synthase

FPP Farnesyl diphosphate

G3P Glyceraldehyde 3-phosphate

GTO Gaussian Type Orbital

HF Hartree Fock

HMG-CoA 3-hydroxy-3-methylglutaryl-CoA

IPP Isopentyl diphosphate

IPP Isopentyl disphosphate

LS Lanesterol synthase

MD Molecular Dynamics

MM Molecular Mechanics

NMR Nuclear Magnetic Resonance

PCR Polymerase Chain Reaction

PMF Potential of Mean Force

QM Quantum Mechanics

QM/MM Combined Quantum Mechanics / Molecular Mechanics

SBMD Stochastic Boundary Molecular Dynamics

SCF Self Consistent Field

SD Steepest descent

SHS Squalene hopene cyclase

STO Slater Type Orbital

WHAM Weighted Histogram Analysis Method

Chapter 1

Introduction

1.1 Introduction to terpene chemistry and aristolochene synthase

1.1.1 Terpene chemistry and biochemistry

Terpenes and terpene derived compounds form the largest class of natural products, with tens of thousands of examples identified in a diverse range of organisms including marine and terrestrial plants, fungi, bacteria, insects and animals. Members of this enormous extended family perform a wide variety of useful and important roles including those of attractants and deterrents, pheromones, anti-tumour and anti-fungal agents and phytotoxins. [1]

Terpenes share the empirical formula $(C_5H_8)_n$ (the isoprene subunit) and are classified by the number of these subunits they contain. The simplest case is that of the hemiterpenes of which isoprene itself is a member - the oxidised terpenoid derivatives prenol (3-methyl-2-butene-1-ol) and isovaleric acid (3-methylbutanoic acid) are also members of this class. Two subunits $(C_{10}H_{16})$ combine to form the monoterpenes, and the sesqui- and diterpenes consist of three and four subunits, with this homologous series continuing into the higher terpenes.

1.1.2   Chemistry and biosynthesis of terpenes

Over 300 sesquiterpenes have been identified from a variety of organisms [1-3] and the study of these compounds and their biosyntheses have been an area of intense interest to a broad range of scientists engaged in natural product chemistry and biochemistry.

Ruzicka first recognised over fifty years ago that the common precursor for all naturally occurring sesquiterpenes is farnesyl diphosphate[1] (FPP, **20** in Figure 1.4) and that all members of the sesquiterpenes may be derived from this common starting material. [4-6]

FPP, which may be derived from isopentyl diphosphate (IPP) which is itself ultimately derived from condensation of multiple acetyl-CoA molecules (a key intermediate in the Krebs cycle). The reaction (shown in Figure 1.1) proceeds *via* the reduction of 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA, **3** in 1.1) to form the mevalonate anion (**4**). Mevalonate is then phosphorylated and diphosphorylated with three units of

---

[1]In older literature pyrophosphate is used instead of diphosphate

ATP, followed by loss of CO_2 to form dimethylallyl diphosphate (DMAP, **7**) which is then isomerised to form IPP(**8**). As the reaction proceeds *via* the formation of the mevalonate ion this synthesis of IPP is known as the mevalonate pathway. [7]
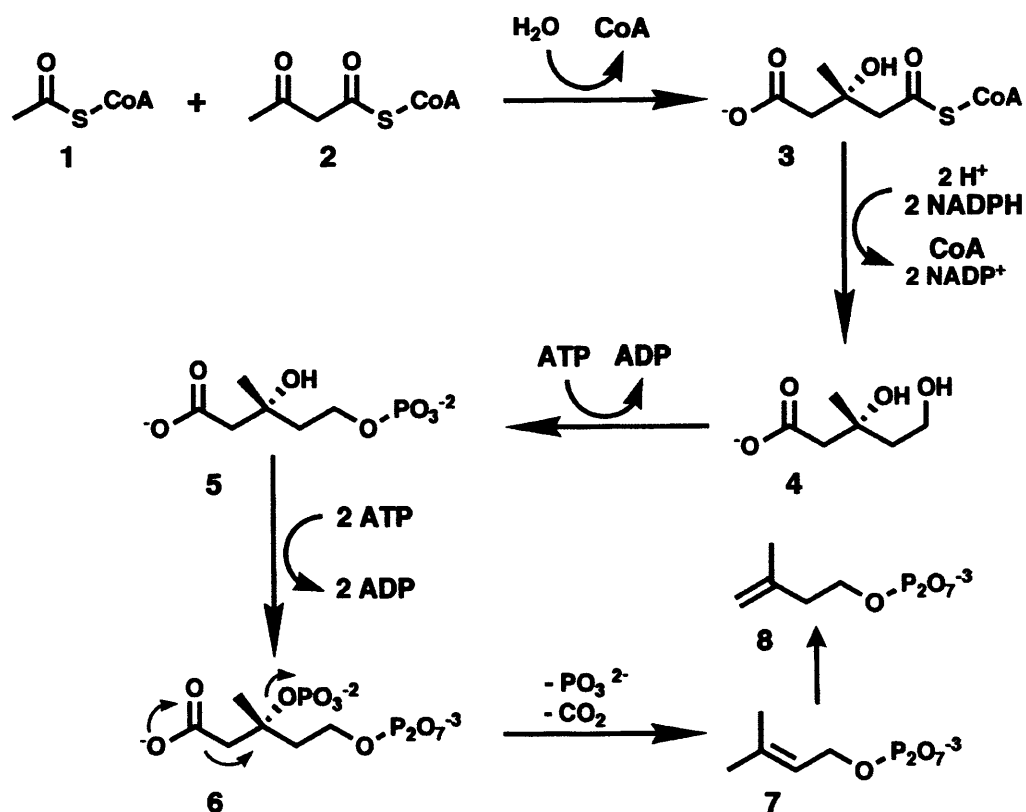
Figure 1.1: *The mevalonate pathway to IPP* - condensation of acetyl-CoA and acetoacetyl-CoA forms HMG-CoA which is then reduced to form the mevalonate ion and hence IPP and DMAP

The discovery that mevalonate is a sterol precursor in liver cells from certain bacteria [8,9] led to the discovery and description of the enzymes that convert acetyl-CoA (**1** in Figure 1.1) and acetoacetyl-CoA (**2**) to mevalonate. [10,11]

Evidence that mevalonate is incorporated into terpenes came from a number of labelling experiments. For example, Bloch and co-workers synthesised mevalonate with ^3H on the C5 CH_2 group and incubated it with liver enzymes. [12] The triterpene

squalene was isolated and found to contain 11 to 12 ^{3}H atoms *i.e.* the tritium was retained during the synthesis of squalene. [12] Around the same time Cornforth performed a series of ^{14}C labelling experiments which were able to describe the synthesis of cholesterol from acetate and therefore conclude that acetyl-CoA is capable of acting as a precursor molecule for terpene synthesis. [13–18] The discovery that HMG-CoA reductase is required for the synthesis of cholesterol in humans has led to the development of statins, the class of drugs which act as HMG-CoA reductase inhibitors and therefore lower blood cholesterol levels. [19,20]

Until around 1990 all plant derived terpenes and terpenoids were assumed to have a precursor olefin diphosphate derived from the IPP metabolic pathway. [21] In addition to the IPP/DMAPP route to terpene precursor molecules such as FPP, there exists another, mevalonate independent, pathway which takes as its starting material pyruvate and glyceraldehyde 3-phosphate (G3P) (**9** in Figure 1.2). These starting materials condense to form 1-deoxy-D-xylulose 5-phosphate (DOXP). DOXP (**10** in Figure 1.2) is then reduced to form 2-C-methylerythritol 4-phosphate which is then converted into its cyclic diphosphate (**14** in Figure 1.2) and thence to IPP or DMAP *via* multiple dehydration steps.

Evidence for a non-mevalonate pathway to IPP can be gathered by considering the fate of ^{13}C atoms in the biosynthesis of IPP from glucose (since glucose is an intermediate in a huge number of biological reactions.) In Figure 1.3 the outline of the mevalonate and non-mevalonate pathways which are displayed separately in Figures 1.1 and 1.2 are shown with carbon centres coloured to indicate possible sites for isotopic labelling. It can be seen that the mevalonate pathway can only transfer two labelled carbon atoms, whereas the non-mevalonate pathway allows for
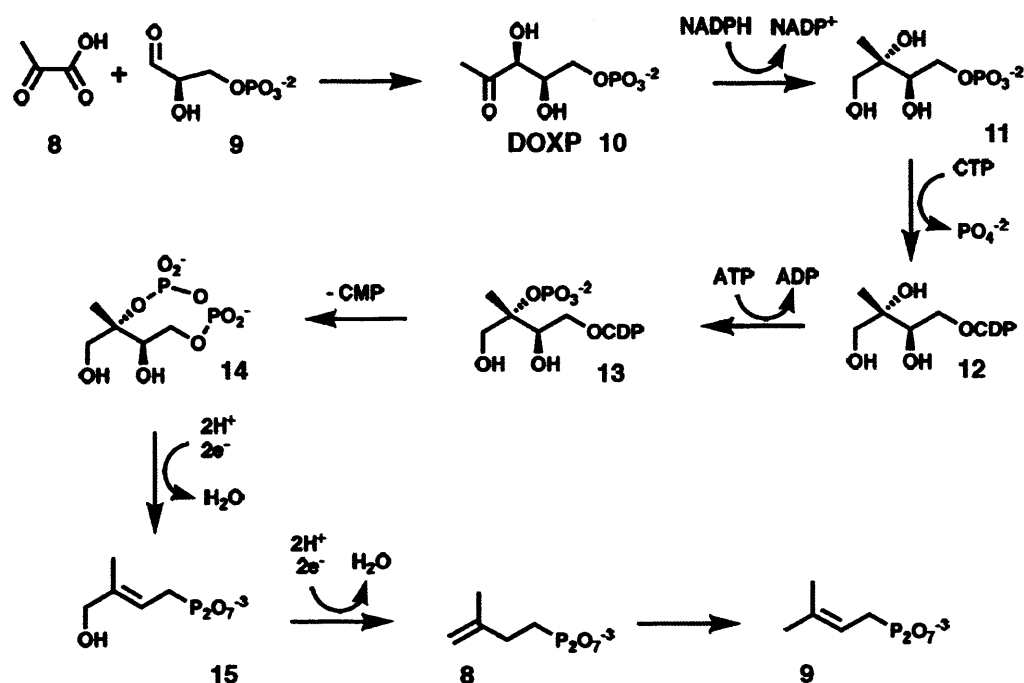
Figure 1.2: *The non-melavonate pathway to IPP* - starting materials G3P and pyruvate condense to form DOXP and thence IPP and DMAP

the transfer of three labelled carbon centres. Detailed ^{13}C labeleling experiments have been performed and helped to confirm the existence of a non-mevalonate pathway. [22]

More evidence for the existence of a non-mevalonate pathway has been obtained from studying the labelling patterns obtained at a number of metabolic end products such as amino acids: since these may be synthesised, in principle, with either acetyl-CoA or pyruvate/G3P as precursor. It is therefore possible to predict labelling patterns in products and make comparisons with experimental results based on feeding organisms with labelled precursor molecules. A number of such experiments have been performed using leucine, valine and tyrosine and used to further demonstrate that the mevalonate pathway is not the unique route to IPP in Nature. [21, 23-29]

mevalonate pathway

non-mevalonate pathway
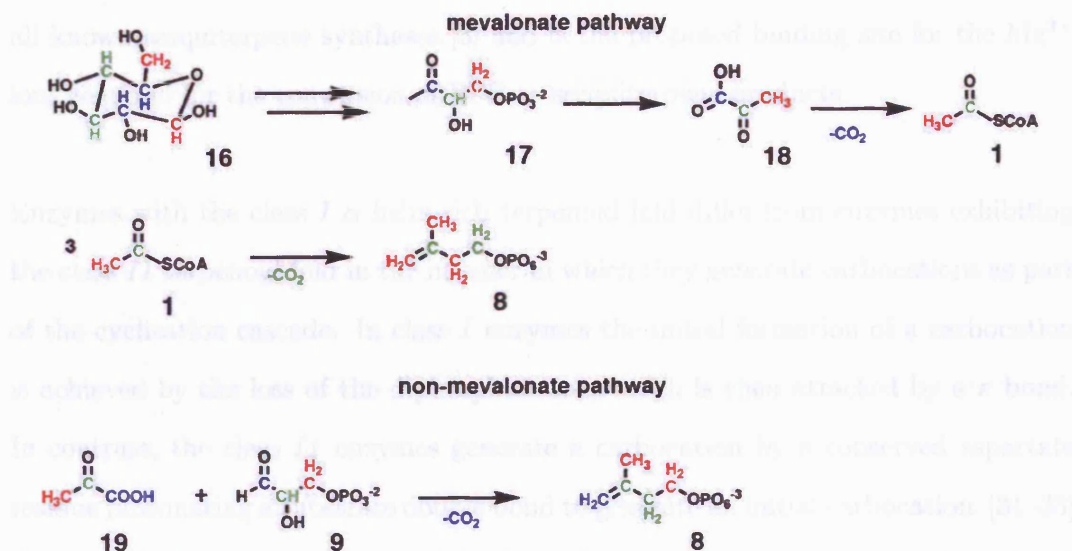
Figure 1.3: ^{13}C *labelling experiments provide evidence for the non-mevalonate pathway to IPP* - Carbon centres coloured to indicate their fate in each pathway

1.1.3 Crystal structures of sesquiterpene synthases

Sesquiterpene synthases are the enzymes that catalyse the conversion of their unique precursor FPP (**20** in Figure 1.4) into C_15 terpene molecules. The mechanisms by which the sesquiterpene synthases take FPP and convert it into a diverse range of natural products, often with high degrees of stereo- and regiocontrol has been a rich topic of research in the last half century.

The first crystal structure of aristolochene synthase obtained was that from the blue cheese mould, *Penicillium roqueforti.* [30] The structure was solved to a resolution of 2.5 Å and revealed that the enzyme has the class *I* α-helical terpenoid cyclase fold [31] and consists of a total of eleven helices (including six α-helices which surround the active site). As with all sesquiterpene synthases there is an aspartate rich DDXX(D/E) domain (starting at Asp 115). This domain is the conserved across

all known sesquiterpene synthases [3] and is the proposed binding site for the Mg^{2+} ions required for the conversion of FPP to sesquiterpene products.

Enzymes with the class I α helix rich terpenoid fold differ from enzymes exhibiting the class II terpenoid fold in the manner in which they generate carbocations as part of the cyclisation cascade. In class I enzymes the initial formation of a carbocation is achieved by the loss of the diphosphate ion, which is then attacked by a π bond. In contrast, the class II enzymes generate a carbocation by a conserved aspartate residue protonating a substrate double bond to generate an initial carbocation. [31–33] There is a high degree of structural similarity (but not necessarily sequence homology) between the class I enzymes. Typically they consist of a number of α-helices, five of which surround the active site of the enzyme and this folding arrangement around the active site has been described as the key feature of the class I terpenoid synthase fold. [34]

The class II enzymes (such as the triterpene synthases squalene-hopene synthase [35] (SHS) and lanesterol synthase (LS) [36]) contain two domains, both consisting of $\alpha\alpha$-barrel folds which are not structurally related to the class I folds described in the previous paragraph. In the higher terpene synthases several class I and/or class II terpene folds may join together to form larger functional active sites, for example in SHS and LS. [32] In the case of the sesquiterpene synthase epi-aristolochene synthase (EAS) both class I and class II folds are present [37] in two separate domains, a feature common to plant sesquiterpene synthases - yet the N-terminal domain in EAS does not have a known catalytic function. This is in contrast to abietadiene synthase from *A. grandis* [38] in which both domains have catalytic activity. In addition it was recently shown that geosmin synthase from *S. coelicolor* [39] is a bifunctional enzyme

in which the N-terminal domain catalyses the formation of germacrene D from FPP and the C-terminal domain completes the biosynthesis of geosmin. Prior to this it was assumed that the C-terminal domain was catalytically inactive.

These structural features (class *I* terpene folds) are common to all of the sesquiterpene synthases for which crystal structures have been solved, trichodiene synthase from *F. sporotrichioides* [40, 41], δ-cadinene synthase from *G. arboreum* [42], pentalenene synthase from *Streptomyces UC5319* [34], 5-*epi*-aristolochene synthase from *N. tabacum* [37], aristolochene synthase from *P. roqueforti* and aristolochene synthase from *A. terreus*. [43, 44] The structures of the products of these enzymes is shown in Figure 1.4.
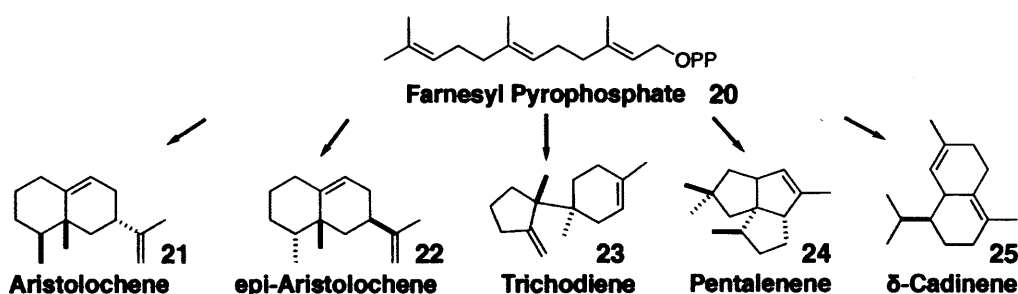
Figure 1.4: *Products of crystallised terpene stynthases* - the structure of the products of the five sesquiterpene synthases for which structures are known to have been solved.

The fungal and bacterial sesquiterpene synthases share low mutual sequence homologies, yet topologically they are all related *via* their class *I* terpenoid fold. It has been noted that avian farnesyl diphosphate synthase [45] also contains a similar α-helical fold to EAS and AS-PR. It has been suggested that enzymes catalysing successive steps in pathways (*e.g.* biosynthesis of FPP followed by enzymatic cyclisation of FPP to form sesquiterpene products) evolved from the enzyme in the previous reaction step through divergence, with function evolving without dramatic change in tertiary structure. [46] This has led to the proposal that all terpene synthases share a common

evolutionary ancestor predating the separation of the various kingdoms of life, thus

explaining the common structure observed between plant, bacterial, fungal and animal

terpene synthases despite the lack of significant sequence homology. [32, 47]

Until very recently it was thought that in addition to their structural similarities

all known sesquiterpene snythases shared the aspartate rich DDXX(D/E) domain

as well as an additional NSE/DTE motif, (L,V)(V,L,A)(N,D)D(L,I,V)X(S,T)XXXE.

[40,48] However, recent crystallographic investigations of DCS from *G. arboreum* have

revealed that this enzyme is apparently unique amongst the sesquiterpene synthases

in lacking the NSE/DTE motif. [42] Instead of a NSE/DTE motif DCS has a second

aspartate rich domain, a feature which it shares with enzymes typically involved in

isoprenoid chain formation such as FPP synthase. [45,49]

1.2 The structure of aristolochene synthase and the mechanism and kinetics of its catalysis

1.2.1 Crystallographic insights into the cyclisation cascade of AS

The isolation, characterisation and bacterial expression of AS-PR was first achieved

[50–52] around the same time as the first set of similar experiments on AS-AT, [48,53]

making it possible to probe kinetic and mechanistic features of AS catalysis in detail.

Following this, crystal structures were solved for AS-AT [43,44] and for AS-PR. [30]

In the published crystal structures of AS-PR no magnesium was present within the crystal structure (Mg^{2+} was not included in the AS containing solution used to obtain crystals), yet magnesium is known to be involved in the cyclisation of FPP to form sesquiterpenes. [54, 55] The role of magnesium is to bind diphosphate and hence initiate catalysis, so in the absence of diphosphate there is no apparent advantage to magnesium binding. The authors of the original crystal structure observed Sm^{3+} binding in the proposed magnesium-binding site, when $SmCl_3$ was introduced to provide an electron rich element for initial phase resolution during crystallography. The authors also note that Sm^{3+} is a good analogue for Mg^{2+} and that Sm^{3+} binds *ca.* 10^3 times more tightly to a carboxylate cluster than Mg^{2+}. [30,56] Despite this, it is difficult to say, precisely, how many magnesium ions bind to allow AS-PR catalysis and what conformation the enzyme adopts as it forms a Michaelis complex (Section 1.2.3).

However, recent work on another aristolochene synthase, obtained from *A. terreus* (AS-AT) has shed new light on this topic. A 2.15 Å crystal structure of AS-AT was solved in 2007 and provided the first insights into the magnesium-phosphate binding motif in AS. [43] In this work, AS-AT was crystallised as a tetramer consisting of two dimeric subunits, labelled A, B, C and D, respectively. This is in contrast to the observation that AS-AT exists as a monomer in solution. [48] In the crystallised form monomer pairs AD and BC join to form the tetrameric structure and it was stated that it is likely that AS-AT (and by extension AS-PR) functions as a dimer in solution, which is a result observed when native gel analysis is performed. [48]

The crystallisation was carried out in the presence of diphosphate anion and Mg^{2+}. Subunit D in the crystal structure contains diphosphate and magnesium ion bound.
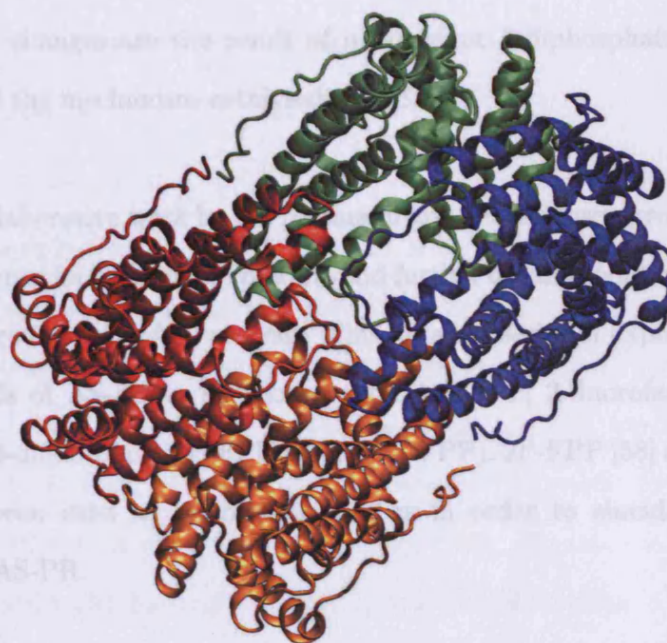
Figure 1.5: *Crystal structure of AS-AT tetramer* - Representation of the crystal structure of AS-AT. [43] Chain A in blue, chain B in red, chain C in green and chain D in orange.

Three magnesium ions bind in the crystal structure of subunit D. Two bind to the carboxylate group of Asp 90 (which is the first residue of the aspartate rich domain motif in AS-AT) and to a diphosphate anion and four crystallographic water molecules. A third magnesium binds to the NSE/DTE motif (Asn 219, Ser 223, Glu 227). This magnesium/diphosphate binding motif is the same as that observed in other terpene cyclases for which comparisons are available, EAS [37], trichodiene synthase [40] and bornyl diphosphate synthase. [57] Subunit D of the crystal structure also exhibits differences in tertiary structure compared to the other subunits; two N-terminal helices adjust their conformations such that the entrance to the active site is capped by amino acid loops. The conformation adopted by chain D of the AS-AT crystal structure will be referred to as the 'closed' structure of AS in the remainder of this work. The differences between chain D and the other subunits suggest that

conformational changes are the result of magnesium / diphosphate binding in the
initial stages of the mechanism catalysed by AS.

Subsequent collaborative work by the Allemann and Christianson groups [44] revealed
a possible sequence for magnesium binding and further details about the conformational
changes that occur during AS catalysis. This detailed series of experiments involved
growing crystals of AS-AT in solutions containing FPP, 2-fluorofarnesyl-FPP (2F-
FPP) and 12,13-difluorofarnesyl-FPP (12,13-DF-FPP). 2F-FPP [58] and 12,12-DFFPP
[59] have all been used as mechanistic probes in order to elucidate the reaction
mechanism of AS-PR.

Allemann, Cane, Christianson and co-workers [44] provided evidence that three magnesium
ions are necessary for catalysis by AS-AT and a possible sequential magnesium binding
mechanism leading to the formation of the Michaelis complex. By combining the
solved crystal structures containing FPP, 2F-FPP and 12,13-DF-FPP, the following
model for substrate and magnesium binding and subsequent catalysis followed by
product release was proposed. Firstl Arg 315 and Tyr 215 bind the diphosphate group
of FPP (via hydrogen bonding interactions) alongside the coordination of Mg_B^{2+}
at the NSE/DTE binding domain. Following this Mg_C^{2+} binds to Asp 90 (of the
aspartate-rich domain) and the diphosphate group of FPP, leading to the formation
of the observed closed configuration of AS-AT. Mg_A^{2+} then coordinates to Asp 90,
producing the conformational changes necessary to form a magnesium binding triad
capable of initiating catalysis. After completion of the cyclisation reaction Mg_A^{2+}
is released, followed by release of aristolochene (plus any other terpene products)
and Mg_C^{2+}. Finally Mg_B^{2+} is released along with diphosphate anion completing the
catalytic cycle and leaving the apo-enzyme free to perform subsequent catalysis.

1.2.2   Kinetic and substrate labelling studies of sesquiterpene synthases

In the absence of purified, isolated, stable and well-characterised samples of enzyme early investigators relied upon isotopic labelling studies to investigate the mechanisms of sesquiterpene synthesis. Typically, these experiments relied on introducing ^{14}C, ^{13}C, ^{2}H, ^{3}H and ^{18}O labelled precursors to the organism, and using radiochemistry, gas chromatography-mass spectrometry and Nuclear Magnetic Resonance (NMR) techniques to study product formation and assign stereochemistry to products. These classic biochemical techniques rely on introducing labelled species to whole organisms and rely on uptake of high concentrations of these exogenous molecules, something which is not always certain. A number of reviews exist detailing this type of work in different terpene classes, including the sesquiterpenes. [60–62]

Isolation and purification of cell-free sample of enzymes allows for much more detailed investigations to be performed, in particular studies of the kinetic parameters of the enzyme. Revolutions in molecular biology, such as the development of site-directed mutagenesis by Smith [63, 64] and the discovery of the polymerase chain reaction (PCR) developed by Mullis [65] earned both Mullis and Smith the Nobel Prize for Chemistry in 1993, and has led to significant advances in the understanding of the kinetics and mechanism of AS.

1.2.3   Kinetic studies on aristolochene synthase

The kinetics of enzymatic catalysis are described by the Michaelis-Menten model [66], subsequently formulated into the steady-state approximation by Briggs and Haldane. [67] Assuming simple first order kinetics and the absence of inhibitors then in the Michaelis-Menten approximation for a given enzyme E and substrate S, there are rate constants for the formation and destruction of an enzyme-substrate reactive complex ES (the Michaelis complex) given by the following equations.

$$E + S \xrightarrow{k_1} ES \tag{1.1}$$

$$ES \xrightarrow{k_2} E + S \tag{1.2}$$
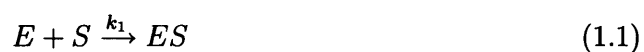
The enzyme-substrate complex can also react to produce product, P and regenerate enzyme.

$$ES \xrightarrow{k_3} P \tag{1.3}$$

The catalytic rate is the product of $[ES]$ and k_3. This quantity is the velocity V of

the reaction and is given by Equation 1.4.

$$V = k_3[ES] \tag{1.4}$$

It is often impossible to determine ES, but it is possible to determine from other rate constants

$$rate_{formation\ of\ ES} = k_1[ES] \tag{1.5}$$

$$rate_{breakdown\ of\ ES} = k_2 + k_3[E][S] \tag{1.6}$$

At this point the steady-state approximation is invoked, *i.e.* the concentration of the intermediate, $[ES]$, is constant (the steady state) whilst the concentration of reactants and products (E, S and P) vary with time. Thus, Equations 1.5 and 1.6 are equal and may be arranged to give Equation 1.7.

$$[ES] = \frac{[E][S]}{(k_2 + k_3)/k_1} \tag{1.7}$$

The rate constants combined in the form shown in Equation 1.7 define the Michaelis

constant K_M, hence

$$[ES] = \frac{[E][S]}{K_M} \tag{1.8}$$

and when $[E] = [E_T]\text{-}[ES]$ (where E_T is the total enzyme concentration) is substituted into Equation 1.8 then the following is true

$$[ES] = [E_T]\frac{[S]}{[S] + K_M} \tag{1.9}$$

and one can substitute this back into equation 1.4 to give

$$V_{max} = k_3[E_T]\frac{[S]}{[S] + K_M} \tag{1.10}$$

V_{max} (the maximum rate of catalysis) is assumed to occur when the active site is saturated with substrate *i.e.*

$$as \quad [S] \rightarrow \infty \quad then \quad \frac{[S]}{[S] + K_M} \rightarrow 1 \tag{1.11}$$

and hence

$$V_{max} = k_3[E_T] \tag{1.12}$$

gives an expression for V_{max} which can then be substituted back into 1.10 giving the standard form of the Michaelis-Menten equation.

$$V = V_{max} \frac{[S]}{[S] + K_M}$$ (1.13)

Since V be measured experimentally for a known $[S]$, values of V_{max} and K_M can be calculated by fitting V and $[S]$ data to the Michaelis-Menten equation using statistical software, or alternatively by means of a Lineweaver-Burke [68] plot created by taking the inverse of the Michaelis-Menten equation and plotting $1/V$ against $1/[S]$ as straight line. The gradient of a Lineweaver-Burke plot is K_M/V_{max}, the x-axis intercept is K_M^{-1} and the y-axis intercept is $1/V_{max}$.

The physical meaning of K_M is that it is the concentration of substrate required to fill half the active sites of an enzyme and therefore a measure of how much affinity the substrate has for the active site: a large value of K_M indicates weak binding and a small value of K_M implies strong binding.

The second important type of data that can be obtained from enzyme kinetics experiments is the turnover number, k_3, which can be calculated using the data in a Lineweaver-Burke plot and Equation 1.12. However, V_{max} is only achieved when all of the enzyme active sites are occupied and usually implies a very high substrate concentration - not normally achieved experimentally or *in vivo*. However, Equations

1.4 and 1.8 may be combined to give the following expression

$$\frac{k_3}{K_M}[E][S] \tag{1.14}$$

and, assuming $[S] \ll K_M^{-1}$ then $[ES] \simeq [E_T]$ then substitution into Equation 1.14 gives Equation 1.15

$$\frac{k_3}{K_M}[E_T][S] \tag{1.15}$$

which implies that the observed rate of catalysis depends on k_3, K_M and $[S]$. The upper limit for the speed of enzyme catalysis therefore depends upon these three variables. If one assumes $[S]$ to be relatively low under physiological conditions (since there is little advantage for an organism to synthesis excess of substrate), then K_M is not reached under these conditions. Therefore k_3 appears to be the limiting factor: however k_3 itself depends upon $[ES]$, which is in fact limited by k_1 - the rate constant for formation of $[ES]$.

There is an upper limit for k_1. This upper limit is k_3, the diffusion limit for successful interactions between enzyme and substrate. This diffusion limit has a value of around 10^8 to 10^9 $M^{-1}s^{-1}$ [69, 70] and thus represents the upper speed limit for enzyme catalysis.

Of course, a simple model of enzyme catalysis involving just three rate constants and a

trivial reaction pathway may not always be valid, but it is, in principle, always possible to measure the maximum observed rate of catalysis (k_{cat}) as [S] reaches saturation along with experimental values of K_M. Therefore, for real systems k_{cat}/K_M is the reported value and can be seen as a measure of enzyme efficiency, with an upper limit of the rate of diffusion controlled encounter between enzyme and its substrate.

Many kinetic studies have been performed on wild type AS-AT and AS-PR and results are shown in Table 1.1. The values of k_{cat}, K_M and hence k_{cat} / K_M are in good agreement for AS-PR and AS-AT and the value for k_{cat} is similar for both enzymes. However, AS-AT and AS-PR have K_M values differing by up to an order of magnitude. The implication is that active site of AS-AT is better able to bind FPP than the AS-PR active site.

The idea that AS-AT binds substrate more effectively than AS-PR offers an explanation for the observation that wild type AS-AT produces a single product in 100% yield [48] whereas AS-PR has been shown to produce a number of other products (*vide infra*) in varying yields. [71-75]

There are in some cases several orders of magnitude difference in the values of k_{cat}/K_M, *e.g.* in reference [71] and [48]. Since these experiments were performed (on AS-PR) independently in different laboratories with protocols described in published literature it is difficult to explain these differences or to suggest which value is the correct value.

Table 1.1: AS kinetic data - *Summary of kinetic data obtained for recombinant and native WT-AS-PR and WT-AS-AT.*

Enzyme	k_{cat} (s^{-1})	K_M (nM)	k_{cat}/K_M (s^{-1}M^{-1})
AS-AT (native) [48]	0.014 ± 0.002	13.5 ± 1.23	1.04×10^6
AS-AT (recombinant) [48]	0.015 ± 0.002	14.8 ± 10.7	1.01×10^6
AS-AT (recombinant) [75]	0.017 ± 0.007	130 ± 10	1.3×10^5
AS-PR (recombinant)) [48]	0.043	520	8.29×10^4
AS-PR (recombinant) [75]	0.043	600 ± 100	7.2×10^4
AS-PR (recombinant) [71]	0.03 ± 0.001	2300 ± 500	$13.9 \pm 3 \times 10^6$

1.3 Mechanistic and mutagenesis studies on AS-PR

In the absence of a model of an enzyme's structure it is difficult to probe the enzymatic reaction mechanism. Kinetic experiments, as described in Section 1.2.3, provide details of the overall rate of the reaction and information about substrates affinity for the active site but little about how the enzyme actually achieves catalysis. Labelling studies can provide information about which biochemical pathway substrate arises from, and substrate labelling studies can provide some information about the stereochemistry of reactions and allow some hypotheses to be formed in the absence of detailed structural information about an enzyme. Mutagenesis, however, allows experiments to be performed which aim to describe precisely which amino acids within an active site are involved in catalysis and at which stage in a mechanism it performs its role.

1.3.1 Previously reported mechanism for the biosynthesis of aristolochene by AS-PR

Figure 1.6 outlines the previously proposed reaction mechanism for AS-PR catalysis.
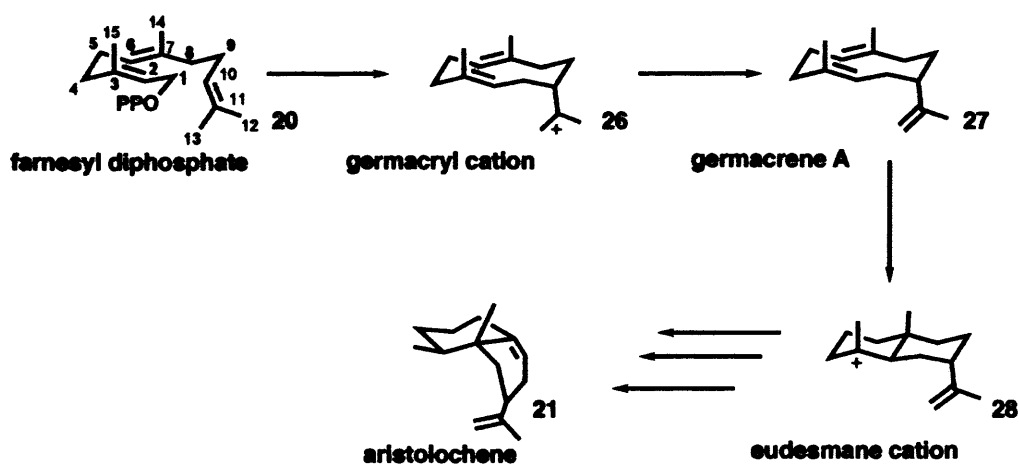
Figure 1.6: *Previously Proposed Reaction Scheme* - FPP is converted to aristolochene with germacrene A as intermediate. [71,76]

The enzyme binds FPP,**20**, in a quasi-cyclic conformation thereby facilitating attack on C1 by the C10-C11 double bond (along with loss of diphosphate) to produce germacryl cation (**26**). [72,74] Proton loss from C12 leads to the production of germacrene A (**27**). [71,76] This was determined by synthesising [12,12,12-^{2}H]FPP and [13,13,13-^{2}H]FPP [55,77]and performing ^{2}H NMR. Upon incubation of [12,12,12-^{2}H]FPP with AS-AT a single olefinic peak at δ4.71 [76] was observed indicating that a proton/deuteron is lost during AS-AT catalysis (Figure 1.7). Conversely incubation of [13,13,13-^{2}H]FPP with AS-AT revealed an alkyl resonance at δ 1.69 consistent with no proton loss at C13.

The stereochemistry of the cyclisation reaction of AS-AT at the C1 position has also
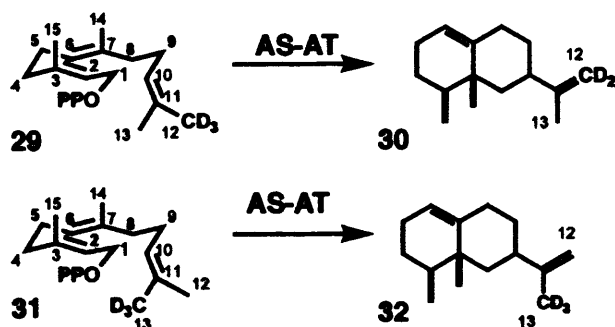
Figure 1.7: *Evidence for proton loss from C12 position* - the synthesis of ^{2}H analogues of FPP followed by incubation with AS-AT was used to probe the mechanism

been studied. Synthesis of $(1R)$-$[1$-^{2}H$]$-FPP and $(1S)$-$[1$-^{2}H$]$-FPP [78] followed by incubation with AS-AT and ^{2}H NMR analysis of the labelled aristolochenes produced demonstrated that inversion of configuration takes place at the C1 position in FPP. [79] The results of this experiment are included as Figure 1.8.
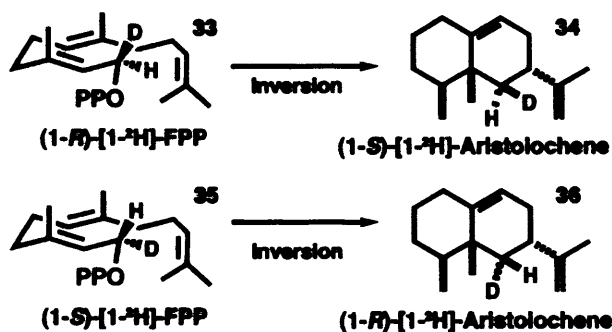
Figure 1.8: *Inversion of stereochemistry at C1 position in AS* - experimental observation of inversion of stereochemistry at C1 position using labelled FPP and AS-AT

Inversion of configuration at the C1 position is consistent with the formation of a ten-membered germacryl ring, and rules out a mechanism involving the initial formation of a six membered ring which has been shown to result in retention of configuration at C1. [80] The frequent assumption made, both implicitly and explicitly, by researchers working on AS-PR is that it shares an identical reaction mechanism with AS-AT.

Following cyclisation, germacrene A (**27** in Figure 1.6) was then postulated to undergo reprotonation of the C6-C7 double bond and a further cyclisation to form the bicyclic eudesmane cation (**28**). Successive 1,2 hydride shift and methyl migration followed by loss of H_{Si} on C8 results in the generation of (+)-aristolochene. Evidence for this came from the synthesis of (4*R*,*8*R)-[4,8 -^{2}H]FPP and (4*R*,8*S*)-[4,8-^{2}H]FPP [81, 82] followed by incubation of these labelled substrates with FPP.
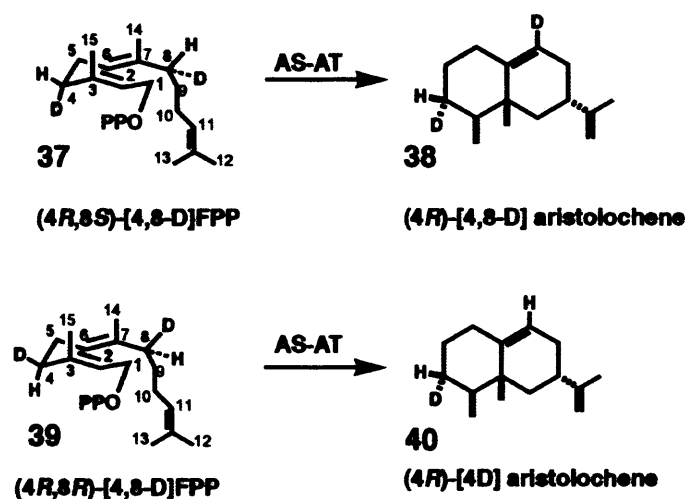
Figure 1.9: *Sterochemistry of C8 deprotonation in AS-AT catalysis* - ^{2}H labelling and NMR experiments demonstrate that final deprotonation occurs exclusively at the $H8_{Si}$ position

In the case of (4*R*,8*S*)-[4,8-^{2}H]FPP the resulting labelled aristolochene contained just a single ^{2}H NMR peak at δ1.36 indicating that $H8_{Si}$ had been lost during the synthesis of aristolochene by AS-AT. In the case of (4*R*,8*R*)-[4,8 -^{2}H]FPP two peaks were observed in the ^{2}H NMR spectra (at δ1.44 and 5.35) indicating that $H8_{Re}$ is conserved by AS-AT. [76] A summary of these experiments is included as Figure 1.9.

1.3.2 Site directed mutagenesis of the active site of AS-PR

Evidence for the mechanism by which AS-PR catalyses its reaction has come mainly through the use of site directed mutagenesis experiments. The Allemann and Cane groups have conducted a number of site directed mutagenesis studies of the active site of AS-PR which have provided evidence for aspects of the mechanism proposed in Figure 1.6. A summary of the results of experiments by the Allemann group (and one set of data from the Cane laboratory) on the active site of these modified AS-PRs is shown in Table 1.2.

AS-PR may be converted into an (E)-β-farnesene synthase by mutation of Tyr 92 into a smaller non-aromatic residue: mutants Y92V, Y92A and Y92C [72,73] produce the linear farnesenes as their major products, although with vastly reduced catalytic efficiency compared to the wild-type - the k_{cat} / K_M of which is in the region of 1 $\times$ 10^5 M^{-1} s^{-1} (Table 1.1) and this is reduced by around four orders of magnitude in the modified proteins. The evidence from these experiments is that the steric bulk of the tyrosine side chain helps guide linear FPP into the required quasi-cyclic conformation required for cyclisation, as alluded to in Figure 1.2. If FPP is not in the appropriate conformation, the initial loss of the diphosphate group, followed by deprotonation at the C15 and C4 position leads to the formation of the linear alkene (E)-β-farnesene. This represents a simple mechanism by which the positive charge can be quenched when farnesyl cation is still in an approximately linear conformation *i.e.* there is no possibility of an interaction between the π electrons of the C10 C11 double bond which is geometrically required for cylcisation to occur.

It has been suggested that the role of Trp 334 in AS-PR is to stabilise eudesmane

cation (**28**) within the active site of the enzyme. [73] By synthesising the W334F,

W334L and W334V mutants the Allemann group was able demonstrate that replacing

the aromatic group with a aliphatic residue produced germacrene A (**27**) as the major

product (100 % in the case of the W334L mutant), whilst replacing the Trp with

another aromatic residue in the W334F mutant still produced aristolochene (although

with a vastly reduced catalytic efficiency). The authors concluded that the role of

an aromatic group at the 334 position was to stabilise the positive charge at the C3

position in eudesmane cation (presumably by interaction between carbocation and

electron system in the ring).

Similar studies suggested a dual rule for Phe 178. The F178Y mutant still produces

aristolochene as its principle product (86.4 %) whilst the F178V mutant produces

germacrene A as its principle product, along with a complex mixture of side products

including both linear (farnesene) products and selinenes (bicyclic aristolochene isomers).

[83] That the F178V mutant can lead to both linear and cyclic products suggest a

dual role for this residue [83] both providing steric bulk in a similar fashion to Tyr

92 and also stabilising carbocations after the formation of eudesmane cation (**28**).

Further evidence that the steric bulk of Phe 178 plays a role in stabilising eudesmane

cation, allowing for the formation of aristolochene, was obtained in a further series

of experiments in which the F178C, F178I, F178Y and F178W mutants of AS-PR

were created and studied in a similar manner to that outlined above. Replacement

of Phe 178 with other bulky aromatic residues (i.e. the F178Y and F178W mutants)

resulted in an enzyme still producing aristolochene as its primary product (*ca.* 90 %)

although the mutant enzymes exhibited an order of magnitude reduction in catalytic

efficiency. [84]
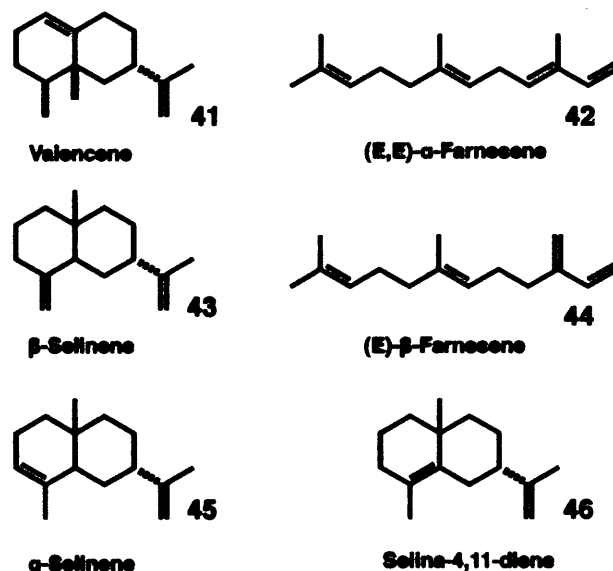
Figure 1.10: *Side products of AS* - compounds detected in product mixtures from mutant and wild-type AS-PR.

The role of Phe 112 was also investigated, on the basis of its position (in the apo-enzyme crystal structure) between Trp 334 and Phe 178. The F112A mutant was observed to principally produce linear farnesenes (total farnesene yield *ca.* 87 %), alongside a smaller yield of germacrene A leading to the conclusion that it is involved in stabilisation of the farnesyl cation produced when pyrophosphate is lost from FPP.

As is implied in Figure 1.6 protonation of the C6 C7 double bond is required to transform germacrene A into eudesmane cation. In 5-*epi*-aristolochene synthase from *N. tabacum* Tyr 520 was proposed, based on detailed analysis of the crystal structure, to be the active site acid which protonates the C6 C7 double bond. A mutagenesis study in which the Y520F mutant was produced was used to test this hypothesis. [37] The Y520F mutant produced germacrene A as its sole product, with a catalytic efficiency approximately 3 % of the wild-type EAS, leading the authors to conclude that the role of Tyr 520 is to act as an acid, capable of protonating the alkene

Table 1.2: AS mutant products - *Summary of Results of Active Site Mutagenesis Experiments performed upon AS-PR.*

Mutant	Major Product	$k_{cat}/\,K_M$ (s^{-1}M^{-1})
Y92F [71]	Aristolochene	12.36 ± 0.8
Y92F [75]	Aristolochene	3.4 ×10^4
Y92V [72]	(*E*)-β-Farnesene	51.7
Y92A [74]	(*E*)-β-Farnesene	10.3 ± 4.3
Y92C [74]	(*E*)-β-Farnesene	16.5 ± 2.4
W334F [73]	(*E*)-β-Farnesene	16.5 ± 2.4
W334V [73]	Germacrene A	37.1 ± 6.3
W334L [73]	Germacrene A	0.3 ± 0.1
F178Y [83]	Aristolochene	196.1± 110
F178V [83]	Germacrene A	1.88± 0.32
F178C [84]	Germacrene A	1678± 221
F178I [84]	Germacrene A	4556± 1754
F178W [84]	Aristolochene	16809± 9102
F112A/F178V [84]	(E)-β-Farnesene	30.2± 2.5

intermediate and facilitating production of *epi*-aristolochene. [85] However, although catalytic efficiency is vastly reduced in the Y520F mutant the enzyme still exhibits turnover to produce *epi*-aristolochene. Therefore the presence of an alternative acid within the active site, or a different function for Tyr 520 cannot be excluded purely on the basis of this result.

Based on analogy with EAS, and study of the solved crystal structure of AS-PR it was proposed that Tyr 92 might also act as the active site acid. [71] Significant

yields (*ca.* 30 %) of germacrene A were achieved when the Y92F variant of AS-PR was incubated with FPP and the reduction of catalytic efficiency was of a similar scale to that measured in the equivalent experiment using the Y520F mutant of EAS. However, in contrast to experiments on EAS, the Y92F variant of AS-PR still produced aristolochene as the major (*ca.* 55 %) product. Such a result suggests that the roles of Tyr 520 and Tyr 92 in EAS and AS-PR are not equivalent as was hypothesised. This observation serves to highlight the fact that whilst there are some striking similarities between the sesquiterpene cyclases, it is not true to say that there is a single mechanistic route to each of the common steps in sesquiterpene catalysis and that the subtleties they exhibit make them challenging enzymes to study.

1.3.3 Site directed mutagenesis of the proposed magnesium binding sites of AS-PR

The proposed magnesium binding sites of AS-PR have also been investigated using mutagenesis studies by the Cane group and the results are summarised in Table 1.3.

The effects on catalytic efficiency are less for mutations occurring outwith the active site, although the D115N, N224L and S248A/E252D proteins had no detectable catalytic properties. Of particular note is that the N244D, S248A, E252D and E252Q mutant enzymes all produce germacrene A as their major product (80 to 100 %). [75] The authors suggest that, on the basis of analogies with trichodiene synthase, mutations in the magnesium binding domains allow for increases in active site volumes, increasing the number of degrees of conformational and spatial freedom

Table 1.3: AS magnesium binding domain - *Summary of Results of mutagenesis experiments performed upon proposed magnesium binding domain of AS-PR. [86]*

Mutant	Major Product	k_{cat}/K_M $(s^{-1}M^{-1})$
D115E	Aristolochene	6.0×10^3
D115N	Inactive	Inactive
D116E	Aristolochene	1.2×10^3
D116N	Aristolochene	1.5×10^3
E119D	Aristolochene	3.0×10^4
E199Q	Aristolochene	3.2×10^3
N244L	Inactive	Inactive
N244D	Germacrene A	36.4
S248A	Germacrene A	2.4×10^2
E252D	Germacrene A	6.5×10^2
S248A/E252D	Inactive	Inactive

and hence either prevent cyclisation or protonation of the C6 C7 double bond and as a results increase the yield of germacrene A compared to the yield of aristolochene. Whilst this does not appear to be an invalid hypothesis there is, as yet, little experimental evidence that this is indeed the case in AS-PR.

1.4 Substrate analogue studies

In Section 1.3.3 the results of mutagenesis experiments were described. In essence these rely on altering the tertiary structure of the enzyme in some way and identifying

the reaction products using chromatography or spectroscopy. From the results of these experiments it is possible to produce evidence for and against proposed reaction mechanisms, based on the assumption that the replacement of a single amino acid residue does not have any effect on the secondary or tertiary structure of the enzyme which might itself have an effect on the catalytic process.

Alternatively substrate analogues, similar to the true substrate but containing different functional groups or atoms in positions of interest may be used. The aim of synthesising these novel molecules is to use their altered properties to investigate the reaction mechanism without altering the tertiary protein structure in a significant way.

The use of fluoro-substituted substrate analogues have emerged as powerful tools for the study of terpene biosynthesis. Alkyl-fluorine bonds' special properties do not alter binding affinities by effecting size or shape changes. [58,87] However, fluorine's place at the extreme of all tables of electronegativity is evidence of an ability to inductively destabilise carbocations on β positions. In addition to destabilisation at the β position there is a stabilisation of carbocations at the α position caused by fluorine's π electron pairs. This ability to stabilise and destabilise carbocations makes it a powerful tool in the study of terpene cyclase reaction mechanisms. [88-90]

1.4.1 Incubation of 2-Fluorofarnesyl-diphosphate with AS-PR

2-Fluorofarnesyl-diphosphate (2F-FPP) has been used to probe the reaction mechanism of AS-PR. 2F-FPP was incubated with AS-PR and the resulting product was identified, as a result of a series of NMR studies, as 2-fluoro-germacrene A. Detailed NMR experiments were necessary because germacrene A is known to undergo thermally induced Cope rearrangements [91] at the temperatures typical during gas chromatography separations and because the ten membered ring is highly fluctional [58] and therefore presents a complicated problem in structural determination.
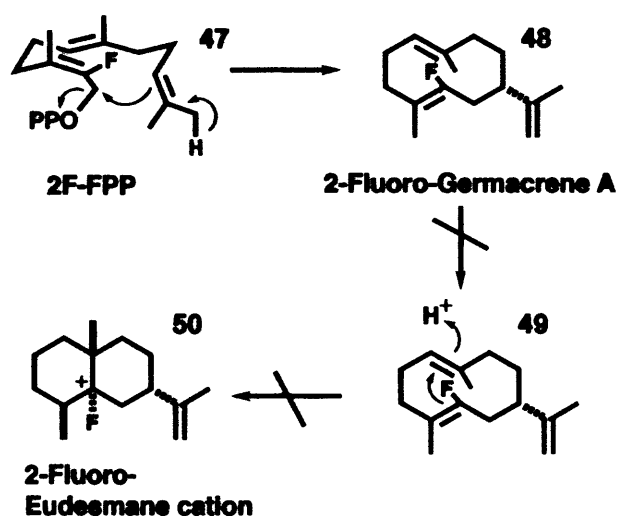
Figure 1.11: *Reaction of 2F-FPP with AS-PR* - incubation of 2F-FPP with AS-PR leads to the formation of 2F-germacrene A. [58]

Subsequent protonation would, according to the mechanism outlined in Figure 1.6, result in the formation of 2-fluoro-eudesmane cation. This species would be slightly stabilised due to the interaction between carbocation and π electrons of the fluorine atom. However, due to the electronegativity of fluorine, the π bond adjacent to fluorine in 2F-germacrene A has a reduced electron density, presumably making the

ring closure step to form 2F-eudesmane cation less favourable.

2F-germacrene A was the sole product isolated from the incubation leading the authors to conclude that germacrene A must be a true intermediate in the conversion of FPP to aristolochene by AS-PR, rather than an off-pathway side product, on the basis that if germacrene A was merely a side product, then other products, aside from 2F-germacrene A would be expected from this reaction. The authors note that germacrene A is not a known substrate of AS-PR, but point out that as an olefin its aqueous solubility is low, resulting in minimal uptake by AS-PR and subsequent lack of detectable turnover.

Considering the initial step proposed in Figure 1.11 one might expect that the electronegative fluorine on the C2 position could prevent the formation of farnesyl cation (formed by initial loss of diphosphate), assuming that loss of diphosphate and subsequent cyclisation occurs in the stepwise fashion, evidence for which was obtained experimentally [71, 74, 83] and is described in Section 1.3.3. The alternative hypothesis is that loss of diphosphate, deprotonation at C12 and subsequent cyclisation to form germacryl cation (**26**) occurs in a concerted process.

1.4.2 Incubation of 12,13-Difluorofarnesyl diphosphate with AS-PR

The stepwise or concerted nature of this initial step in aristolochene biosynthesis has also been addressed by the use of fluorinated substrate mimics. 12,13-Difluorofarnesyl

diphosphate (12,13-DFFPP) was synthesised and used to investigate the cyclisation of FPP in AS-PR. [59]
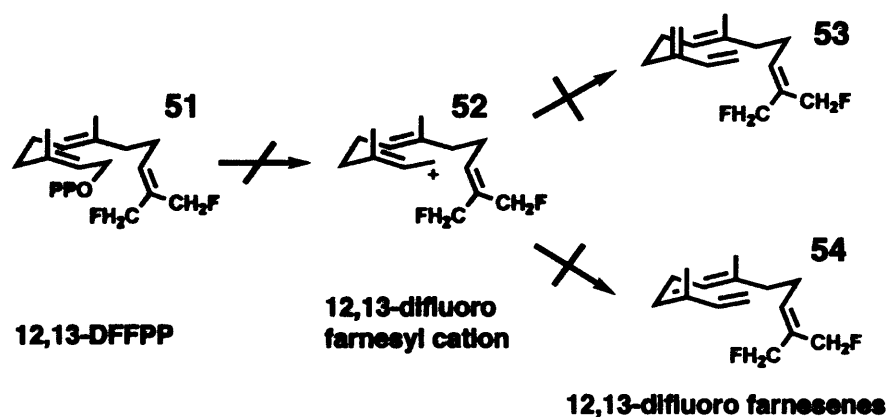
Figure 1.12: *Reaction of 12,13-DFFPP with AS-PR* - 12,13-DFFPP is an inhibitor of AS-PR, with no turnover to produce fluoro-farnesene products, strongly suggesting a concerted pathway for the conversion of FPP to germacryl cation during aristolochene biosynthesis by AS-PR. [59]

Results of this experiment are included as Figure 1.12. If the reaction proceeded by stepwise loss of diphosphate from 12,13-DFFPP to form a 12,13-difluoro farnesyl cation then it would be expected that the inductive destabilisation of the C10-C11 bond would lead to the formation of linear difluoro-farnesenses by the quenching of the unstable cationic intermediate implied by this mechanism (Figure 1.12).

However, no products were detected from incubation of 12,13-DFFPP with AS-PR and instead, 12,13-DFFPP was shown to be potent reversible competitive inhibitor of the enzyme with the inhibition constant K_I calculated to be 0.8 ± 0.2 μM compared with K_M values of between 0.5 and 2.3 μM reported for AS-PR (Table 1.1 and references cited therein). This suggested that aristolochene biosynthesis proceeds *via* a concerted reaction mechanism in which attack on C1 by the C10 C11 double bond leads to loss of diphosphate (and inversion of stereochemistry at C1) and production

of germacryl cation.

This conclusion is different from the one suggested by previous mutagenesis studies (Section 1.3.3) in which the F112A mutant enzyme produced linear farnesene products when incubated with FPP. [84] One possible explanation for this apparently contradictory result is that the replacement of Phe with Ala alters the volume and contour of the AS-PR active site in such a way that stepwise loss of pyrophosphate occurs, and that the active site geometry is such that the normal chaperoning of FPP into germacryl cation is not possible.

The difference in the conclusions between mutagenesis experiments and substrate labelling experiments gives reason to be cautious about interpretation of the results of any one experiment in AS chemistry. It is often not possible to demonstrate absolute proof of a reaction mechanism and each experiment serves only to add evidence for or against a particular hypothesis. In addition, the interesting new results from the use of fluorinated analogues of FPP have demonstrated that this approach is an excellent tool in providing new information about the reaction catalysed by AS-PR.

1.4.3   Incubation of 6-fluorofarnesyl diphosphate and 14-fluoro farnesyl diphosphate with AS-PR

The synthesis of 6-fluorofarnesyl diphosphate (6F-FPP) and 14-fluorofarnesyl diphosphate (14F-FPP) and subsequent incubation with AS-PR and analysis using a series of NMR experiments has provided evidence for the intermediacy of germacrene A during

catalysis by AS-PR. In addition, these experimental results have been used to support the proposal that formation of eudesmane cation (**28** in Figure 1.6) following protonation of the C6 C7 double bond occurs in a concerted fashion.

Figure 1.13 shows the relevant result of these experiments: in both cases the fluorinated analogue (6F- or 14F-germacrene A) was identified as the sole product of the reaction.
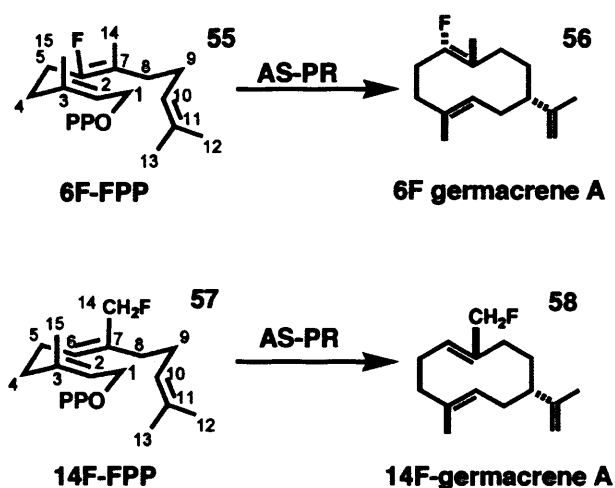
Figure 1.13: *Reaction of 6F-FPP and 14F-FPP with AS-PR* - in both cases the fluorinated analogue of germacrene A was identified

This in turn shed light on the possible mechanism for the formation of eudesmane cation from germacrene A; if this occurs *via* a stepwise mechanism (Figure 1.14) then the developed carbocation on the C7 position is greatly destabilised by being β to a fluorine atom. [58,59,87]

The alternative hypothesis presented was that formation of eudesmane cation may occur *via* a concerted reaction mechanism in which protonation of the C6 C7 double bond occurs simultaneously with attack of the developing positive charge on C7 by the π electron system between C2 and C3, thus avoiding the protonated germacrene A intermediate envisaged in Figure 1.14. The prerequisite for this concerted mechanism
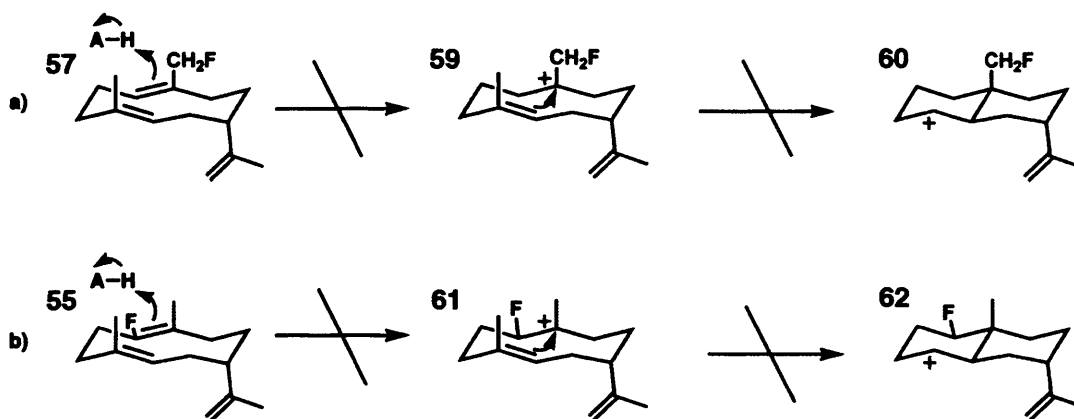
Figure 1.14: *Stepwise formation of germacrene A with 6F-FPP and 14-FPP* - in a stepwise mechanism formation of a protonated germacrene is disfavoured due to the destabilisation of the carbocation β to fluorine in 14F-FPP (a) and 6F-FPP (b)

is that AS-PR forces germacrene A into a favourable conformation to allow for effective electronic orbital overlap between C2 and C7 and hence formation of eudesmane cation. In the analysis by Miller *et. al.* [58, 59] it is suggested that withdrawal of electron density from the C6 C7 double bond prevents this favourable overlap and subsequent σ bond formation.
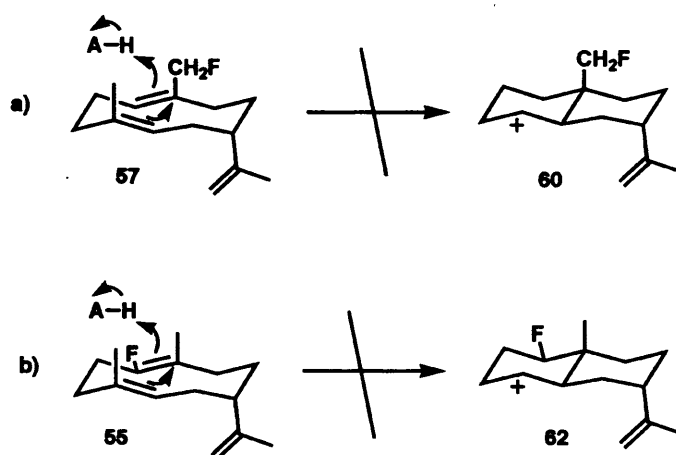
Figure 1.15: *Concerted formation of germacrene A with 6F-FPP and 14-FPP* - in a concerted mechanism formation of a fluorinated eudesmane cation is prevented by weak electronic overlap between C7 and C2 in 14F-FPP (a) and 6F-FPP (b)

1.5 Key questions in the study of AS-PR structure and function

The results of kinetics, mutagenesis, substrate labelling and substrate analogue studies described in the previous sections have all been used to shed light on the reaction mechanism catalysed by aristolochene synthase. However, there is still much uncertainty and debate regarding many aspects of this reaction. A summary of these forms the basis of the following paragraphs. In addition, the broader question of how sesquiterpene synthases are able to catalyse their reactions is an area of considerable interest and the study of AS-PR serves as a model for understanding the other members of this enzyme family.

1.5.1 The nature of the acid required in AS catalysis

As discussed in section Section 1.3.2 and alluded to in Figure 1.6 an acid is required to protonate the C6 C7 double bond of germacrene A and allow for the formation of eudesmane cation and hence aristolochene.

The identity of this acid has never been established. Tyr 92 (either on its own or *via* a proton shuttle mechanism) has previously been suggested [30, 71] on the basis of experimental results and comparisons with EAS from *N. tabacum.* [85] In the case of EAS (Section 1.3.2) the Y520F protein had its catalytic efficiency reduced to 3 % of WT-EAS and germacrene A was identified as the sole product. [85] Therefore, if Tyr 92 played a role in AS-PR directly analogous to that of Tyr 520 in EAS then one would

Table 1.4: Products identified in the product mixture of Y92F AS-PR - *Aristolochene production is still significant and hence Y92F is unlikely to act as the active site acid.* [71]

Product	Percentage of Yield
Germacrene A	28.6
Aristolochene	56.2
Valencene	5.9
β-Farnesene	1.8
β-Selinene	1.8
Selina-4,11-diene	2.1
α-selinene	3.6

expect the Y92F variant of AS-PR to exhibit similar behaviour - in particular one would expect no post-germacrene A derived species (Figure 1.6) in the final product mixture. However, the experiments of Calvert *et. al.* [71] reveal that only 30.4% of the products (germacrene A and β-farnesene in Table 1.4) identified are consistent with Y92 acting as the acid capable of protonating the C6 C7 double bond.

Therefore, since nearly 70 % of the products obtained from Y92F AS-PR are products which require the protonation of C6 C7 it is difficult to agree with the hypothesis that Tyr 92 plays an identical role to Tyr 520 in EAS.

Other active site acids have been proposed; the diphosphate group is one such possibility based on analysis of the crystal structure of AS-AT it was suggested that it may be positioned to act as a proton source. [43] However subsequent work, very recently published, and performed to challenge the hypotheses examined in this work have demonstrated that in AS-PR, protonation must occur above the plane of the germacryl

ring system. [92] Therefore, assuming AS-PR and AS-AT diphosphate binding occurs in a similar fashion, the diphosphate anion cannot act as the acid-base required for catalysis in AS-PR since its position is known to be beneath the plane of the germacryl ring system based on data from the binding geometry of fluoro-FPP analogues with AS-AT. [44]

Hydroxonium ion is another potential candidate for the acid required. The presence of a water molecule performing this role is not unprecedented; recent work on the maize sesquiterpene synthases TPS6 and TPS11 have shown that water may act as the acid in the protonation of a neutral intermediate [93] and thus cannot be excluded *a priori* from the reaction catalysed by aristolochene synthase.

1.5.2 The role of magnesium during AS catalysis

Clearly the generation of the diphosphate anion results in the formation of a charged species within the active site, requiring four units of positive charge to achieve neutrality - so the role of two magnesiums might be clear. Yet, it appears that AS-AT [43, 44] binds three magnesium cations - at least when it binds a single pyrophosphate in one subunit of a tetrameric form. [43] An explanation for the apparent need for a third magnesium was described in Section 1.2.1. This work suggested that, in AS-PR, two magnesium ions were required to bind diphosphate and the third closed the active site into an active form. [44] Yet this result must be interpreted with caution since it refers only to AS-AT with a substrate analogue and to the best of current knowledge a similar experiment has not been performed with AS-PR.

Studies have, however, suggested a role for all three magnesium ions in AS-AT [44] in which the third magnesium is required to trigger the cyclisation cascade, but it is difficult to envisage a laboratory experiment that could test this hypothesis. In addition the recent discovery that DCS does not contain the NSE motif [42] calls into question whether this model of magnesium binding and subsequent triggering of catalysis is valid across all the sesquiterpene synthases, since it is no longer true to say that all of the known members of this family exhibit the same metal binding ligands.

1.5.3   Intramolecular proton transfer mechanisms

The mechanism proposed in the literature and summarised in Figure 1.6 requires that germacryl cation is quenched to form germacrene A, then reprotonated to form eudesmane cation. This mechanism apparently requires the enzyme to form a high energy carbocation, quench it to form an olefin product, then protonate an alkene bond. Such a process typically requires a strong acid and therefore appears energetically inefficient and one possibility that has not been considered in AS-PR is an intramolecular proton transfer process as outlined in Figure 1.16.

Such a direct intramolecular proton transfer process (*e.g.* pathway *b* (Figure 1.16) would not be unprecedented in terpene biosynthesis: intramolecular proton transfer has been demonstrated in taxadiene synthase from the Pacific yew *Taxus brevifola* [94] and in abietadiene synthase from the grand fir tree *Abies grandis*. [95] Nor would water acting as the proton source (*e.g.* pathway *a* (Figure 1.16) lack precedent, as recent evidence from experiments involving maize sesquiterpene synthases suggest. [93]
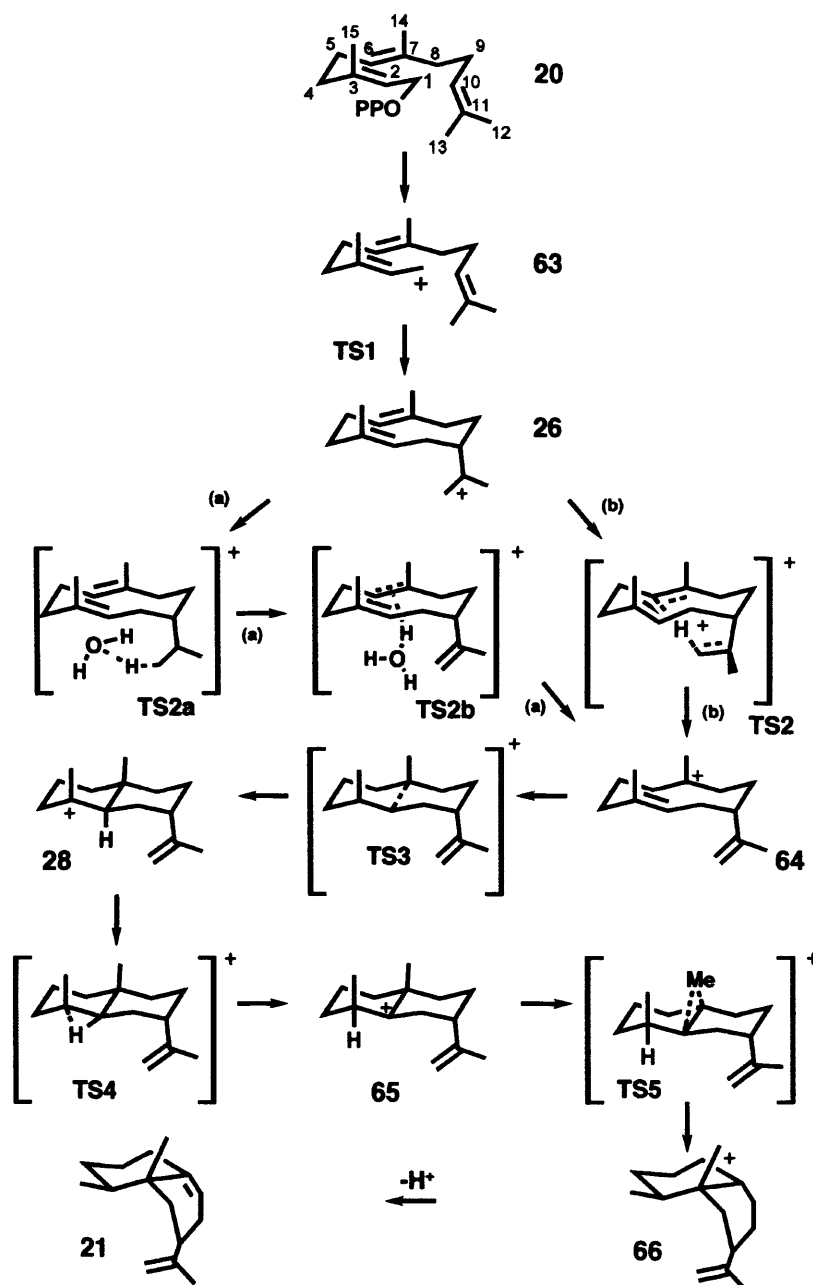
Figure 1.16: *Proposed intramolecular pathways for the formation of aristolochene* - proposed reaction mechanisms for biosynthesis of aristolochene either involving either (b) direct intramolecular proton transfer or (a) proton transfer mediated by a water molecule.

1.5.4 The pathway from eudesmane cation to aristolochene

The mechanism outlined in Figure 1.16 also elaborates on the proposed reaction mechanism following the protonation of the C6 C7 bond (irrespective of the proton source). The formation of bicyclic eudesmane cation (**28**) occurs by attack of the carbocation on C7 by the π electrons on the C2 C3 bond. Following formation of **28** a 1,2-hydride shift *via* **TS4** occurs to form intermediate **65**, which then undergoes a 1,2-methyl shift *via* **TS5** to form the subsequent intermediate **66**. Deprotonation of H_8 exclusively from the *Si* position [76] is the final step leading to the production of aristolochene (**21**).

1.6 Summary

Understanding the subtle details of how AS-PR controls this complex series of steps - from the folding of FPP into quasi-cyclic conformation, followed by cyclisation, possible protonation of an alkene followed by a series of carbocation rearrangements occurring with exquisite precision is undoubtedly a challenge for science. As has been outlined in the previous sections many complex and detailed experiments have been performed often combining techniques from molecular biology, crystallography, enzymology and synthetic and physical chemistry, in order to answer a great number of fundamental questions in this area.

The answers to such questions are of interest to anyone seeking to explain the physical basis and mechanistic details of AS-PR, and by extension, terpene synthase catalysis:

for example it is known that wild type AS-PR can produce valencene (Figure 1.10) as a minor side product [71] yet in the reaction mechanism in Figure 1.16 there is no reason *a priori* why the deprotonation of **66** should not occur at the C6 position leading to formation of valencene in significant yields. Yet, as discussed in the previous paragraphs AS-PR is specific enough to deprotonate almost exclusively at the $H8_{Si}$ position during the final step in the reaction. [76]

However, there remain a series of question which have proven difficult or impossible to answer using laboratory based techniques, yet must still be answered if a more complete understanding of AS-PR and the biosynthesis of aristolochene from it is to be gained. Therefore, the aim of this work is to expand upon previous work by using computational chemistry methods to investigate the reaction mechanism of AS-PR and to explore and answer questions raised in Section 1.5.

Chapter 2

Computational Chemistry Techniques and Applications

2.1 Introduction to Computational Methods

The growth in computational chemistry is inextricably linked to increase in power and decrease in price of computing hardware. New and more detailed theories of electronic structure have been developed to utilise this enormous computational power and it is fair to say that there is not an area of chemistry today which has not been investigated using computational methods.

Practical models exist to represent reality in a manner that has some application to a particular task. Even the act of representing aryl groups as a hexagon of alternating

single and double lines in a tradition 'arrow pushing' mechanistic diagram is an attempt at representing the complex π electron system that leads to the features associated with aromaticity. Despite the known flaws in the Kekulé structure (*i.e.* it does not represent the delocalised nature of the π electron system and cannot rationalise the apparently high hydrogenation energy of benzene) it is still a useful and powerful tool in explaining the reactivity of aromatic species.

Therefore it seems a model in chemistry must possess two qualities, it must represent some physical property and it must represent that property in such a way that can be practically used in the right circumstances.

Computational chemistry has evolved along several strands, each aiming to provide a useful, practical description of chemical properties, which aside from having theoretical interest may also be used to describe, predict and explain the properties of chemical species.

2.2 Molecular Mechanics

Since atoms and molecules possess both mass and charge and can exhibit vibrational (*i.e.* spring-like) properties it is possible to use the ideas of Newtonian mechanics to formulate models of molecular structure. The basic premise of molecular mechanics (MM) is that a description of the potential energy of a molecule can be defined as the sum of the potential energy functions which might make up a complete description of the molecule. Such terms include, but are not limited to, bond stretching, bond

angles, electrostatic forces and van der Waals interactions. By producing functions which are in good agreement with experimental data (and data obtained using more sophisticated calculations) over a suitably large reference set of molecules a useful force field can be produced - with the assumption that its functional form is applicable for wider range of species than that included in the reference set of molecules.

2.2.1 Bond stretching terms

Consider a diatomic molecule, AB with an internuclear distance (perhaps representing a single σ bond) r. This internuclear distance (which for most chemical purposes also describes a bond length) has some equilibrium value r^o at which the potential energy, U_{ab} is at a minimum. Any increase in the internuclear distance will increase the potential energy of the molecule, which will attempt to return to its equilibrium position.

Such a system can be considered, mathematically to be equivalent to two masses connected by a spring, and the functional form of the potential is well known in physics as Hooke's Law and is shown here in Equation 2.1

$$U_{ab} = \frac{1}{2}K_{ab}(r_{ab} - r_{ab}^o)^2 \qquad (2.1)$$

where K_{ab} is the spring constant for AB. One can therefore imagine how it might be possible to study a whole series of molecules, perhaps spectroscopically, in order

to determine force constants for each of them. However it is well known that bond lengths for a given functional group (*e.g.* the C C bonds in the n-alkanes) are all approximately equal (*ca.* 1.55 Å) and their vibrational properties are similar too (as evidenced by the IR spectra of simple alkanes). Therefore, it is not surprising to discover that it is possible to describe the bond stretching properties of many groups of molecules using the same spring constants.

Equation 2.1 also reveals an intrinsic problem with the MM approach. The functional form of the equation suggest that as the two nuclei are pulled apart their potential energy will continue to increase as $r_{ab} \rightarrow \infty$. Chemically, this does not make sense, since at infinite separation there exists no measurable force between two atoms. Thus the dissociation energy of a species can be described as the point at which the potential energy between the two atoms has reached its maximum as the nuclei are moved further apart.

With this in mind, it is more useful to use the Morse [96] potential to describe the bond stretching potential between two atoms, as shown in in Equation 2.2

$$U_{ab} = D_{ab}[1 - e^{-\alpha_{ab}(r_{ab}-r_{ab}^0)}]^2 \qquad (2.2)$$

which has a functional form much more closely resembling a dissociation potential energy curve. However, Hooke's law type descriptions of bond stretching can still be of use in modelling bond properties for the following reasons. First, in order to write fast, efficient software simple addition or multiplication operations are quicker to perform

than evaluating an exponential term such as the one described in Equation 2.1. This is not a real concern when considering small molecules on modern multi GHz processors but can quickly become a bottleneck when performing large molecular dynamics simulations on protein structures (Section 2.9). Second, near the equilibrium region (the region of interest in problems most amenable to MM calculations) there is little difference between Hooke's Law and the Morse potential. This initial example highlights the balance, in all computational chemistry, between theoretical rigour, computational cost and the desire to generate data that compare well with experimental observation.

2.3   Bond angle terms

2.3.1   Valence bond angles

Aside from bond length, bond angle is another property for which it is relatively easy to obtain experimental data (for example from crystallography) and which is crucial to modelling molecular structures. A typical functional form for the angle, $\theta = ABC$ is shown as Equation 2.3.

$$U_{\theta ABC} = \frac{1}{2}[k_{ABC} + k'_{ABC}(\theta_{ABC} - \theta^o_{ABC}) + k''_{ABC}(\theta_{ABC} - \theta^o_{ABC})^2 \ldots + k'^n_{ABC}(\theta_{abc} - \theta^o_{abc})^n]$$

(2.3)

Again, a decision must be taken as to how many terms to include when parameterising a force field. The balance between accuracy and computational speed means that a variety of different cut-off points are used when designing bond angle functions. Truncation at the quadratic term is used for the AMBER forcefield [97], the quadratic term alone is used in the CHARMM22 parameter set [98], whilst MM3 continues up until the 6th power. [99]

2.3.2 Dihedral angles

The dihedral angle, ω described by four atoms, ABCD is usually fitted to a function of the following form.

$$U_{\omega(ABCD)} = \frac{K_n}{2}\left(1 + \cos\left(n\omega - \gamma\right)\right) \tag{2.4}$$

The inclusion of a cosine term is required to describe the periodic nature of the potential energy function, apparent when one considers something such as the energy changes associated with eclipsed/staggered conformers of ethane. The constants in the expression alter the form of this cosine function: n is some real number, ω the periodicity of the function (*i.e.* how many cosine functions occur in one complete circular rotation), and γ some real number which shifts the location of the maxima and minima of the cosine function away from 0 and 180 degrees.

2.4 Non bonded interactions

So far a description has been given of the terms which describe the bonding between atoms in a molecule yet non bonded interactions, the interaction between charges on atoms, are required for a more complete picture of the potential energy function of a molecule. The following section discusses the methods MM employs to describe these interactions.

2.4.1 Electrostatic forces

Classically, the electrostatic energy of two point charges separated by a large distance is described as the product of the multipole moments on each charge. Since there is no limit to the order of multipole moments and each increase in order produces a more complicated tensor to describe the interaction, some approximation is required

This estimation of this interaction potential, U, can be made less complex if one assigns a charge, q_a and q_b to each atom a,b of a diatomic system and applies the following formula

$$U_{ab} = \frac{q_a q_b}{\epsilon_{ab} r_{ab}} \tag{2.5}$$

r_{ab} is the internuclear distance (bond length), and ϵ_{ab} is the effective dielectric constant between the two nuclei.

2.4.2 van der Waals interactions

Two atoms of a noble gas at infinite separation have no forces acting between them (except gravity, which is negligible for chemical purposes). Treatment of these atoms as hard spheres of equal radius means that when $r_{ab} < 2r_a$ (*i.e.* after the two spheres touch) the potential rapidly increases as the atoms are forced into each others radii. However, as the two atoms approach each other their electronic motions couple, such that an attractive force is generated, with a minimum of potential energy at some value of r_{ab}. This attractive (or dissipative) force is the van der Waals interaction or London force.

The functional form which best describes this interaction is known as the Lennard-Jones [100] potential the formula for which is given by Equation 2.6

$$U_{ab} = \frac{a_{ab}}{r_{ab}^{12}} - \frac{b_{ab}}{r_{ab}^{6}} \qquad (2.6)$$

where U is the size of the potential, r_{ab} the internuclear distance between atoms a and b and a_{ab} and b_{ab} are constants. The utilisation of even powers on r means that computation is speeded up considerably as r^{12} can be treated as $(r^6)^2$, which in computing terms is a trivial operation. In a similar fashion to the treatment of bond stretching in Section 2.2.1 some details of accuracy offered by exponential functions are sacrificed in order to obtain functional forms which produce fast computer code.

2.5 Overview of CHARMM

Section 2.2 detailed the common functional forms that an MM force field might take. A great deal of work, however, is required to obtain a set of parameters for all of the different atom types that might be of interest in a particular set of chemical problems. The CHARMM22 force field [98] as part of the CHARMM (Chemistry at HARvard Macromolecular Mechanics) [101] software package is used exclusively throughout this work. CHARMM is described by its authors as having been

> developed with a primary focus on the study of molecules of biological interest, including peptides, proteins, prosthetic groups, small molecule ligands, nucleic acids, lipids, and carbohydrates, as they occur in solution, crystals, and membrane environments.[1]

The choice of CHARMM and the CHARMM22 force field was guided by two principal factors. First, its stated purpose (study of biological macromolecules) corresponds with the aim of this work to study AS-PR. There is a large body of published literature in which CHARMM22 has been used to study biological phenomena and this gives confidence that it may be a useful tool for this study. Second, the Allemann group has experience of using CHARMM [102] and thus there is a practical motive for its use in this work.

There are, of course, a number of other force fields and software packages which could be used for the study of AS-PR. For example AMBER [97] and GROMOS [103] are

---

[1]http://www.charmm.org

both widely used. Although it is not possible to perform an *ad hoc* prediction of which force field would be 'best' for this particular problem there are a number of attempts to try and compare the utility of the various force fields and software packages in the literature. [104–106]

2.5.1 The CHARMM22 force field

In CHARMM the energy function for the vector, $\overrightarrow{R}$ of the co-ordinates of atoms is given by the following expression

$$
\begin{aligned}
U(\overrightarrow{R}) = &\sum_{bonds} K_b(b - b_o)^2 + \sum_{UB} K_{UB}(S - S_o)^2 + \sum_{angle} K_\theta(\theta - \theta_o)^2 + \\
&\sum_{improper} K_{imp}(\psi - \psi_o)^2 + \sum_{dihedrals} K_\chi(1 + cos(n\chi - \delta)) + \\
&\sum_{nonbond} \epsilon \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon_l r_{ij}}
\end{aligned}
\tag{2.7}
$$

where K_b, K_{UB}, K_θ, K_χ, and K_{imp} are the bond, Urey-Bradley, angle, dihedral angle, and improper dihedral angle force constants, respectively. The variables b, S, ϕ, ψ and χ are the bond length, Urey-Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle, respectively, with the subscript zero representing the equilibrium values for the individual terms. Coulomb and Lennard-Jones terms contribute to the external or non-bonded interactions; ϵ is the Lennard-Jones well depth and R_{min} is the x axis intercept on the Lennard-Jones potential curve; q_i is

the partial atomic charge; ϵ_l is the effective dielectric constant, and r_{ij} is the distance between atoms i and j. [98]

The Urey-Bradley term is included in CHARMM22 as an additional term to describe the bond stretching motion between 1,3 pairs of atoms. It is not included for all atom pairings, but introduced in order to ensure a better fit with experimentally obtained spectroscopic data where such data exist.

In general the values for the various equilibria contained within Equation 2.7 were obtained by study of crystallographic data, with some averaging if multiple values were obtained. In addition some *ab initio* (see Section 2.6) data were used where experimental data were lacking or inconsistent. The force constants for the geometrical parameters were fitted to experimental infrared and Raman spectra, again *ab initio* data were introduced where data were unavailable or inconsistent.

The CHARMM22 force field has found wide and heavy use in simulation of macromolecules, and aside from the theoretical rigour involved in its construction the fact that the original work has been cited over 2,400 times in the literature[2] is testimony to the acceptance this particular MM force field has gained in computational science.

2.5.2 Summary of MM methodology

Molecular Mechanics makes a few reasonable assumptions about the physical nature of atoms, molecules and the interactions between them and uses the mathematics of

---

[2]Information from Web of Science, accessed September 2009

classical mechanics to build up parameter sets. These parameter sets aim to reproduce experimental observables for the range of chemical species of interest. Theoretical rigour, and a degree of accuracy are sacrificed at the expense of producing fast, efficient code capable of treating tens of thousands of atoms.

Molecular mechanics in its simplest form does not treat the electronic structure of atoms in any detail, no information is available about the nature of the orbitals which contribute to chemical bonds, or the nature of chemical species far from minima (for example, transition state structures) and thus quantitative data regarding these important aspects of chemical behaviour are not considered. Although modifications to MM can be made to try and investigate these phenomena are possible (see, for example the review by Gao *et. al.* [107], it is often more useful to employ quantum mechanics based calculations.

2.6 *Ab initio* quantum mechanical methods

It is now well known that the Newtonian view of mechanics is insufficient to explain many phenomena. The black-body radiation problem led Planck to his hypothesis that energy was quantised, which in turn led to de Broglie to suggest that electrons have a discrete wavelength depending on their momentum. That electrons have an calculable wavelength ultimately led to Schrödinger formulation of wave mechanics and the essential equation which serves as a starting point for most discussion of

quantum mechanics.

$$E\Psi = \hat{H}\Psi \qquad (2.8)$$

That is, the energy of a wavefunction, Ψ, is an eigenvalue of some (Hamiltonian) operator, $\hat{H}$. All that is required to obtain the energy (or any other quantity) of a system is a wavefunction and an appropriate operator. The product of the wavefunction with itself (or strictly speaking it complex conjugate) integrated over all space is unity (*i.e.* the particle exists somewhere in space). The only other conditions which need to be imposed on the wavefunction are that it must be continuous, differentiable and single valued over all space.

Furthermore, Schrödinger demonstrated [108,109] that for any approximate wavefunction, Φ operating on a Hamiltonian, the energy eigenvalue obtained is always greater than the exact energy E_o. This is expressed in the following equation.

$$E[\Psi] = \frac{\langle \Phi | \hat{H} | \Phi \rangle}{\langle \Phi | \Phi \rangle} \geq E_o \qquad (2.9)$$

The power of the variational method is that as long as a few simple conditions are obeyed then the route to the best wavefunction is to attempt to minimise the energy value obtained by application of Equation 2.8. With respect to atoms, there is no requirement that the orbitals might make up a wavefunction bear any relationship to the canonical set of atomic orbitals. Indeed there are excellent mathematical reasons

why this should not be the case.

What about the form of the Hamiltonian? It must consider the contribution to the energy of a molecule as being made up of the kinetic energy of the nuclei and electrons, plus the attractive and repulsive forces that exist between them. Conceptually this is somewhat similar to the formulation of MM force fields which also account for the attractive and repulsive forces between atoms. Equation 2.10 shows an operator which may serve as a QM Hamiltonian.

$$H = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k}\nabla_i^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i<k} \frac{e^2}{r_{ij}} + \sum_k \frac{e^2 Z_k Z_l}{r_{kl}} \tag{2.10}$$

where i and j are electrons, k and l nuclei. The mass, m, of the electron is m_e and m_k is the mass of nucleus k (the charge of which is Z_k) and $\hbar$ is the Dirac constant. ∇^2 is the Laplace operator defined in Equation 2.11. Thus the Hamiltonian in Equation 2.10 is made up of two terms representing kinetic energy of electrons and nuclei, followed by a term representing the attraction between nuclei and electrons. The final two terms represent internuclear and interelectron repulsion - Coloumbic nature and similar to the expression in Equation 2.5.

$$\nabla^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \tag{2.11}$$

2.6.1 The Born-Oppenheimer approximation

Equation 2.10 explicitly includes a kinetic energy term associated with the nuclei of a molecule. Since the mass of even the hydrogen nucleus is far, far greater than the mass of an electron it is possible to introduce an approximation in which we assume that on the time-scale of electronic motion, the nucleus is essentially fixed - the Born-Oppenheimer approximation. [110] Without this approximation it would not be possible to talk about optimum geometries for a molecule, since these assume that the nuclei of the elements occupy fixed positions. In essence, the Born-Oppenheimer approximation ignores the quantum mechanical nature of the nucleus and treats it as a point charge, which for most chemical purposes is valid and reasonable. Thus it is possible to recast Equation 2.10 into the following form

$$(\hat{H}_{el} + V_n)\Psi_{el} = E_{el}\Psi_{el} \tag{2.12}$$

where V_n is the nuclear-nuclear repulsion term and the wavefunction Ψ_{el} is purely an electronic wavefunction. The Born-Oppenheimer approximation, coupled with the variational principle, means that the only challenge to compute the energy of a species is to find a suitable wavefunction which produces a value for E_{el}. Upon finding such a wavefunction the only improvement we can make is to attempt to find another, lower energy, wavefunction, since this will better approximate the true energy.

2.6.2   Molecular orbitals and the LCAO method

In the case of simple, one electron systems (for example the hydrogen atom) there is no need to guess a trial wavefunction and attempt to improve upon this using the variational theorem. The exact solutions for this problem are the s,p,d, etc orbitals - the functional forms of which are readily available in the literature (*e.g.* Atkins' classic textbook.) [111]

Chemically, it makes sense to consider that in a molecule the orbitals within that molecule might be made up from some combination of atomic orbitals. Thus a trial molecular wavefunction, Ψ can be constructed by some combination of atomic wavefunctions, ϕ in Equation 2.13.

$$\Psi = \sum_{i=1}^{N} = a_i \phi_i \tag{2.13}$$

In which there are N sets of atomic wavefunctions and a is some coefficient. Because Ψ is formed by adding together these atomic wavefunctions this is called a linear combination of atomic orbitals (LCAO). The set of ϕ orbitals is known as the basis

set. This expression for Ψ can be substituted into 2.9 to give

$$E = \frac{\left\langle \left(\sum_i a_i \psi_i\right) \middle| H \middle| \left(\sum_j a_j \psi_j\right) \right\rangle}{\left\langle \left(\sum_i a_i \psi_i\right) \middle| \left(\sum_j a_j \psi_j\right) \right\rangle}$$

and since the coefficients are numerical constants.

$$= \frac{\sum_{ij} a_i a_j \left\langle \psi_i \middle| H \middle| \psi_j \right\rangle}{\sum_{ij} a_i a_j \left\langle \psi_i \middle| \psi_j \right\rangle} \tag{2.14}$$

$$= \frac{\sum_{ij} a_i j_i H_{ij}}{\sum_{ij} a_i a_j S_{ij}}$$

The integral H_{ij} and S_{ij} are known as the resonance and overlap integrals. The variational principle allows us to minimise the coefficients a_{ij} so that the following is true.

$$\frac{\delta E}{\delta a_{ij}} = 0 \tag{2.15}$$

Equation 2.15 produces a set of equations which must satisfy the following equation

for all value of k.

$$\sum_{i=1}^{N} a_i(H_{ki} - ES_{ki}) = 0 \tag{2.16}$$

Solving this differential equation will give a set of N solutions, each of which corresponds to a value of E.

2.6.3 The Hartree-Fock Self Consistent Field method

The starting point for almost all practical applications of the concepts discussed in this section is the Hartree-Fock (HF) self consistent field method. [112–114] It is is possible to construct a wavefunction Ψ as the product of a number of electronic wavefunctions, χ_i.

$$\Psi = \chi_1\chi_2\chi_3\cdots\chi_n \tag{2.17}$$

Each electron occupying these orbitals has the property of spin (here given the label α_i) thus the orbitals represented in Equation 2.17 are properly called spin orbitals. Each χ_i must satisfy the Pauli exclusion principle. [115] The Pauli exclusion principle can be maintained by representing Ψ as the Slater determinant [116,117] of

Ψ consisting of N spin orbitals χ.

$$\Psi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\alpha_1) & \chi_2(\alpha_1) & \chi_3(\alpha_1)\dots & \chi_N(\alpha_1) \\ \chi_1(\alpha_2) & \chi_2(\alpha_2) & \chi_3(\alpha_2)\dots & \chi_N(\alpha_2) \\ \chi_1(\alpha_3) & \chi_2(\alpha_3) & \chi_3(\alpha_3)\dots & \chi_N(\alpha_3) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \chi_1(\alpha_N) & \chi_2(\alpha_N) & \chi_3(\alpha_N)\dots & \chi_N(\alpha_N) \end{vmatrix} \tag{2.18}$$

Each of the χ_i orbitals are eigenfunctions of Equation 2.19

$$f_i = -\frac{1}{2}\nabla_i^2 - \sum_{k=1}^{M} \frac{Z_k}{r_i k} + V_i(j) \tag{2.19}$$

where the final $V_i(j)$ term is the potential associated with each electron interacting with the potential of other electrons. Thus if one creates a set of χ_i orbitals - by logic, intuition or guesswork - one can create a related set of f_i too. The Schrödinger equation for each χ_i is then solved. As a result a new set of χ_i is generated - and because of the variational principle evaluation of the energy provides information on how close one is to the 'true' value. Once the χ_i cease to produce a pre-determined change in energy with each iteration the self-consistency condition is reached.

This f_i can be used to compute the energy of the system and a new set of χ_i with which it is possible to re-compute the energy. Because of the variational principle, as long as the computed value of E is getting smaller with each iteration, the set of

orbitals (and the $V_i(j)$ term in Equation 2.19) are becoming closer to the true value.

The principle drawback in the HF approach (aside from the computational expense which can, by brute-force approaches, be overcome) is that each electron does not interact *directly* with the other electrons in the system, instead it feels the potential in an averaged way (Equation 2.19). Considering phenomena such as van der Waals forces, it is apparent that electrons do couple their motions so as to minimise repulsion. This inability to take into account the electron correlation problem is the principal source of inaccuracy in the HF method.

2.6.4 Basis sets

Equation 2.13, which can be applied to the HF equations introduced a requirement for a basis set of orbitals which can be used to construct molecular orbitals. A basis set of infinite size would be able to compute energy to the limit of accuracy of the HF method, but this is not a practical proposition. Instead basis sets of various sizes have been constructed which aim to produce accurate results whilst minimising computational cost.

Slater type orbitals (STOs) [118] vary as e^{-r}, but are limited by the fact that analytical solutions to integrals involving this function, as applied to problems in quantum mechanics are not available. Reliance on numerical solutions is a computational expensive task, and thus Gaussian type orbitals (GTOs, Equation 2.20) [119,120] are

preferred

$$\phi(x, y, z) = f(x, y, z)e^{(-\zeta r^{-2})} \tag{2.20}$$

where $f(x, y, z)$ is a complicated polynomial expression, and ζ describes the width of the GTO. GTOs may be mathematically preferable, but their functional form does not reflect reality in the same manner as STOs; GTOs decay to zero too quickly, and do not describe the nuclear cusp. Therefore the challenge is to produce combinations of GTOs which closely resemble STOs. A linear combination of 3 GTOs produces a function which approximates a STO and is labelled STO-3G (Slater Type Orbital from three Gaussians).

The minimal basis set uses the equivalent of one STO to describe each molecular orbital and is the least accurate basis set. To improve accuracy additional functions are included. For example, the 6-31G basis set uses one STO function (made up of six Gaussian functions) to describe 'core' electrons (*i.e.* ones not typically involved in chemical bonding) and a further set of two STOs for valence electrons, one containing three Gaussians, the other a single Gaussian. Furthermore, additional functions may be added. For example, the basis set 6-31+G(d,p) includes p and d polarisation functions on atoms (*i.e.* p polarisation on atoms with valence electrons in the s-shell, d polarisation on atoms with valence electrons in the p-shell). The $+$ indicates that a diffusion factor has been included to better model STO behaviour at regions further from the nucleus.

2.6.5 Semi empirical methods

Finding solutions to the HF equations with appropriately large basis sets is computationally demanding, and routine, accurate, calculations using this method are currently limited to the study of molecules containing at most dozens of non-hydrogen atoms. However unlike the MM methods the power of HF, and HF-derived techniques is their ability to access information about non-equilibrium structures. For example, describing a transition state or analysis of the vibrational modes of an excited state molecule are not possible with an MM force field.

A number of semi-empirical methods have been developed which attempt to reduce the complexity of the SCF calculations described in Section 2.6.3 with the aim of speeding up calculations whilst avoiding too great a loss of accuracy. Common to these methods are the exclusion of all electrons other than valence electrons and the setting of the resonance integral, S, in Equation 2.14 to unity, and all non-diagonal element in the overlap integral, H, to zero. This vastly reduces the number of integral steps which need to be performed, and the value of those integrals is parameterised over a suitable range of systems.

Most modern methods are based on the NDDO (Neglect of Differential Diatomic Overlap) model. [121, 122] The key feature of this method is that all integrals are set as zero, except for the integrals between identical orbitals on the same atom, and identical orbital on another atom *i.e.* integrals involving overlap between different basis functions is ignored. That is to say, integrals are computed in $(\mu\nu|\lambda\sigma)$ if μ and ν are on the same atom and λ σ are on the same atom (but not necessarily a different atom from μ and ν). MNDO (Modified Differential Overlap) [123, 124] is based on

the NNDO formalism, and uses spectroscopic data to evaluate two electron one centre integrals. Two centre, two electron integrals are treated as multipoles - that is the interaction between two s orbitals is modelled as a point charge, the *sp* interaction by a dipole across the *sp* axis and the *pp* interaction treated as a quadrapole.

MNDO tends to overestimate repulsion between atoms (*i.e.* it does not treat dispersion effects correctly). Austin Model 1 (AM1) [125] includes additional attractive terms in the MNDO model to attempt to overcome this problem, and PM3 (Parameterised Model 3 [126]) was a systematic re-parameterisation of AM1.

The advantage of the semi-empirical methods is the speed compared to *ab initio* Hartree-Fock calculations (especially those with large basis sets), whilst retaining the ability to investigate molecular structures far from equilibrium positions (*e.g.* transition state structures). However, if a particular chemical problem does not have a semi-empirical parameter set that is well suited to the problem then problems arise. Although if one recognises and accepts the theoretical underpinnings of a given model, then it is possible to reparameterise some or all of the method and use specific reaction parameters (SRP) [127] designed for a smaller subset of problems.

2.7 Density Functional Theory

There is no direct physical interpretation of a wavefunction. It is a mathematical creation, which when operated upon by a suitable Hamiltonian, returns an eigenvalue which represents a given physical quantity. On the other hand, electron density (ρ) is a

simple concept. The relevant breakthrough occurred when Kohn and Hohenberg [128] demonstrated that the electron density of a system maps exactly the ground state wavefunction of a system and that the ground state density also minimises the total energy of the system, *i.e.*

$$E[\rho] \leq E[\rho_0] = E_0 \tag{2.21}$$

$E[\rho]$ is a functional - a function which takes another function as its argument and returns a scalar. However, Kohn and Hohenberg only produced an existence theorem, that is to say a functional $E[\rho_0]$ must exist but says nothing about what form such a functional might take.

Kohn (who won that 1998 Nobel Prize in Chemistry for his work on developing DFT) and Sham were able to develop this existence theorem in order to create a tool with practical applications. [129] In a manner analogous to HF theory (section 2.6.3) a set of orbitals, χ_i are obtained which minimise the energy of the system.

$$h_i^{KS}\chi_i = \epsilon_i\chi_i \tag{2.22}$$

These orbitals are operated upon by the Kohn-Sham (KS) operator.

$$h_i^{KS} = -\frac{1}{2}\nabla^2 - \sum_k^{nuclei} \frac{Z_k}{|r_i - r_k|} + \int \frac{\rho(r')}{|r_i - r'|}dr' + V_{xc} \tag{2.23}$$

The V_{xc} term is a derivative of the exchange-correlation energy with respect to ρ. The exchange-correlation energy is a correction to the kinetic energy term for non-interacting electrons (because the kinetic energy term is a Slater determinant of one electron orbitals) and the correction for electron-electron repulsion caused by Pauli's exclusion principle. The exact exchange energy (and hence V_{xc}) is unknown and as a result it is possible that the Kohn-Hohenberg variational property may be broken once this approximation is introduced. Therefore DFT is an exact theory (as long as the exchange energy is known) but its application gives approximate results, where as HF theory is an approximation whose equations may be solved exactly.

Thus, in order to achieve the most accurate results using DFT one needs to determine the exchange-correlation functional precisely. The Kohn-Sham and Kohn-Hohenberg theories provide no direct clues as to how this might be accomplished. Many approaches have been attempted: the local density approximation (LDA) assumes that the exchange energy at a given point is the same as the exchange energy of a uniform electron gas (*i.e.* an ideal gas in which electrons are completely dissociated from their ions) of a density equal to the density at the point of interest. [130, 131]

The electron density in a real molecular system is non-uniform, so the LDA may not be sufficiently accurate. To improve upon accuracy, the generalised gradient approximation (GGA) has been developed. Essentially this adds a further correction to the LDA functional to better account for electronic exchange. Several of these exchange functionals (including Becke's 'B' functional [132]) include a further parameter obtained from empirical data (which also means that very often DFT is not strictly an *ab initio* method). Correlation energy is treated in a similar manner, with various empirical and non-empirical functional forms existing.

Hybrid exchange-correlation functionals are based around a combination of HF and LDA based exchange and a functional for the correlation energy. Of these, by far the most popular is B3LYP [132,133] which includes HF and the Becke exchange alongside LDA exchange, combined with Lee, Yang and Parr's correlation functional. [134]

Unlike *ab initio* methods, where it is, in principle, possible to continue to improve the accuracy of results until such a point that they are indistinguishable from the results of the Schrödinger equation, this is not possible with DFT methods. There is no 'best' functional and no systematic improvements are possible. Numerous studies have been performed in an effort to assess the qualities of the range of DFT methods available [135–141] and, in general, the hybrid GGA based DFT methods offer the best accuracy for a given problem (compared to high level *ab initio* results). Despite this there is no clear consensus as to what the most suitable functional is for a given problem. It is therefore important to make informed choices regarding the selection of a DFT method for a given problem and, in general, avoid relying on using the same functional for all problems.

Despite these disadvantages DFT is an extremely powerful and popular technique in modern computational chemistry. For the same computational effort, a DFT calculation produces a more accurate result than en equally expensive Hartree Fock derived *ab initio* result. As a result DFT is routine tool for exploring many chemical problems. Indeed, in solid state physics (where DFT has its historical roots) and materials science it is the *de facto* standard computational method.

2.8 Quantum mechanical / molecular mechanics (QM/MM) methods

Molecular mechanics offers a fast (compared to *ab initio* and DFT techniques) method for predicting the behaviour of large molecular systems in regions close to their equilibrium, but does not readily allow us to make quantitative statements about structures far from equilibrium (*e.g.* transition states) or provide any information about the electronic environment which might make up a particular molecule - for this information we need to use the computationally demanding DFT, or *ab initio* / semi-empirical molecular orbital theories. Even the use of semi-empirical MO theory is restricted to the routine study of molecules considerably smaller than enzymes such as AS-PR. Since it is not usually practical to apply semi-empirical QM methods to large biomolecules the combined quantum mechanics / molecular mechanics (QM/MM) approach has been developed.

The basic premise for all QM/MM methods is that a small region of the system being studied, for example the active site of an enzyme, is treated using a QM based methods, with the remainder of the system, including bulk solvent, being treated using an MM force field. Such an approach aims to combine the speed of MM with the detail and accuracy of QM methods. The obvious difficulty is how to describe the interaction between the QM region and the MM region.

2.8.1   The Generalised Hybrid Orbital method

Of the numerous methods for describing the QM/MM boundary, the one which is currently well supported in the CHARMM software package is the Generalised Hybrid Orbital (GHO) method. Developed by Gao and co-workers [142], the GHO method divides a system up into two regions, one to be treated with QM, the other with MM. A single atom is treated with both QM and MM (atom B in Figure 2.1) and this atom connects to both the purely QM atoms and the MM region (as shown in Figure 2.1). The bond between atoms B and QM is usually a σ bond and atom B generally has sp^3 hybridisation, thus B and QM could easily represent a single bond between two carbon atoms.
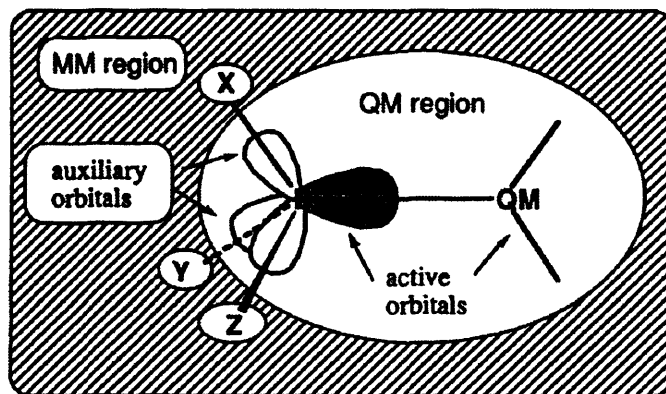
Figure 2.1: Schematic representation of the division between QM and MM regions in the GHO method. [142]

There are four orbitals on atom B, and only one of them (as indicated in Figure 2.1) is involved in QM calculations, with the remaining three sp^3 type orbitals forming a source of potentials felt by the QM atoms. This active orbital is fully involved in SCF calculations, allowing charge to transfer across the boundary. GHO theory is applicable to any SCF based methodology, but in most instances is used with semi-empirical MO theory.

There are other QM/MM methods in use (for example the link atom method which caps a valence orbital on the QM and MM parts with a hydrogen atom [143,144]) and there have been attempts to compare the quality of results from these different methods. [145,146] In general GHO gives good results, but in this work its use is primarily dictated by the fact the it is the only fully functional QM/MM implementation available in CHARMM v29.

2.9 Molecular Dynamics

Enzymes are dynamic systems, in constant motion within their solvent environment. Many enzymes undergo conformational changes in order to perform catalysis. To fully understand the molecular basis for catalytic processes there must be accounting for this dynamic behaviour.

Molecular dynamics is the method employed to study the motion of enzymes as a function of time: since the mass of atom is known and it is possible to assign velocities to each atom's co-ordinates, this allows the total momentum to be determined. For a system whose initial co-ordinates and momenta are defined, it is possible to apply Newton's equations of motion and describe the trajectory of the system over time.

The accuracy of such a calculation depends in part upon the time step δt used to propagate the next step in a dynamics simulation. A smaller time step results in a more precise calculation, but with an increase in computational cost. Different methods exist for the fast, efficient solution of this problem. The Verlet 'leapfrog'

algorithm [147] calculates new velocities **v** and positions **r** based on the following equations.

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v} \left(t + \frac{1}{2}\delta t \right)$$
$$\mathbf{v}\left(t + \frac{1}{2}\delta t \right) = \mathbf{v}\left(t - \frac{1}{2}\delta t \right) + \delta \mathbf{a}(t) \tag{2.24}$$

Thus, position is based on the velocity (**v**) half a time step (δt) away. Acceleration, **a**, (obtained from the force field calculations) is only computed every other half time step and the velocity to be used is derived from acceleration. The principle advantage of the Verlet leapfrog algorithm is its relative speed compared to other methods which have greater stability (for example Runge-Kutta methods) but require the expensive computation of high-order derivatives. As always with MD based methods there must be a trade-off between speed and practicality.

2.9.1 The SHAKE algorithm and Langevin and stochastic boundary dynamics

Despite the use of MM forcefields and propagation technique such as Verlet's leapfrog algorithm (*vide supra*), MD simulation on a large macromolecule can still be considered computationally demanding. Therefore a number of techniques and models have been developed to reduce the complexity of calculations further still, with the aim of avoiding as much loss of accuracy is possible.

As the length of time steps tends towards zero the accuracy of an MD simulation increases, but for practical purposes time steps in the region of 1 fs are usual. In order to increase speed and accuracy, high frequency C-H bond stretching vibrational modes (which do not generally contribute much to the overall dynamic motion of a protein) may be frozen using the SHAKE algorithm. [148,149]

Langevin dynamics are an attempt to include the effects of friction and the random forces occurring through the interaction of solvent with solute; the basic formalism is described in the following equation

$$F(\mathbf{x}) = -V\mathbf{x} - \gamma v(\mathbf{x}) + R(\mathbf{t}) \qquad (2.25)$$

where F is the force, V is the potential generated by the MM force field, γ is the diffusion coefficient and R is a random force acting upon the system (itself a Gaussian distribution with a mean value of zero).

Stochastic boundary molecular dynamics (SBMD), first developed by Brooks III and Karplus [150] is used to treat large systems in which only atoms within a given radius of a point are assumed to be of interest (for example the active site of an enzyme might form the centre of an SBMD simulation). Outside there is a boundary region where Langevin dynamics are applied before the outer radius is reached - at the actual boundary a deformable potential is applied preventing loss of solvent molecules. Atoms outside the boundary are not explicitly included in calculations, therefore regions far from an enzyme active site and presumably not involved in overall

reactivity need not be included explicitly in calculations.

2.10   Free energy sampling in QM/MM systems

As previously discussed, chemical problems are not confined to regions around the equilibrium geometry of a given structure. The study of bond making and breaking processes, the concept of transition state and activation energy all require the investigation of non-equilibrium structures of molecules.

The probability, P, of a given value of some chemical quantity (for example a bond length or dihedral angle) occurring is proportional to the relative free energy of the conformation. Thus, over infinite time all possible conformations will occur. In an MD simulation only a single group of molecules of reactants are studied, and therefore the probability of observing random, non-equilibrium events such as bond breaking is vanishingly small, especially as the longest routine MD experiments simulates at most a few dozen nanoseconds. Therefore, in order to solve problems of practical interest to chemists methods to sample non-equilibrium phenomena are required.

Umbrella sampling is a method which allows rare events - such as bond breaking - to be investigated during an MD simulation. A reaction co-ordinate, x, is defined so as to represent the process of interest, and the reaction co-ordinate is divided up into suitably sized segments. A biasing potential, U, which is a function of reaction co-ordinate is applied at each section of the reaction co-ordinate, such that most sampling occurs within a very narrow region. Typically this biasing potential is a

simple quadratic function and after each simulation a function of the following form
is obtained

$$W(x) = -k_b T \ln P(x) - U(x) + F \qquad (2.26)$$

Where W is the potential of mean force (PMF) which is free energy as a function of
the reaction co-ordinate, x.

After performing an appropriate number of simulations across the range of the reaction
co-ordinate for enough time at each step to produce a statistically significant set of
data, the final step is to combine these data to produce the true potential of mean force
(PMF), which is free energy as a function of reaction co-ordinate. A common method
is Kumar and Kollman's Weighted Histogram Analysis Method [151] (WHAM) which
combines the results of the simulations to produce a combined PMF.

2.10.1 The Weighted Histogram Analysis Method (WHAM)

The essential features of the WHAM method is that $P(x)$, the best estimate of the
unbiased probability, is given by Equation 2.10.1

$$P(x) = \frac{\sum_{i=1}^{N_{sims}} n_i(x)}{\sum_{i=1}^{N_{sims}} e^{([F_i - U_{bias,i}(x)] / k_B T)}} \qquad (2.27)$$

where N_{sims} is the number of simulations, $n_i(x)$ is the number of counts in the histogram bin associated with x and $U_{bias,i}$ (the biasing potential from i) and F_i (the free energy shift at i) are unknowns. F_i is defined by Equation 2.10.1.

$$F_i = -k_B T ln(\sum_{x_{bins}} P(x) e^{[-U_{bias,i}(x)/k_B T]})$$
(2.28)

To perform a WHAM free energy simulation the chosen reaction coordinate is divided up into a number of bins, the minimum of the biasing potential is placed at the centre of each chosen point on the reaction coordinate. After a suitably large sampling time the biasing potential is moved to a new position along the reaction coordinate. The results of the simulations (*e.g.* MD simulations) are used to created a histogram of relative $P(x)$ against x. The WHAM equations are then used to calculate the PMF by solving the equations until self-consistency.

2.11 Application of computational techniques to terpene chemistry

The techniques described in the previous paragraphs have been employed in order to investigate the reaction mechanisms catalysed by sesquiterpene synthases. In particular, the Tantillo group have undertaken a number of such studies [152–157] principally employing gas phase DFT. The results of some of these experiments are

of particular relevance to a study of AS-PR.

2.11.1   Computational studies of trichodiene synthase

Trichodiene synthase has been previously proposed to catalyse the synthesis of trichodiene by the mechanism outlined by pathway **A** in Figure 2.2. In this pathway FPP isomerises to form nerolidyl diphosphate which then loses diphosphate and cyclises to form bisabolyl cation. [157] Alkene-cation cyclisation leads to **69** and a 1,4 hydride transfer leads to the formation of curpenyl cation. A further series of 1,2 methyl shifts and a final deportation step then leads to the formation of trichodiene.

Formation of **69** involves the formation of a secondary cation from a tertiary cation - in general an unfavourable step. However, in the Tantillo group's study of trichodiene synthases [157] it was suggested that this apparently unfavourable energy barrier could be compensated for by the formation of a σ bond from a π bond. However, in a series of DFT calculations at the B3LYP/6-31+G(d,p) level it was not possible to locate a minimum energy structure for **69**.

In pathway **B** however, **70** was formed which would result from a cyclisation step and 1,2-methyl shift from bisabolyl cation to **70**. Following formation of **70** a 1,4-hydride shift followed by a 1,2-methyl shift (the reverse of the shift which led to the formation of **B**) occurs leading to curpenyl cation and hence trichodiene. The steps involved in the pathway to curpenyl cation *via* **70** are described as being 'asynchronous but concerted' *i.e.* a temporary methyl shift occurs in order to prevent the formation of a secondary carbocation.
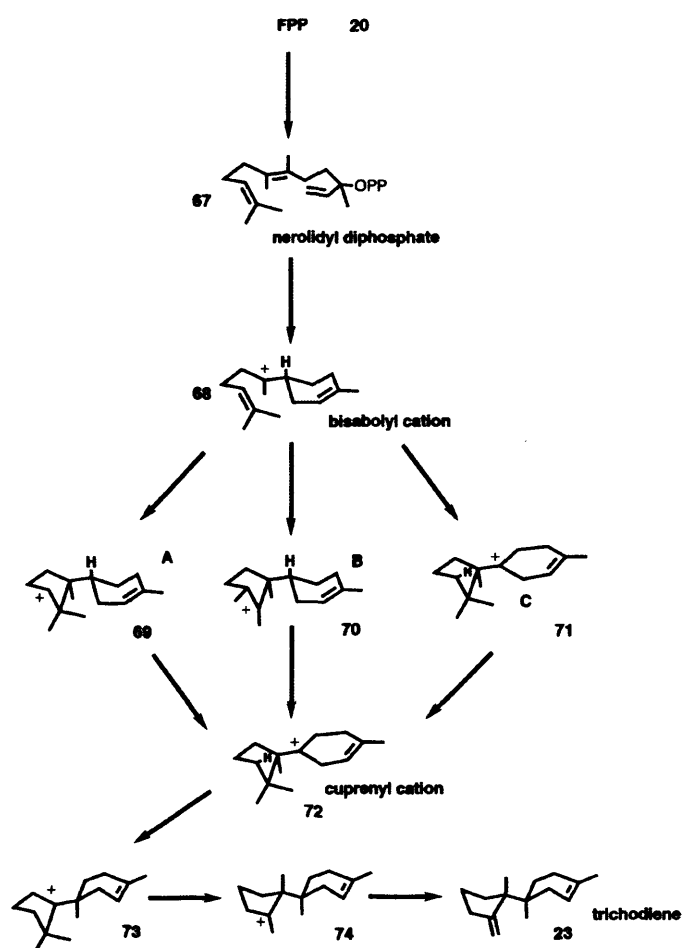
Figure 2.2: *Proposed mechanisms for the synthesis of trichodiene synthase -* computational results suggest that the intramolecular proton transfer in pathway C is more likely. [157]

There is a third possible pathway (pathway **C**), which involves an intramolecular 1,5-proton transfer from bisabolyl cation to form **71** (of which there are a number of possible conformers). In this pathway a π bond acts as the base (in a manner superficially analogous to the pathway *via* **TS2** in Figure 1.16). Crucially, the calculated energy barrier for pathway **C** (*via* transition states identified in the study) was lower by *ca.* 20 kcal mol^{-1} compared to pathway **B** - the exact value depends upon which conformer of **71** is considered.

Although experimental evidence [41] does not preclude any of pathways **A**, **B** or **C** occurring the authors note that that none of the deportation products of **69** or **70** have been detected (and on-pathway premature deprotonation products are observed in other sesquiterpene synthases *e.g.* AS-PR) but iso-γ-bisabolene (Figure 2.3) has recently [158] been detected in the product mixture. The presence of iso-γ-bisabolene may be explained by considering that it is a premature deprotonation product of **71**. Hence, computational data combined with detailed study of experimental data has been used to demonstrate that intramolecular proton transfer mechanisms are feasible and on balance, more likely, for at least one sesquiterpene. The implication of this for the proposed intramolecular proton transfer mechanisms in AS-PR are that it provides a clear precedent for using gas phase experimental data as a starting point for a detailed exploration of the possible pathways to aristolochene.

2.11.2 Computation studies of pentalenene synthase

The commonly proposed [34, 78, 159–161] mechanism for the reaction catalysed by pentalenene synthase involves the formation of farnesyl cation, followed by ring closure
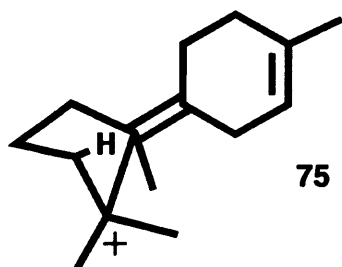
Figure 2.3: *Iso-γ-bisabolene* - the presence of this product in the trichodiene synthase product mixture is evidence in favour of an intramolecular proton transfer mechanism.

to form the humilyl cation (intermediate **76** in Figure 2.4) followed by a ring closure step (the calculated barrier height for which was *ca.* 4-6 kcal mol^{-1}) [156] forming **80** (in Figure 2.4) followed by hydride shift to give the secoilludyl cation, which then undergoes attack by the π system in **79** to form **81** followed by deprotonation to give pentalenene **24**.

The Tantillo group's study failed to identify structure **78**. Instead they were able to identify a pathway involving the formation of intermediate **80**, which was a concerted process forming the tricyclic system made of a four, five and six membered ring. The barrier for this was 5-6 kcal mol^{-1} [156]. To form **81** from **80** was *via* a transition state, **TS6**, involving a hydride and methyl shift. This too was a concerted process, but one in which the ring opening and cyclisation occur before the functional group migrations. There is a high (compared to other steps in this mechanism) barrier of approximately 30 kcal mol^{-1} associated with this step. Thus, if this process does occur in the enzyme, this step may well be the formal rate determining step and the role of pentalenene synthase may be to to stabilise the formation of **TS6**.

In failing to observe **78**, there is a comparison to be made with AS-PR. In Figure 1.16
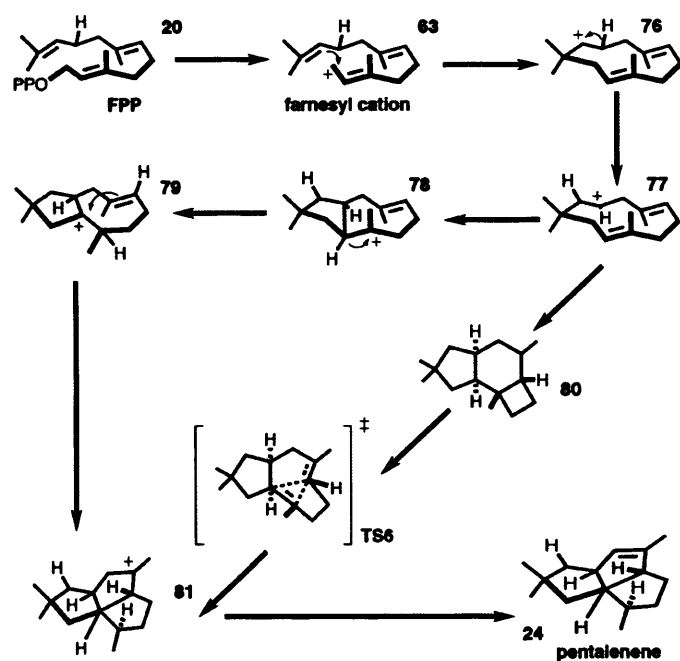
Figure 2.4: *Proposed and novel mechanisms for the formation of pentalenene -* intermediates **78** and **79** in previously proposed mechanism are avoided by the formation of intermediate **80**

the formation of eudesmane cation (**28** in Figure 1.16) may proceed *via* structure **64** and **TS3**. However, the computational evidence here suggests that a similar step (*i.e.* the formation of a ring by the attack of a carbocation by a π system) is likely to be concerted and may proceed without formation of the predicted intermediates.

These examples from the literature demonstrate the use of gas-phase computational techniques for the study of specific problems in sesquiterpene chemistry and their ability to add new evidence and insight into the basis for sesquiterpene catalysis. The aim of the work described in the following chapters is to continue this theme and to then attempt to address the problem by modelling the complete enzyme environment.

Chapter 3

Gas Phase Investigation of the Conversion of Farnesyl Diphosphate to Aristolochene

3.1 Introduction

Chapter 1 outlined questions regarding the reaction mechanism catalysed by AS-PR and the experimental evidence for and against proposed reaction mechanisms. Chapter 2 described how computational chemistry can be used to answer questions which might not otherwise be addressed experimentally or to provide a theoretical basis to explain empirical observations. In this chapter a series of gas phase experiments are described, with the purpose of addressing those questions raised in Chapter 1 and

to shed new light on the chemistry of AS-PR catalysis.

3.1.1 Gas phase studies of intermediates and transition states of AS-PR catalysis

A series of gas phase calculations were performed in order to investigate the structures, electronic and free energy differences between intermediates and transition states proposed to be part of the biosynthesis of aristolochene by AS-PR. It is understood that AS-PR's role is to force FPP into an appropriate conformation, trigger catalysis then guide the reactive carbocation down a particular pathway leading to aristolochene rather than any of the possible other neutral terpene products. Therefore, a gas phase study will provide data on the inherent reactivity of these intermediates and transitions states. These data can then be compared against experimental data and the results of QM/MM MD free energy simulations (Chapter 4) which attempt to model the reaction within the enzyme system.

The starting point for these simulations was farnesyl cation (**63** in Figure 1.16). As discussed in Section 1.3.2 although there is experimental evidence (mostly gained subsequently to, and partly as a result of, questions inspired by this work) that the cyclisation of FPP to form germacryl cation **26** occurs in a concerted fashion [58,59] it is also reported that germacryl cation is a true intermediate [30,71,74,83] and hence it was of interest to study the gas phase barrier to this cyclisation *via* **TS1** using a number of the computational techniques described in Chapter 2.

Figure 1.16 describes the two intramolecular pathways to be considered: a direct intramolecular proton transfer *via* **TS2** and a water mediated proton transfer *via* **TS2a** and **TS2b**. For both cases investigation of the magnitude of the barriers to the conversion of **26** to **28** *via* these transition states was computed.

That eudesmane cation **28** is a true intermediate of the conversion of FPP to aristolochene by AS-PR has never been questioned in the literature and hence the calculation of this intermediate's energy is of interest. The following steps in the reaction mechanism (formation of intermediates **65** and **66** *via* **TS5** and **TS6** respectively) were also studied computationally in order to provide information on the barrier heights and the structures of the intermediates and transition states.

The final step, loss of H^+ from the *Si* position of H8 [76] was not computed. First, a base is required to abstract this proton and would therefore need to modelled in all previous calculations to be able to compute relative free energy differences. Since this identity of this base is not known, it is difficult to know how to model this reaction step. Computing the abstraction of free H^+ is not satisfactory since this species is of a very high energy and not isolated in aqueous chemical reactions. In addition, DFT treats the energy of H^+ as being exactly 0 (since a species with no electron density cannot have an energy associated with this density).

3.2 Methods

3.2.1 Computational details

Gaussian 03 [162], revisions D.01 and B.05 were used for all electronic energy,E, and free energy, G,[1] calculations. All calculations were performed on the Helix computational and visualisation facility at Cardiff University [2] or at the Minnesota Supercomputing Institute of the University of Minnesota.[3]

Structure modification and visualisation were performed using Molden [163] unless otherwise stated.

A variety of semi-empirical and density functional theory methods were employed, namely AM1 [125], PM3 [126], MPW1PW [164] and MPWB1K. [165] Energy calculations were also made at the MP2 [166]//MPWB1K level. The 6-31+G(d,p) basis set was used exclusively in all DFT and *ab initio* calculations.

The use of the semi-empirical methods (AM1 and PM3) was guided by the requirements of the following stage of this investigation (Chapter 4) in which a combined QM/MM MD simulation was performed. Software limitations and computational resources limit the choice of QM method to AM1 or PM3 and therefore gas-phase calculations at this level were performed to asses their performance against high level QM (MP2)

---

[1]IUPAC recommends the thermodynamic property G is given the title Gibbs Energy or Gibbs Function, but the more common term free energy is used throughout this work.

[2]http://www.helix.cf.ac.uk/

[3]http://www.msi.umn.edu/

and DFT (MPW1PW and MPWB1K) methods.

The choice of DFT functionals was guided by the observation that the MPW derived functionals perform better than, for example, B3LYP [131, 134, 167] in estimating energetics in carbocation cyclisations. MPWB1K is noted for its performance in considering non-bonded interactions [165] and hence was included in this work.

Full MP2 calculations were not possible due to computational demand and software performance problems and therefore single point energy calculations at the MP2//MPWB1K level were performed in order to obtain the improved theoretical rigour of MP2 at an affordable computational cost.

The choice of the 6-31+G(d,p) basis set was made on its previous published application in similar studies and the fact that it allowed DFT and MP2 calculations to made without too great a demand on available hardware resources.

When calculating the free energy of each species the harmonic oscillator approximation was applied. For the lowest energy vibrational modes, the frequencies of all vibrational modes below 100 cm^{-1} were raised to 100 cm^{-1} prior to the free energy calculations as a way to approximately account for anharmonicity.[4]

---

[4]This approach was suggested as a result of private communication with Professors Donald Truhlar and Jiali Gao at the University of Minnesota

3.2.2 Calculation of intermediate and transition state structures

An initial structure for germacryl cation (**26** in Figure 1.16) was constructed using Hyperchem [168] and this initial geometry was optimised using the MM2 force field [169] and AM1. [125] The resulting structure was then modified to produce an approximate geometry for the farnesyl cation (**63** in Figure 1.16) by increasing the C1 C10 distance. The geometry of the approximate models of **26** and **63** were then optimised using the semi-empirical, *ab initio* and DFT methods listed in Section 3.2.1.

The proposed transition state involved in the formation of germacryl cation **26** (**TS1** ins Figure 1.16) was calculated by taking the optimised geometry of **63** and performing a relaxed scan of the potential energy surface (PES), using AM1, with the only fixed coordinate being the C1 C10 internuclear distance which was decreased in 0.1 Å increments. On completion of this scan the approximate coordinates of **TS1** were identified and a full transition state search was performed using AM1. This locates a transition state with a single negative vibrational mode and was then used as a starting point to calculate the geometry and energy of **TS1** using the other computational methods.

The optimised structure of **26** was used as a starting point to identify **TS2**. A reaction coordinate based upon a C6 hydrogen atom and C12 was then defined and a relaxed potential energy surface scan along this coordinate in 0.1 Å increments was performed at the AM1 level. The approximate location of this transition state was identified. AM1 was then used to calculate the transition state and having positively identified it as a transition state (by the presence of a single negative vibrational mode which described a reasonable chemical process) its structure was used as the

starting point for transition state searches using the other computational methods.

In the reaction illustrated in Figure 1.16 there are two proposed species between **TS2** and **28** (protonated germacrene A (**64**) and a transition state (**TS3**) corresponding to the attack of the carbocation on **64** by the C2 C3 π system which leads to the formation of eudesmane cation **28**. Despite numerous repeated attempts it was not possible to produce stable geometries for either **26** or **TS3**. This result immediately hinted at a highly concerted conversion of germacryl cation to eudesmane cation (*vide infra*).

Following identification of a stable eudesmane cation geometry a relaxed PES scan was performed at the AM1 level, where the distance from the axial hydrogen on the C2 position to the carbocation on C3 was reduced by 0.1 Å increments. From the resulting data the approximate location of **TS4** was identified, and further AM1 calculations were performed to isolate this structure. Upon completion of this calculation the AM1 structure of **TS4** was used to calculate geometries of this transition state using the other computational methods.

The PES scan described in the previous paragraph was also used to identify the approximate structure of **65**. This approximate structure was minimised using AM1 and the geometry of this minimum was then used to obtained minimum energy structures using the other techniques described in Section 3.2.1.

Using the optimised structure of carbocation **65** and in a similar manner to the methods described in previous paragraphs a reaction co-ordinate based upon the distance between C14 and C2 was used to perform a relaxed PES scan with the C14

C2 interatomic distance decreasing in 0.1 Å increments using AM1. From the data approximate geometries for **TS5** and **66** were identified and optimised using AM1. From these AM1 geometries further geometry optimisations were performed using the other methods described previously.

3.3 Conversion of farnesyl cation to aristolochene *via* direct intramolecular proton transfer

The free energy profile for the conversion of **63** to **66** was plotted against a dimensionless reaction coordinate (Figure 3.1). Using the computed free energy data it was possible to partition the free energy into enthalpic and entropic components and these were plotted in a similar fashion (Figures 3.2 and 3.3). These data are also displayed in Tables 3.1, 3.2 and 3.3 respectively.

Overall, broad agreement is observed between the *ab initio* and DFT derived results and the semi-empirical data. This overall broad agreement between computationally expensive methods and semi-empirical theories demonstrates that the use of these methods is justified when performing QM/MM based dynamics simulations.
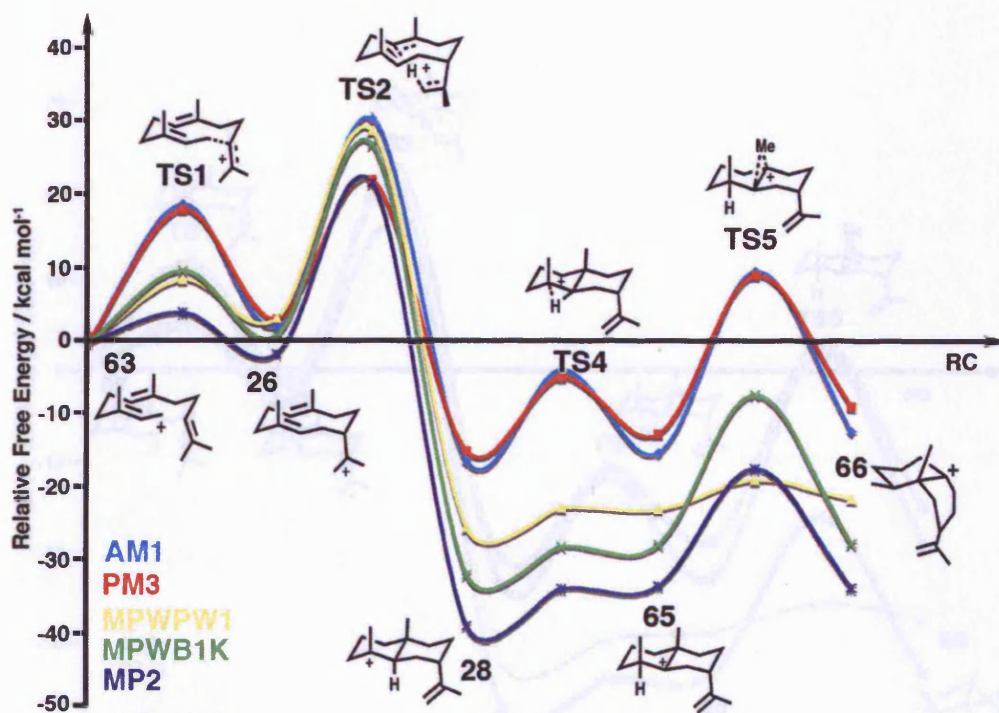
Figure 3.1: *Gas phase free energy profile* - for the conversion of **63** to **66** relative to free energy (ΔG°_{298} of farnesyl cation. MP2 data are single point energies at the MPWB1K//MP2 level

Table 3.1: Gas phase free energy ΔG°_{298} *of intermediates and transition states in Figure 1.16. All values, in kcal mol^{-1}, are relative to the farnesyl cation,* **63**. *MP2 data are ΔE values calculated at the MP2//MPWB1K level*

Method	63	TS1	26	TS2	28	TS4	65	TS5	66
				ΔG°_{298} (kcal / mol^{-1})					
AM1	0	18.9	2.2	30.6	-16.2	-3.6	-15.0	9.8	-12.0
PM3	0	18.2	3.2	22.3	-14.7	-4.7	-12.4	9.3	-8.6
MPWPW1	0	8.6	3.3	29.0	-25.2	-22.5	-22.7	-18.6	-21.2
MPWB1K	0	9.9	0.9	27.0	-31.7	-27.8	-27.5	-7.0	-27.3
				ΔE (kcal / mol^{-1})					
MP2//MPWB1K	0	4.2	-1.5	21.8	-38.6	-33.5	-33.2	-17.1	-33.3
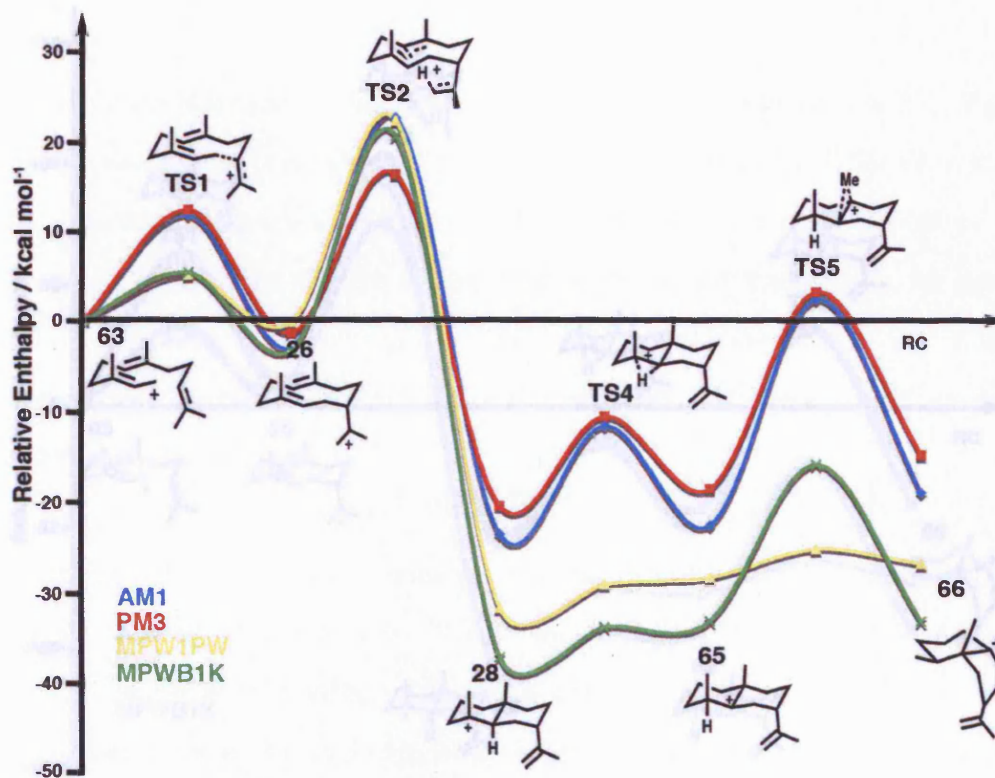
Figure 3.2: *Gas phase enthalpy profile* - ΔH°_{298} for the conversion of **63** to **66** relative to enthalpy of farnesyl cation.

Table 3.2: Gas phase enthalpy ΔH°_{298} *of intermediates and transition states in Figure 1.16. All values, in kcal mol^{-1} are relative to the farnesyl cation, 63.*

Method	63	TS1	26	TS2	28	TS4	65	TS5	66
AM1	0	11.6	-3.1	22.4	-23.7	-11.7	-22.7	2.3	-19.0
PM3	0	12.1	-15	16.1	-20.6	-10.7	-18.6	3.1	-14.9
MP1PW1	0	5.0	-0.2	22.1	-31.9	-29.2	-28.5	-25.3	-27.0
MPWB1K	0	5.2	-3.4	20.3	-37.3	-34.0	-33.3	-16.0	-33.3
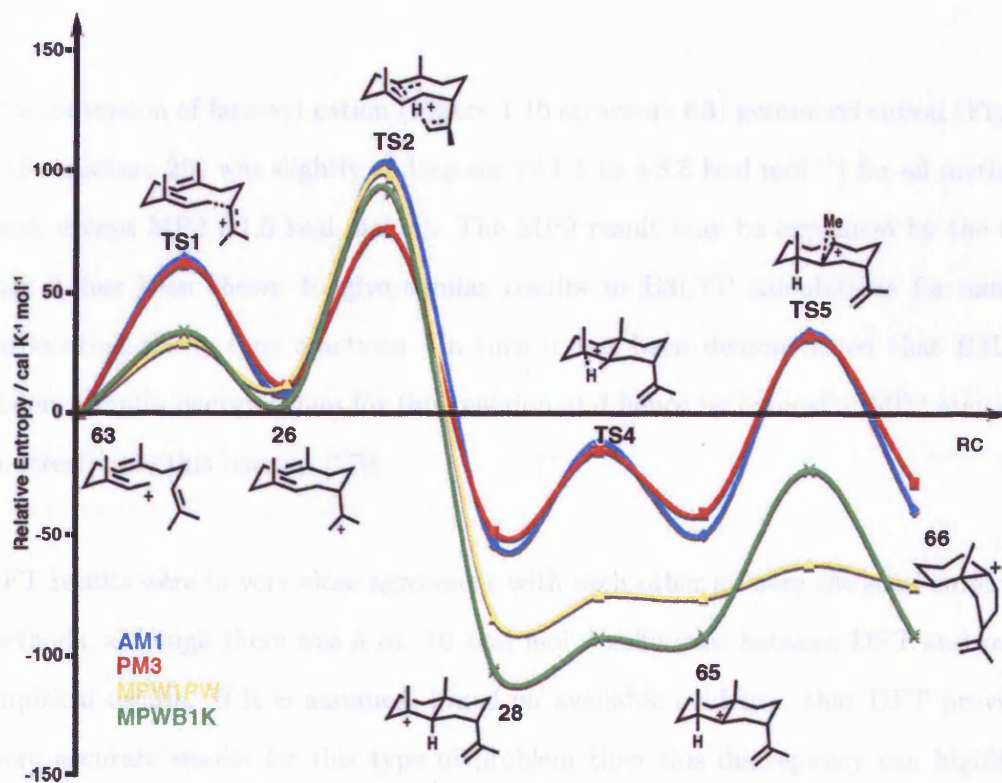
Figure 3.3: *Gas Phase entropy Profile* - $\Delta S^{\circ}{}_{298}$ for the conversion of **63** to **66** relative to entropy of farnesyl cation.

Table 3.3: Gas phase entropy $\Delta S^{\circ}{}_{298}$ *of intermediates and transition states in Figure 1.16. All values, in cal mol^{-1} K^{-1} are relative to the farnesyl cation, **63**.*

Method	63	TS1	26	TS2	28	TS4	65	TS5	66
AM1	0	63.5	7.3	102.6	-54.2	-12.2	-50.4	32.9	-40.2
PM3	0	61.1	10.9	74.8	-49.4	-15.9	-41.7	31.2	-28.9
MP1PW1	0	29.0	11.2	97.2	-84.5	-75.4	-75.6	-62.5	-71.2
MPWB1K	0	33.1	3.11	90.8	-106.2	-93.3	-92.2	-23.6	-91.7

3.3.1 Conversion of farnesyl cation to germacryl cation

The conversion of farnesyl cation (Figure 1.16 structure **63**) germacryl cation (Figure 1.16 structure **26**) was slightly endergonic ($+1.1$ to $+3.8$ kcal mol^{-1}) for all methods used, except MP2 (-1.5 kcal mol^{-1}). The MP2 result may be explained by the fact that it has been shown to give similar results to B3LYP calculations for similar carbocation-alkene type reactions. In turn it has been demonstrated that B3LYP underestimates energy values for this reaction and hence by extension MP2 may also underestimate this barrier. [170]

DFT results were in very close agreement with each other as were the semi-empirical methods, although there was a *ca.* 10 kcal mol^{-1} difference between DFT and semi-empirical results. If it is assumed, based on available evidence, that DFT provides more accurate results for this type of problem then this discrepancy can highlight an inherent weakness in the semi-empirical approach to QM problems and emphasise the need for caution when interpreting the results of QM calculations.

Table 3.4 shows how the C1 C10 interatomic distance varied between farnesyl cation and **26**. The π electron system between C10 and C11 is presumably attracted to the positive charge C1 and this provides an electronic driving force for the ring closure. Overall the geometries of **63**, **TS1** and **26** are similar irrespective of the method used. The C1 C10 bond distance in **26** was 1.5 Å and this is entirely consistent with the formation of a carbon carbon σ bond between C1 and C10 . The length of this bond in **TS1** was approximately 2.0 Å for all methods except MPWB1K. That the MPWB1K structure of **TS1** was more 'reactant like' (*i.e.* the C1 C10 distance was closer to to **63** than **26**) and also had the lowest ΔG^{o}_{298} is consistent with the
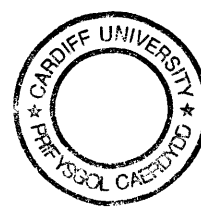
Table 3.4: *C1 C10 interatomic distance in **63**, **26** and **TS2** - all values in Å.*

Method	63	TS1	26
AM1	3.9	2	1.5
PM3	4.1	1.9	1.5
mPW1PW	4.0	2.0	1.5
MPWB1K	3.6	2.7	1.5

Hammond postulate [171] which *inter alia* states that if two species possess closely matching energy then the inter-conversion between these two states will involve only a small amount of molecular reorganisation.

3.3.2 Conversion of germacryl cation to eudesmane cation *via* TS2

The key step in this proposed mechanism is the intramolecular proton transfer from C12 to the C6 C7 double bond, avoiding germacrene A as a neutral intermediate. The free energy difference between **26** and **TS2** was between 19.1 (PM3) and 28.0 (AM1) kcal mol^{-1} and the conversion of **26** to **28** was exergonic by approximately 15 kcal mol^{-1} for all methods.

Figure 1.16 shows an intermediate **64** and transition state **TS3** on the pathway to eudesmane cation **28**. Despite numerous and repeated attempts neither of these species could be identified using the methods employed in this gas-phase study. The conclusion is that the conversion of TS2 to **28** is a concerted process occurring rapidly

due to the attractive force between the C2 C3 π electron system and the positive charge developed on C7 which drives the formation of a thermodynamically favourable ten-membered ring. Alternatively, this may reflect a weakness in the methodology and/or software used for these calculations and it is possible, but not certain, that higher levels of theory coupled with more accurate basis sets could have identified structures for **64** and **TS3**. However, given the experimental evidence recently obtained [92, 172, 173] and discussed in Section 1.4.3 it is reasonable to conclude that the calculations presented here represent the first computational evidence that the formation of eudesmane cation by AS-PR is indeed highly concerted.

In the Marcus theory of proton transfer [174] the free energy of activation is partitioned into two parts: one part associated with the proton transfer's bond making-bond breaking process and another part comprising the changes in molecular geometry necessary for proton transfer to take place. In order for a proton transfer to occur there must not only be a reasonable overall barrier but the free energy of rearrangement must also be a small component of the overall free energy of activation.

The isopropyl group in **26** must rotate so that a methyl proton of C12 is in close proximity to the C6 C7 double bond before the transition state can be reached. To evaluate the energetic cost of this rotation the potential energy as a function of the C9 C10 C11 C12 dihedral angle, β, in Figure 3.5, was calculated. The energy difference between **3***, at which the value of β is the same as that in **TS2**, was between 0.02 (MPWB1K) and 1.9 (AM1) kcal mol^{-1} higher than the energy of **26**, demonstrating that the rotation about the C10 C11 single bond does not contribute significantly to the overall energetic barrier between **26** and **28**.
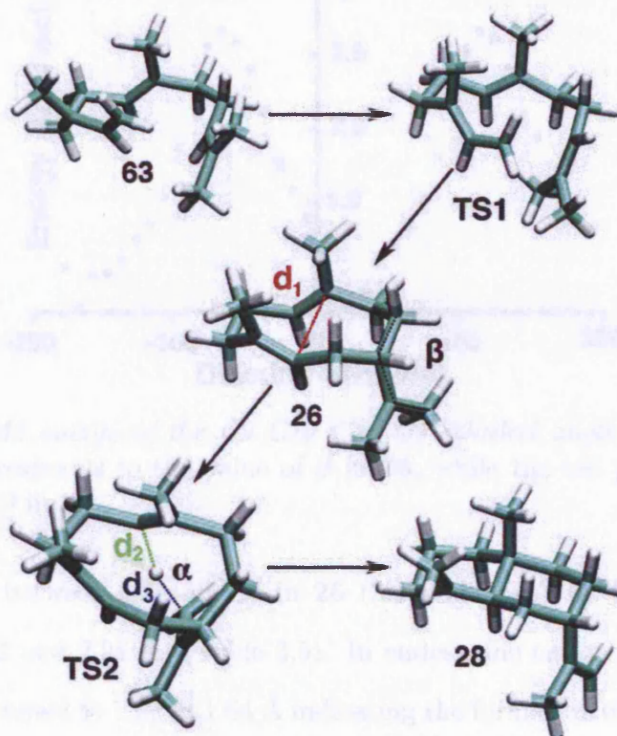
Figure 3.4: *Structures of intermediate and transition states in conversion of farnesyl cation to eudesmane cation via **TS2** - Calculated using MPWB1K. The distances d_1, d_2 and d_3 and the angles α and β are included in Table 3.5*

Table 3.5: *Distances d_1, d_2 and d_3 and angles α and β for the conversion of **26** to **28**, the dihedral angle β between C9, C10, C11and C12 in **26** and in **3*** (**26** with β as calculated for **TS2**) (as defined in Figure 3.4), and the energy barrier for the conversion of **26** to **3***

Method	d_1 (Å)			d_2 (Å)	d_3 (Å)	α (deg)	β (deg)		ΔE (kcal mol^{-1})
	26	**TS2**	**28**	**TS2**			**26**	**3***	**26 → 3***
AM1	2.92	2.99	1.55	1.45	1.45	151	23.0	-25.2	1.7
PM3	3.09	3.03	1.55	1.45	1.45	151	15.9	-29.9	1.5
mPW1PW	3.1	3.02	1.66	1.46	1.44	161	18.4	-19.9	1.0
MPWB1K	3.00	2.98	1.64	1.47	1.46	159	15.9	-18.3	0.02
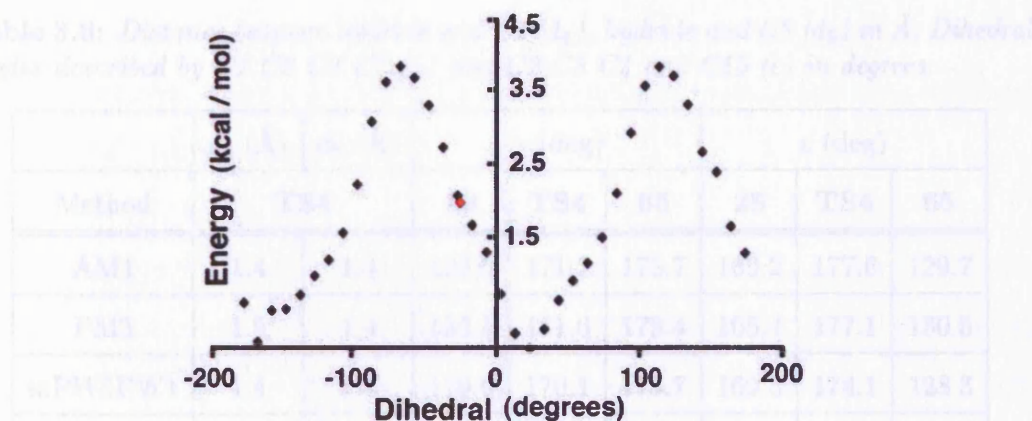
Figure 3.5: *AM1 energy of the C9 C10 C11 C2 dihedral angle, β - in which E= 0 kcal mol^{-1} corresponds to the value of β in* **26**, *while the red point indicates the energy of* **26** *for β in* **3***

The distance d_1 between C2 and C7 in **26** (3.1 and 2.9 Å by DFT) and in **TS2** were similar (3.02 and 2.98 Å) (Table 3.5). In eudesmane cation (**28**) however, this distance was shortened to 1.66 or 1.64 Å indicating the formation of a σ bond between C2 and C7.

The migrating proton was approximately equidistant between C12 and C6 in **TS2** at all levels of theory (Table 3.5) and the angle α (Figure 3.4) between C6, the transferring proton and C12 was between 151and 161 degrees suggesting the stereochemistry of the transfer does not require a large distortion from linearity (Figure 3.4).

3.3.3   Conversion of 28 to 65 *via* TS4

Following formation of eudesmane cation the next step in the reaction mechanism is a 1,2 hydride shift from the C2 position to the carbocation on C3 *via* transition state **TS4**.

Table 3.6: *Distance between hydride and C2 (d_4), hydride and C3 (d_5) in Å. Dihedral
angles described by C1 C2 C3 C7 (γ) and C2 C3 C4 and C15 (ϵ) in degrees.*

	d_4 (Å)	d_5 (Å)	γ (deg)			ϵ (deg)		
Method	**TS4**		**28**	**TS4**	**65**	**28**	**TS4**	**65**
AM1	1.4	1.4	129.0	171.1	178.7	169.2	177.6	129.7
PM3	1.5	1.4	133.1	171.6	179.4	165.4	177.1	130.5
mPW1PW1	1.4	1.4	119.6	170.1	178.7	169.3	174.1	128.3
MPWB1K	1.3	1.4	117.6	170.1	180.0	177.9	173.1	137.1

The free energy differences between **28** and **65** are very similar (Figure 3.1 and Table
3.1). The process was slightly endergonic in all cases and the relative free energy
change was between +4.2 kcal mol^{-1} (MPWB1K) and +1.2 kcal mol^{-1} (AM1). Small
differences in relative free energy were expected since both **28** and **65** are both tertiary
carbocations and it is difficult to suggest *a priori* which of the two carbocationic
species is energetically more favourable. The magnitude of the barrier between +12.6
kcal mol^{-1} (AM1) and +0.2 kcal mol^{-1} (MPW1B1K, Table 3.1), and the DFT barriers
were *ca.* 10 kcal mol^{-1} lower than the semi-empirical barrier - a feature which may
be explained in the same manner as was given in Section 3.3.1.

At the transition state **TS4** the value of d_5, the distance between the transferring
hydride and C2 and d_6, the distance between the hydride and C3, (shown in Figure
3.3.3) were equal or approximately equal irrespective of method (Table 3.6). This
demonstrates that the transition state occurs when the hydride is approximately
equidistant between donor and acceptor carbons. This finding can be again be
rationalised using Hammond's postulate: the difference in free energy between reactant
and transition state was small, as was the free energy difference between transition

state and product, hence one expects the degree of molecular reorganisation between the two to be minimal, which leads to the conclusion that the transition state structure obtained for **TS4** is reasonable.

The eudesmane cation **28** has an approximately tetrahedral geometry at the C2 position, as described by the dihedral angle γ (C1 C2 C3 C7). After undergoing the 1,2 hydride shift required to form **65**, the geometry of C2 becomes planar. At C3 the opposite occurs: the initially planar C3 carbocation, described by dihedral ϵ (C2 C3 C4 C15) becomes tetrahedral as the hydride migrates from C2 to C3. The data in Table 3.6 demonstrate how these angles change, reflecting the change in charge and hence geometry at hydride donor and acceptor carbons.
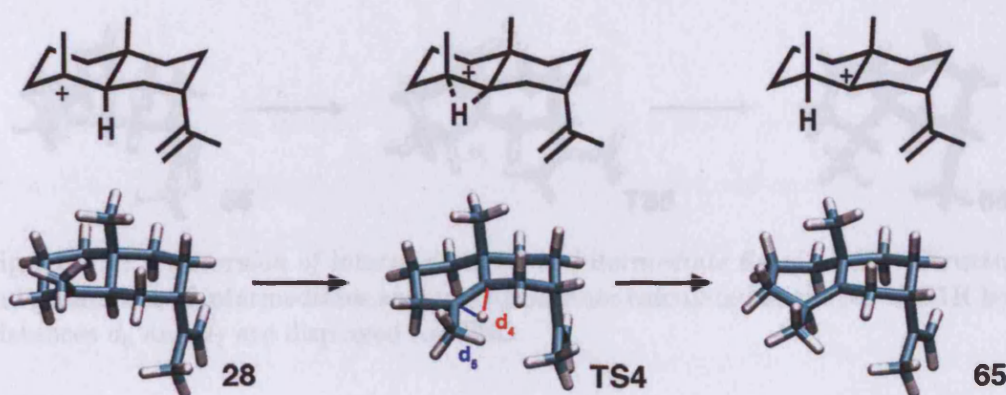
Figure 3.6: *Conversion of eudesmane cation **28** to intermediate **65** via **TS4** -* structures and geometries of intermediates and transition state calculated at the MPWB1K level. Distances d_4 and d_5 are displayed for **TS4**.

3.3.4 Conversion of intermediate 65 intermediate 66 *via* TS5

The conversion of **65** to **66** was slightly endergonic (+0.2 (MPWB1K) to +3.8 (PM3) kcal mol^{-1}) for all calculated free energies (Table 3.1). One explanation for this

apparent decrease in energetic stability is slight steric hindrance between the methyl groups on the C2 and C3 positions in **66** which is not present in **65**.

The free energy barrier to reach **TS5** was approximately 20 to 25 kcal mol^{-1} for all methods, except mPW1PW in which the barrier was significantly smaller (4.1 kcal mol^{-1}) but visual inspection of the geometry for the transition state reveals little difference between the structures, so it is difficult to speculate on the cause of this discrepancy. Given the general agreement in geometry and the observation that the MP2//MPW1B1K ΔE value of the same barrier was 16 kcal mol^{-1} this suggests that the MPW1PW results is less reliable in this case than the other methods, which are all in broad agreement.
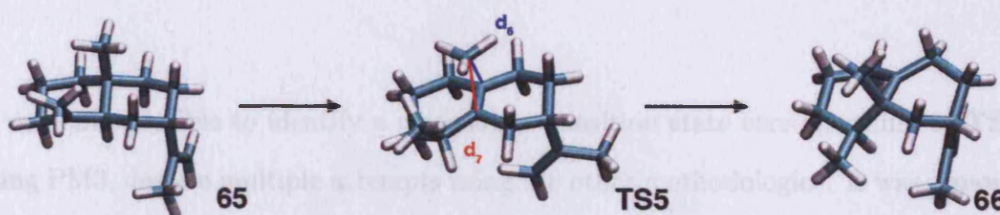
Figure 3.7: *Conversion of intermediate **65** to Intermediate **66** via **TS5*** - Structures and geometries of intermediates and transition state calculated at the MPWB1K level. Distances d_6 and d_7 are displayed for **TS5**.

Figure 3.7 shows the geometries of intermediates **65** and **66** and the transition state **TS5** between them. The transition state corresponds to the methyl group on C14 migrating between C7 and C2, consistent with this step being a 1,2 methyl shift (Figure 1.16). The distance, d_6 was defined as the distance between C14 and C7 and another distance, d_7 was defined as the distance between C14 and C2. At the transition state the distances, d_6 and d_7 were equal (Table 3.7), whereas at **65** d_6 was approximately 1.5 Å corresponding to a typical carbon-carbon single bond, and at **66** d_7 was again close to 1.5 Å a typical length for a carbon carbon bond.

Table 3.7: *Distance between migrating methyl group and C7 (d_6) and migrating methyl group and C2 (d_7) in Å.*

	d_6 (Å)	s_7 (Å)	d_6 (Å)	s_7 (Å)	d_6 (Å)	s_7 (Å)
Method	**65**		**TS5**		**66**	
AM1	1.5	2.5	2.1	2.1	2.4	1.5
PM3	1.5	2.4	2.1	2.1	2.4	1.5
MPW1PW	1.6	2.4	2.1	2.1	2.3	1.6
MPWB1K	1.6	2.3	2.1	2.1	2.3	1.6

3.4 Conversion of farnesyl cation to aristolochene *via* water mediated proton transfer

It was only possible to identify a reasonable transition state corresponding to **TS2a** using PM3, despite multiple attempts using the other methodologies. It was, however possible to identify **TS2b** using all of the methods. Single point MP2//MPWB1K calculations were not possible as the software repeatedly ran out of memory when performing the SCF calculation. The structure of **TS2a** and the MPWB1K structure of **TS2b** are shown as Figures 3.4 and 3.4.

The structure of transition state **TS2a** (Figure 3.4) clearly shows the abstraction of a proton from the C12 position and the protonation of a water molecule. It is clear from this diagram too that there is a proton on the nascent hydroxonium ion that is well positioned to protonate the C6-C7 double bond (*cf*. **TS2** in Figure 3.4).

Following formation of **TS2a** protonation of the C6 C7 double bond can occur as

shown in the structure of **TS2b**. In **TS2b** a proton from the water molecule transfers onto C6 breaking the double bond and forming **64** (or proceeding directly to **28** in a concerted reaction mechanism).
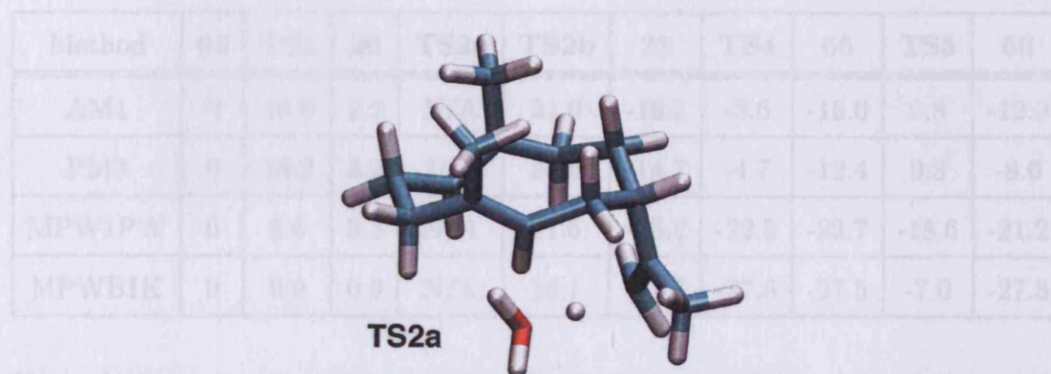
Figure 3.8: *Calculated structure of transition state TS2a* - Geometry of **TS2a** calculated using PM3
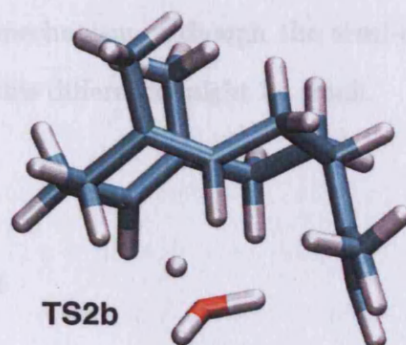
Figure 3.9: *Calculated structure of transition state TS2b* - Geometry of **TS2a** calculated using MPWB1K

The free energy data obtained are shown in Table 3.8. In the case of AM1 and PM3 the barrier to protonation of the C6 C7 double bond *i.e. via* **TS2b** (31.0 and 24.5 kcal mol^{-1}) is almost equal to the direct **TS2** intramolecular proton transfer case (30.6 and 22.3 kcal mol^{-1}). However, the DFT data show that protonation of C6 C7 *via* **TS2b** is lower in energy than *via* **TS2** by 7.4 (MPW1PW) and 10.9 (MPWB1K) kcal mol^{-1}.

Table 3.8: Gas phase free energy for water mediated proton transfer - $\Delta G^\circ{}_{298}$ *of intermediates and transition states in Figure 1.16. All values, in kcal mol^{-1}, are relative to the farnesyl cation,* **63**. *MP2 data are ΔE values calculated at the MP2//MPWB1K level*

Method	63	TS1	26	TS2a	TS2b	28	TS4	65	TS5	66
AM1	0	18.9	2.2	N/A	31.0	-16.2	-3.6	-15.0	9.8	-12.0
PM3	0	18.2	3.2	19.8	24.5	-14.7	-4.7	-12.4	9.3	-8.6
MPW1PW	0	8.6	3.3	N/A	21.6	-25.2	-22.5	-22.7	-18.6	-21.2
MPWB1K	0	9.9	0.9	N/A	16.1	-31.7	-27.8	-27.5	-7.0	-27.3

These data suggest that it is possible that proton transfer, mediated by a water molecule, offers a lower energy pathway to the formation of eudesmane cation, **28**, than the direct transfer mechanism, although the semi-empirical (AM1 and PM3) QM results suggest that this difference might be small.

3.5 Discussion

There exists little direct experimental thermodynamic information on aristolochene biosynthesis and the gas phase results outlined in the preceding paragraphs represent the first attempt to obtain detailed free energy information for the reaction mechanism. Given the reaction proceeds largely by way of a series of carbocation rearrangements, which are generally extremely rapid there is little prospect of experimentally observing many of the proposed intermediates directly (although recently it has been possible to use bromo-substituted analogues of IPP to trap carbocationic intermediates in higher terpenoid synthases [175]) and thus computational techniques offer the best chance of

obtaining insights into the energetic barriers involved in catalysis by AS-PR as well as structural details of the proposed intermediates and transition state structures.

3.5.1 Direct or water mediated proton transfer and the intermediacy of germacrene A

The evidence obtained from these gas phase calculations suggests that intramolecular proton transfer, either direct, or mediated by a water molecule is a feasible method for the protonation of the C6 C7 double bond, a necessary and key step in the reaction catalysed by AS-PR. The lowest free energy barriers calculated for direct intramolecular transfer were 22.3 kcal mol^{-1} (Table 3.1) and 16.1 kcal mol^{-1} for the water mediated proton transfer (Table 3.8). In the case of water mediated proton transfer it is also possible that a dual transition state exists; *i.e.* one in which proton abstraction at C12 occurs simultaneously with protonation of C6 C7. This transition state was not identified during calculations, despite several attempts, but its existence cannot be excluded purely on the basis that it was not observed in the series of QM calculations described in this work.

In any case, the C6 C7 protonation schemes outlined here share a common feature - germacrene A is not a true intermediate, and the positive charge generated by the loss of diphosphate and cyclisation to form germacryl cation (**26**) is not quenched by the formation of germacrene A. These intramolecular processes maintain overall positive charge throughout the reaction scheme and therefore might be seen to be more efficient than a reaction involving the formation, quenching and regeneration of

a carbocation.

The fact that germacrene A has been observed in the product mixture of AS-PR [73,83,84,176] can be explained if, rather than being a true intermediate, germacrene A is a side product resulting from the premature deprotonation of C12, without any subsequent proton transfer onto C6. This proposal is consistent with the observation that germacrene A is not a substrate of AS-PR[5].

In addition such a proposal is consistent with all previous mutagenesis experiments. Previous results which were interpreted to conclude that Tyr 92 might act as an active site acid [72] might now be reinterpreted by suggesting that the aromatic side portion of Tyr 92 acts to stabilise the nascent positive charge which builds up on C7 following intramolecular proton transfer. This is consistent with the observation that the Y92F mutant produces aristolochene as its primary product [75, 176], since Phe is also aromatic and thus similarly able to stabile positive charge. This hypothesis also explains why the Y92F mutant is still active (Table 1.2), whereas if Tyr 92 acts as an acid, protonating the C6 C7 bond then the Y92F mutant (which contains no phenolic proton) should not turn over. Thus, based on mutagenesis and kinetic studies intramolecular proton transfer is a better hypothesis than Y92F acting as a proton source.

The intramolecular proton transfer hypothesis is also consistent with the proposals outlined in Section 1.3.2 in which the steric bulk of Tyr 92 (or Phe 92) is required to ensure that substrate is folded into the correct geometry for cyclisation; incorrect folding results in the formation of linear farnesene compounds. However, partly in

---

[5]Allemann group, Cardiff University. Unpublished data

response to publication of the results described in this chapter [92, 172] the use of of fluoro-substitued analogues to investigate the reaction mechanism of AS-PR has provided indirect evidence that germacrene A is a true, on-path intermediate in the synthesis of aristolochene by AS-PR. [58, 59, 173]

3.5.2 Concerted reaction mechanisms

In the case of water mediated proton transfer, a concerted process is possible, in which proton abstraction on C12 occurs simultaneously with protonation of the C6 C7 double bond, the transition state structure containing both protonation and deprotonation components. As mentioned in Section 3.4 this dual transition state was not identified during calculations, but it is still possible that water mediated proton transfer might occur in such a highly concerted fashion.

It was not possible to identify a stable geometry for **64** or **TS3**. Following the protonation of C6 C7 (irrespective of method) it appears that the formation of bicyclic eudesmane cation is strongly favoured, and the driving force is the breaking of the C2 C3 π bond and the subsequent formation of a σ bond between C2 and C7. Tables 3.1 and 3.8 and Figure 3.1 show that formation of **28** is strongly exergonic, and thus on the basis of these calculations formation of eudesmane cation is likely to be a concerted process, with fusing of the ring system acting as the thermodynamic driving force.

Following formation of eudesmane cation, there is a smaller (relative to C6 C7 double bond protonation) barrier associated with the 1,2 hydride shit (Section 3.3.3). The

DFT (and MP2) results suggest that this process may occur in a concerted process: formation of eudesmane cation is followed rapidly by the migration of a hydride to quench the positive charge developed on C3. This, of course, generates another carbocation on the C2 position, which is then quenched by the 1,2 methyl shift *via* **TS5**.

The MPWPW1 results showed a relatively low barrier of 4.1 kcal mol^{-1}, although the other methods had calculated barrier heights in the region of 20 kcal mol^{-1}. On balance it appears that there is a appreciable barrier from **28** to **65** and this process is not concerted to a significant degree.

Despite this, it is possible to view the entire conversion of **28** to aristolochene as occurring *via* one concerted process, driven thermodynamically by the final step (deprotonation of C8 on **66** to form aristolochene). After protonation of the C6 C7 double bond, ring closure drives both the hydride shift step and the methyl shift step leading to deprotonation to form aristolochene. Evidence that these steps occur rapidly with low thermodynamic barriers and precise control from the active site comes from the study of the product mixture of wild type AS-PR (Section 1.3.2). The presence of selinenes (Figure 1.10) would suggest that it is possible for carbocation quenching to occur after formation of eudesmane cation, but before the hydride shift, methyl shift, deprotonation steps required for to produce aristolochene occur. Selinenes have never been reported in the product mixture of WT-AS and therefore the post-eudesmane cation steps in aristolochene biosynthesis must occur rapidly and with high precision to avoid off-pathway selinene side products.

3.5.3 Comparison with previous mutagenesis experiments

The results of previous mutagenesis experiments have been interpreted based upon a reaction mechanism which involves protonation of C6 C7 double bond by an amino acid side chain acting as the proton source with a series of discrete intermediates existing from eudesmane cation to aristolochene. For any of the hypotheses included in this chapter to be valid, they must be consistent with the results of previously reported mutagenesis experiments.

As discussed in Section 3.5.1 the role of Tyr 92 is better explained by suggesting that its aromaticity is employed to stabilise a positive charge developing on C6, rather than it acting as a proton source as had been previously suggested. The Y92F mutant is still an active enzyme; the reduced catalytic efficiency is explained by the fact that replacement of Tyr with Phe induces structural changes within the active site which affect the rate of catalysis. In the case of the Y92A, Y92V and Y92C series of mutants, the lack of any aromatic group at the appropriate position prevents the formation of eudesmane cation and hence the driving force for formation of aristolochene. Therefore premature deprotonation occurs, resulting in the formation of farnesene side products.

Intramolecular proton transfer based on the evidence of the calculations outlined above and the lack of strong evidence in favour of any of the proposed proton sources suggested by previous studies, is a more attractive hypothesis than one which involves an amino acid side-chain (Tyr 92 or otherwise) acting as an acid.

The formation of germacrene A by the W334V and W334L mutants does not necessarily

require that germacrene A be a true intermediate in the synthesis of aristolochene by AS-PR. Instead, the lack of an aromatic side-chain in Val and Leu may prevent the stabilisation of positive charge on C3 or C7, therefore preventing steps following formation eudesmane cation, but allowing formation of eudesmane cation as an intermediate. In this situation germacrene A is formed not as intermediate, but rather as an off-pathway product caused by premature deprotonation at the C12 position.

The presence or absence of an aromatic side-chain at position 178 is also key in determining the major product of mutant AS enzymes. The F178V, F178I and F178C mutants all produce germacrene A as their major product - however this may also be explained by germacrene A being an off pathway aborted mechanism product, rather than as a true intermediate. On this basis it appears that Phe 178 stabilises positive charge before formation of eudesmane cation. The requirement for an aromatic group at position 178 is demonstrated by the observation that F178Y and F178W mutants both produce aristolochene as their major product.

3.5.4 Stereochemical consequences of intramolecular proton transfer

In principle a reaction mechanism which involves intramolecular proton transfer as described in this chapter can involve transfer of a proton from either the C12 or C13 position. Work by the Cane group has demonstrated that deprotonation occurs exclusively at the C12 position. [76] Gas phase results presented above (Section 3.3.2) demonstrate the low intrinsic barrier to rotation of the ispropylidene cation containing

C12 and C13. Therefore it is clear that the active site of AS-PR is constructed in such a way as to either increase the barrier to rotation or to favour deprotonation at a unique position by some other mechanism. A unique position of deprotonation does not rule out an intramolecular proton transfer mechanism, but rather reinforces the evidence that AS-PR exhibits precise control over the course of the reaction in order to produce aristolochene in high yield.

Stereochemical evidence designed to test this hypothesis has since been obtained which demonstrates that intramolecular transfer does not, in the form described in this work, occur in AS-PR. In an elegant series of experiments [12,12,12,13,13,13-^{2}H$_6$]-FPP was incubated with AS-PR in the both H$_2$O and D$_2$O. [92] If the source of the proton which incorporates itself into the C6 C7 double bond is C12 or C13 then one would expect the resultant aristolochene to contain a singe deuteron, whether the experiment is performed in H$_2$O or D$_2$O. In the presence of D$_2$O deuterium is incorporated, but this sheds little light on the reaction since it must be assumed that all the protons in the solvent and enzyme have fully exchanged with ^{2}H. However, when the experiment is performed in H$_2$O no ^{2}H is observed (by ^{2}H NMR) in the resultant aristolochene. This provides compelling evidence that direct intramolecular proton transfer *via* **TS2** does not occur.

In addition detailed NMR experiments in the same set of experiments revealed that when [12,12,12,13,13,13-^{2}H$_6$]-FPP is incubated with AS-PR in the presence of D$_2$O the deuteron that is incorporated does so above the plane of the ring (*i.e.* the opposite face from that which contains the isropropylidene group). This result rules out a water molecule, on its own, binding tightly beneath the plane of the ring and acting as the acid; the pathway *via* **TS2a** and **TS2b**. In addition it also rules out diphosphate

acting as the general acid (a previous proposal for AS-AT [43]) since it too is bound underneath the ten-membered ring. Figure 3.10 provides details of the key results from these experiments.
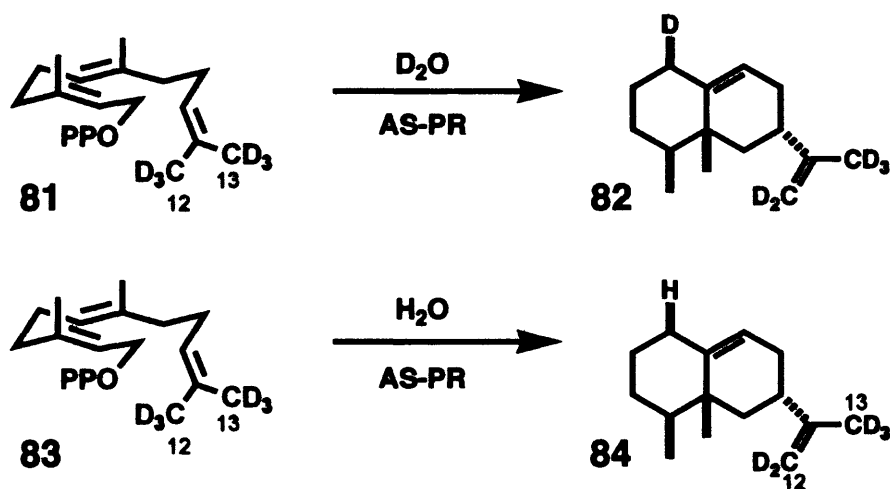
Figure 3.10: *Stereochemistry of C6 protonation in AS-PR* - ^{2}H labelling of FPP and incubation in D_2O and H_2O and subsequent NMR reveals that the proton source is neither C12 nor a bound water molecule beneath plane of ring. [92]

Thus, as a result of this labelling study, intramolecular proton transfer, as described in Figure 1.16 does not occur in AS-PR. This is despite the fact that such mechanisms have precedent [32,38,93,95] and either would be consistent with much of the previously reported work on AS-PR.

The identity of this active site acid remains unknown, and it is now possible to add intramolecular proton transfer to the list of candidates which have been tested but found not to agree with experimental evidence. The recent proposal that Lys 206 might perform this role [92] is worth examining. It has been suggested that the (protonated) NH_3^+ group Lys 206 can act as the acid, with a network of hydrogen bonds between Lys 206 Asp 203 and Arg 200 having previously been identified [30] and it was suggested that a proton shuttle mechanism might exist in which solvent

protonates Arg 200 which is then able to protonate Asp 203 which can transfer charge onto Lys 206 and allow for protonation of the C6 C7 double bond. Some evidence for this comes from data[6] in which the K206Q protein exhibited a threefold drop in catalytic activity and the K206R variant exhibit a 4 fold reduction in activity, *i.e.* that as the ability to accept a proton decreased so does catalytic activity.

3.6 Conclusions

In this chapter the first computational study of the intermediates and transition states involved in the conversion of farnesyl diphosphate to aristolochene was carried out. Two novel proposals concerning the nature of the acid involved in the protonation of the C6 C7 double bond were tested using high level *ab initio* and semi-empirical molecular orbital theory, as well as DFT, to demonstrate that these mechanisms are both plausible, and consistent with previously published mutagenesis studies. Following protonation of C6 C7 evidence was presented for a highly concerted process to form eudesmane cation and the subsequent intermediates required to reach the end of the reaction. These computational results demonstrating a concerted reaction are also consistent with previously reported experimental studies.

The limitation of a gas phase study of a biological system is, of course, that no consideration of the solvent and enzymatic environment is given. The obvious next stage in this work is to model the key steps in the reaction mechanism in an enzyme based model. As discussed in previous chapters the combined QM/MM molecular

---

[6]Forcat S., Taylor S., and Allemann R., unpublished data

dynamics simulation offers a method for modelling the reaction mechanism within the enzyme system. In performing a QM/MM dynamics simulation, at present only the semi-empirical QM methods are practical, as other higher level theories are too computationally demanding. The overall broad agreement between the DFT, MP2 and semi-empirical results obtained in this gas phase study suggest that either PM3 or AM1 would be suitable methods for use in a QM/MM investigation.

Chapter 4

Docking and QM/MM Studies of Aristolochene Synthase

4.1 Introduction

The aim of the set of experiments described in this chapter is to investigate further the reaction mechanism catalysed by AS-PR, building upon the gas phase data obtained in Chapter 3. In particular the aim is to discover why AS-PR seemingly favours the formation of germacrene A as a true intermediate and utilises a proton source on the opposite face of the ten membered ring from the isopropylidene group, rather than one of the intramolecular proton transfer mechanisms (direct or *via* a water molecule), which have precedent in the other terpene synthases and are are energetically feasible (Chapter 3).

In addition both experiment and theory suggest strongly that eudesmane cation (**28** in Figure 1.16) is formed in a concerted manner and computational confirmation of this in the condensed phase would further demonstrate that **64** and **26** are not intermediates in the reaction catalysed by AS-PR.

Purely gas phase studies do not allow for an exploration of the role of magnesium ions in AS-PR catalysis. All sesquiterpene synthases are presumed to require three magnesium ions for catalysis and recent work has described a possible sequence by which magnesium ions in AS-PR bind to trigger catalysis. In addition, by analogy with AS-AT, AS-PR is proposed to enter a closed conformation in which catalysis takes place. By construction of three different models of AS-PR (*vide infra*) and performing a series of QM/MM simulations, the role of magnesium in AS-PR catalysis may be explored.

4.2 Methods

The crystal structure of AS-PR (PDB entry 1DGP) [30] was used as the starting point for the generation for all docked structures. This structure contains a farnesol molecule which was deleted from the pdb file to create a structure with a vacant active site. The MPWB1K gas phase structure of **TS2** was kept rigid and used in the docking calculations. Insight II (Biosym Technologies, San Diego, California, USA) was used for visualisation and structural adjustments. The structure of **TS2** was an arbitrary choice and **26** may have been an equally appropriate choice. CHARMM [101] was used for energy minimisation calculations. A combined QM/MM potential

was used. [142, 177] The ligand was treated using PM3 [125] and the CHARMM22 MM force field [98] was used for the enzyme, pyrophosphate anion and magnesium cations. PM3 was chosen since it is the semi empirical method whose barrier height for the key step (formation of **TS2**) most closely matched the DFT results (see, for example, Figure 3.1).

The CHARMM implementation of the TIP3 model [178] was used for water molecules. Following docking, the **TS2** ligand was replaced by germacryl cation **26** and the docking procedure repeated. All docking was performed manually, using chemical intuition. All amino acids were assigned an initial protonation state associated with pH 7. The QM region had a charge of +1 in all cases. The HBUILD module of CHARMM [179] was used to add hydrogens to protein atoms unless otherwise specified.

4.2.1 Preparation of holoenzyme model A

All solved crystal structures of AS-PR show the enzyme in an open and unproductive form, compared to the closed pyrophosphate bound structure of AS-AT recently solved (PDB entry 2OA6, subunit D). [43] It is only this closed conformation, which contains three magnesium ions and a bound pyrophosphate anion that is assumed to be the catalytic form.

The initial results from molecular docking followed closely the results obtained by mutagenesis studies that identified the key residues involved in Mg^{2+} ion binding. [75] Two cations were docked into the structure to quench the charge on the pyrophosphate.

Figure 4.1 shows the active site of AS-PR in this active site. The first Mg^{2+} ion is coordinated to the pyrophosphate, Ser 248 and Glu 252, as well as two water molecules. Asn 244, one of the three residues in the proposed magnesium binding triad forms a hydrogen bonding network by donating a hydrogen bond to the pyrophosphate and accepting one from Ser 248.
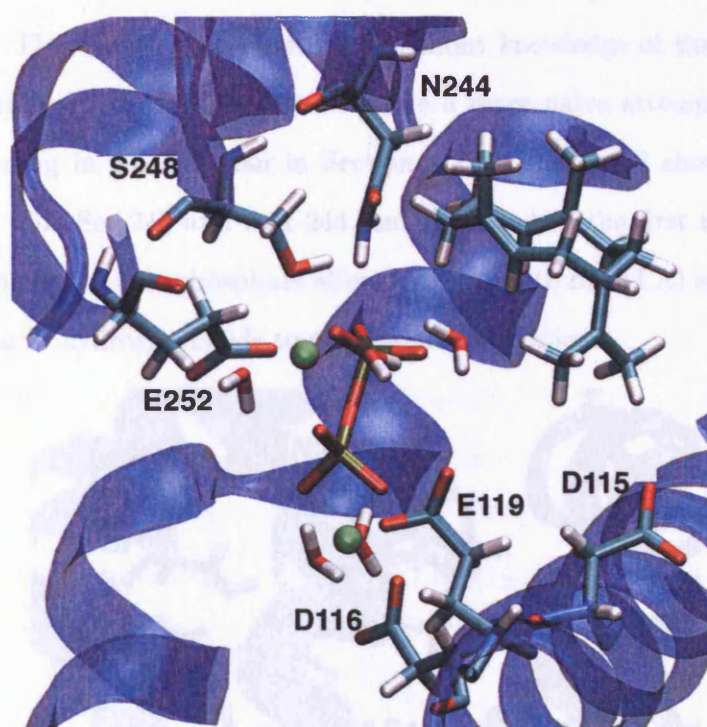
Figure 4.1: *Active site of model A* - showing key residues in proposed magnesium binding domain.

This hydrogen bonding network within the docked structure is consistent with the experimental observation that Asn 244 is essential for the catalytic activity of AS-PR. [75] The second Mg^{2+} ion is coordinated to the other oxygen atom of the pyrophosphate anion, exposed to solvent water and close to the aspartate rich motif of Asp 115, Asp 116 and Asp 119. [30]

4.2.2 Preparation of holoenzyme model B

A third magnesium ion was added to the structure of model A in a position approximately equidistant between to two magnesium cations already present and in a position that was accessible to solvent in an attempt to model a magnesium entering AS-PR from solvent. This model was constructed without knowledge of the structure and magnesium binding in AS-AT and is therefore a more naïve attempt at predicting the Mg^{2+} binding in AS-PR than in Section 4.2.3. Figure 4.2 shows that in this structure Glu 252, Ser 248 and Asn 244 remain bound to the first magnesium ion. The orientation of the pyrophosphate alters (compared to model A) so that a second oxygen atom now hydrogen bonds to two magnesium cations.
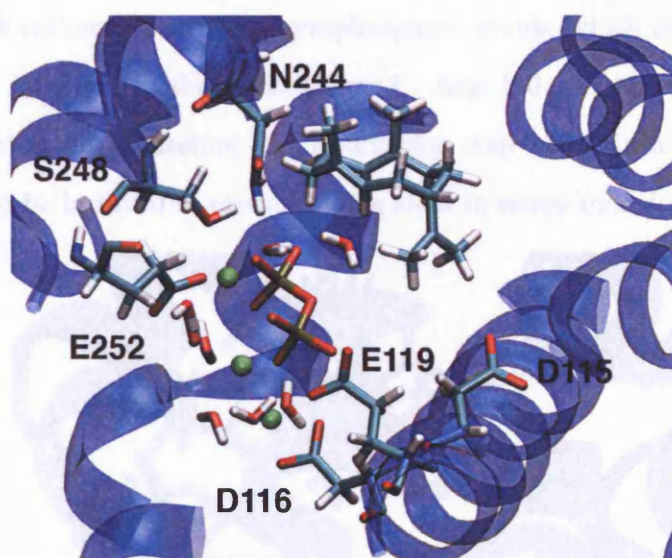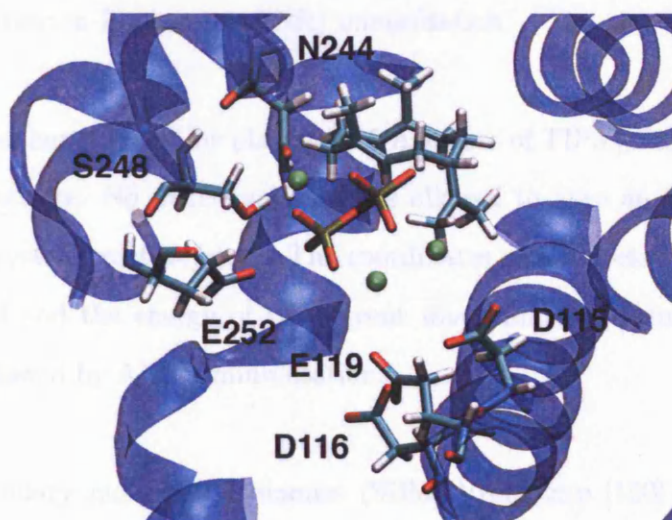

Figure 4.2: *Active site of model B* - showing key residues in proposed magnesium binding domain

4.2.3 Preparation of holoenzyme model C

The coordinates of model B and the crystal structure of chain D (the closed, pyrophosphate bound subunit) of AS-AT were overlayed by superimposing the backbone atoms of the aspartate rich domains of each structure using the SWISS-MODEL code. [180, 181] The coordinates of pyrophosphate and magnesium atoms in model B were replaced by the coordinates of pyrophosphate and magnesium from the resulting overlay and formed the coordinate set for a new docked structure, model C.

In this model one Mg^{2+} remains bound to Glu 252, Ser 248 and Asn 244. However, the position of the remaining magnesium cations is different from that of model B. In model C the cations are close to pyrophosphate group, which is itself orientated differently and is closer to the docked ligand. Asp 116 and Glu 119 remain well positioned to bind to magnesium and in addition Asp 115 is in a more favourable conformation to be involved in metal binding than in either model A or B.

Figure 4.3: *Active site of model C* - showing key residues in proposed magnesium binding domain

4.2.4   Molecular dynamics simulations

CHARMM [101] (version c29b2) was used for all calculations unless otherwise stated. The force field employed was the CHARMM27 force field. [98] Missing parameters for pyrophosphate were created by analogy with existing parameters in the CHARMM27 force field.

The initial docked structure was truncated to a 25 Å sphere centred on the substrate. All residues with at least one heavy atom within the sphere were preserved; as a result five residues were deleted. Charged residues near the surface had their charges scaled to give a total charge of zero [1] and also to retain correct hydrogen bonding properties at the surface region.

Following truncation the energy of the enzyme, pyrophosphate and magnesium ions was minimised using 500 steps of steepest descent (SD) followed by 1000 steps of adopted basis Newton-Raphson (ABNR) minimisation.

The system was then solvated by placing a 25 Å sphere of TIP3 [178] water molecules around the substrate. No water molecule was allowed to have an initial position $\leq$ 2.2 Å from a crystal structure atom. The coordinates of the docked structure atoms were then fixed and the energy of the solvent was minimised using 500 then 1250 steps of SD followed by ABNR minimisation.

Stochastic boundary molecular dynamics (SBMD)(reference [150] and Chapter 2) were performed on the water molecules only. The friction coefficient for water oxygen

---

[1]Using a software patch kindly supplied by Dr Jolanta Żurek.

was 62 ps^{-1}. The non-bonded interaction cutoff distance was set to 13 Å since this offered the optimum balance between simulation time and accuracy. All other MM parameters were the defaults within CHARMM. The system was heated from 10 K to 298 K over 10 ps using 10,000 steps (*i.e.* a 1 fs time step). Following heating, 30 ps of equilibration dynamics at 298 K and 250/1000 steps of SD/ABNR energy minimisation were performed.

The coordinates of the enzyme, pyrophosphate and magnesium were then unfrozen (*i.e.* only the substrate was left fixed) and SBMD were performed on the whole system. The friction coefficient for protein heavy atoms was set at 250 ps^{-1}. The buffer radius was 21 Å, with the buffer region (21 to 25 Å) additionally divided into four zones, with force constraints applied to atoms as described in Table 4.1[2]. The SBMD conditions outlined here were used in all subsequent simulations unless otherwise stated.

The system was heated from 10 K to 298 K over three separate 10 ps SBMD simulations (*i.e.* 10 K to 100 K, followed by 100 K to 200 K, followed by 200 K to 298 K). Following heating, the holoenzyme was equilibrated for 30 ps at 298 K. The resulting set of coordinates was used as the starting point for QM/MM simulations described in the following section.

---

[2]These constants were used as a result of personal communication with Dr Jolanta Żurek

Table 4.1: Force constants used in buffer region of SBMD

Atom Type	Backbone	β Carbons	γ Atoms (C, O, S)	Other Heavy Atoms
Buffer 1: 21 to 22Å	0.27	0.24	0.23	0.33
Buffer 2: 22 to 23Å	0.85	0.78	0.73	0.69
Buffer 3: 23 to 24Å	1.44	1.31	1.22	1.18
Buffer 4: > 24 Å	1.70	1.55	1.45	1.4

4.2.5 QM/MM minimisation and molecular dynamics

Before performing free energy sampling simulations it was necessary to introduce a QM region into the system, and describe the interaction between the QM region and the MM region. The general principles and features of the QM/MM approach have been described fully in the literature [177] and elsewhere in this work (Chapter 2).

There were three separate QM regions and each was used within each model. First for the case of the intramolecular proton transfer the substrate alone formed the QM region and no description of a covalent link between QM and MM region was required. Second, for the case of water mediate proton transfer, the substrate and the water molecule formed the QM region, with no covalencey between molecules.

Finally, to investigate to role of Lys 206 the GHO method [142] was used. The GHO method requires a QM atom to be selected which will be the only atom allowed to interact with the MM region. In this work the boundary atom was the backbone carbon and side chain as shown in Figure 4.4.
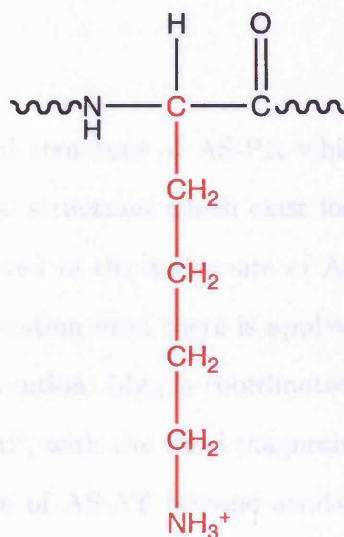
Figure 4.4: *GHO boundary for Lys 206* - GHO lysine showing QM atoms in red, MM atoms in black. The Boundary atom is the red backbone carbon.

4.2.6 Investigating the role of specific amino acids in AS-PR catalysis

Residues proposed to be involved in magnesium ion binding have been identified, and as discussed in Chapter 1 and the references cited therein. These are contained within an 'aspartate-rich' domain (Asp 115 to Glu 119) and "NSE/DTE" motif (in AS-PR consisting of Asn 244, Ser 248 and Glu 252) which are common to many sesquiterpene synthases.

The trajectories obtained during the simulations described in this chapter were analysed using software VMD and related packages [182–189] and the distance between the closest magnesium ion and Asp 115, Asp 116, Glu 119, Asn 244, Ser 248 and Glu 252 (the proposed magnesium ligands) were measured, with the distance measured from the atom which one would expect to most strongly ligate Mg^{2+} (*i.e* carboxylate

oxygen or hydroxyl oxygen).

Although there is no crystal structure of AS-PR which contains magnesium ions, there are a number of crystal structures which exist for AS-AT [43,44] in which the three magnesium ions observed in the active site of AS-AT are labelled Mg_A, Mg_B and Mg_C. The naming convention used there is applied here for AS-PR and results in the following naming convention: Mg_A is coordinated directly to E119 and Mg_B is coordinated to the NSE motif, with the third magnesium (if present) labelled Mg_C. Although a crystal structure of AS-AT became available during the course of this work it was not used as the Allemann group has more direct experience with AS-PR and there are more mutagenesis and substrate analogue experiments with which to compare computational data.

4.2.7   Free energy simulations

Having defined the relevant QM section all restraints on atoms were removed and 500/1250 steps of SD/ABNR QM/MM minimisation were performed. PM3 [126] was used as the semi-empirical QM code for all calculations. The QM region consisted of 15 heavy atoms (no water involved in proton transfer) or 16 heavy atoms (water involved). The coordinates of this structure were saved and visually inspected using VMD [182] to ensure no unreasonable geometries were generated.

Following minimisation, QM/MM SBMD dynamics were performed for 30 ps at 298 K and the resulting coordinates and velocities were used as the starting point for free energy sampling simulations.

Umbrella sampling along a reaction coordinate was used to investigate the free energy changes for each step of the reaction. [190] A description of the techniques is included as Section 2.10. A reaction coordinate for each step was chosen and restrained to the initial starting value. QM/MM SBMD was restarted using the coordinates and trajectories from the previous dynamics step, with the same SBMD parameters as previously described.

A force of 200 kcal mol^{-1} Å^{-2} was used as umbrella potential for each simulation, and a 13 Å cut off was used for the generation of non-bonded interactions unless otherwise indicated. The system was equilibrated for 1 ps, and the dynamics continued for another 9 ps. This total of 10 ps was chosen as it allowed for an acceptable length of simulation with little apparent loss of accuracy compared to 20 and 30 ps simulations which were conducted to test the MD protocol. In each subsequent simulation, the value of the reaction coordinate was adjusted by $\pm$ 0.1 Å in the direction of the product and the dynamics restarted using the coordinates and velocities saved after 1 ps equilibration of the previous dynamics simulation. Reaction coordinate values during the 9 ps of production dynamics were saved and analysed using the Weighted Histogram Analysis Method [151] (WHAM) to produce free energy profiles.

4.3 Results - free energy simulation

4.3.1 Direct proton transfer

The reaction coordinate, RC, for the conversion of germacryl cation **26** to intermediate **64** *via* **TS2** was defined using Equation 4.3.1

$$RC = r_{DH} - r_{AH} \tag{4.1}$$

where r_{DH} is the magnitude of the vector between donor atom D (C12 in this case) and transferred atom H (proton attached to C12) and r_{AH} is the magnitude between H and acceptor atom A (C6 in this case).

From the results of the QM/MM equilibration dynamics the initial value of RC was -2.9, -3.4 and -3.2 Å for models A, B and C respectively. Sampling was started at this point and increased by 0.1 Å until sampling was stopped at $RC = 2.9$, 3.4 and 2.7 Å since at these values of RC the reaction step being described, by visual inspection, had clearly been reached and passed.

4.3.2   QM/MM free energy calculation for the conversion of 26 to 64 *via* TS2.

Data were collected and analysed using WHAM and plots of relative free energy against *RC* are displayed in Figure 4.5, Figure 4.6 and Figure 4.7.
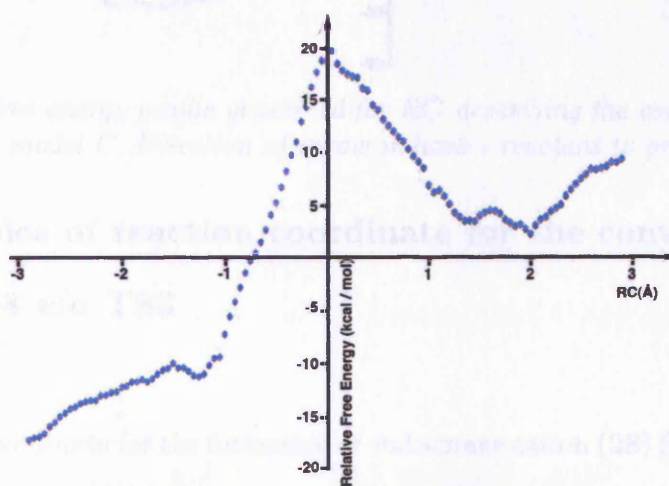
Figure 4.5: *Free energy profile generated for RC describing the conversion of **26** to **64** via **TS2** in model A. Direction of arrow indicates reactant to product transition*
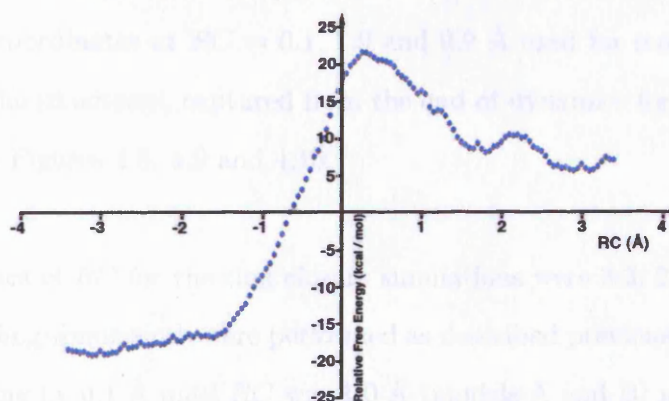
Figure 4.6: *Free energy profile generated for RC describing the conversion of **26** to **64** via **TS2** in model B. Direction of arrow indicates reactant to product transition*
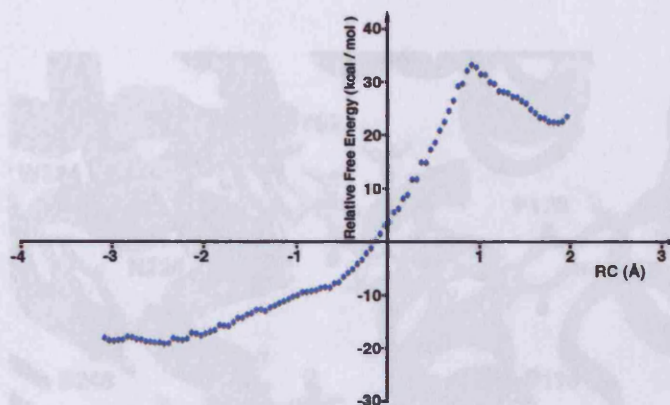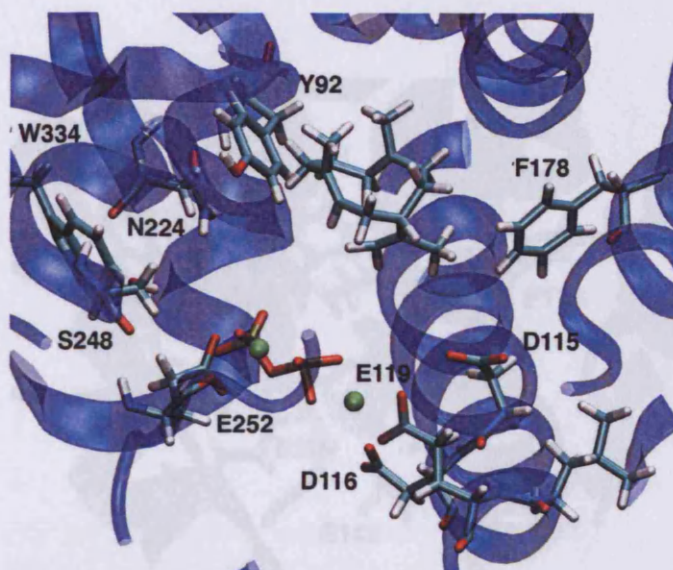
Figure 4.7: *Free energy profile generated for RC describing the conversion of **26** to **64** via **TS2** in model C. Direction of arrow indicates reactant to product transition*

4.3.3 Choice of reaction coordinate for the conversion of 64 to 28 *via* TS3

The reaction coordinate for the formation of eudesmane cation (**28**) from intermediate **64** *via* **TS3** was defined as the distance between C2 and C7. The starting point for each calculation was taken from the previous set of umbrella sampling calculations with the final coordinates at $RC = 0.1$, 1.9 and 0.9 Å used for models A, B and C respectively. The structures, captured from the end of dynamics for the value of RC are included as Figures 4.8, 4.9 and 4.10.

The initial values of RC for the ring closure simulations were 3.3, 2.9 and 3.2 Å and umbrella sampling simulations were performed as described previously with the value of RC decreasing by 0.1 Å until RC was 1.0 Å (models A and B) and 1.1 Å (model C). WHAM was used to produce free energy profiles and these are shown as Figure 4.11, Figure 4.12 and Figure 4.13.
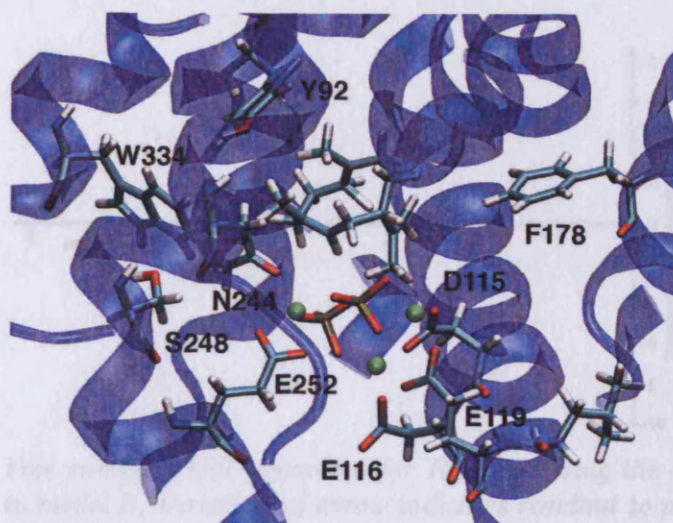
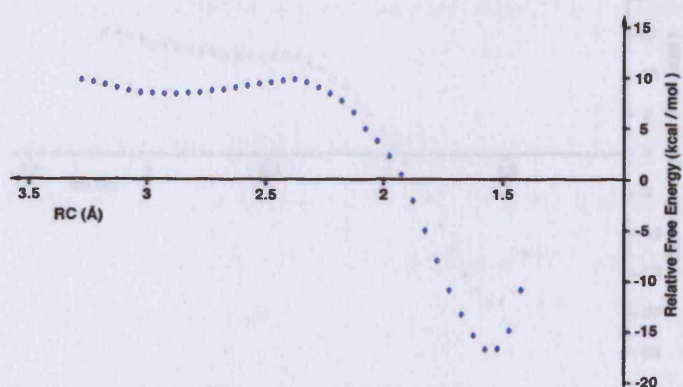**Figure 4.8:** *Structure of model A used as starting point for sampling simulation in Section 4.3.2 - RC = 0.1 Å*
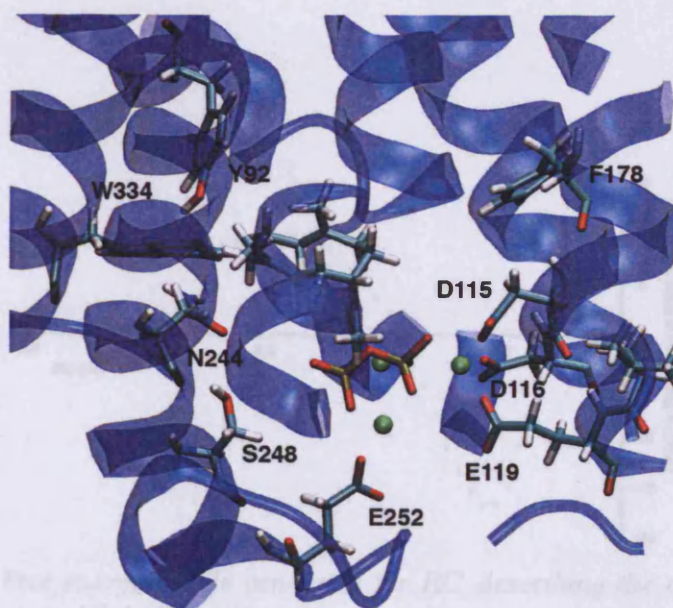
Figure 4.9: *Structure of model B used as starting point for sampling simulation in Section 4.3.2 - RC = 1.9 Å*
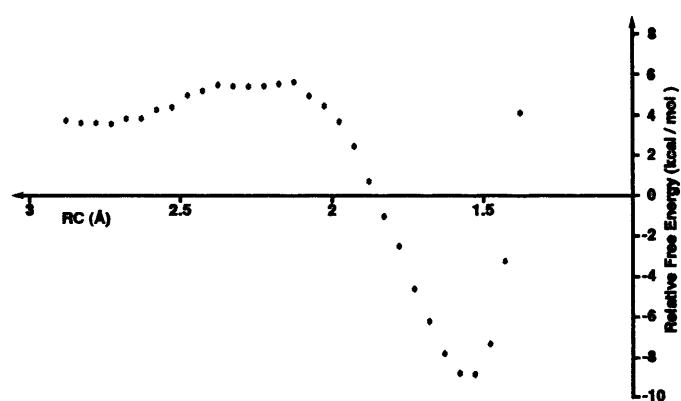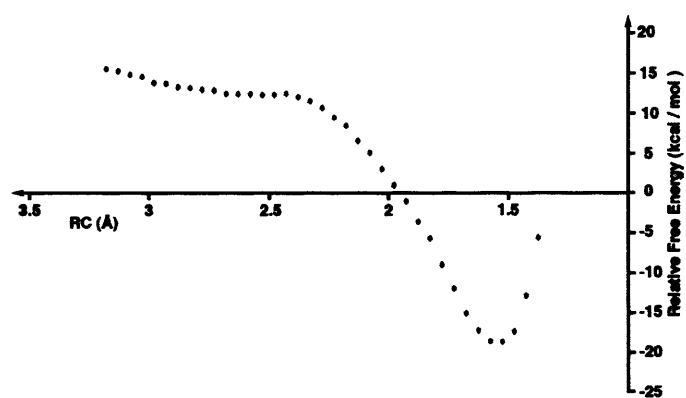
Figure 4.10: *Structure of model C used as starting point for sampling simulation in Section 4.3.2 - RC =0.9 Å*

Figure 4.11: *Free energy profile generated for RC describing the conversion of 64 to 28 via TS4 in model A. Direction of arrow indicates reactant to product transition*

Figure 4.12: *Free energy profile generated for RC describing the conversion of 64 to 28 via TS4 in model B. Direction of arrow indicates reactant to product transition*

Figure 4.13: *Free energy profile generated for RC describing the conversion of 64 to 28 via TS4 in model C. Direction of arrow indicates reactant to product transition*

4.3.4 Choice of reaction coordinate for the conversion of eudesmane cation, 28, to intermediate 65 *via* TS4

The structure obtained at the end of dynamics with with $RC = 1.5$ Å from the previous set of calculations was chosen to represent eudesmane cation (**28**) in the subsequent set of calculations. The structures associated with this value of RC are included as Figures 4.14, 4.15 and 4.16 respectively.
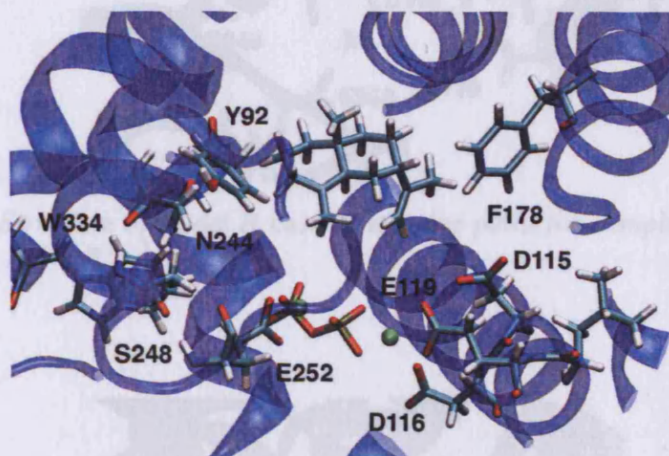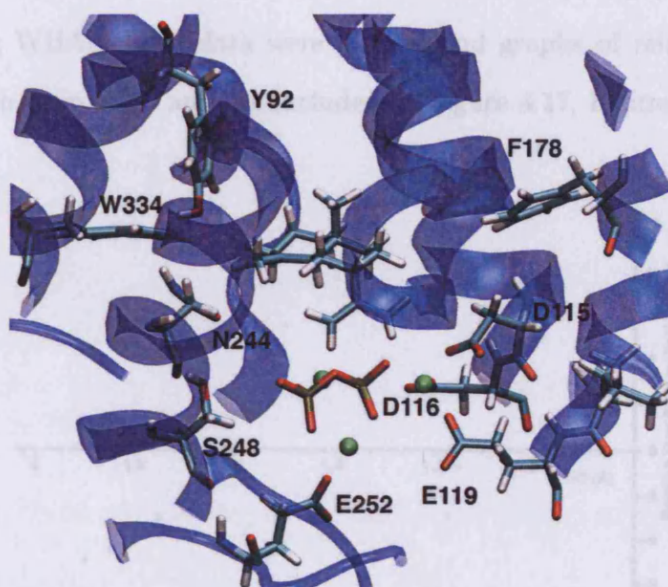

Figure 4.14: *Structure of model A used as starting point for sampling simulation in Section 4.3.4 - RC = 1.5 Å*

A new reaction coordinate was defined for models A and B, in which RC was the distance between the transferring hydride on C3 and C2. Initial values of RC in this simulation were 1.9 (model A) and 2.0 Å (model B) and RC was decreased by 0.1 Å until RC was 0.9 Å for both models.

For model C RC was defined using Equation 4.3.1 with D, A and H set to be C3, C2 and the hydride of C3 respectively. RC started at 1.9 Å and was decreased by 0.1 Å until $RC = 0.8$ Å.

Figure 4.15: *Structure of model B used as starting point for sampling simulation in Section 4.3.4 - RC = 1.5 Å*

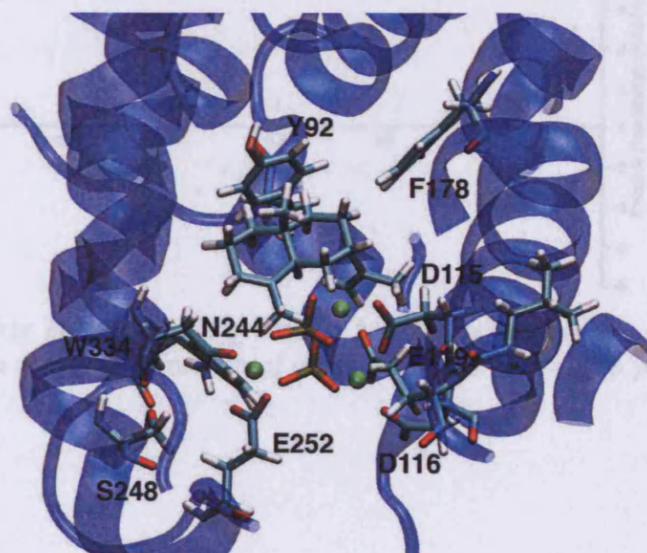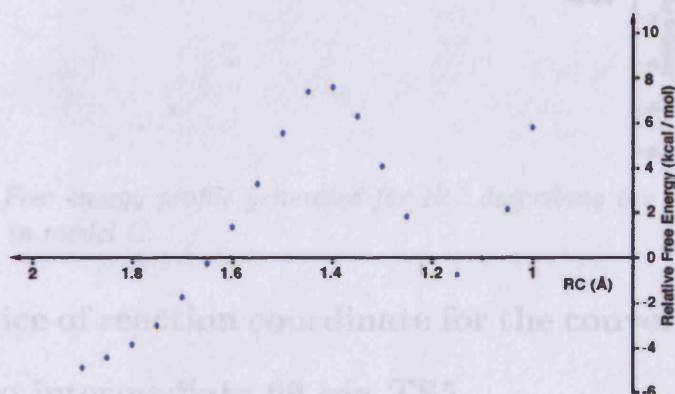**Figure 4.16:** *Structure of model C used as starting point for sampling simulation in Section 4.3.4 - RC = 1.5 Å*

Umbrella sampling simulations were performed as described previously and the results analysed using WHAM. The data were plotted and graphs of relative free energy against reaction coordinate and are included as Figure 4.17, Figure 4.18 and Figure 4.19.

Figure 4.17: *Free energy profile generated for RC describing the conversion of 28 to 65 via TS4 in model A. Direction of arrow indicates reactant to product transition*

Figure 4.18: *Free energy profile generated for RC describing the conversion of 28 to 65 via TS4 in model B. Direction of arrow indicates reactant to product transition*

Figure 4.19: *Free energy profile generated for RC describing the conversion of **28** to **65** via **TS4** in model C.*

4.3.5 Choice of reaction coordinate for the conversion of intermediate 65 to intermediate 66 *via* TS5

The structures from the end of the previous set of dynamics with RC 1.4, 1.2 and 1.0 Å (for models A, B and C respectively) were used as the starting points in the subsequent set of dynamics. A reaction coordinate to describe the C14 methyl group shifting from the C7 to the C2 position was described using Equation 4.3.1, where D, H and A were defined as C7, C14 and C2 respectively.

The starting value for RC for models A, B and C in this simulation was -1.0 Å and this value was increased by 0.1 Å until $RC = 1.0$ Å (model A) or 0.9 Å (models B and C). The data obtained were analysed using WHAM and plots of relative free energy against RC are included as Figure 4.23, Figure 4.24 and Figure 4.25.

Figure 4.20: *Structure of model A used as starting point for sampling simulation in Section 4.3.5 - RC = 1.4 Å*

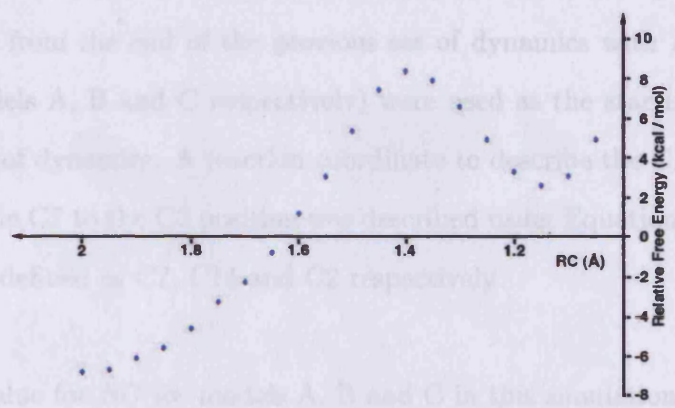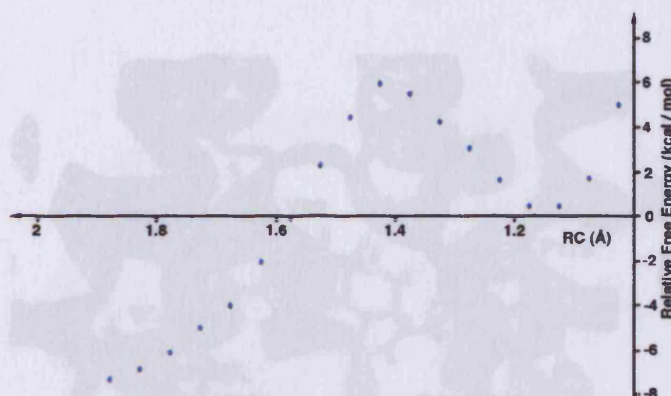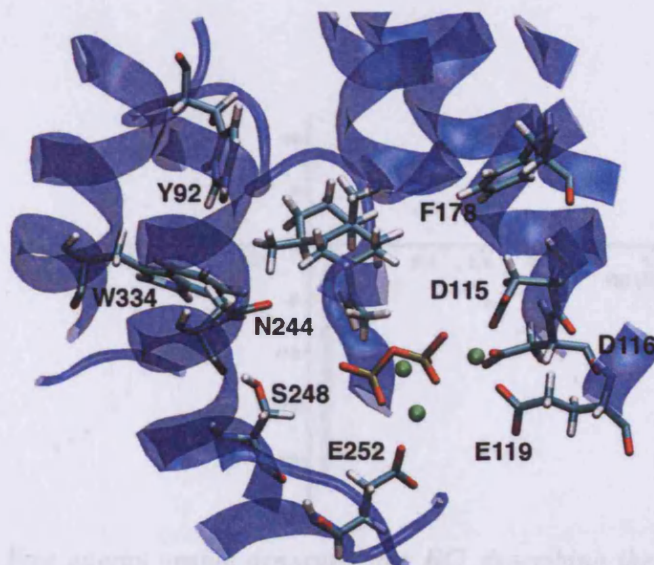**Figure 4.21:** *Structure of model B used as starting point for sampling simulation in Section 4.3.5 - RC = 1.0 Å*
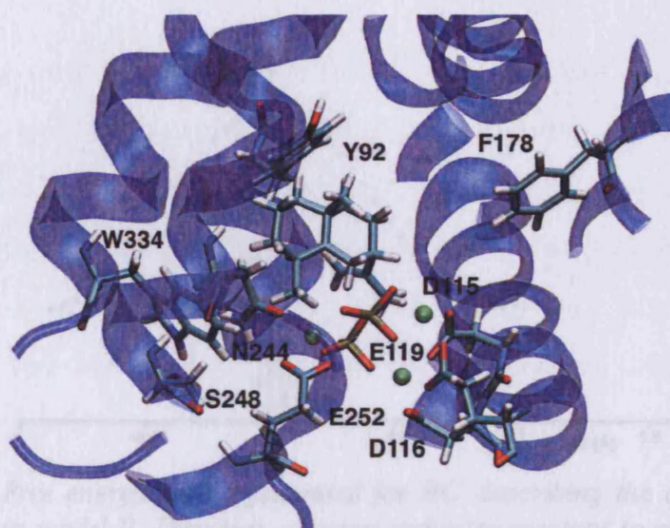
Figure 4.22: *Structure of model C used as starting point for sampling simulation in Section 4.3.5 - RC = 1.2 Å*
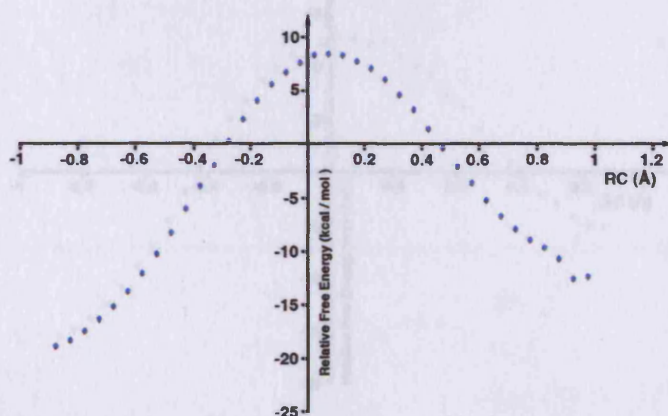
Figure 4.23: *Free energy profile generated for RC describing the conversion of **65** to **66** via **TS5** in model A. Direction of arrow indicates reactant to product transition*

Figure 4.24: *Free energy profile generated for RC describing the conversion of* **65** *to* **66** *via* **TS5** *in model B. Direction of arrow indicates reactant to product transition*

Figure 4.25: *Free energy profile generated for RC describing the conversion of* **65** *to* **66** *via* **TS5** *in model C. Direction of arrow indicates reactant to product transition*

4.3.6 Free energy profile for combined reaction coordinate

The free energy profiles were then combined to show a free energy profile for a complete reaction coordinate from **26** to **TS5**. The free energy of the structure in the initial umbrella sampling simulation (*i.e.* **26** to **64** *via* **TS2**) was set to 0 kcal mol^{-1} and a suitable minimum energy point being used as the starting free energy value when the subsequent free energy profile was included. The x axis was normalised such that each step had equal width, even though the actual number data points differs for each simulation step.

A combined profile, from **26** to **TS5** was produced in this manner and a plot of relative free energy against this non-dimensional reaction coordinate is shown as Figure 4.26.

Figure 4.26: *Combined free energy profile for the conversion of 26 to TS5* - Free energy of **26** set to 0 kcal mol^{-1}.

The data used to make Figure 4.26 were altered so that the eudesmane cation **28** structure was set to be the zero point of a new free energy profile, included as Figure 4.27.

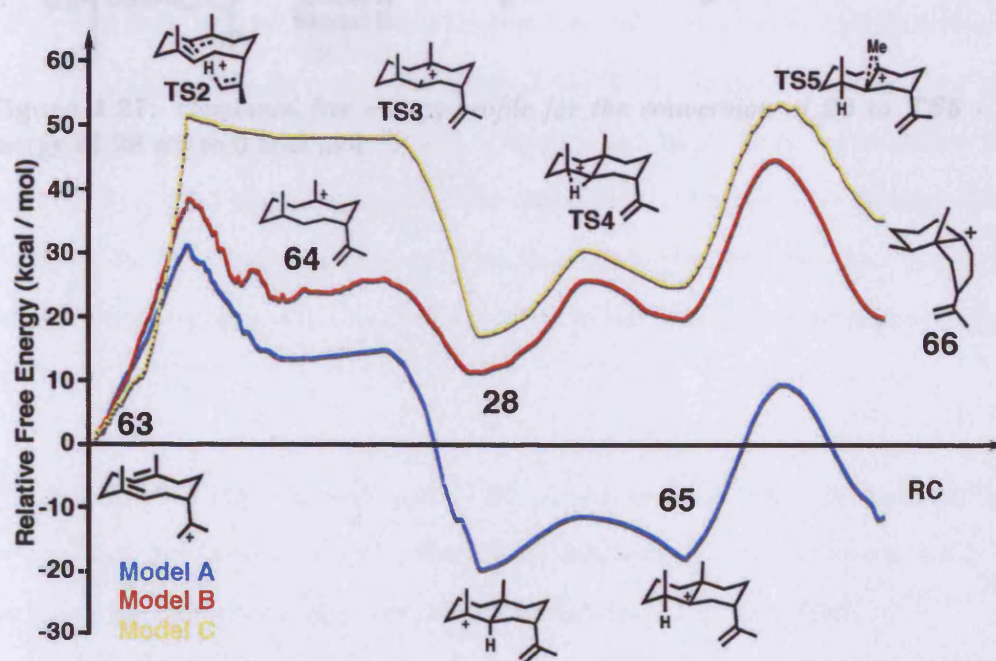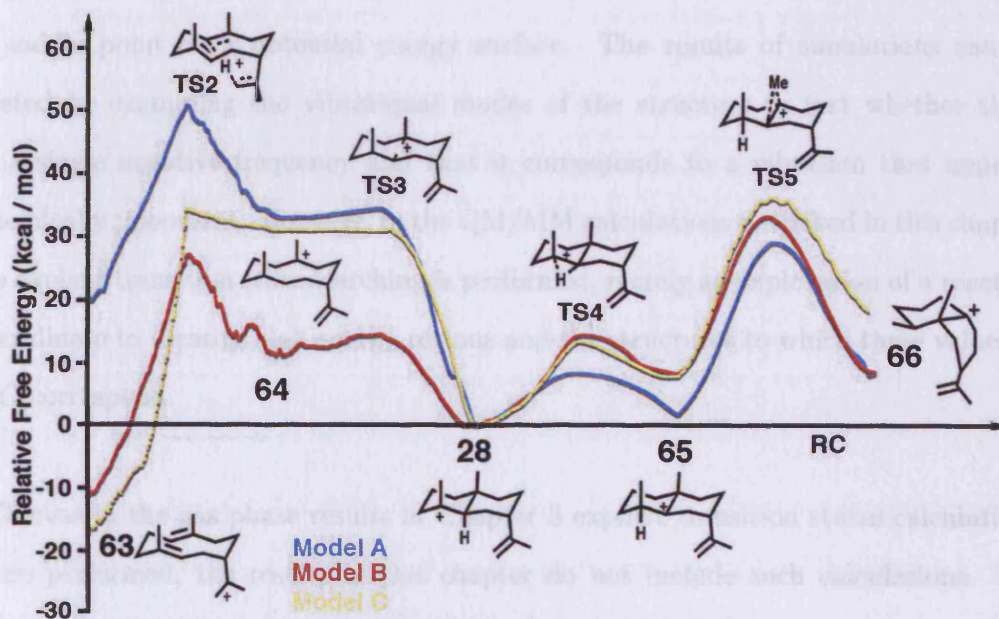**Figure 4.27:** *Combined free energy profile for the conversion of **26** to **TS5** -* Free energy of **28** set to 0 kcal mol^{-1}.

4.3.7 Structures of transition states and intermediates

A gas phase QM calculation aims to locate a 'true' transition state structure - *i.e.* a saddle point on a potential energy surface. The results of simulations can be tested by examining the vibrational modes of the structure to test whether there is a single negative frequency and that it corresponds to a vibration that appears chemically important. However, in the QM/MM calculations described in this chapter no explicit transition state searching is performed, merely an exploration of a reaction coordinate to identify high energy regions and the structures to which these values of *RC* correspond.

Whereas in the gas phase results in Chapter 3 explicit transition states calculations were performed, the results in this chapter do not include such calculations. For ease of nomenclature, however, the labels **TS2**, **TS3**, etc have been used to name structures throughout this work - but this does not imply that the structure itself is a true transition state in the QM/MM calculations. Although no transition state searching has been performed, it is still possible to compare the geometric parameters described in Chapter 3 with the equivalent data in the QM/MM simulations contained here.

The structures of the germacryl cation, **26**, in models A, B and C immediately prior to the start of the umbrella sampling simulations described in Section 4.3.2 were examined and the values of d_1 (as defined in Section 3.3.2) measured.

The values of *RC* corresponding to maximum free energy in Figures 4.5, 4.6 and 4.7 (0.0 Å, 0.3 Å and 0.9 Å respectively) were visually inspected and were superficially

Table 4.2: Distances d_1, d_2 and d_3 and angle α for the conversion of **26** to **28**

	d_1 (Å)			d_2 (Å)	d_3 (Å)	α (deg)	β (deg)
Method	**26**	**TS2**	**28**	**TS2**			**26**
AM1	2.92	2.99	1.55	1.45	1.45	151	23.0
PM3	3.09	3.03	1.55	1.45	1.45	151	15.9
mPW1PW	3.10	3.02	1.66	1.46	1.44	161	18.4
MPWB1K	3.00	2.98	1.64	1.47	1.46	159	15.9
QM/MM (model A)	2.96	3.30	1.55	1.51	1.47	158	17.2
QM/MM (model B)	3.03	3.20	1.53	1.3	1.62	158	16.3
QM/MM (model C)	2.95	3.27	1.52	1.48	1.51	160	16.0

similar to the structure of **TS2** calculated in Chapter 3 (Figure 3.4) and therefore values of distances d_1, d_2 and angles α and β (again, as defined in Section 3.3.2) were recorded and are included here as Table 4.2.

In a similar manner, the structures of substrate corresponding to minimum energy values of RC in Figures 4.14, 4.15 4.15 (1.5 Å in all cases) were visually inspected and found to be similar to the structure of eudesmane cation **28** calculated in Section 4.3.2 and hence the value of d_1 was obtained from those values of RC.

The data displayed in Figures 4.17, 4.18 and 4.19 indicated a maximum free energy at RC =1.4 Å for all models, and following visual inspection of the substrate geometry at this value of RC, which revealed a structure consistent with the structure of **TS4** calculated in Section 3.3.3, distances d_4 and d_5 and dihedrals γ and ϵ were measured. In a similar fashion the substrate structure at RC =1.1 Å for all models was inspected for similarity to **65** in Section 3.3.3 and values of dihedrals γ and ϵ measured. The

Table 4.3: Distance between hydride and C2 (d_4), hydride and C3 (d_5) in Å. Dihedral angles described by C1 C2 C3 C7 (γ) and C2 C3 C4 and C15 (ϵ) in degrees.

	d_4 (Å)	d_5 (Å)	γ (deg)			ϵ (deg)		
Method	TS4		28	TS4	65	28	TS4	65
AM1	1.4	1.4	129.0	171.1	178.7	169.2	177.6	129.7
PM3	1.5	1.4	133.1	171.6	179.4	165.4	177.1	130.5
mPW1PW1	1.4	1.4	119.6	170.1	178.7	169.3	174.1	128.3
MPWB1K	1.3	1.4	117.6	170.1	180.0	177.9	173.1	137.1
QM/MM (model A)	1.5	1.4	137.0	178.0	177.4	177.5	174.0	137.6
QM/MM (model B)	1.7	1.3	135.5	174.4	177.9	161.2	152.1	136.3
QM/MM (model C)	1.4	1.5	134.2	164.8	178.1	178.6	177.5	128.1

data obtained are included as Table 4.3.

The maximum value of free energy in Figures 4.23, 4.24 and 4.25 corresponded to values of RC =0.1 Å (model A) and RC =0.0 Å (models B and C) and the substrate geometry was visually similar to that observed in Figure 3.7 and the distances d_6 and d_7 were recorded in Table 4.4. In a similar manner the structures represented by RC =0.9 Å for all three models was assessed against the structure of **66** in Figure 3.7 and values of distances d_6 and d_7 were recorded in Table 4.4.

The data contained with Table 4.2, 4.3 and 4.4 demonstrate that the key geometric parameters described in the true transition states and intermediates identified in Chapter 3 share geometric parameters with the high and low free energy regions observed in the results of the combined QM. On that basis it is reasonable to make comparisons between the free energy barriers described in Chapter 3 with the acceptance

Table 4.4: Distance between migrating methyl group and C7 (d_6) and migrating methyl group and C2 (d_7) in Å.

	d_6 (Å)	s_7 (Å)	d_6 (Å)	s_7 (Å)	d_6 (Å)	d_7 (Å)
Method	65		TS5		66	
AM1	1.5	2.5	2.1	2.1	2.4	1.5
PM3	1.5	2.4	2.1	2.1	2.4	1.5
MPW1PW	1.6	2.4	2.1	2.1	2.3	1.6
MPWB1K	1.6	2.3	2.1	2.1	2.3	1.6
QM/MM (model A)	1.5	2.5	2.0	1.9	2.5	1.6
QM/MM (model B)	1.6	2.5	2.0	1.9	2.5	1.6
QM/MM (model C)	1.6	2.5	2.0	1.9	2.4	1.5

that there may be small differences between structures and hence free energies in the two sets of experiments.

4.3.8 Interaction between Asp 115 and magnesium ion during direct intramolecular proton transfer

Figures B.1 to B.3 show the interaction between Mg_B and Asp 115. Models B and C show a remarkably similar distance relationship between the magnesium ion and Asp 115, with a distance of approximately 1.9 Å throughout the course of the QM/MM free energy sampling. Such a distance is indicative of a tightly bound Asp - Mg^{2+} complex. In contrast, model A shows a much weaker binding with the mean distance approximately 4 Å. There was no clear variation of Mg Asp 115 distance corresponding to the formation of the various intermediates. Asp 115 is of particular interest because

it has been demonstrated that when pyrophosphate is present in an 'active' AS-AT complex Asp 90 (the analogue to Asp 115 in AS-PR) binds in a *syn-syn* bidentate fashion to both Mg_A and Mg_C [44], whereas in the absence of pyrophosphate Asp 90 binds in a bidentate fashion to Mg_C. This is not present in model A (since there is no Mg_C). Nor is this *syn-syn* bidentate ligation present during the QM/MM simulation described in the chapter form model B (see, for example, Figures 4.15 and 4.21). However, model C does exhibit the behaviour observed in in AS-AT and examples of this can be seen in Figures 4.16 and 4.57.

4.3.9 Interaction between Asp 116 and magnesium ion during direct intramolecular proton transfer

Figures B.4 to B.6 show the interaction between Mg_B and Asp 116. In contrast to the results described in Section 4.3.8 all three models show a similarly tight binding (Mg Asp 116 distance of *ca.* 1.9 Å). In a similar fashion to the results in Section 4.3.8 there was no relationship between the Asp 116 Mg_B distance and the formation of any of the sesquiterpene intermediates.

4.3.10 Interaction between Glu 119 and magnesium ion during direct intramolecular proton transfer

Glu 119 behaved in a similar manner to Asp 115: there was evidence of tight binding in Models B and C, but this binding was weaker, if present at all, in model A where

the mean distance was in the region of 5 Å and this binding became weaker as the simulation progressed. However, in Models B and C the binding was tighter (mean distance *ca.* 2.0 Å) and did not vary during the progress of the simulation.

4.3.11 Interaction between Asn 244 and magnesium ion during direct intramolecular proton transfer

In model C, Figure B.12, there is good evidence for binding between Asn 244 and Mg_B, with a mean Asn-Mg distance of *ca.* 2 Å. In the case of Models A and B (Figures B.10 and B.11) there is no real evidence of binding. In the case of model A there was a trend for the Asn 244 Mg distance to increase as the simulation progressed and in the case of Models B and C there was a trend for the Mg-Asn 244 distance to increase during the course of the simulation. This binding in Models B and C is slightly tighter (2.9 Å) than that observed in subunit D of AS-AT. [43]

4.3.12 Interaction between Ser 248 and magnesium ion during direct intramolecular proton transfer

Ser 248 is unique in that none of the Models in Figures B.13, B.14 or B.15 showed any evidence of tight binding to Mg_B during the simulations - this is in contrast to the 2.2 Å Ser 223 Mg_B distance observed in subunit D of the crystal structure of AS-AT published by the Christianson group. [43]

4.3.13 Interaction between Glu 252 and Mg_C ion during direct intramolecular proton transfer

Figures B.16 to B.18 show the interaction between the magnesium ion and Glu 252. All three models showed a tight binding between Mg_C and Glu 252, with little variation in the mean distance of *ca.* 1.9 Å throughout the simulation - this binding is tighter than that reported in the crystal structure of an AS-AT - pyrophosphate complex of 2.4 Å [43] - an indication that Glu 252 is able to bind more tightly either as a result of having a fully occupied active site, or as a result of the weakened interaction with Ser 248.

4.3.14 Discussion of interaction between Phe 178 Trp 334 and Tyr 92 and C7, C3 and C2 in direct intramolecular transfer mechanism

Figures A.10 to A.18 provide information about the role of Trp 334 during during the simulations of the direct intramolecular proton transfer mechanism. In general, the carbon centre which shows the greatest variation of distance over the course of the reaction is C7.

Either before formation of **64** (Figures A.10 and A.11) or during formation of **4** (A.12) Trp 334 is closest to C7 (a distance of approximately 4 Å (model C) to 9 Å (model A). This is consistent with the conclusions drawn by Deligeorgopoulou *et al* [73] and discussed in Section 1.3.2 in which a proposed role of Trp 334 is to stabilise the

reaction prior to formation of eudesmane cation **28**. In addition, there is a decrease in W334 C7 distance in all three models as **66** is formed - this species contains a positive charge on C7 and therefore on the basis of this result it appears W334 may serve a hitherto unreported dual role, in which it aids formation of eudesmane cation, then stabilises the positive charge on the final charges species of the cascade.

There is no clear relationship in the case of the Trp 334 C3 distance. A carbocation is formed on C3 following the formation of eudesmane cation until the formation of **TS4**. There has been no known suggestion of a role for Trp 334 in the previous literature and the results of these simulations do not suggest such a role.

A similar argument regarding the relationship between Trp 334 and C2 can be made - there is no consistent relationship between the distance and no experimental results to suggest this should be the case and therefore the results of the combined QM/MM simulation can be seen as consistent with experiment.

The dual role proposed for Phe 178 [83] discussed in Section 1.3.2 requires that Phe 178 stabilises both the cyclisation process which forms germacryl cation **26** and also the formation of eudesmane cation **28**. There is less evidence for this arising from these simulations which is consistent with the general observation that smaller reductions of k_{cat}/K_M are produced with F178 mutants than with W334 mutants (Table 1.2). However the distance relationship between F178 and C7 (Figures A.1, A.2 and A.3) provides further evidence for this hypothesis - in particular both model A and model C (in Figures A.1 and A.3) show that F178 moves closer to C7 during the process that leads to formation of **28**, although this is less in model B.

In the cases of C2 and C3, there is little consistent evidence that F178 undergoes any motion which suggests it plays a significant part in catalysis following formation of eudesmane cation, and this is again consistent with experimental results described in Section 1.3.2.

Based on the results described in this chapter, the role of Tyr 92 is similar to that of Phe 178. In Figures A.19 to A.21 there is some evidence that the distance between C7 and Tyr 92 decreases in all models in order to stabilise a positive charge on C7 which will occur prior to the formation of eudesmane cation, although in common with the results for the Trp 334 result this is less true for model A.

There is not a clear relationship for Tyr 92 and C3 and C2 in any of the models. Although this is consistent with previous hypotheses regarding the role of Tyr 92 it does not shed any light on the observed production of valencene side products [176] in Tyr 92 mutants of AS-PR as one might expect Tyr 92 to show some affinity for stabilising a charge on C7 during the formation of **66** which might be disrupted by the replacement of Tyr by another amino acid leading to valencene rather than aristolochene synthesis.

4.4 Water mediated proton transfer

In order to investigate the water mediated proton transfer mechanism in model A a water molecule present in the initial docked structure was made part of the QM region, following setup, solvation, heating and equilibration as described previously in Section 4.2.4. This water molecule was chosen based upon proximity to the active site, although, in principle, any water molecule could be taken an docked in a suitable position.

Models B and C were modified so that an additional water molecule, with the same approximate position as the water molecule in the modified model A described in the previous paragraph. The same setup, solvation, heating and equilibration procedure was applied, again, with the additional water molecule included in the QM region but with no covalent bonding required between water and substrate.

In the modified model A the reaction coordinate for the formation of germacryl cation **26** to a charged complex containing germacrene A and a hydroxonium cation *via* **TS2a** was defined using Equation 4.3.1 with H the proton closest to the oxygen atom on C12 (atom D) and A defined as the oxygen atom of the water molecule. The initial value of RC was *ca.* -1.0 Å and umbrella sampling was performed as described previously with the value of RC varied from -1.0 to +1.0 in 0.1 Å increments. WHAM analysis of the resulting data was performed and the results of this simulation are included as Figure 4.28.

The same definition of RC was employed with model B, and the value of RC had

an initial value of *ca.* -1.5 Å and was varied between -1.5 and +1.5 Å in 0.1 Å increments during umbrella sampling. In the modified model C the initial value of *RC* was *ca.* -3.2 Å and the value of *RC* was varied between -3.2 and +1.6 Å in 0.1 Å increments during umbrella sampling. Following application of the WHAM the results were plotted and are included as Figure 4.29 and Figure 4.30 respectively.
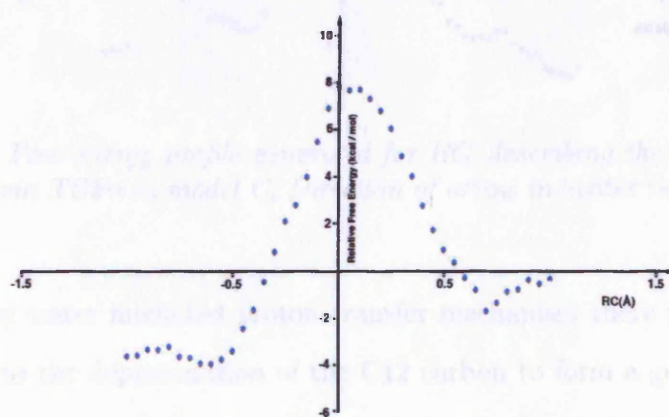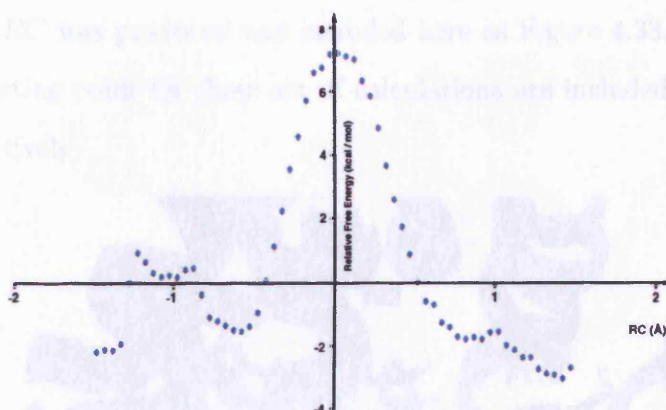

Figure 4.28: *Free energy profile generated for RC describing the deprotonation of germacrene A via **TS2a** in model A. Direction of arrow indicates reactant to product transition*
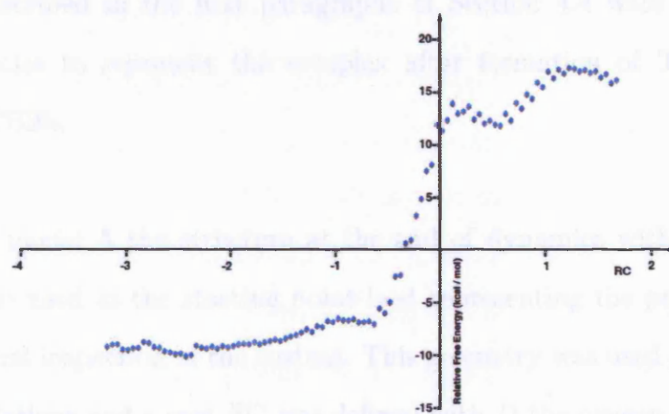
Figure 4.29: *Free energy profile generated for RC describing the deprotonation of germacrene A via **TS2a** in model B. Direction of arrow indicates reactant to product transition*

Figure 4.30: *Free energy profile generated for RC describing the deprotonation of germacrene A via **TS2a** in model C. Direction of arrow indicates reactant to product transition*

In the proposed water mediated proton transfer mechanism there is a step roughly corresponding to the deprotonation of the C12 carbon to form a germacrene A and hydroxonium ion, followed the protonation of the C6 C7 double bond by the acid water to form **TS4** *via* **TS2b** and regenerate the water molecule. The results of the simulations described in the first paragraphs of Section 4.4 were used to identify a suitable species to represent the complex after formation of **TS2a** but before formation of **TS2b**.

In the case of model A the structure at the end of dynamics with $RC = 0.3$ Å in Figure 4.28 was used as the starting point best representing the proposed structure based on a visual inspection of the system. This geometry was used as a new starting point for simulations and a new RC was defined with D the oxygen atom on the QM hydroxonium ion, A as C6 and H the proton on the hydroxonium ion closest to C6. The initial value of RC was *ca.* -3.0 Å and RC was varied from -3.0 to +3.0 in 0.1 Å increments, with umbrella sampling performed as previously described. Following the umbrella sampling simulations WHAM was performed and a plot of relative free

energy against RC was produced and included here as Figure 4.33. The structures used as the starting point for these set of calculations are included as Figures 4.31 and 4.32 respectively.
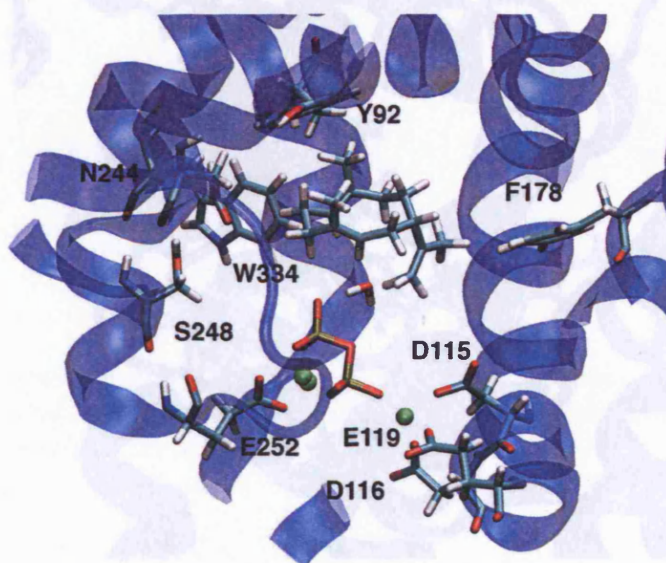

Figure 4.31: *Structure of model B used as starting point for sampling simulation in Section 4.4 - RC = 0.4 Å*

In the cases of model B and model C the structures described at the end of dynamics with $RC = 0.4$ Å in Figure 4.29 and $RC = 0.5$ Å in Figure 4.30 were used as the starting point for the next series of simulations for each model respectively. A new RC was defined in the same manner as in the previous paragraph. The initial value of RC was *ca.* -3.5 Å and RC was varied from -3.5 to +3.5 Å in 0.1 Å increments. Umbrella sampling was performed and WHAM used to produce plots of relative free energy against RC and these are included as Figure 4.34 and Figure 4.35.

The results of these two separate sets of simulations were combined to produce a plot of relative free energy against an arbitrary reaction coordinate. This plot is included as Figure 4.36
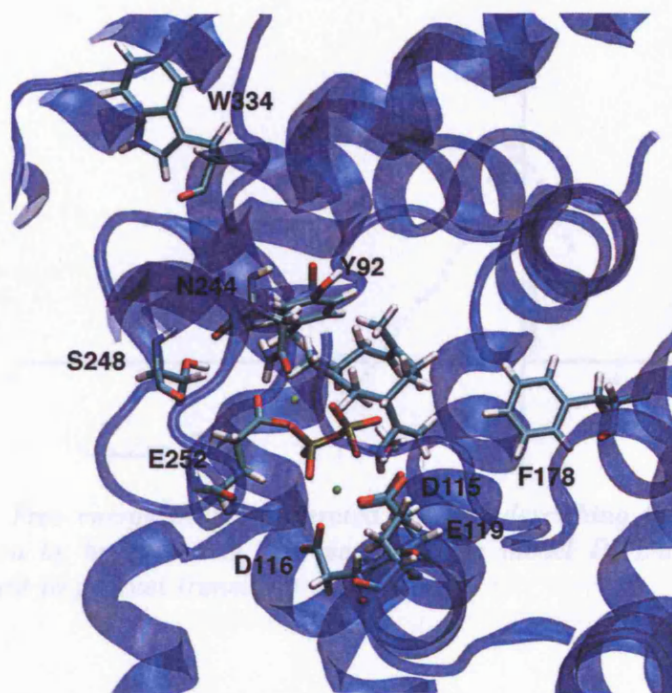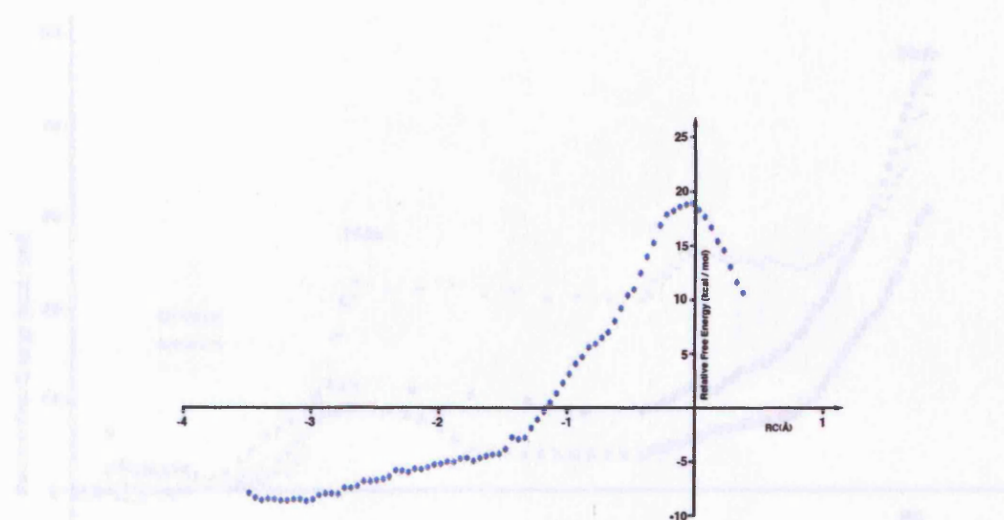
Figure 4.32: *Structure of model C used as starting point for sampling simulation in Section 4.4 - RC = 0.5 Å*
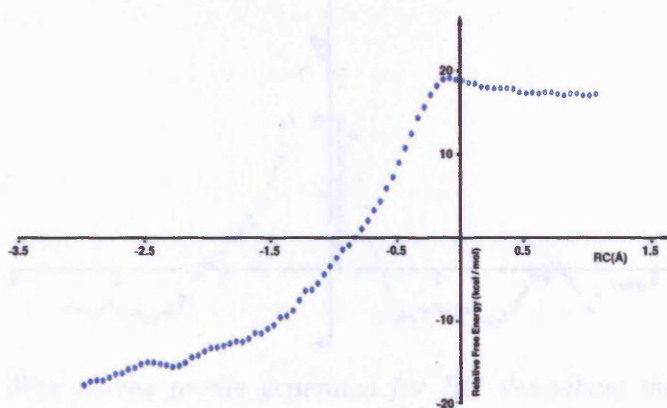
Figure 4.33: *Free energy profile generated for RC describing the protonation of germacryl cation by hydroxonium ion via* **TS2b** *in model A. Direction of arrow indicates reactant to product transition*

Figure 4.34: *Free energy profile generated for RC describing the protonation of germacryl cation by hydroxonium ion via* **TS2b** *in model B. Direction of arrow indicates reactant to product transition*

Figure 4.35: *Free energy profile generated for RC describing the protonation of germacryl cation by hydroxonium ion via* **TS2b** *in model C. Direction of arrow indicates reactant to product transition*
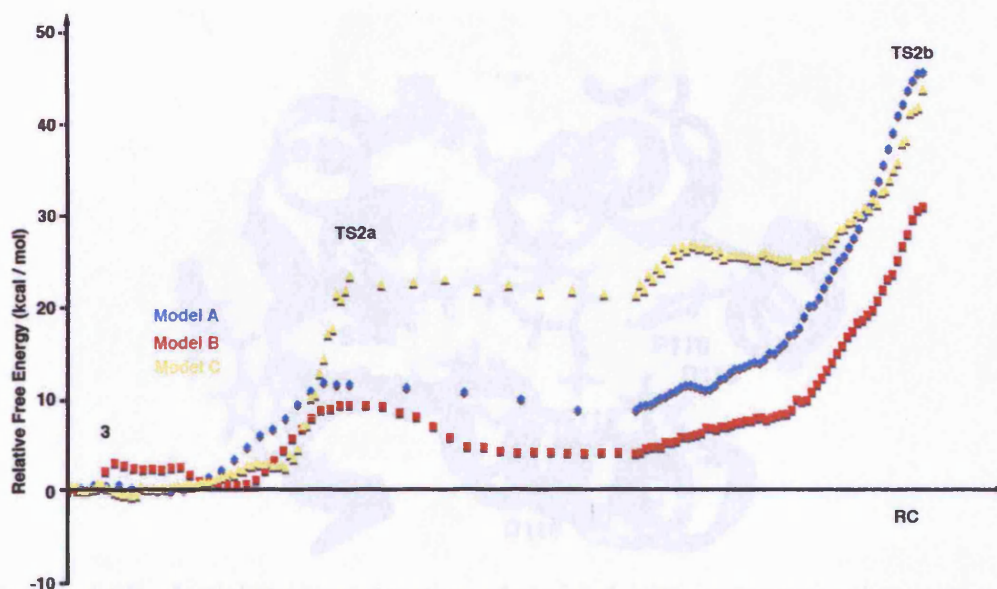
Figure 4.36: *Combined free energy profile for the conversion of germacryl cation to* ***TS2b*** *via* ***TS2a*** *in a water mediated proton transfer mechanism*

4.5 Lys 206 as a possible active site acid.

In order to investigate the mechanism of aristolochene biosynthesis in which it has been suggested that Lys 206 is the acid involved in the protonation of neutral germacrene A, modifications were made to models A, B and C described in Section 4.2.1.

In order to model germacrene A within the active site, a proton from C12 of germacryl cation in each of the models was removed from the raw pdb file. Following this the setup, solvation, equilibration and QM/MM dynamics procedure described in Section 4.2.4 was performed in which the QM region was defined as germacrene A and the section of Lys 206 as shown in Figure 4.4. The interface between QM and MM region was described using the GHO method. [142] The initial starting structures used are included as Figures 4.37, 4.38 and 4.39 respectively.
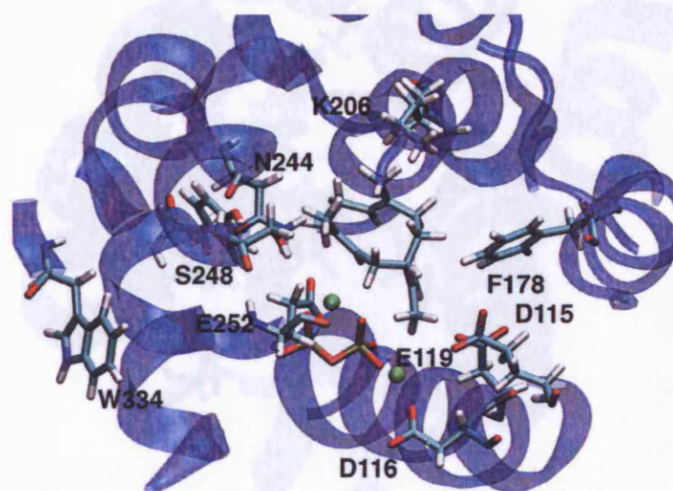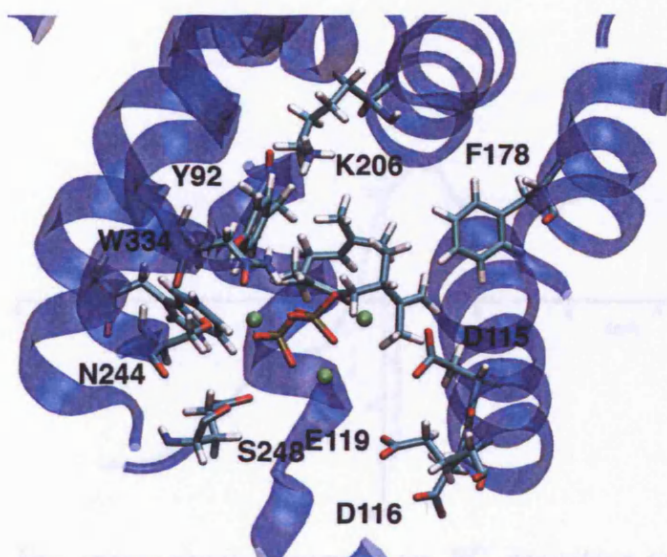
Figure 4.37: *Initial starting structure of model A with germacrene A in active site - Lys 206 acting as active site acid.*
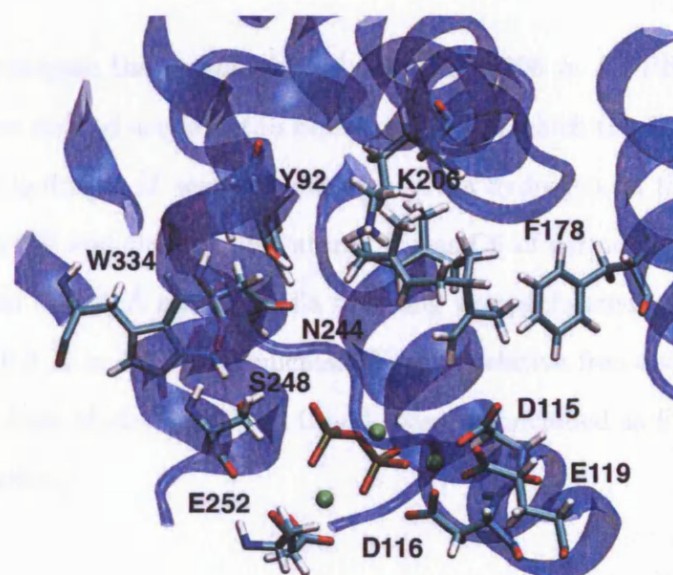
Figure 4.38: *Initial starting structure of model B with germacrene A in active site - Lys 206 acting as active site acid.*
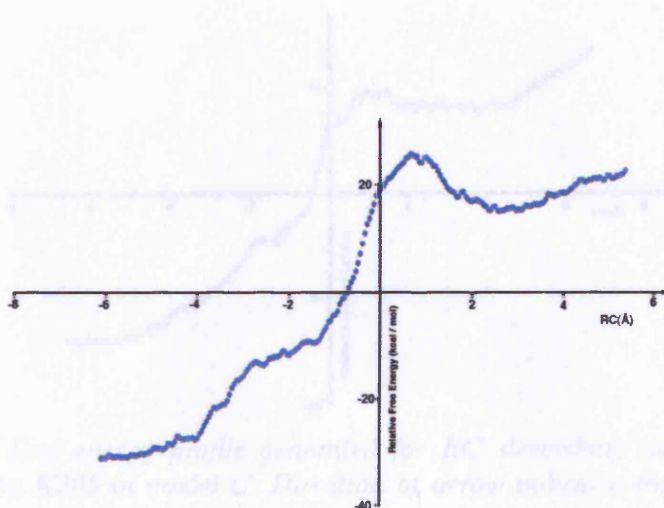
Figure 4.39: *Initial starting structure of model C with germacrene A in active site - Lys 206 acting as active site acid.*

4.5.1 Protonation of C6 by Lys 206

In order to investigate the protonation of C6 by Lys 206 in AS-PR, a reaction coordinate RC was defined according to Equation 4.3.1 in which the donor atom D and the transferred hydrogen H were the nitrogen and a hydrogen on the NH_3 group of protonated Lys 206 and the acceptor atom, A, was C6 of germacrene A. The initial value of RC was *ca.* 6.0 Å and umbrella sampling was performed with values of RC from -6.0 to +6.0 Å in 0.1 Å increments. Plots of relative free energy against RC were produced from Models A, B and C and these are included as Figures 4.40, 4.41 and 4.42 respectively.

Figure 4.40: *Free energy profile generated for RC describing the protonation of germacrene A by K206 in model A. Direction of arrow indicates reactant to product transition*
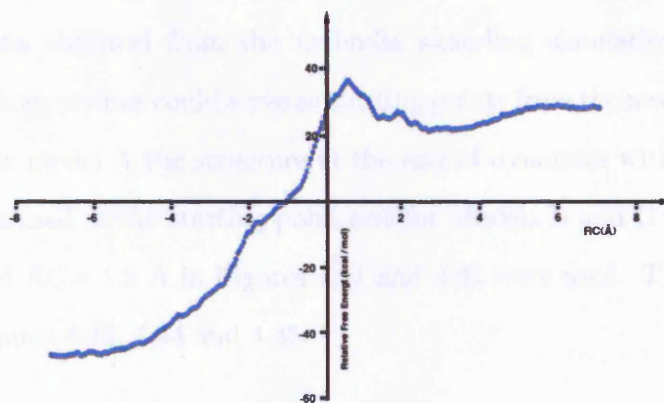
Figure 4.41: *Free energy profile generated for RC describing the protonation of germacrene A by K206 in model B. Direction of arrow indicates reactant to product transition*

Figure 4.42: *Free energy profile generated for RC describing the protonation of germacrene A by K206 in model C. Direction of arrow indicates reactant to product transition*

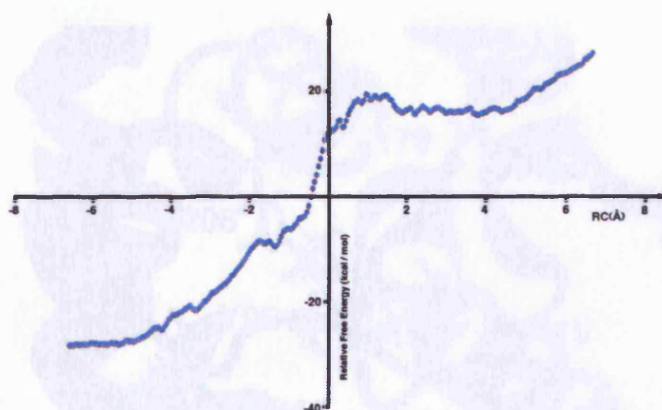## 4.5.2 Ring closure following protonation of C6 by Lys 206

Following protonation of C6 by Lys 206, and in common with all other proposed mechanisms, the next step in the reaction mechanism is the formation of eudesmane cation **28** from intermediate **64** *via* **TS3**. Visual inspection of the structures represented by the free data obtained from the umbrella sampling simulations were used to determine which structures could serve as starting points from the next set of QM/MM simulations. For model A the structure at the end of dynamics with $RC = 2.4$ Å in Figure 4.40 was used as the starting point and for Models B and C the structures at $RC = 0.6$ Å and $RC = 1.9$ Å in Figures 4.41 and 4.42 were used. The structures are included as Figures 4.43, 4.44 and 4.45.

A new reaction coordinate RC was defined as the distance between C2 and C7 in the structure approximating **64**. The initial value of this RC was *ca.* 3.0 Å in Models A and C and *ca.* 2.7 Å in model B. The values 3.0, 3.0 and 2.7 Å were used as the starting points for umbrella sampling, the value of RC was decreased in 0.1 Å
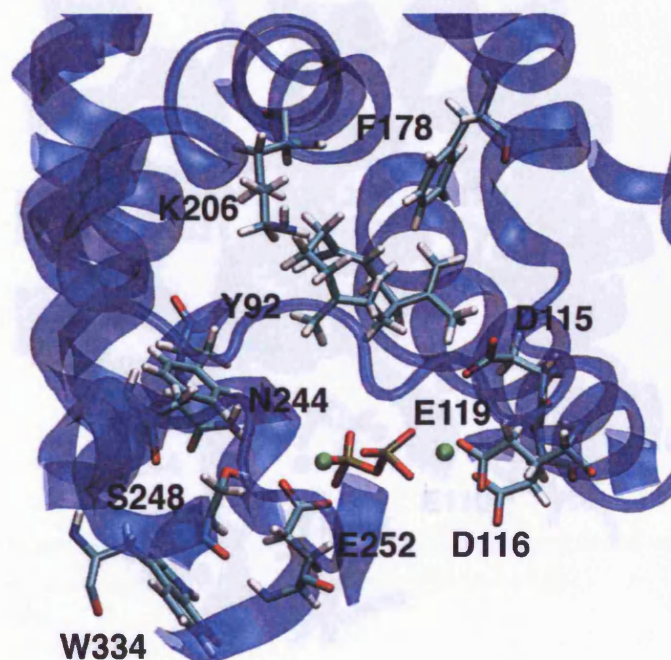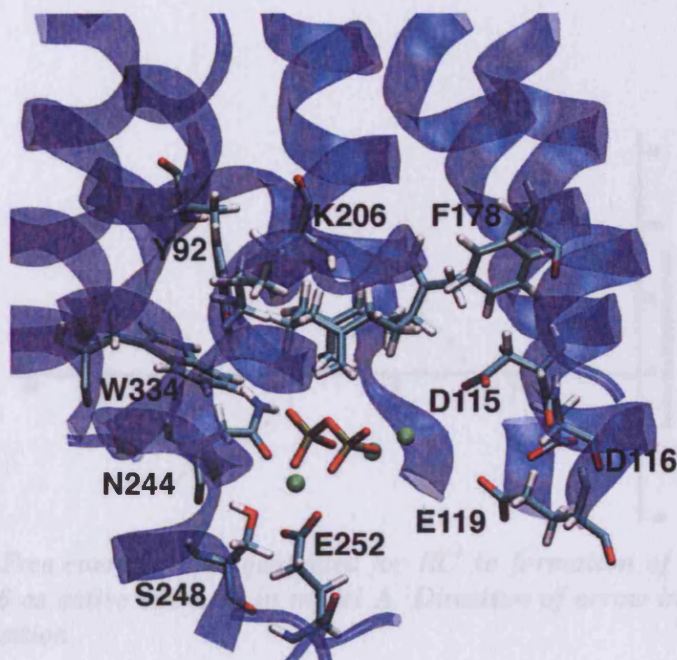
Figure 4.43: *Structure of model A used as starting point for sampling simulation in Section 4.5.2 - RC = 2.4 Å*
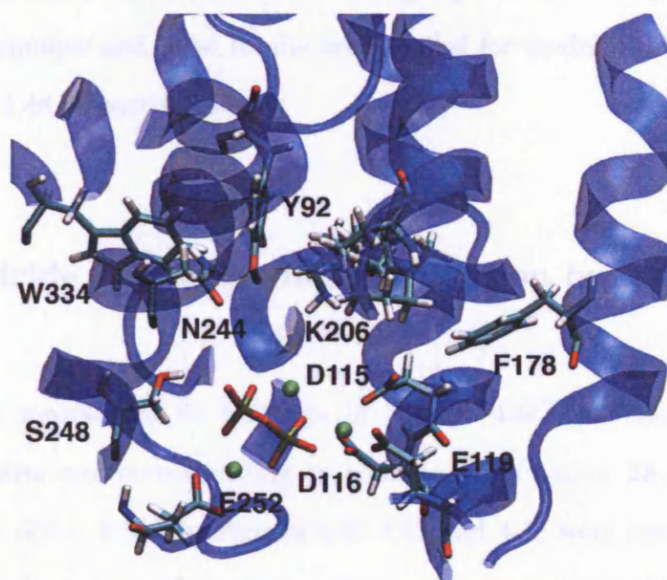
**Figure 4.44:** *Structure of model B used as starting point for sampling simulation in Section 4.5.2 - RC = 0.6 Å*

Figure 4.45: *Structure of model B used as starting point for sampling simulation in Section 4.5.2 - RC =1.9 Å*

increments until $RC = 1.0$ Å. Plots of free energy against RC were produced by using the WHAM technique and these results are included for model A, B and C as Figure 4.46, 4.47 and 4.48 respectively.

4.5.3   Hydride shift following protonation by Lys 206

The structures represented by the data in Figures 4.46 4.47 and 4.48 were used to identify a structure corresponding to to eudesmane cation **28**. The structures represented by $RC = 1.5$ Å in Figures 4.46 4.47 and 4.48 were used as the starting points for the following set of simulations designed to investigate the formation of **65** from **28** *via* **TS4**. The structures represented after dynamics by this value of RC in
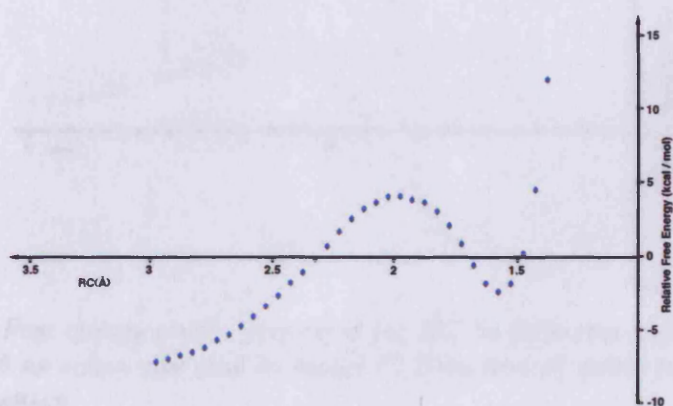
Figure 4.46: *Free energy profile generated for RC to formation of 28 from 64 via TS3 with K206 as active site acid in model A. Direction of arrow indicates reactant to product transition*
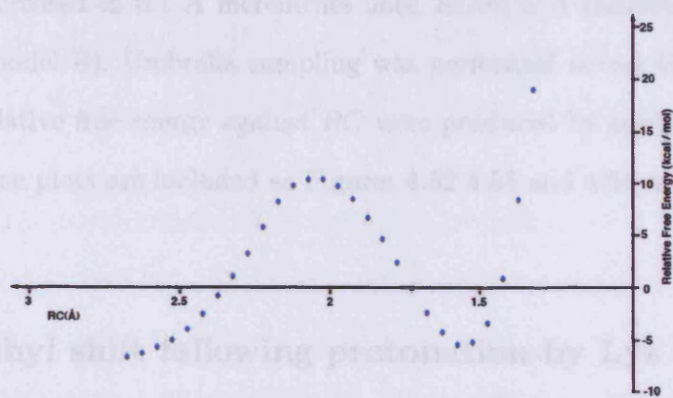
Figure 4.47: *Free energy profile generated for RC to formation of 28 from 64 via TS3 with K206 as active site acid in model B. Direction of arrow indicates reactant to product transition*
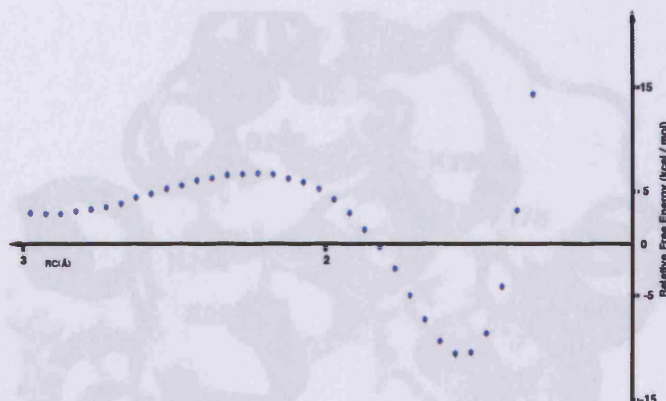
Figure 4.48: *Free energy profile generated for RC to formation of **28** from **64** via **TS3** with K206 as active site acid in model C. Direction of arrow indicates reactant to product transition*

Section 4.5.2 are included as Figures 4.49, 4.50 and 4.51.

A reaction coordinate RC was defined according to Equation 4.3.1 where D was C2, A was C3 and the migrating proton H was on the axial position of C2. The initial value of RC was *ca.* -1.0 Å in model A and C and *ca.* -0.9 Å in model B. These values of RC were increased in 0.1 Å increments until RC=1.0 Å (models A and C) and RC =1.5 Å (model B). Umbrella sampling was performed across this range of RC and plots of relative free energy against RC were produced by applying the WHAM equations. These plots are included as Figures 4.52 4.53 and 4.54 respectively.
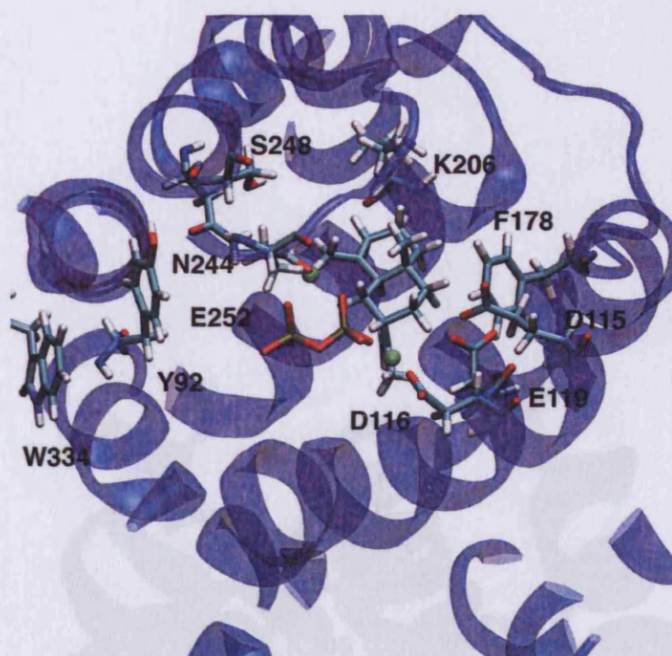
4.5.4 Methyl shift following protonation by Lys 206

The structures represented by the data in Figures 4.52, 4.53 and 4.54 were used to identify a structure corresponding to intermediate **65**. This was then used to investigate the next step in the proposed mechanism: the formation of intermediate
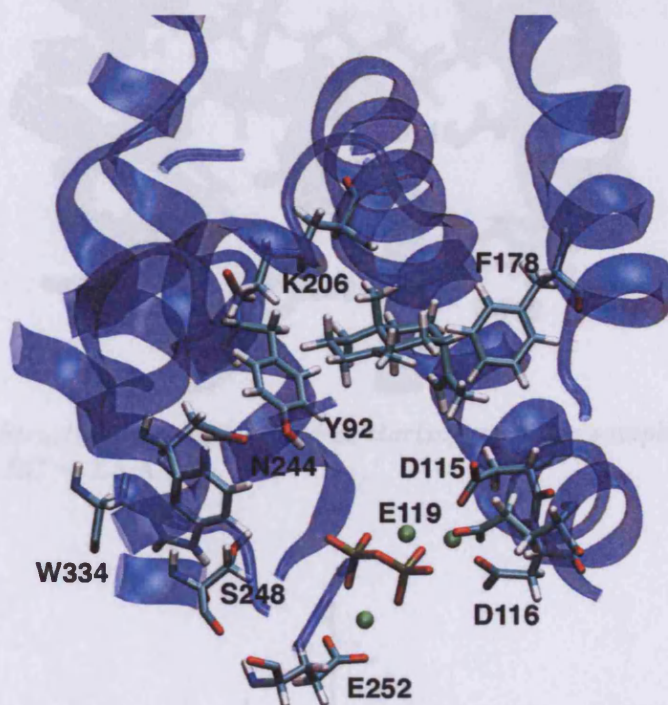
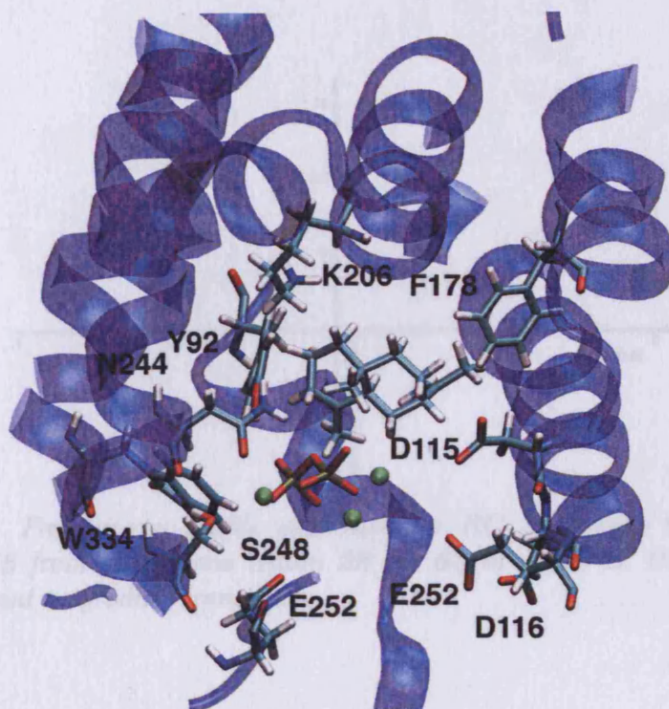**Figure 4.49:** *Structure of model A used as starting point for sampling simulation in Section 4.5.2 - RC = 1.5 Å*

66 *via* **TS5**. The structure represented by $RC = 0.8$ Å in Figures 4.52 and 4.54 and $RC = 1.0$ Å in Figure 4.53 were used as the starting point for the next set of simulations in models A,C and B respectively. The structures represented by these value of RC in section 4.5.3 are included as Figures 4.55, 4.56 and 4.57

A new reaction coordinate was defined according to Equation 4.3.1 in which D was C7, A was C2 and H was C14. The initial value of RC for model A was *ca.* -0.8 Å and -0.9 Å for models B and C. Umbrella sampling was applied as previously described and in the case of model A the value of RC was increased for -0.8 Å to 0.9 Å in 0.1 Å increments. The corresponding ranges for Models B and C were from -0.9 Å, 1.1 and 1.0 Å respectively. Following umbrella sampling, WHAM was performed and the resulting data were used to generate plots of relative free energy against RC, included here as Figures 4.58, 4.59 and 4.60 for models A, B and C respectively.

Figure 4.50: *Structure of model A used as starting point for sampling simulation in Section 4.5.2 - RC = 1.5 Å*

Figure 4.51: *Structure of model A used as starting point for sampling simulation in Section 4.5.2 - RC = 1.5 Å*
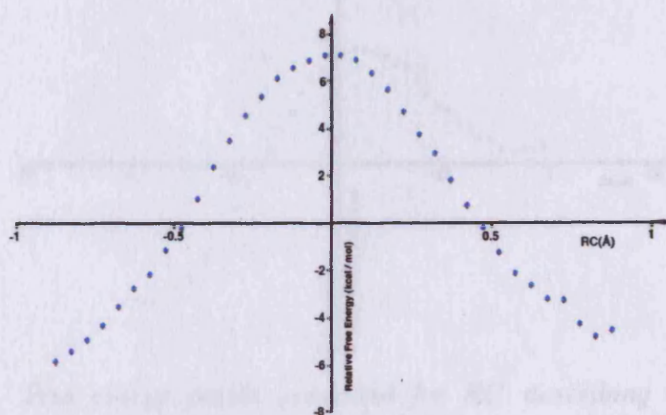
Figure 4.52: *Free energy profile generated for RC describing the formation of intermediate **65** from eudesmane cation **28** via **64** in model A. Direction of arrow indicates reactant to product transition*

Figure 4.53: *Free energy profile generated for RC describing the formation of intermediate **65** from eudesmane cation **28** via **64** in model B. Direction of arrow indicates reactant to product transition*

Figure 4.54: *Free energy profile generated for RC describing the formation of intermediate **65** from eudesmane cation **28** via **64** in model C. Direction of arrow indicates reactant to product transition*

**Figure 4.55:** *Structure of model A used as starting point for sampling simulation in Section 4.5.4 - RC = 0.8 Å*

Figure 4.56: *Structure of model B used as starting point for sampling simulation in Section 4.5.4 - RC = 1.0 Å*
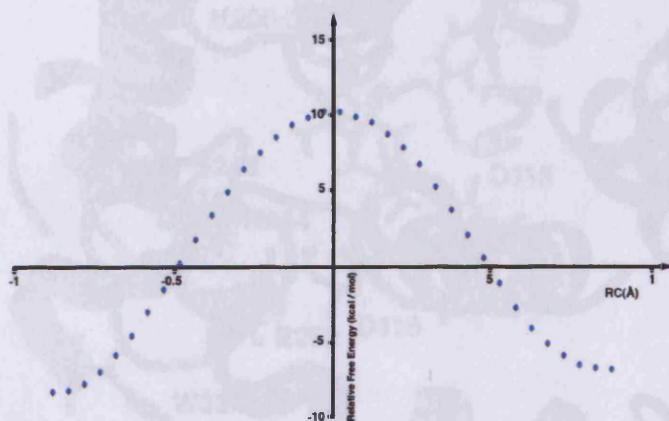
Figure 4.57: *Structure of model C used as starting point for sampling simulation in Section 4.5.4 - RC = 1.0 Å*

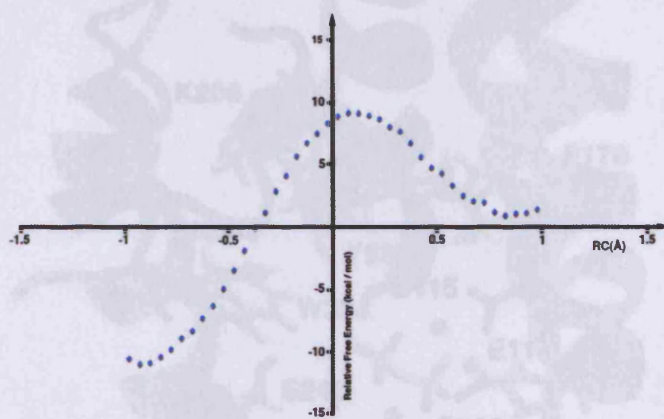**Figure 4.58:** *Free energy profile generated for RC describing the formation of intermediate **66** from **65** via **TS5** in model A. Direction of arrow indicates reactant to product transition*

Figure 4.59: *Free energy profile generated for RC describing the formation of intermediate **66** from **65** via **TS5** in model B. Direction of arrow indicates reactant to product transition*
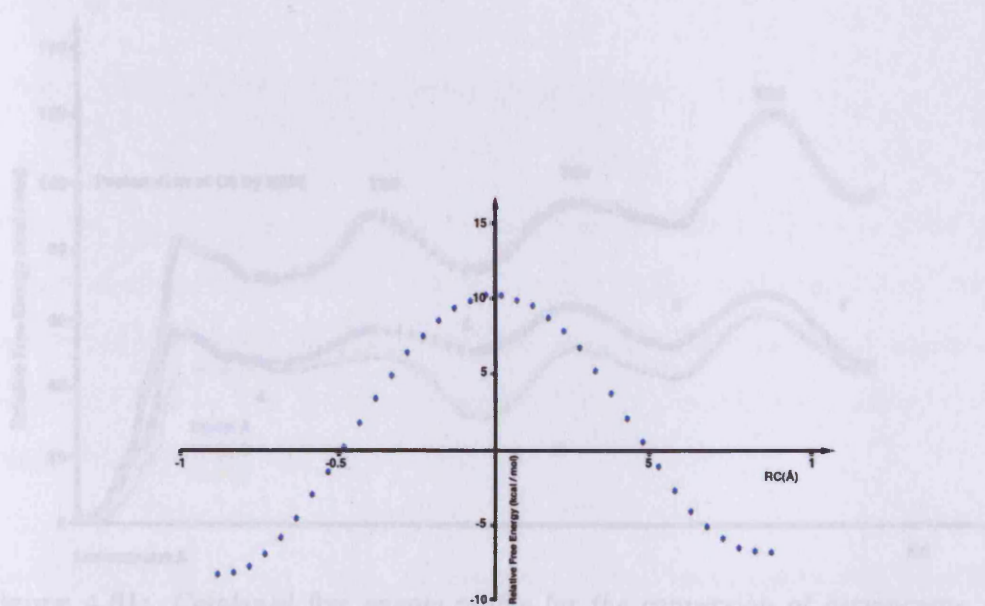
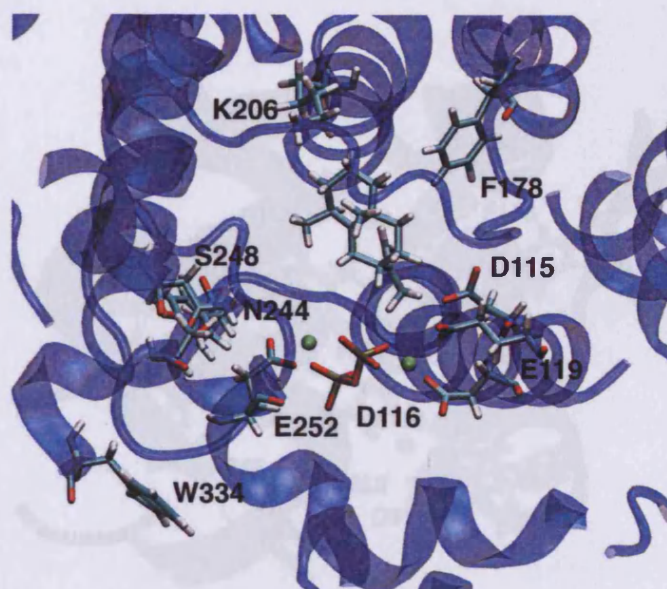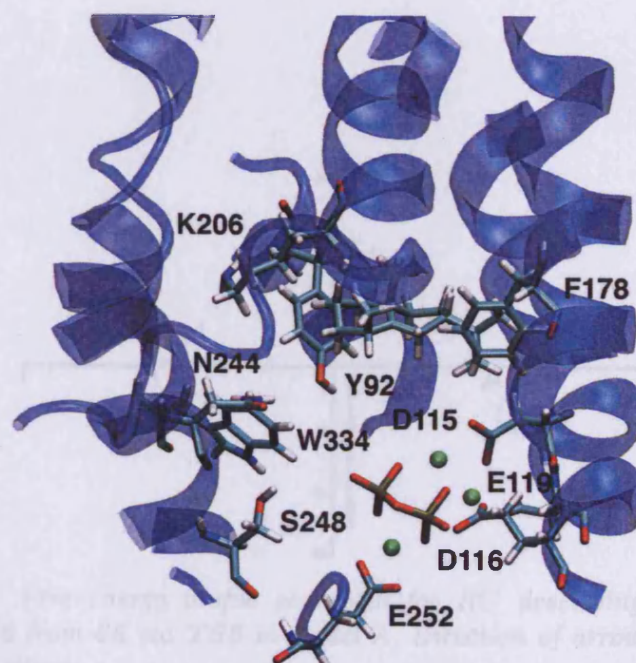

Figure 4.60: *Free energy profile generated for RC describing the formation of intermediate **66** from **65** via **TS5** in model C. Direction of arrow indicates reactant to product transition*

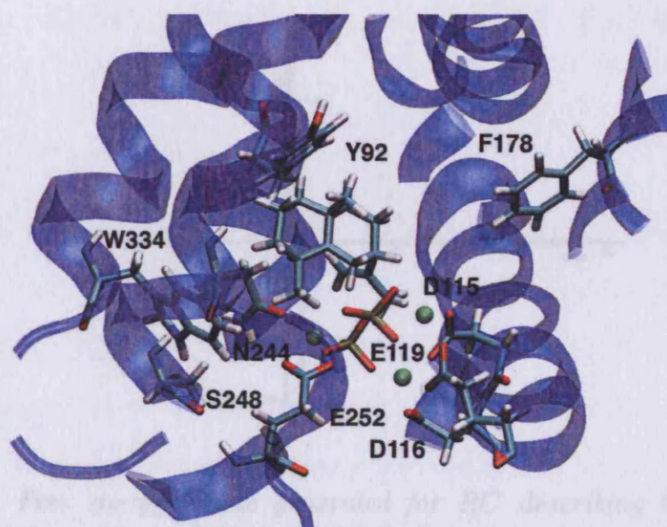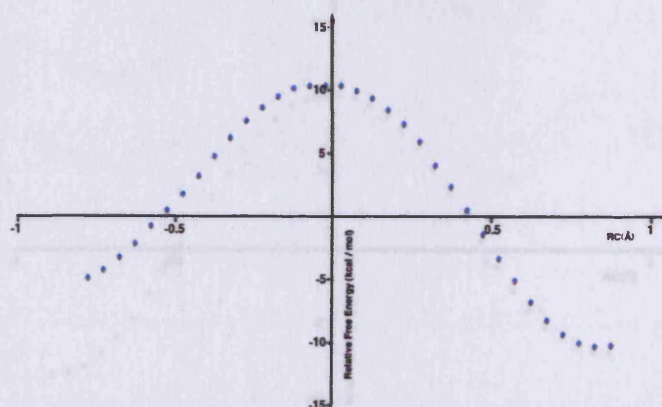### 4.5.5   Combined free energy profile for the conversion of germacrene A to aristolochene in which Lys 206 acts as active site acid

The free energy data obtained from the simulations described in Section 4.5 were combined to produce a complete free energy profile for the conversion of germacrene A to **66** in which Lys 206 acts as an active site acid. Two plots were produced, one in which the energy of the enzyme-substrate complex containing germacrene A was defined as the zero point of free energy, and another in which the energy of the enzyme-substrate complex containing eudesmane cation was defined as the zero point of free energy. These plots are included as Figure 4.61 and Figure 4.62 respectively.
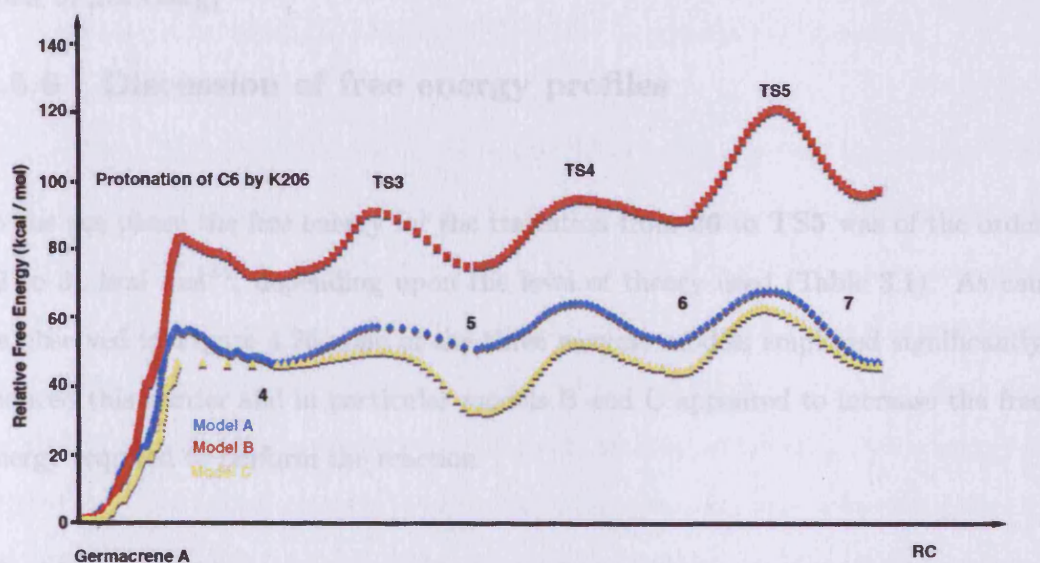
Figure 4.61: *Combined free energy profile for the conversion of germacrene A to aristolochene* - energy of germacrene A-enzyme complex set as arbitrary zero point of free energy
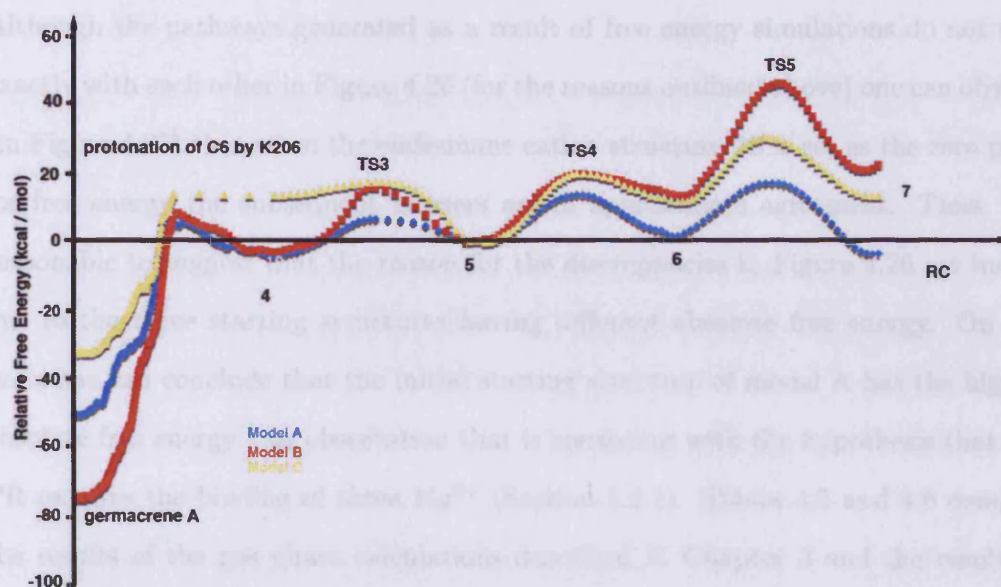
Figure 4.62: *Combined free energy profile for the conversion of germacrene A to aristolochene* - energy of eudesmane cation-enzyme complex set as arbitrary zero point of free energy

4.5.6   Discussion of free energy profiles

In the gas phase the free energy for the transition from **26** to **TS5** was of the order 22 to 31 kcal mol^{-1}, depending upon the level of theory used (Table 3.1). As can be observed in Figure 4.26 none of the three enzyme models employed significantly reduced this barrier and in particular models B and C appeared to increase the free energy required to perform the reaction.

The highest total barrier occurs with model C. In all cases the barrier from **26** to **TS2** is the highest total barrier. The difficulty in comparing the different docked structures directly is that the relative free energy of each starting structure is not known.

Although the pathways generated as a result of free energy simulations do not map exactly with each other in Figure 4.26 (for the reasons outlined above) one can observe (in Figure 4.27) that when the eudesmane cation structure **28** is set as the zero point for free energy the subsequent barriers are in approximate agreement. Thus, it is reasonable to suggest that the reason for the discrepancies in Figure 4.26 are indeed due to the three starting structures having different absolute free energy. On this basis one can conclude that the initial starting structure of model A has the highest absolute free energy - an observation that is consistent with the hypothesis that AS-PR requires the binding of three Mg^{2+} (Section 1.2.1). Tables 4.5 and 4.6 compare the results of the gas phase calculations described in Chapter 3 and the results of QM/MM simulations described in this chapter. Table 4.6 demonstrates that when the free energy of eudesmane cation is set as the zero point of free energy there is a good agreement between gas phase and QM/MM enzymatic calculations.

In any case, none of the docked structures lowers the energy for the formation of **TS2** from germacryl cation **26**, compared to the gas-phase reaction. The role of enzymes is to lower the free energy of activation for the formation of transition states compared to the equivalent reaction in solvent [107] and on the basis of these computational results (and assuming that the gas phase activations energy are higher than the equivalent reactions in the aqueous environment) there is no lowering of free energy of activation, implying no advantage is afforded to this route in the enzyme compared to the uncatalysed. This is in agreement with the experimental findings which seemingly rule out this mechanism *in vivo.* [92] However, the role of AS-PR is not solely to increase the rate of reaction compared to the uncatalysed reaction - it is also the chaperone reactive carbocations down a specific pathway and prevent the formation of other sesquiterpene products. [32, 54]

Table 4.5: Free Energy Comparisons - $\Delta G^\circ{}_{298}$ of intermediates and transition states in gas phase and QM/MM models. All values, in kcal mol^{-1}, are relative to germacryl cation, **26**.

Method	26	TS2	28	TS4	65	TS5	66
AM1	0.0	28.4	-18.3	-5.3	-17.2	7.6	-14.2
PM3	0.0	19.1	-18.0	-8.0	-15.6	6.1	-11.8
MPWPW1	0.0	25.6	-18.5	-25.8	-26.1	-22.0	-24.6
MPWB1K	0.0	26.1	-32.6	-28.7	-28.4	-8.0	-28.2
QM/MM (model A)	0.0	31.1	-11.8	0.6	-7.7	19.5	-1.4
QM/MM (model B)	0.0	38.3	11.4	25.6	19.6	44.5	20.2
QM/MM (model C)	0.0	51.3	17.2	30.5	25.0	52.9	35.3

Table 4.6: Free Energy Comparisons - $\Delta G^\circ{}_{298}$ of intermediates and transition states in gas phase and QM/MM models. All values, in kcal mol^{-1}, are relative to eudesmane cation, **28**.

Method	26	TS2	28	TS4	65	TS5	66
AM1	18.3	46.7	0.0	12.5	1.1	26.0	4.2
PM3	18.0	37.0	0.0	10.0	2.3	24.0	6.1
MPWPW1	28.5	54.2	0.0	2.7	2.5	6.6	4.0
MPWB1K	32.6	58.7	0.0	3.9	4.1	24.6	4.3
QM/MM (model A)	11.8	42.9	0.0	12.5	4.1	31.4	10.4
QM/MM (model B)	-11.4	27.0	0.0	14.2	8.2	33.1	8.8
QM/MM (model C)	-17.2	34.1	0.0	13.3	7.8	35.7	18.1

Irrespective of the identity of the proton source required in AS-PR catalysis, there are several other features of interest. Gas phase studies, as discussed elsewhere, were unable to isolate a structure corresponding to either **64** or **TS3**, despite repeated attempts. It has been suggested that the fusing of the ten-membered ring to form subsequent bicyclic intermediates is a concerted process requiring little or no free energy of activation. Figure 4.11, Figure 4.12 and Figure 4.13 show that for all three models the total barrier to the formation of **28** is *ca.* 2 kcal mol^{-1}, and that the formation of eudesmane cation is highly endergonic, therefore kinetically a very fast step, and thermodynamically highly favourable. These computational results demonstrate that formation of eudesmane cation is a concerted process similar to the concerted processes described in work by the Tantillo group and described in Section 2.11.

Following the formation of **28**, the magnitudes of the barriers resulting from QM/MM free energy simulations and the results of the gas phase calculation discussed previously are in excellent agreement. The general conclusion is that AS-PR does not play a specific role in the catalysis of subsequent reaction steps, and that these take place rapidly (as judged by their low calculated barrier heights). The role of the enzyme is therefore to provide a proton source to protonate the C6-C7 double bond, after which the reaction proceeds without any covalent interactions until the final deprotonation step. This is in agreement with the conclusions of previous researchers and consistent with the fact that germacrene A has been detected in the product mixture resulting from incubation of AS-PR and FPP. [176]

In the case of water mediated proton transfer (Section 4.4) the free energy for the deprotonation of germacryl cation to form germacrene A calculated using the combined

QM/MM method was *ca.* 12, 19 and 9 kcal mol^{-1} in models A, B and C respectively. This compares with a value of 19.8 kcal mol^{-1} calculated in the gas phase (Table 3.8). This is an excellent agreement between gas and condensed phase value, but again, there is no evidence of the lowering of free energy that would suggest that AS-PR catalyses the formation of aristolochene *via* a water mediated proton transfer reaction.

The subsequent step in which a germacrene A-hydroxonium ion complex protonates the C6 C7 double bond *i.e.* the route to aristolochene *via* **TS2b** had a barrier height for the formation formation of **TS2b** of *ca.* 46, 44 and 31 kcal mol^{-1} for models A, B and C respectively (Figure 4.36). This compares with values of 16.1 to 31.0 calculated in the gas phase (Table 4.35). Therefore there appears to be no lowering of the free energy barrier in enzyme compared to the gas phase, suggesting that this is not a likely reaction mechanism *in vivo*. The data displayed in Figure 4.36 also show that the enzyme model with the highest barrier to formation of **TS2b** was model A. In a similar fashion to the discussion of the direct intramolecular proton transfer hypothesis in the preceding paragraphs the results of the simulations described in Section 4.4 also demonstrate that the presence of a third magnesium ion lowers the overall free energy for reaction in AS-PR - again there is a more detailed discussion at Section 4.6.

Investigating the role of Lys 206 as a potential active site acid is somewhat more challenging as there is no comparable gas phase calculation - it would not be a good comparison to treat the system in the gas phase as consisting of an isolated lysine residue and germacrene A - although previous researchers have attempted to use ammonia as a model for lysine (acting as a base) in the gas phase. [154] However the

results contained in Section 4.5 have some large free energy barriers which suggest the Lys 206 does not, in isolation, act as the active site acid capable of protonating neutral germacrene A in AS-PR catalysis. The absolute magnitude of the free energy calculated to protonate germacrene A from the simulations outlined in Section 4.5.1 was *ca.* 82, 55 and 46 kcal mol^{-1} for models A, B and C (Figure 4.61) - large values which show great variation between the models employed, and certainly 82 kcal mol^{-1} is an extremely endergonic process to be catalysed by an enzyme, especially when one considers the low inherent gas phase barrier calculated for this reaction.

However Figure 4.62 demonstrates, in a manner similar to the direct intramolecular proton transfer case that if eudesmane cation is set as the zero point of free energy then the barrier heights to the formation of the structures corresponding to the transition states and intermediates following **28** are in broad agreement with the results obtained in the gas phase (Table 4.7) and in this case the QM/MM model A results in fact offer the closest agreement with gas phase data, and in particular Models B and C tend to be in greatest disagreement with the barrier height to formation of **65**, **TS5** and **66**, although the reason for this is not clear.

Studying the QM/MM results included in this chapter, and the gas phase data from Chapter 3 it is possible to conclude that AS-PR may actually have very little influence over the formation of aristolochene following formation of eudesmane cation. The evidence for this is reflected in the observation that, in general there is good agreement between the free energy barriers in the gas phase and as calculated in the QM/MM simulations, and in the experimental results which have not been able to 'trap' selinenes in product mixtures to any great extent, either in WT AS-PR or in any of the numerous mutagenesis experiments which have been performed (Section

Table 4.7: Free Energy Comparisons - ΔG°_{298} of intermediates and transition states in gas phase and QM/MM models. All values, in kcal mol^{-1}, are relative to eudesmane cation, **28**.

Method	28	TS4	65	TS5	66
AM1	0.0	12.5	1.1	26.0	4.2
PM3	0.0	10.0	2.3	24.0	6.1
MPWPW1	0.0	2.7	2.5	6.6	4.0
MPWB1K	0.0	3.9	4.1	24.6	4.3
QM/MM (model A)	0.0	13.5	1.6	16.9	-3.7
QM/MM (model B)	0.0	19.5	13.3	45.9	20.7
QM/MM (model C)	0.0	19.7	11.5	29.3	13.0

1.3.2).

As for the role of Lys 206 as the active site acid involved in protonating the C6 C7 double bond, the QM/MM results outlined here suggest that it is unlikely that Lys 206 performs this role alone - the energy barrier to protonation (46 to 82 kcal mol^{-1}) is large, even when one considers that the reaction is, essentially, the protonation of an alkene.

4.6 Results - enzyme influence on reaction pathway

As discussed in Chapter 1 a number of amino acid residues have been studied using site directed mutagenesis in order to investigate their role aristolochene biosynthesis. These experiments offer a static view of the reaction *i.e.* one creates a novel variant

of AS-PR then calculates the resultant change in kinetic parameters and observes the new product distribution then uses these results to improve, create or refine hypotheses about the role of that amino acid.

It has been suggested that the aromatic side chains of Tyr 92, Phe 178 and Trp 334 [72, 73] serve to stabilise the positive charges which are built up on C7, C3 and C2 during the reaction mechanism - that these carbon centres become positively charged does not vary across the reaction mechanisms which are proposed in the literature and in this work. This is not without precedent as π electron stabilisation of terpene synthesis has been demonstrated in an elegant series of experiments by the Hoshino group. [191]

The MD simulations performed in this work make it possible to study the motion of individual atoms relative to any other atom over the entire time-scale of the dynamics simulation - as is outlined in Chapter 2 the trajectories recorded during an MD simulation may be used to extract this useful information - it is not even required to specify which details are of interest before a simulations is started since the trajectory records all the relevant information.

Therefore, the QM/MM simulations described in Section 4.3 were also used to measure several key distances over the course of the simulation. The distances measured were those between C7, C3, C2 and the amino acid residues Tyr 92, Phe 178 and Trp 334. The trajectories were loaded into the visualisation software VMD [182] from the start to finish values of RC for each simulation as described in Section 4.3. VMD and its associated MD analysis packages [183–189] were used to measure each of these distances for the reaction mechanisms (direct intramolecular transfer or proton

transfer by way of Lys 206) studied in this work. Plots of the relevant distances for each model are included in Appendix A.

4.6.1 Interaction between Asp 115 and magnesium ion during proton transfer by Lys 206

Models A and C showed a much longer mean distance between Mg_A and Asp 115 compared with model B (ca. 4 Å compared with *ca.* 1.9 Å) and this is different to the results in Section 4.3.8, with the approximate mean distance the same across all three models. None of models A and B and C exhibit the syn-syn bidentate co-ordination to Mg_C exhibited in AS-AT, which may in part explain the high energy barrier obtained in Section 4.5.5, because incorrect binding makes the efficient transfer of a proton onto the C6-C7 double bond more difficult.

4.6.2 Interaction between Asp 116 and magnesium ion during proton transfer by Lys 206

The interaction between Mg_A and Asp 116 was strongest in models A and B (mean distance *ca.* 1.9 Å in Figures B.22 and B.22) whereas in model C the interaction was not evident, but the Mg_A-D116 distance did decrease from *ca.* 11 Å to 7 Å during the course of the simulation.

4.6.3 Interaction between Glu 119 and magnesium ion during proton transfer by Lys 206

The only model which displayed a strong interaction between Glu 119 and Mg_A was model B (Figure B.26), the binding was weakest in model A, and both model A and model C displayed a large significant change in Mg_A Glu 119 distance during the formation of eudesmane cation **28**.

4.6.4 Interaction between Asn 244 and magnesium ion during proton transfer by Lys 206

The relationship between Asn 244 and Mg_B in the Lys 206 protonation mirrors that one observed in the direct intramolecular proton transfer route (Section 4.3.11) in that neither models A nor B demonstrated a strong interaction between Asn 244 and Mg_B, but Asn 244 in model C does (mean distance $ca.$ 2 Å).

4.6.5 Interaction between Ser 248 and magnesium ion during proton transfer by Lys 206

Binding between Ser 248 and Mg_B in model C is present (mean distance $ca.$ 2 Å) but was not observed in models A or B. This is different from the results obtained in Section 4.3.12 in which none of the models showed any significant interaction between

S248 and Mg_B, hence this is the only evidence of S248 acting as a magnesium binding ligand in these simulations and therefore the first direct example of S248 acting in its proposed role in AS-PR.

4.6.6 Interaction between Glu 252 and magnesium ion during proton transfer by Lys 206

Figures B.34 to B.36 show the interaction between Glu 252 and Mg_B. In common with the results obtain in section 4.3.13 all three models showed a strong interaction between Mg_B and Glu 252 with a mean distance of *ca.* 2 Å in all three models.

4.6.7 Discussion of interaction between Phe 178, Trp 334 and Tyr 92 and C7, C3 and C2 in Lys 206

The results for the interaction between Phe 178, Trp 334 and Tyr 92 and C7, C3 and C2 are much less compelling and the less obvious nature of the interactions can go towards explaining the large free energy barriers associated with the Lys 206 protonation scheme, the results of which are described in Section 4.5.1. Of particular note is Trp 334 and model A - the distances between all three carbon centres studied and the side-chain are very large and in some cases increase during the simulation which is the opposite to what one would expect if Trp 334 were able to influence the course of the reaction with the mechanism in question.

Phe 178 does, however, appear to act in a similar fashion as described in Section 4.3.14 with Phe 178 moving closer to C7 in order to stabilise formation of eudesmane cation, although this is less pronounced in model C. No clear, consistent relationship between the Phe 178 C2 and C3 distances were observed and this is in agreement with the results for the direct proton transfer mechanism in Section 4.3.14.

Chapter 5

Summary and Outlook

5.1 Summary of findings

5.1.1 Comparison of free energy profiles and geometric data

The structures obtained from the gas phase QM and DFT simulations described in Chapter 3 and the structures which resulted from the combined QM/MM simulations in Chapter 4 are in agreement in that the structures are geometrically very similar (Tables 4.2, 4.3 and 4.4). These geometric similarities give confidence that the commentary regarding the free energy differences between intermediates and the structures labelled as transition states are valid. The discussion of the free energy differences is contained within Section 4.5.6 and these results add weight to the hypothesis that the formation of eudesmane cation **28** is a concerted process following

the protonation of the C6 C7 double bond. In addition the QM/MM results in Chapter 4 also agree with the gas phase results in Chapter 3 that there are noticeable barriers to the conversion of **28** to **65** (the hydride shift step) and from **65** to **66** (the methyl shift step). These observations are in contrast to previous hypotheses which have suggested that the biosynthesis of aristolochene by AS-PR is a fully concerted process from the protonation of the C6 C7 double bond until the completion of the reaction.

If AS-PR does not rely on a highly concerted mechanism from **28** to **66** how does the enzyme control the formation of aristolochene as its product? Yields in excess of 90% [71, 76, 79] are observed and, as described in Chapter 1.6, there are many possible side products, such as valencenes, selinenes and farnesenes (Figure 1.10) resulting from slight variations in the reaction mechanism at each stage. How AS-PR and the other sesquiterpene synthases control this precise sequence is a key question in the understanding of how the current terpenome evolved.

5.1.2 The role of the enzymatic environment on aristolochene biosynthesis

The purpose of performing the QM/MM simulations in Chapter 3 was, aside from the investigation of the absolute free energy changes for the reaction mechanisms, to gain insights into how those amino acid residues identified by other researchers working on AS-PR influenced the course of the reaction. The results of these simulations are included in Sections 4.6. These results went some way to confirming the conclusions

made in previous studies on AS-PR (as described in Section 1.3.2) but there were several contradictory results. Previously reported experimental results (Section 1.3.2) are also on occasion not in good agreement and this is a feature that has not been fully explored in the literature. Based on these observations it would seem that each step of aristolochene biosynthesis is not controlled by a single dominating interaction, but by the combined effect of the total enzyme environment with a few key residues having a larger effect than others.

For example, the disruption of the magnesium binding domain by the creation of the D115N, N244L and S248A/E252D proteins [75] produces an enzyme which is completely inactive. It appears, therefore, that these residues' role in magnesium binding is so crucial that drastic disruptions (in terms of functional group sterics and effective charge) are able to prevent catalysis. More modest disruption (*e.g.* the N244D protein) can still yield a functioning enzyme, albeit with a reduction of k_{kat}/K_M of approximately four orders of magnitude (Tables 1.1 and 1.3).

This hypothesis is an explanation for the fact that, AS-PR and AS-AT show little overall sequence homology but a number of key active site residues are conserved across both enzyme and this is consistent with Keasling and co-workers concept of 'plasticity residues' [192] which they have applied to another sesquiterpene cyclase, γ humelene synthase.

However, unlike *epi*-aristolochene synthase [85], the identity of the C6 C7 protonating acid is still unresolved. Despite their demonstrated plausibility in the gas phase neither the direct nor water mediated transfer reaction mechanisms has their free energy barriers lowered by the presence of enzyme in place of solvent (which would

be required for it to be consistent with current theories of enzymatic processes [107])
and more importantly detailed NMR experiments have seemingly ruled out both
direct intramolecular proton transfer and (at least in the form described in this work)
water mediated proton transfer. [92]

In *epi*-aristolochene synthase, germacrene A has been shown to be a true intermediate
and the role of Tyr 520 acting as the active site acid in C6 C7 protonation has also
been demonstrated. [85] The situation in AS-PR is much less clear; there is evidence
for the intermediacy of germacrene A, but it is not true to say it has been conclusively
identified as an intermediate. Nor has any single candidate proven itself as the active
site acid in the key protonation step. In Section 4.5 the possible role of Lys 206 was
investigated and it was found that Lys 206 is not, on its own, likely to perform this
role.

In total four possible proton sources have now been proposed and and rejected (Tyr
92, Lys 206, direct and water mediated proton transfer). Study of the active site of
AS-PR (and AS-AT) suggests that there are no other immediately viable candidates
able to act on their own. Indeed, it is remarkable that in *epi*-aristolochene synthase
the hydroxyl group of Tyr can seemingly act as an acid powerful enough to protonate
an olefinic bond, leading to the conclusion that these enzymes are able to alter the
electronic environment in order to make such chemically difficult reactions possible.

Given that there is no single, viable, candidate for the acid required in this key
step, it is possible that a complicated proton shuttle mechanism might exist, with
a proton from the enzyme surface (and ultimately solvent) being passed along from
the surface to the active site, with each proton transfer step maintaining helping to

maintain charge within the enzyme. If, for example, Lys 206 was to have one of its terminal (*i.e.* NH_3^+) protons abstracted by the C6 C7 bond whilst simultaneously being reprotonated by an adjacent amino acid capable of changing protonation state relatively easily and this proton source was ultimately replenished for external solvent then Lys 206 could conceivably act as the required acid. Experimental support (laboratory or computational) for such a mechanism might prove difficult, but the possibility must be considered.

5.1.3   Role of magnesium binding domains

Three models of AS-PR were created from the previously solved crystal structure of AS-PR. [30] The key difference in each model was the number and location of bound magnesium ions. Evidence from AS-AT [43,44] and *epi*-aristolochene synthase [85] suggests that three magnesium ions are required for catalysis, yet the enzyme may be able to bind substrate with just two magnesium ions present, a third being required to trigger the cyclisation cascade. [44] In addition the existence of 'closed' and 'open' forms of AS-AT hint that AS-PR may also be catalytically active with three magnesiums present and in a closed form.

In model C and with the direct intramolecular transfer mechanism (Section 4.3.8,4.3.9 and 4.3.10) an excellent agreement with the magnesium ion binding motifs described in AS-AT which gives confidence in the hypothesis that AS-PR is likely to bind to magnesium in a catalytic sequence analogous to the described in AS-AT. Furthermore, given that the magnesium binding regions are conserved across all known sesquiterpene cyclases there is every reason to suggest that this is a feature common across this

group of enzymes.

However, despite the observed differences in magnesium binding in the three models studied in Chapter 4 no common theme emerged regarding the free energy differences associated with the presence of a third magnesium. Therefore, given the validity of the magnesium binding model used (at least in model C) it is possible to conclude that the failure to fully approximate the as yet unobserved closed form of AS-PR is responsible for this weakness in the simulation results. Whilst this weakness must be acknowledged, the fact that the structures in the QM/MM simulation matched the gas phase structures (Section 4.3.7) and that excellent agreement with the proposed magnesium binding domains has been achieved makes it possible to conclude that the change in conformation from open to closed is key to making AS-PR a catalytically active enzyme.

5.2 Outlook

The results presented in this work are, to the best of the author's knowledge, the first attempts to study the dynamic behaviour of AS-PR computationally using the combined QM/MM MD approach to study the structure and dynamics of the enzyme and the free energy profile during the reaction mechanism. As such there are limited reference points with which to assess either the validity of the models used or the results obtained. In particular one must always exercise caution when considering data obtained directly or indirectly from an X-ray crystal structure starting point, for the reasons discussed in an insightful study by Garcia-Viloca *et al.* [193] in which

the different results obtained by QM/MM simulation depend entirely on the choice of initial starting structure made, the conclusions of which has broad implications for all workers who choose to use X-ray crystal data as the starting point for further simulation or modelling.

Therefore, in order to continue this work and further explore the mechanism of AS-PR more detailed structural studies are required. It is, in principle, possible to improve upon the resolution of X-ray crystallographic experiments or alternatively, and perhaps more interestingly, it may be possible to conduct extended time-scale MD experiments to study the dynamic motion of AS-PR and to investigate more fully the conformational changes which take AS-PR from apo-enzyme to its catalytic form.

A possible series of experiments could consist of multi-nanosecond MD simulations on both AS-PR and AS-AT designed to test the hypotheses presented regarding the reaction mechanism catalysed by these enzymes [30, 43, 44] might consist of the following. First, a study of the long term dynamic motion of the solvated enzyme in the absence of magnesium or substrate, the purpose of which is to understand the differences between the crystallographic structures and the conformations the enzymes adopt in a solvated environment.

Following this the binding of FPP within the active site of FPP can be simulated - the introduction of a magnesium ion (Mg_B) is proposed to be the first ion to bind at the NSE/DTE domain [44] and it is interesting to consider what effect, if any, this has on the conformation which FPP adopts within the active site. The binding of Mg_C is the final ion in the proposed sequence, and following this binding the conversion

of FPP to aristolochene can be studied in a manner similar to the described in the previous chapter.

However, the initial step (loss of diphosphate ion) was not studied in the previous chapter and a set of simulations and a simulation of the binding of FPP to AS could be used a starting point to investigate the energetic and conformational differences between the stepwise and concerted mechanisms for the initial cyclisation.

In performing an expanded QM/MM study of AS-PR more study is required to elucidate the exact mechanism by which AS-PR is able to protonate an alkene bond as is catalytically required; it fis fascinating how *epi*-aristolochene synthase can utilise Tyr 520 but AS-PR does not act in the same manner. Indeed, AS-PR, and probably AS-AT, both seem to actively avoid utilising intramolecular proton transfer reactions which avoid the need to quench and reform neutral intermediates.

The final step, deprotonation at $H8_{Si}$, forming aristolochene was not studied in this work, but it is in principle possible to investigate this step using the QM/MM method and this would lead to an enzyme structure in which aristolochene, diphosphate and magnesiums are bound. The release of products would complete the catalytic cycle and MD simulation could be used to investigate the process by which aristolochene is released from the active site and apo-enzyme is regenerated.

Is the mechanism catalysed by AS-PR fundamentally the same as AS-AT? If so, does it provide a general model for sesquiterpene biosynthesis *i.e.* can the same substrate binding and release mechanisms be applied across the range of sesquiterpenes. If QM/MM studies shed further light onto the mechanistic features of the sesquiterpene

synthases then it may be possible to fully understand how the few highly conserved regions of these enzymes are able to direct the cyclisation of FPP into such diverse products.

Such increased knowledge might help understand and predict mutations that convert an aristolochene synthase into a δ-cadinene synthase. If such a transformation is possible it would be major breakthrough in understanding the physical basis of enzyme catalysis and a crucial step in one of the major goals of modern chemical biology - the design of novel enzymes *in silico* and their use synthesising commercially or medically important biomolecules.

Bibliography

[1] Glasby, J.S. *Encyclopedia of Terpenoids* (Wiley, 1982).

[2] Cane, D.E. Isoprenoid biosynthesis - Stereochemistry of the cyclization of allylic pyrophosphates. *Acc. Chem. Res.*, **18**, (1985), 220–226.

[3] Cane, D.E. Enzymatic formation of sesquiterpenes. *Chem. Rev.*, **90**, (1990), 1089–1103.

[4] Ruzicka, L. Perspectivies of the biosynthesis and chemistry of the terpines - Review with 60 references. *Pure Appl. Chem.*, **6**, (1963), 493–523.

[5] Ruzicka, L. Faraday Lecture - History of the isoprene rule. *Proc. Chem. Soc. London*, **11**, (1959), 341–360.

[6] Ruzicka, L. The isoprene rule and the biogenesis of terpenic compounds. *Experientia*, **9**, (1953), 357–367.

[7] Stryer, L. *Biochemistry* (W.H. Freeman and Co., New York, 1995), 4th ed.

[8] Wright, L. & Cleland, M. Factors influencing incorporation of mevalonic acid into cholesterol by rat liver homogenates. *Proc. Soc. Exp. Biol. Med.*, **96**, (1957), 219–224.

[9] Tavormina, P. & Gibbs, M. The metabolism of β,γ-dihydroxy-β-methylvaleric acid by liver homogenates. *J. Am. Chem. Soc.*, **78**, (1956), 6210–6210.

[10] Ferguson, J. & Rudney, H. Biosynthesis of β-hydroxy-β-methylglutaryl CoA in yeast. 1. Identification and purification of the hydroxymethylglutaryl coenzyme condensing enzyme. *J. Biol. Chem.*, **234**, (1959), 1072–1075.

[11] Rudney, H. & Ferguson, J. Biosynthesis of β-hydroxy-β-methylglutaryl CoA in yeast. 2. Formation of hydroxymethylglutaryl CoA *via* the condensation of acetyl CoA and acetoacetyl CoA. *J. Biol. Chem.*, **234**, (1959), 1076–1080.

[12] Rilling, H., Tchen, T.T. & Bloch, K. On the mechanism of squalene biogenesis. *Proc. Natl. Acad. Sci. USA*, **44**, (1958), 167–173.

[13] Cornforth, J.W., Hunter, G.D. & Popjak, G. Distribution of acetate carbon in the ring-structure of cholesterol. *Biochem. J.*, **53**, (1953), 34–35.

[14] Cornforth, J.W., Hunter, G.D. & Popjak, G. Studies of cholesterol biosynthesis. 1. A new chemical degradation of cholesterol. *Biochem. J.*, **54**, (1953), 590–597.

[15] Cornforth, J.W., Hunter, G.D. & Popjak, G. The biosynthesis of cholesterol from acetate. *Arch. Biochem. Biophys.*, **42**, (1953), 481–482.

[16] Cornforth, J.W., Hunter, G.D. & Popjak, G. A new chemical degradation of cholesterol. *Biochem. J.*, **53**, (1953), 34.

[17] Cornforth, J.W., Hunter, G.D. & Popjak, G. Studies of cholesterol biosynthesis. 2. Distribution of acetate carbon in the ring structure. *Biochem. J.*, **54**, (1953), 597–601.

[18] Cornforth, J.W., Cornforth, R.H., Popjak, G. & GORE, I.Y. Studies on the biosynthesis of cholesterol. 5. Biosynthesis of squalene from d, l-3-hydroxy-3-methyl [2-^{14}C] pentano-5-lactone. *Biochem. J.*, **69**, (1958), 146–155.

[19] Stancu, C. & Sima, A. Statins: Mechanism of action and effects. *J. Cell. Mol. Med.*, **5**, (2001), 378–387.

[20] Slater, E.E. & MacDonald, J.S. Mechanism of action and biological profile of HMG CoA reductase inhibitors.A new therapeutic alternative. *Drugs*, **36**, (1988), 72–82.

[21] Eisenreich, W., Bacher, A., Arigoni, D. & Rohdich, F. Biosynthesis of isoprenoids *via* the non-mevalonate pathway. *Cell. Mol. Life. Sci.*, **61**, (2004), 1401–1426.

[22] Eisenreich, W., Menhard, B., Hylands, P.J., Zenk, M.H. & Bacher, A. Studies on the biosynthesis of taxol: The taxane carbon skeleton is not of mevalonoid origin. *Proc. Natl. Acad. Sci. USA*, **93**, (1996), 6431–6436.

[23] Eisenreich, W., Schwarz, M., Cartayrade, A., Arigoni, D., Zenk, M.H. & Bacher, A. The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms. *Chem Biol*, **5**, (1998), 221–33.

[24] Eisenreich, W., Rieder, C., Grammes, C., Hessler, G., Adam, K.P., Becker, H., Arigoni, D. & Bacher, A. Biosynthesis of a neo-*epi*-verrucosane diterpene in the liverwort *Fossombronia alaskana*. A retrobiosynthetic NMR study. *J. Biol. Chem.*, **274**, (1999), 36312–36320.

[25] Rieder, C., Strauss, G., Fuchs, G., Arigoni, D., Bacher, A. & Eisenreich, W. Biosynthesis of the diterpene verrucosan-2-β-ol in the phototrophic eubacterium

Chloroflexus aurantiacus. a retrobiosynthetic NMR study. *J. Biol. Chem.*, **273**, (1998), 18099–18108.

[26] Klapa, M., Park, S., Sinskey, A. & Stephanopoulos, G. Metabolite and isotopomer balancing in the analysis of metabolic cycles: 1. Theory. *Biotechnol. Bioeng.*, **62**, (1999), 375–391.

[27] Park, S.M., Klapa, M.I., Sinskey, A.J. & Stephanopoulos, G. Metabolite and isotopomer balancing in the analysis of metabolic cycles: 2. Applications. *Biotechnol. Bioeng.*, **62**, (1999), 392–401.

[28] Schmidt, K., Marx, A., de Graaf, A., Wiechert, W., Sahm, H., Nielsen, J. & Villadsen, J. [13]C tracer experiments and metabolite balancing for metabolic flux analysis: Comparing two approaches. *Biotechnol. Bioeng.*, **58**, (1998), 254–257.

[29] Szyperski, T. Biosynthetically directed fractional ^{13}C labeling of proteinogenic amino acids. an efficient analytical tool to investigate intermediary metabolism. *Eur. J. Biochem.*, **232**, (1995), 433–448.

[30] Caruthers, J.M., Kang, I., Rynkiewicz, M.J., Cane, D.E. & Christianson, D.W. Crystal structure determination of aristolochene synthase from the blue cheese mold, *Penicillium roqueforti. J. Biol. Chem.*, **275**, (2000), 25533–25539.

[31] Wendt, K. & Schulz, G. Isoprenoid biosynthesis: Manifold chemistry catalyzed by similar enzymes. *Structure*, **6**, (1998), 127–133.

[32] Christianson, D.W. Unearthing the roots of the terpenome. *Curr. Opin. Chem. Biol.*, **12**, (2008), 141–150.

[33] Wendt, K., Schulz, G., Corey, E. & Liu, D. Enzyme mechanisms for polycyclic triterpene formation. *Angew. Chem. Int. Ed. Engl.*, **39**, (2000), 2812–2833.

[34] Lesburg, C.A., Zhai, G.Z., Cane, D.E. & Christianson, D.W. Crystal structure of pentalenene synthase: Mechanistic insights on terpenoid cyclization reactions in biology. *Science*, **277**, (1997), 1820–1824.

[35] Wendt, K.U., Poralla, K. & Schulz, G.E. Structure and function of a squalene cyclase. *Science*, **277**, (1997), 1811–1815.

[36] Thoma, R., Schulz-Gasch, T., D'Arcy, B., Benz, J., Aebi, J., Dehmlow, H., Hennig, M., Stihle, M. & Ruf, A. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature*, **432**, (2004), 118–122.

[37] Starks, C.M., Back, K.W., Chappell, J. & Noel, J.P. Structural basis for cyclic terpene biosynthesis by tobacco 5-*epi*-aristolochene synthase. *Science*, **277**, (1997), 1815–1820.

[38] Peters, R.J., Flory, J.E., Jetter, R., Ravn, M.M., Lee, H.J., Coates, R.M. & Croteau, R.B. Abietadiene synthase from grand fir (*Abies grandis*): Characterization and mechanism of action of the "pseudomature" recombinant enzyme. *Biochemistry*, **39**, (2000), 15592–15602.

[39] Jiang, J., He, X. & Cane, D.E. Biosynthesis of the earthy odorant geosmin by a bifunctional *Streptomyces coelicolor* enzyme. *Nat. Chem..Biol.*, **3**, (2007), 711–715.

[40] Rynkiewicz, M.J., Cane, D.E. & Christianson, D.W. Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. *Proc. Natl. Acad. Sci. USA*, **98**, (2001), 13543–13548.

[41] Vedula, L.S., Jiang, J.Y., Zakharian, T., Cane, D.E. & Christianson, D.W. Structural and mechanistic analysis of trichodiene synthase using site-directed

mutagenesis: Probing the catalytic function of tyrosine 295 and the asparagine 225/serine 229/glutamate 233 Mg_B^{2+} motif. *Arch. Biochem. Biophys.*, **469**, (2008), 184–194.

[42] Gennadios, H.A., Gonzalez, V., Di Costanzo, L., Li, A., Yu, F., Miller, D.J., Allemann, R.K. & Christianson, D.W. Crystal structure of $(+)$-δ-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. *Biochemistry*, **48**, (2009), 6175–6183.

[43] Shishova, E.Y., Di Costanzo, L., Cane, D.E. & Christianson, D.W. X-ray crystal structure of aristolochene synthase from *Aspergillus terreus* and evolution of templates for the cyclization of farnesyl diphosphate. *Biochemistry*, **46**, (2007), 1941–1951.

[44] Shishova, E.Y., Yu, F.L., Miller, D.J., Faraldos, J.A., Zhao, Y.X., Coates, R.M., Allemann, R.K., Cane, D.E. & Christianson, D.W. X-ray crystallographic studies of substrate binding to aristolochene synthase suggest a metal ion binding sequence for catalysis. *J. Biol. Chem.*, **283**, (2008), 15431–15439.

[45] Tarshis, L.C., Yan, M.J., Poulter, C.D. & Sacchettini, J.C. Crystal-structure of recombinant farnesyl diphosphate synthase at 2.6 Angstrom resolution. *Biochemistry*, **33**, (1994), 10871–10877.

[46] Reardon, D. & Farber, G.K. Protein motifs. 4. The structure and evolution of α - β barrel proteins. *FASEBJ*, **9**, (1995), 497–503.

[47] Mau, C.J.D. & West, C.A. Cloning of casbene synthase cDNA - Evidence for conserved structural features among terpenoid cyclases in plants. *Proc. Natl. Acad. Sci. USA*, **91**, (1994), 8497–8501.

[48] Cane, D.E. & Kang, I. Aristolochene synthase: Purification, molecular cloning, high-level expression in *Escherichia coli*, and characterization of the *Aspergillus terreus* cyclase. *Arch. Biochem. Biophys.*, **376**, (2000), 354–364.

[49] Hosfield, D.J., Zhang, Y., Dougan, D.R., Broun, A., Tari, L.W., Swanson, R.V. & Finn, J. Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. *J. Biol. Chem.*, **279**, (2004), 8526–9.

[50] Hohn, T.M. & Plattner, R.D. Purification and characterization of the sesquiterpene cyclase aristolochene synthase from *Penicillium roqueforti*. *Arch. Biochem. Biophys.*, **272**, (1989), 137–143.

[51] Cane, D.E., Wu, Z., Proctor, R.H. & Hohn, T.M. Overexpression in *Escherichia coli* of soluble aristolochene synthase from *Penicillium roqueforti*. *Arch. Biochem. Biophys.*, **304**, (1993), 415–419.

[52] Proctor, R.H. & Hohn, T.M. Aristolochene synthase - Isolation, characterization, and bacterial expression of a sesquiterpenoid biosynthetic gene (ari1) from *Penicillium roqueforti*. *J. Biol. Chem.*, **268**, (1993), 4543–4548.

[53] Croteau, R.B., Shaskus, J.J., Renstrom, B., Felton, N.M., Cane, D.E., Saito, A. & Chang, C. Mechanism of the pyrophosphate migration in the enzymatic cyclization of geranyl and linalyl pyrophosphates to (+)- and (-)-bornyl pyrophosphates. *Biochemistry*, **24**, (1985), 7077–7085.

[54] Christianson, D.W. Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.*, **106**, (2006), 3412–3442.

[55] Cane, D.E., Abell, C., Harrison, P.H., Hubbard, B.R., Kane, C.T., Lattman, R., Oliver, J.S. & Weiner, S.W. Terpenoid biosynthesis and the stereochemistry of

enzyme-catalysed allylic addition-elimination reactions. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **332**, (1991), 123–129.

[56] Needham, J.V., Chen, T.Y. & Falke, J.J. Novel ion specificity of a carboxylate cluster Mg(II) binding site - Strong charge selectivity and weak size selectivity. *Biochemistry*, **32**, (1993), 3363–3367.

[57] Whittington, D.A., Wise, M.L., Urbansky, M., Coates, R.M., Croteau, R.B. & Christianson, D.W. Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc. Natl. Acad. Sci. USA*, **99**, (2002), 15375–15380.

[58] Miller, D.J., Yu, F.L. & Allemann, R.K. Aristolochene synthase catalyzed cyclization of 2-Fluorofarnesyl-Diphosphate to 2-Fluorogermacrene A. *ChemBioChem*, **8**, (2007), 1819–1825.

[59] Yu, F., Miller, D.J. & Allemann, R.K. Probing the reaction mechanism of aristolochene synthase with 12,13-difluorofarnesyl diphosphate. *Chem. Commun.*, **40**, (2008), 4155–4157.

[60] Banthorpe, D.V. & Branch, S.A. The biosynthesis of C5-C20 terpenoid compounds. *Nat. Prod. Rep.*, **2**, (1985), 513–524.

[61] Beale, M.H. & MacMillan, J. The biosynthesis of C5-C20 terpenoid compounds. *Nat. Prod. Rep.*, **5**, (1988), 247–264.

[62] Banthorpe, D. & Branch, S. The biosynthesis of C5-C20 terpenoid compounds. *Nat. Prod. Rep.*, **4**, (1987), 157–173.

[63] Winter, G., Fersht, A.R., Wilkinson, A.J., Zoller, M. & Smith, M. Redesigning enzyme structure by site-directed mutagenesis - Tyrosyl transfer-RNA synthetase and ATP binding. *Nature*, **299**, (1982), 756–758.

[64] Hutchison, C.A., Phillips, S., Edgell, M.H., Gillam, S., Jahnke, P. & Smith, M. Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.*, **253**, (1978), 6551–6560.

[65] Saiki, R.K., S., S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. & Arnheim, N. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle-cell anemia. *Science*, **230**, (1985), 1350–1354.

[66] Michaelis, L. & Menten, M. The kinetics of the inversion effect. *Biochem. Z.*, **49**, (1913), 333–369.

[67] Briggs, G. & Haldane, J. A note on the kinetics of enzyme action. *Biochem. J.*, **19**, (1925), 338–339.

[68] Lineweaver, H. & Burk, D. The determination of enzyme dissociation constants. *J. Am. Chem. Soc.*, **56**, (1934), 658–666.

[69] Alberty, R. & Hammes, G. Application of the theory of diffusion-controlled reactions to enzyme kinetics. *J. Phys. Chem.*, **62**, (1958), 154–159.

[70] Wolfenden, R. & Snider, M. The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res.*, **34**, (2001), 938–945.

[71] Calvert, M.J., Ashton, P.R. & Allemann, R.K. Germacrene A is a product of the aristolochene synthase mediated conversion of farnesyl pyrophosphate to aristolochene. *J. Am. Chem. Soc.*, **124**, (2002), 11636–11641.

[72] Calvert, M.J., Taylor, S.E. & Allemann, R.K. Tyrosine 92 of aristolochene synthase directs cyclisation of farnesyl pyrophosphate. *Chem. Commun.*, **20**, (2002), 2384–2385.

[73] Deligeorgopoulou, A., Taylor, S.E., Forcat, S. & Allemann, R.K. Stabilisation of eudesmane cation by tryptophan 334 during aristolochene synthase catalysis. *Chem. Commun.*, 2162–2163.

[74] Deligeorgopoulou, A. & Allemann, R.K. Evidence for differential folding of farnesyl pyrophosphate in the active site of aristolochene synthase: A single-point mutation converts aristolochene synthase into an (E)-β-farnesene synthase. *Biochemistry*, **42**, (2003), 7741–7747.

[75] Felicetti, B. & Cane, D.E. Aristolochene synthase: Mechanistic analysis of active site residues by site-directed mutagenesis. *J. Am. Chem. Soc.*, **126**, (2004), 7212–7221.

[76] Cane, D.E., Prabhakaran, P.C., Oliver, J.S. & McIlwaine, D.B. Aristolochene biosynthesis - stereochemistry of the deprotonation steps in the enzymatic cyclization of farnesyl pyrophosphate. *J. Am. Chem. Soc.*, **112**, (1990), 3209–3210.

[77] Aberhart, D.J. Microbiological preparation of (S)-(+)-2,3-dihydroxy-3-methylbutanoic acid by *syn* dihydroxylation of 3-methylcrotonic acid. *J. Org. Chem.*, **45**, (2002), 5218–5220.

[78] Harrison, P.H.M., Oliver, J.S. & Cane, D.E. Pentalenene biosynthesis and the enzymatic cyclization of farnesyl pyrophosphate - Inversion at C1 during 11-membered ring formation. *J. Am. Chem. Soc.*, **110**, (1988), 5922–5923.

[79] Cane, D.E., Prabhakaran, P.C., Salaski, E.J., Harrison, P.H.M., Noguchi, H. & Rawlings, B.J. Aristolochene biosynthesis and enzymatic cyclization of farnesyl pyrophosphate. *J. Am. Chem. Soc.*, **111**, (1989), 8914–8916.

[80] Croteau, R. Biosynthesis and catabolism of monoterpenoids. *Chem. Rev.*, **87**, (1987), 929–954.

[81] Cane, D., McIlwaine, D. & Oliver, J. Absolute-configuration of (-)-β-trans-bergamotene. *J. Am. Chem. Soc.*, **112**, (1990), 1285–1286.

[82] Ito, M., Kobayashi, M., Koyama, T. & Ogura, K. Stereochemical analysis of prenyltransferase reactions leading to (*Z*)- and (*E*)-polyprenyl chains. *Biochemistry*, **26**, (1987), 4745–4750.

[83] Forcat, S. & Allemann, R.K. Dual role for phenylalanine 178 during catalysis by aristolochene synthase. *Chem. Commun.*, **18**, (2004), 2094–2095.

[84] Forcat, S. & Allemann, R.K. Stabilisation of transition states prior to and following eudesmane cation in aristolochene synthase. *Org. Biomol. Chem.*, **4**, (2006), 2563–2567.

[85] Rising, K., Starks, C., Noel, J. & Chappell, J. Demonstration of germacrene A as an intermediate in 5-*epi*-aristolochene synthase catalysis. *J. Am. Chem. Soc.*, **122**, (2000), 6526–6526. J. Am. Chem. Soc.

[86] Felicetti, B. & Cane, D.E. Aristolochene synthase: Mechanistic analysis of active site residues by site-directed mutagenesis. *J. Am. Chem. Soc.*, **126**, (2004), 7212–7221.

[87] O'Hagan, D. Understanding organofluorine chemistry. An introduction to the C-F bond. *Chem. Soc. Rev.*, **37**, (2008), 308–319.

[88] Poulter, C.D. & Rilling, H.C. The prenyl transfer reaction. Enzymic and mechanistic studies of the 1'-4 coupling reaction in the terpene biosynthetic pathway. *Acc. Chem. Res.*, **11**, (1978), 307–313.

[89] Cane, D.E., Yang, G., Xue, Q. & Shim, J.H. Trichodiene synthase. Substrate specificity and inhibition. *Biochemistry*, **34**, (1995), 2471–2479.

[90] Jin, Y., Williams, D., Croteau, R. & Coates, R. Taxadiene synthase catalyzed cyclization of 6-Fluorogeranylgeranyl diphosphate to 7-Fluoroverticillenes. *J. Am. Chem. Soc.*, **127**, (2005), 7834–7842.

[91] Weinheim, A., Youngblow, W., Washeche, P., TKB, K. & Ciereszk, L. Isolation of elusive (-)-germacrene A from gorgonian, *Eunicea mammosa* chemistry of colelenterates. *Tetrahedron Lett.*, **7**, (1970), 497.

[92] Miller, D.J., Gao, J., Truhlar, D.G., Young, N.J., Gonzalez, V. & Allemann, R.K. Stereochemistry of eudesmane cation formation during catalysis by aristolochene synthase from *Penicillium roqueforti*. *Org. Biomol. Chem.*, **6**, (2008), 2346–2354.

[93] Kollner, T.G., Schnee, C., Li, S., Svatos, A., Schneider, B., Gershenzon, J. & Degenhardt, J. Protonation of a neutral (S)-β-bisabolene intermediate is involved in (S)-β-macrocarpene formation by the maize sesquiterpene synthases TPS6 and TPS11. *J. Biol. Chem.*, **283**, (2008), 20779–20788.

[94] Williams, D.C., Carroll, B.J., Jin, Q., Rithner, C.D., Lenger, S.R., Floss, H.G., Coates, R.M., Williams, R.M. & Croteau, R. Intramolecular proton transfer in the cyclization of geranylgeranyl diphosphate to the taxadiene precursor of taxol catalyzed by recombinant taxadiene synthase. *Chem. Biol.*, **7**, (2000), 969–977.

[95] Ravn, M., Coates, R., Jetter, R. & Croteau, R. Stereospecific intramolecular proton transfer in the cyclization of geranylgeranyl diphosphate to (-)-abietadiene catalyzed by recombinant cyclase from grand fir (*Abies grandis*). *Chem. Commun.*, **1**, (1998), 21–22.

[96] Morse, P.M. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Phys. Rev.*, **34**, (1929), 57–64.

[97] Cornell, W., Cieplak, P., Bayly, C., Gould, I., Merz, K., Ferguson, D., Spellmeyer, D., Fox, T., Caldwell, J. & Kollman, P. A 2nd generation force-field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, (1995), 5179–5197.

[98] MacKerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F., Mattos, C., Michnick, S., Ngo, T., Nguyen, D., Prodhom, B., Reiher, W., Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, (1998), 3586–3616.

[99] Allinger, N., Yuh, Y. & Lii, J. Molecular mechanics - the MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.*, **111**, (1989), 8551–8566.

[100] Jones, J.E. On the determination of molecular fields. I. From the variation of the viscosity of a gas with temperature. *Proc. R. Soc. Lond. A. Math. Phys. Sci.*, **106**, (1924), 441–462.

[101] Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S. & Karplus, M. CHARMM - A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, (1983), 187–217.

[102] Pang, J., Pu, J., Gao, J., Truhlar, D.G. & Allemann, R.K. Hydride transfer reaction catalyzed by hyperthermophilic dihydrofolate reductase is dominated by quantum mechanical tunneling and is promoted by both inter- and intramonomeric correlated motions. *J. Am. Chem. Soc.*, **128**, (2006), 8015–23.

[103] Schuler, L., Daura, X. & van Gunsteren, W. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.*, **22**, (2001), 1205–1218.

[104] Forrest, L.R. & Honig, B. An assessment of the accuracy of methods for predicting hydrogen positions in protein structures. *Proteins*, **61**, (2005), 296–309.

[105] Mackerell, Jr, A.D., Feig, M. & Brooks, 3rd, C.L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.*, **25**, (2004), 1400–15.

[106] Trbovic, N., Kim, B., Friesner, R.A. & Palmer, 3rd, A.G. Structural analysis of protein dynamics by MD simulations and NMR spin-relaxation. *Proteins*, **71**, (2008), 684–94.

[107] Gao, J., Ma, S., Major, D.T., Nam, K., Pu, J. & Truhlar, D.G. Mechanisms and free energies of enzymatic reactions. *Chem. Rev.*, **106**, (2006), 3188–3209.

[108] Schrödinger, E. Quantization as a problem of eigenvalue. *Ann. Phys.*, **81**.

[109] Schrödinger, E. Quantisation as an eigenvalue problem. *Ann. Phys.*, **79**, (1926), 361.

[110] Born, M. & Oppenheimer, R. Quantum theory of molecules. *Ann. Phys.*, **84**, (1927), 457–484.

[111] Atkins, P. *Physical Chemistry* (Oxford University Press, Oxford, UK, 1988), 3rd ed.

[112] Hartree, D. The wave mechanics of an atom with a non-Coulomb central field. Part I theory and methods. *Proc. Camb. Philol. Soc.*, **24**, (1928), 89–110.

[113] Hartree, D. The wave mechanics of an atom with a non-Coulomb central field. Part II some results and discussion. *Proc. Camb. Philol. Soc.*, **24**, (1928), 111–132.

[114] Hartree, D. The wave mechanics of an atom with a non-Coulomb central field. Part III Term values and intensities in series an optical spectra. *Proc. Camb. Philol. Soc.*, **24**, (1928), 426–437.

[115] Pauli, W. On the connection of the arrangement of electron groups in atoms with the complex structure of spectra. *Z. Phys.*, **31**, (1925), 765–783.

[116] Slater, J. The theory of complex spectra. *Phys. Rev.*, **34**, (1929), 1293–1322.

[117] Slater, J. Physical meaning of wave mechanics. *Journal of the Franklin Institute*, **207**, (1929), 449–455.

[118] Slater, J. Atomic shielding constants. *Phys. Rev.*, **36**, (1930), 57–64.

[119] Boys, S. Electronic wave functions. 2. A calculation for the ground state of the beryllium atom. *Proc. R. Soc. Lond. A. Math. Phys. Sci.*, **201**, (1950), 125–137.

[120] Boys, S. Electronic wave functions. 1. A general method of calculation for the stationary states of any molecular system. *Proc. R. Soc. Lond. A. Math. Phys. Sci.*, **200**, (1950), 542–554.

[121] Pople, J. & Segal, G. Approximate self consistent molecular orbital theory. 2. Calculations with complete neglect of differential overlap. *J. Chem. Phys.*, **43**, (1965), 136.

[122] Pople, J., Santry, D. & Segal, G. Approximate self consistent molecular orbital theory. i. Invariant procedures. *J. Chem. Phys.*, **43**, (1965), 129.

[123] Dewar, M. & Thiel, W. Ground states of molecules. 38. MNDO method - Approximations and parameters. *J. Am. Chem. Soc.*, **99**, (1977), 4899–4907.

[124] Dewar, M. & Thiel, W. Ground states of molecules . 39. MNDO results for molecules containing hydrogen, carbon, nitrogen, and oxygen. *J. Am. Chem. Soc.*, **99**, (1977), 4907–4917.

[125] Dewar, M., Zoebisch, E., Healy, E. & Stewart, J. The development and use of quantum mechanical molecular models. 76. AM1 - A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, **107**, (1985), 3902–3909.

[126] Stewart, J. Optimization of parameters for semiempirical methods. 2. Applications. *J. Comput. Chem.*, **10**, (1989), 221–264.

[127] Rossi, I. & Truhlar, D. Parameterization of NDDO wave functions using genetic algorithms - An evolutionary approach to parameterizing potential

energy surfaces and direct dynamics calculations for organic reactions. *Chem. Phys. Lett.*, **233**, (1995), 231–236.

[128] Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev. B*, **136**, (1964), 864.

[129] Kohn, W. & Sham, L. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, **140**, (1965), 1133.

[130] Perdew, J. & Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, **45**, (1992), 13244–13249.

[131] Vosko, S., Wilk, L. & Nusair, M. Accurate spin dependent electron liquid correlation energies for local spin-density calculations - A critical analysis. *Can. J. Phys.*, **58**, (1980), 1200–1211.

[132] Becke, A. Density-functional exchange energy approximation with correct asymptotic behavior. *Phys. Rev. A*, **38**, (1988), 3098–3100.

[133] Stephens, P., Jalkanen, K., Devlin, F. & Chabalowski, C. *Ab initio* calculation of vibrational circular dichroism spectra using accurate post self-consistent field force fields - Trans-2,3-dideuteriooxirane. *J. Phys. Chem.*, **97**, (1993), 6107–6110.

[134] Lee, C., Yang, W. & Parr, R. Development of the Colle-Salvetti correlation energy formula into a functional of the electron density. *Phys. Rev. B*, **37**, (1988), 785–789.

[135] Curtiss, L., Redfern, P., Raghavachari, K. & Pople, J. Assessment of Gaussian2 and density functional theories for the computation of ionization potentials and electron affinities. *J. Chem. Phys.*, **109**, (1998), 42–55.

[136] Curtiss, L., Raghavachari, K., Trucks, G. & Pople, J. Gaussian2 theory for molecular energies of 1st row and 2nd row compounds. *J. Chem. Phys.*, **94**, (1991), 7221–7230.

[137] Curtiss, L., Raghavachari, K., Redfern, P. & Pople, J. Gaussian3 theory using scaled energies. *J. Chem. Phys.*, **112**, (2000), 1125–1132.

[138] Johnson, B., Gill, P. & Pople, J. The performance of a family of density functional methods. *J. Chem. Phys.*, **98**, (1993), 5612–5626.

[139] Scheiner, A., Baker, J. & Andzelm, J. Molecular energies and properties from density functional theory: Exploring basis set dependence of Kohn-Sham equation using several density functionals. *J. Comput. Chem.*, **18**, (1997), 775–795.

[140] Boese, A., Martin, J. & Handy, N. The role of the basis set: Assessing density functional theory. *J. Chem. Phys.*, **119**, (2003), 3005–3014.

[141] Proynov, E., Chermette, H. & Salahub, D. New τ-dependent correlation functional combined with a modified Becke exchange. *J. Chem. Phys.*, **113**, (2000), 10013–10027.

[142] Gao, J., Amara, P., Alhambra, C. & Field, M. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A*, **102**, (1998), 4714–4721.

[143] Field, M., Bash, P. & Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.*, **11**, (1990), 700–733.

[144] Singh, U. & Kollman, P. A combined *ab initio* quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems - Applications to the CH_3Cl^+ Cl exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.*, **7**, (1986), 718–730.

[145] Reuter, N., Dejaegere, A., Maigret, B. & Karplus, M. Frontier bonds in QM/MM methods: A comparison of different approaches. *J. Phys. Chem. A*, **104**, (2000), 1720–1735.

[146] Rodriguez, A., Oliva, C., Gonzalez, M., van der Kamp, M. & Mulholland, A.J. Comparison of different quantum mechanical/molecular mechanics boundary treatments in the reaction of the hepatitis C virus NS3 protease with the NS5A/5B substrate. *J. Phys. Chem. B*, **111**, (2007), 12909–12915.

[147] Verlet, L. Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, **159**, (1967), 98.

[148] Yoneya, M., Berendsen, H. & Hirasawa, K. A non iterative matrix method for constraint molecular dynamics simulations. *Mol. Simul.*, **13**, (1994), 395–405.

[149] Ryckaert, J., Ciccotti, G. & Berendsen, H. Numerical integration of Cartesian equations of motion of a system with constraints - Molecular dynamics of *n*-alkanes. *J. Comput. Phys.*, **23**, (1977), 327–341.

[150] Brooks, C. & Karplus, M. Deformable stochastic boundaries in molecular dynamics. *J. Chem. Phys.*, **79**, (1983), 6312–6325.

[151] Kumar, S., Bouzida, D., Swendsen, R., Kollman, P. & Rosenberg, J. The weighted histogram analysis method for free energy calculations on biomolecules. 1. The method. *J. Comput. Chem.*, **13**, (1992), 1011–1021.

[152] Gutta, P. & Tantillo, D.J. A promiscuous proton in taxadiene biosynthesis? *Org. Lett.*, **9**, (2007), 1069–1071.

[153] Ho, G.A., Nouri, D.H. & Tantillo, D.J. Carbocation rearrangements in aspernomine biosynthesis. *Tetrahedron Letters*, **50**, (2009), 1578–1581.

[154] Lodewyk, M.W., Gutta, P. & Tantillo, D.J. Computational studies on biosynthetic carbocation rearrangements leading to sativene, cyclosativene, α-ylangene, and β-ylangene. *J. Org. Chem.*, **73**, (2008), 6570–6579.

[155] Ho, G., Nouri, D. & Tantillo, D. The cationic cascade route to longifolene. *Journal of Organic Chemistry*, **70**, (2005), 5139–5143.

[156] Gutta, P. & Tantillo, D. Theoretical studies on farnesyl cation cyclization: Pathways to pentalenene. *J. Am. Chem. Soc.*, **128**, (2006), 6172–6179.

[157] Hong, Y.J. & Tantillo, D.J. Which is more likely in trichodiene biosynthesis: Hydride or proton transfer? *Organic Letters*, **8**, (2006), 4601–4604.

[158] Cool, L.G. Sesquiterpenes from *Cupressus macrocarpa* foliage. *Phytochemistry*, **66**, (2005), 249–260.

[159] Cane, D.E., Oliver, J.S., Harrison, P.H.M., Abell, C., Hubbard, B.R., Kane, C.T. & Lattman, R. Biosynthesis of pentalenene and pentalenolactone. *J. Am. Chem. Soc.*, **112**, (1990), 4513–4524.

[160] Cane, D.E., Abell, C., Lattman, R., Kane, C.T., Hubbard, B.R. & Harrison, P.H.M. Pentalenene biosynthesis and the enzymatic cyclization of farnesyl pyrophosphate - anti stereochemistry in a biological se' reaction. *J. Am. Chem. Soc.*, **110**, (1988), 4081–4082.

[161] Seemann, M., Zhai, G.Z., de Kraker, J.W., Paschall, C.M., Christianson, D.W. & Cane, D.E. Pentalenene synthase. analysis of active site residues by site-directed mutagenesis. *J. Am. Chem. Soc.*, **124**, (2002), 7681–7689.

[162] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, Jr., J.A., Vreven, T., Kudin, K.N., Burant, J.C., Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J.E., Hratchian, H.P., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Ayala, P.Y., Morokuma, K., Voth, G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Gonzalez, C. & Pople, J.A. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

[163] Schaftenaar, G. & Noordik, J. Molden: A pre- and post-processing program for molecular and electronic structures. *J. Comput. Aided Mol. Des.*, **14**, (2000), 123–134.

[164] Adamo, C. & Barone, V. Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J. Chem. Phys.*, **108**, (1998), 664–675.

[165] Zhao, Y. & Truhlar, D. Hybrid meta density functional theory methods for thermochemistry, thermochemical kinetics, and noncovalent interactions: The MPW1B95 and MPWB1K models and comparative assessments for hydrogen bonding and van der Waals interactions. *J. Phys. Chem. A*, **108**, (2004), 6908–6918.

[166] Møller, C. & Plesset, M. Note on an approximation treatment for many electron systems. *Phys. Rev.*, **46**, (1934), 618–622.

[167] Becke, A.D. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys. C*, **98**, (1993), 5648–5652.

[168] Inc, H. Hyppercube 7.0. 1115 NW 4th Street, Gainsville, FL 32601, USA.

[169] Allinger, N. Conformational analysis. 130. MM2 - hydrocarbon force field utilizing V_1 and V_2 torsional terms. *J. Am. Chem. Soc.*, **99**, (1977), 8127–8134.

[170] Matsuda, S.P.T., Wilson, W.K. & Xiong, Q. Mechanistic insights into triterpene synthesis from quantum mechanical calculations. Detection of systematic errors in B3LYP cyclization energies. *Org. Biomol. Chem.*, **4**, (2006), 530–43.

[171] Hammond, G.S. A correlation of reaction rates. *J. Am. Chem. Soc.*, **77**, (1955), 334–338.

[172] Allemann, R.K., Young, N.J., Ma, S.H., Truhlar, D.G. & Gao, J.L. Synthetic efficiency in enzyme mechanisms involving carbocations: Aristolochene synthase. *J. Am. Chem. Soc.*, **129**, (2007), 13008–13013.

[173] Miller, D.J., Yu, F., Knight, D.W. & Allemann, R.K. 6- and 14-Fluoro farnesyl diphosphate: mechanistic probes for the reaction catalysed by aristolochene synthase. *Org. Biomol. Chem.*, **7**, (2009), 962–75.

[174] Marcus, R.K. Energetic and dynamical aspects of proton transfer reactions in solution. *Faraday Symposia of the Chemical Society*, **10**, (1975), 60.

[175] Lu, Y.P., Liu, H.G. & Liang, P.H. Different reaction mechanisms for cis- and trans-prenyltransferases. *Biochem. Biophys. Res. Commun.*, **379**, (2009), 351–5.

[176] Calvert, M.J., Ashton, P.R. & Allemann, R.K. Germacrene A is a product of the aristolochene synthase mediated conversion of farnesyl pyrophosphate to aristolochene. *J. Am. Chem. Soc.*, **124**, (2002), 11636–11641.

[177] Gao, J. Methods and applications of combined quantum mechanical and molecular mechanical potentials. *Reviews in Computational Chemistry*, **7**, (1995), 119–185.

[178] Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, (1983), 926–935.

[179] Brunger, A.T. & Karplus, M. Polar hydrogen positions in proteins: Empirical energy placement and neutron diffraction comparison. *Proteins*, **4**, (1988), 148–156.

[180] Guex, N. & Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, **18**, (1997), 2714–2723.

[181] Schwede, T., Kopp, J., Guex, N. & Peitsch, M.C. Swiss-model: An automated protein homology-modeling server. *Nucleic Acids Res.*, **31**, (2003), 3381–3385.

[182] Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Grap.*, **14**, (1996), 33–38.

[183] Eargle, J., Wright, D. & Luthey-Schulten, Z. Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics*, **22**, (2006), 504–506.

[184] Frishman, D. & Argos, P. Knowledge-based secondary structure assignment. *Proteins*, **23**, (1995), 566–579.

[185] Sanner, M., Olsen, A. & Spehner, J.C. Fast and robust computation of molecular surfaces. In *Proceedings of the 11th ACM Symposium on Computational Geometry*, C6–C7 (ACM, New York, 1995).

[186] Sharma, R., Zeller, M., Pavlovic, V.I., Huang, T.S., Lo, Z., Chu, S., Zhao, Y., Phillips, J.C. & Schulten, K. Speech/gesture interface to a visual-computing environment. *IEEECGA*, **20**, (2000), 29–37.

[187] Stone, J. An Efficient Library for Parallel Ray Tracing and Animation. Master's thesis, Computer Science Department, University of Missouri-Rolla (1998).

[188] Stone, J., Gullingsrud, J., Grayson, P. & Schulten, K. A system for interactive molecular dynamics simulation. In J.F. Hughes & C.H. Séquin, eds., *2001 ACM Symposium on Interactive 3D Graphics*, 191–194 (ACM SIGGRAPH, New York, 2001).

[189] Varshney, A., Brooks, F.P. & Wright, W.V. Linearly scalable computation of smooth molecular surfaces. *IEEE Computer Graphics and Applications*, **14**, (1994), 19–25.

[190] Kottalam, J. & Case, D. Dynamics of ligand escape from the heme pocket of myoglobin. *J. Am. Chem. Soc.*, **110**, (1988), 7690–7697.

[191] Morikubo, N., Fukuda, Y., Ohtake, K., Shinya, N., Kiga, D., Sakamoto, K., Asanuma, M., Hirota, H., Yokoyama, S. & Hoshino, T. Cation-π interaction in the polyolefin cyclization cascade uncovered by incorporating unnatural amino acids into the catalytic sites of squalene cyclase. *J. Am. Chem. Soc.*, **128**, (2006), 13184–13194.

[192] Yoshikuni, Y., Ferrin, T.E. & Keasling, J.D. Designed divergent evolution of enzyme function. *Nature*, **440**, (2006), 1078–1082.

[193] Garcia-Viloca, M., Poulsen, T.D., Truhlar, D.G. & Gao, J. Sensitivity of molecular dynamics simulations to the choice of the X-ray structure used to model an enzymatic reaction. *Protein Sci.*, **13**, (2004), 2341–2354.

Appendix A

Interaction Between Active Site Amino Acid Residues and Carbocation Centres

A.1 Active site residues in direct intramolecular transfer

A.1.1 Interaction between F178 and C7, C3 and C2 in direct intramolecular transfer mechanism
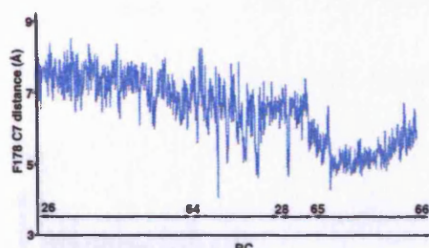
Figure A.1: *Distance between F178 and C7 in model A during direct intramolecular proton transfer - start and end points of simulations and corresponding structures indicated by double headed arrows*
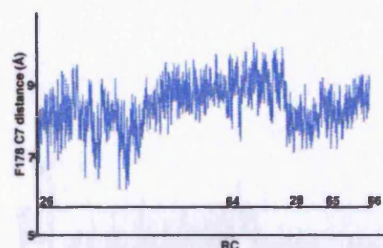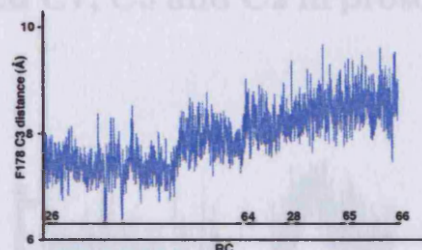

Figure A.2: *Distance between F178 and C7 in model B during direct intramolecular proton transfer - start and end points of simulations and corresponding structures indicated by double headed arrows*
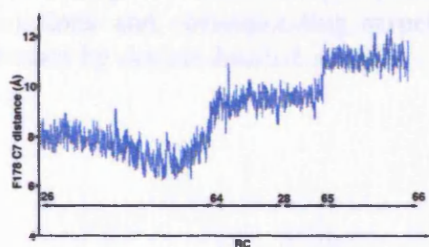
Figure A.3: *Distance between F178 and C7 in model C during direct intramolecular proton transfer - start and end points of simulations and corresponding structures indicated by double headed arrows*
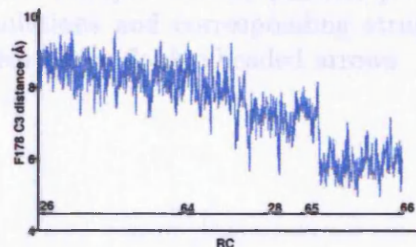
Figure A.4: *Distance between F178 and C3 in model A during direct intramolecular proton transfer - start and end points of simulations and corresponding structures indicated by double headed arrows*
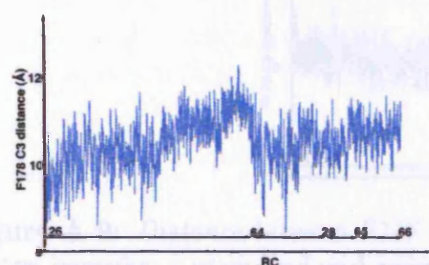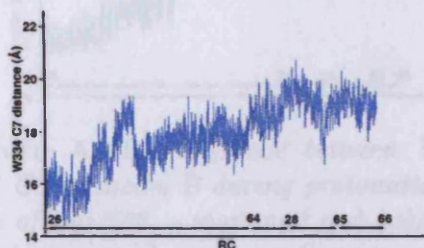

Figure A.5: *Distance between F178 and C3 in model B during direct intramolecular proton transfer - start and end points of simulations and corresponding structures indicated by double headed arrows*
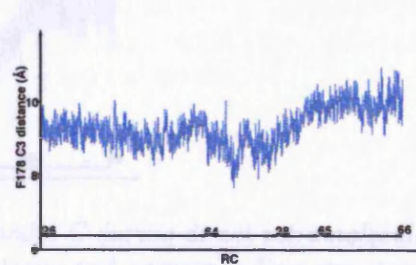
Figure A.6: *Distance between F178 and C3 in model C during direct intramolecular proton transfer - start and end points of simulations and corresponding structures indicated by double headed arrows*
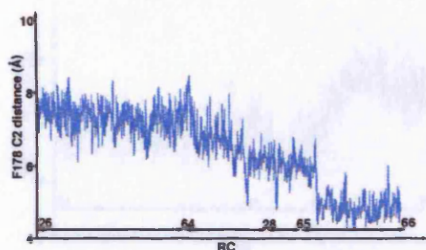
Figure A.7: *Distance between F178 and C2 in model A during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
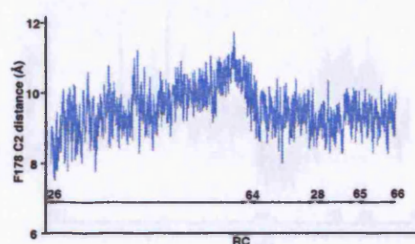
Figure A.8: *Distance between F178 and C2 in model B during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure A.9: *Distance between F178 and C3 in model C during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows

## A.1.2 Interaction between W334 and C7, C3 and C2 in direct intramolecular transfer mechanism
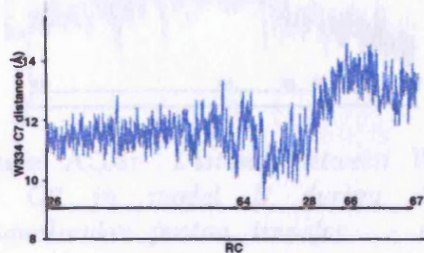
Figure A.10: *Distance between W334 and C7 in model A during direct intramolecular proton transfer - start and end points of simulations and corresponding structures indicated by double headed arrows*
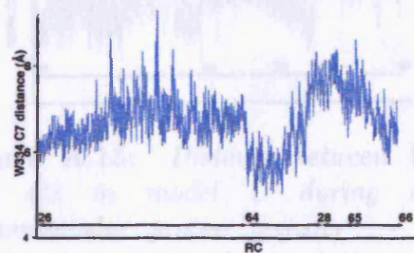
Figure A.11: *Distance between W334 and C7 in model B during direct intramolecular proton transfer - start and end points of simulations and corresponding structures indicated by double headed arrows*
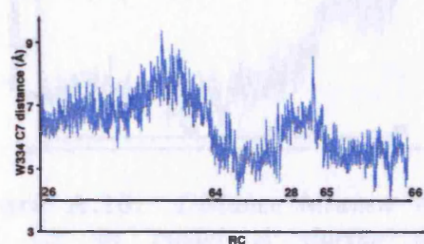
Figure A.12: *Distance between W334 and C7 in model C during direct intramolecular proton transfer - start and end points of simulations and corresponding structures indicated by double headed arrows*
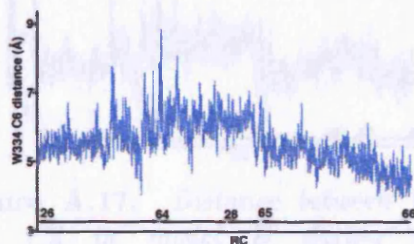
Figure A.13: *Distance between W334 and C3 in model A during direct intramolecular proton transfer - start and end points of simulations and corresponding structures indicated by double headed arrows*
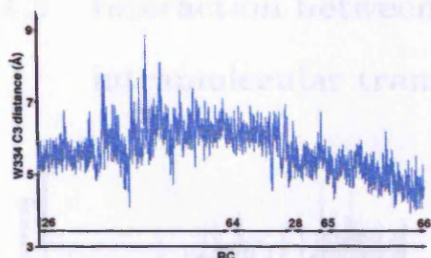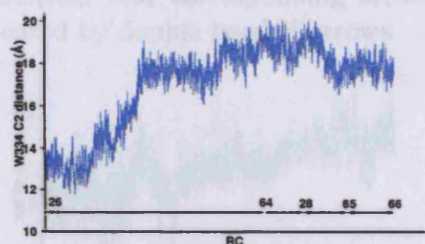
Figure A.14: *Distance between W334 and C3 in model B during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
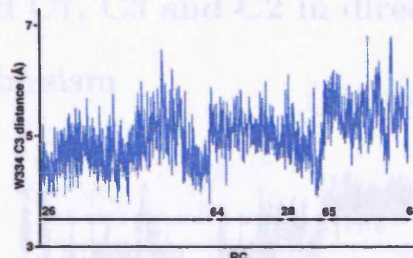
Figure A.15: *Distance between W334 and C3 in model C during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
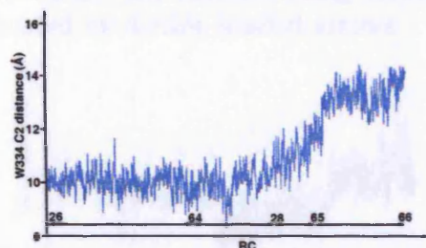
Figure A.16: *Distance between W334 and C2 in model A during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
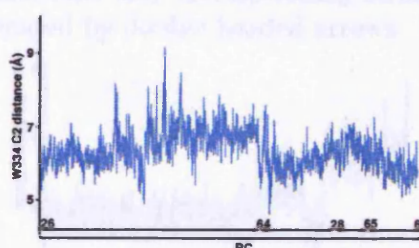
Figure A.17: *Distance between W334 and C2 in model B during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
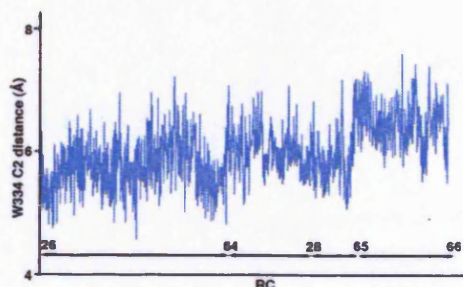
Figure A.18: *Distance between W334 and C2 in model C during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows

A.1.3 Interaction between Y92 and C7, C3 and C2 in direct intramolecular transfer mechanism
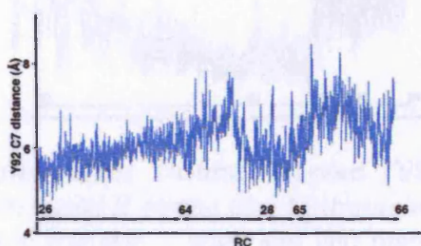
Figure A.19: *Distance between Y92 and C7 in model A during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
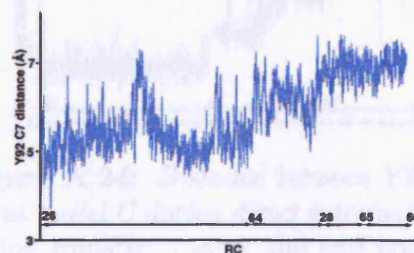
Figure A.20: *Distance between Y92 and C7 in model B during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
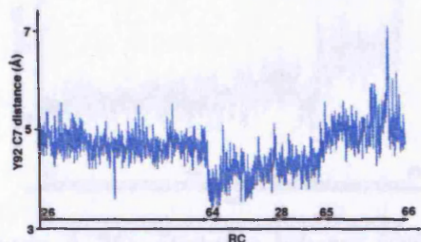
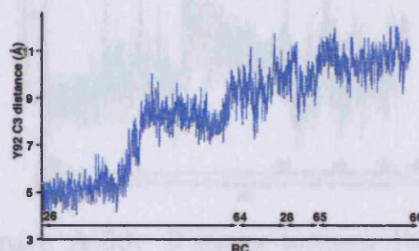**Figure A.21:** *Distance between Y92 and C7 in model C during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
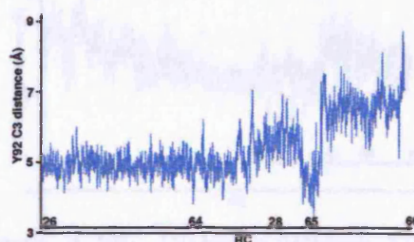
Figure A.22: *Distance between Y92 and C3 in model A during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
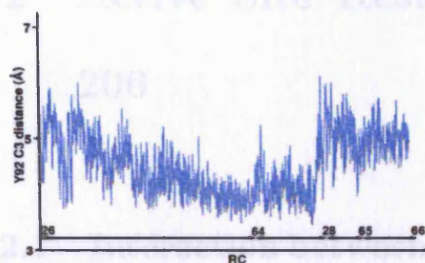
Figure A.23: *Distance between Y92 and C3 in model B during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
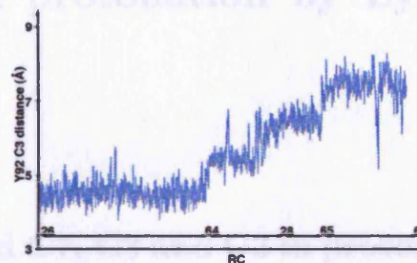
Figure A.24: *Distance between Y92 and C3 in model C during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
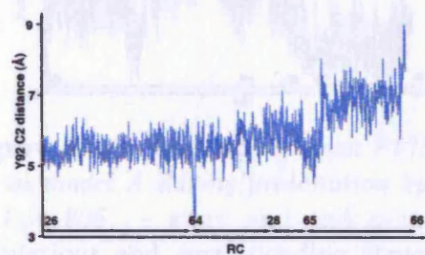
Figure A.25: *Distance between Y92 and C2 in model A during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
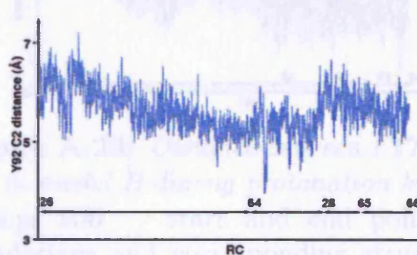
Figure A.26: *Distance between Y92 and C2 in model B during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows
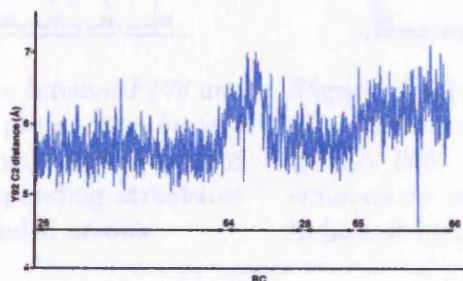
Figure A.27: *Distance between Y92 and C2 in model C during direct intramolecular proton transfer* - start and end points of simulations and corresponding structures indicated by double headed arrows

A.2 Active Site Residues in protonation by Lys 206

A.2.1 Interaction between F178 and C7, C3 and C2 in protonation by way of Lys 206
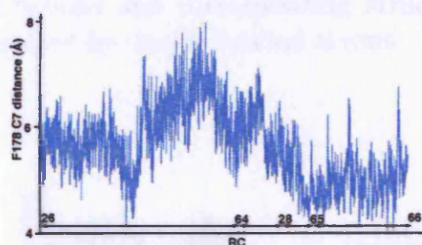
Figure A.28: *Distance between F178 and C7 in model A during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
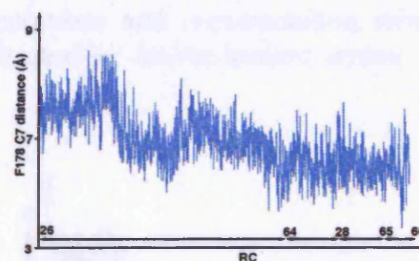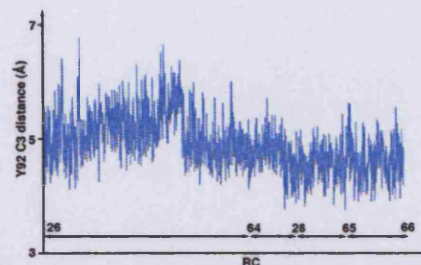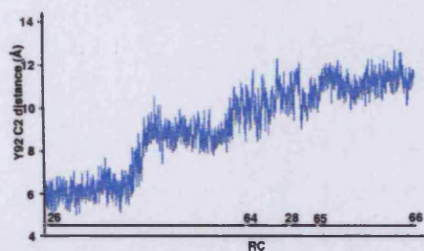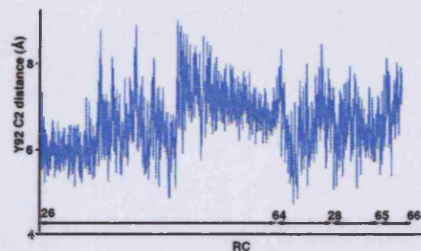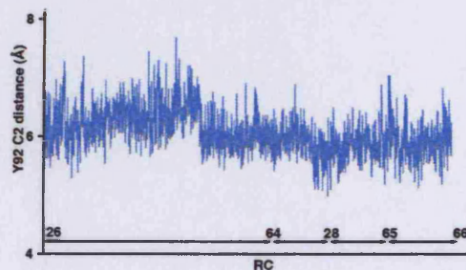

Figure A.29: *Distance between F178 and C7 in model B during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
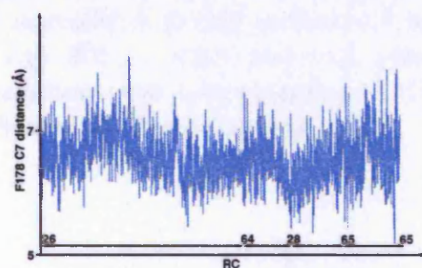
Figure A.30: *Distance between F178 and C7 in model C during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
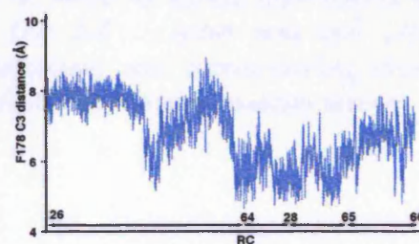
Figure A.31: *Distance between F178 and C3 in model A during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure A.32: *Distance between F178 and C3 in model B during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure A.33: *Distance between F178 and C3 in model C during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
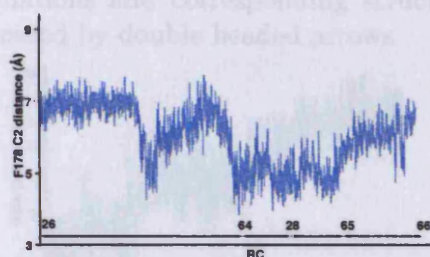
Figure A.34: *Distance between F178 and C2 in model A during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
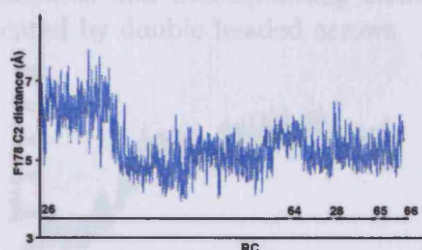
Figure A.35: *Distance between F178 and C2 in model B during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
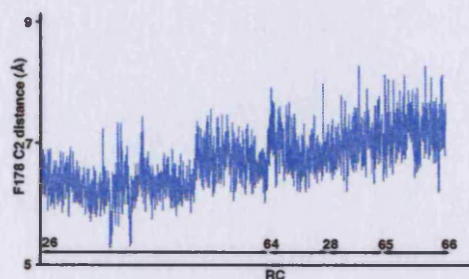
Figure A.36: *Distance between F178 and C3 in model C during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows

A.2.2 Interaction between W334 and C7, C3 and C2 in protonation by way of Lys 206

Figure A.37: *Distance between W334 and C7 in model A during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
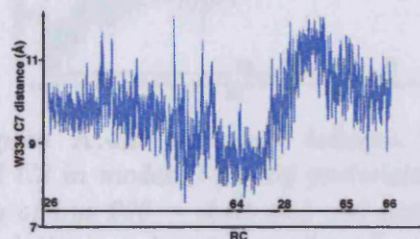
Figure A.38: *Distance between W334 and C7 in model B during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
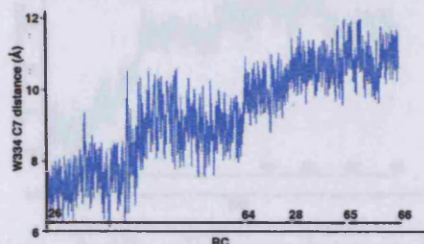
Figure A.39: *Distance between W334 and C7 in model C during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
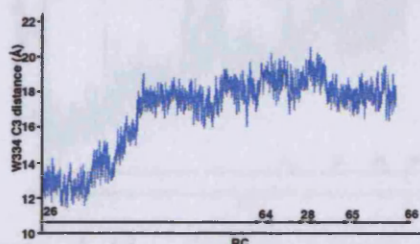
Figure A.40: *Distance between W334 and C3 in model A during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure A.41: *Distance between W334 and C3 in model B during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
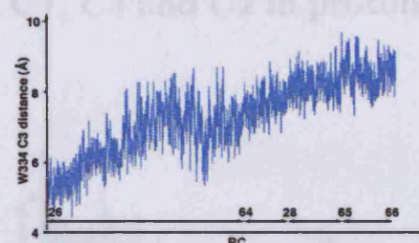
Figure A.42: *Distance between W334 and C3 in model C during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure A.43: *Distance between W334 and C2 in model A during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
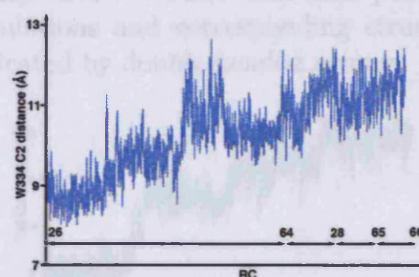
Figure A.44: *Distance between W334 and C2 in model B during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
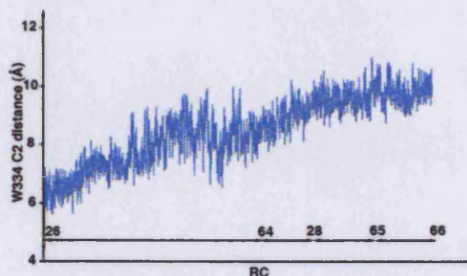
Figure A.45: *Distance between W334 and C3 in model C during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows

A.2.3 Interaction between Y92 and C7, C3 and C2 in protonation by way of Lys 206
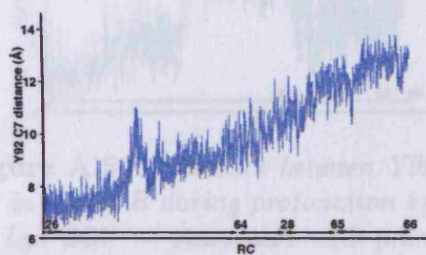
Figure A.46: *Distance between Y92 and C7 in model A during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
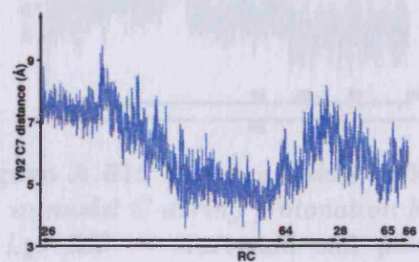
Figure A.47: *Distance between Y92 and C7 in model B during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
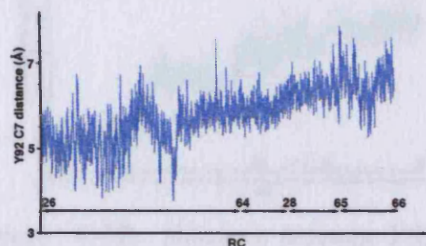
Figure A.48: *Distance between Y92 and C7 in model C during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
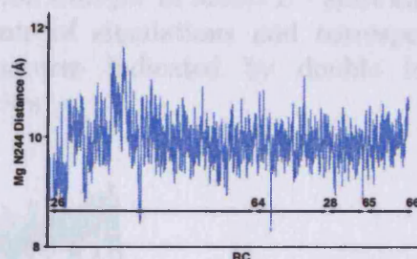

Figure A.49: *Distance between Y92 and C3 in model A during protonation by way of Lys 206* - start and end points of simulations and corresponding structures indicated by double headed arrows
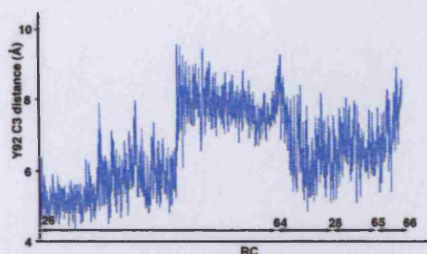
Figure A.50: *Distance between Y92 and C3 in model B during protonation by way of Lys 206 - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure A.51: *Distance between Y92 and C3 in model C during protonation by way of Lys 206 - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure A.52: *Distance between Y92 and C2 in model A during protonation by way of Lys 206 - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure A.53: *Distance between Y92 and C2 in model B during protonation by way of Lys 206 - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure A.54: *Distance between Y92 and C3 in model C during protonation by way of Lys 206 - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure B.1: *Distance between Mg$_A$ and D115 during direct intramolecular proton transfer in Model A* - start and end points of simulations and corresponding structures indicated by double headed arrows
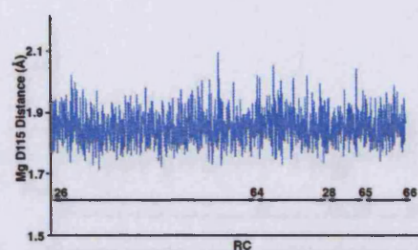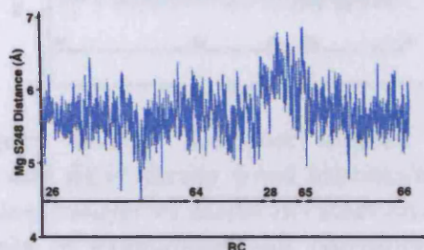

Figure B.2: *Distance between Mg$_A$ ion and D115 during direct intramolecular proton transfer in Model B* - start and end points of simulations and corresponding structures indicated by double headed arrows
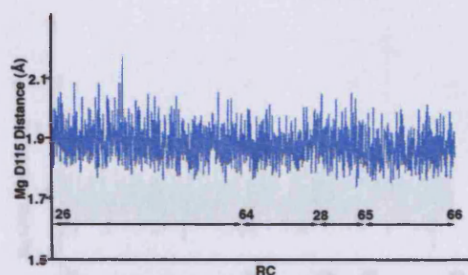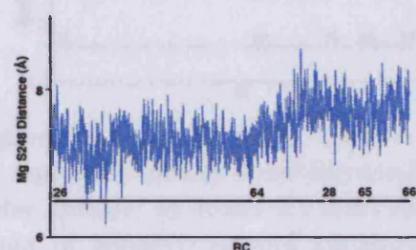

Figure B.3: *Distance between Mg$_A$ ion and D115 during direct intramolecular proton transfer in Model C* - start and end points of simulations and corresponding structures indicated by double headed arrows

Appendix B

Interaction Between Magnesium

Binding Residues and Carbocation
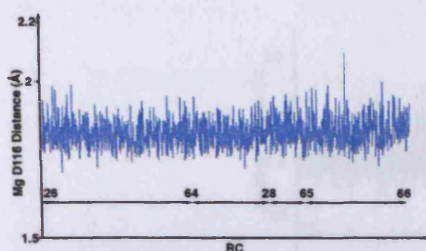
Figure B.4: *Distance between Mg_A ion and D116 during direct intramolecular proton transfer in Model A - start and end points of simulations and corresponding structures indicated by double headed arrows*
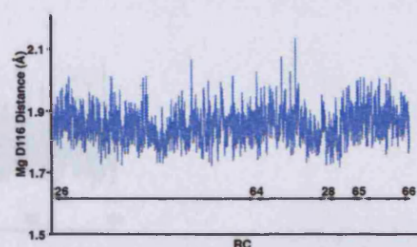
Figure B.5: *Distance between Mg_A ion and D116 during direct intramolecular proton transfer in Model B - start and end points of simulations and corresponding structures indicated by double headed arrows*
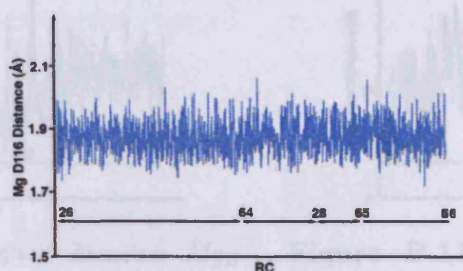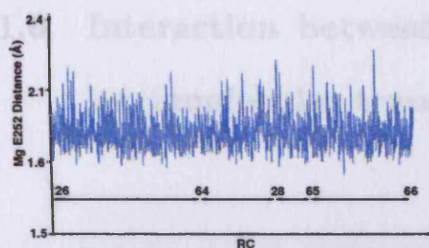

Figure B.6: *Distance between Mg_A ion and D116 during direct intramolecular proton transfer in Model C - start and end points of simulations and corresponding structures indicated by double headed arrows*
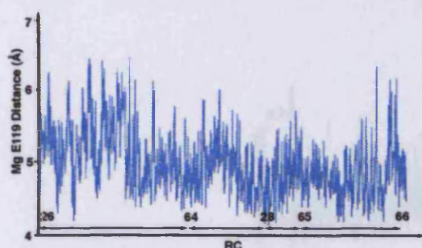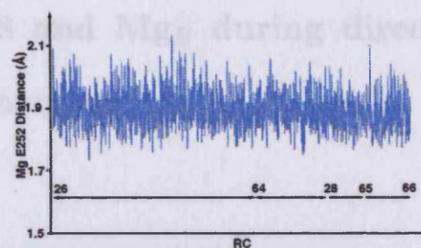

Figure B.7: *Distance between Mg_A ion and E119 during direct intramolecular proton transfer in Model A - start and end points of simulations and corresponding structures indicated by double headed arrows*
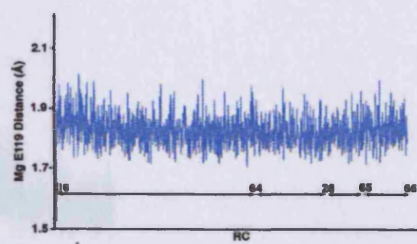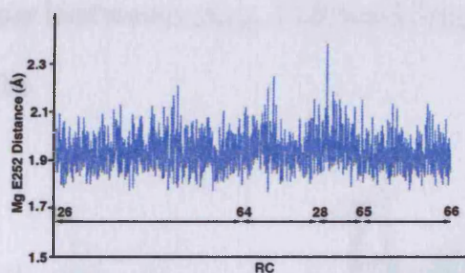
Figure B.8: *Distance between Mg_A ion and E119 during direct intramolecular proton transfer in Model B - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure B.9: *Distance between Mg_A ion and E119 during direct intramolecular proton transfer in Model C* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure B.10: *Distance between Mg_B ion and N244 during direct intramolecular proton transfer in Model A* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure B.11: *Distance between Mg_B ion and N244 during direct intramolecular proton transfer in Model B* - start and end points of simulations and corresponding structures indicated by double headed arrows
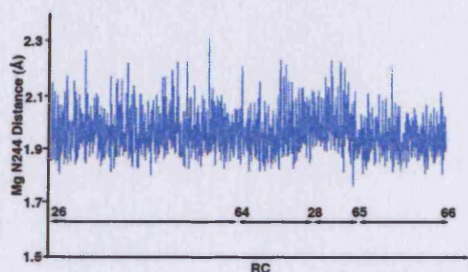
Figure B.12: *Distance between Mg_B ion and N244 during direct intramolecular proton transfer in Model C* - start and end points of simulations and corresponding structures indicated by double headed arrows

B.1.5   Interaction between Ser 248 and Mg_B during direct intramolecular transfer mechanism

Figure B.13: *Distance between Mg_B ion and S248 during direct intramolecular proton transfer in Model A* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure B.14: *Distance between Mg_B ion and S248 during direct intramolecular proton transfer in Model B* - start and end points of simulations and corresponding structures indicated by double headed arrows
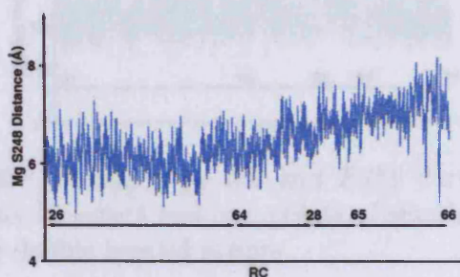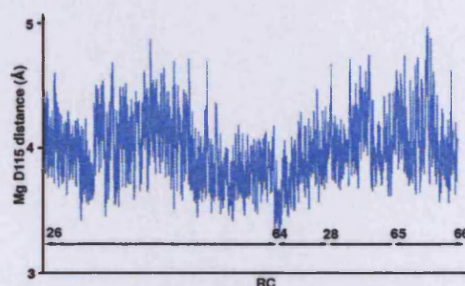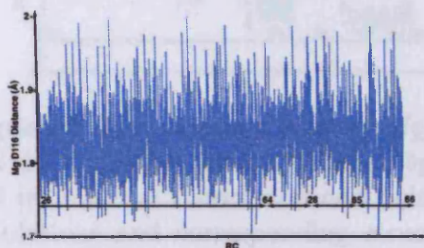

Figure B.15: *Distance between Mg_B ion and S248 during direct intramolecular proton transfer in Model C* - start and end points of simulations and corresponding structures indicated by double headed arrows
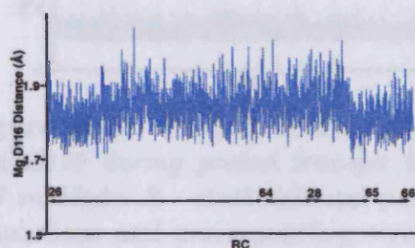
Figure B.16: *Distance between Mg_B ion and E252 during direct intramolecular proton transfer in Model A* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure B.17: *Distance between Mg_B ion and E252 during direct intramolecular proton transfer in Model B* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure B.18: *Distance between Mg_B ion and E252 during direct intramolecular proton transfer in Model C* - start and end points of simulations and corresponding structures indicated by double headed arrows

B.1.6 Interaction between Ser 248 and Mg$_B$ during direct intramolecular transfer mechanism

B.2 Magnesium Binding Residues and Protonation by Lys 206

B.2.1 Interaction between Asp 115 and Mg$_A$ during protonation by Lys 206
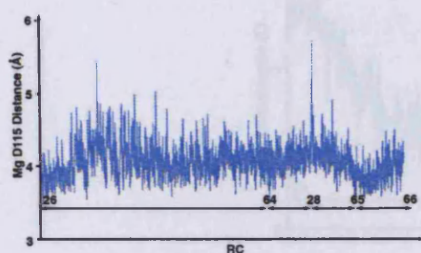
Figure B.19: *Distance between Mg$_A$ ion and D115 during proton transfer by Lys 206 in Model A* - start and end points of simulations and corresponding structures indicated by double headed arrows
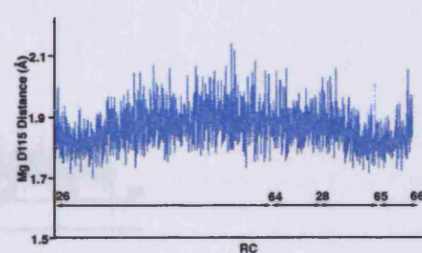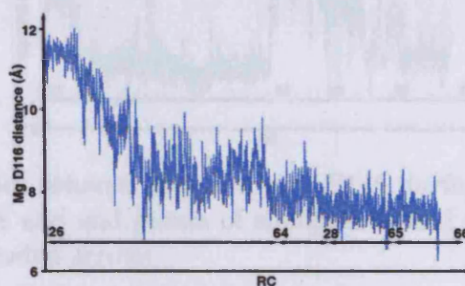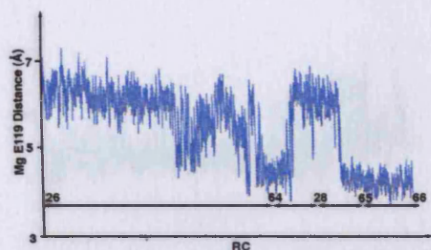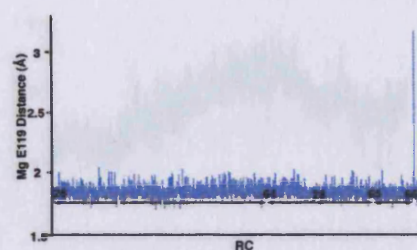

Figure B.20: *Distance between Mg$_A$ ion and D115 during proton transfer by Lys 206 in Model B* - start and end points of simulations and corresponding structures indicated by double headed arrows
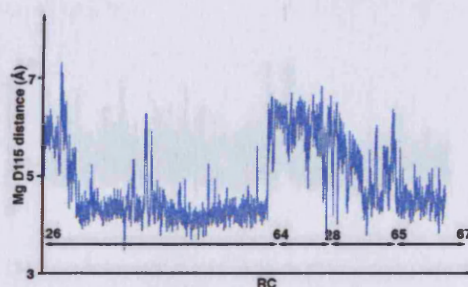

Figure B.21: *Distance between Mg$_A$ ion and D115 during proton transfer by Lys 206 in Model C* - start and end points of simulations and corresponding structures indicated by double headed arrows

B.2.2   Interaction between Asp 116 and Mg$_A$ during protonation by Lys 206

Figure B.22: *Distance between Mg$_A$ ion and D116 during proton transfer by Lys 206 in Model A* - start and end points of simulations and corresponding structures indicated by double headed arrows
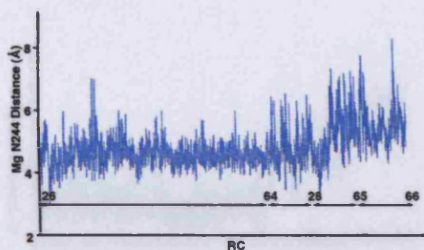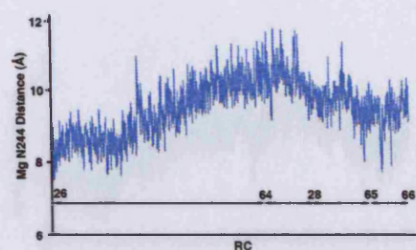

Figure B.23: *Distance between Mg$_A$ ion and D116 during proton transfer by Lys 206 in Model B* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure B.24: *Distance between Mg$_A$ ion and D116 during proton transfer by Lys 206 in Model C* - start and end points of simulations and corresponding structures indicated by double headed arrows

Figure B.25: *Distance between Mg_A ion and E119 during proton transfer by Lys 206 in Model A - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure B.26: *Distance between Mg_A ion and E119 during proton transfer by Lys 206 in Model B - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure B.27: *Distance between Mg_A ion and E119 during proton transfer by Lys 206 in Model C - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure B.28: *Distance between Mg$_B$ ion and N244 during proton transfer by Lys 206 in Model A* - start and end points of simulations and corresponding structures indicated by double headed arrows
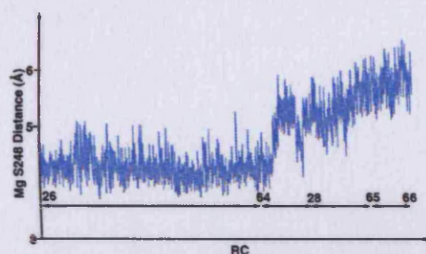
Figure B.29: *Distance between Mg$_B$ ion and N244 during proton transfer by Lys 206 in Model B* - start and end points of simulations and corresponding structures indicated by double headed arrows
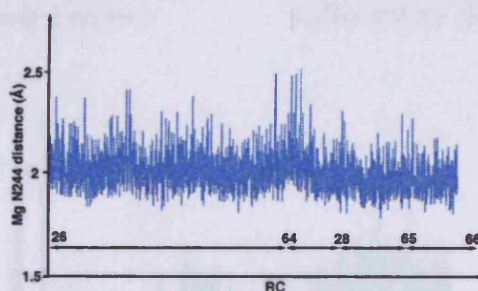
Figure B.30: *Distance between Mg$_B$ ion and N244 during proton transfer by Lys 206 in Model C* - start and end points of simulations and corresponding structures indicated by double headed arrows

B.2.3 Interaction between Glu 119 and Mg$_A$ during protonation by Lys 206

## B.2.4 Interaction between Asn 244 and Mg$_B$ during protonation by Lys 206

B.2.5 Interaction between Ser 248 and Mg$_B$ during protonation by Lys 206

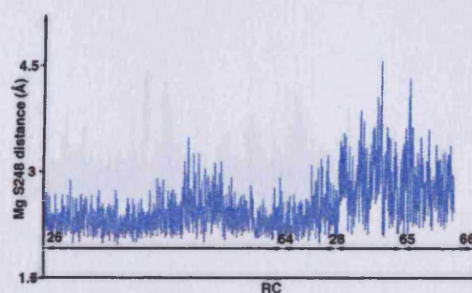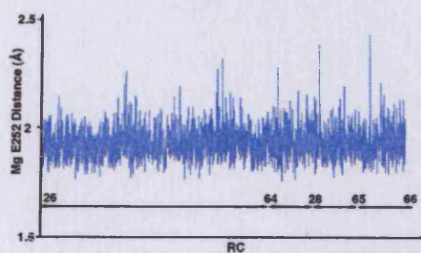## B.2.6 Interaction between Glu 252 and Mg$_B$ during protonation

Figure B.31: *Distance between Mg_B ion and S248 during proton transfer by Lys 206 in Model A - start and end points of simulations and corresponding structures indicated by double headed arrows*
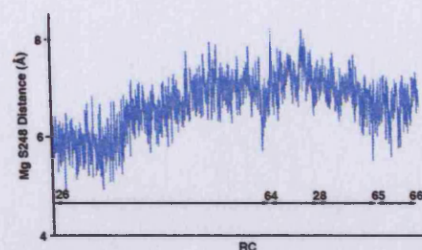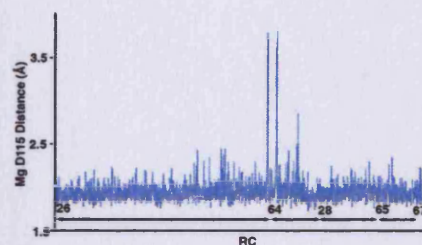

Figure B.32: *Distance between Mg_B ion and S248 during proton transfer by Lys 206 in Model B - start and end points of simulations and corresponding structures indicated by double headed arrows*
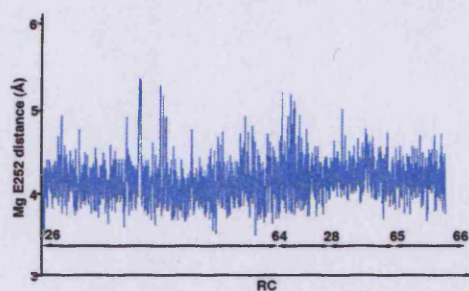

Figure B.33: *Distance between Mg_B ion and S248 during proton transfer by Lys 206 in Model C - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure B.34: *Distance between Mg_B ion and E252 during proton transfer by Lys 206 in Model A - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure B.35: *Distance between Mg_B ion and E252 during proton transfer by Lys 206 in Model B - start and end points of simulations and corresponding structures indicated by double headed arrows*

Figure B.36: *Distance between Mg_B ion and E252 during proton transfer by Lys 206 in Model C* - start and end points of simulations and corresponding structures indicated by double headed arrows