APPLICATION OF COMPUTATIONAL METHODS IN ELUCIDATING THE ISOMERIZATION STEP IN THE BIOSYNTHESIS OF COUMARINS.

A thesis submitted in fulfilment of the requirements for the degree

Of

DOCTOR OF PHILOSOPHY
IN CHEMISTRY

of
RHODES UNIVERSITY, SOUTH AFRICA

Department of chemistry
Faculty of science

By

TENDAMUDZIMU TSHIWAWA

May 2018

RHODES UNIVERSITY
Where leaders learn

To:

the sun in my world Takalani Tshiwawa L.G.U

“So long as men breathe, or eyes can see,
So long lives this, and this gives life to thee”

W. Shakespeare, Sonnet 18

To:

my supervisors Dr K.A Lobb and Prof P. T Kaye

All clay is just a lump of soil except in the hands of a potter,
not only did you mould this project and my research skills,
but you also moulded my character.
With all my heart, thank you.

ABSTRACT

The identity of the enzyme(s) responsible for the biosynthetic transformation of cinnamic acid derivatives to important, naturally occurring coumarins has yet to be established. This study constitutes a high-level theoretical analysis of the possibility that a recently reported molecular mechanism of the synthesis of coumarins from Baylis-Hillman adducts, may provide a viable model for three critical phases in the biosynthetic pathway Particular attention has been given to the first of these phases:

i) *E*→*Z* isomerisation of the cinnamic acid precursor; ii) Cyclisation (lactonisation) to the hemi-acetal intermediate; and ii) Dehydration to afford the coumarin derivative.

In order to accomplish this analysis, an enzyme capable, theoretically, of effecting this *E*→*Z* isomerisation required identification, and its potential involvement in the transformation mechanism explored. Combined Molecular Mechanics and high-level Quantum Mechanical/DFT calculations were used to access complementary models of appropriate complexes and relevant processes within the enzyme active sites of a range of eleven Chalcone Isomerase (CHI) enzyme candidates, the structures of which were downloaded from the Protein Data Bank. Detailed B3LYP/6-31+G(d,p) calculations have provided pictures of the relative populations of conformations within the ensemble of conformations available at normal temperatures. Conformations of several protonation states of cinnamic acid derivatives have been studied in this way, and the results obtained showed that coupled protonation and deprotonation of (*E*)-*o*-coumaric acid provides a viable approach to achieve the *E*→*Z* isomerization. *In silico* docking of the B3LYP/6-31+G(d,p) optimized (*E*)-*o*-coumaric acid derivatives in the active sites of each of the candidate CHI enzymes (CHI) revealed that (*E*)-*o*-coumaric acid fits well within the active sites of *Medicago Sativa* CHI crystallographic structures with 1FM8 showing best potential for not only accommodating (*E*)-*o*-coumaric acid , but also providing appropriate protein active site residues to effect the simultaneous protonation and deprotonation of the substrate , two residues being optimally placed to facilitate these critical processes. Further exploration of the chemical properties and qualities of selected CHI enzymes, undertaken using High Throughput Virtual Screening (HTVS), confirmed 1FM8 as a viable choice for further studies of the enzyme-catalysed *E*→*Z* isomerization of (*E*)-*o*-coumaric acid. A molecular dynamics study, performed to further evaluate the evolution of (*E*)-*o*-coumaric acid in the CHI active site over time, showed that the ligand in the 1FM8 active site is not only stable, but also that the desired protein-ligand interactions persist throughout the simulation period to facilitate the *E*→*Z* isomerization. An integrated molecular orbital and molecular mechanics (ONIOM) study of the 1FM8-(*E*)-*o*-coumaric acid complex, involving the direct protonation and deprotonation of the ligand by protein residues; has provided a plausible mechanism for the *E* → *Z* isomerization of (*E*)-*o*-coumaric acid within the 1FM8 active site; a transition state complex (with an activation energy of *ca*. 50 kCal.mol^{-1}) has been located and its connection with both the (*E*)- and (*Z*)-*o*-coumaric acid isomer has been confirmed by Intrinsic Reaction Coordinate (IRC) calculations. More realistic models of the 1FM8-(*E*)-*o*-coumaric acid complex, with the inclusion of water solvent molecules, have been obtained at both the QM/MM and adaptive QM/MM levels which simulate the dynamic active site at the QM level. The results indicate that the simultaneous protonation and deprotonation of (*E*)-*o*-coumaric acid within the CHI enzyme is a water-mediated process – a conclusion consistent with similar reported processes. Visual inspection of the 1FM8-(*Z*)-*o*-coumaric acid complex reveals *both* the necessary orientation of the phenolic and carboxylic acid moieties of the (*Z*)-*o*-coumaric acid *and* the presence of appropriate, proximal active site residues with the potential to permit catalysis of the subsequent lactonisation and dehydration steps required to generate coumarin.

ACKNOWLEDGEMENTS

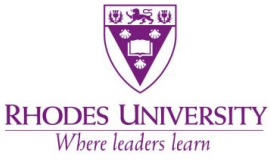
I am sincerely grateful to my supervisor and mentor Doctor Kevin Lobb for his tireless effort in teaching me the computational skills required for this work, for providing excellent computational facilities, and above all his pure kindness and patience. Great appreciation to Adele, Jess and Claire Lobb for the time they sacrificed for me during the summer holiday so I can finish up my work, and for hosting me on Christmas day.

I would also like to express my gratitude to my co-supervisor Professor Perry Kaye, for his guidance and unwavering faith in us seeing this project to completion.

I appreciate and acknowledge the financial assistance from the National Research Foundation (NRF) without which none of this would have been possible. The office of the dean (in particular Prof Tony Booth) in the faculty of science at Rhodes University also funded the extended time I needed to finish writing up this thesis, for that I am truly thankful.

For the computational facilities, gratitude's to my host university (Rhodes University, in particular the chemistry department), for the Perkin server made available to me for this project. The bulk of the computational work in this project was possible because of the computational time made available to our research group by the Center of High Performance Computing in Cape Town (CHPC), and therefore I am highly grateful to the CHPC.

Great gratitude's also to the colleagues in my research group for their great support, In particular Arthur Sarron for the chalcone derivatives used in the High Throughput Virtual Screening, Vuyani Moses for assisting me with the CHARMM scripts, Luxolo Joji for borrowing me this computational time when necessary, Faez Khan for helping me out with GROMACS, Lester Sigauke for helping me trouble shoot Python scripts, and Kamogelo Mafokwana for the emotional support.



DEDICATION

I dedicate this thesis to:

- kufhondo kwanga Takalani Tshiwawa
- my late mother Nyamutshagole Tshiwawa
- my mentor and supervisor, Dr K. Lobb and his family
- my co-supervisor Prof P. Kaye
- and my family and friends, especially my loving brothers Emmanuel Manyaga, Mutshidzi Manyaga, and my best friend Ndamulelo Ndou.

They are my inspiration every day.

ABBREVIATIONS

CHI – chalcone isomerase enzyme

QM – Quantum Mechanics

MM – Molecular Mechanics

QM/MM – hybrid Quantum Mechanics and Molecular Mechanics

MD – Molecular Dynamics

MO – Molecular Orbital

HOMO – Highest occupied molecular orbital

LUMO – Lowest occupied molecular orbital

QSAR-Quantitative structure-activity relations

HF- Hartree Fock method

MP - Möller-Plesset (MP) perturbation theory

DFT - Density Functional Theory

HTVS – High Throughput Virtual Screening

ONIOM – Our own n-layered integrated molecular orbital and molecular mechanics

TABLE OF CONTENTS

Contents

CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction to the chapter

The study focuses on the E → Z isomerization of *o*-coumaric acid by isomerase enzymes, and this chapter outlines the theoretical background and literature review relevant to the study. This includes a review of the properties of coumarins, the importance of the bioorganic E → Z transformation, details of a computational study of molecules in general, and an introduction to the computational approaches utilized in this study.

1.2 Coumarins

Coumarins are an extensive class of phenolic compounds isolable from a wide range of natural products; they have also been detected in bacteria and fungi and as well as animal sources (1,2). They were first reported and isolated in the 1820's, when they were recognised as the hay-like, sweet aroma of the Tonka bean (1–3). In higher plants coumarins contribute to the resistance by plants against a wide variety of problems, in particular that of phytopathogens or abiotic stresses (such as oxidative stress) and/or possibly chemical regulation including that of hormones (3).

The coumarin structure consists of a benzene ring fused to an α-pyrone ring (Figure 1.1). Many coumarins of natural product origin are substituted with oxygen-containing substituents (such as methoxy or hydroxy groups) at one or more of the six substituent positions.

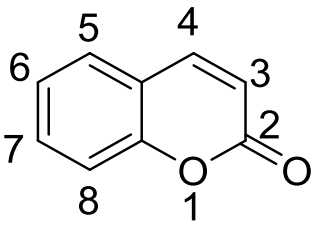
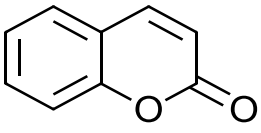
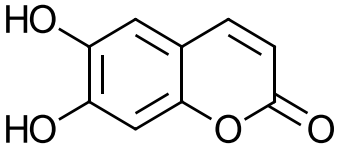
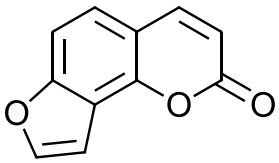
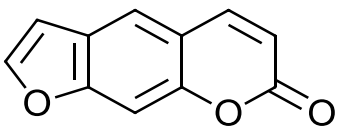
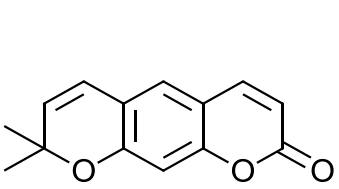
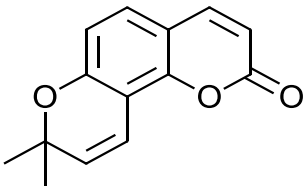
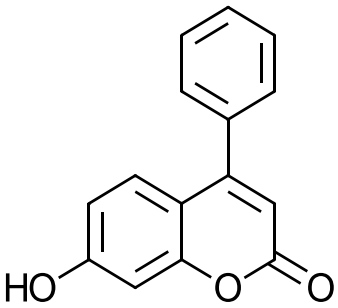
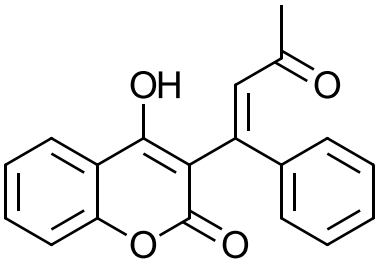


Figure 1.1. General structure of coumarin and the numbering utilized in this study.

Coumarins are so structurally diverse that they are now divided into different categories, examples of which are simple coumarins (including simple substituents), furanocoumarins (with additional fused 5-memberd furan ring), pyranocoumarins (with the fused 6-membered pyran ring), phenylcoumarins (where a substituent is an aromatic ring) and the pyrone-substitued coumarins (where substitution is specific to positions 3 and 4 on the coumarin) (Table 1.1) (3–6).

Table 1.1. Some examples of classification of coumarins.

Classification	Structural features	Examples	
Simple coumarins	Hydroxylated, alkoxylated or alkylated on the benzene ring	 coumarin	 6,7-dihydroxycoumarin
Furanocoumarins	5-membered furan ring attached to benzene ring; ring fusion may be angular or linear	 angelicin	 psoralen

Pyranocoumarins	6-membered pyran ring attached to benzene ring; ring fusion may be angular or linear	 xanthyletin  seselin
Phenylcoumarins	Phenyl substitution present on coumarin	 7-Hydroxy-4-phenylcoumarin
Pyrone-substitued coumarins	Substitution on pyrone ring,	 warfarin

Coumarins where used are used in diverse fields ranging from biology and medicine (with a multitude of bioactive coumarins), and polymer science (where it is the reactivity of coumarin that is of importance) to the field of laser dyes (where characteristics of the excited molecule come into play). They are also used in perfumes and cosmetics, due to their characteristic aromas. As such they also contribute to the taste or smell of particular cigarettes and alcoholic beverages (7,8). The fast development of resistance by pathogens and cancer cells to drugs that are currently in clinical use is a worrying development in medicine. It has become imperative to search for new drugs with activity against disease. Biological activities of coumarin derivatives include antitumor, anti-inflammatory, antioxidant, antimicrobial, anti-HIV, and more (1,9,10). Due to their diverse biological activities the study of the biological formation of coumarins is also significant.

Cancer in its various forms comes with a high cost both in terms of treatment and fatalities, accounting for 8.8 million deaths in 2015 (11). 7-Hydroxycoumarin (Figure 1.2a) also known as umbelliferone has attracted attention due to its antitumor activities (12–14). Coumarin and 7-hydroxycoumarin, its major biotransformation product in humans successfully, inhibits several types of human tumour cells *in vitro* (6). *In vivo,* coumarin also has antitumor activities in various types of cancer cells (1,15–17). Several other coumarins, such as esculetin (Figure 1.2b) and scopoletin (Figure 1.2c), have been reported to have anti-inflammatory activities (5,18–20) which appear to be associated with their antioxidant activities (21–24). Coumarins have also shown diverse antimicrobial activities, such as activity against both Gram-positive and negative bacteria including *Staphylococcus aureus* and *Bacillus cereus* (in the former case) and *Escherichia coli* and *Pseudomonas aeruginosa* (in the latter) (5,25,26). An example of a coumarin with antimicrobial activities is isoscopoletin (Figure 1.2d) (5,25,26).

Vitamin K is essential in many respects within the human body, such as its requirement as a cofactor in the complete synthesis of coagulation proteins, and is comprised by several structurally similar, fat-soluble compounds. 4-Hydroxycoumarins (Figure 1.2e) have been used for over 50 years as vitamin K antagonists. Most coumarins that are anticoagulants are derivatives of 4-hydroxycoumarin (27–29).

An estimated 0.8% of adults aged 15-49 are living with HIV worldwide (30,31). The burdens of the epidemic continue to vary, and over 35 million people have died from HIV since its first discovery. Several phytochemical and synthetic coumarins interact in a unique way with the HIV-1 RT enzyme, and this inhibition provides an opportunity for action where drug resistance is becoming evident (32–35). A variety of coumarins have been found to not only efficiently inhibit the HIV-1 RT enzyme but also the protease and integrase enzymes critical to the HIV-1 life cycle (33,36–38). Coumarins with HIV-1 action have a range of structural features such as the pyranocoumarin khellactone and its derivatives, tetracyclic dipyranocoumarins, multimer coumarins and others. Another example of a potent anti-HIV coumarin is the pyranocoumarin suksdorfin (Figure 1.2f), isolated from the plant *Lomatium sucksdorfii* (32). This compound has shown promising activity *in vitro*.

There are many routes to the syntehsis of coumarins, and these include the Perkin, von Pechman, Reformatsky, Knoevenagel and Wittig reactions (39,40). Several studies have been undertaken regarding the synthesis of coumarins and of testing their biological activities at

Rhodes University (34,41–45). Other studies have looked at the mechanism of formation and the subsequent reactivity of coumarins. An example includes a ^{1}H NMR-based kinetic study of the acid catalysed lactonization of salicylaldehyde-derived Baylis-Hillman adducts to form coumarins (41). The mechanism proposed for this transformation has implications in the study of the biosynthetic formation of coumarin.

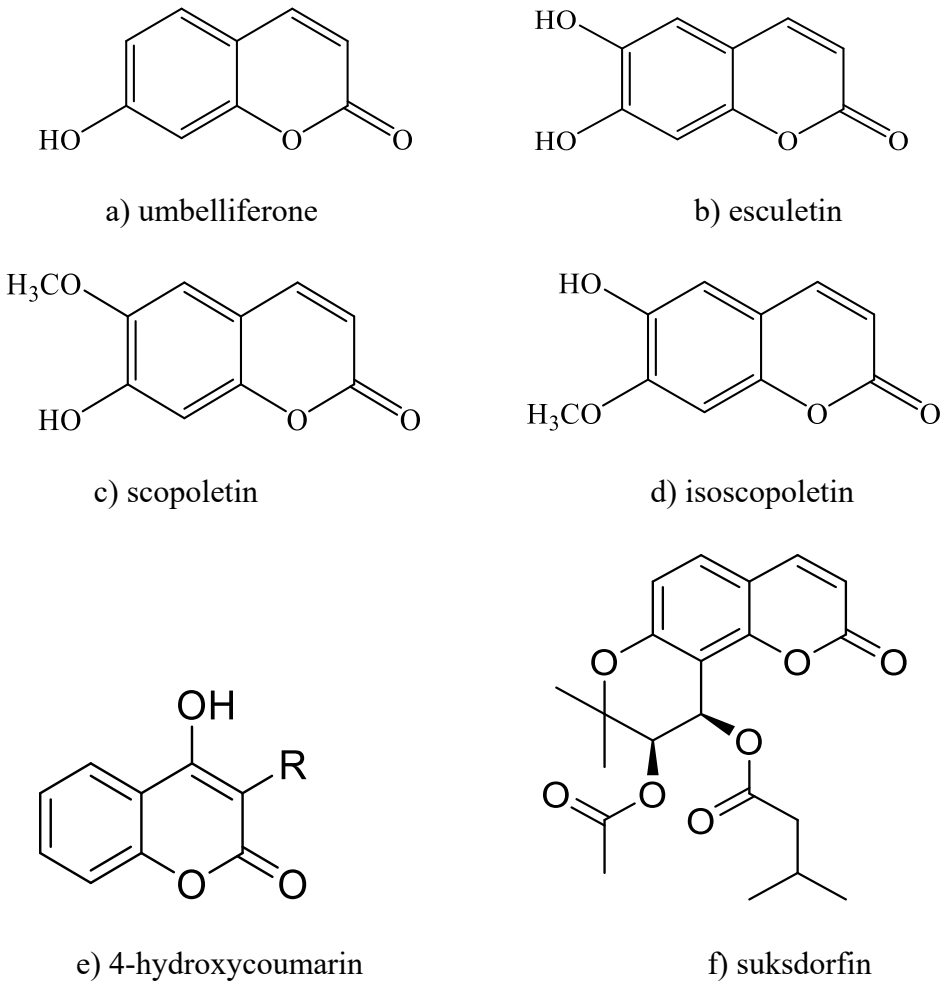


Figure 1.2. Biologically active coumarins from literature.

1.3 Bioorganic transformation of cinnamic acids to coumarins

In the reported study of lactonization of Baylis-Hillman adducts to form coumarins (41), the authors hypothesised that a similar reaction might occur in nature as anenzyme-catalyzed process. Figure 1.3 shows the putative structures suggested by the authors of the paper to be involved in the enzyme-catalysed E→Z isomerization of o-coumaric acid, and the subsequent lactonization of the Z-isomer to form coumarin (41).

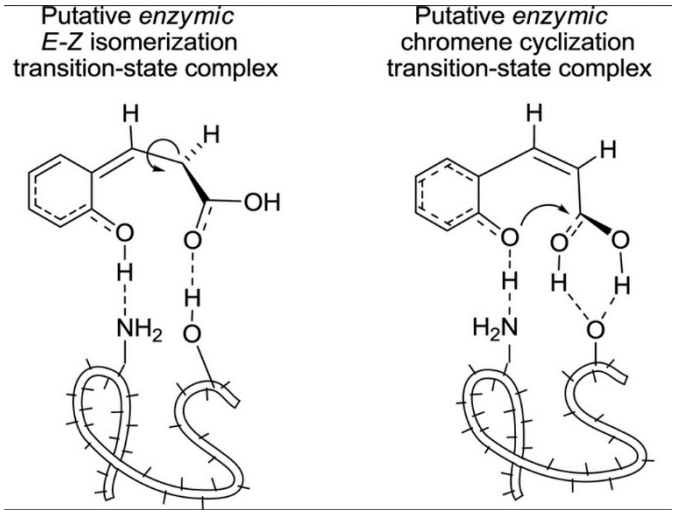


Figure 1.3. Putative structures involved in a suggested enzyme catalysed *E* → *Z* isomerisation of *o*-coumaric acid and the subsequent lactonization of the *Z*-isomer to form coumarin (41)**. Reprinted (adapted) with permission from (*J. Org. Chem.*, 2016, 81, 109). Copyright (2016) American Chemical Society.**

It is expected that the catalysis of these transformations to coumarin by enzymes will be very efficient. Although coumarins are largely found as plant secondary metabolites, the biological transformation of substituted cinnamic acids to coumarins is not well elaborated in the literature (46,47). Most of the P450-dependent enzymatic steps required for this process are still hypothetical (3). This study gives particular attention to the three critical phases proposed for the biological transformation of cinnamic acid precursors to coumarins:

i) *E* → *Z* isomersation of the cinnamic acid precursor;
ii) Cyclisation (lactonization) to the hemi-acetal intermediate; and
iii) Dehydration to afford the coumarin derivative.

1.3.1 Biosynthesis of Coumarins

In an extensive article written by Bourgaud and colleagues in 2006, the key steps in the biosynthesis of coumarins in plants were reviewed (3). There are many enzymes involved in the bioorganic transformation of cinnamic acids to coumarins, and it is clear that several of these enzymes are still hypothetical (indicated in Figure 1.4 with a question mark), particularly those involved in the *E*→*Z* isomerization of coumaric acid (2-hydroxycinnamic acid) and in the lactonization step (Figure 1.4).

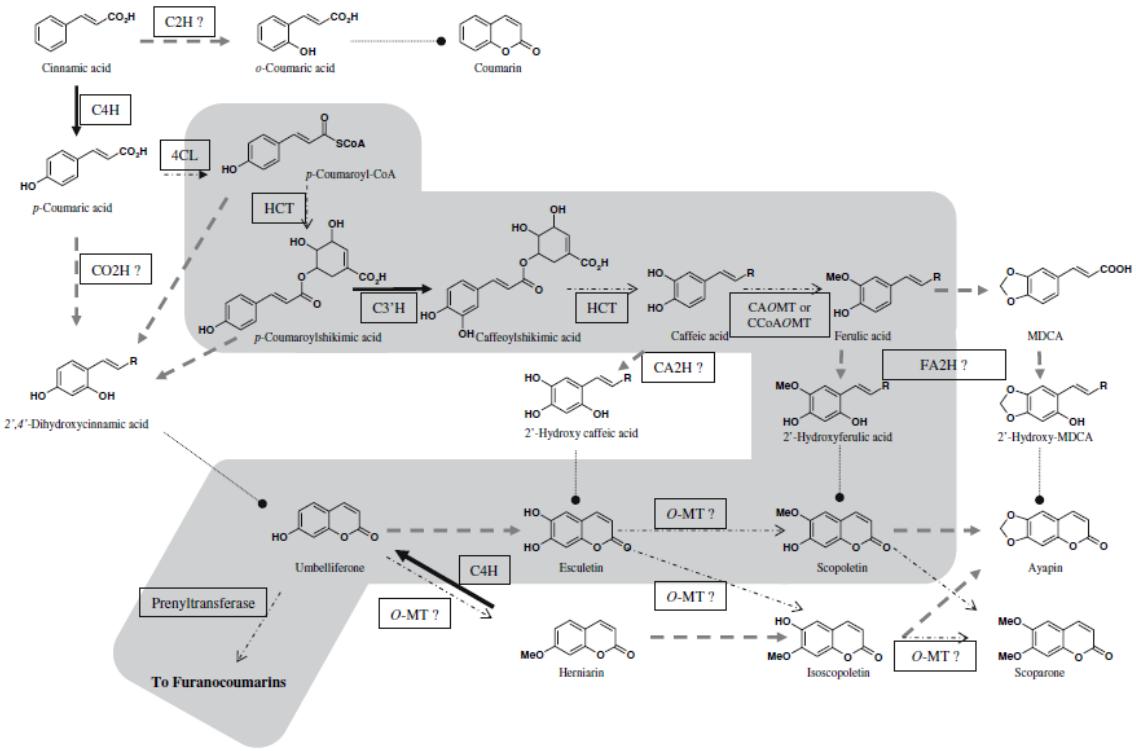
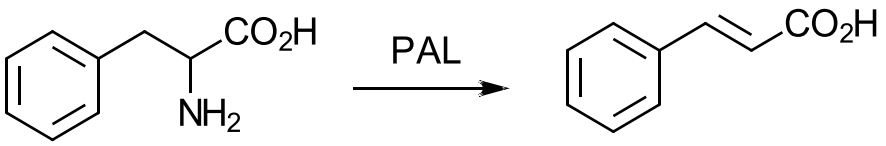


Figure 1.4. Biosynthetic pathway of cinnamic acid to coumarin. Several hypothetical enyzmes are included, but these are assigned with question marks. R = CO_2H or CO-SCoA; C2H, cinnamic acid 2-hydroxylase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CO2H, 4-coumaric acid 2-hydroxylase; HCT, hydroxycinnanoyl-transferase; CA*O*MT, caffeic acid *O*-methyltransferase; CCoA*O*MT, caffeoyl CoA *O*-methyltransferase; CA2H, caffeic acid 2-hydroxylase; FA2H, ferulic acid 2-hydroxylase; MDCA2H, methylene edioxycinnamic acid 2-hydroxylase; *O*-MT, *O*-methyltransferase. (——►) P450, (– —►) putative P450, (·····►) non-P450 enzyme, (·····●) lactonisation (3)**. From Bourgaud *et al. Phytochem Rev.* 2006, 5,293, used with permission.**

The literature review will therefore explore what is known about this transformation in the biological context, with the aim of highlighting what is not yet known. Isolated enzymes, which

have been identified to be crucial in the coumarin biosynthetic pathway (48) include phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate:CoA ligase (4CL), 4-coumaroyl CoA 2′-hydroxylase (C2′H) (49), psoralen synthase, angelicin synthase (50,51) and bergaptol O-methyltransferase (BMT) (52).

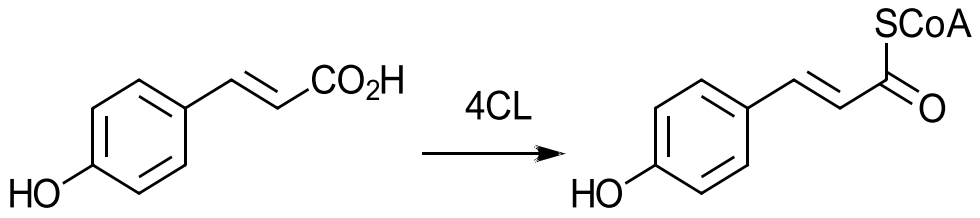
1.3.1.1 Phenylalanine ammonia lyase (PAL)



phenylalanine (*E*)-cinnamic acid

The question is raised within the context of the biosynthesis of coumarins as to how the substituted cinnamic acid precursors are provided. Phenylalanine ammonia lyase (PAL) catalyses the transformation of phenylalanine to (*E*)-cinnamic acid and ammonia (53–55). PAL was first discovered in 1961 by Koukol and Conn (56–59), and has been isolated and characterized from higher plants (60–62), fungi (56,63), and prokaryotes (64–66). In recent studies PAL has been considered for the treatment of phenylketonuria (a genetic metabolic disease characterised by elevated, neurotoxic levels of phenylalanine resulting from a decrease or loss of phenylalanine hydroxylase) (67–69). It is also a potential therapeutic enzyme for cancer treatment due to its significant cytotoxic effects towards specific cells (70,71). This enzyme has also been implicated in inter-species communication between cotton and maize in a recent study of PAL enzymatic activity together with gene expression induction in damaged and undamaged seedlings (72). There are several theoretical studies concentrating on PAL (73,74), and recent computational work involving docking of compounds to *Av*-PAL highlighted the important role of Tyr78 in the enzyme reaction (75).

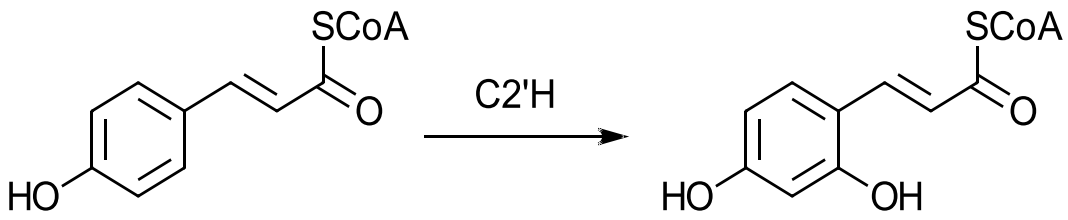
1.3.1.2 4-Coumarate:CoA ligase (4CL)



p-coumaric acid *p*-coumaroyl CoA

4-coumarate:CoA ligase (4CL) catalyses hydroxycinnamate CoA ester formation, and plays an essential role at the divergence point from general phenylpropanoid metabolism to the major branch pathway of coumarins (76–78). 4CL belongs to the family of ligases responsible for forming carbon-sulphur bonds (acid-thiol ligases). It catalyses the transformation of coumaric acid and cinnamic acids such as caffeic acid and ferulic acid into their corresponding CoA thiol esters. These are used in the biosynthesis of several phenylpropanoid-derived compounds, such as flavonoids, isoflavonoids, lignin, suberins, and also in the biosynthesis of coumarins (79,80). Recently three putative 4CL genes (Pp4CL1, Pp4CL7, Pp4CL10) were isolated from *Peucedanum praeruptorum*. Biochemical characterisation of the recombinant proteins revealed that *p*-coumaric acid and ferulic acid are used by Pp4CL1 as two major substrates for coumarin biosynthesis in *P. praeruptorum* (48). The study included the homology modelling and docking of Pp4CL1 with adenosine 5'-(3-(4-hydroxyphenyl)propyl)phosphate (APP) (48). There are several theoretical studies in the literature on 4CL (81–83), and these include a study of enzyme mechanismbased on the crystal structure of *Populus tomentosa* 4CL (84).

1.3.1.3 4-Coumaroyl CoA 2′-hydroxylase (C2′H)



p-coumaroyl CoA 2,4-dihydroxycinnamic acid

4-Coumaroyl CoA 2′-hydroxylase (C2′H) is an enzyme that is involved in the conversion of *p*-coumaroyl CoA to 2,4-dihydroxycinnamic acid. 2,4-Dihydroxycinnamic acid is subsequently

transformed into umbelliferone following the spontaneous closure of the lactone ring by a non-enzymatic reaction that occurs under acidic or neutral conditions (49). The activity of C2′H in the synthesis of umbelliferone was demonstrated in a study of the 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* (49).

1.3.2 Enzyme mediated transformations involving coumarins

There are some known enzymes that take coumarins from biosynthesis to further functionalization. Some examples of these are the cytochrome P450 enzymes, psoralen synthase and angelicin synthase (which transform substituted coumarins into angular furanocoumarins) and bergaptol O-methyltransferase, which simply imparts methoxy functionality to a specific coumarin.

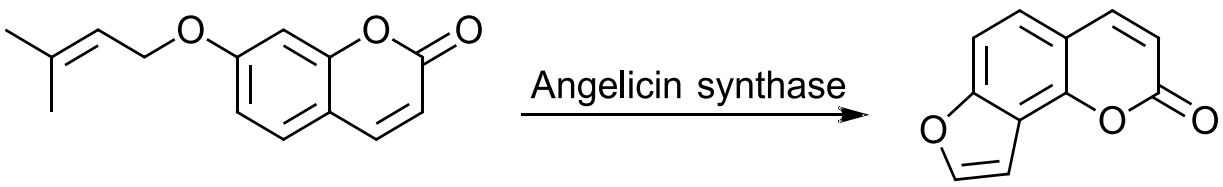
1.3.2.1 Psoralen synthase and Cytochromes

marmesin psoralen

Psoralen synthase is an enzyme that is responsible for the conversion of marmesin to psoralen. Psoralen is a linear furanocoumarin that is often found accumulated on the surface of wax tissues of plants such as *Ammi majus* (50,85–88), *Petroselinum crispum* (89,90) and *Coronilla* genera (50). The surface coating of psoralen is considered essential for protection against fungal infections and herbivore attacks (50).

Cytochromes P450 (CYPs) are super-family of proteins containing heme as a cofactor, and therefore they are hemoproteins (91,92). In recent studies psoralen synthase was identified as a member of the CYP71 family, which is the largest plant CYP family and contains numerous duplicated genes (50). In another study psoralen synthase was considered as one of the two subclades functionally assigned to the biosynthesis of furanocoumarins, and shown to be part of two distinct clades within the CYP71AJ subfamily (93).

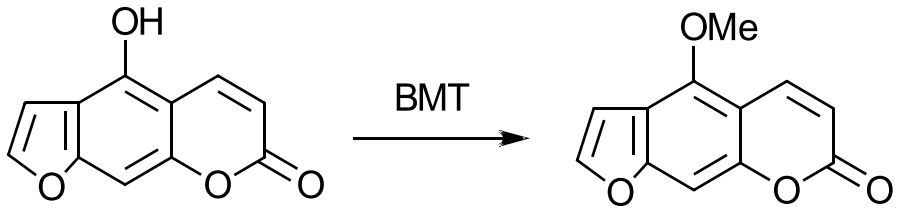
1.3.2.2 Angelicin synthase



8-prenylumbelliferon angelicin

Angelicin synthase, like psoralen synthase, is a cytochrome P450 enzyme isolated from *Pastinaca sativa* (parsnip). This is angelicin synthase was the first isolated P450 gene committed to angular furanocoumarin biosynthesis (51). It catalyses the conversion of 8-prenylumbelliferone to the angular furanocoumarin commonly known as angelicin. A study of this enzyme investigated expression of CYP71AJ4-like genes and their relation to the evolution of an angular furanocoumarin pathway (51). In the molecular cloning and functional characterization study of psoralen synthase, angelicin synthase was suggested to have evolved from psoralen synthase. Analogy modelling and docking were used to define the conditions for high affinity substrate binding and to predict the minimal requirements to accommodate, within the active site, a related angular furanocoumarin, (+)-columbianetin (which is a minor product of enzyme catalysis, whilst also remaining a competitive inhibitor of the angelicin synthase) (51).

1.3.2.3 Bergaptol O-methyltransferase (BMT)



bergaptol bergapten

Bergaptol O-methyltransferase (BMT) is an enzyme that catalyses the transformation of bergaptol to bergapten. Methoxylated psoralens such as bergapten or xanthotoxin are accumulated in plants belonging to the *Apiaceae* or *Rutaceae* family as the final products of their furanocoumarin biosynthesis (52,94–96). Studies on BMT include the constructive expression of bergaptol O-methyltransferase in *Glehnia littoralis* cell cultures (94) and furanocoumarin biosynthesis in *Ammi majus* L, which included the cloning of bergaptol O-methyltransferase (50,52).

1.3.3 Chalcone isomerase enzymes

Isomerase enzymes are bioorganic catalysts that facilitate the structural transformation of molecules from one isomer to another. The product of the isomerisation has the same molecular formula as its substrate but differs in spatial arrangement or bond connectivity. Isomerase enzymes are used in many biochemical pathways, such as the citric acid cycle and glycolytic pathway (97,98). Isomerase enzymes are classified into racemases, cis-trans isomerases, intramolecular oxidoreductases, intramolecular transferases, and intramolecular lyases (99–101). As will become apparent later in the introduction, the possible involvement of chalcone isomerase-type lyases in coumarin biosynthesis was to be explored.

1.3.3.1 Introduction to chalcone isomerase enzymes

Chalcone isomerase enzymes (CHI), also known as chalcone-flavone isomerases, are plant enzymes that catalyse the intramolecular cyclization of chalcones to flavanones (Figure 1.6). CHI's were first isolated from soybeans. In plants, the biosynthesis of flavones is essential because the flavonoid pigment functions in pollination and seed dispersal (they attract insects and birds by their colours) (102,103). Flavones are also essential for protection in plants as they absorb UV light and act as sunscreen to protect plant DNA from UV damage (102,103). The intramolecular cyclization of chalcone to flavone with CHI follows second-order kinetics (k_{cat}/K_m) that approaches the diffusion-controlled limit with a turnover rate that exceeds the spontaneous conversion rate by 10^7-fold (102,103).

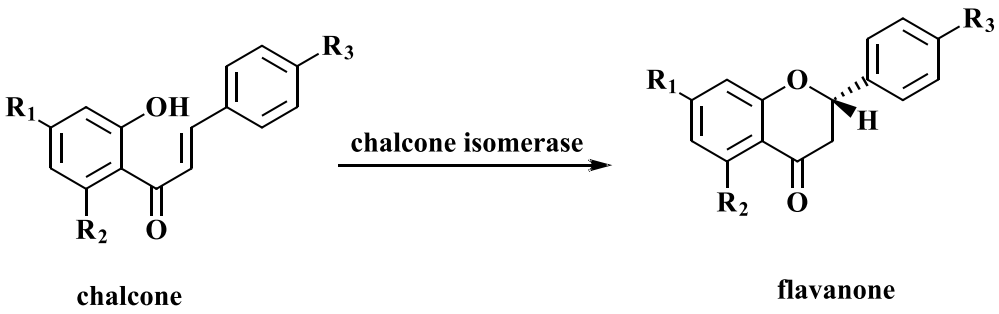


Figure 1.6. Chalcone isomerase catalysed transformation of chalcone to flavanone.

Most of the CHIs reported in the literature were isolated from either plants or bacteria (*Eubacterium ramulus*). The CHIs explored in this study are of plant origin with protein data bank (PDB) ids: 1FM7, 1FM8, 1JEP, 1EYQ, 1EYP, 4DOI, 4DOK, 4DOL, 4DOO, 1JXO (Y106F mutant), and 1JX1 (T48A mutant). The sets of X-ray coordinates of CHIs reported in this study, whose PDB name starts with the number '1', were isolated from the species *Medicago sativa*, and all those with a PDB name starting with the number '4' were isolated from the species *Arabidopsis thaliana*. These sets of X-ray coordinates of CHIs in the PDB include both the apoenzyme and enzyme–flavanone complexes. CHIs crystalize as dimeric structures for all pdbs, with each set of coordinates containing two active sites.

1.3.3.2 Chalcone isomerase enzymes from *Medicago sativa*

There is much literature on the study of CHIs (103–115). One of the most relevant studies for the current work is the study of CHI reaction mechanisms reported by Jez and colleagues (104) focused mainly on CHI 7,4-dihydroxyflavanone (PDB code; 1FM7), CHI 7,4-dihydroxyflavanone (PDB code; 1FM8), and CHI 4-dihydroxyflavanone (PDB code; 1JEP). Comparisons of the structural variation in flavanone binding showed that the CHI active site is dynamic and catalysis may be due to structural variation in the active site geometry (104). Figure 1.7 illustrates a crystal structure of CHI 7,4-dihydroxyflavanone (PDB code; 1FM8) complexed with a flavanone.

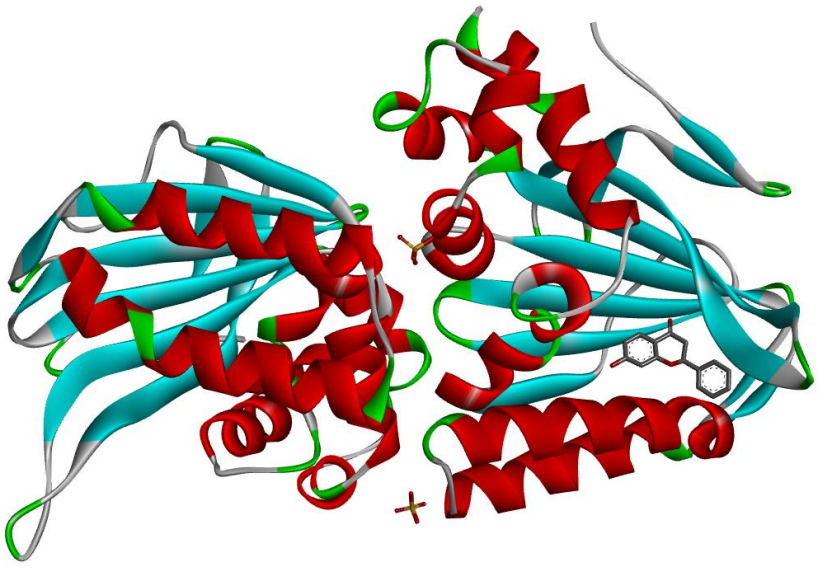


Figure 1.7. Crystal structure of CHI 7,4-dihydroxyflavanone complexed with 5,4′-dideoxyflavanone (PDB code; 1FM8).

Jez and colleagues proved that in the CHI catalysed formation of flavones from chalcone, a water molecule deprotonates the chalcone phenolic OH and the resulting oxyanion undergoes nucleophilic attack on the α,β-unsaturated double bond. Following this, a water molecule acts as a general acid to stabilize the enolate, resulting in the formation of a flav-3-en-4-ol intermediate that tautomerizes into the expected reaction product (Michael addition) Figure 1.8 (104). The structural and kinetic studies emphasized the importance of the protonation state of the substrate in the reaction and a proper docking pose (104).

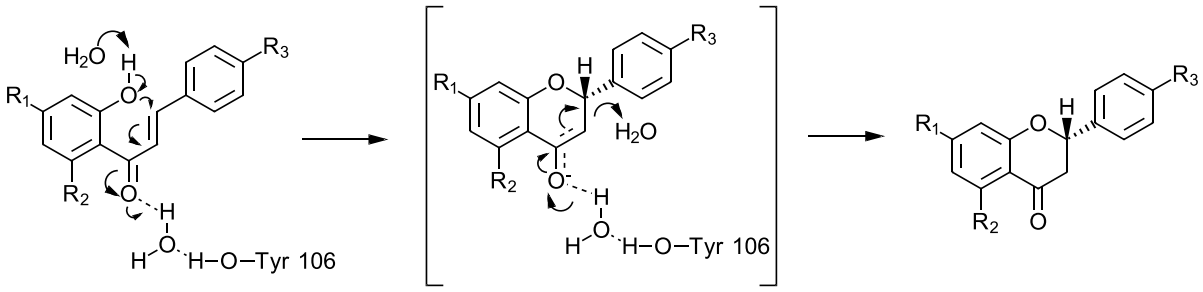


Figure 1.8. Cyclization reaction catalysed by CHI (104).

1.3.3.3 Chalcone isomerase enzymes from *Arabidopsis thaliana*

CHIs from *A. thaliana* that were considered in this study are those which were the object of a study by Ngaki and colleagues on the evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis (116). *A. thalina* contains five actively transcribed genes encoding CHI-fold proteins, three of which additionally encode the amino-terminal chloroplast-transit sequence (116). The expression profiles of the three CHI-fold proteins correlate with those of core fatty-acid biosynthetic enzymes. Their maximum expression occurs in seeds and coincides with increased fatty acid storage in developing embryos (116). These proteins are fatty-acid-binding proteins *in vitro* (116). Figure 1.9 shows an example of CHI from *A. thaliana* PDB code: 4DOI.

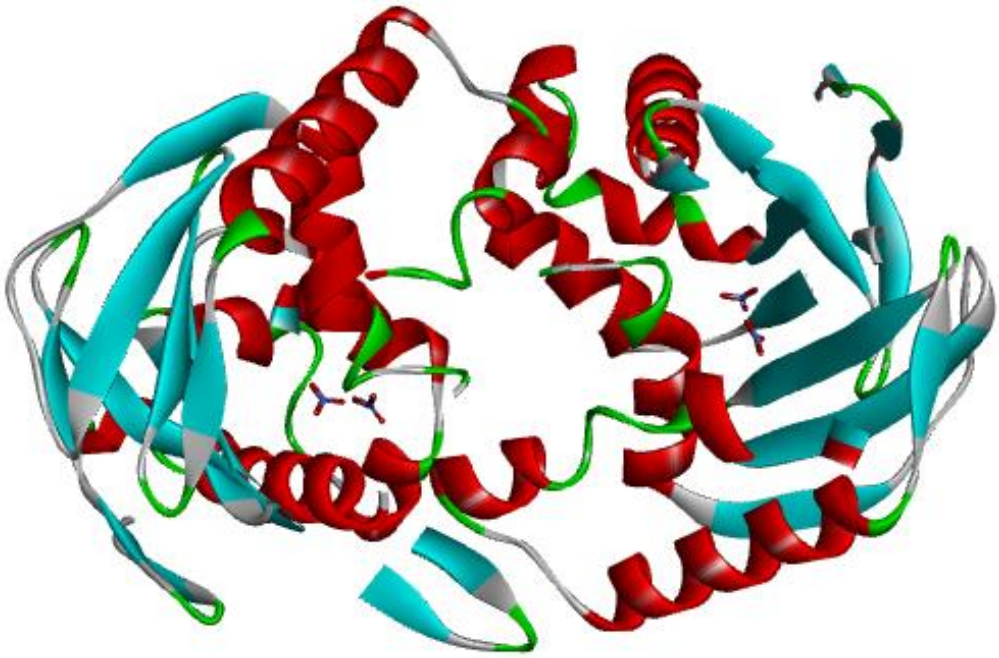


Figure 1.9. CHI 7,4-dihydroxyflavanone (PDB code; 4DOI).

The conversion of chalcone to flavanone with a CHI from *A. thaliana* involves a combination of electrostatic catalysis and water-mediated charge stabilization during a stereospecific Michael addition reaction (a reaction similar to those of other CHIs) (Figure 1.10) (116).

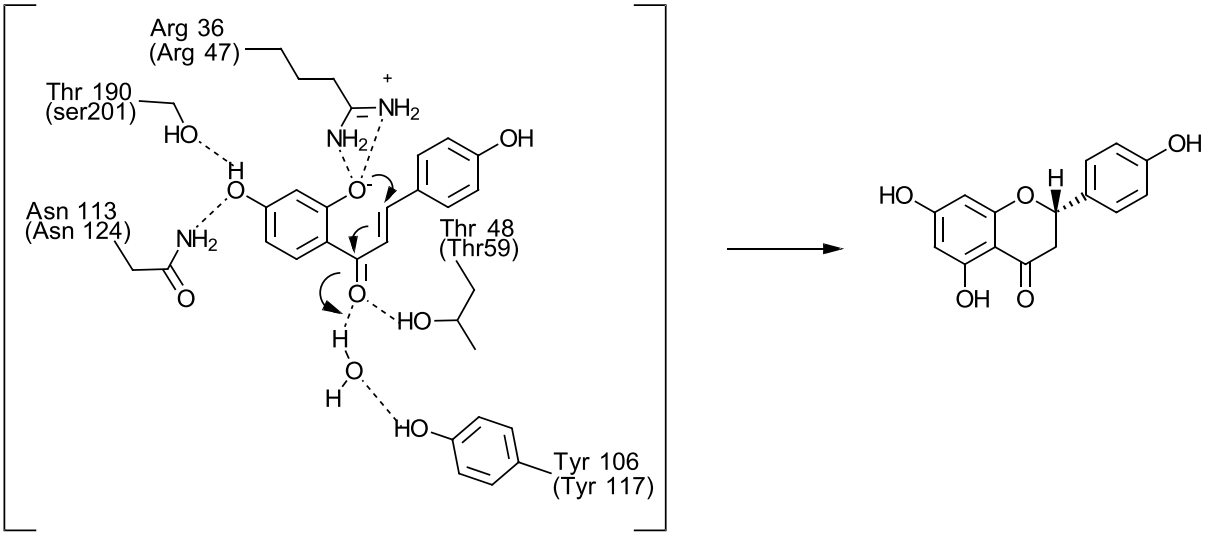


Figure 1.10. Cyclization reaction catalysed by CHI (116).

Tyr 106 seems to have an essential role in the catalytic action of CHIs; in the reaction mechanisms for CHI from *M. sativa* and CHI from *A. thaliana*, Tyr 106 is one of the catalytic residues involved in the stabilization of the intermediate.

1.3.3.4 Chalcone isomerase enzymes in this study

The current study focusses particularly on the E→Z isomerization of the Z isomer of coumaric acid, within the scheme of subsequent lactonization and dehydration to form coumarin (Figure 1.5). No enzymes responsible for these steps were found in literature and, therefore, known isomerase enzymes (which are capable of effecting transformations on chemically similar compounds) were used in this study.

The uncertainty in the biological transformation of cinnamic acid to coumarin is illustrated in Figure 1.5. Given the dynamic nature of the CHI active sites, the feasibility of CHI-like enzyme involvement, not only in lactonization of the chemically related (*Z*)-coumaric acid, but also in a concerted, prior E → Z isomerization step, was explored in this study.

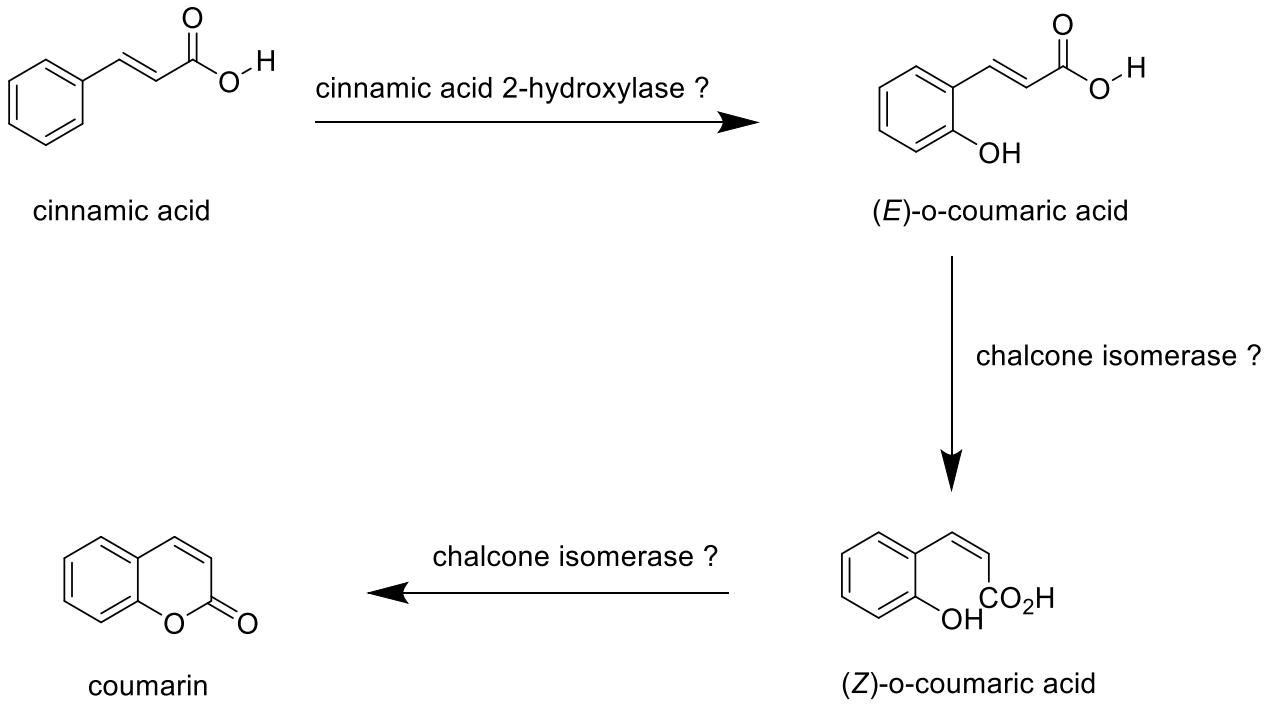


Figure 1.5. Proposed enzyme catalysed transformation pathway of cinnamic acid to coumarin. The enzymes are represented with a question mark because they are hypothetical.

1.4 Computational approach to the study

Computational chemistry studies the fundamental structures and properties of atoms, molecules, solids, and chemical reactions using theoretical chemistry approaches and calculations integrated into efficient computer programmes (117–121).

1.4.1 Importance of computational study in bioorganic transformations

The computational study of systems is important in chemistry as it rationalizes at a molecular level the chemical behaviour of the system. The predictions of chemical properties and reaction processes through computer programmes potentially allow cost saving by synthesising only the molecules that were predicted, through computational methods, to have the desired chemical properties and by testing experimentally only reaction mechanisms that are predicted to be practically feasible through computational methods. Virtual screening of chemical libraries provides a means of identifying compounds that complement targets of known biological structures instead of traditional high-throughput screening involving the experimental testing of large libraries of compounds, many of which might not bind effectively, thus saving time and resources (122–125). Pharmacophore modelling and Quantitative Structure Activity Relations (QSAR) methods can also be employed when a set of active ligand molecules are known and little or no structural information is available for the targets (123,126–130).

1.4.2 Computational approaches utilized in this study

Computational chemistry is based primarily on Molecular Mechanics and Quantum Mechanics approaches. Molecular Mechanics, Quantum Mechanics, and hybrid Quantum Mechanics/Molecular Mechanics approaches were utilized in this research project, and these are addressed in the following sections.

1.4.2.1 Molecular Mechanical (MM) methods

Molecular Mechanics models molecular systems using potential functions that are derived using classical mechanics, *i.e*., the calculation of the potential energy of the system using what are termed "force fields" (131–135). Molecular Mechanics is based on the following assumptions.

i) The nucleus itself is a perfect sphere
ii) The molecular bonds are treated as springs.
iii) The potential function relies on experimental parameters such as force constants and equilibrium bond-length bond-angle or dihedral angle values.
iv) The potential energy function is the sum of individual functions for bond stretching, angle bending and torsional energies and non-bonded interactions.

The force field is defined by the exact terms used in calculating this potential, and also with the exact parameters used for bond-lengths, bond-angles, and the torsion angles, and non-bonding parameters such as van der Waals and the electrostatic interaction parameters. These molecular interaction potentials are usually parameterized from quantum mechanical or experimental data (125,128,136). The potential energy obtained through Molecular Mechanics has no absolute meaning, but is useful in a relative context, such as for comparing different conformations of a molecule (131–135). MM force fields have an extremely reduced computational cost compared to Quantum Mechanical calculations thus making MM calculation possible for simulations on large systems (131–135). Some commonly used force field parameters are the AMBER (137), CHARMM (138), and GROMOS (139) force fields. In this research project, the force fields used were the CHARMM36 (C36) for molecular dynamics (MD) simulations with CHARMM and Amber03 force fields for MD simulations with GROMACS.

Because of their computational affordability MM force fields are methods of choice for simulations of bioorganic systems such as proteins. MM force fields are used in Computational Structure-Based Drug Discovery (CSBDD) which involves drug design based on the three-dimensional structure of a biomolecular target such as a protein (128,133). Molecular docking and Molecular dynamics are approaches to CSBDD which rely on MM force fields.

1.4.2.1.1 Molecular Docking

One of the most frequently used methods within CSBDD is Molecular Docking (140–142). Molecular Docking approaches simulate the molecular recognition process by predicting the preferred orientation of a particular molecule when it is bound in the form of a stable complex. This approach can be used to model the interaction between a small molecule and a protein at the atomic level. In Molecular docking, the pose (the position, the orientation and the conformation of the ligand in the protein) and the binding affinity of the ligand are determined with sampling methods and scoring schemes. In most docking approaches the location of the binding site is known before the docking process, and the receptor and the ligand can both be treated as rigid bodies (early elucidation), the ligand as flexible and receptor rigid (computationally affordable), or both receptor and ligand as flexible (more computationally expensive). A docking approach without knowledge of the binding site is referred to as blind docking. If the binding site is not known then cavity selection programmes or online servers such as GRID (143), POCKET (144), SurfNet (145), PASS (146), and MMC (147) can be utilized to determine possible binding sites. There are many simulation software packages available for molecular docking, but the ones selected for this study are AutoDock 4.2 (148) and AutoDock Vina (149,150) developed by the Scripps Research Institute. This software automates the docking of ligands to macromolecules using a Lamarkian Generic Algorithm (for Autodock 4.2) or the Iterated Local Search Global Optimizer Algorithm (for Vina) together with Empirical Free Energy Scoring Functions (150).

1.4.2.1.1.1 Reference docking studies from the literature

Docking is used extensively in rational drug design. An example of a docking study on coumarin derivatives is one in which the system 11a was docked into a receptor site of HIV-1 protease (PR) (34). The study revealed potential hydrogen-bonding interactions between proximate protein residues and both the coumarin and the azidothymide (AZT) moieties (Figure 1.11f). The *in silico* docking results also showed close correspondence between the docked conformation of compound 11a and the X-ray structure of the known inhibitor, ritonavir, in the active site of HIV PR. Docking of several derivatives (11a-11e) into the pocket site of HIV-1 reverse transcriptase (RT) revealed that the considered ligands exhibit potential hydrogen-bonding interactions with amino acid residues that line the pocket, and the coumarin

moiety occupies the same cavity as efavirenz, these observations showed the possibility of the considered compounds to act as non-nucleotide RT inhibitor.

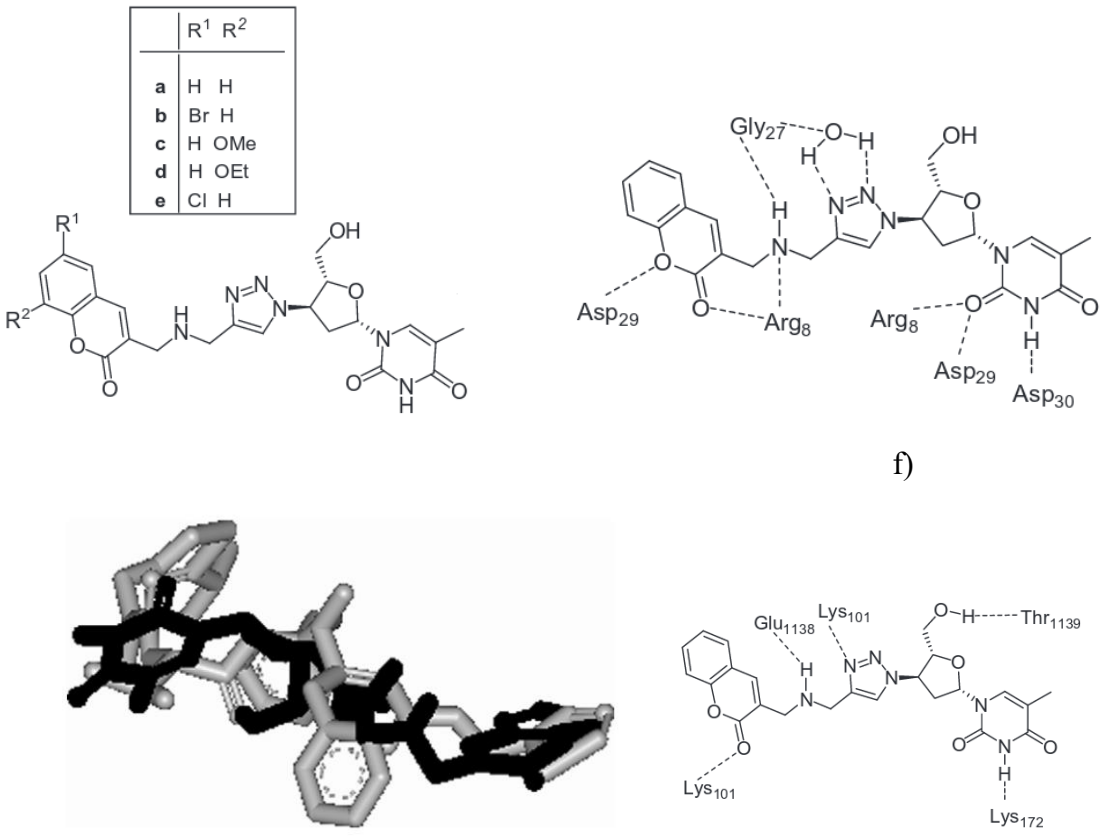


Figure 1.11. f) Schematic representation of potential hydrogen-bonding interactions (<4Å) between compound 11a and the residues in the receptor cavity of HIV-1 protease (PDB code: 1HXW). g) Overlay of the docked conformations of compound 2a (in black) and ritonavir (in grey, as found in the crystal structure) in the HIV-1 PR active site (PDB code: 1HXW). h) Schematic representation of hydrogen-bonding interactions (<4Å) between AZT-coumarin product 2a and the residues in the non-nucleoside pocket of HIV-1 reverse transcriptase (PDB code: 1lkw) (34)**. used with permission.**

This example illustrates the range of chemical information available from the molecular docking approaches.

Molecular docking is the least computationally expensive method of CSBDD. Because of its ability to predict binding interactions and orientations [very accurately in some cases (133)], Molecular Docking has found wide application in “rational drug design”, however the flexibility of the target binding site is frequently overlooked due to the computational expense

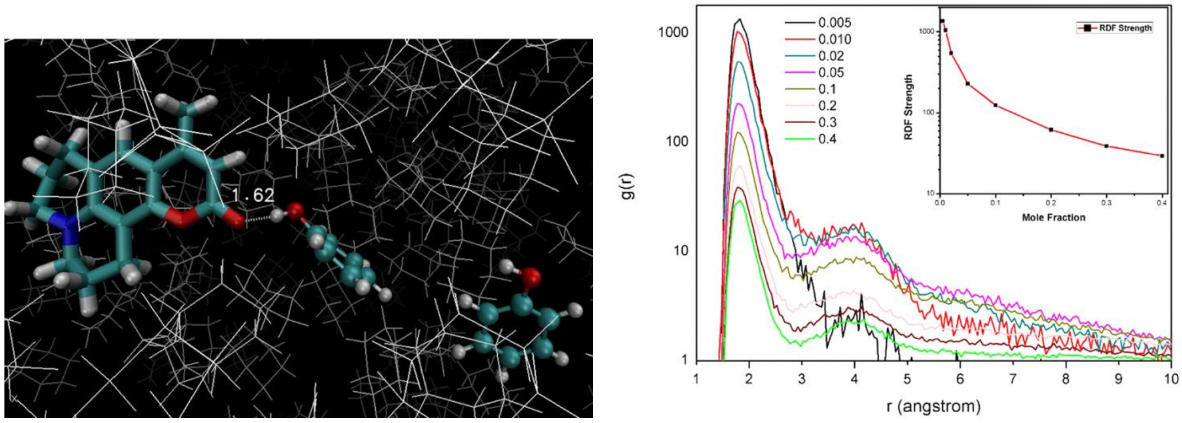
as the number of flexible groups in a simulation increase. During the molecular recognition process the receptors can undergo conformational changes which make the flexibility of the receptor target binding site an essential aspect to be considered during simulations. These flexibility aspects can be adequately taken into account with Molecular Dynamics, which is a stochastic rather than systematic approach.

1.4.2.4.2 Molecular dynamics

Molecular dynamics is a computer simulation technique that specifies the position and speed of each atom in a system by applying Newtonian equations of motion as described in classical mechanics (125,128,136). Newtonian dynamics determines the net force and acceleration experienced by each atom in a molecule. Each atom i at position $r_{i,}$ is treated as a point with a mass m_i and a fixed charge q_i. MD describes the evolution of a system subject to the Newtonian dynamics where forces may be calculated at each step in the simulation using MM or even QM approaches.

1.4.2.4.2.1 Reference molecular dynamics studies from the literature

Hossen *et al.*,utilized molecular dynamics (MD) to elucidate the hydrogen-bonding environment of coumarin 102 in a phenol–cyclohexane mixture (151). In this study MD simulations provided an overview of the H-bonding environment around a particular coumarin containing a fluorophore at various phenol mole fractions in the phenol-cyclohaxane mixture. Although some QM calculations were performed in this study, the molecular dynamics forces were calculated at the molecular mechanics level using the GAFF force field. This provided statistical information on the radial distribution of hydrogen-bond acceptors around the coumarin molecule. Ultimately the study showed that a single coumarin-phenol hydrogen-bond interaction is very important to inducing fluorescence quenching.



b)

Figure 1.12. a) Shows a representative MD simulation snapshot showing coumarin – phenol H-bonding in the phenol – cyclohexane mixture at a phenol mole fraction of 0.005, containing a particular coumarin. The mixture is denoted with LICORICE (coumarin), CPK (phenol), and lines (cyclohexane) respectively. b) Shows the radial distribution function, *g(r)*, of the distance between the carbonyl oxygen of coumarin 102 and the hydroxyl hydrogen of phenol. The inset shows the *g(r)* peak strength against the mole fraction (151). Used with permission.

Another molecular dynamics simulation, now highlighting the use of molecular dynamics with biological systems was in a study of the stereoselectivity of formation of chalcone derivatives (152). In this study the initial coordinates of the enzyme substrate complexes were obtained from the crystal structure of a chalcone isomerase (CHI, PDB ID 1FM7) and this was complexed with 7,4′-dihydroxyflavanone in the active site. The study utilized the RESP protocol as implemented in the Amber9 package. The results showed that CHI exhibits strong strereoselectivity with the binding affinities of CHI(II) with the *s*-trans conformers of chalcone compounds being much stronger than with *s*-cis. Figure 1.13 illustrates the plots of MD-simulated internuclear distance and RMSD versus simulation time for CHI(II) with *s*-cis/*s*-trans chalcone compounds. This study highlights how MD may be used to explore the long-term stability of systems.

There are many examples of MD simulation studies not cited here but used as references at appropriate points in the Discussion. (27,109,153–155).

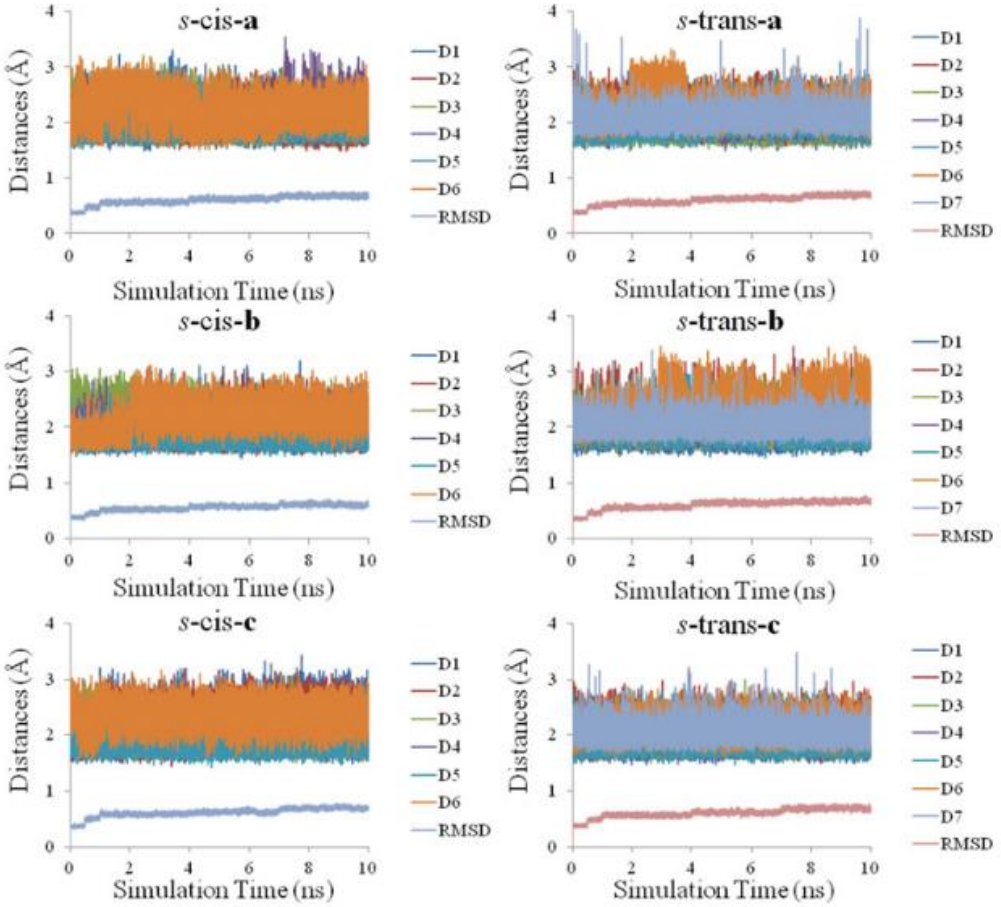


Figure 1.13. Plots of MD-simulated internuclear distance and RMSD versus simulation time for CHI(II) with *s*-cis/*s*-trans chalcone compounds a-c (152). Used with permission.

Molecular dynamics, particularly with systems as large as enzymes, are dependent on Molecular Mechanical methods as the underlying computational approach. However, because molecular mechanics does not account for electrons or for electronic structure, MM is not useful for the study of molecular properties which rely on the electronic structure of the system, or in cases such as the molecular modelling of chemical reactions.

1.4.2.2 Quantum Mechanical (QM) Methods

Quantum Mechanical methods determine the electronic structure of atoms and molecules by solving the Schrödinger equation or Schrödinger wave equation (SWE) (Equation 1.1).

$$H\Psi = E\Psi \tag{1.1}$$

H is the Hamiltonian operator, E is the "eigenvalue" or the energy of the system, and ψ is the wavefunction (118,156,157). The Schrödinger equation cannot be solved exactly for a multi-electron system due to the many-body problem (118,156,157), but it can be solved approximately using *ab-initio* or semi-empirical (essentially *ab initio* with experimental parameters) approaches (117,118,157–161).

The Latin term *ab initio*, meaning from first principles, denotes that equation 1.1 is solved using only the values of the fundamental constants and the atomic numbers of the nuclei utilizing a chosen model (157,161). The model chosen for the wavefunction determines the accuracy of the *ab initio* approach.

1.4.2.2.1 *Ab initio* methods

The first approximation used in *ab-initio* approaches is the Born-Oppenheimer approximation which assumes that:
The electronic wavefunction depends on the position of the nuclei and not on their velocity. i.e., the motion of the nuclei is much smaller than the motion of electrons such that the nuclei can be considered fixed.
The nuclear motion such as rotational and vibrational sees a smeared out potential from the speedy electrons.
Therefore for fixed set of locations *R* of the nuclei the Schrödinger equation takes the form of Equation 1.2 (157)

$$H\Psi(r;R) = E(R)\Psi(r;R) \tag{1.2}$$

The electronic wavefunction is dependent on the electronic coordinates *r* and parametrically on *R*. The Hamiltonian is given by Equation 1.3.

$$H = -\frac{\hbar^2}{2m_e}\sum_i^n \nabla_i^2 - \sum_i^n \sum_I^N \frac{Z_I e^2}{4\pi\varepsilon_0 r_{Ii}} + \frac{1}{2}\sum_{ij}^n \frac{e^2}{4\pi\varepsilon_0 r_{ij}} \tag{1.3}$$

Conventionally, in molecular structure calculations, the nucleus-nucleus repulsion term is added as a classical term at the end of the calculation, and thus not included in *H*.

One of the most direct *ab initio* approaches is the Hartree-Fock method. The Hartree-Fock (HF) model is a self-consistent method, which solves the Schrödinger equation (SWE) using a wavefunction that is a product of one-electron wavefunctions. An antisymmetric form of the wave function ψ is built by introducing the *Slater determinant* (Equation 1.4) (118,156,157,162)

$$\psi(1,\ldots,N) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(1) & \ldots & \phi_1(N) \\ . & . & . \\ . & . & . \\ \phi_N(1) & \ldots & \phi_N(N) \end{vmatrix} \tag{1.4}$$

The exchange of two particles is equivalent to the exchange of two columns which produces a change of sign, which nicely describes the antisymmetric nature of electrons. If two rows are equal, the determinant is zero, thus all ϕ_i 's must be different. This demonstrate Pauli's exclusion principle which states that *two (or more) identical fermions cannot occupy the same state* (118,156,157,162). In modern HF calculations, the one electron wavefunctions are approximated by Linear Combination of Atomic Orbitals (LCAO). It is desirable for these orbitals to be Slater type orbitals. Optimizations with Slater type orbitals, however are computationally expensive and, therefore, in most optimizations the atomic orbitals used are composed of linear combinations of one or more Gaussian-type orbitals (118,156,157,162). Because HF uses one determinant and the smeared electron cloud integration, it approximates treatment of the electron correlation. HF overestimates the electron-electron repulsion and therefore gives "absolute" energies that are too high. The best possible HF energy (the HF limit) is greater than the exact energy from the SWE solution by an amount that is referred to as the electron correlation energy.

Post-Hartree-Fock methods are *ab initio* methods that improve on the Hartree-Fock method by accounting for electron correlation. These include the Configuration Interaction (CI) method and the Møller-Plesset perturbation theory (MP) (118,156,157,162).

Configuration Interaction treats the electron correlation by constructing wavefunctions which are linear combinations of configurational state functions built from spin orbitals. It includes new, product basis functions by promoting one or more electrons from occupied molecular orbitals to unoccupied molecular orbitals; this can result in a matrix of a size that is beyond available computer power (118,156,157,162). The CI expansion is truncated to some manageable length firstly by either deleting some of the some unoccupied MOs preventing their potential occupation, or freezing some core electron MOs to contain two electrons in all configurations for the CI expansion. Another approach is that of truncating the CI expansion, by limiting the excitation, such that the single excited configuration will not mix with the HF reference (Brillouin's theorem). The doubles excited configurations are therefore crucial in the CI expansion (Condon-Slater rules). A practical CI wavefunction will include the reference configuration together with all doubles excited configurations (CID). The most commonly used CI method includes single and double configurations (CISD) (118,156,157,162). CI is good for calculating excited electronic states because it can calculate excited states with the same spin and symmetry as the ground state, it is also good for radicals (118,156,157,162).

Møller-Plesset perturbation theory accounts for the electron correlation by adding the electron correlation effects to the HF method by means of the Rayleigh-Schrödinger perturbation theory (RS-PT) (118,156,157,162). Møller-Plesset perturbation methods are based on the idea that the problem at hand is that the energy obtained from HF calculations is higher than the "exact energy" because the electron-electron repulsion terms are not taken into account when solving the Schrödinger equation.

Mathematically a Hamiltonian consisting of two parts, a reference (H_0) and a perturbation (H') can be defined on the premise that the H' is in some sense smaller than H_0. Assuming that the Schrödinger equation for the reference Hamiltonian operator has been solved then Equations 1.5 hold.

$$H = H_0 + \lambda H' \qquad (1.5)$$

$$H_0\Phi_i = E_i\Phi_i \qquad i= 0, 1, 2,\ldots, \text{infinity}$$

λ is a parameter determining the strength of the perturbation. The perturbed Schrödinger equation is given as Equation 1.6.

$$H\Psi = W\Psi \tag{1.6}$$

If $\lambda = 0$ them $H = H_0$ (118,156,157,162). The Rayleigh-Schrödinger perturbation theory can be either to the second (MP2), third (MP3), or Fourth (MP4) order (118,156,157,162). MP methods are computationally expensive, but they give accurate bond lengths and bond angles even for radicals, quite accurate electron affinities and very good frequencies that do not need to be scaled (118,156,157,162).

1.4.2.2.2 Density Functional Theory (DFT)

Density Functional Theory is a variational method that focuses on the electron density ρ rather than the wavefunction, ψ, to compute the electronic structure of matter (118,156,157,162). The electronic wavefunction is dependent on 3n variables: the x, y, and z coordinates of each electron. The electron density is a function of position dependent on just three variables: x, y, and z positions in space; this makes $\rho(r)$ much simpler than ψ. The energy of the molecule is a function of the electron density $E[\rho]$. The functional part of DFT comes from the fact that the energy of the molecule is a function of electron density and this density $\rho(r)$ is a function of spatial coordinates (118,156,157,162). DFT has its conceptual roots in early but useful approximate models, such as the one developed by E. Fermi and L.H Thomas which emerged in the 1920's – the Thomas-Fermi model – and the Hartree-Fock-Slater model. It was in 1964 that DFT was put on a firm theoretical footing when P. Hohenberg and W. Kohn gave the formal proof that the ground-state energy and all other ground-state electronic properties are uniquely determined by the electron density – the Hohenberg-Kohn theorem does not describe the dependence of energy on the electron density but proves that such a functional exists. W. Kohn and L. J. Sham's proposal that the form of the functional is that in Equation 1.7 allows for the calculation of the energy *via* what is termed the DFT method.

$$E[\rho(r)] = T_{e'}[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}\rho(r)] + E_{xc}[\rho(r)] \tag{1.7}$$

Where $T_{e'}$ is the kinetic energy of the non-interacting electrons whose density is the same as the density of real electrons, V_{ne} is the nuclear – electron attraction term, V_{ee} is the classical

electron-electron repulsion term, E_{xc} is the exchange-correlation functional (156). The electron density is constructed with occupied orbitals, Equation 1.8, and are calculated from Kohn-Sham equations (Equation 1.9), which are similar to the HF equations except for the exchange-correlation potential (V_{XC}).

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{n} |\psi_i(\boldsymbol{r})|^2 \tag{1.8}$$

$$\left\{ -\frac{\hbar^2}{2m_e}\nabla_1^2 - j_0 \sum_{I=1}^{N} \frac{Z_I}{r_{I1}} + j_0 \int \frac{\rho(\boldsymbol{r}_2)}{r_{12}}\, \mathrm{d}\boldsymbol{r}_2 + V_{\mathrm{XC}}(\boldsymbol{r}_1) \right\} \psi_i(\boldsymbol{r}_1) = \varepsilon_i \psi_i(\boldsymbol{r}_1) \tag{1.9}$$

The first term on the left of Equation 1.9 is the usual one-electron kinetic and potential energy contribution and the second term is the potential energy of repulsion between electrons 1 and 2 (161). Kohn-Sham equations are solved iteratively and self-consistently (161). V_{XC} is a functional derivative of the exchange-correlation energy given by Equation 1.10.

$$V_{\mathrm{XC}}[\rho] = \frac{\delta E_{\mathrm{XC}}[\rho]}{\delta \rho} \tag{1.10}$$

The approximate nature of E_{XC} is the main source of error in DFT. The E_{XC} functional is often separated into an exchange functional representing the energy and a correlation functional representing dynamic correlation energy. A variety of exchange-correlation functionals have been developed for use in DFT, such as the BLYP, BP91, and PBE. The most popular is the BLYP functional which is a combination of the gradient-corrected exchange functional developed by A.D. Becke and the gradient-corrected correlation functional developed by C. Lee, W. Yang, and R.G. Parr. The B3LYP functional is one of the 'hybrid' DFT functionals that use HF corrections in conjunction with the density functional correlation and exchange. DFT is computationally affordable and takes into account the correlation effect. In some cases DFT gives results that are more in agreement with experimental values than those obtained from HF procedures, and this is the case particularly with *d*-metal complexes (161).

1.4.2.2.3 Semi-empirical methods

Semi-empirical methods are methods that use a simplified form of the Hamiltonian and adjustable parameters obtained from experimental data (161). Since the cost for performing HF based or DFT calculations increases as the size of the system and the basis function increases, semi-empirical methods are simplified versions of HF using corrections derived from experimental data in order to improve performance (118,156,157,162). The high computational cost of HF based methods is mostly due to the many integrals that need to be calculated, particularly the two electron integrals (118,156,157,162). Semi-empirical methods reduce the computational cost by replacing many of the integrals with approximate functions with empirical parameters which are adjusted to improve the agreement with experiment (118,156,157,162). In semi-empirical methods only the valence electrons are considered explicitly and only a minimal set of valence orbitals are considered on each atom. Core orbitals are not treated by semi-empirical methods since they do not change significantly during chemical reactions. The core electrons are accounted for by reducing the nuclear charge or by introducing functions that model the combined repulsion due to the nuclei and core electrons. The Zero Differential Overlap (ZDO) approximation, which is one of the main features of semi-empirical methods, neglects all products of basis functions that depend on the same electron coordinates when located on different atoms. (156). The most frequently used methods (such as MNDO, AM1, and PM3) are also based on the Neglect of Differential Diatomic Overlap (NDDO) integral approximation.

1.4.2.2.4 The Basis sets

To solve for energy and wavefunction in HF calculations, the one electron wave functions are approximated by Linear Combination of Atomic Orbitals (LCAO); these Atomic Orbitals (AO) must be specified. According to the variational principle, if a set of AOs is infinite then we will obtain the HF limit when using HF theory. An infinite set of AOs is impractical and, therefore, a choice is made on how to truncate the expansion; this choice of AOs defines the basis set (156). In theoretical and computational chemistry, a basis set is a set of mathematical descriptions of orbitals (a set of basis functions) of a system which are used for approximate theoretical calculations or modelling (163,164). There are two types of basis functions: the Slater Type Orbitals (STO) and the Gaussian Type Orbitals (GTO) (118). STOs are similar to

atomic orbitals of the hydrogen atom (118,156,157,162). The general expression for STOs is given by Equation 1.11.

$$\chi_{\zeta,n,l,m}(r, \theta, \varphi) = NY_{l,m}(\theta, \varphi)r^{n-1}e^{-\zeta r} \quad (1.11)$$

where N is the normalization constant and $Y_{l,m}$ are the spherical harmonic functions. STOs are generally used for atomic and diatomic systems which require high accuracy (118). Because STOs can be computationally expensive, Frank Boys suggested the approximation of STOs by linear combination of Gaussian Orbitals (165). GTOs can be written in term of polar coordinates in Equation 1.12.

$$\chi_{\zeta,n,l,m}(r, \theta, \varphi) = NY_{l,m}(\theta, \varphi)r^{2n-2-l}e^{-\zeta r2} \quad (1.12)$$

The r^2 dependence of the exponential makes GTOs inferior to STOs because the GTOs do not represent proper behaviour near the nucleus, and the tail of the wavefunction is represented poorly (118). Most computer calculations use GTOs because they are less computationally expensive and, to compensate for accuracy, more Gaussian equations are combined. Once the choice of the type of basis function has been made then one has to decide on the number of functions to be used.

The smallest number of possible function is called a *minimal basis set* and they are composed of a minimum number of basis functions required to represent all electrons on a neutral atom. Most common minimal basis sets have the form STO-nG, where n is an integer (118,156,157,162). Extended basis sets have a much more detailed description; these include the Double-Zeta, Triple-Zeta, Quadruple-Zeta, Split-Valence, Polarized Sets, and Diffuse sets.

Each function from the minimal basis set is doubled to generate a Double-Zeta basis set; each atomic orbital is expressed as the sum of two functions (118,156,157,162). One set is tighter (closer to the nucleus, has large exponents), and the other set is looser (further from the nucleus, has smaller exponents). The Double-Zeta basis set allows for radial flexibility in describing the electron cloud. The Triple-Zeta and Quadruple-Zeta basis sets work in a similar way to the Double-Zeta basis set except that three and four functions are used instead of two (118,156,157,162).

Split-Valence basis sets result from the fact that the Double-Zeta expansion takes too much effort to calculate, and is therefore simplified by calculating a Double-Zeta basis set only for the valence orbitals, since they are the most vital in the calculations. The inner orbitals are described with a single Slater orbital since they are nearly the same both in atoms and in molecules. The most common Split-Valence basis sets are the 3-21G, 4-31G and the 6-31G basis sets. The first number in the acronyms represents the number of Gaussian functions summed to describe the inner shell orbitals, the second number represents the number of Gaussian functions that comprise the first STO of the Double-Zeta, and the third number represents the number of Gaussian functions that comprise the second STO of the Double-Zeta basis set (118,156,157,162).

Polarization functions in basis sets allow for electron density to be unsymmetric around a particular nucleus. When atoms come closer their charge distribution causes a polarization effect (the positive charge is drawn to one side while the negative charge is drawn to another) which distorts the shapes of atomic orbitals. The 's' orbital tend to have some 'p' character, and the 'p' orbital tends to take on more 'd' character. These polarised functions may also allow for angular flexibility (118,156,157,162). The higher angular momentum 'p' functions are usually added on hydrogen and the 'd' functions are added on heavy atoms. For example, the 6-31G Split-Valence basis set can be written as 6-31G(d) or 6-31G* when 'd' functions are added on heavy atoms, and 6-31G(d,p) or 6-31G** when 'd' functions are added on heavy atoms and 'p' functions are added on hydrogen (118,156,157,162).

Diffuse sets are important when studying anions (with loosely bound electrons), highly electronegative atoms, calculating electron affinities, and gas phase acidities. The '+' sign means that we are accounting for the 'p' orbitals, while the '++' sign means that we are accounting for the 'p' and 's' orbitals (118,156,157,162).

Typical basis sets used in this study are the 6-31G(d,p) and the 6-31+G(d,p) as implemented in Gaussian 09 (166) and Orca (167).

1.4.2.2.5 Reference QM studies from literature

QM calculations may be used to calculate potential energy surfaces related to reaction pathways. For example, in a study which elucidated latent mechanistic complexity in competing acid-catalysed reactions of salicylaldehyde-derived Baylis–Hillman adducts (41). The authors used DFT [B3LYP/6-31G(d)] as implemented in Gaussian 03 to study three independent reaction pathways (Figure 1.14), which were investigated and which permitted rationalization of the experimental data. Insights were provided into the possible mechanisms for the E–Z isomerization and cyclization of the Baylis-Hillman adducts and the subsequent dehydration to afford coumarins.

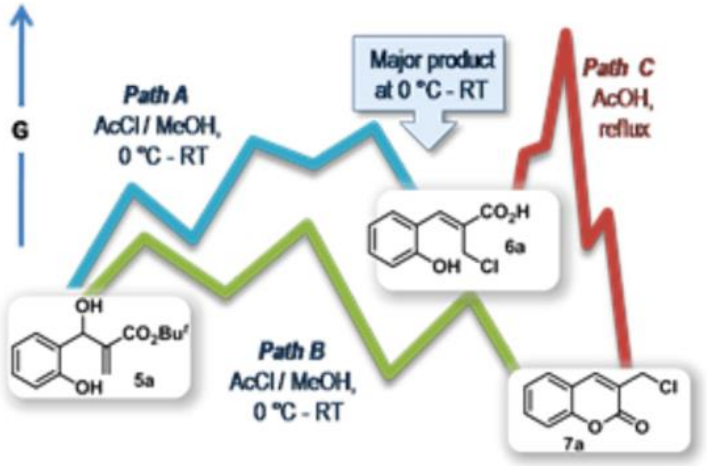


Figure 1.14. Three independent reaction pathways investigated (41). **Used with permission.**

QM studies may also be useful in exploring molecular properties where such properties involve orbital energy levels. A recent study of the dual fluorescence and solvatochromic properties of 3-acyl coumarins (168) showed that the introduction of an acyl group at the C3 position of the coumarin (Figure 1.1) decreases the dipole moment (Figure 1.15). The DFT(B3LYP/6-31G**) study also confirmed the existence of $\pi \rightarrow \pi^{*}$ transitions in spectra of these compounds.

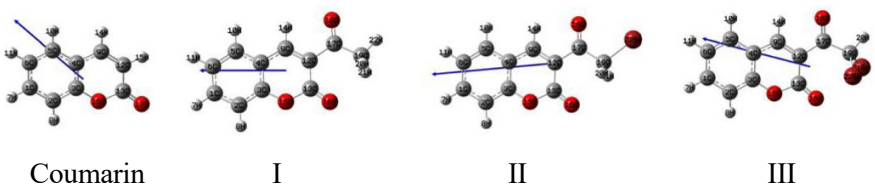
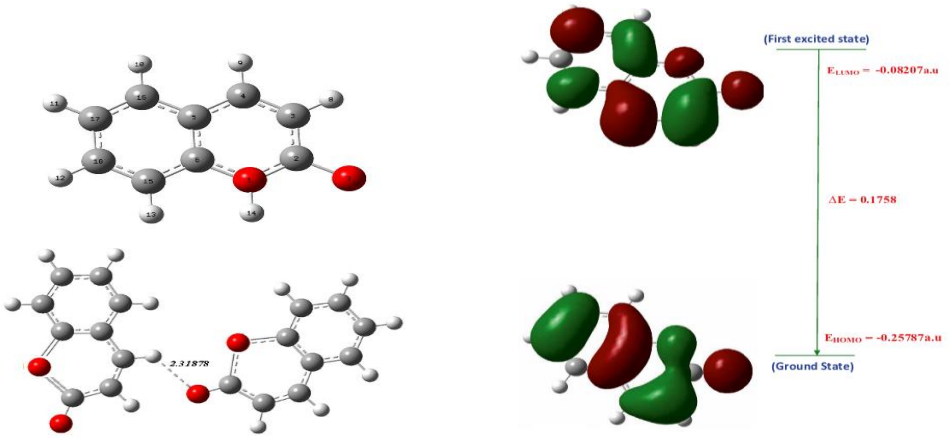


Figure 1.15. Direction of dipole moment obtained from DFT calculations. I (168). used with permission.

The properties amenable to exploration by QM are extensive. In a study of the synthesis, characterization and theoretical properties of coumarin, the non-linear optical (NLO) behaviour of single crystals was investigated using DFT (B3LYP/6-311G++(d,p)) as implemented in Gaussian (169). The computational study provided energy gap values which provided the supporting evidence as to why the coumarin molecule is a better NLO material than urea (Figure 1.16). The computed electronic spectra using the TD-DFT method revealed an intense transition band in the coumarin.



a) b)

Figure 1.16. a) Optimized monomeric and dimeric molecular structures of coumarin. b) HOMO-LUMO plot of coumarin (169). used with permission.

1.4.2.3 QM/MM methods

QM/MM methods are hybrid methods that permit the modelling of bond-making and bond-breaking processes of systems too large to be treated by QM alone or for systems which lack MM parameters (170–174). In QM/MM methods the system is partitioned into a QM region and an MM region. The QM region describes the part of the system that is involved in chemical reactions (the "Reaction center" or active zone) where QM information is required, and the MM region which describes the bulk of the system (the "Spectator region" or environmental zone). Since the active zone (e.g. the active site of an enzyme) is of primary interest in most QM/MM calculations, it is also referred to as the primary sub-system (PS) and the bulk of the system is referred to as the secondary sub-system (SS). The potential energy term $V_{QM/MM}$ has three classes of interactions. i.e., interactions between atoms in the QM region, interactions between atoms in the MM region, and interactions between QM and MM atoms. The potential

energy of the QM interactions is calculated using any of the *ab initio* or semi-empirical methods. The potential energy of the MM interactions is calculated using any of the MM force field calculations. The potential energy of the QM/MM interactions is calculated through subtractive or additive schemes. In the subtractive QM/MM coupling scheme, the potential energy term for the interactions between the QM and the MM regions is computed at the MM level. The MM calculation will compute the potential energy of both the active zone and the environmental zone, and then subtract the potential energy of the active zone calculated at MM, and then add the potential energy of the QM region calculated at the QM level (Equations 1.13).

$$V_{QM/MM} = V_{MM}(MM+QM) + V_{QM}(QM) - V_{MM}(QM) \quad (1.13)$$

In additive QM/MM coupling, the potential energy of the active zone calculated at the QM level is added to the potential energy of the environmental zone calculated at the MM and to the potential energy term for the interaction between the active and the environmental zone.

$$V_{QM/MM} = V_{QM}(QM) + V_{MM}(MM) + V_{QM\text{-}MM}(QM+MM) \quad (1.14)$$

The challenge in additive QM/MM coupling is in the calculation of the $V_{QM\text{-}MM}(QM+MM)$ term. In conventional QM/MM there are several schemes for describing the interactions between the PS and the SS, *viz*., Mechanical embedding, electrostatic embedding, polarizable embedding, and flexible embedding.

The most basic embedding scheme is the mechanical embedding in which all interactions are handled at the force field level (175,176). Chemical bonds between the PS and the SS are modelled by harmonic potentials (V^{bond}), angles defined by one QM atom and two MM atoms are described by the harmonic potentials as (V^{angles}), torsion angles involving at most two QM atoms are usually modelled by a periodic potential function ($V^{torsion}$), non-bonded interactions, such as the Van der Waals interactions, are modelled by a Lennard-Jones potential (V^{LJ}) and the electrostatic interactions by the Coulomb potential (V^{Coul}) (175,176). Mechanical embedding requires an accurate set of MM parameters such as atom-centered point charges for both the PS and the SS. The charge distribution in the PS can change as the reaction progresses, and this could result in errors if only a single set of parameters is used (175,176). Another shortcoming of mechanical embedding is that the electronic wavefunction is evaluated for an

isolated QM system, and therefore the SS cannot induce polarization of the electron density in the PS.

The electrostatic embedding mechanism improves on the mechanical embedding by including the polarization effects (176). Electrostatic embedding does not require MM electrostatic parameters for the PS because the electrostatic interactions between the PS and the SS are treated by including one-electron terms in the QM Hamiltonian (Equation 1.15).

$$h_i^{\mathrm{QM-MM}} = h_i^{\mathrm{QM}} - \sum_{J}^{M} \frac{e^2 Q_J}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_J|} \tag{1.15}$$

In this equation, r_i and R_J identify the coordinates of the interacting electron *i* together with the MM atom *J*, h_i^{QM} remains as the one-electron operator providing the energy that describes both kinetic and nuclear attraction of this electron and the summation over *M* relates to all MM atoms which have associated partial charge Q_J (175). There is a problem with the use of the very strong MM charges on atoms in the descriptions of the MM region in that there is a very real risk of over-polarization at the QM/MM boundary resulting in electron density exceeding the bounds of the QM region. The electrostatic interactions are not the only interactions between the PS and the SS; in reality, the interactions include polarization, dispersion and charge transfer, and QM effects such as Pauli repulsion (175).

Polarizable embedding improves further the embedding mechanisms by allowing the MM atoms to some extent to be polarizable, such that the PS and the SS can mutually polarize each other. There are several approaches for modelling the MM atoms' polarization, including the induced dipole and fluctuating charge models, but also including the charge-on-a-spring model (175,176). Total QM/MM energy calculations polarizable embedding paradigm involves the computation of the MM polarizations in conjunction with the QM self-consistent field (SCF) iterations (175,176).

Flexible embedding is the preferred of all reported embedding approaches because it goes a step further by allowing the summation of charges of both the PS and the SS regions to change during the course of modelling, thus allowing for accurate simulation of charge transfer that extends across both regions (175,176).

The more sophisticated the embedding approach is, the more costly it becomes computationally. Therefore the QM/MM approaches in this study utilized electronic embedding, as implemented in the ONIOM method in GAUSSIAN 09 (166).

1.4.2.3.1 Reference QM/MM studies from the literature

An interesting study from the literature is the comparative computational analysis of different active-site conformations and substrates in a chalcone isomerase catalysed reaction (177). The study focused on a rate-limiting, intramolecular Michael addition step of a 2′-oxyanion to the chalcone α,β-double bond. Two different active site conformations of CHI (initial coordinates from PDB code: 1EYQ) observed in the crystallographic structures (named A and B in the reference study) were used to calibrate the influence of the active site conformation on the free energy barrier. QM/MM analysis showed that the reaction proceeds with a charge flow from the O2′-oxyanion to the C^{β} atom. As such the charge–charge interaction between the substrate and Lys97 becomes stronger in the transition state (TS) than that in the Michaelis Complex (MC), especially in the B active site.

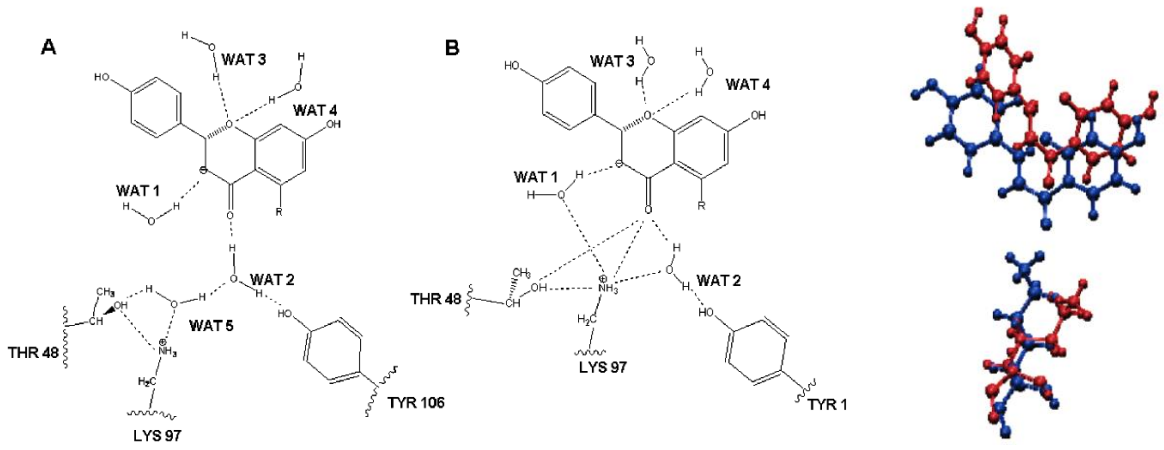
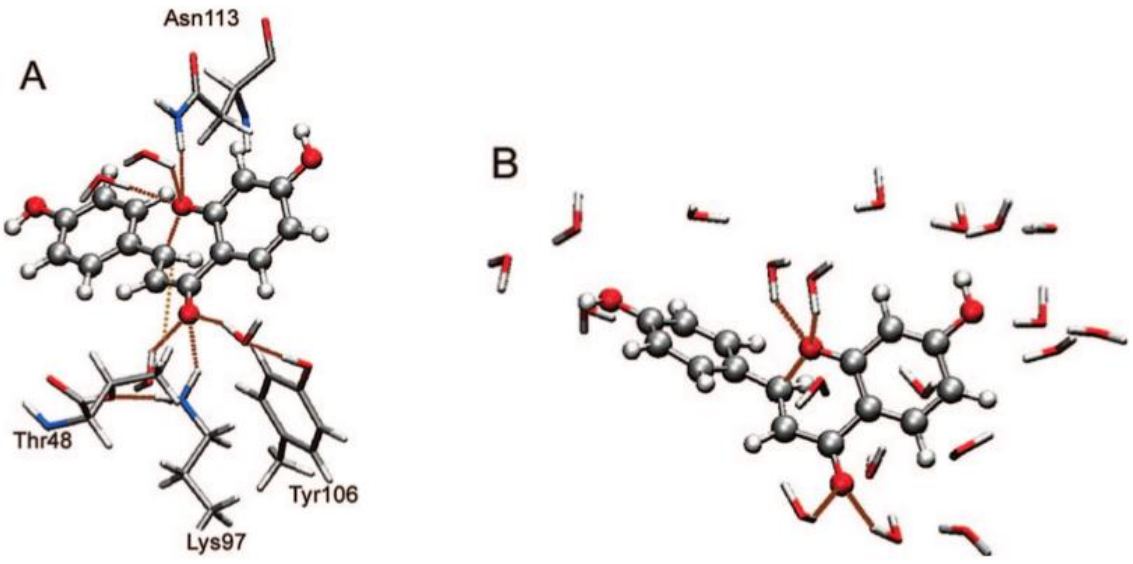


Figure 1.17. a) and b) Schematic views of the two transition state conformations (A and B) for the CHI active site. c) Superposition of the TS structures of 6′-deoxychalcone in A (red) and B (blue) conformations of the active site showing only the substrate and Lys97 (177). Used with permission.

Another study, also on chalcone isomerase enzymes that employed the QM/MM approach is the study of dynamic effects on the reaction rate in a Michael addition catalysed by these enzymes (178). In this study a comparison was made between the reaction dynamics for the uncatalysed reaction in aqueous solution and for the reaction catalysed by CHI. The reactant constituted the QM subsystem and the MM subsystem was the enzyme (PDB code: 1EYQ); the crystallization water and solvating water molecules were described using OPLS-AA and TIP3P potentials. The study showed that the Michael addition involves a large charge transfer to the C^{α} and, to a lesser degree, charge transfer to the carbonyl oxygen.



a) b)

Figure 1.18. a) Transition structures for the transformations: a) in the CHI; and b) in solution with only the atoms included in the control space defined in the Hessian matrix shown in the picture (178). Used with permission.

In the context of the current study, the E→Z isomerization process is a chemical (electronic) process that involves the protonation of the acyl oxygen and deprotonation of the hydroxyl OH of coumaric acid, mediated by enzymes. A QM approach is necessary to understand the transformation, but the system including the enzyme is too large for QM. Therefore this study uses hybrid QM/MM methods where the system is partitioned into a QM region and the MM region.

1.4.2.4 Adaptive QM/MM methods

1.4.2.4.1 Discrete and Continuum solvation methods

The QM/MM calculations considered thus far are mostly *in vacuo*. In nature, reaction mechanisms occur in a solvated medium, and the most abundant solvent in nature is water. Adaptive partitioning is focused on QM/MM hybrid simulations of reactive species in a condensed phase in which the role of the solvent molecules are critical. In this discussion on adaptive partitioning, water is generally the reference solvent. There are two ways of simulating a solvated system; solvent molecules can be simulated implicitly using continuum models or explicitly using solvent particles.

A continuum medium is used as the basis for implicit solvent models. The idea behind these implicit solvent models is that if we had a solute-solvent system the effects of the bulk solvent on the solute tend towards an average effect; as such is it possible to effectively integrate over all individual solvent molecule coordinates. Thus the solvent effect is condensed into a single dielectric constant (ε), and this allows for determination of the solute-solvent interaction based on the presence of solute in a solvent field (179–185). This simplifies the computation of the solute-solvent system's energy. In explicit solvation models the solute is treated by the creation of a cavity within a medium that is polarizable in order to hold the solute molecule. The destabilization of energy resulting from the removal of solvent-solvent interactions is ΔG_{cav}. The solute is then modelled in the solvent cavity, which results in both electrostatic and non-electrostatic contact between the solute and the continuum medium. The electric charge distribution of the solute will polarize the dielectric medium, inducing charge moments which will, in turn, act back on the molecule, thereby producing an electrostatic stabilization. There is induced polarization from the solute on the continuum medium and the medium induces in kind polarization on the solute. The attractive terms and the energetic destabilization as a result of polarizing the solute and solvent is taken into account by the electrostatic term ΔG_{elec}. Dispersion is the main non-electrostatic contribution to the solvation energy ($\Delta G_{nonelec}$) although there are other contributing factors, including the non-electrostatic component of hydrogen bonding (179–185). The total solvation energy ($\Delta G_{solvation}$) is given as a sum of these energy contributions (Equation 1.16).

$$\Delta G_{solvation} = \Delta G_{elec} + \Delta G_{cav} + \Delta G_{nonelec} \quad (1.16)$$

The dielectric constant is the only parameter characterizing the solvent, therefore solvents having the same dielectric constant are treated equally. The simplest shape of the hole in the continuum is a sphere (or an ellipsoid) which is easier to handle because the interaction between the solute and the continuous dielectric medium can be calculated analytically. For more realistic models the cavity is molecular-shaped; generally these are generated by interlocking spheres located at each nucleus of the solute. The Polarized Continuum Method (PCM) developed by Tomasi and his group is one of the most commonly utilized continuum models (179–185). Since its development there have been several improvements on the PCM model such as the IEF-PCM, IPCM and SCIPCM models. In these methods the potential energy is solved iteratively, and therefore giving the self-consistent reaction field (SCRF) term.

Implicit solvent models are quite useful in the simulation of biomolecules, but they have problems with the parameterization and treatment of ionization effects. These continuum models are more computationally affordable, but because the medium is considered as a continuum the coordinates of each solvent molecule are not determined and, therefore, the intramolecular solute – solvent interactions are not properly defined (179–185).

Explicit solvent models depends on utilizing a significantly large number (hundreds or thousands) of discrete solvent molecules. This is a more realistic way of simulating a solvated system in which there are direct and specific solvent interactions with the solute. Including explicit solvent molecules in a QM simulations means increasing the number of electrons in the system, and the Schrödinger equation becomes more and more difficult to solve. Explicit solvation with QM is computationally expensive, therefore most explicit solvation models are used in MM approaches. Commonly used solvents such as water often have idealized models generated to reduce the number of coordinates requiring evaluation in the energy term while maintaining the accuracy of the model. The most common of these models include TIPXP (such as TIP3P where X is an integer relating to the number of points used within the calculation of the energy) (179–186), together with the simple charge model (SPC) (179–186). Simulating an explicitly solvated system with a hybrid QM/MM approach means also partitioning the solvent molecules such that some of the solvent molecules will be part of the PS and the bulk water solvents will be part of the SS. The interaction between the PS (the solute active zone and the solvent molecules that have penetrated the QM region) and the SS (the solute environmental zone and the bulk solvent molecules) are taken into account with the embedding methods.

A QM/MM approach for a solvated medium should also consider the fact that solvent can move between the QM and the MM region during simulation. Several QM/MM techniques have been altered and improved for condensed phases such as Constrained QM/MM, and Adaptive Partitioning (Discontinuous Adaptive QM/MM and Continuous Adaptive QM/MM) (171,175,187–190).

1.4.2.4.1 Constrained QM/MM

One of the most common constrained QM/MM techniques is the Flexible Inner Region Ensemble Separator (FIRES) which constrains solvent from moving between regions, but has no prerequisite for a defined QM volume (171,175,187–190). In FIRES QM-defined solvent (in particular this solvent furthest from the QM center) in turn determines the radius of the sphere around the active zone, and the constraining potential is located on the sphere. This allows flexibility in that the QM zone self-corrects and adjusts to the overall free energy surface in terms of molecular density since the QM radius is not constant throughout the simulation. The introduction of constraints (harmonic) in the FIRES method alters the partition function; however the effects of this are structural distortions at the QM/MM boundary (171,175,187–190). The Boundary based on Exchange Symmetry Theory (BEST) method improves on the FIRES method by including all possible scenarios of exchange of solvent with respect to both QM and the MM regions (171,175,187–190).

1.4.2.4.2 Discontinuous Adaptive QM/MM

The main idea is to find an approach that will dynamically allow for diffusion of solvent molecules from the QM to the MM region, and *vice versa* during simulation. The most common discontinuous adaptive QM/MM approaches include the Abrupt, ONIOM-XS, Hot Spot adaptive partitioning, and Buffered Force.

The Abrupt method as suggested by its name involves the abrupt swapping between QM and MM descriptions as solvent molecules move spatially through the QM/MM boundary. As a result of this movement, the molecule is simply either included in the QM or in the MM set for calculation. Abrupt changes in the exact QM/MM partitioning in results in an abrupt changes in the calculated potential energy because the QM and the MM energies are different and, therefore, there is no conservation of the combined total energy of the simulation (171,175,187–190).

To reduce the problems of conservation of energy resulting from abrupt changes within the QM/MM partition, ONIOM-XS method introduces a transitional region between the PS and SS. i.e., the buffer region. The solvent molecules in the buffer region can be characterised as either QM or MM semi-continuously and, consequently, they have partially QM and partially MM character. ONIOM-XS introduces to some level continuity into the transition between the QM and the MM partition, but some degree of discontinuity remains; this is why it is till classified as discontinuous. Because of its discontinuity ONIOM-XS does not conserve energy, but the total momentum is conserved (171,175,187–190).

Like the ONIOM-XS method, the Hot Spot method introduces a buffer region between the active zone (the primary sub-system) and the environmental zone (the secondary sub-system). But, unlike the ONIOM-XS method, the Hot Spot method has a combined active and transition region. The molecules comprising the active zone and the buffer zone are QM and the molecules in the environmental zone are MM. Like the ONIOM-XS method, the Hot Spot method is discontinuous and therefore suffers the same consequence of non-conservation of energy (171,175,187–190).

The Buffered Force method, like its predecessors (the ONIOM-XS and the Hot Spot methods), does not conserve energy during simulation.

1.4.2.4.3 Continuous Adaptive QM/MM

Continuous Adaptive QM/MM methods provide the best approach to a QM/MM study of a solvated system because they ensure a complete continuous transition of molecules between the active zone (PS) and the environmental zone (SS). These methods include Permuted Adaptive Partitioning (PAP), Sorted Adaptive Partitioning (SAP), Difference-based Adaptive solvation (DAS), and Size-Consistent Multi-Partitioning (SCMP). Like most of the discontinuous methods these methods also rely on defining a transitional region linking the PS and the SS where the molecules have fractional QM and MM character based on their positions. Continuous methods entails the computation of an exhaustive array of partitions of the solvent molecules. All molecules in the buffer region potentially have QM character and, therefore, contributions of many QM/MM partitions are included in an adaptive potential energy

expression that allocates the desired partial QM character (171,175,187–190). Continuous methods define this adaptive potential energy $V^{ad}(r)$ as

$$V^{ad}(r) = \sum_n \sigma^{(n)}(r) V^{(n)}(r) \qquad (1.17)$$

Where $\sigma^{(n)}(r)$ are the weight functions; when molecules of solvent (m_i) increase in distance from the QM center then the weight function that describes m_i QM tends towards zero as the m_i transits through the buffer zone to the SS. The difference in these continuous methods is in the nature of $\sigma^{(n)}(r)$, even though the way the potential energy is defined is similar (171,175,187–190). Most continuous adaptive methods conserve both energy and momentum.

Permuted Adaptive Partitioning (PAP) is the most commonly used continuous method, because it assigns the weight function $\sigma^{(n)}(r)$ in the most intuitive manner. In PAP exhaustive generation of partitions of solvent in the buffer region is required to include all relevant molecules in two sets (QM and MM) for their contribution to adaptive potential energy in Eq 1.17. If the number of these molecules in the buffer zone equals the number M at a particular point in the simulation, it follows that there will be up to 2^M possible PAP partitions with significant contribution. The switching function λ is necessary at each point in the simulation for all M buffer zone molecules to obtain the weight function of these partitions. Although this method is considered more intuitive, it requires calculation of a significantly large number of computationally exhaustive QM/MM partitions for each conformation (as mentioned before this is of the order of 2^M possible partitions) (171,175,187–190).

In Sorted Adaptive Partitioning (SAP) the number of relevant partitions in the buffer region is reduced with the introduction of a weight function $\sigma^{(n)}(r)$ that biases such that fewer partitions have contribution – the design is that only $M + 1$ of the 2^M now contribute. In these 'ordered' partitions, QM-defined buffer solvent are spatially closer to the QM core, while the MM-defined atoms are further from this core, therefore as such the number of QM solvent molecules in each partition differs (171,175,187–190).

Difference-based Adaptive solvation (DAS) has similarities with the SAP in that only these 'ordered' partitions contribute. DAS is mostly applied combined with the non-energy conserving approach, where the expression for the forces does not contain the gradient for the

weight function $\sigma^{(n)}(r)$ and therefore there would be no problems with the discontinuities. DAS simulations conserve the total energy better than PAP and SAS when large time-steps are used (171,175,187–190).

In the Size-Consistent Multi-Partitioning (SCMP) method the user predefines the number of contributing partitions *N*. The switching functions λ are introduced at each time-step for all molecules in order obtain the weight functions. Because each partition takes relatively the same time to calculate, the SCMP method can run efficiently in parallel. Therefore SCMP allows consideration of a fairly large number of partitions, which would lead to a smooth change in QM character of molecules inside the buffer zone.

This study considered the Difference-based Adaptive solvation (DAS) method as implemented in the CP2K package (191).

1.4.2.4.4 Boundary position in adaptive QM/MM

Defining the size and shape of the QM region in solvated systems is a challenge with QM/MM methods in condensed phases. In the methods discussed so far, the QM region has been considered as a sphere with a 'centre' and a radius. The 'centre' is usually located relative to the centre of mass of relevant atoms in the simulation. When simulating biomolecules, a fixed point in space can be used as a centre for the QM region. In order to conserve momentum, large mass virtual atom is assigned to be the QM centre (171,175,187–190). The size of the QM can be determined in three ways:

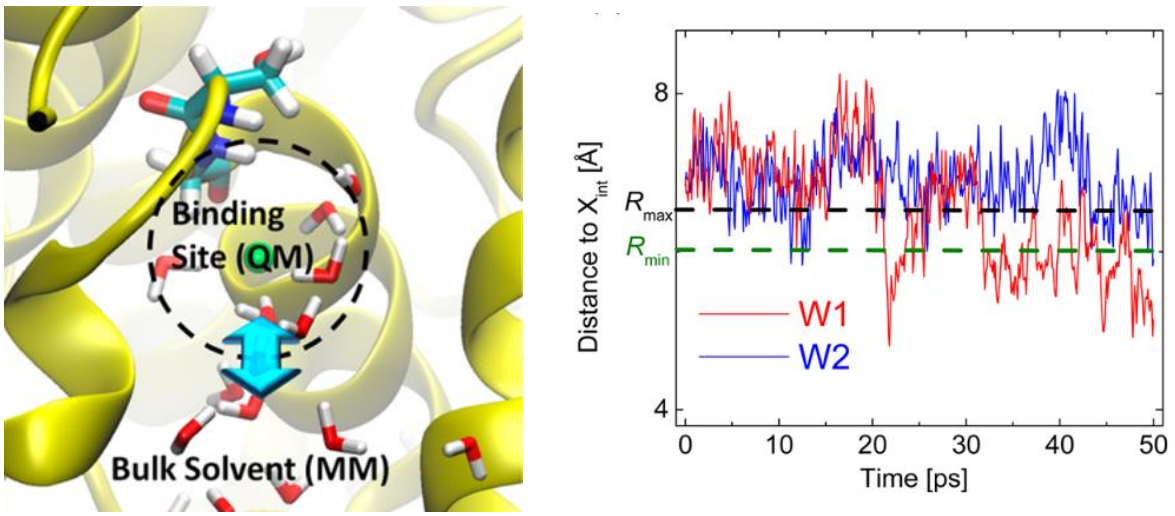
i) Distance-based (Distance-Adaptive Multi-scale: DAMS)
ii) Number-based (Number-Adaptive Multi-Scale: NAMS)
iii) Density-based (Density-based Adaptive QM/MM: DBA)

The original and most frequently used approach is DAMS, in which the size of the sphere is constant during the course of dynamics. In NAMS, however, the solvent population of the PS is kept constant. In DBA the QM character of the molecule also depends on orientation, therefore the number of QM solvent molecules may be reduced relative to those of DAMS and

NAMS making DBA more computationally affordable. In theory, these selections of size are independent of the choice of adaptive QM/MM method (171,175,187–190).
In this research project, DAMS and NAMS were used as implemented in the CP2K package (191).

1.4.2.4.5 Reference adaptive partition QM/MM studies from the literature

Adaptive partitioning QM/MM has been used recently in a simulation of water molecules as they diffuse in and out of the active site of a protein (192). The MD simulations were done on two binding sites of *Escherichia coli* CLC Cl^- ion transport protein. In this study, the authors used CHARMM22 force fields for biomacromolecule atoms while the water was defined using the TIP3P model. A modified PAP scheme was performed with the QM/MM programme which is called Tinker for the MD and NIDO for the QM. Pseudoatoms X_{int} were introduced to mark the active site of the intracellular binding site S_{int}. The position of X_{int} were independent of the motion of ligand or of residues within the active site. Figure 1.19a shows the active zone considered, and Figure 1.19b shows the time-steps of two water molecules relative to X_{int}. The study highlighted exchange processes with solvent within the binding site of the protein.



a) b)

Figure 1.19. a) *Escherichia coli* CLC binding site showing the QM and part of the MM region. (192). b) Distance from X_{int} to W1 and from X_{int} to W2 monitored during the course of simulation. Dashed lines indicate the boundary region (192). used with permission.

Adaptive partitioning QM/MM was also used recently in a simulation of proton hopping in bulk water (188). In this study a proton indicator was proposed to serve as the reference point. Using a modified PAP, the movement of a single proton within bulk solvent water was simulated in which a QM sub-system was defined spherically with a 4.0 Å radius centered on the proton. This procedure together with a modified PAP scheme, yielded acceptable results for this proton solvation. The results obtained also concurred with those obtained from a previous study which utilized a larger QM radius of 12 Å.

1.5 Aims and objectives of the present study

The project focuses initially on the application of results from recently published work regarding the transformation of cinnamic acids into coumarins (43), in order to explore the author's suggested enzyme-catalysed transformation of cinnamic acids to coumarins. Particular attention is given to the bioorganic transformation of *o*-coumaric acid with chalcone isomerase enzymes. There are three critical phases involved in the transformation of cinnamic acid derivatives to coumarins.

i) E → Z isomersation of the *o*-hydroxycinnamic acid precursor;
ii) Cyclisation (lactonisation) to the hemi-acetal intermediate; and
iii) Dehydration to afford the coumarin derivative.

The research involves the application of combined, low-level Molecular Mechanics and high-level Quantum Mechanical/DFT calculations to access complementary models of the overall reaction.

The detailed aims of the project are as follows:

Studying the several protonation states of different derivatives of *o*-hydroxycinnamic acid to explore and determine which mechanism (protonation/deprotonation) is most theoretically favourable to achieve the partial C(7)=C(8) double bond.
Docking the *o*-hydroxycinnamic acid derivatives into the chalcone isomerase enzymes to determine if the *o*-hydroxycinnamic acid molecules are compatible ligands for the chalcone isomerase enzyme receptors, and also to explore the types of receptor-ligand interactions in the complex.
High throughput virtual screening of the binding of thousands of chalcone derivatives to the chalcone isomerase enzymes to determine the chemical properties of the chalcone molecules that influence their binding affinities towards chalcone isomerase enzymes.
Exploring the dynamic evolution of *o*-coumaric acid in the active site of chalcone isomerase enzymes, to determine if this ligand remains in the CHI active site over time, and the types of receptor-ligand interaction that are persistent over a particular time frame.
Using the ONIOM model to study the protonation and deprotonation of the *o*-coumaric acid with protein residues in order to achieve a partial *o*-coumaric acid C(7)=C(8) double bond, and the subsequent *E*→*Z* isomerization of the protonated/deprotonated system.
Studying the evolution of the *o*-coumaric acid – chalcone isomerase enzyme complex with adaptive partitioning QM/MM in a condensed phase, to see which types if receptor ligand interaction are more favourable when the active site is computed at QM level, and to establish whether the system considered would spontaneously favour the protonation/deprotonation, *E*→*Z* isomerization, and lactonization steps required for the overall transformation from *o*-coumaric acid to coumarin.

This study also makes reference to previous computational studies of molecules at Rhodes University (Department of Chemistry) in the Computational Mechanistic Chemistry and Drug Discovery (CMCDD) research group (34,193–195).

CHAPTER 2
COMPUTATIONAL DETAILS

2.1 Introduction to the chapter

This chapter provides the computational approaches and procedures used in detail in this study. The procedures discussed are detailed in this section but relate to later chapters in which the emphasis is on the results of the study.

2.2 Computational approaches/practical procedures

The following sections relate to specific chapters and, as such, the title for the section matches the title for each of the subsequent chapters in the thesis.

2.2.1 Population analysis of cinnamic acid derivatives (Chapter 3)

Population analysis of different protonation states of cinnamic acid derivatives employed a QM approach. These electronic structure methods were appropriate for calculation of the properties of the cinnamic acid derivatives, such as the total energy, vibrational frequencies and dipole moments (196–198). Although these methods are computationally expensive, the derivatives are between 20 and 30 atoms in size and, therefore, they lend themselves to this type of study.

The study was carried out both *in vacuo* and with solvent correction using Gaussian 09 D01 (166). Solvent-corrected calculations were performed using the polarizable continuum model (PCM) (180) for water-solvated and chloroform-solvated media. The Z matrix (in some cases the Cartesian coordinates) giving the internal coordinates of the system was prepared by building the input geometry of the structure of the molecule in GaussView. Density Functional Theory (DFT) with the 6-31+G(d,p) basis set was utilised in all geometry optimizations. All

optimized structures were visualized using GaussView (199). The outputs obtained from the optimizations were analysed, in each case, for the relative energy and the energies of the Highest Occupied Molecular Orbital (HOMO) and of the Lowest Unoccupied Molecular Orbital (LUMO). The energies obtained were in Hartrees and were converted to kcal/mol by multiplying the value obtained using the conversion factor 627.509 kcal/mol/Ha. The HOMO-LUMO energy gap was obtained by tabulating orbital energy levels prior to a simple difference calculation between HOMO and LUMO energy values within Excel. The optimised geometries and the molecular orbital diagrams represented in the text were obtained from GaussView (199). The figures showing the energy trends were made utilizing Excel. Qualitative and quantitative analysis of the results are provided in Chapter 3.

2.2.2 *In silico* docking (Chapter 4)

Molecular docking is a computational method which predicts the binding orientation of one molecule to the next, the binding affinities, binding interactions and other receptor-ligand complex properties. The aims in this section are to identify a putative receptor and analyse the binding pose and the interactions of the ligand with this receptor. *In silico* docking was done on the 9 wild and 2 mutant chalcone isomerase enzymes discussed in Chapter 1 Section 1.3.1.1 *i.e*., CHI PDB structures: 1FM7, 1FM8, 1JEP, 1EYQ, 1EYP, 4DOI, 4DOK, 4DOL, 4DOO, 1JXO (Y106F mutant), and 1JX1 (T48A mutant).

The three-dimensional crystal structures were downloaded from the RCSB protein Data Bank (200). The CHIs are dimeric structures and, therefore, Chain A was selected for the computational calculations. Discovery Studio Visualizer (201) was utilized to determine the centre of mass of the crystal structure with the ligand bound to the A Chain. These coordinates were used to identify the active site in the docking procedure. Other crystal structure components such as Chain B of the dimer, crystal structure water, and the crystal structure ligand were also removed with Discovery Studio Visualizer (201). The ligand structures utilized for docking were the B3LYP/6-31+G(d,p) optimized geometries from the population analysis study, which were saved as PDB structures for compatibility with the docking programme.

Docking was done using Autodock Vina (202) employing a simple scoring function and a rapid gradient-optimization conformational search. Docking was controlled with the use of scripts, in which the exhaustiveness for the docking was set to 64.

The results were visualized using Discovery Studio Visualizer (201) in order to examine and analyse binding affinities, docking poses, and receptor ligand interactions.

2.2.3 High Throughput Virtual screening (HTVS) (Chapter 5)

High throughput virtual screening (HTVS) done in this section was an extension of the *in silico* docking followed in Chapter 4. HTVS was done on 8 064 chalcone and chalcone derivatives docked into the 11 CHI crystal structures which were utilised for docking *o*-coumaric acid [1FM7, 1FM8, 1JEP, 1EYQ, 1EYP, 4DOI, 4DOK, 4DOL, 4DOO, 1JXO (Y106F mutant), and 1JX1 (T48A mutant)]. Because the same CHI crystallographic structures were utilised, the preparation of this receptors has already been discussed in the preceding section, and the active-site coordinates used for the HTVS are the same as those used for docking.

All chalcones and chalcone derivatives utilised in the HTVS study were obtained from a virtual library created by Arthur Sarron (a PhD student within this research group). In the creation of the chalcone virtual library, he selected acetophenone, benzaldehyde and alkyl halides from the Acros website (203), based on criteria of purchasability and length of the alkyl chains to avoid steric overload during docking. All structures in this library had been optimized at the AM1 level using Gaussian prior to release.

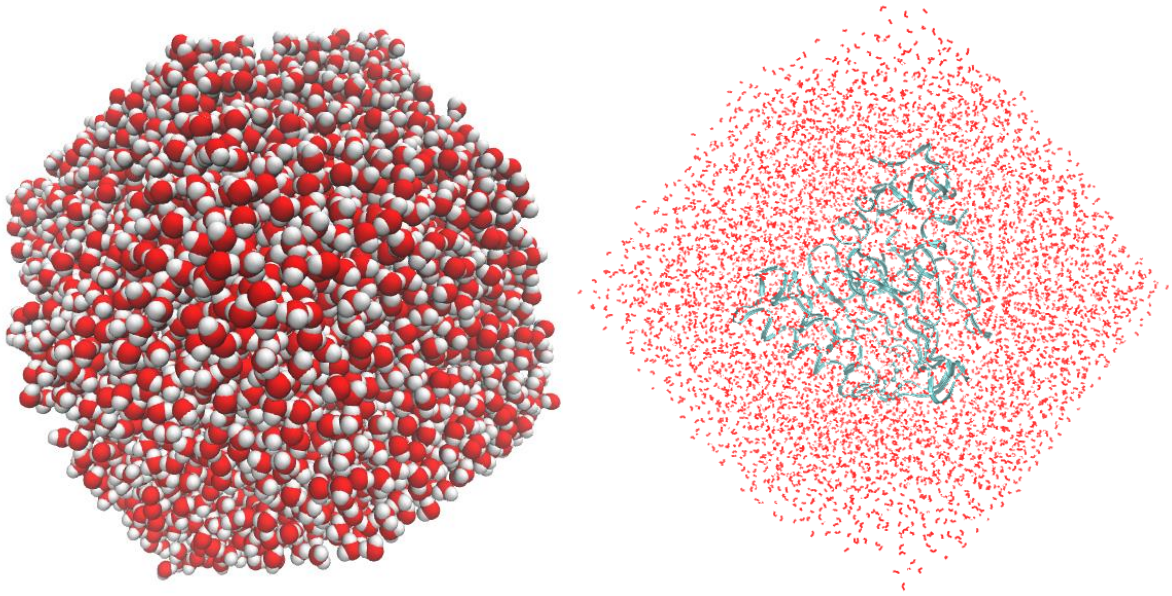
HTVS also used Autodock Vina (202) for docking, with an exhaustiveness of 64, but the management of the docking was done using scripts to create input and job-files for submission to the batch queuing system at the Center for High Performance Computing. The results were visualized using Discovery Studio Visualizer (201) and analysed (also using scripts) for the same properties as obtained for the docking results, namely, best binding affinities, docking poses, and receptor ligand interactions.

2.2.4 Molecular dynamics (Chapter 6)

The chemical structures obtained from docking *o*-coumaric acid with CHI receptors were further studied using molecular dynamics. The receptor-ligand complexes were solvated with water molecules under periodic boundary conditions, neutralized and subjected to dynamics calculations using CHARMM (138,204) and GROMACS (205,206). Initial calculations were done using CHARMM for short simulations (the available programme did not scale well beyond 16 cpu cores); final calculations used the GROMACS suite for simulations up to 100 ns.

2.2.4.1 Models for the CHARMM simulation and computational approach

Eleven CHARMM simulations were performed using the *E*-*o*-coumaric acid with 20 atoms docked to the 11 chalcone isomerase X-ray crystallographic structures. The ligand structure used was the *o*-coumaric acid with the best binding affinity for each CHI from the docking study. The ligand topology was obtained by processing the ligand through the CHARMM General Force Field (CGENFF) program (207) which performs atom typing and assignment of parameters and charges by analogy in a fully automated fashion. The protonation of the protein was done by uploading and processing the PDB topologies of the 11 chalcones to the H++ website (208) of Virginia Tech, the protonation states of the histidine (HIS) were checked manually and all HIS were found to be HSD (histidine with hydrogen on the delta nitrogen), and HIS was changed to HSD on the processed PDB files. All receptor-ligand complexes were solvated in a rhombic dodecahedron (RHDO) with the crystal dimension of 69.0475 69.0475 69.0475; water molecule coordinates were extracted from a cubic box containing 46656 equilibrated TIPS and water molecules were removed where atom clashes occurred. The system was neutralized with 18 Cl^- and 24 K^+ ions, minimized by steepest decent, and then heated in stages from 110 K to 310 K. After heating, the system was subjected to an equilibration step at 310K, and finally MD production dynamics was performed for 30 ns. The results were visualized with VMD (209). Figure 2.1 shows a model of an *E*-*o*-coumaric acid-CHI complex solvated in an RHDO water environment.



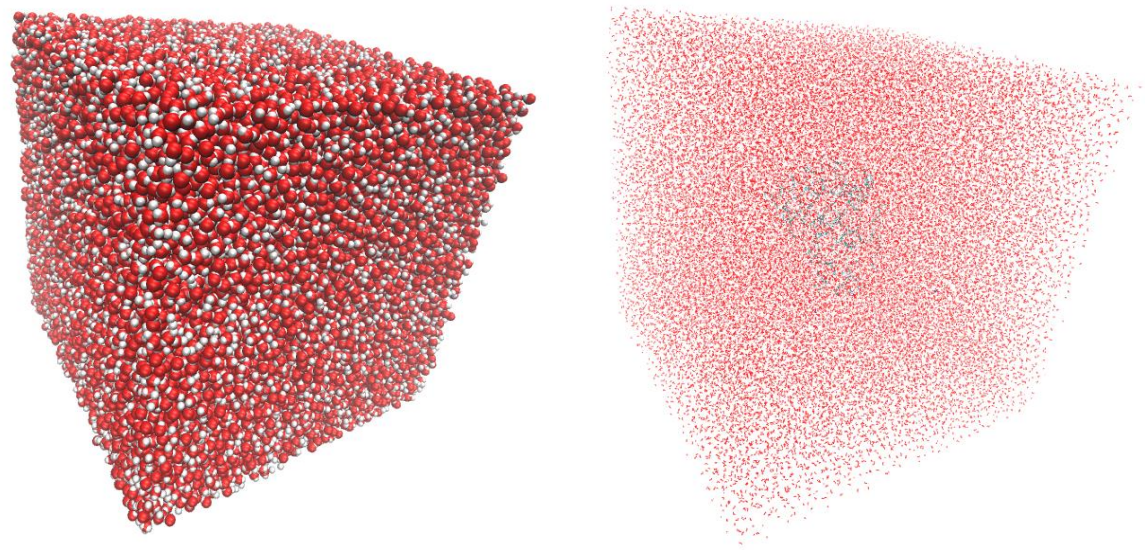
a. b.

Figure 2.1. Example of the CHARMM model that was used for simulation (E-o-coumaric acd-1EYQ complex).
(a) The RHDO water box illustrated with VDW Edgy material from VMD. (b) The RHDO water box with the protein-ligand complex in the centre of the box; the ligand is illustrated with the CPK model, the protein is illustrated with ribbons and water is illustrated with the line system from VMD.

2.2.4.2 Models for the GROMACS simulation and computational approach

The procedure that was used for GROMACS (210) simulations is quite similar to that used for CHARMM. The same initial structures used for CHARMM were used in the GROMACS simulation. The difference between the GROMACS and CHARMM simulations arises in the subsequent solvation and molecular dynamics preparation. For the CHARMM simulation the complex was placed in a rhombic dodecahedron for simulation while for GROMACS the complex system was solvated in a cubic solvent water environment (both systems incorporated periodic boundary conditions). For GROMACS, the ligand topology (itp and gmx files) were prepared using ACPYPE (211). The protein topology was prepared with GROMACS using the AMBER03 force fields. The receptor-ligand complex was centered in the explicitly solvated cubic water box, extending past the macromolecule for a further 3.0 Å in all dimensions. All systems were neutralized with Na^+ ions equivalent to their negative charge, minimized,

equilibrated and then subjected to the final MD simulation. Figure 2.2 shows an example of the GROMACS model used in the simulation.



a. b.

Figure 2.2 Example of a GROMACS model used for simulation (E-o-coumaric acd-1FM7 complex).
(a) The cubic water box illustrated with VDW Edgy material from VMD. (b) The cubic water box with the protein-ligand complex in the centre of the box; the ligand is illustrated with the CPK model, the protein is illustrated with ribbons and water is illustrated with the line system from VMD.

2.2.5 ONIOM QM/MM (Chapter 7)

The ONIOM study complements the MD study because the MD study only informs of the dynamics of the CHI- *o*-coumaric acid complexes, but cannot be used to describe reactivity or reactions at the active site, Therefore, the ligand and the receptor models were those used in the dynamics study (the ligand is the best docked ligand from docking and the receptors are the same PDB crystallographic structures used for docking). 1FM8- *o*-coumaric acid complex was chosen for the ONIOM reactivity study.

The docked complexes were saved as a combined PDB file with Discovery studio (201). The PDB coordinates were opened with Gauss View (199). The ligand and identified protein residues (close enough to protonate and deprotonate the ligand) were selected as the high layer, while the bulk of the protein was kept as the low layer in the two layer ONIOM calculations.

Two separate ONIOM calculations were performed; firstly where the higher layer was set at the semi-empirical AM1 level, and secondly where it was set at the hybrid DFT level with B3LYP/6-31G. In both cases the lower layer was set at the MM level, using the Amber 03 force field. For potential energy surface scans, redundant coordinates were set for exploring reactivity associated with protonation/deprotonation (involving residues that were modelled in the high layer). The Cartesian coordinates (or Z matrix) of the system were optimised (to minima, transition states, and under redundant coordinate conditions) with Gaussian 09 revision E01 (166). The results were visualised and analysed with Gauss View (199).

2.2.6 Adaptive partitioning QM/MM (Chapter 8)

The adaptive partitioning QM/MM was a way to combine and study the dynamics of the system while accounting for reactivity at the active-site. Two CHI-*o*-coumaric acid complexes were immersed in a water solvated, neutralized environment with the active-site defined at QM level and the bulk of the system defined at the MM level. The ligand and the receptor models used in this section were those for which data was available for comparative purposes from both MD and ONIOM QM/MM analyses. These complexes were the 1FM8 and 1EYQ complexes with *o*-coumaric acid. The initial complex coordinates were the same as were generated in Chapter 4 and used in Chapter 6. The protein was protonated with H++ (208), with salinity 0.15, internal dielectric 10, external dielectric 80, and pH of 6.5. HIS hydrogen atoms were derived from van der Waal's and H-bonding contacts. Under this more accurate approach, HIS195 in both cases was doubly deprotonated and so the residue name was changed to HIP prior to working with Ambertools for both 1FM8 and 1EYP.

The adaptive partitioning QM/MM study was carried out with the CP2K programme (191). The protein topology was prepared with Ambertools 16 using the GAFF (Generalized Amber Force Field), and the topology of the ligand was prepared using Antechamber. The protein-ligand complex topology was combined with Tleap, which also solvated (84.271 84.324 84.462 water box for 1FM8 and 84.291 84.824 84.432 water box for 1EYP) and neutralized the system with five Na+ for both 1EYP and 1FM8 complexes. The energy was minimised under Periodic Boundary Conditions (PBC) using the limited memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm optimizer for 1000 iterations with convergence criteria

MAX_DR of 0.01, RMS_DR of 0.005, MAX_FORCE 0.01 and RMS_FORCE 0.005. Poisson Ewald parameters were set to type SPME with GMAX 80 and ALPHA 0.40. Constant NVT dynamics was performed for 10000 steps at 298K using canonical sampling through a velocity rescaling (CSVR) thermostat. The final NPT MD was computed for 50000 steps, with the time-step of 0.5 fs, 1.0 bar pressure and 298 K using the CVSR thermostat.

The system was partitioned into the Primary Subsystem (PS), the Buffer region, and the Secondary Subsystem (SS). As already stated, the PS consists of the species in the active-site that are of interest, the Buffer region is a small region separating the PS and the SS, and the SS constitutes the bulk system. The PS was set in two different ways;

i) Only ligand as PS
ii) Ligand, tyrosine and serine amide NH, and glutamic acid backbone C=O as PS

The system with only the ligand as PS, was considered for QM/MM calculations only. The system with the ligand, tyrosine and serine amide NH and glutamic acid backbone C=O as PS were considered for QM/MM calculations, and for Adaptive QM/MM. The PS was computed at the QM level with DFTB+ with a radius from 0.0 to 4.5, the Buffer region extending from 4.5 to 7.0, and the SS was computed at the MM level with GAFF. Analysis was through the use of scripts and VMD for visualization.

CHAPTER 3
CONFORMATIONAL POPULATION ANALYSIS OF CINNAMIC ACID DERIVATIVES

3.1 Introduction to the chapter

Because this study focuses on the E → Z isomerization of *o*-coumaric acid, which is a derivative of cinnamic acid, this chapter concerns the conformational population analysis of several derivatives of cinnamic acid including *o*-coumaric acid. Some other derivatives of cinnamic acid were considered because they serve as references for coumaric acid. The E → Z isomerization may be viewed as a chemical process involving the protonation and deprotonation of coumaric acid (Figure 3.1) and therefore this chapter will also focus on the theoretical descriptors that are used to determine the reactivity of a chemical system. These descriptors include the ionization potential for removing an electron from coumaric acid (related to the energy of the highest occupied molecular orbital), the electron affinity for attaching an additional electron to coumaric acid (related to the energy of the lowest unoccupied molecular orbital), and molecular hardness (related to the energy difference of the frontier orbitals) and charge distribution (158,212).

3.2 Conformational preferences

Most molecules can adopt more than one low-energy conformation, and this conformational flexibility is available to cinnamic acid derivatives (213,214). When considering the molecular structures of cinnamic acid derivatives and their reactivity (in particular the ability of o-coumaric acid to be protonated and deprotonated), it was imperative to consider as many low energy conformations as possible, taking into account minor electronic and orbital variations as well as the usual steric influences. The conformers identified correspond to minima on the Potential Energy Surface (215), and the energies of these conformers are reported relative to the energy of the global minimum. The fractional population distribution of all minima must

follow the Boltzmann distribution (181,216). The conformer corresponding to the global minimum is expected to be the most highly populated. If the relative energy between the global minimum and the second most highly populated local minimum is too small, then both conformers are likely to have relatively high populations nature. Generally, the global minimum determines molecular properties such as chemical reactivity, but reaction rates may be dependent on the *reacting* conformation of the reactant (217).

3.2.1 Selection of conformations

Conformational isomers are stereoisomers which can be interconverted by internal rotation about a single bond, by extension or contraction of the bond length, deformation of a bond angle or combinations of these changes. Internal rotation is the most common means of conformational change. The interconversion between such conformers is restricted by a rotational energy barrier which must be overcome to interconvert one conformer (rotamer) to another (218). In the past decades, methods have been developed for sampling as many local minima as possible; these include replica-exchange molecular dynamics, metadynamics, and simulated annealing (219). The choice of a particular method is connected to the biological and physical characteristics of the system, most importantly system size (219). Coumaric acid has 4 rotatable single bonds (dihedrals whose rotation can result in significant conformational changes, as illustrated in figure 3.1). The current study takes into account these critical geometric features. Because of the practical rigidity of the benzene ring in cinnamic acid, conformers differ mostly in the orientations of the substituent groups with respect to this benzene ring for orientations resulting from rotations about the dihedrals identified in Figure 3.1.

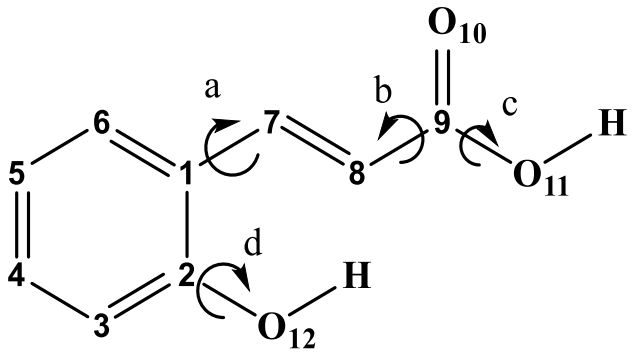


Figure 3.1. Structure of *o*-coumaric acid and the numbering utilized in this study, showing rotatable dihedrals.

Conformation may have a bearing on the E→Z isomerization of coumaric acid (which is one of the critical steps in the bioorganic transformation of coumaric acid to coumarin). The E→Z isomerization requires rotation around a double bond, and rotation around a double bond may require partial delocalization of electrons on the double bond. If this is the case, delocalization of the C(7)=C(8) double bond of coumaric acid could be achieved by protonation at O(10) or, increasingly unlikely, at C(8) or at C(7), and this could be coupled with deprotonation at O(12). Attention was also given to deprotonation of the coumaric acid structure at either O(11) or O(12), and to deprotonation at *both* O(11) and O(12). Chemically, we would expect protonation to be easier at, for instance O(10). However, given the microenvironments within the context of an enzyme active site, all possible protonation/deprotonation centres were considered.

The protonated coumaric acid structures have higher numbers of atoms than the simultaneously protonated and deprotonated, or deprotonated analogues considered in this study; they also have more conformational preferences that are of interest. As a result, conformers of the various protonated coumaric acid structures serve as reference structures in the selection of conformers of all protonated or deprotonated structures in this study.

When two neighbouring functional groups are oriented towards each other, the orientation is referred to as a *syn*, and when they are oriented away from each other the orientation is referred to as *anti* (Figure 3.2). Because the conformers considered in this study are rotamers, the labelling of conformers uses these *syn* and *anti* descriptors to define relative orientations about the rotatable bonds. The conformers of interest are those that differ in the orientation of the OH groups (these are conformers differing with rotation around bonds “c”, “d” and “e”). The changes in conformation resulting from rotation of the OH groups (bonds “c”, “d”, and “e”) are explored with other rotatable bonds kept constant. For the protonated coumaric acid structure, rotating bond “b” is irrelevant because from C(9) the structure is bilaterally symmetrical (resonance spreads the positive charge over both oxygen atoms), therefore for structures that were not protonated at O(10) the conformations which would result from rotation around bond “b” were not considered. Rotating around bond “a” results in significant changes in the conformation of the system. Therefore all conformers with bond “a” in a *syn* arrangement were denoted as “S”, and all conformers where bond “a” had an *anti*-arrangement were denoted “A”. Figure 3.2 shows an example of an “S” and an “A” arrangement.

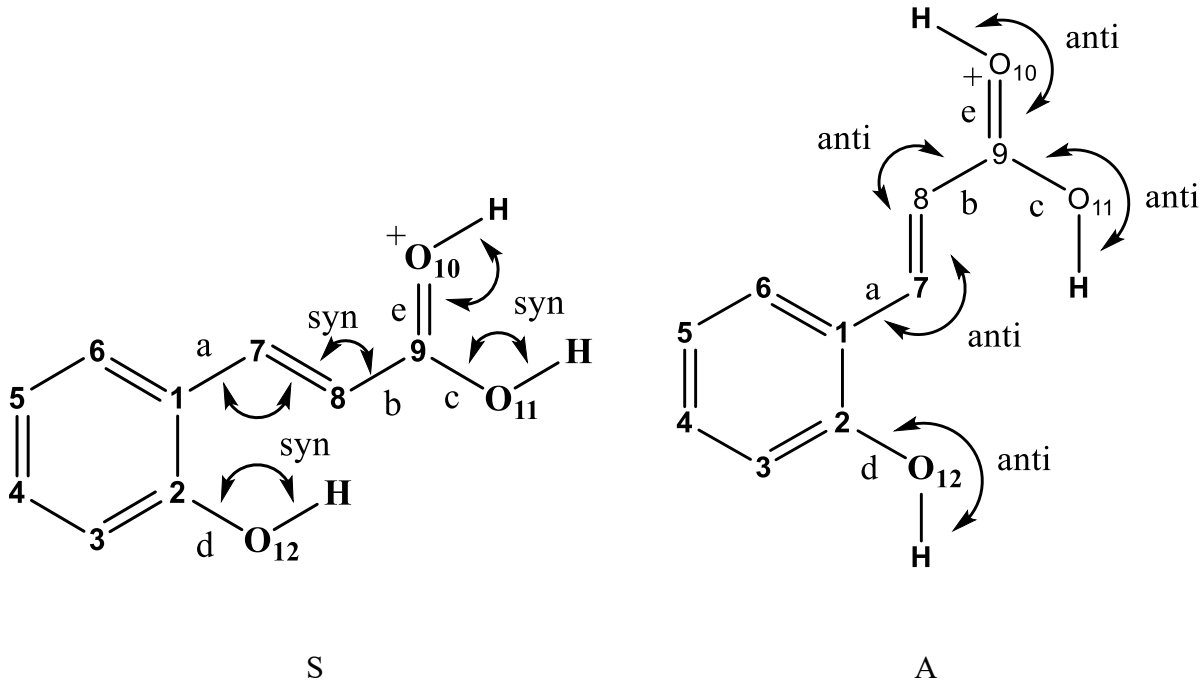


Figure 3.2. Structure of *o*-coumaric acid and the numbering utilized in this study, showing rotatable dihedrals and the "S" and "A" arrangement.

Figure 3.3a shows several "S" conformers of coumaric acid protonated at O(10); the numbering at the bottom of each conformer which, from this point on will be used to denote each particular conformer. The numbers assigned to the conformers do not indicate priority in terms of energy but were assigned to each conformer arbitrarily before optimization. Figure 3.3b shows the "S" conformers of *o*-coumaric acid differing in the orientation of OH groups and considered as starting geometries. The labelling of these conformers is derived from the similarities of the orientation of the OH groups with those of the protonated coumaric acid in Figure 3.3a. Figure 3.3c shows "S" conformers of *o*-coumaric acid deprotonated either at O(11) or at O(12) differing in the orientation of the OH groups and considered as starting geometries. Figure 3.3d "S" shows conformers of *o*-coumaric acid protonated at O(10) and deprotonated at O(12) differing by the orientation of OH groups, and considered as starting geometries.

Figures 3.3e to 3.3h show the conformers designated "A" which are similar to conformers designated "S" in Figure 3.3a to 3.3d but differ by orientation of the alkenoic acid chain.

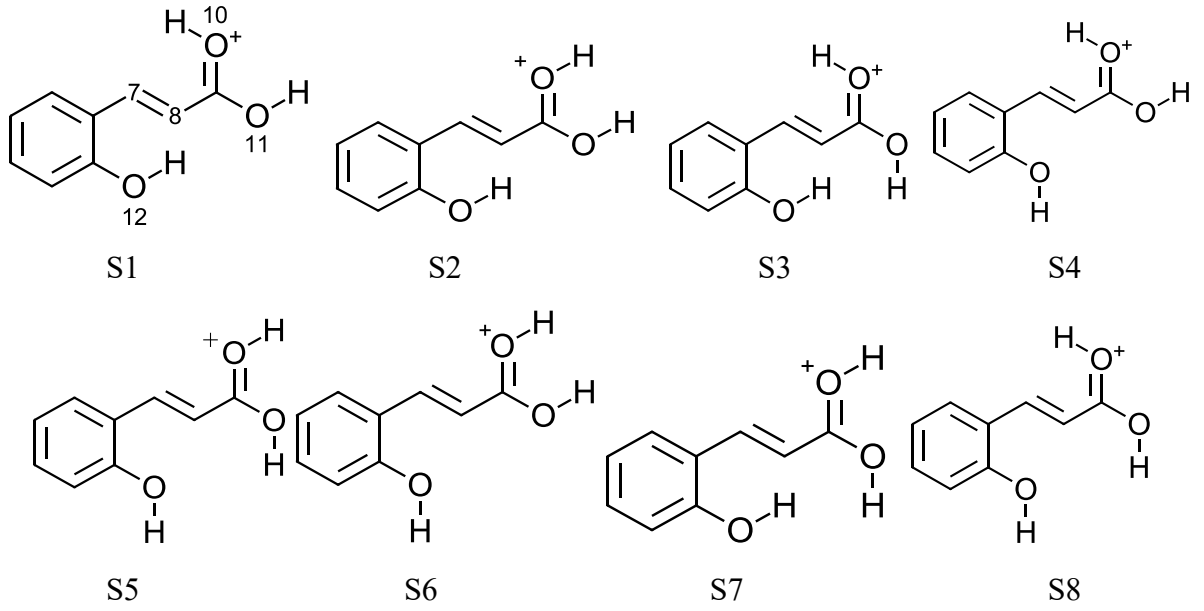


Figure 3.3a. Conformers of *o*-coumaric acid protonated at O(10) showing different possibilities considered as starting geometries. Conformers differ in the orientation of OH groups.

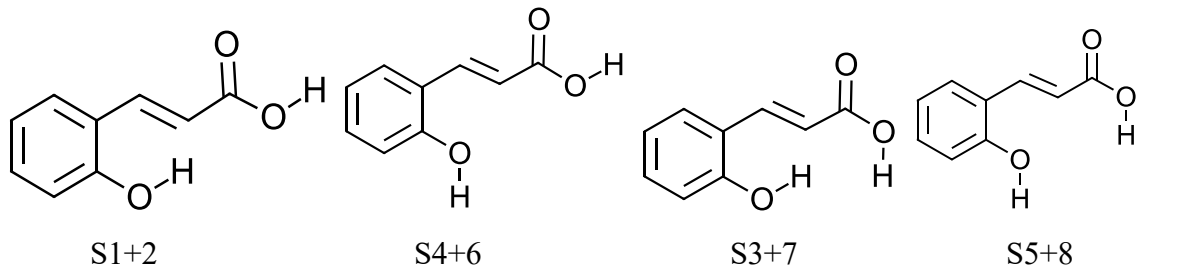


Figure 3.3b. Conformers of *o*-coumaric acid that are neither protonated or deprotonated showing different possibilities considered as starting geometries. Conformers differ in the orientation of OH groups. The naming of the conformers results from the similarity of that particular conformer to conformers in Figure 3.3a. For example conformer S1+2 is conformationally similar to conformer S1 and S2 in the absence of the hydrogen at O(10). Similar conformers of structures protonated at C(7) or C(8) have similar corresponding names.

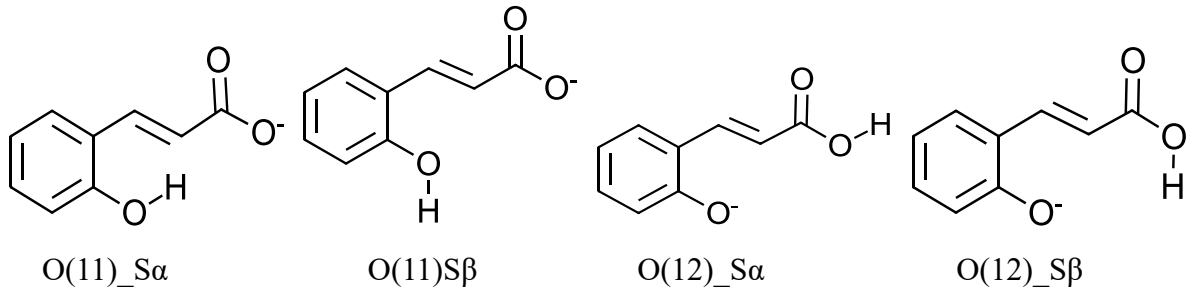


Figure 3.3c. Conformers of *o*-coumaric acid deprotonated at O(11) or O(12) showing different possibilities considered as starting geometries. Conformers differ in the orientation of OH groups and the location of the anionic oxygen. α and β denotes the orientation of the available OH.

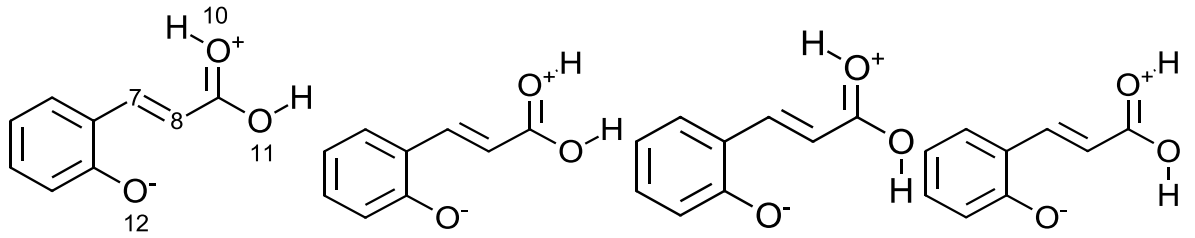


Figure 3.3d. Conformers of *o*-coumaric acid protonated at O(10) and deprotonated at O(12) showing different possibilities considered as starting geometries. Conformers differ in the orientation of OH groups. The naming of the conformers results from the similarity of that particular conformer to the corresponding conformers in Figure 3.3a. because of the absence of the hydrogen at O(12) the conformers illustrated in this figure are similar to two conformers in Figure 3.3a, but however takes the name of the corresponding conformer with the lowest numerical value. For example conformer S1 is similar to S1 and S4 and is denoted as S1'. Where prime symbol denotes the absence of a hydrogen at O(12).

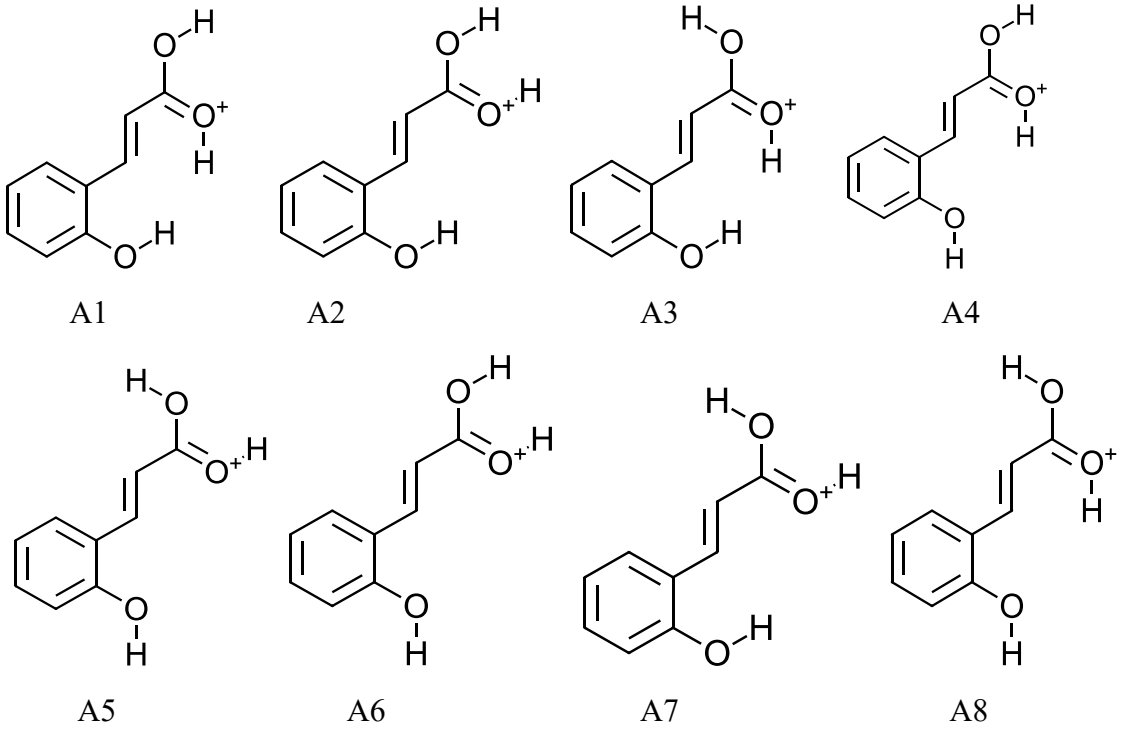


Figure 3.3e. Conformers of *o*-coumaric acid protonated at O(10) and in which the alkenoic acid chain is oriented away from the 12-OH group (i.e. an *anti* arrangement) showing different possibilities considered as starting geometries. Conformers differ by the orientation of OH groups.

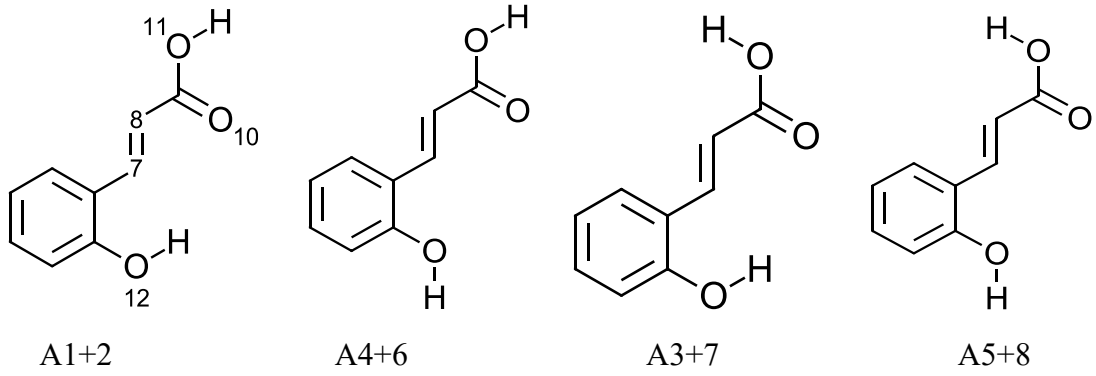


Figure 3.3f. Conformers of *o*-coumaric acid that are neither protonated nor deprotonated but in which the alkenoic acid chain oriented away from the 12-OH in group showing different possibilities considered as starting geometries. Conformers differ in the orientation of OH groups.

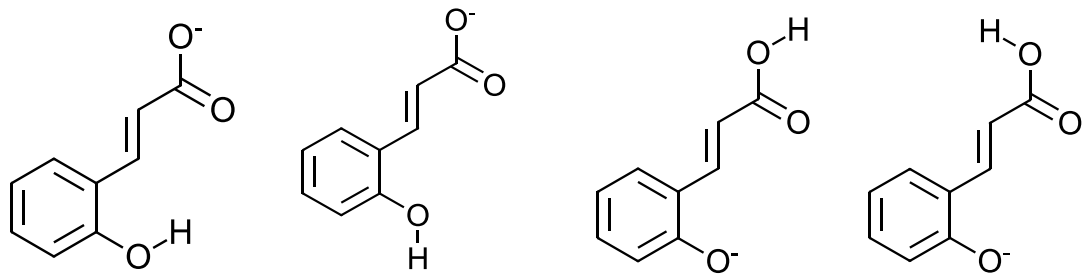


Figure 3.3g. Conformers of *o*-coumaric acid deprotonated at O(11) or O(12)in which the alkenoic acid chain is oriented away from the 12-OH group showing different possibilities considered as starting geometries. Conformers differ in the orientation of OH groups. α and β denotes the orientation of the available OH.

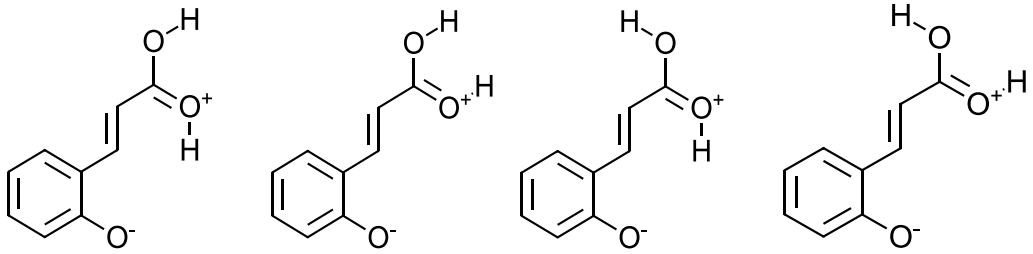


Figure 3.3h. Conformers of *o*-coumaric acid protonated at O(10) and deprotonated at O(12) and in which the alkenoic acid chain oriented away from the 12-OH group showing different possibilities considered as starting geometries. Conformers differ in the orientation of OH groups. Prime symbol denotes the absence of a hydrogen at O(12).

For the sake of consistency and to facilitate immediate connections among different types of information, the conformers of the various derivatives considered in this research are denoted using a similar naming system to that of the *o*-coumaric acid conformers. The main features of these systems involve the orientation of the OH groups or their *o*-methylated ether analogues; the naming system denotes in each case the relative orientation of other OH and OMe groups. For different positions (*ortho*, *meta*, and *para*) of the phenolic OH group, the numbering system follows the designated orientations of the OH groups in *o*-coumaric acid.

3.2.2 Computational approach

The study was carried out *in vacuo* and with solvent correction at the hybrid DFT level using the B3LYP/6-31+G(d,p) model chemistry. Solvent-corrected calculations were performed using the polarizable continuum model (PCM) (180) for water-solvated and chloroform-solvated systems. All of the calculations were performed with Gaussian 09, revision E01 (220,221). The results were visualized using GaussView (199).

3.2.3 Results

The reported results were obtained from the optimization and conformational exploration of three hydroxy derivatives of cinnamic acid, *ortho*-, *meta*- and *para*- coumaric acid, and two derivatives with through-conjugation electron donating and withdrawing groups, viz., *p*-methoxycinnamic acid, and *p*-nitrocinnamic acid respectively. These derivatives of cinnamic acid were either protonated at C(7), C(8) or O(10), and were also deprotonated either at O(11) or O(12) or at both O(11) and O(12). Protonation and deprotonation resulted in derivatives differing in the number of atoms. To compensate for the missing protons in the structures, the energy of the proton(s) removed was added to the total energy of the system in kcal/mol. The energy value for the proton used for this correction (-6.28 kcal/mol) was obtained from the literature (222).

3.2.3.1 Relative energies *in vacuo*

For illustrative purposes, Figures 3.4- 3.8 show the geometry-optimised conformers for the 16 *o*-coumaric acid arrangements, while Tables 3.1-3.5 summarise the complementary energies. The geometry-optimised structures of the cinnamic acid derivatives and their relative energies are provided in Appendix A (supplementary material in the disc). The results show that O(10) protonated *o*-coumaric acid conformers have lower energies relative to the mass balanced *o*-coumaric acid conformers with different protonation or deprotonation states (Table 3.1). Generally from these results, we can deduce that the increase in relative energy follows the following general pattern.

Structures protonated at O(10) < *o*-coumaric acid protonated at either C(7) or C(8) < *o*-coumaric acid that is neither protonated nor deprotonated < *o*-coumaric acid that is protonated at O(10) and deprotonated at O(12) < *o*-coumaric acid that is deprotonated at O(12) < *o*-coumaric acid that is deprotonated at O(11) < *o*-coumaric acid that is deprotonated at both O(11) and O(12)

The global minimum of all the energy minima obtained in the conformational population analysis of *o*-coumaric acid is a conformer protonated at O(10) with all three OH groups

oriented in the same direction (conformer S5). The energy of conformer S5 of *o*-coumaric acid is relatively close to that of conformer A5 which differs from it in the orientation of the alkenoic acid chain. Because the energy difference between conformer S5 and conformer A5 is less than 1 kcal/mol, theoretically, conformer A5 will be relatively as highly populated as conformer S5. For further computational studies involving *o*-coumaric acid protonated at O(10), both conformers S5 and A5 were considered as suitable starting points for further study.

The relative energy data indicates that it is theoretically easier to protonate *o*-coumaric acid than to deprotonate it. *i.e.* protonation of *o*-coumaric acid results in a system that is lower in mass-balanced energy than the neutral system, but deprotonation of *o*-coumaric acid results in a system that is higher in mass balanced energy than the neutral system. To simultaneously protonate and deprotonate *o*-coumaric acid (at O(10) and O(12) respectively) only about 20 kcal/mol energy is required (in terms of mass-balanced energy) relative to the neutral *o*-coumaric acid substrate.

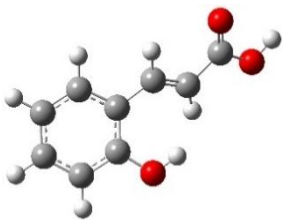
In this study, the effects of protonating or deprotonating o-coumaric acid were explored so as to theoretically determine an energetically favourable mechanism for the weakening of the C(7)=C(8) double bond in order to permit the E→Z isomerization which is a critical step in the interconversion of coumaric acid to coumarins. Consequently, attention was focused on the lengths of the C(7)=C(8) double bond of the different conformers of all protonation states of *o*-coumaric acid. Table 3.6 shows the calculated C(7)=C(8) bond lengths corresponding to all conformers of all *o*-coumaric acid structures with different protonation states. From the literature, a normal C-C single bond has a length of about 1.54 Å and a normal double bond has a length of about 1.35 Å (223). The C(7)=C(8) double bond in a neutral *o*-coumaric acid is at about 1.35 Å for all minima sampled in this study. All sampled, deprotonated conformers have a bond length around 1.3 Å, meaning that deprotonation of *o*-coumaric acid does not weaken the C(7)=C(8) double bond significantly (but may have the opposite effect). Protonation of *o*-coumaric acid at O(10), however, increases this bond length and, hence, decreases the strength of the C(7)=C(8) double bond; the average length of the C(7)=C(8) double bond in the O(10) protonated *o*-coumaric acids is about 1.38 Å. Simultaneous protonation at O(10) and deprotonation at O(12), however, weakens the C(7)=C(8) double bond more than protonation at O(10) alone (in agreement to the proposal from literature (43)). Trends in the weakening of this C(7)=C(8) bond may be followed as follows. As start, the average C(7)=C(8) double bond length of sampled neutral *o*-coumaric acid minima is about 1.35 Å, but

this increases to about 1.38 Å when protonation is at O(10). Protonation at O(10) and deprotonation at O(12) increases this even further to about 1.42 Å. Finally, protonating on either C(7) or C(8) directly weakens the C(7)=C(8) double bond by effectively breaking the π-bond. The C(7)=C(8) has the average of 1.48 Å when protonation is at C(8) – in line with literature value for the length of a single C-C bond. Protonation at C(7) provides the greatest effect with the C(7)=C(8) bond having an average length of about 1.50 Å.

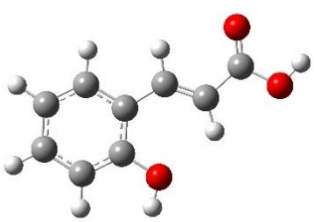
Most of the sampled conformations maintained geometry throughout optimization such that the geometry of the optimized conformer was similar to that of the input structure (Figure 3.4). However, significant structural changes were observed in the conformers which had been protonated at C(7). Protonation at C(7) seems to be very unfavourable in o-coumaric acid and most of these structures were transformed to C(8) protonated structures during optimization, due to migration of the proton at C(7) to C(8). In structures were proton migration did not occur, the system stabilized itself by forming heterocyclic structures, the side chain either forms a five membered ring fused to the benzene ring, with the hydrogen of O(12) being somewhat shared between O(12) and O(10), or it forms a Z-isomer stabilized by the hydrogen of O(12) being shared between O(12) and O(10) (See Figure 3.4, o-coumaric_C7_S5+8, for example).

The 1.50 Å bond length from C(7) protonation suggests the ease of a C(7)=C(8) protonated *E*-coumarin to be interconverted to a Z-coumarin via rotations on the C(7)=C(8) bond. Theoretically, protonating at either C(7) and C(8) weakens and lengthens the C(7)=C(8) double bond better that the other approaches considered, but such protonated intermediates could be problematic practically in the synthesis of coumarins because they might cyclise into unwanted bicyclic structures as observed with these geometry optimizations. It is likely that the activation energy required for this protonation, or the ease of access of protonation may be factors that work against these rearrangements.

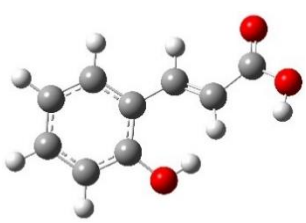
S-conformers



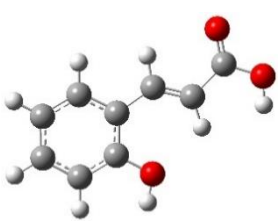
o-coumaric_S1+2
214.36700 kcal/mol



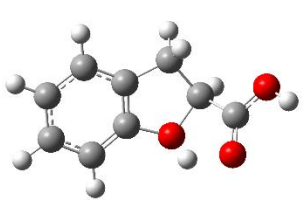
o-coumaric_S4 +6
213.24722 kcal/mol



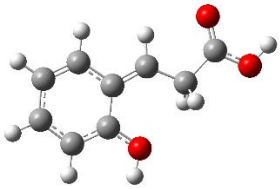
o-coumaric_S3 +7
220.87847 kcal/mol



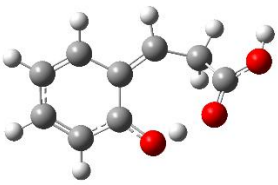
o-coumaric_S5+8
218.92449 kcal/mol



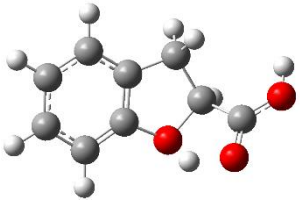
o-coumaric_C7_ S1+2
19.25656 kcal/mol



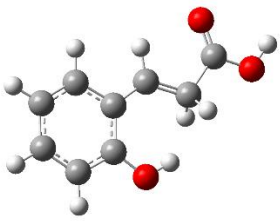
o-coumaric_ C7_ S4+6
8.75894 kcal/mol



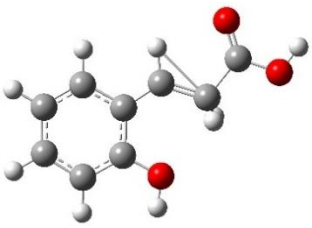
o-coumaric_C7_S3+7
14.83266 kcal/mol



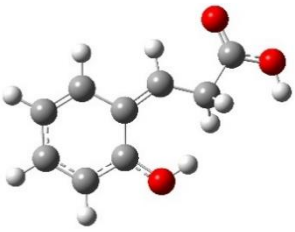
o-coumaric_C7_S5+8
26.87200 kcal/mol



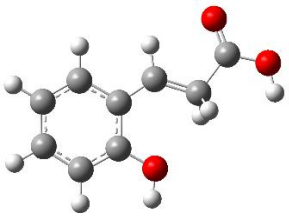
o-coumaric_C8_S1+2
15.77775 kcal/mol



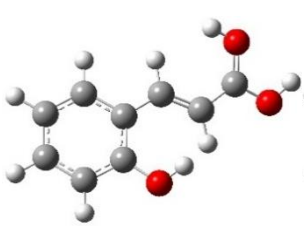
o-coumaric_C8_S4 +6
8.75900 kcal/mol



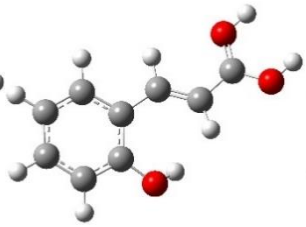
o-coumaric_C8_S3+7
25.32571 kcal/mol



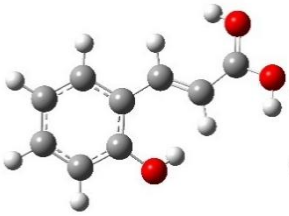
o-coumaric_C8_S5+8
17.56293 kcal/mol



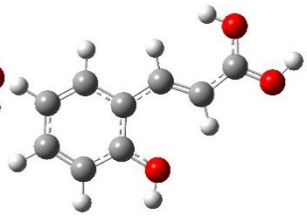
o-coumaric_S1
10.62952 kcal/mol



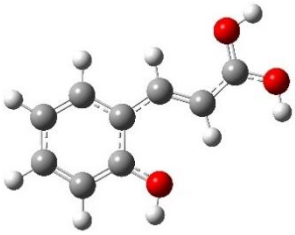
o-coumaric_S2
12.46949 kcal/mol



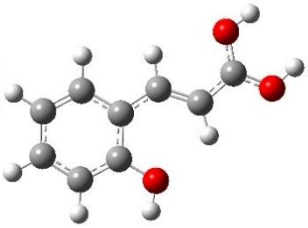
o-coumaric_S3
14.34202 kcal/mol



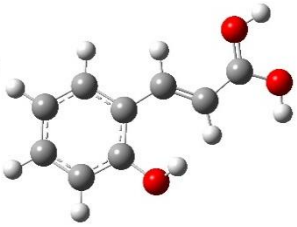
o-coumaric_S4
3.01145 kcal/mol



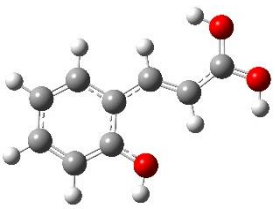
o-coumaric_S5
0.00000 kcal/mol



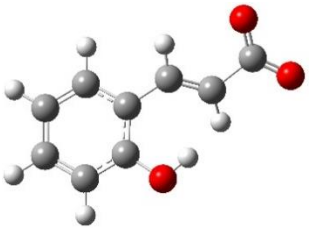
o-coumaric_S6
5.20118 kcal/mol



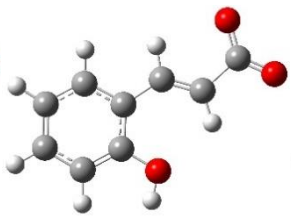
o-coumaric_S7
8.01830 kcal/mol



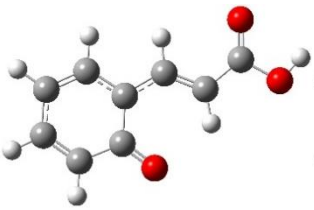
o-coumaric_S8
6.16214 kcal/mol



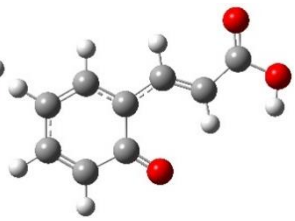
o-coumaric_ O(11)_Sα
551.72835 kcal/mol



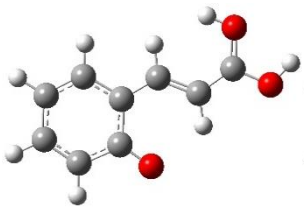
o-coumaric_ O(11)_Sβ
558.73090 kcal/mol



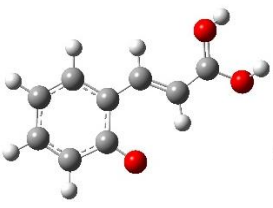
o-coumaric _ O(12)_Sα
543.26509 kcal/mol



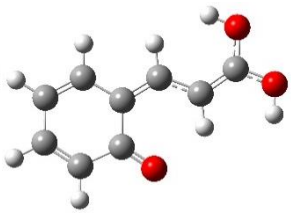
o-coumaric _ O(12)_Sβ
545.09690 kcal/mol



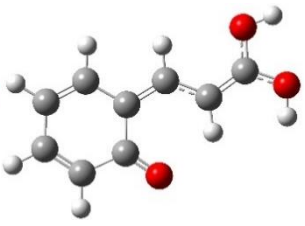
o-coumaric_ S1'
243.41468 kcal/mol



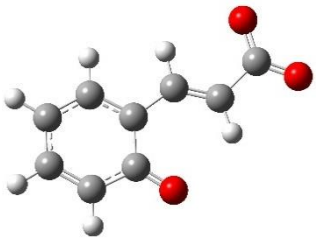
o-coumaric_ S2'
246.97187 kcal/mol



o-coumaric_ S3'
243.25115 kcal/mol

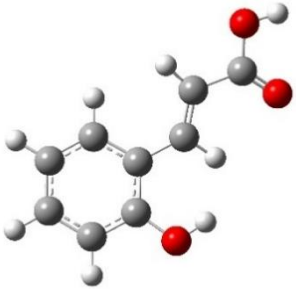


o-coumaric_ S5'
240.71346 kcal/mol

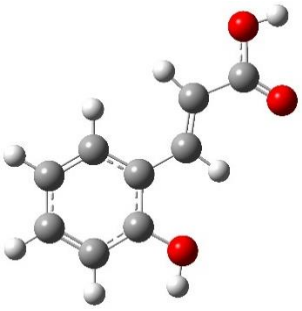


o-coumaric_ deprot_S_O11/O12
960.26460 kcal/mol

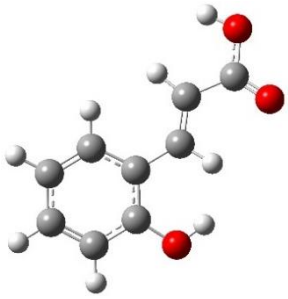
A-conformers



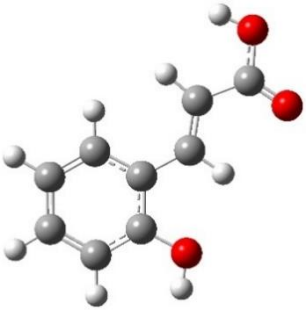
o-coumaric_A1+2
214.44537 kcal/mol



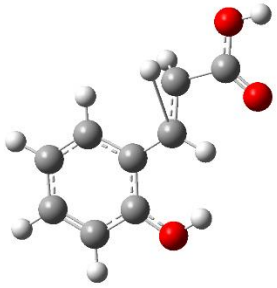
o-coumaric_A4 +6
213.64612 kcal/mol



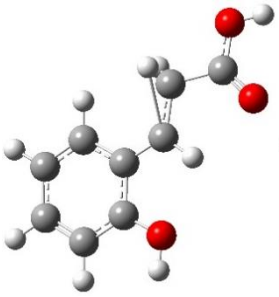
o-coumaric_A3 +7
220.75008 kcal/mol



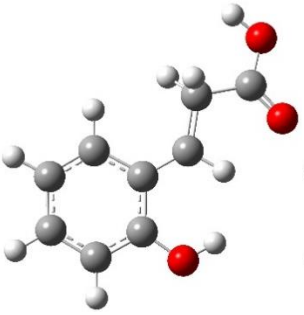
o-coumaric_A5+8
220.01772 kcal/mol



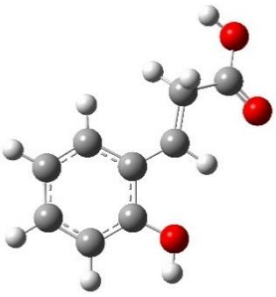
o-coumaric_C7_ A1+2
13.61286 kcal/mol



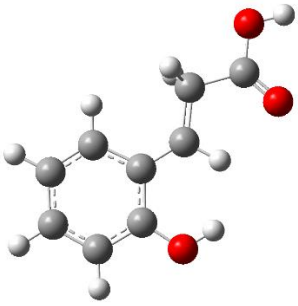
o-coumaric_ C7_ A4+6
9.57036 kcal/mol



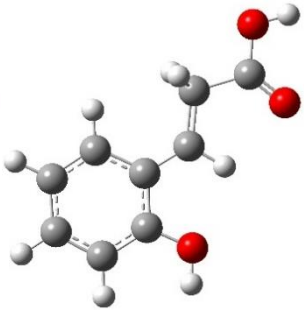
o-coumaric_C7_A3+7
22.16805 kcal/mol



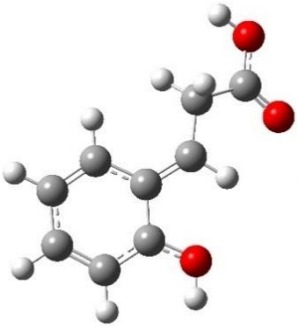
o-coumaric_C7_A5+8
18.23015 kcal/mol



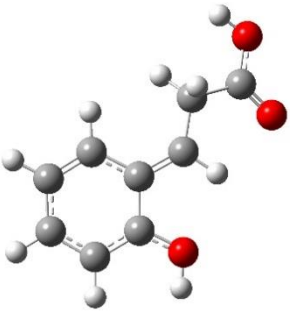
o-coumaric_C8_A1+2
13.61292 kcal/mol



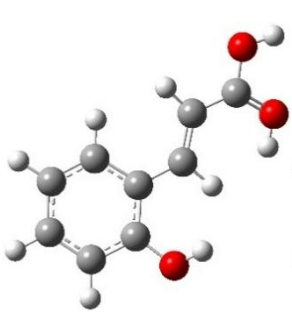
o-coumaric_C8_A4 +6
9.57457 kcal/mol



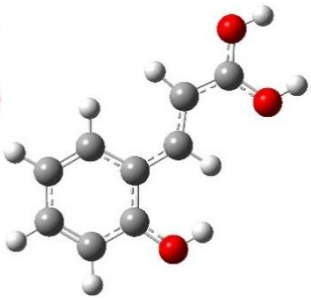
o-coumaric_C8_A3+7
22.16805 kcal/mol



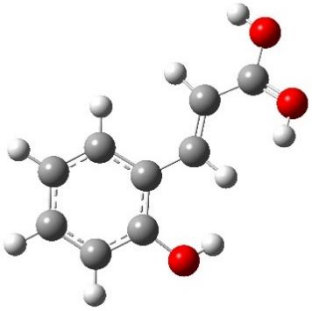
o-coumaric_C8_A5+8
18.23015 kcal/mol



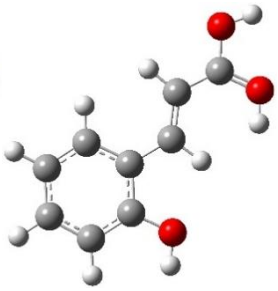
o-coumaric_A1
10.70464 kcal/mol



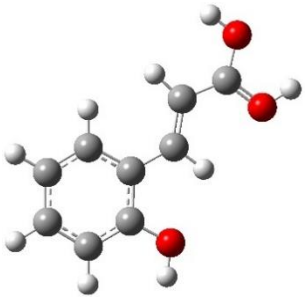
o-coumaric_A2
11.63052 kcal/mol



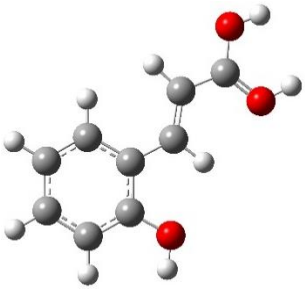
o-coumaric_A3
14.35451 kcal/mol



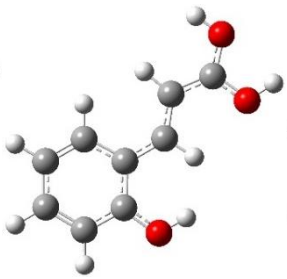
o-coumaric_A4
2.78348 kcal/mol



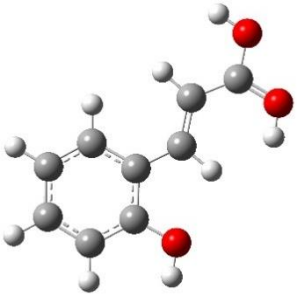
o-coumaric_A5
2.78348 kcal/mol



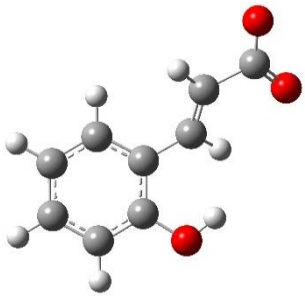
o-coumaric_A6
5.61904 kcal/mol



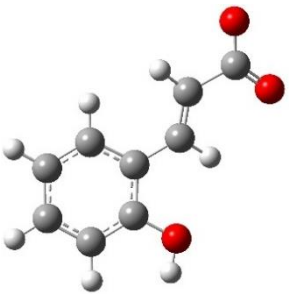
o-coumaric_A7
6.97570 kcal/mol



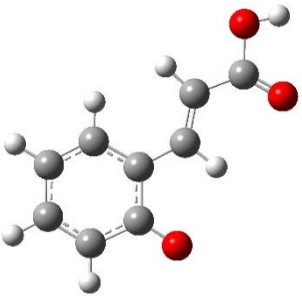
o-coumaric_A8
6.49980 kcal/mol



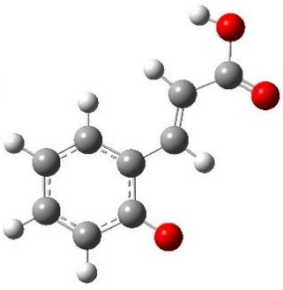
o-coumaric_ O(11)_Aα
552.24943 kcal/mol



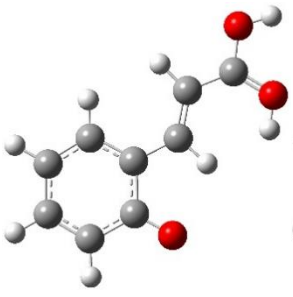
o-coumaric_ O(11)_Aβ
557.99484 kcal/mol



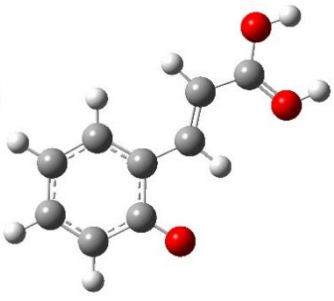
o-coumaric _ O(12)_Aα
544.33994 kcal/mol



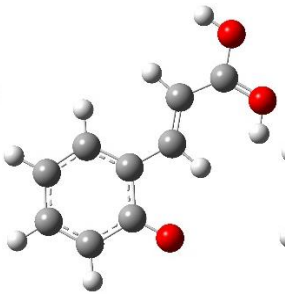
o-coumaric _ O(12)_Aβ
548.70573 kcal/mol



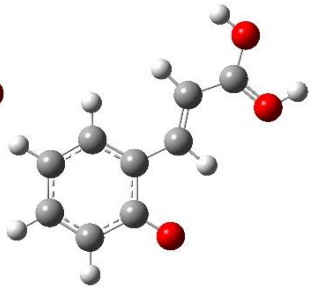
o-coumaric_ A1'
242.32288 kcal/mol



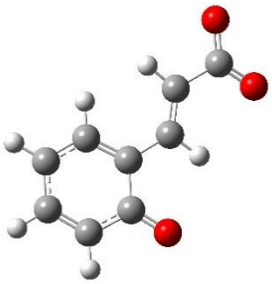
o-coumaric_ A2'
247.72744 kcal/mol



o-coumaric_ A3'
244.45903 kcal/mol



o-coumaric_ A5'
243.94159 kcal/mol



o-coumaric_ deprot_A_O11/O12
958.82530 kcal/mol

Figure 3.4. Calculated conformers of o-coumaric acid protonated at either C7, C8 or the O(10) oxygen. DFT/B3LYP/6-31+G(d,p) results in *vacuo*.
In the acronyms, the first letter (o-coumaric) informs that this is the ortho substituted coumaric acid. "S" denotes a *syn* arrangement, and "A" denotes an *anty* arrangement. The numbers 1 to 8 represent the orientation of the OH groups. The terms 1+2, 3+7, 4+6 and 5+8 result from protonation is at C7 or C8 in which case the orientation of the O(10) oxygen is symmetrical, therefore the conformers of coumaric acid will differ by the orientation of the alkenoic OH and the phenol OH, and their conformers are compared with their similar conformer when protonation is at the O(10) oxygen with the H of the O(10) oxygen oriented to the left or the right. α and β denotes the orientation of OH for deprotonated structures. Prime (') denotes deprotonation at O(12).

The first part if this section focused particularly on *o*-coumaric acid because it is the main structure of interest. The *m*- and *p*-coumaric acid where also studied in a similar manner to *o*-coumaric acid, because they are structural isomers of o-coumaric acid, and therefore, could provide information on how other spatial arrangements of coumaric acid affect the theoretical descriptors considered in this study. The additional results from the calculation of the similar conformers of all three substitutions of coumaric acid also increase the statistical information available and therefore reduce the probability of assumptions based on biased results. The other structural effects taken into account were having an electron withdrawing or electron donating functional group substituting the para- OH of *p*-coumaric acid. To take into account the effects of an electron withdrawing group, *p*-methoxycinnamic acid was considered, and to take into account the effects of an electron donor *p*-nitrocinnamic acid was considered.

Considering all conformers of these four additional structures increases dramatically the data available, it is neither practical nor illuminative to discuss the results on an individual basis. Therefore for simplicity, similar conformers of *p*-coumaric acid, *m*-coumaric acid, *p*-methoxycinnamic acid, *p*-nitrocinnamic where discussed together in comparison with the results of o-coumaric acid. Figures 3.5-3.8 shows the conformers considered for each structure and the name of that particular conformer and it relative energy are given below the figure. Table 3.2-3.5 gives the relative energies of these sets of conformers. The results in Table 3.2-

3.5 show that the trends of the energies of the conformational structures of all five cinnamic acid derivatives are the same although the exact values are different, *i.e.* they follow the same general pattern in mass corrected energy values:

Structures protonated at O(10) < cinnamic acid protonated at either C(7) or C(8) < cinnamic acid that is neither protonated nor deprotonated < cinnamic acid that is protonated at O(10) and deprotonated at O(12) < cinnamic acid that is deprotonated at O(12) < cinnamic acid that is deprotonated at O(11) < cinnamic acid that is deprotonated at both O(11) and O(12)

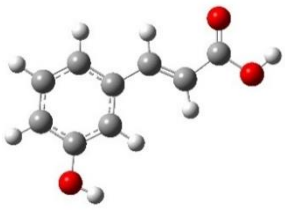
The bond lengths also follow the same general pattern when compared across similar conformers of the five cinnamic acid derivatives, as the one discussed for *o*-coumaric acid. The bond length increase across structures in the following general pattern (table 3.7-3.10).

The neutral structure, cinnamic acid deprotonated at O(11) and those that are deprotonated at both O(11) and O(12) – generally around 1.34 Å < cinnamic acid derivatives deprotonated at O(12) generally with the bond length of about 1.35 Å < cinnamic acid derivatives protonated at O(10) bond length generally around 1.37 Å < cinnamic acid derivatives protonated at O(10) and deprotonated at O(12) bond length generally around 1.38 Å < acid protonated at either C(7) or C(8) bond length generally around 1.47 Å

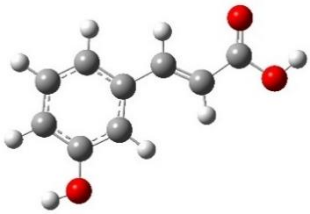
Although the patterns are the same the exact bond length values do vary.

The other trends observed across structures are the change in structural arrangements of the optimised geometries of structures protonated at C(7) (Figures 3.5-3.8). As before it is generally observed that structures protonated act C(7) rearrange during optimization, the proton that originates at C(7) either migrates to C(8), or the structure stabilizes by formation of an additional ring; The additional rign may either be a 5 membered heterogeneous ring with O(13) or a three membered ring made up of C(1), C(7), and C(8).

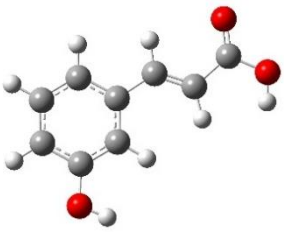
S-Conformer



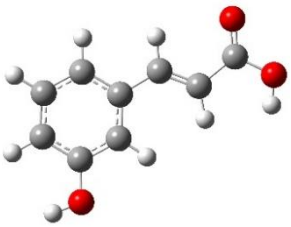
m-coumaric_S1+2
208.6691 kcal/mol



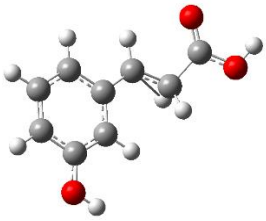
m-coumaric_S4 +6
208.7769 kcal/mol



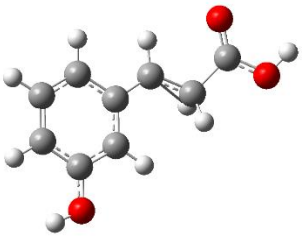
m-coumaric_S3 +7
215.2515 kcal/mol



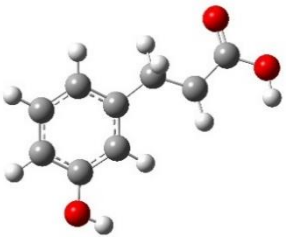
m-coumaric_S5+8
215.0167kcal/mol



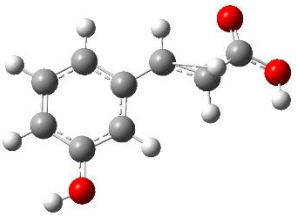
m-coumaric_C7_ S1+2
12.0996 kcal/mol



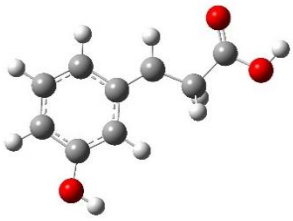
m-coumaric_ C7_ S4+6
12.0236 kcal/mol



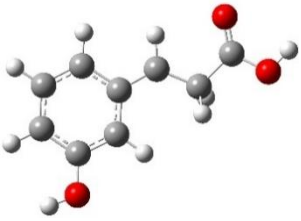
m-coumaric_C7_S3+7
58.2116 kcal/mol



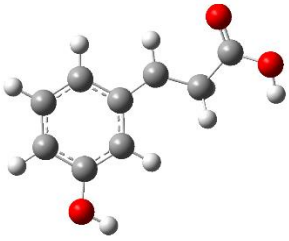
m-coumaric_C7_S5+8
20.6483 kcal/mol



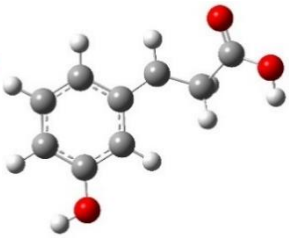
m-coumaric_C8_S1+2
12.0996 kcal/mol



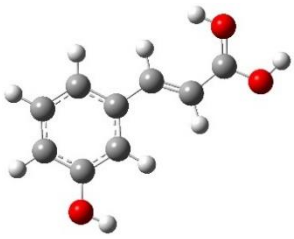
m-coumaric_C8_S4 +6
12.0236 kcal/mol



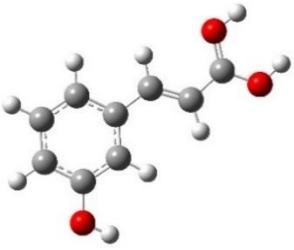
m-coumaric_C8_S3+7
21.0134 kcal/mol



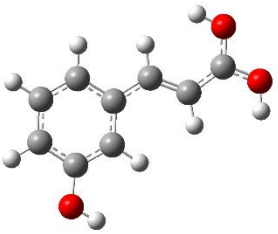
m-coumaric_C8_S5+8
20.6483 kcal/mol



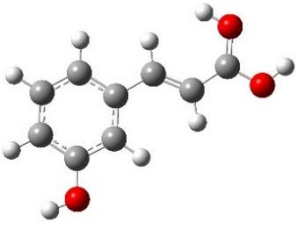
m-coumaric_S1
3.3673 kcal/mol



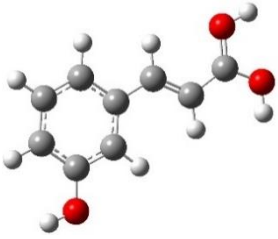
m-coumaric_S2
5.5055 kcal/mol



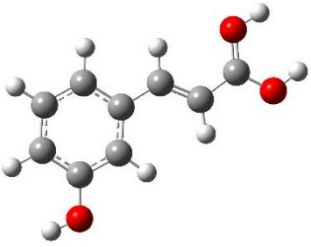
m-coumaric_S3
7.3460 kcal/mol



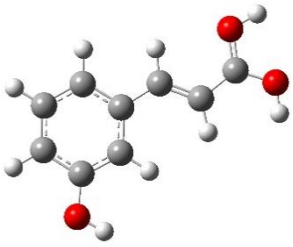
m-coumaric_S4
2.5823 kcal/mol



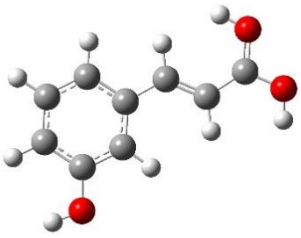
m-coumaric_S5
0.0000 kcal/mol



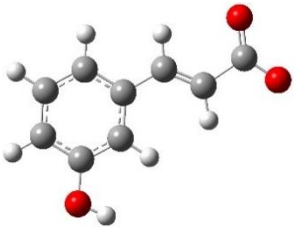
m-coumaric_S6
4.7285 kcal/mol



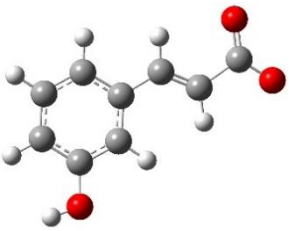
m-coumaric_S7
1.1186 kcal/mol



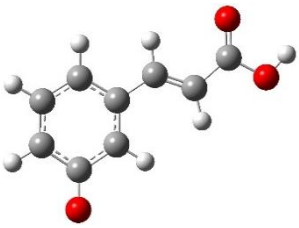
m-coumaric_S8
6.1916 kcal/mol



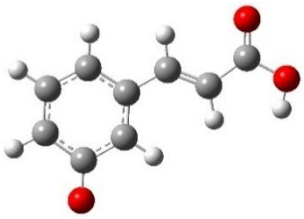
m-coumaric_ O(11)_Sα
549.0997 kcal/mol



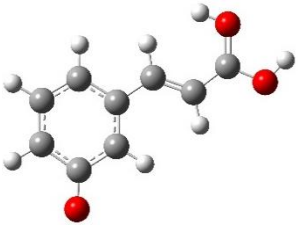
m-coumaric_ O(11)_Sβ
550.7850 kcal/mol



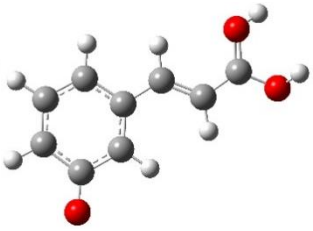
m-coumaric _ O(12)_Sα
546.2696 kcal/mol



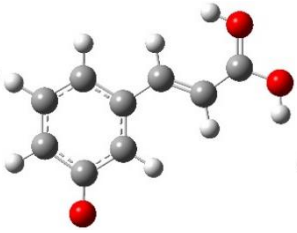
m-coumaric _ O(12)_Sβ
549.9381 kcal/mol



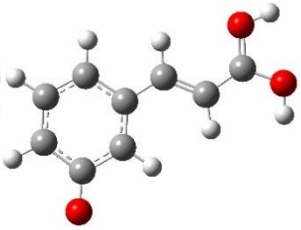
m-coumaric_ S1'
265.8136 kcal/mol



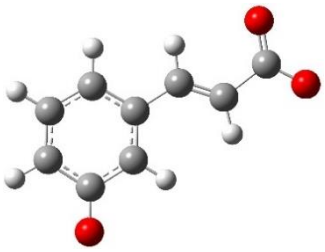
m-coumaric_ S2'
270.1057 kcal/mol



m-coumaric_ S3'
267.0484 kcal/mol

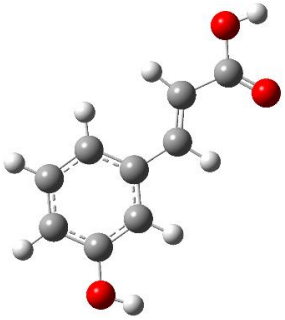


m-coumaric_ S5'
264.2891kcal/mol

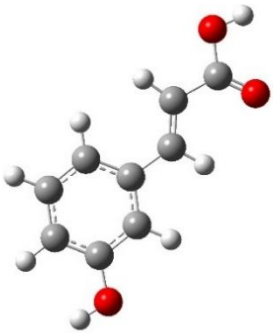


m-coumaric_ deprot_S_O11/O12
949.0888 kcal/mol

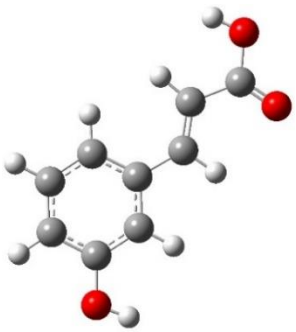
A-conformers



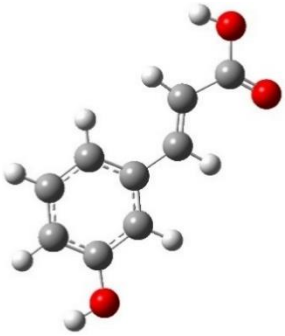
m-coumaric_A1+2
208.7389 kcal/mol



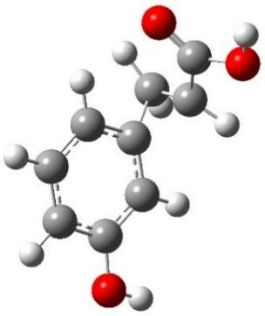
m-coumaric_A4 +6
208.8368 kcal/mol



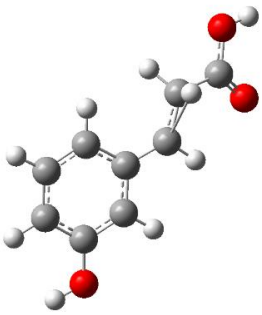
m-coumaric_A3 +7
215.0649 kcal/mol



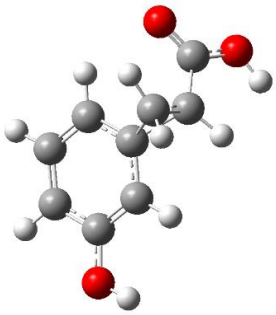
m-coumaric_A5+8
215.2218 kcal/mol



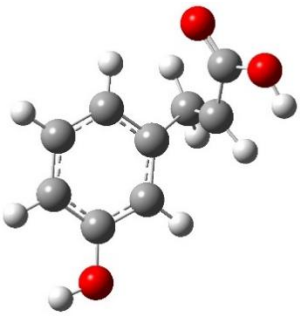
m-coumaric_C7_ A1+2
24.9156 kcal/mol



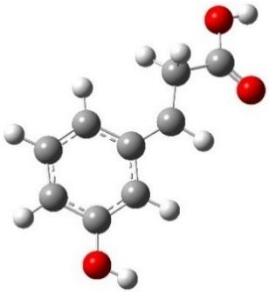
m-coumaric_ C7_ A4+6
12.5569 kcal/mol



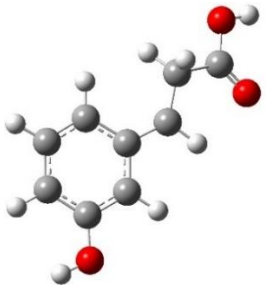
m-coumaric_C7_A3+7
33.8180 kcal/mol



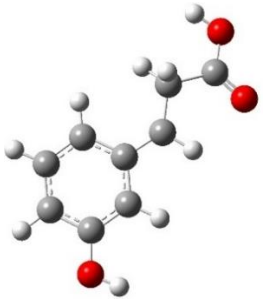
m-coumaric_C7_A5+8
33.2533 kcal/mol



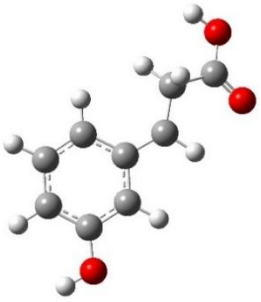
m-coumaric_C8_A1+2
12.4718 kcal/mol



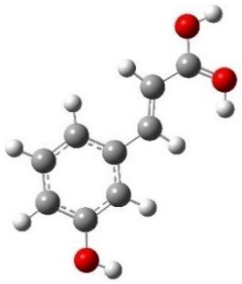
m-coumaric_C8_A4 +6
12.5569 kcal/mol



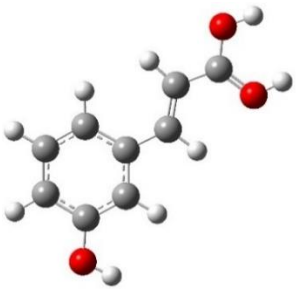
m-coumaric_C8_A3+7
21.1502 kcal/mol



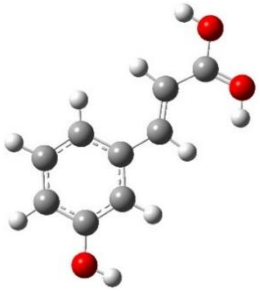
m-coumaric_C8_A5+8
21.2947 kcal/mol



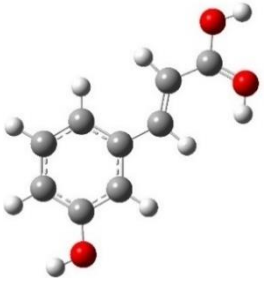
m-coumaric_A1
3.8136 kcal/mol



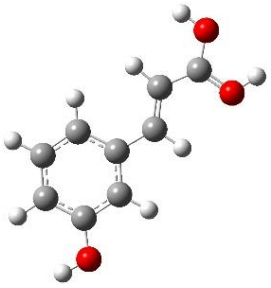
m-coumaric_A2
5.7368 kcal/mol



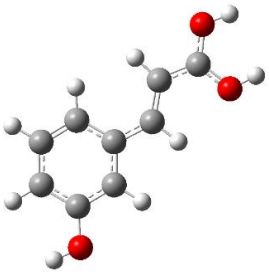
m-coumaric_A3
7.5408 kcal/mol



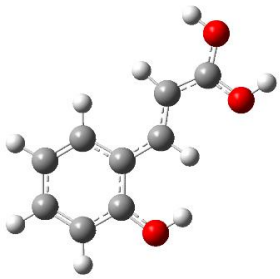
m-coumaric_A4
2.7815 kcal/mol



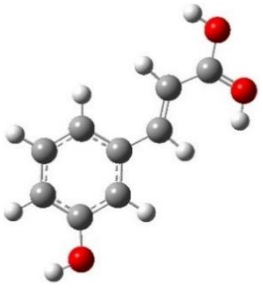
m-coumaric_A5
0.4158 kcal/mol



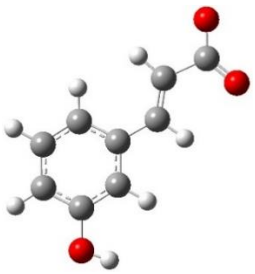
m-coumaric_A6
4.9922 kcal/mol



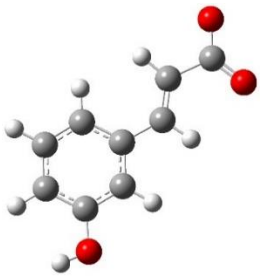
m-coumaric_A7
1.1186 kcal/mol



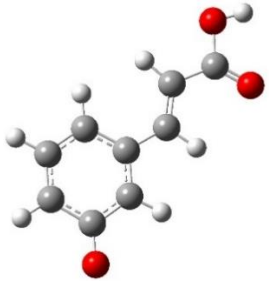
m-coumaric_A8
6.5508 kcal/mol



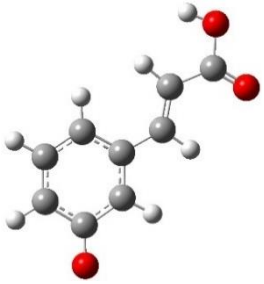
m-coumaric_ O(11)_Aα
549.3462 kcal/mol



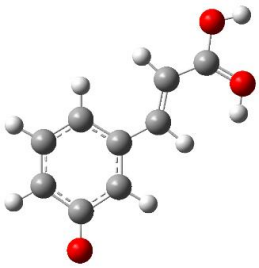
m-coumaric_ O(11)_Aβ
550.5838 kcal/mol



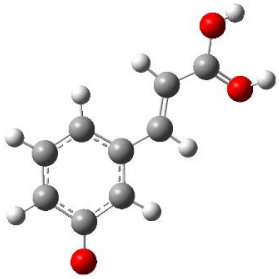
m-coumaric _ O(12)_Aα
546.2692 kcal/mol



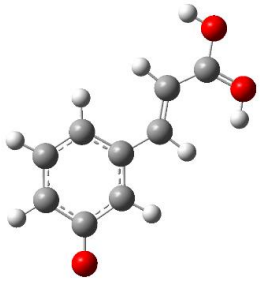
m-coumaric _ O(12)_Aβ
550.5708 kcal/mol



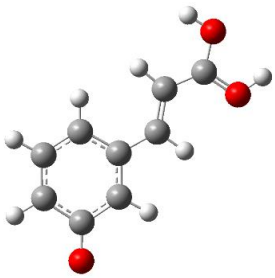
m-coumaric_ A1'
266.7794 kcal/mol



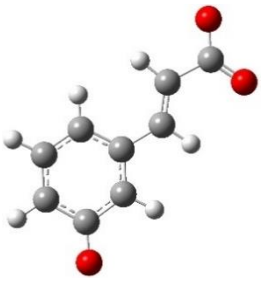
m-coumaric_ A2'
271.3305 kcal/mol



m-coumaric_ A3'
268.8662 kcal/mol

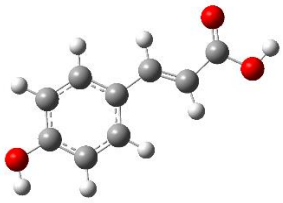


m-coumaric_ A5'
266.3645 kcal/mol

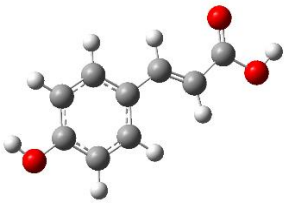


m-coumaric_ deprot_A_O11/O12
947.9905 kcal/mol

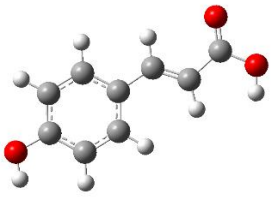
Figure 3.5 Calculated conformers of *m*-coumaric acid protonated at either C7, C8 or the O(10) oxygen. DFT/B3LYP/6-31+G(d,p) results in *vacuo*.
In the acronyms, the first letter (*m*-coumaric) informs that this is the meta substituted coumaric acid. "S" denotes a *syn* arrangement, and "A" denotes an *anty* arrangement. The numbers 1 to 8 represent the orientation of the OH groups. The terms 1+2, 3+7, 4+6 and 5+8 result from protonation is at C7 or C8 in which case the orientation of the O(10) oxygen is symmetrical, therefore the conformers of coumaric acid will differ by the orientation of the alkenoic OH and the phenol OH, and their conformers are compared with their similar conformer when protonation is at the O(10) oxygen with the H of the O(10) oxygen oriented to the left or the right. α and β denotes the orientation of OH for deprotonated structures. Prime (') denotes deprotonation at O(12).



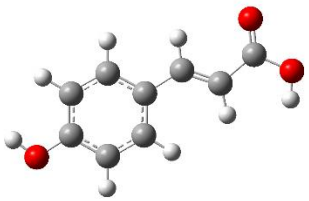
p-coumaric_S1+2
213.1845 kcal/mol



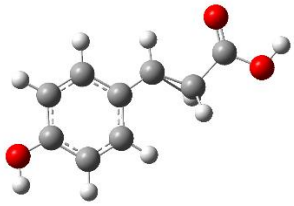
p-coumaric_S4 +6
213.0744 kcal/mol



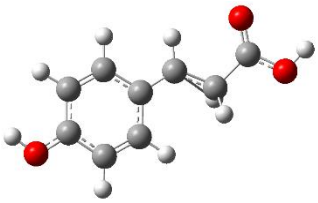
p-coumaric_S3 +7
219.5634 kcal/mol



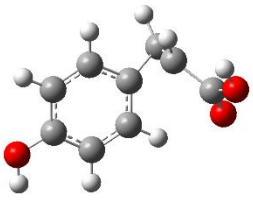
p-coumaric_S5+8
219.3584 kcal/mol



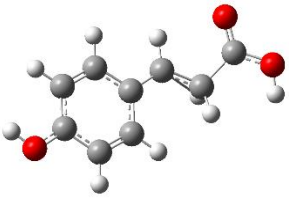
p-coumaric_C7_ S1+2
6.8169 kcal/mol



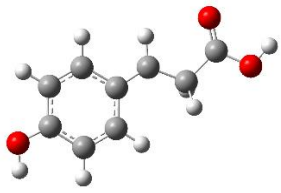
p-coumaric_ C7_ S4+6
51.2316 kcal/mol



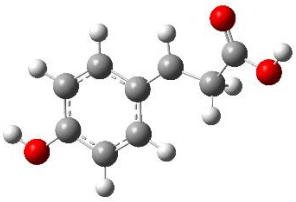
p-coumaric_C7_S3+7
27.0678 kcal/mol



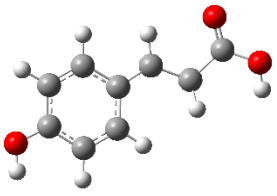
p-coumaric_C7_S5+8
60.7579 kcal/mol



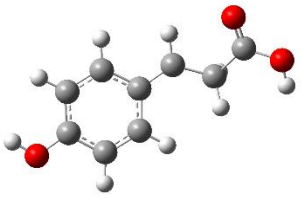
p-coumaric_C8_S1+2
6.8170 kcal/mol



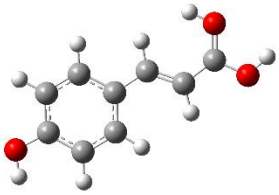
p-coumaric_C8_S4 +6
6.5563 kcal/mol



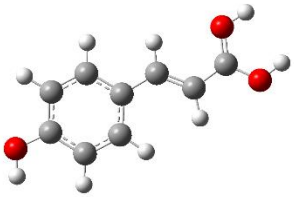
p-coumaric_C8_S3+7
15.5059 kcal/mol



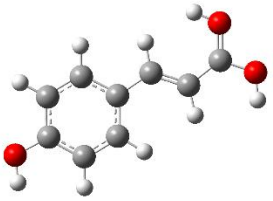
p-coumaric_C8_S5+8
15.1417 kcal/mol



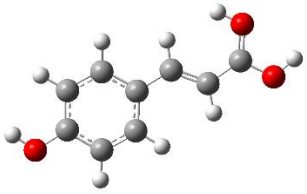
p-coumaric_S1
2.7059 kcal/mol



p-coumaric_S2
4.8749 kcal/mol



p-coumaric_S3
6.5195 kcal/mol



p-coumaric_S4
2.5955 kcal/mol

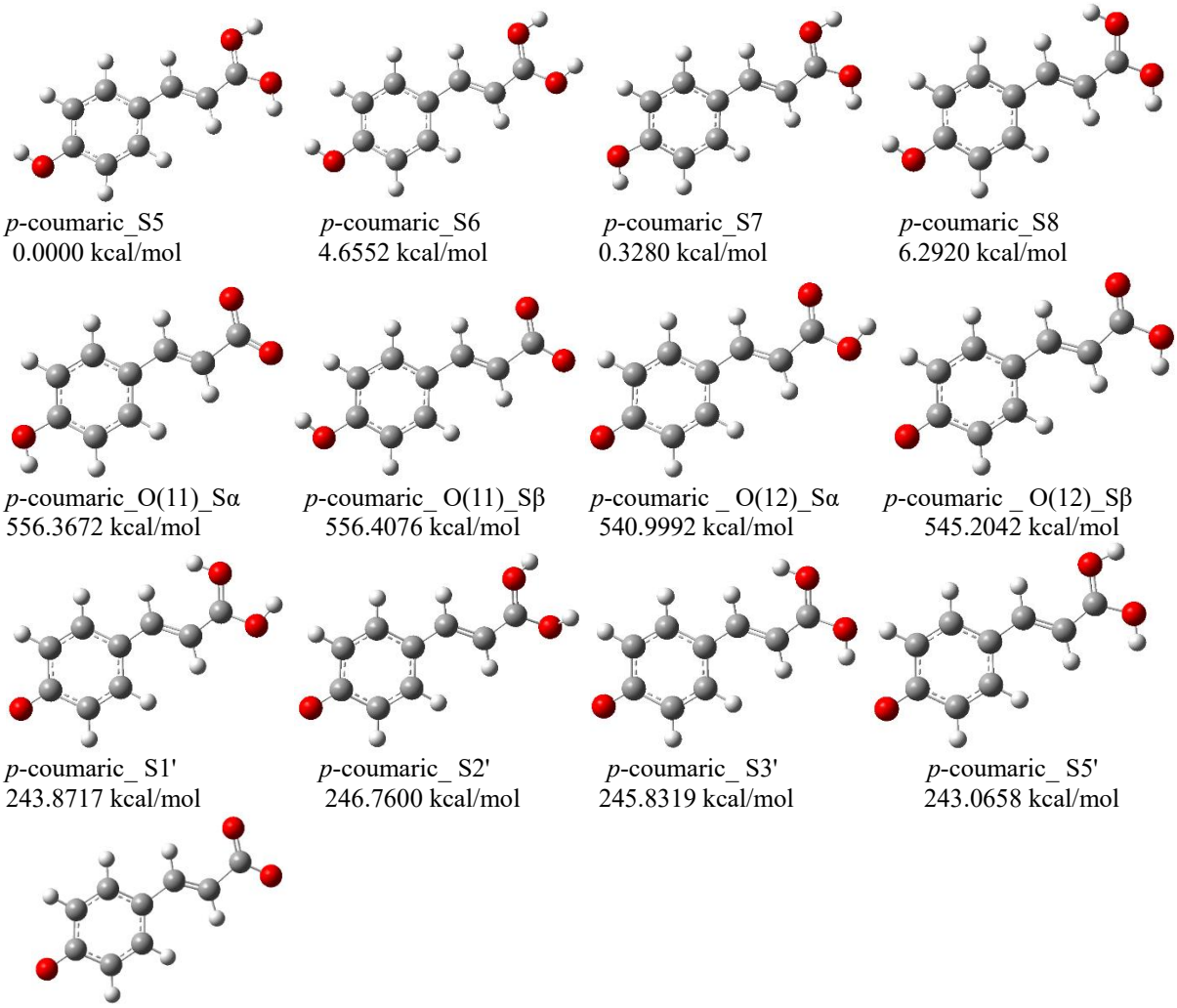
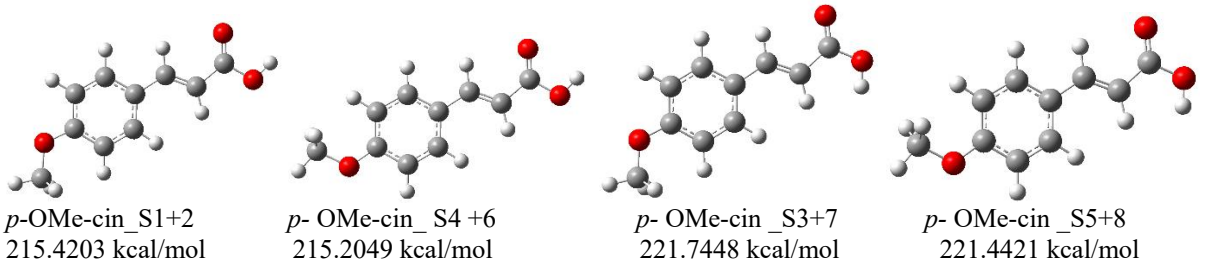


Figure 3.6. Calculated conformers of *p*-coumaric acid protonated at either C7, C8 or the O(10) oxygen. DFT/B3LYP/6-31+G(d,p) results in *vacuo*.
In the acronyms, the first letter (*p*-coumaric) informs that this is the para substituted coumaric acid. "S" denotes a *syn* arrangement. The numbers 1 to 8 represent the orientation of the OH groups. The terms 1+2, 3+7, 4+6 and 5+8 result from protonation is at C7 or C8 in which case the orientation of the O(10) oxygen is symmetrical, therefore the conformers of coumaric acid will differ by the orientation of the alkenoic OH and the phenol OH, and their conformers are compared with their similar conformer when protonation is at the O(10) oxygen with the H of the O(10) oxygen oriented to the left or the right. α and β denotes the orientation of OH for deprotonated structures. Prime (') denotes deprotonation at O(12).



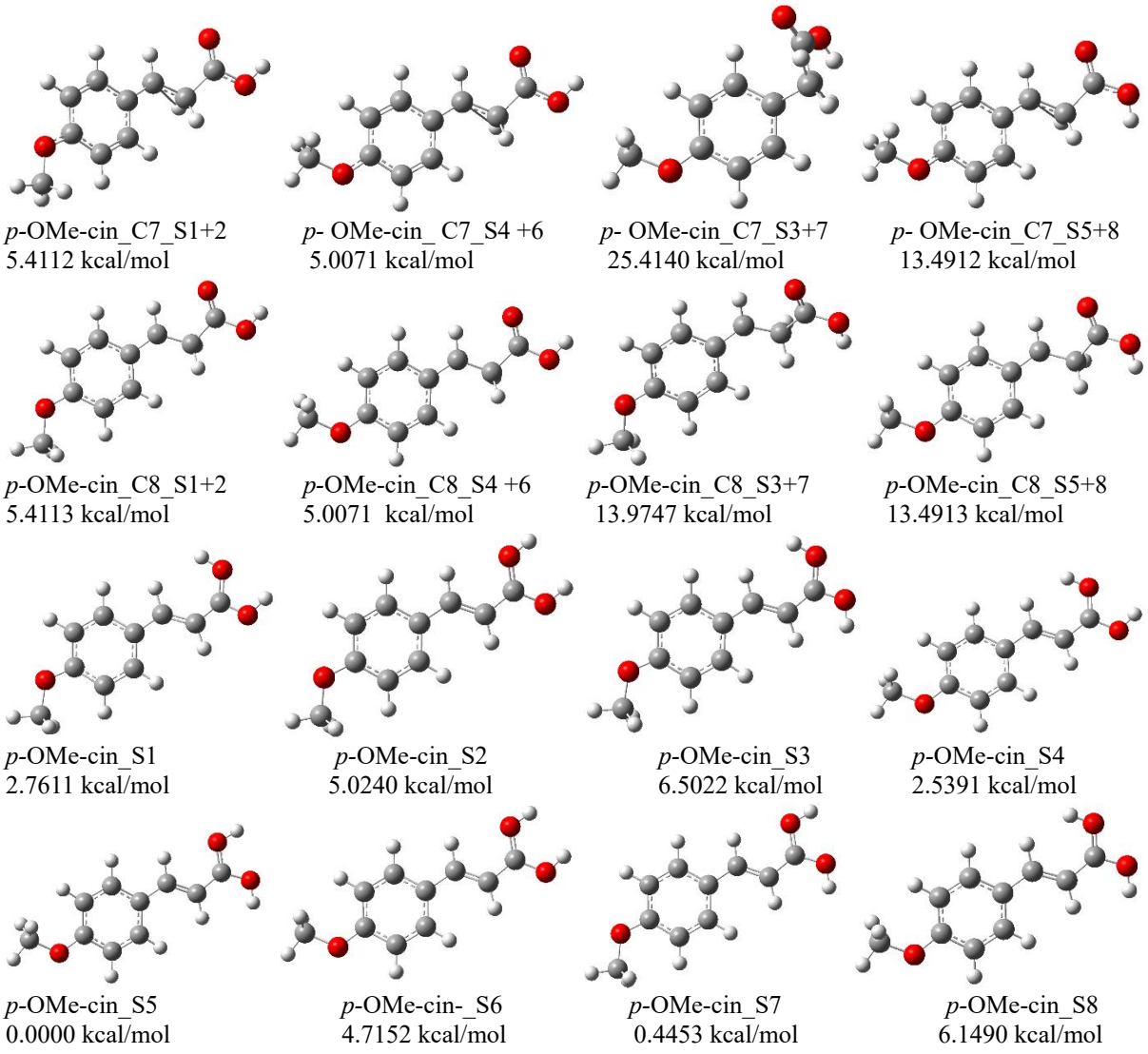
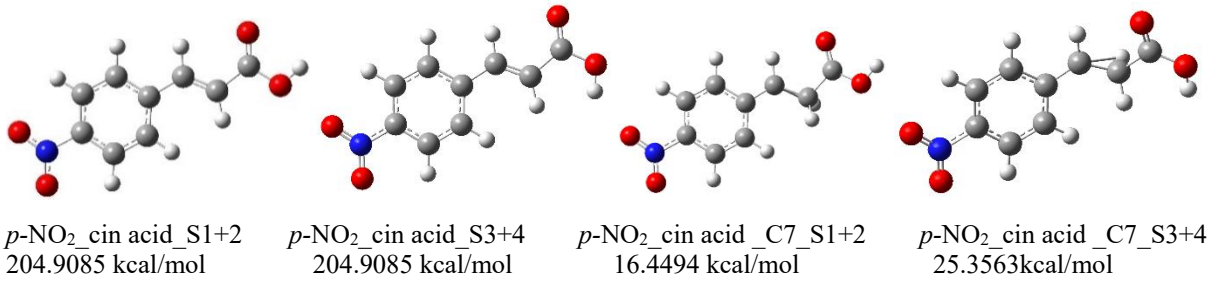
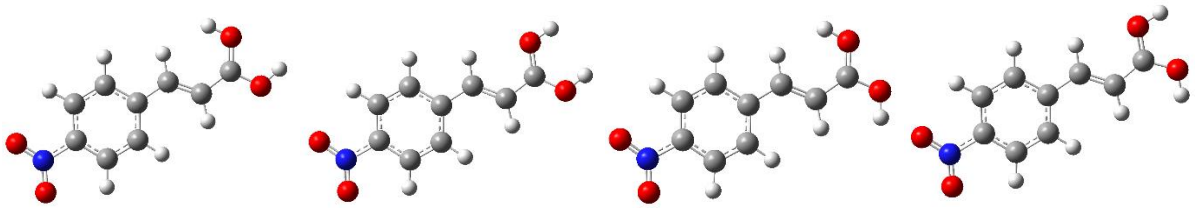


Figure 3.7. Calculated conformers of *p*-methoxycinnamic acid protonated at either C7, C8 or the O(10) oxygen. DFT/B3LYP/6-31+G(d,p) results in *vacuo*.
In the acronyms, the first letter (*p*-OMe-cin) informs that this is the *p*-methoxycinnamic acid. "S" denotes a *syn* arrangement. The numbers 1 to 8 represent the orientation of the OH groups. The terms 1+2, 3+7, 4+6 and 5+8 result from protonation is at C7 or C8 in which case the orientation of the O(10) oxygen is symmetrical, therefore the conformers of coumaric acid will differ by the orientation of the alkenoic OH and the phenol OH, and their conformers are compared with their similar conformer when protonation is at the O(10) oxygen with the H of the O(10) oxygen oriented to the left or the right.





p-NO_2_cin acid_S1 2.6620 kcal/mol | *p*-NO_2_cin acid_S2 4.2756 kcal/mol | *p*-NO_2_cin acid_S3 6.7240 kcal/mol | *p*-NO_2_cin acid_S4 0.0000 kcal/mol

Figure 3.8. Calculated conformers of *p*-nitrocinnamic acid protonated at either C7, C8 or the O(10) oxygen. DFT/B3LYP/6-31+G(d,p) results in *vacuo*.

In the acronyms, the first letter (*p*-NO_2_cin acid) informs that this is *p*-nitrocinnamic acid. "S" denotes a *syn* arrangement. The numbers 1 to 8 represent the orientation of the OH groups. The terms 1+2, 3+7, 4+6 and 5+8 result from protonation is at C7 or C8 in which case the orientation of the O(10) oxygen is symmetrical, therefore the conformers of coumaric acid will differ by the orientation of the alkenoic OH and the phenol OH, and their conformers are compared with their similar conformer when protonation is at the O(10) oxygen with the H of the O(10) oxygen oriented to the left or the right.

Table 3.1: Corrected relative energy of different conformers of o-coumaric acid with different protonation states. Because structures protonated at O(10) have more conformational preferences that are of interest and their conformers serve as reference for conformers of the other structures, the columns corresponding to other protonation states have the rows corresponding to such similar conformers. For example, conformer S1 and S2 are conformationally similar to S1+2 of structures protonated at C7, C8 and the neutral structures, and therefore row corresponding S1 and S2 are combined for structures protonated at C7, C8 and neutral structures. This applies to all tables.

Type of conformer (*o*)	**Corrected relative energy of conformers (kcal/mol)**							
	Protonated at O(10) (S1 to A8)	**Protonated at O(10) and deprotonated at O(12) (S1' to A5')**	**Protonated at C(7) (S1+2 to A5+8)**	**Protonated at C(8) (S1+2 to A5+8)**	**Neutral *o*-coumaric acid (S1+2 to A5+8)**	**Deprotonated at O(11) (Sα, Sβ, Aα, and Aβ)**	**Deprotonated at O(12) (Sα, Sβ, Aα, and Aβ)**	**Deprotonated at both O(11) and O(12) (S and A)**
S1	10.62952	243.41468	19.25656	15.77775	214.36700	551.72835	543.26509	960.26460
S2	12.46949	246.97187						
S4	3.01145		8.75894	8.75900	213.24722			
S6	5.20118							
S3	14.34202	243.25115	14.83266	25.32571	220.87847	558.73090	545.09690	
S7	8.01830							
S5	0.00000	240.71346	26.87200	17.56293	218.92449			
S8	6.16214							
A1	10.70464	242.32288	13.61286	13.61292	214.44537	552.24943	544.33994	958.82530
A2	11.63052	247.72744						
A4	2.78348		9.57036	9.57457	213.64612			
A6	5.61904							
A3	14.35451	244.45903	22.16805	22.16805	220.75008	557.99484	548.70573	
A7	6.97570							
A5	0.98487	243.94159	18.23015	18.23015	220.01772			
A8	6.49980							

Table 3.2: Corrected relative energy of different conformers of *m*-coumaric acid with different protonation states.

Type of conformer (*m*)	Corrected relative energy of conformers (kcal/mol)							
	Protonated at O(10) (S1 to A8)	Protonated at O(10) and deprotonated at O(12) (S1' to A5')	Protonated at C(7) (S1+2 to A5+8)	Protonated at C(8) (S1+2 to A5+8)	Neutral *m*-coumaric acid (S1+2 to A5+8)	Deprotonated at O(11) (Sα, Sβ, Aα, and Aβ)	Deprotonated at O(12) (Sα, Sβ, Aα, and Aβ)	Deprotonated at both O(11) and O(12) (S and A)
S1	3.3673	265.8136	12.0996	12.0996	208.6691	549.0997	546.2696	949.0888
S2	5.5055	270.1057						
S4	2.5823		12.0236	12.0236	208.7769			
S6	4.7285							
S3	7.3460	267.0484	58.2116	21.0134	215.2515	550.7850	549.9381	
S7	1.1186							
S5	0.0000	264.2891	20.6483	20.6483	215.0167			
S8	6.1916							
A1	3.8136	266.7794	24.9156	12.4718	208.7389	549.3462	546.2692	947.9905
A2	5.7368	271.3305						
A4	2.7815		12.5569	12.5569	208.8368			
A6	4.9922							
A3	7.5408	268.8662	33.8180	21.1502	215.0649	550.5838	550.5708	
A7	1.1186							
A5	0.4158	266.3645	33.2533	21.2947	215.2218			
A8	6.5508							

Table 3.3: Corrected relative energy of different conformers of *p*-cinnamic acid with different protonation states.

Type of conformer (*p*)	Corrected relative energy of conformers (kcal/mol)							
	Protonated at O(10) (S1 to A8)	Protonated at O(10) and deprotonated at O(12) (S1' to A5')	Protonated at C(7) (S1+2 to A5+8)	Protonated at C(8) (S1+2 to A5+8)	Neutral *p*-coumaric acid (S1+2 to A5+8)	Deprotonated at O(11) (Sα, Sβ, Aα, and Aβ)	Deprotonated at O(12) (Sα, Sβ, Aα, and Aβ)	Deprotonated at both O(11) and O(12) (S and A)
S1	2.7059	243.8717	6.8169	6.8170	213.1845	556.3665	540.9992	950.4977
S2	4.8749	246.7600						
S4	2.5955		6.5564	6.5563	213.0744			
S6	4.6552							
S3	6.5195	245.8319	27.0678	15.5059	219.5634	556.4076	545.2042	
S7	0.3280							
S5	0.0000	243.0658	15.1417	15.1417	219.3584			
S8	6.2920							

Table 3.4: Corrected relative energy of different conformers of *p*-OMe,cinnamic acid with different protonation states.

Type of conformer (*p*)	Corrected relative energy of conformers (kcal/mol)			
	Protonated at O(10) (S1 to A8)	Deprotonated at C(7) (S1+2 to A5+8)	Deprotonated at C(8) (S1+2 to A5+8)	Neutral *p*-OMe,cinnamic acid (S1+2 to A5+8)
S1	2.7611	5.4112	5.4113	215.4203
S2	5.0240			
S4	2.5391	5.0071	5.0071	215.2049
S6	4.7152			
S3	6.5022	25.4140	13.9747	221.7448
S7	0.4453			
S5	0.0000	13.4912	13.4913	221.4421
S8	6.1490			

Table 3.5: Corrected relative energy of different conformers of *p*-NO_2,cinnamic acid with different protonation states.

Type of conformer (o)	Corrected relative energy of conformers (kcal/mol)			
	Protonated at O(10) (S1 to A8)	Deprotonated at C(7) (S1+2 to A5+8)	Deprotonated at C(8) (S1+2 to A5+8)	Neutral *p*-NO_2,cinnamic acid (S1+2 to A5+8)
S1	2.6620	16.4494	25.3570	204.9085
S2	4.2756			
S3	6.7240	25.3563	25.3568	204.9085
S4	0.0000			

Table 3.6: C(7)=C(8) bond lengths *o*-coumaric acid

Type of conformer (*o*)	C(7)=C(8) bond length (Å)							
	Protonated at O(10) (S1 to A8)	Protonated at O(10) and deprotonated at O(12) (S1' to A5')	Protonated at C(7) (S1+2 to A5+8)	Protonated at C(8) (S1+2 to A5+8)	Neutral *p*-coumaric acid (S1+2 to A5+8)	Deprotonated at O(11) (Sα, Sβ, Aα, and Aβ)	Deprotonated at O(12) (Sα, Sβ, Aα, and Aβ)	Deprotonated at both O(11) and O(12) (S and A)
S1	1.38590	1.41463	1.54864	1.48985	1.35334	1.34736	1.37088	1.35355
S2	1.39082	1.42432						
S4	1.37990		1.47844	1.47828	1.35057			
S6	1.38469							
S3	1.38521	1.41281	1.50462	1.49251	1.35322	1.34831	1.36828	
S7	1.38991							
S5	1.38433	1.41689	1.55230	1.48239	1.34998			
S8	1.37930							
A1	1.38253	1.41613	1.48760	1.48747	1.349181	1.34429	1.37343	1.35344
A2	1.38840	1.42650						
A4	1.37949		1.48303	1.48451	1.34922			
A6	1.38487							
A3	1.38331	1.41621	1.48803	1.48236	1.34897	1.34652	1.37257	
A7	1.38887							
A5	1.38511	1.42180	1.48253	1.48240	1.34899			
A8	1.37957							

Table 3.7: C(7)=C(8) bond lengths of *m*-coumaric acid

Type of conformer (*m*)	C(7)=C(8) bond length (Å)							
	Protonated at O(10) (S1 to A8)	Protonated at O(10) and deprotonated at O(12) (S1' to A5')	Protonated at C(7) (S1+2 to A5+8)	Protonated at C(8) (S1+2 to A5+8)	Neutral *p*-coumaric acid (S1+2 to A5+8)	Deprotonated at O(11) (Sα, Sβ, Aα, and Aβ)	Deprotonated at O(12) (Sα, Sβ, Aα, and Aβ)	Deprotonated at both O(11) and O(12) (S and A)
S1	1.37707	1.39934	1.47773	1.47764	1.34784	1.34521	1.35984	1.34728
S2	1.38152	1.40845						
S4	1.37566		1.47651	1.47641	1.34738			
S6	1.38009							
S3	1.37724	1.39882	1.42942	1.47789	1.34769	1.34536	1.35856	
S7	1.38215							
S5	1.38058	1.40353	1.47601	1.47599	1.34714			
S8	1.37568							
A1	1.37669	1.39976	1.44037	1.47783	1.34752	1.34513	1.36031	1.34796
A2	1.38107	1.40916						
A4	1.37553		1.47650	1.47637	1.34741			
A6	1.38000							
A3	1.37677	1.39957	1.43582	1.47755	1.34731	1.34549	1.35919	
A7	1.38215							
A5	1.38054	1.40477	1.43958	1.47598	1.34716			
A8	1.37551							

Table 3.8: C(7)=C(8) bond lengths of *p*-coumaric acid

Type of conformer (*p*)	C(7)=C(8) bond length (Å)							
	Protonated at O(10) (S1 to A8)	Protonated at O(10) and deprotonated at O(12) (S1' to A5')	Protonated at C(7) (S1+2 to A5+8)	Protonated at C(8) (S1+2 to A5+8)	Neutral *p*-coumaric acid (S1+2 to A5+8)	Deprotonated at O(11) (Sα, Sβ, Aα, and Aβ)	Deprotonated at O(12) (Sα, Sβ, Aα, and Aβ)	Deprotonated at both O(11) and O(12) (S and A)
S1	1.38369	1.42361	1.48367	1.48360	1.34980	1.34483	1.37869	1.35283
S2	1.37634	1.43069						
S4	1.38368		1.48383	1.48382	1.34977			
S6	1.38877							
S3	1.38396	1.42410	1.45855	1.48429	1.34965	1.34477	1.37792	
S7	1.38875							
S5	1.38866	1.42710	1.48425	1.48416	1.34963			
S8	1.38391							

Table 3.9: C(7)=C(8) bond lengths *p*-OMe cinnamic acid

Type of conformer (p)	Energy of conformer at protonated position (Å)			
	Protonated at O(10) (S1 to A8)	Deprotonated at C(7) (S1+2 to A5+8)	Deprotonated at C(8) (S1+2 to A5+8)	Neutral *p*-OMe,cinnamic acid (S1+2 to A5+8)
S1	1.38562	1.48497	1.48501	1.32862
S2	1.39114			
S4	1.38567	1.48552	1.48551	1.35021
S6	1.39113			
S3	1.38587	1.46095	1.48588	1.32883
S7	1.39067			
S5	1.39069	1.48571	1.48572	1.35005
S8	1.38597			

Table 3.10: C(7)=C(8) bond lengths of *p*-NO_2 cinnamic acid

Type of conformer (o)	Corrected relative energy of conformer at protonated position (Å)			
	Protonated at O(10) (S1 to A8)	Deprotonated at C(7) (S1+2 to A5+8)	Deprotonated at C(8) (S1+2 to A5+8)	Neutral *p*-NO_2,cinnamic acid (S1+2 to A5+8)
S1	1.3711	1.4724	1.47242	1.3463
S2	1.3753			
S3	1.3712	1.4716	1.4714	1.3462
S4	1.3759			

3.2.3.2 Relative energy considering a solvated environment

To account for solvent effects on the descriptors considered in this research, the calculations in *vacuo* were repeated twice, the first considering a solvated water medium and the second with a chloroform media using the PCM model as implemented in Gaussian 09. The water simulated environment was chosen because it is the most biologically relevant environment. Chloroform simulated environment was chosen because chloroform is a non-polar solvent, in contrast to polar water, and therefore a chloroform simulated environment can account for a non-polar medium. The input geometries for both water and chloroform environment were structurally similar to those of the calculations *in vacou*, and the same level of theory and basis set were utilised [B3LYP/6-31+G(d,p)].

To avoid bulkiness of the chapter the table and figures with data corresponding to this section are given in Appendix A (Supplementary material in the disc). In summary the trends that are observed *in vacou* were also repeated in a water and a chloroform solvated environment. The general trends shortly presented are those observed across the entire set of calculations for comparisons across different derivatives of cinnamic acid in vacuo and in the two solvent media.

The energy of structures follows exactly the same pattern as stated before:

Structures protonated at O(10) < cinnamic acid protonated at either C(7) or C(8) < cinnamic acid that is neither protonated nor deprotonated < cinnamic acid that is protonated at O(10) and deprotonated at O(12) < cinnamic acid that is deprotonated at O(12) < cinnamic acid that is deprotonated at O(11) < cinnamic acid that is deprotonated at both O(11) and O(12)

The bond lengths also follow the same general pattern when compared across similar conformers of the five cinnamic acid derivatives, as the one discussed for *o*-coumaric acid. The bond length increase across structures also follow the general pattern:

The neutral structure, cinnamic acid deprotonated at O(11) and those that are deprotonated at both O(11) and O(12) – generally around 1.34 Å < cinnamic acid derivatives deprotonated at O(12) generally with the bond length of about 1.35 Å < cinnamic acid derivatives protonated at O(10) bond length generally around 1.37 Å < cinnamic acid derivatives protonated at O(10)

and deprotonated at O(12) bond length generally around 1.38 Å < acid protonated at either C(7) or C(8) bond length generally around 1.47 Å

However, although the patterns are the same there is noticeable variation in the exact bond length values across the different series.

The trends for rearrangement during geometry optimization of structures protonated at C(7) is still followed. Even in polar solvent media (such as with water) the charges are not mediated to the point of preventing this rearrangement. The resultant structures following rearrangement matched those in unsolvated calculations.

The trends observed shows that protonating on either carbon of the C(7)=C(8) double bond results in a structure stabilized by more than 200 kcal/mol energy than the neutral structure, all media and all cinnamic acid derivatives considered. Again, either C(7) or C(8) protonation results in structures with an elongated C(7)=C(8) bond (1.48 Å across all systems).

Protonating cinnamic acid derivatives at O(10) also results in a system that is stabilised by 200 kcal/mol relative to the neutral structure, albeit with the lengthening of the C(7)=C(8) bond length reduced slightly to 1.37 Å. This suggests that protonating a system at O(10) could be one of the better options to achieve the E→Z isomerization.

Even for the general series with solvent media, protonating at O(10) and deprotonating at a O(12) results in a system that is generally about 20 kcal/mol higher than the neutral structure, but the C(7)=C(8) double bonds of these system are still elongated (about 1.4 Å form about 1.3 Å). There are no structural rearrangements observed with these systems. The data analysis suggests that this approach could also be favourable for achieving a delocalized C(7)=C(8) double bond that is rotatable to achieve the desired Z isomer of coumaric acid is the protonation at O(10) and deprotonation at O(12).

Based on this evidence it is possible to choose to either to protonate at O(12) or to simultaneously protonate at O(10) and deprotonate at O(12) in order to achieve E→Z isomerization of the appropriate acid. However, further information on which to base this choice is available from the electronic structure calculations in the form of other chemical descriptors. It is prudent to also use this information in making this decision.

3.3 Frontier orbitals

Some of the most revealing descriptors when exploring the reactivity of molecules, are the energies of the frontier Molecular Orbitals (MO's) [i.e., the Highest Occupied Molecular Orbital (HOMO) and the Lowest Unoccupied Molecular Orbital (LUMO)]. Together with this information are the energy difference of the frontier orbitals (HOMO-LUMO energy gaps), and the localization of the frontier molecular orbitals on a chemical structure.

The relevance of the frontier orbitals is largely because they are responsible for many of the chemical and spectroscopic properties of the molecule. Although a critical step in the synthesis of coumarins is the E → Z isomerisation of the cinnamic acid precursor, and we understand from section 3.2 the effects of protonating, deprotonating or both protonating and deprotonating this system in the weakening of the crucial bond, the other two critical steps in this synthesis are

i) Cyclisation (lactonisation) to the hemi-acetal intermediate; and
ii) Dehydration to afford the coumarin derivative.

The act of protonation, together with these two steps are all chemical processes involving the transfer of electrons. The study of the HOMO, LUMO, HOMO-LUMO energy gaps and visualization of electron densities on cinnamic acid derivatives was done to acquire information on the theoretical reactivity patterns of their protonated, deprotonated or both protonated and deprotonated systems (i.e., to understand their proton donating and proton accepting tendencies).

An ideal mechanism of weakening the C(7)=C(8) double bond would be one that not only weakens the double bond, but also result on a system with reasonable HOMO-LUMO energy gaps and the MOs localized in a manner that's suggest favour of the cyclization to the hemi-acetal intermediate with dehydration to afford coumarins.

This section focuses on the analysis of the various trends of the energies of the HOMO, the LUMO, the HOMO-LUMO energy gap and the visualised MO figures. Because these results

were obtained from the optimised geometries discussed in section 3.2, then the discussion of these results will follow the similar manner. These results are discussed in a comparative manner with sections 3.3.1 focusing of the comparisons of the results *in vacuo* and sections 3.3.2 focusing on comparisons of results in solvated environments. To avoid bulkiness of the chapter the figures of MO and the tables given in this section are of the study of *o*-coumaric acid *in vacuo*, the structure of interest in this study. The rest of the figures and tables with the results discussed in this section are given in Appendix B (Supplementary material in the disc).

3.3.1 Results *in vacou*

The results discussed in this section are of structures with different protonation states, in the discussion in section 3.2. Although before their relative energies were mass balanced due to the energy each proton contributes to the total energy of the system, energy corrections were not necessary with the comparisons of the energies of the frontier orbitals and their energy gaps, because the energy of each frontier orbital is dependant of on the energy of the electrons, and all the different protonation states of cinnamic acid derivatives had the same number of electrons (i.e., the solutions of the Schrödinger equation are wavefunctions, and wave functions refers to MO (224)). Therefore, the energies of the HOMO, the LUMO and the HOMO-LUMO energy gaps were directly compared across different protonation states of cinnamic acid derivatives considered in this study. The results (Table 3.11-3.13) show that the HOMO energies of the deprotonated systems are generally higher than the HOMO energies of the neutral system by more than 100 kcal/mol (over 150 kcal/mol for systems deprotonated at two positions). The HOMO energies of the protonated systems are lower than the energies of a neutral system (as low as 100 kcal/mol less in some cases). Compared to the energy changes resulting from either protonating or deprotonating cinnamic acid derivatives, simultaneous protonation and deprotonation of cinnamic acid derivatives does not cause a drastic change in the HOMO energy. These trends are observed with all the cinnamic acid derivatives considered (*o*-, *m*-, *p*- coumaric acid, OMe-cinnamic acid and NO_2-cinnamic acid). Minor deviations from the from the relative energy trends were observed with systems protonated at C(7). Such deviations could be attributed to the difference of the optimised geometries which resulted in the change of spatial localization of electrons (the structures protonated at C(7) had proton migration or geometry rearrangements during optimisation.). By definition the energies of the

LUMO are higher than the energies of the HOMO, however, the trends of the changes in LUMO energy are similar to the trends of the changes of the HOMO energies, and we may rewrite the above but with LUMO as focus:

The LUMO energies of the deprotonated systems are generally higher than the LUMO energies of the neutral system by more than a 100 kcal/mol (over 150 kcal/mol for systems deprotonated at two positions). The LUMO energies of the protonated systems are lower than the energies of a neutral system (as low as 100 kcal/mol less in some cases). Compared to the energy changes resulting from either protonating or deprotonating a cinnamic acid derivative, simultaneous protonation and deprotonation of cinnamic acid derivatives does not cause a drastic change in the LUMO energy.

The HOMO-LUMO energy gaps relate to several chemical properties of a molecule, one of which, molecular hardness is of interest in our study. Large HOMO-LUMO energy gaps informs that the molecule is hard (less chemically reactive), while smaller HOMO-LUMO energy gaps informs that the molecule is soft (more reactive). The protonated *o*-coumaric acid will be intermediate in the formation of coumarins, therefore the ideal intermediate would be one with the HOMO-LUMO energy gap lower than that of the neutral structure (a more reactive intermediate). For the HOMO-LUMO energy gap to be higher than that of the neutral structure, this would mean that protonated state is chemically harder than the neutral system (thus the intermediate would be less reactive than the starting structure precursors). The results show that the HOMO-LUMO energy gaps of the deprotonated systems are lower than those of the neutral structures (more than 20 kcal/mol lower). This suggests the deprotonated systems are more reactive that the neutral structures (this was expected because of the lone pairs of electrons on the deprotonated oxygens). The structures that were deprotonated in two positions have the lowest energy gaps of the deprotonated systems (the system that is deprotonated twice has two oxygens with lone pairs), and thus a reactivity increased by the forced removal of a second proton. The HOMO-LUMO energy gaps of the protonated structures were also lower than the HOMO-LUMO energy gaps of the neutral structures, but with the following general trend:

Protonated and deprotonated structures had the smallest energy gaps (generally more than 30 kcal/mol smaller than the neutral system). The structure protonated at O(10) and those protonated at C(8) had HOMO-LUMO energy gaps relatively within the same range (generally

about 25 kcal/mol smaller than those of the neutral structures). Finally, structures protonated at C(7) had HOMO-LUMO energy gaps generally higher than the other protonated systems (generally about 10 kcal/mol smaller energies than the neutral structures)

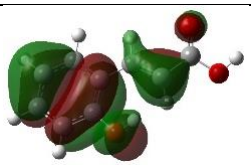
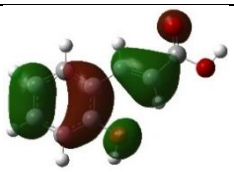
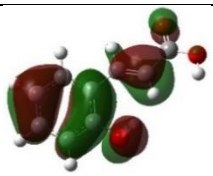
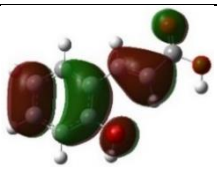
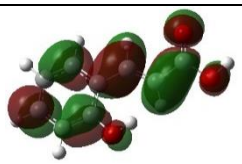
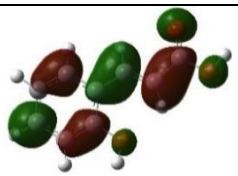
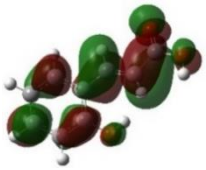
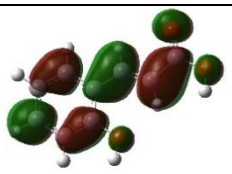
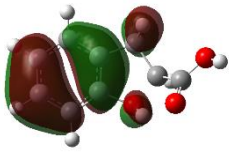
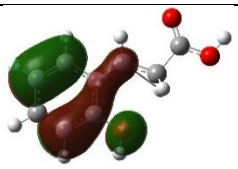
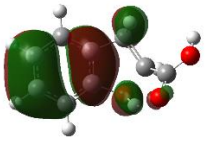
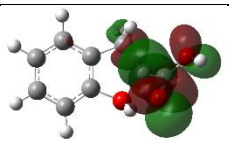
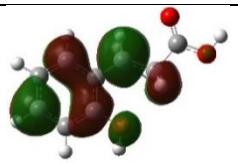
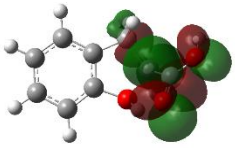
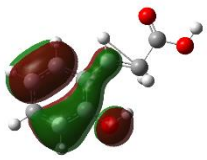
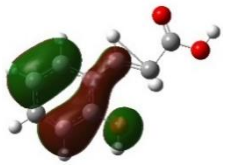
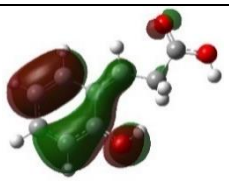
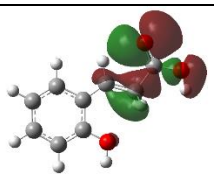
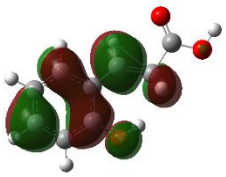
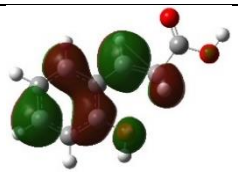
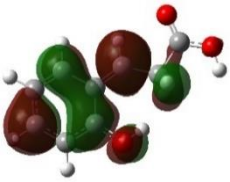
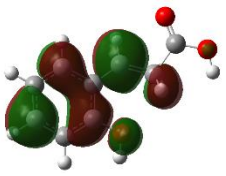
The reason why structures protonated at C(7) have the HOMO-LUMO energy gaps that are not within the same range as the structure protonated at C(8) and those protonated at O(10) could be because the optimised structures had rearranged (either by hydride shift or by cyclization to bicyclic systems).

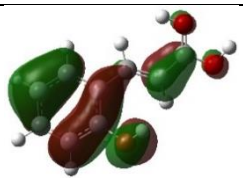
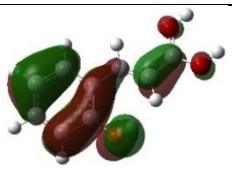
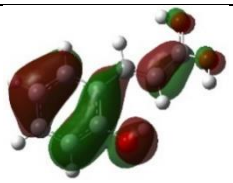
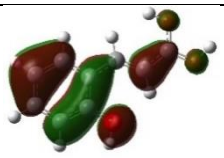
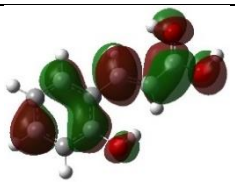
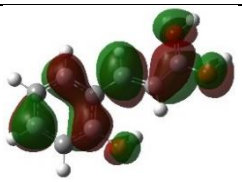
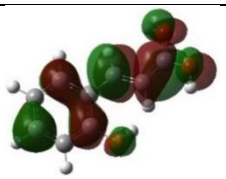
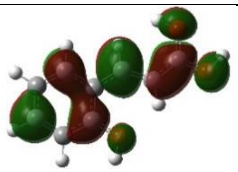
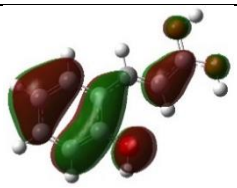
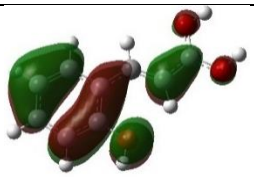
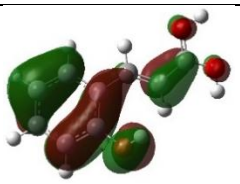
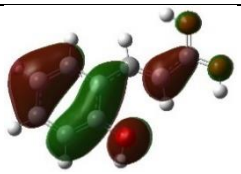
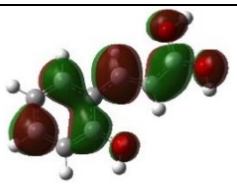
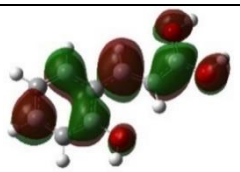
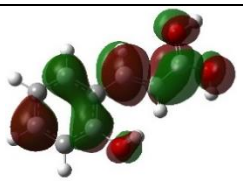
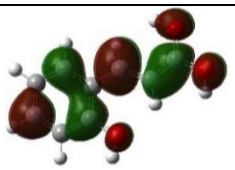
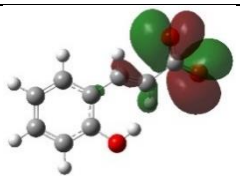
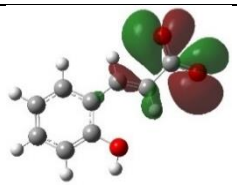
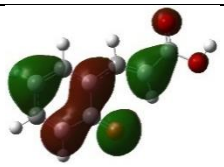
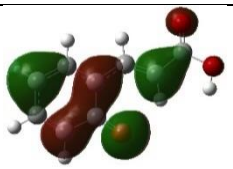
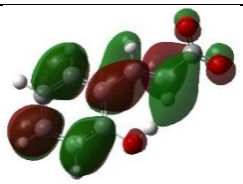
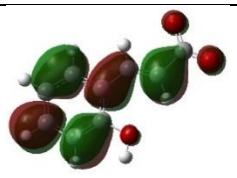
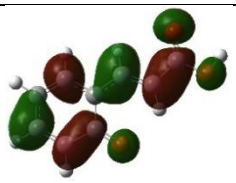
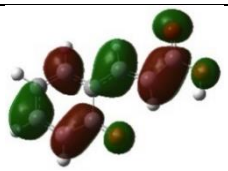
The knowledge of molecular hardness of the chemical structures alone may not be enough, in chemistry also knowing which part of that system is more reactive is essential. Therefore this study also analysed the MO figures of the individual conformers of the different protonation states of cinnamic acid derivatives. The localization of the HOMO informs that that particular part of the structure is the more electron donating. The localization of the LUMO informs that that particular part of the structure is more electron accepting (225). The results shows that generally the HOMO is more localized on the heavy atoms of the entire system, however there are exceptions to this rule (as seen in Figure 3.9),

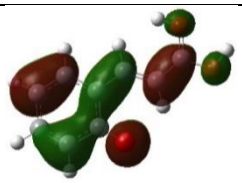
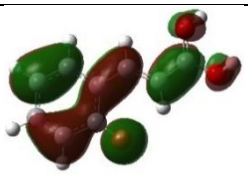
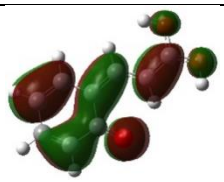
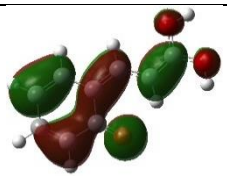
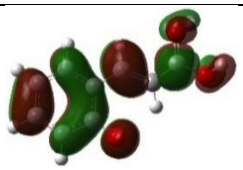
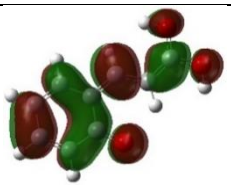
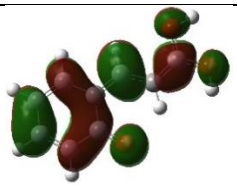
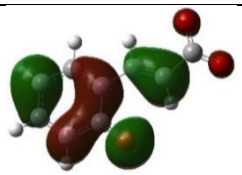
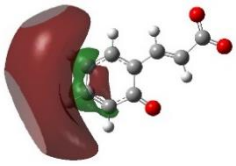
The HOMO of the C(7) protonated systems is more localised on the benzene ring and not much on the alkenoic acid. The HOMO of structures deprotonated at O(11), however, are more localized on the alkenoic acid chain than on the benzene ring (this is expected because the lone pairs of electrons on the deprotonated oxygen makes this part of the molecule more electron donating).

The localization of the LUMO in most structures is generally on the heavy atoms (as for the HOMO). On bicyclic systems, the LUMO is more localised on the heterogeneous part of the system than on the benzene moiety.

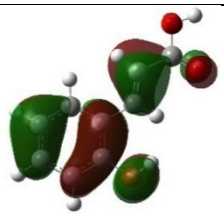
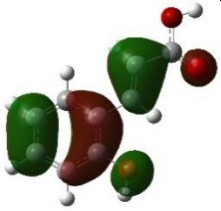
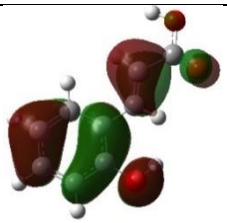
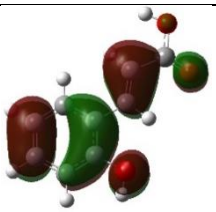
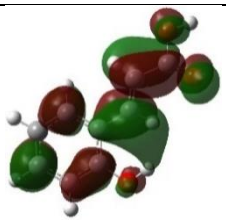
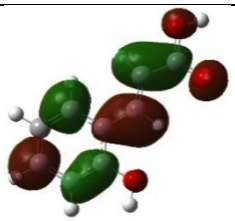
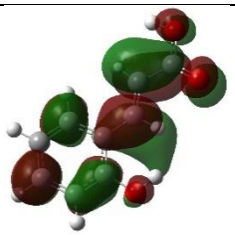
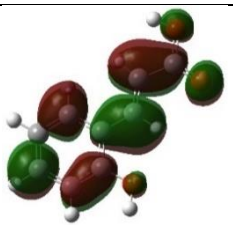
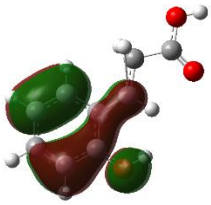
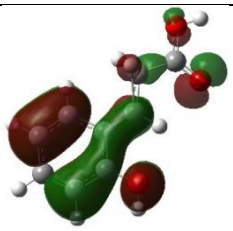
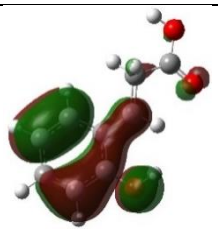
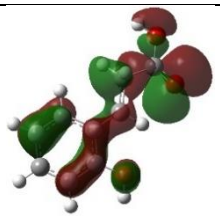
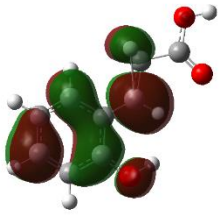
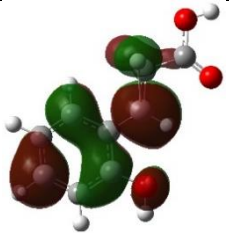
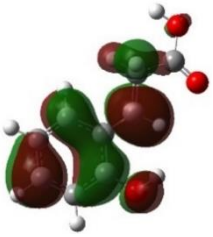
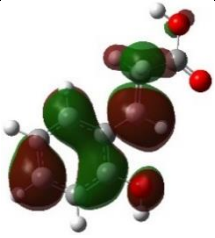
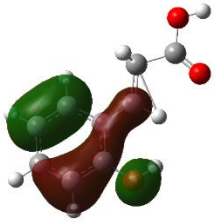
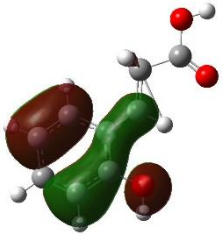
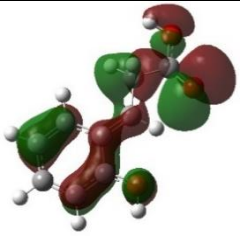
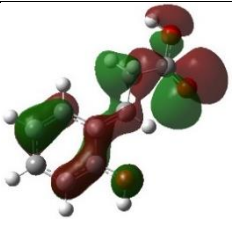
S-conformers

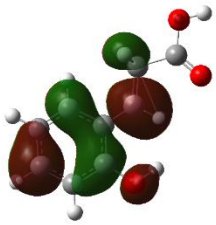
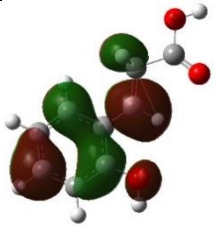
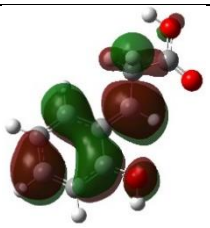
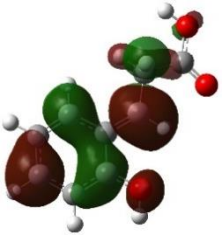
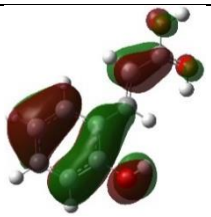
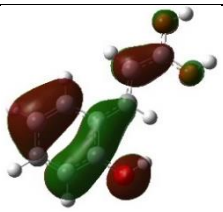
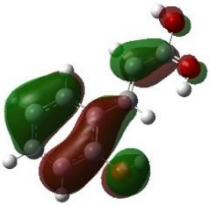
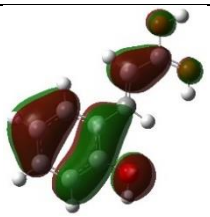
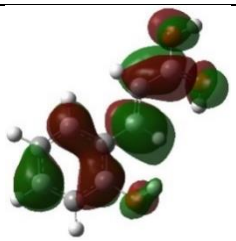
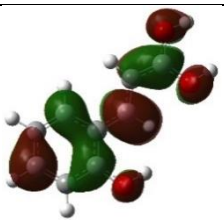
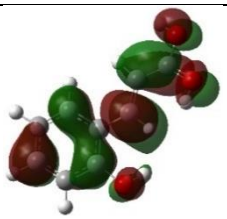
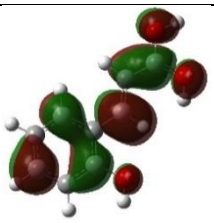
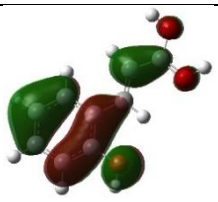
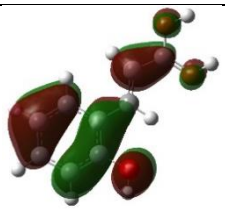
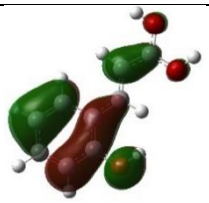
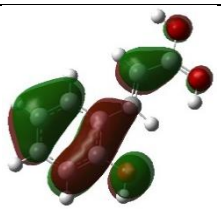
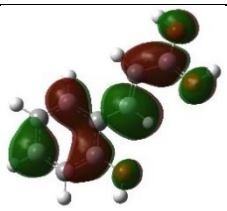
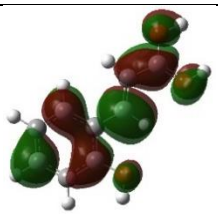
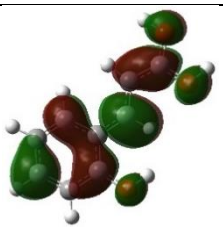
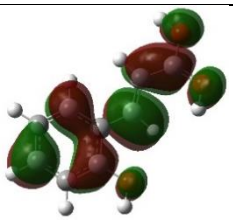
		Type of conformer showing MO			
***o*-coumaric acid**	**HOMO**	 *o*-coumaric_S1+2	 *o*-coumaric_S4 +6	 *o*-coumaric_S3 +7	 *o*-coumaric_5+8
	LUMO	 *o*-coumaric_S1+2	 *o*-coumaric_S4 +6	 *o*-coumaric_S3 +7	 *o*-coumaric_S5+8
o-coumaric acid protonated at C(7)	**HOMO**	 o-coumaric_C7_S1+2	 o-coumaric_C7_S4 +6	 o-coumaric_C7_S3+7	 o-coumaric_C7_S5+8
	LUMO	 o-coumaric_C7_S1+2	 o-coumaric_C7_S4 +6	 o-coumaric_C7_S3+7	 o-coumaric_C7_S5+8
***o*-coumaric acid protonated at C(8)**	**HOMO**	 *o*-coumaric_C8_S1+2	 *o*-coumaric_ C8_S4 +6	 *o*-coumaric_C8_S3+7	 *o*-coumaric_C8_S5+8
	LUMO	 *o*-coumaric_C8_S1+2	 *o*-coumaric_C8_S4 +6	 *o*-coumaric_C8_S3+7	 o-coumaric_C8_S5+8

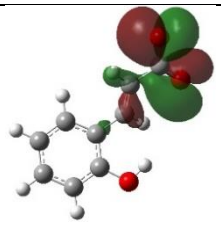
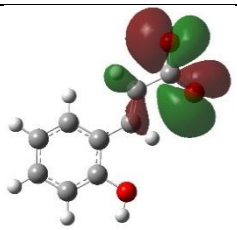
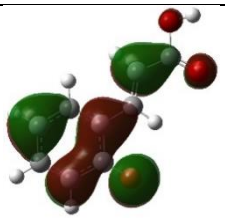
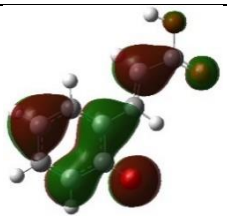
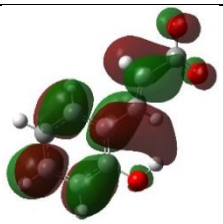
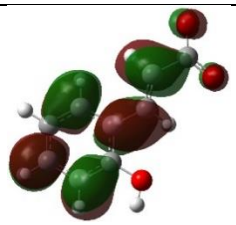
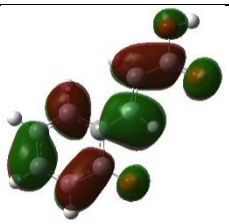
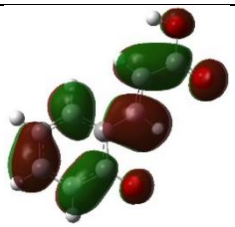
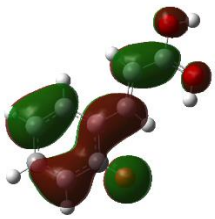
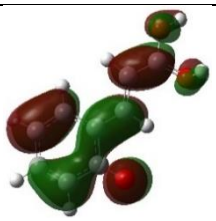
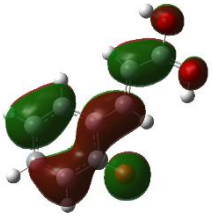
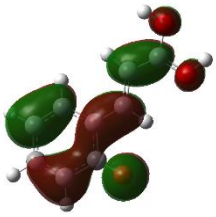
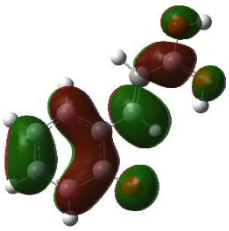
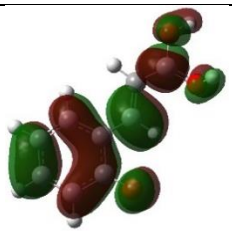
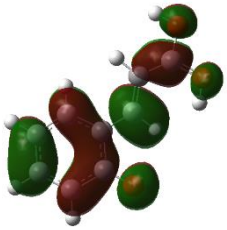
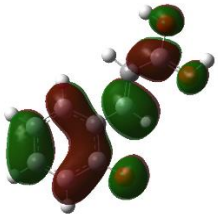
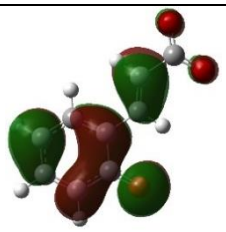
o-coumaric acid protonated at O(10)	**HOMO**	 o-coumaric_S1	 o-coumaric _S2	 o-coumaric_S3	 o-coumaric_S4
	LUMO	 *o*-coumaric_S1	 *o*-coumaric_S2	 *o*-coumaric_S3	 *o*-coumaric_S4
	HOMO	 *o*-coumaric_S5	 *o*-coumaric_S6	 *o*-coumaric_S7	 *o*-coumaric_S8
	LUMO	 *o*-coumaric_S5	 *o*-coumaric_S6	 *o*-coumaric_S7	 *o*-coumaric_S8
***o*-coumaric acid deprotonated at either O(11) or O(12)**	**HOMO**	 *o*-coumaric_ O(11)_Sα	 *o*-coumaric_ O(11)_Sβ	 *o*-coumaric _ O(12)_Sα	 o-coumaric _O(12)_Sβ
	LUMO	 o-coumaric_ O(11)_Sα	 o-coumaric_ O(11)_Sβ	 o-coumaric _ O(12)_Sα	 o-coumaric _O(12)_Sβ

***o*-coumaric acid protonated at O(10) and deprotonated at O(12)**	**HOMO**	 *o*-coumaric_ S1'	 *o*-coumaric_ S2'	 *o*-coumaric_ S3'	 *o*-coumaric_ S5'
	LUMO	 *o*-coumaric_ S1'	 *o*-coumaric_ S2'	 *o*-coumaric_ S3'	 *o*-coumaric_ S5'
***o*-coumaric acid deprotonated at O(11) and O(12)**	**HOMO**	 *o*-coumaric_ deprot_S_O11/O12			
	LUMO	 *o*-coumaric_ deprot_S_O11/O12			

A-conformers

		Type of conformer showing MO			
***o*-coumaric acid**	**HOMO**	 *o*-coumaric_A1+2	 *o*-coumaric_A4 +6	 *o*-coumaric_A3 +7	 *o*-coumaric_A5+8
	LUMO	 *o*-coumaric_A1+2	 *o*-coumaric_A4 +6	 *o*-coumaric_A3 +7	 *o*-coumaric_A5+8
***o*-coumaric acid protonated at C(7)**	**HOMO**	 *o*-coumaric_C7_A1+2	 *o*-coumaric_ C7_A4 +6	 *o*-coumaric_C7_A3+7	 o-coumaric_C7_A5+8
	LUMO	 *o*-coumaric_C7_A1+2	 *o*-coumaric_C7_A4 +6	 *o*-coumaric_C7_A3+7	 *o*-coumaric_C7_A5+8
***o*-coumaric acid protonated at C(8)**	**HOMO**	 *o*-coumaric acid_C8_A1+2	 *o*-coumaric acid_ C8_4 +6	 *o*-coumaric acid _C8_3+7	 *o*-coumaric acid_C8_5+8

	LUMO	 *o*-coumaric_C8_A1+2	 *o*-coumaric_C8_A4 +6	 *o*-coumaric_C8_A3+7	 o-coumaric_C8_A5+8
o-coumaric acid protonated at O(10)	**HOMO**	 o-coumaric_A1	 o-coumaric_A2	 o-coumaric_A3	 o-coumaric_A4
	LUMO	 *o*-coumaric_A1	 *o*-coumaric_A2	 *o*-coumaric_A3	 *o*-coumaric_A4
	HOMO	 *o*-coumaric_A5	 *o*-coumaric_A6	 *o*-coumaric_A7	 *o*-coumaric_A8
	LUMO	 *o*-coumaric_A5	 *o*-coumaric_A6	 *o*-coumaric_A7	 *o*-coumaric_A8

***o*-coumaric acid deprotonated at either O(11) or O(12)**	**HOMO**	 *o*-coumaric_ O(11)_Aα	 *o*-coumaric_ O(11)_Aβ	 *o*-coumaric _ O(12)_Aα	 *o*-coumaric _O(12)_Aβ
	LUMO	 *o*-coumaric_ O(11)_Aα	 *o*-coumaric_ O(11)_Aβ	 *o*-coumaric _ O(12)_Aα	 *o*-coumaric _O(12)_Aβ
***o*-coumaric acid protonated at O(10) and deprotonated at O(12)**	**HOMO**	 *o*-coumaric_A1'	 *o*-coumaric_A2'	 *o*-coumaric_A3'	 *o*-coumaric_A5'
	LUMO	 *o*-coumaric_A1'	 *o*-coumaric_A2'	 *o*-coumaric_A3'	 *o*-coumaric_A5'
***o*-coumaric acid deprotonated at O(11)**	**HOMO**	 *o*-coumaric_ deprot_A_O11/O12			

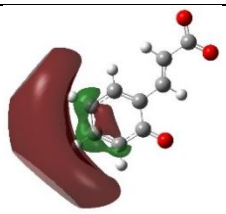
	LUMO	 *o*-coumaric_ deprot_A_O11/O12			

Figure 3.9. Calculated conformers of *o*-coumaric acid protonated at either C7, C8 or the O(10) oxygen. DFT/B3LYP/6-31+G(d,p) results in *vacuo*.
In the acronyms, the first letter (*o*-coumaric) informs that this is the ortho substituted coumaric acid. "S" denotes a *syn* arrangement, and "A" denotes an *anty* arrangement. The numbers 1 to 8 represent the orientation of the OH groups. The terms 1+2, 3+7, 4+6 and 5+8 result from protonation is at C7 or C8 in which case the orientation of the O(10) oxygen is symmetrical, therefore the conformers of coumaric acid will differ by the orientation of the alkenoic OH and the phenol OH, and their conformers are compared with their similar conformer when protonation is at the O(10) oxygen with the H of the O(10) oxygen oriented to the left or the right. α and β denotes the orientation of OH for deprotonated structures. Prime (') denotes deprotonation at O(12).

Table 3.11: HOMO Energy of different conformers of *o*-coumaric acid with different protonation states.

Type of conformer (*o*)	HOMO Energy of conformers (kcal/mol)							
	Protonated at O(10) (S1 to A8)	Protonated at O(10) and deprotonated at O(12) (S1' to A5')	Protonated at C(7) (S1+2 to A5+8)	Protonated at C(8) (S1+2 to A5+8)	Neutral *m*-coumaric acid (S1+2 to A5+8)	Deprotonated at O(11) (Sα, Sβ, Aα, and Aβ)	Deprotonated at O(12) (Sα, Sβ, Aα, and Aβ)	Deprotonated at both O(11) and O(12) (S and A)
S1	-242.2977	-129.2154	-257.5147	-267.5108	-152.2260	-41.8670	-30.3147	48.8699
S2	-238.3507	-129.9433						
S4	-240.0513		-267.4104	-267.4104	-149.6532			
S6	-236.0917							
S3	-245.8557	-131.2799	-260.3761	-269.5627	-155.4890	-32.8247	-33.2012	
S7	-241.7895							
S5	-239.1037	-127.5839	-257.0252	-266.5193	-152.9915			
S8	-243.1574							
A1	-243.0382	-130.2759	-267.3477	-267.3477	-151.2847	-41.4026	-30.6284	48.3177
A2	-237.8236	-130.7654						
A4	-240.3023		-266.7076	-266.9774	-149.0822			
A6	-235.8784							
A3	-245.1905	-132.0643	-269.3745	-266.5821	-153.9328	-34.1362	-32.9125	
A7	-240.0324							
A5	-238.2064	-127.9730	-266.5821	-266.5821	-151.8683			
A8	-242.7809							
Average	-240.6219	-129.8876	-264.0423	-267.3116	-152.0659	-37.5576	-31.7642	48.5938

Table 3.12: LUMO Energy of different conformers of *o*-coumaric acid with different protonation states.

Type of conformer (*o*)	LUMO Energy of conformers (kcal/mol)							
	Protonated at O(10) (S1 to A8)	Protonated at O(10) and deprotonated at O(12) (S1' to A5')	Protonated at C(7) (S1+2 to A5+8)	Protonated at C(8) (S1+2 to A5+8)	Neutral *m*-coumaric acid (S1+2 to A5+8)	Deprotonated at O(11) (Sα, Sβ, Aα, and Aβ)	Deprotonated at O(12) (Sα, Sβ, Aα, and Aβ)	Deprotonated at both O(11) and O(12) (S and A)
S1	-170.1537	-61.5580	-127.6843	-204.1016	-53.8398	40.6559	44.7535	117.2113
S2	-165.9808	-61.5204						
S4	-164.3744		-187.5104	-187.5104	-47.5083			
S6	-160.2517							
S3	-171.9735	-63.3339	-186.4751	-190.5036	-56.2682	41.9611	43.3291	
S7	-167.4492							
S5	-161.7891	-58.8284	-135.3147	-193.6412	-50.0245			
S8	-166.3197							
A1	-170.9946	-62.6436	-192.2669	-192.2669	-55.1010	40.3924	40.0598	117.7384
A2	-166.4263	-61.9596						
A4	-165.5918		-188.6023	-188.9474	-50.4889			
A6	-161.2808							
A3	-172.7516	-64.0681	-193.9926	-190.3530	-57.5295	39.4323	38.1836	
A7	-167.6249							
A5	-162.6237	-58.9539	-190.3530	-190.3530	-53.1119			
A8	-167.4366							
Average	-166.4389	-61.6082	-175.2749	-192.2097	-52.9840	40.6104	41.5815	117.4748

Table 3.13: HOMO-LUMO Energy gap of different conformers of *o*-coumaric acid with different protonation states.

Type of conformer (*o*)	HOMO-LUMO Energy of conformers (kcal/mol)							
	Protonated at O(10) (S1 to A8)	Protonated at O(10) and deprotonated at O(12) (S1' to A5')	Protonated at C(7) (S1+2 to A5+8)	Protonated at C(8) (S1+2 to A5+8)	Neutral *m*-coumaric acid (S1+2 to A5+8)	Deprotonated at O(11) (Sα, Sβ, Aα, and Aβ)	Deprotonated at O(12) (Sα, Sβ, Aα, and Aβ)	Deprotonated at both O(11) and O(12) (S and A)
S1	72.1440	67.6574	129.8304	63.4092	98.3862	82.5229	75.0682	68.3414
S2	72.3699	68.4229						
S4	75.6769		79.9000	79.9000	102.1449			
S6	75.8400							
S3	73.8822	67.9460	73.9010	79.0591	99.2208	74.7858	76.5303	
S7	74.3403							
S5	77.3146	68.7555	121.7105	72.8782	102.9670			
S8	76.8377							
A1	72.0436	67.6323	75.0807	75.0807	96.1837	81.7950	70.6882	69.4207
A2	71.3973	68.8057						
A4	74.7105		78.1053	78.0300	98.5933			
A6	74.5976							
A3	72.4389	67.9962	75.3819	76.2291	96.4033	73.5685	71.0961	
A7	72.4076							
A5	75.5827	69.0191	76.2291	76.2291	98.7564			
A8	75.3443							
Average	74.1830	68.2794	88.7674	75.1019	99.0819	78.1680	73.3457	68.8810

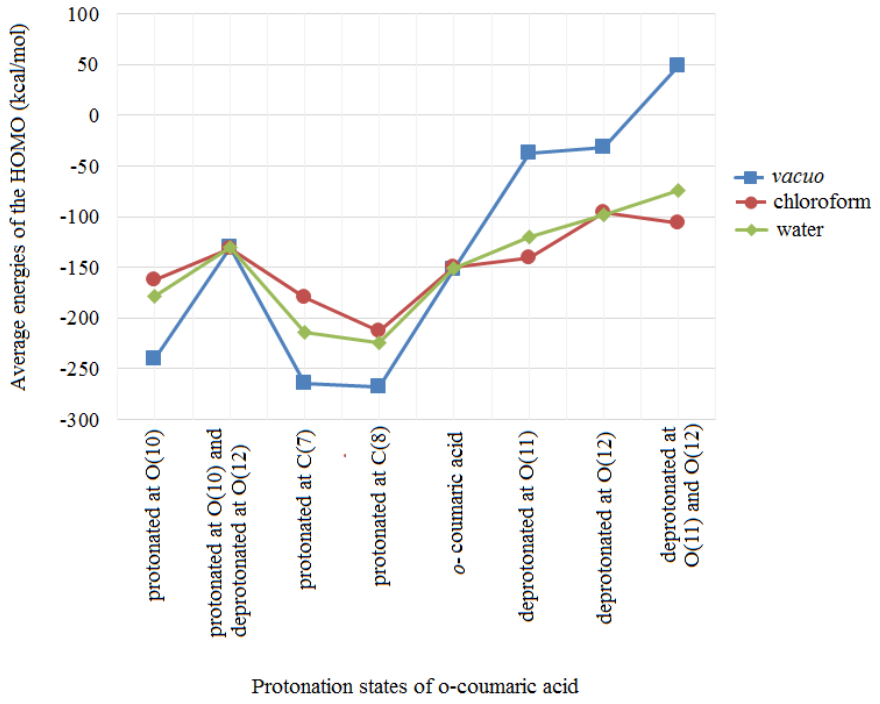
3.3.2 Results considering a solvated environment

For illustrative purposes the average energies of the HOMO and the energies of the LUMO and the HOMO-LUMO of energy gaps *o*-coumaric acid are provided in 2D plots to show figuratively the trends of the values across different protonation states of different derivatives of cinnamic acid.

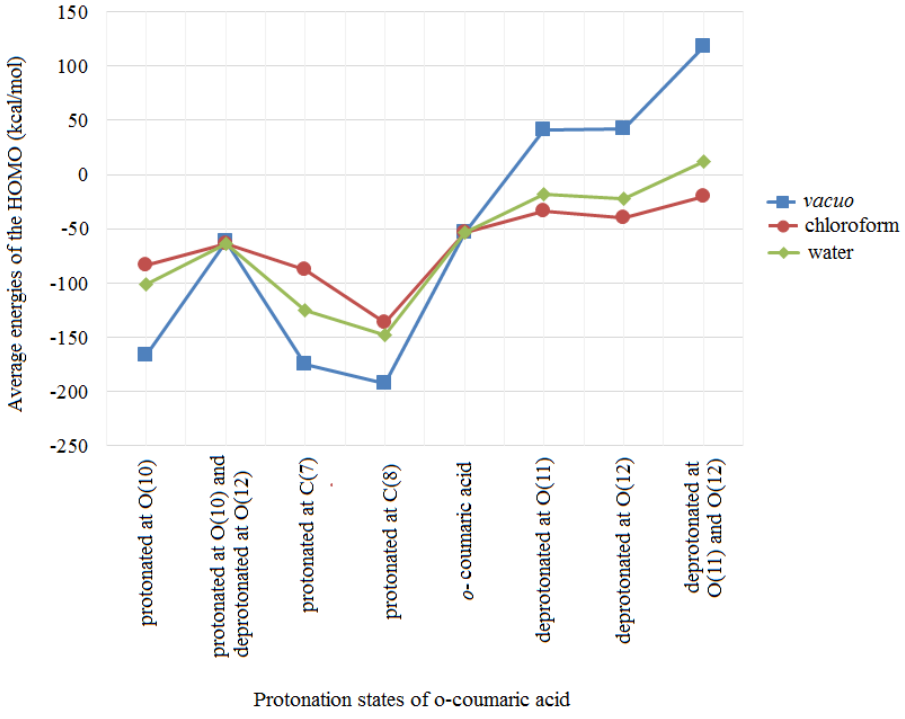
Figure 3.10 gives an overview of the trends of the frontier MO's energies and energy gaps. The figures and discussions in this section are based on the analysis of the data in tables in Appendix B2 and B3 (Supplementary materials in the disc). The results from these tables show that most of the trends of the frontier MO's energy in water and in chloroform are similar those observed *in vacuo,* but the range of values most cases are different. For protonated systems the energy of the frontier MO's *in vacuo* is generally lower than the energy of similar conformers in water, and the frontier MO's energy in water are less than the energy of the MO's from conformers in chloroform. The MO's energies of the neutral systems and the protonated and deprotonated system are relatively the same across different media. The MO's energies of the deprotonated systems in solvated environments are much lower than the energy of the deprotonated systems in vacuo.

The values of the HOMO-LUMO energy gaps have the same general trends in solution as those discussed for the systems studied in vacuo. The structure deprotonated at O(11) and those deprotonated simultaneously at O(11) and O(12) did not follow the general trends in solvated environments (the their energy differences were much higher than those in vacuo, meaning that they are less reactive in in solution).

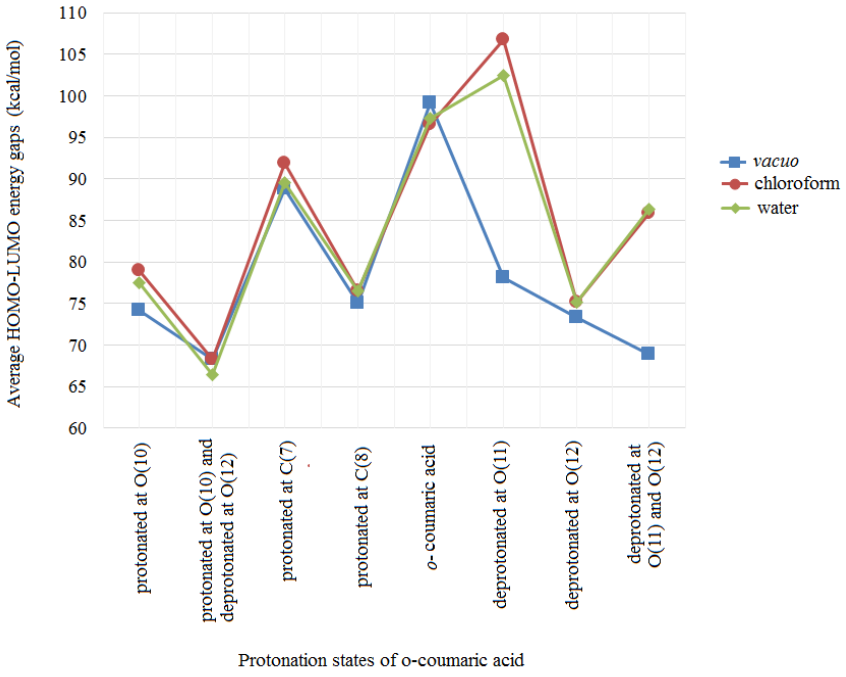
Most of the trends of the MO's figure in solution were similar to those observed *in vacuo*.
Ultimately the simultaneous protonation/deprotonation of these structures perturbs the HOMO-LUMO gap in a way to most likely induce reactivity.


Average energies of the HOMO (kcal/mol)
100
50
0
-50
-100
-150
-200
-250
-300
protonated at O(10)
protonated at O(10) and deprotonated at O(12)
protonated at C(7)
protonated at C(8)
o- coumaric acid
deprotonated at O(11)
deprotonated at O(12)
deprotonated at O(11) and O(12)
vacuo
chloroform
water
Protonation states of o-coumaric acid

(a)


Average energies of the HOMO (kcal/mol)
150
100
50
0
-50
-100
-150
-200
-250
protonated at O(10)
protonated at O(10) and deprotonated at O(12)
protonated at C(7)
protonated at C(8)
o- coumaric acid
deprotonated at O(11)
deprotonated at O(12)
deprotonated at O(11) and O(12)
vacuo
chloroform
water
Protonation states of o-coumaric acid

(b)



(c)

Figure 3.10. The average HOMO-LUMO energy gaps of *o*-coumaric acid of the different protonation states.

3.3.3 Summary of the chapter

The aim of this research project is to find an enzyme that could facilitate in the bioorganic transformation of coumaric acid to coumarin. This chapter aimed at studying the different protonation states of cinnamic acid derivative, in particular *o*-coumaric acid, to determine which of the protonation states is likely to occur, and which of the protonation states is likely to induce reactivity. A mechanism that involves simultaneous protonation and deprotonation of coumaric acid fulfils both of these requirements based on the following reasons.

The relative energy of the simultaneously protonated and deprotonated system is intermediate compared to the relative energy of the other protonation states; although not as stabilized as some other protonation states the requirement for simultaneous protonation/deprotonation is about 20 kcal/mol relative to the neutral structure. Other structures with greater stabilization relative to the neutral system undergo irreversible rearrangement that would not rationalize the transformations that proceed toward coumarin.

For this protonation/deprotonation at O(10) and O(12) respectively, there is induced a bond length of around 1.38 Å for C(7)=C(8). This bond length is second weakest to that resulting from protonating at either C(7) or C(8), but as already observed these latter protonations are not appropriate for further reactivity toward coumarins.

Compared to the energy changes resulting from either protonating or deprotonating a cinnamic acid derivative, simultaneous protonation and deprotonation of cinnamic acid derivatives does not cause a great change in the LOMO energy. The LUMO energies of the deprotonated systems are generally higher than the LUMO energies of the neutral system by more than a 100 kcal/mol (over 150 kcal/mol for systems deprotonated at two positions). The LUMO energies of the protonated systems are lower than the energies of a neutral system (as low as 100 kcal/mol less in some cases).

The HOMO-LUMO energy gaps results showed that generally protonated and deprotonated systems had the smallest HOMO-LUMO energy gaps as compared to the other protonation states.

Further computational studies presented use the global minima obtained from the QM results discussed in this chapter. These studies involve *in silico* docking of the minimized structures from this study, and also theoretically explore the reaction mechanisms that could be involved in the E→Z isomerization of coumarins.

At this point the next step in this process is the identification of an enzyme that could simultaneously protonate and deprotonate *o*-coumaric acid.

CHAPTER 4
IN SILICO DOCKING

4.1 Introduction to the chapter

In recent decades, *in silic*o docking has attracted a lot of attention because it is a reasonably affordable molecular recognition simulation process that predicts the preferred orientation of one molecule relative to another when they are bound in the form of a stable complex. Experimental methods for determining the structures of ligand-bound complexes are difficult and expensive. Molecular docking, which predicts the putative binding modes of a ligand, also provides the binding affinities, and indicates the interaction of the ligand with the enzyme. This chapter is concerned with the *in silic*o docking of the cinnamic acid derivatives discussed in Chapter 3 to eleven wild and mutated chalcone isomerase enzyme (CHI) crystal structures from the protein data bank (PDB) (226,227). *In silic*o docking of the cinnamic acid derivatives in the CHI enzymes were undertaken to determine if these derivatives are theoretically viable substrates for the CHI structures, and also to identify possible reactive-center interactions in the active-site of the CHI structures. The procedure was to explore and establish whether ligand binding to the active-site was acceptable, and then extend the analysis to include a search for protein residues in the active-site that are in close proximity to the ligand and thus capable of facilitating the *E*→*Z* isomerization of coumaric acid.

4.2 Chalcone isomerase (CHI) enzymes and cinnamic acid derivatives

Chalcone isomerase enzymes were selected as putative structures that might facilitate the interconversion of *o*-coumaric acid to coumarins because of the reasons outlined earlier (chapter 1 section 1.3). Chapter 1 section 1.3 also covers. in more detail, the selection of the CHI PDB structures: 1FM7, 1FM8, 1JEP, 1EYQ, 1EYP, 4DOI, 4DOK, 4DOL, 4DOO, 1JXO (Y106F mutant), and 1JX1 (T48A mutant). The full names, description and origin of these CHI

crystal structures are given in Table 4.1. The systematic approach to the preparation of the protein structures are outlined in Chapter 2 Section 2.2.2. Chain A was selected for all crystal structures.

Table 4.1: PDB crystal structures considered in this study [227].

CHI crystal structure (PDB code)	Full name	Description	Origin
1FM7	CHI·7,4′-dihydroxyflavanone	CHI complexed with 7,4′-dihydroxyflavanone	*Medicago sativa*
1FM8	CHI·7,4′-dihydroxyflavanone	CHI complexed with 7,4′-dihydroxyflavanone	*Medicago sativa*
1EYP			
1EYQ			
1JEP	CHI·4′-dihydroxyflavanone	CHI complexed with 4′-dihydroxyflavanone	*Medicago sativa*
4DOO	AtFAP1	Fatty-acid-binding protein	*Arabidopsis thaliana*
1JX0		Y106F mutant	mutant
1JX1		T48A mutant	mutant
4DOI	AtCHI	CHI gene At3g55120	*Arabidopsis thaliana*
4DOL	AtFAP3	Fatty-acid-binding protein	*Arabidopsis thaliana*
4DOK	AtCHIL	CHI gene At5g05270	*Arabidopsis thaliana*

Each ligand considered in this section was obtained from the optimised structures discussed in Chapter 3 and their preparation was discussed in Chapter 2 Section 2.2.1. Figure 4.1 shows the conformation of the ligands utilized as starting structures for all dockings. These conformers were selected because they were the minima of all sampled *in vacu*o conformations of the respective ligands. The outputs from B3LYP/6-31+G(d,p) optimizations were saved as PDB files prior to docking.

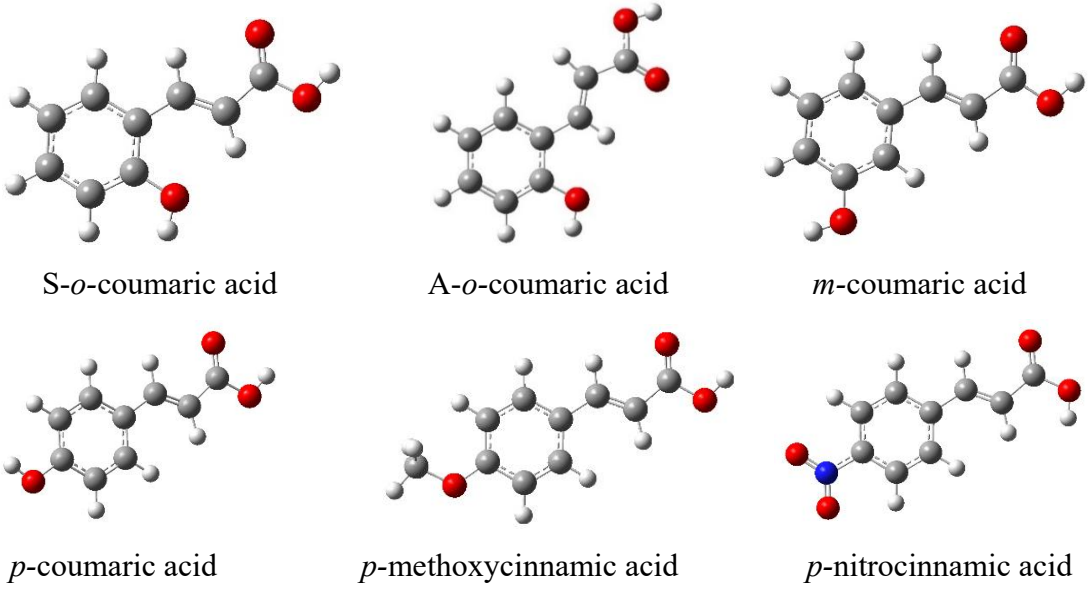


Figure 4.1. Structures of cinnamic acid derivatives optimized with DFT/B3LYP/6-31+G(d,p) *in vacuo*.

4.3 Computational approach

As described in detail earlier, docking was performed using a turnkey computational docking program based on a simple scoring function and rapid gradient-optimization conformational search – Autodock Vina. The results were visualized using Discovery Studio Visualizer. The active-site coordinates were defined by the center of mass of the crystal structure ligands. The crystal structures of 4DOI and 4DOL from the PDB were not complexed with a ligand, but both structures were superimposable with 1FM7. The PDB ligand coordinates of 1FM7 were used as coordinates when docking S-*o*-coumaric acid to 4DOI and 4DOL.

4.4 Results

The results reported in this section were obtained from *in silico* docking of the three derivatives of cinnamic acid, *viz*., coumaric acid (*ortho, meta* and *para*), *p*-methoxycinnamic acid and *p*-nitrocinnamic acid, to 11 crystal structures of CHI. Because *o*-coumaric acid is our primary focus of interest, the structure docking was conducted using two, low-energy starting conformations, *viz*., the "S" conformer with the *syn* arrangement and the "A" conformer with the *anti* arrangement, as obtained from the QM study (Chapter 3).

4.4.1 S-*o*-coumaric acid docked to Chalcone isomerase enzymes

Figures 4.2 (a-l) show S-*o*-coumaric acid docked to 11 crystal structures of CHI and their corresponding receptor-ligand interactions. CHI crystal structures from *M. sativa* (1FM7, 1FM8, 1JEP, 1EYQ) and the mutants (1JX0 and 1JX1) are superimposable and are relatively similar in structure, with their active-sites lying within approximately the same radius. This simplifies comparison of the different docking results. Docking S-*o*-coumaric acid to *M. sativa* CHI and the mutant CHI structures resulted in acceptable docked conformations where S-*o*-coumaric acid was docked within the active-site as shown by figures 4.2 (a-e) (for *M. Sativa* enzymes), and figure 4.2 (h) (for one of the mutant enzymes). These observations are supported by the data in table 4.3 which shows the XYZ coordinates of the center of mass of the ligand from PDB and the XYZ coordinates of the center of mass of the docked ligand. Although docking S-*o*-coumaric acid to the mutant 1JX1 [Figure 4.2(i)] resulted in a relatively good binding energy the ligand did not dock to the active site of this enzyme (the XYZ coordinates of the center of mass of the docked S-*o*-coumaric acid are significantly different from those of the PDB ligand). Table 4.2 gives the binding affinities of all receptor-ligand complexes. The 1FM7- S-*o*-coumaric acid complex had the best binding affinity (-7.1 kcal/mol) of all CHI-S-*o*-coumaric acid complexes considered in this study followed by the 1FM8- S-*o*-coumaric acid complex (-7.0 kcal./mol). In general, CHIs from *M. sativa* have relatively better binding affinities for this ligand than the CHIs from *A. thaliana*. Table 4.2 also details interactions between residues and the ligand that may not be directly apparent in Figure 4.2. Of particular interest is the pervasiveness of the PHE47-ligand π-π interaction, where the PHE47 also serves to hold the orientation of the ligand within the active-site. In Figure 4.2, this is highlighted with a bold, red-dashed line. Also of interest is the apparent lack of interaction with TYR106. The

literature indicates that the CHI-catalysed biochemical formation of chalcone is a water-mediated process in which TYR106 has a critical role. If water is involved, the requirement of proximity of TYR106 to the ligand may be relaxed.

The S-*o*-coumaric acid complexes with *A. thaliana* CHIs (and with the mutant 1JX1) are not as favourable as those with *M. sativa* CHIs, as evidenced not by the binding energies but by the final position of the ligand outside the active site in all cases of the *A. thaliana* CHIs. These observations are shown in figures 4.2 (f), 4.2 (i, j) and 4.2(l). The XYZ coordinates of the center of mass of the crystal structure ligand are quite different from the XYZ coordinates of the docked S-*o*-coumaric acid ligand (table 4.3). It is interesting to note that for crystal structures of 4DOI and 4DOL, where the crystal structure itself does not contain a ligand, binding of S-*o*-coumaric acid was outside of the active site (the active site identified in locality by superimposition with 1FM7).
There are a number of possible inferences to be drawn from this; it is possible that the conformation of the protein is such that the active site is closed to substrates for these two crystal structures, or it may be that the coumaric acid is not a good substrate for these proteins.

It is interesting to note that the dockings in all cases was highly reproducible. Thus, superimposition of the top ten binding conformations in all binding experiments shows no difference in docking position or orientation, but only differences in the conformation of the S-*o*-coumaric acid OH groups (Figure 4.3).

Due to the possibility that the CHI-catalysed biochemical formation of chalcone is a water-mediated process involving Tyr106, it was worth analysing the Solvent Accessibility Surface (SAS) of the pocket containing the docked ligand. Figure 4.4 shows the SAS of the pocket site and the SAS value is given at the bottom of the corresponding figure. Figure 4.5 also shows the distance between Tyr106-OH and O(10) of S-*o*-coumaric acid. The figure shows that the benzene ring of the ligand is generally in a more water accessible environment than the carboxylic acid chain. When the ligand is not docked into the active site, it is observed that the ligand is encompassed by the SAS, and this highlights the lack of meaningful interaction of the ligand with the protein. An unusual case is with the docking of the ligand to 4DOO, in which the binding is not to the active site but deep within the folds of the protein. As such the ligand is virtually inaccessible to solvent as highlighted by the SAS. However, in the desired case

where the ligand binds to the active site, there is a balance between the SAS and the protein-ligand binding surface.

Because of the role Tyr106 plays on the activity of CHI enzyme, this section also focuses on the analysis of several descriptors of Tyr106 (table 4.4). 1FM7 had the best binding affinity, but the *o*-coumaric acid O(10)···HO Tyr106 distance is 8.105 (Å). 1FM8 had the second best binding affinity and its *o*-coumaric acid O(10)···HO Tyr106 distance is 3.687 (Å) which is a distance close enough for hydrogen bonding to occur, or for the Tyr106 to be involved directly in protonation (without the need for water mediation). 1EYQ and 1JEP also have a close *o*-coumaric acid O(10)···HO Tyr106 distance of about 3 (Å), also close enough for protonation. *A. thaliana* CHI and 1JX0 did not have an analogous Tyr106 residue. The hydrophobicity of Tyr106, when present in the CHIs, is -1.3 and its pKa value is 10. Tyr106 of 1JX1 had the highest solvent accessibility of 39.095 and the % solvent accessibility of 17.805, consistent with the observation of solvent accessibility encompassing the ligand; both are highly solvent accessible due to non-binding to the active site. The remaining cases where SAS values are available for Tyr106 are where the ligand is binding. Thus with bound S-*o*-coumaric acid in the active site the SAS value for the Tyr106 is very close to 25 and the % solvent accessibility of this residue is around 11%.

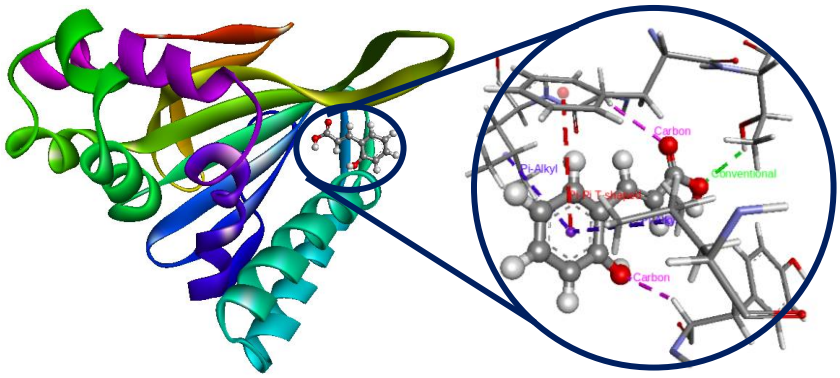


Figure 4.2a. 1FM7 with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interaction.

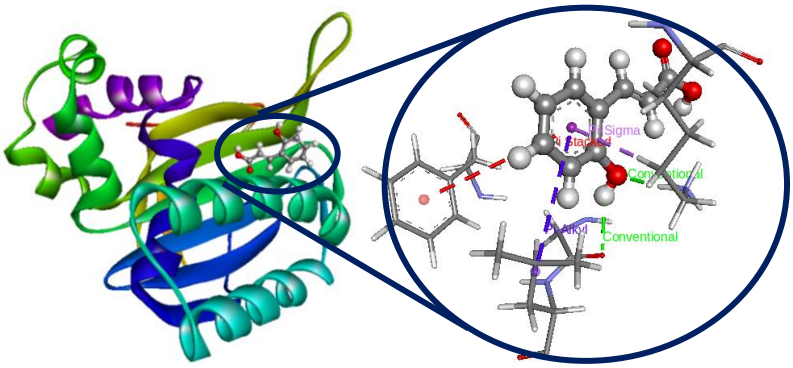


Figure 4.2b. 1FM8 with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interaction.

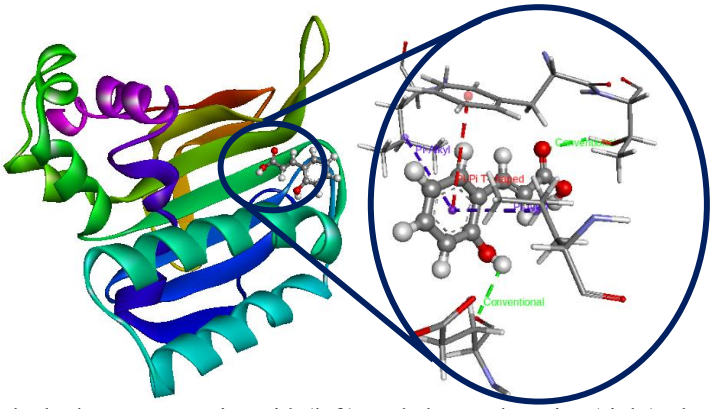


Figure 4.2c. 1EYP with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interactions.

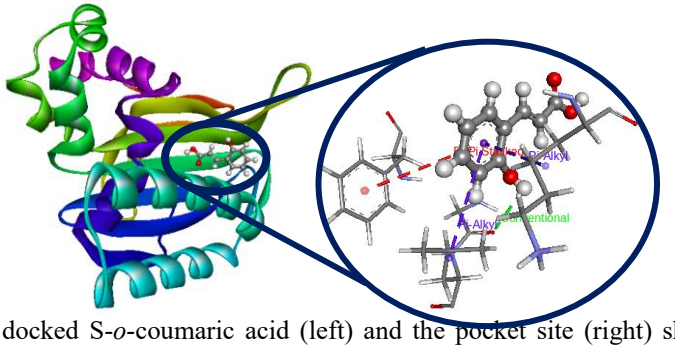


Figure 4.2d. 1EYQ with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interactions.

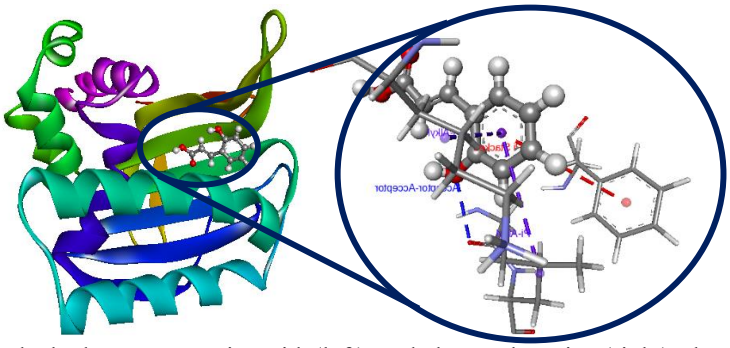


Figure 4.2e. 1JEP with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interaction.

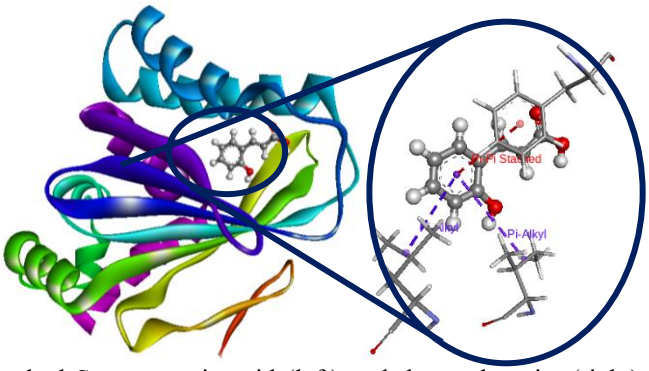


Figure 4.2f. 4DOO with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interactions.

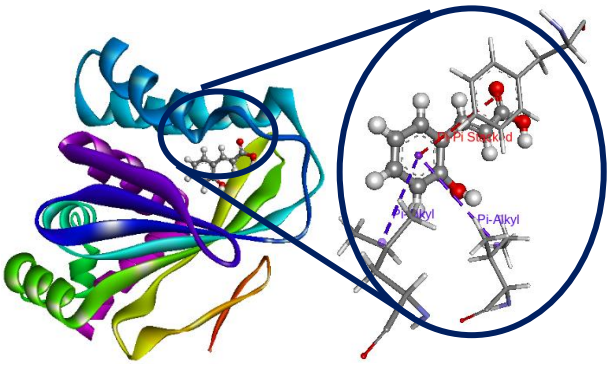


Figure 4.2g. 4DOO_other with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interactions.

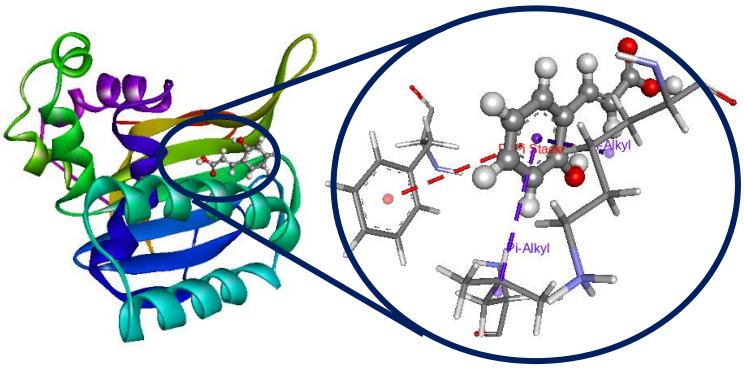


Figure 4.2h. 1JX0 with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interactions.

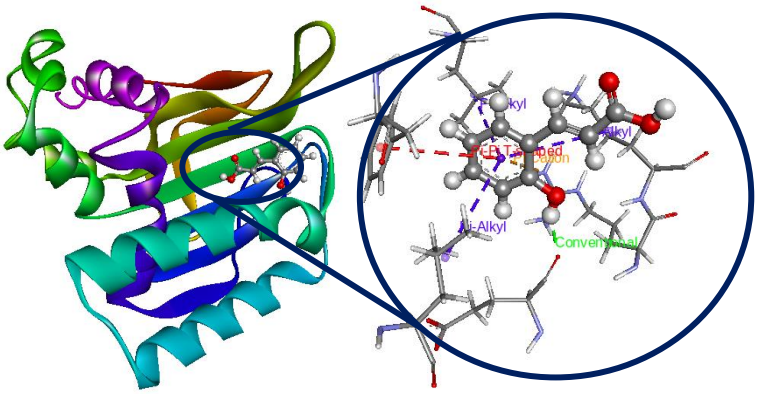


Figure 4.2i. 1JX1 with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interactions.

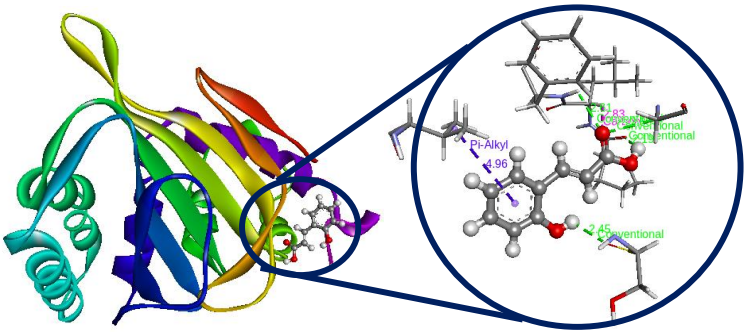


Figure 4.2j. 4DOI with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interactions.

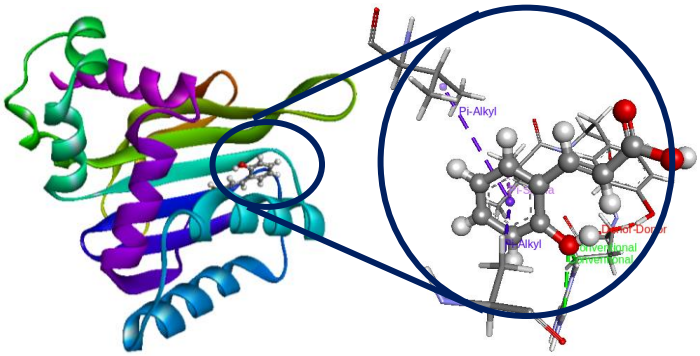


Figure 4.2k. 4DOL with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interactions.

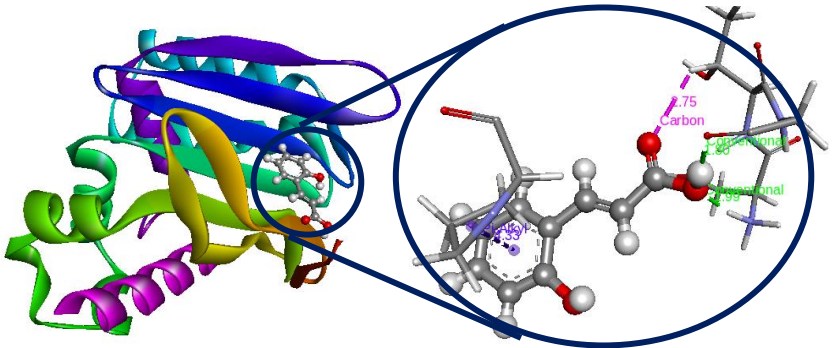


Figure 4.2l. 4DOK with docked S-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interactions.

Figure 4.2. Docked complexes *o*-coumaric acid with 11 crystallographic structures of chalcone isomerase enzymes.

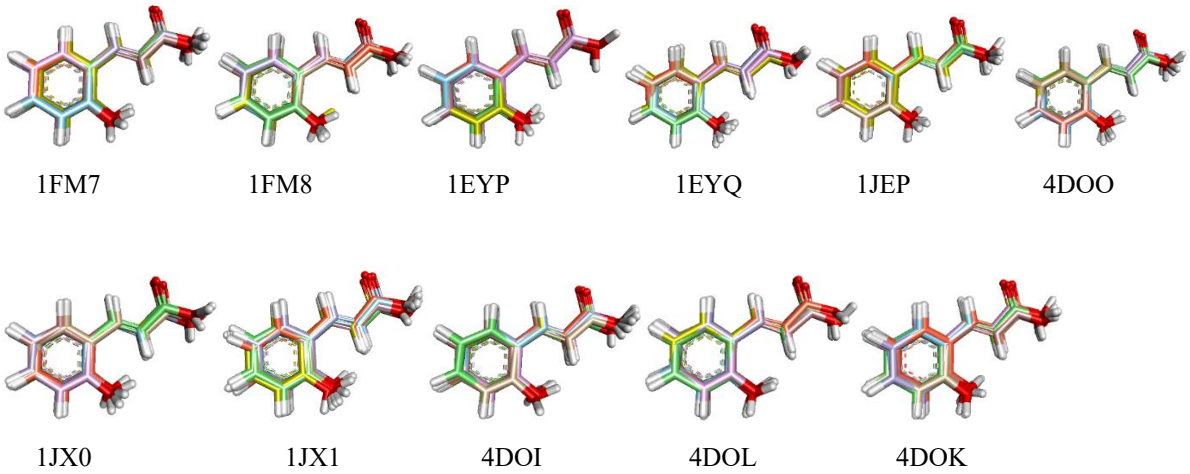


Figure 4.3. Best docked poses of S-*o*-coumaric acid within the chalcone isomerase enzyme.

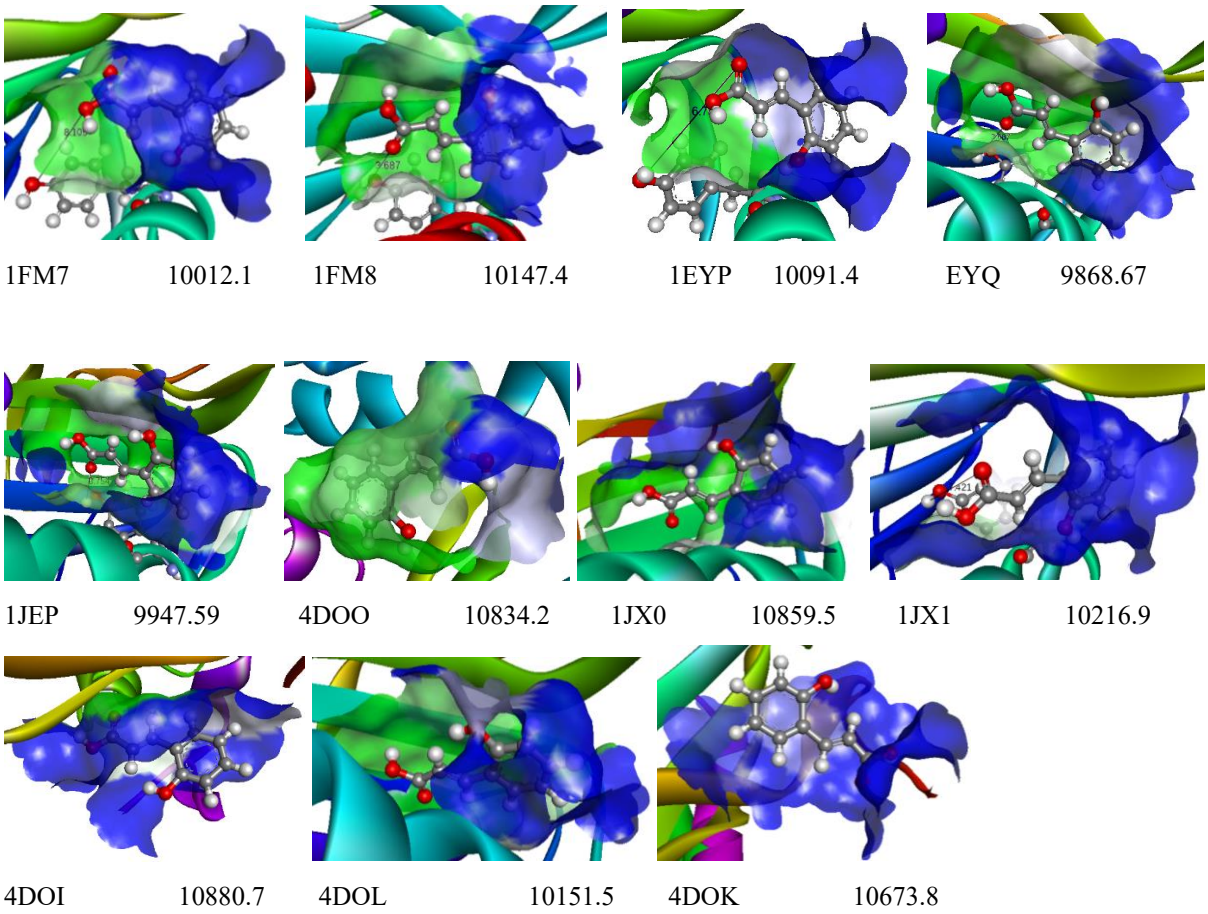


Figure 4.4. Solvent Accessibility Surfaces (SAS) of the ligands.

Table 4.2: Molecular docking binding affinities and the binding interactions of S-*o*-coumaric acid with chalcone isomerase enzymes crystal structures from PDB.

CHI crystal structure (PDB code)	Binding Affinity with best docked conformation (kcal/mol)	Types of interaction between receptor and ligand	Residue interacting with ligand and interaction distance
1EYQ	-6.7	Pi-Pi Stacked	PHE47 5.86 Å
		Pi-Alkyl	LEU38 5.32 Å LYS109 4.57 Å
		Conventional	GLY37 2.72 Å
1FM7	-7.1	Pi-Pi Stacked(T-shaped)	PHE47 4.80 Å
		Pi-Alkyl	LEU101 4.46 Å LEU38 4.79Å
		Conventional	THR48 2.40 Å
		Carbon	GLY37 2.91 Å TYR106 2.58 Å
1FM8	-7.0	Pi-Pi Stacked	PHE47 5.90 Å
		Pi-Alkyl	LEU38 5.30 Å
		Pi-Sigma	LYS109 2.67 Å
		Conventional	GLY37 2.65 Å
1JEP	-6.7	Pi-Pi Stacked	PHE47 5.86 Å
		Pi-Alkyl	LEU38 5.36 Å LYS109 4.65 Å
		Aceptor-Aceptor	GLY37 2.92 Å
1JX0	-6.5	Pi-Pi Stacked	PHE47 5.69 Å
		Pi-Alkyl	LEU38 5.28 Å LYS109 4.29 Å
1EYP	-6.8	Pi-Pi Stacked(T-shaped)	PHE47 4.75
		Pi-Alkyl	LEU101 4.42 LEU38 4.96
		Conventional	GLU105 2.57 THR48 2.28
1JX1	-6.5	Pi-Pi Stacked(T-shaped)	PHE47 4.90 Å
		Pi-Alkyl	LEU38 4.67 Å LEU101 4.80 Å LYS109 5.36 Å
		Pi-Cation	ARG108 3.96 Å
		Conventional	GLY105 1.97 Å
4DOL	-6.2	Pi-Sigma	VAL55 2.85
		Pi-Alkyl	VAL108 5.11 Å ALA116 3.95 Å
		Donor-Donor	TYR56 2.47 Å
		Conventional	ARG44 2.66 Å ARG44 1.86
4DOO	-6.6	Pi-Pi Stacked	PHE134 4.66 Å
		Pi-Alkyl	ILE146 5.45 Å LEU148 5.38 Å

4DOO_other	-6.6	Pi-Pi Stacked	PHE134 4.66 Å
		Pi-Alkyl	ILE146 5.45 Å LEU148 5.38 Å
4DOI	-6.2	Carbon	LEU39 2.83
		Pi-Alkyl	VAL16 4.96
		Conventional	GLY67 2.80 PHE40 2.61 SER234 2.45 PRO38 2.19
4DOK	-4.1	Carbon	THR3 2.75
		Pi-Alkyl	PRO14 4.33
		Conventional	SER2 1.99 GLY4 1.80

Table 4.3: XYZ coordinates of PDB ligand and docked S-*o*-coumaric acid within chalcone isomerase enzyme crystallographic structures.

CHI crystal structure (PDB code)	**Coordinates of the PDB crystal structure ligand and the docked S-*o*-coumaric acid ligand**					
	Ligand COM coordinates from PDB			**S-*o*-Coumaric acid docked COM coordinates**		
	x	**y**	**z**	**x**	**Y**	**z**
1EYQ	26.4094	21.7857	32.9777	25,8295	22,6182	32,1554
1FM7	27.2413	22.1962	32.9759	28,9532	24,2568	29,9638
1FM8	27.2701	21.7562	33.2601	28,3645	23,19	31,21
1JEP	26.8902	22.328	33.4699	28,1667	24,3591	30,3454
1JX0	27.534	21.5262	32.7994	27,5137	22,7495	32,6127
1EYP	26.4094	21.7857	32.9777	28,0558	23,237	31,9809
1JX1	26.4094	21.7857	32.9777	49,3143	37,8528	61,5787
4DOI	27.2413	22.1962	32.9759	31,7395	20,8657	35,5826
4DOK	27.2701	21.7562	33.2601	19,4778	15,1953	24,3849
4DOL	27.2413	22.1962	32.9759	20,8305	0,823356	-17,0523
4DOO	-2.20893	2.51464	-2.86768	-6,3207	6,78212	-2,40977
4DOO_other	1.57793	0.132286	-1.99757	-6,34547	6,84195	-2,41245

Table 4.4: Quantitative properties of Tyr106 related to its activity in the active site of CHI.

CHI crystal structure (PDB code)	O(10)···HO Tyr 106 Hydrogen bond (Å)	Tyr 106 hydrophobicity	Tyr 106 pKa	Tyr 106 Residue solvent accessibility	Tyr 106 % solvent accessibility
1EYQ	3.687	-1.3	10	25,147	11,453
1FM7	8.105	-1.3	10	25,76	11,732
1FM8	3.687	-1.3	10	24,384	11,106
1JEP	3.754	-1.3	10	25.65	11.682
1JX0	N/A	N/A	N/A	N/A	N/A
1EYP	6.750	-1.3	10	24.159	11.003
1JX1	7.421	-1.3	10	39,095	17,805
4DOI	N/A	N/A	N/A	N/A	N/A
4DOK	N/A	N/A	N/A	N/A	N/A
4DOL	N/A	N/A	N/A	N/A	N/A
4DOO	N/A	N/A	N/A	N/A	N/A
4DOO_other	N/A	N/A	N/A	N/A	N/A

4.4.2 A-*o*-coumaric acid docked to chalcone isomerase enzymes

A-*o*-coumaric acid is a conformer of *o*-coumaric acid differing form S-*o*-coumaric acid in the orientation of the alkenoic acid chain. Docking of both the "A" and the "S" conformers was done to see if the orientation of the alkenoic acid chain has a significant influence on the docking pose of the ligand within the receptor and the resulting protein-ligand interactions. This conformational change is not handled by the docking algorithm during docking conformational searching and so it is necessary to separately dock both conformations to the respective enzymes. The results of this section are therefore discussed in comparison to the results of the S-*o*-coumaric acid, where the best binding between A-*o*-coumaric acid or S-*o*-coumaric acid and the respective enzyme is representative of the binding between *o*-coumaric acid and the enzyme. Figure 4.5 shows the docking pose of A-*o*-coumaric acid with 1FM8 (1FM8 was chosen as a representative docking pose because it had the best binding affinity across all of the A-*o*-coumaric acid complexes with of other CHI crystal structures). The remaining figures showing the docking poses with other crystal structures are available in Appendix C (Supplementary material in the disc). Analysis of the ligand docking poses shows that just like the S-*o*-coumaric acid conformer the A-*o*-coumaric acid also maintains its initial

(in this case “A”) conformation when docked to the CHI crystal structures (Figure 4.5). This means that, during docking *o*-coumaric acid does not undergo internal rotation about the single bond connecting the benzene ring and the alkenoic acid chain. Figure 4.5 also shows an example of the protein ligand interactions of A-*o*-coumaric acid with 1FM8; the absence of the π-π interactions with PHE47 is also apparent. The other CHI-A-*o*-coumaric acid interactions are given in Table 4.5 and the figures illustrating such interactions are given in Appendix C, section C1 (Supplementary material in the disc). The data in Table 4.5 together with the poses illustrated Appendix C1 show that the CHI-A-*o*-coumaric acid interactions are similar to the types of interactions observed with CHI-S-*o*-coumaric acid complexes, with the exception of the interaction of A-*o*-coumaric acid with 4DOI, in which case the 4DOI- A-*o*-coumaric acid complex had fewer conventional hydrogen bonding interactions than the 4DOI-S-*o*-coumaric acid complex. However, although CHI-A-*o*-coumaric acid had similar types of interactions as those observed with CHI-S-*o*-coumaric acid, the interactions with A-*o*-coumaric acid were weaker (Table 4.5 shows the length of the interactions in Å, i.e., the further apart the interacting species the weaker the interactions). Figure 4.5 also shows that the docked pose of A-*o*-coumaric acid with 1FM8 has the alkenoic acid chain lies more into the protein and the phenyl group more towards the surface of the protein when compared with S-*o*-coumaric acid (Figure 4.2) . This trend is observed with all A-*o*-coumaric acid complexes with CHIs from *M. sativa* (in comparison to the CHI-S-*o*-coumaric acid complexes with CHI from *M. sativa*). The *A. thaliana* CHI-A-*o*-coumaric acid complexes did not always follow this trend; in fact, in most cases the A-*o*-coumaric acid is docked more into the tail of the protein than on the active compared with the CHI-S-*o*-coumaric acid complexes. Given that neither the S-*o*-coumaric acid nor the A-*o*-coumaric acid bind to the active site for *A. thaliana* CHIs, this variability is not unexpected.

Table 4.5 also provides the binding affinities corresponding to the CHI-A-*o*-coumaric acid complexes. The binding affinities follow the same trend as that observed with CHI-S-*o*-coumaric acid complexes. The differences in the binding affinity of CHI-A-*o*-coumaric acid complexes from that of CHI-S-*o*-coumaric acid complexes is less than 1 kcal/mol (the CHI-A-*o*-coumaric acid complexes generally had the lower value), The 4DOI-A-*o*-coumaric acid complex has a more pronounced energy difference compared to the other CHI complexes [4DOI-A-*o*-coumaric acid complex has a binding affinity of -5.7 kcal/mol while the 4DOI-S-*o*-coumaric acid complex (from Table 4.2) has a binding affinity of -6.2 kcal/mol]

Figure 4.6 shows the superimposed geometries of the docked poses of the ligand in each of the ten most stable CHI-A-*o*-coumaric acid complexes. This figure shows that the orientation of some of the alkenoic acid chains are so different that the top 10 docked poses are not superimposable; this is a very different situation from what was observed with the CHI-S-*o*-coumaric acid complexes (where the only variation in docking pose was observed in the orientation of OH groups). The greater variability in the docking pose within the active site translates to a greater uncertainty in the binding for this conformer when compared to binding for the S-*o*-coumaric acid. These observations, together with the somewhat greater binding affinities for the S-*o*-coumaric acid, and the slight preference for the S-*o*-coumaric acid (as observed in the conformer population analysis studies) suggests that S-*o*-coumaric acid is the conformation that is more likely to interact with the CHI.

Table 4.6 also shows the XYZ coordinates used for docking the ligand and the XYZ coordinates of the center of mass of the docked ligand for A-*o*-coumaric acid complexes. Comparison of these XYZ coordinates suggests that the ligands generally follow docking trends which are similar to those observed with the S-*o*-coumaric acid complexes.
The distance between the of TYR106 OH hydrogen atom and the oxygen atom of the carboxylate carbonyl of A-*o*-coumaric acid (Table 6.7) is generally about 5 Å (this is the optimal bond distance that may be attained with this conformation of coumaric acid). This is much worse than the distance in the corresponding S-*o*-coumaric acid complexes, where a distance of 3 Å was attained in some cases. This is another reason for considering reactivity of the S-*o*-coumaric acid conformer rather than on the A-*o*-coumaric acid in these systems.

The SAS trends for the CHI-A-*o*-coumaric acid complexes were similar to those observed with CHI-S-*o*-coumaric acid complexes, and do not provide further evidence either way either of the conformers. The SAS data for these complexes is presented in Appendix E (supplementary material in the disc).

From these observations it could be hypothesised that in an ideal system where both the “S” and the “A” conformers of *o*-coumaric acid are highly populated, there should be very slight preference for binding the S conformer (the difference between the binding affinities being less than 1 kcal/mol). Any A conformer will be less tightly bound and be slightly more prone to either dissociation from the enzyme or conformational change close to or within the active site to the favoured S conformer. These differences are very small and, arguably, less than the

expected error for these methods. However, if coumaric acid is a viable substrate of this enzyme, then reactivity towards this substrate will only be observed on the S conformer in terms of participating residue alignment for this conformer. Consequently, it was decided that further study at QM/MM level should be concentrated on the S conformer.

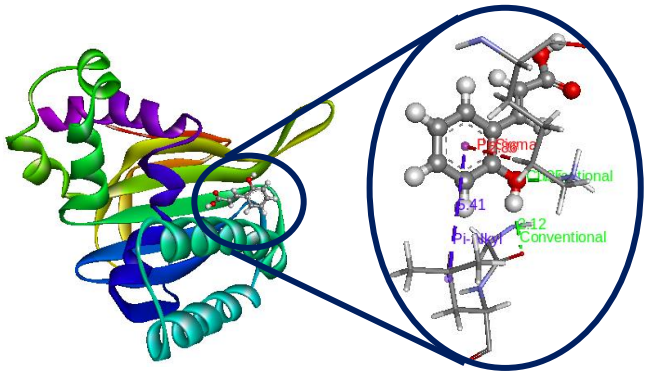


Figure 4.5. 1FM8 with docked A-*o*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interaction.

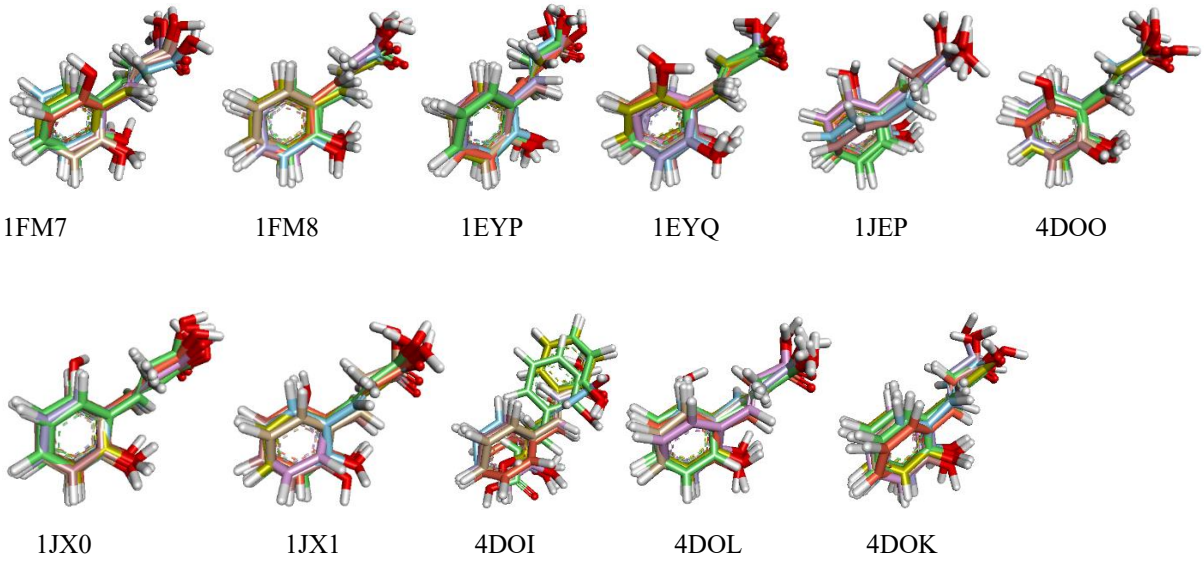


Figure 4.6. Best docked poses of A-*o*-coumaric acid within the chalcone isomerase enzyme.

Table 4.5: Molecular docking binding affinities and the binding interactions of A-*o*-coumaric acid with chalcone isomerase enzyme crystal structures from PDB.

CHI crystal structure (PDB code)	Binding Affinity with best docked conformation (kcal/mol)	Types of interaction between receptor and ligand	Residue interacting with ligand and interaction distance
1FM7	-6.9	Pi-Pi Stacked	PHE47 5.54 Å
		Pi-Alkyl	LEU101 5.43 Å LEU38 5.25 Å LYS109 4.48 Å
		Conventional	LYS109 2.67 Å LYS109 2.37 Å
1FM8	-6.9	Pi-Sigma	LYS109 2.68 Å
		Pi-Alkyl	LEU38 5.41 Å
		Conventional	GLY37 2.12 Å LYS109 1.95 Å
1EYP	-6.6	Pi-Pi Stacked	PHE47 4.78 Å
		Pi-Alkyl	LEU38 5.03 Å LEU101 4.37 Å
		Conventional	THR48 1.74 Å
1EYQ	-6.8	Pi-Alkyl	LEU101 5.50 Å LEU38 5.33 Å LYS109 4.44 Å
		Conventional	TYR106 2.44 Å
1JEP	-6.7	Pi-Pi Stacked	PHE47 5.99 Å
		Pi-Alkyl	LEU38 5.41 Å LYS109 4.49 Å
4DOO	-6.3	Pi-Pi Stacked(T-shaped)	PHE124 4.36
		Pi-Alkyl	VAL103 5.18 ILE140 4.95
1JX0	-6.6	Pi-Alkyl	ILE50 5.20 Å MET191 5.06 Å VAL110 4.49 Å
		Conventional	LYS109 2.68 Å ASN113 2.68 Å ARG36 2.27 Å ASN113 2.12 Å ARG36 2.09 Å
1JX1	-6.6	Pi-Pi T-shaped	TYR106 4.73 Å
		Pi-Alkyl	ALA48 5.02 Å ARG36 3.73 Å
		Carbon	TYR106 2.98 Å
		Conventional	ARG108 2.75Å
4DOI	-5.7	Pi-Alkyl	VAL16 4.93 Å
		Conventional	ASP235 3.33 Å PRO38 2.23 Å
4DOL	-6.3	Pi-Pi T-shaped	PHE113 5.11 Å
		Pi-Alkyl	LEU104 5.25 Å

			VAL108	4.79 Å
		Conventional	ARG44	2.41 Å
4DOK	-4.1	Pi-Sigma	GLN16	2.98 Å
		Conventional	THR	1.99 Å

Table 4.6: XYZ coordinates of PDB ligand and docked A-*o*-coumaric acid within chalcone isomerase enzyme crystallographic structures.

CHI crystal structure (PDB code)	Coordinates of the PDB crystal structure ligand and the docked A-*o*-coumaric acid ligand					
	Ligand coordinates from PDB			A-*o*-coumaric acid docked coordinates		
	x	y	z	x	y	z
1FM7	26.4094	21.7857	32.9777	26.6995	23.152	32.7074
1FM8	27.2413	22.1962	32.9759	27.2429	22.6844	32.6475
1EYP	27.2701	21.7562	33.2601	29.0795	24.8302	29.5995
1EYQ	26.8902	22.328	33.4699	26.2714	22.7066	32.0832
1JEP	27.534	21.5262	32.7994	26.5098	22.7366	32.8356
4DOO	26.4094	21.7857	32.9777	-5.30181	7.57945	-1.80283
1JX0	27.2413	22.1962	32.9759	26.1382	19.3536	34.6372
1JX1	27.2701	21.7562	33.2601	50.0511	40.7141	61.2479
4DOI	27.2413	22.1962	32.9759	24.2755	21.54	41.327
4DOL	-2.20893	2.51464	-2.86768	19.9708	0.929354	-16.8265
4DOK	1.57793	0.132286	-1.99757	18.9658	18.0419	26.0632

Table 4.7: Quantitative properties of Tyr106 related to its activity in the active site of CHI.

CHI crystal structure (PDB code)	O(10)···HO Tyr106 Hydrogen bond (Å)	Tyr106 hydrophobicity	Tyr106 pKa	Tyr106 Residue solvent accessibility	Tyr106 % solvent accecibility
1FM7	5.268	-1.3	10	25,76	11,732
1FM8	5.634	-1.3	10	24,384	11,106
1EYP	5.245	-1.3	10	24.159	11.003
1EYQ	5.449	-1.3	10	25,147	11,453
1JEP	5.605	-1.3	10	25,65	11,682
4DOO	N/A	N/A	N/A	N/A	N/A
1JX0	N/A	N/A	N/A	N/A	N/A
1JX1	9.204	-1.3	10	39,095	17,805
4DOI	N/A	N/A	N/A	N/A	N/A
4DOL	N/A	N/A	N/A	N/A	N/A
4DOK	N/A	N/A	N/A	N/A	N/A

4.4.3 Other cinnamic acid derivatives docked to Chalcone isomerase enzymes

The other cinnamic acid derivatives docked to CHI crystal structures included the *meta*- and the *para*-coumaric acid structures, and the *p*-methoxy and *p*-nitro cinnamic acids. The *meta*- and the *para*-coumaric acids were used in order to explore if the position of the OH on a coumaric acid had a significant effect on the protein-ligand interactions and the binding affinities. *p*-Methoxycinnamic acid and *p*-nitrocinnamic acid were docked to see the effects of either electron-donating or electron-withdrawing groups on the binding to the CHI enzyme. For clarity, the effects of these groups were analysed simultaneously. The results show that the *m*- and the *p*- position of the OH on coumarin have no significant effects on the trends observed with the *o*-coumaric acid (Figure 4.7-4.14, and Table 4.8-4.19).

Figure 4.8, 4.10, and 4.12 showed that the docking is less reproducible with the top ten docking conformations of the *m*- and *p*- coumaric acids and the *p*-methoxycinnamic acid ligand not being perfectly superimposable). Figure 4.14 shows the reproducibility of the *p*-nitrocinnamic acid derivative, with the top 10 docked poses being highly superimposable and differing only in the orientation the nitro and the OH groups.

Table 4.8, 4.11, 4.14, and 4.17 show that ligands binding with 1FM7 have the best binding affinities (*ca.* 6 kcal/mol) followed by 1FM8 (the difference of 0.1 kcal/mol for the *m*- and *p*-coumaric acid, and 0.2 kcal/mol for the nitro and methoxycinnamic acid derivatives). 4DOK complexes have the weakest binding affinities for all cinnamic acid derivatives (*ca*. 4 kcal/mol). Table 4.8, 4.11, 4.14, and 4.17 also shows that these cinnamic acid derivatives are held to the CHI receptors by several interactions, the most common of which are the π-alkyl interactions observed between the benzene ring of the ligand and LEU38 for the *M.sativa* CHI. The π-π interactions with PHE47 observed with the *o*-coumaric acid were seldom observed with these other cinnamic acid derivatives. The nitro group in the *p*-nitrocinnamic acid derivatives contributes to the stabilization of the CHI- *p*-nitrocinnamic acid complex by forming hydrogen bonds between the nitro group and protein residue side-chain OH groups.

Tables 4.9, 4.12, 4.15, and 4.18 show that the ligand in *M.sativa* CHI docked to the active site (mass centered XYZ coordinates of the docked ligand relatively close to the mass centered XYZ coordinates of the crystal structure ligand).

Tables 4.10, 4.13, 4.16, and 4.19 show that the O(10)···HO Tyr106 distance is relatively large for the other cinnamic acid derivatives as compared to the O(10)···HO Tyr106 distance with *o*-coumaric acid derivatives with CHI. The O(10)···HO Tyr106 distance in the 1FM8- *m*-coumaric acid complex is 5 Å, but 8 Å in complexes with *p*-coumaric acid and *p*-methoxycinnamic acid, 10Å for *p*-nitrocinnamic acid. Tables 4.10, 4.13, 4.16, and 4.19 also show that the SAS of Tyr106 in all complexes is averaged around 25 with the %SAS avaraged *ca.* 11%.

The results with *p*-methoxycinnamic acid also follow these trends. The *p*-nitrocinnamic acid showed some deviations, i.e., the nitro group was generally more embedded within the protein after docking, with the alkenoic acid chain being oriented towards the surface of the protein (Figure 4.13 shows *p*-nitroxycinnamic acid docked to 1FM8). Because the nitro group is deeper within the enzyme, the alkenoic acid chain of the *p*-nitroxycinnamic acid does not interact with the reactive residues in the active site of the enzyme. This informs us that if the CHIs do mediate the E→Z isomerization, the addition of an electron withdrawing group on the coumaric acid will result in a substrate that is not processed by this enzyme (i.e. it will no longer be a viable substrate). However, the binding affinity of the *p*-nitroxycinnamic acid is very slightly better than the binding affinity of the other cinnamic acid derivatives (the best *p*-nitroxycinnamic acid binding affinity is 7.2 kcal/mol and that of *o*-coumaric acid is 7.1 kcal/mol).

m-Coumaric acid docked to chalcone isomerase enzymes.

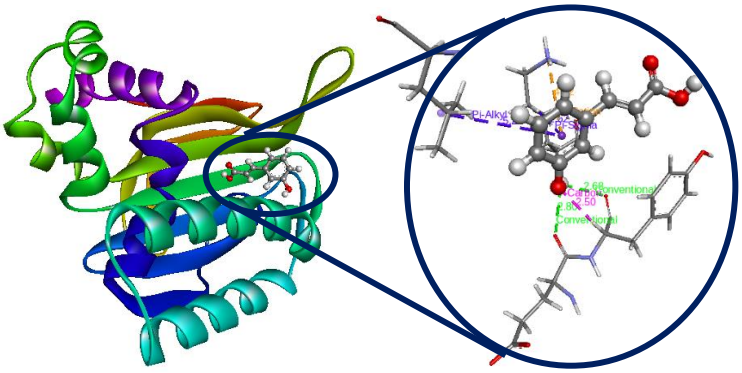


Figure 4.7. 1FM8 with docked *m*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interaction.

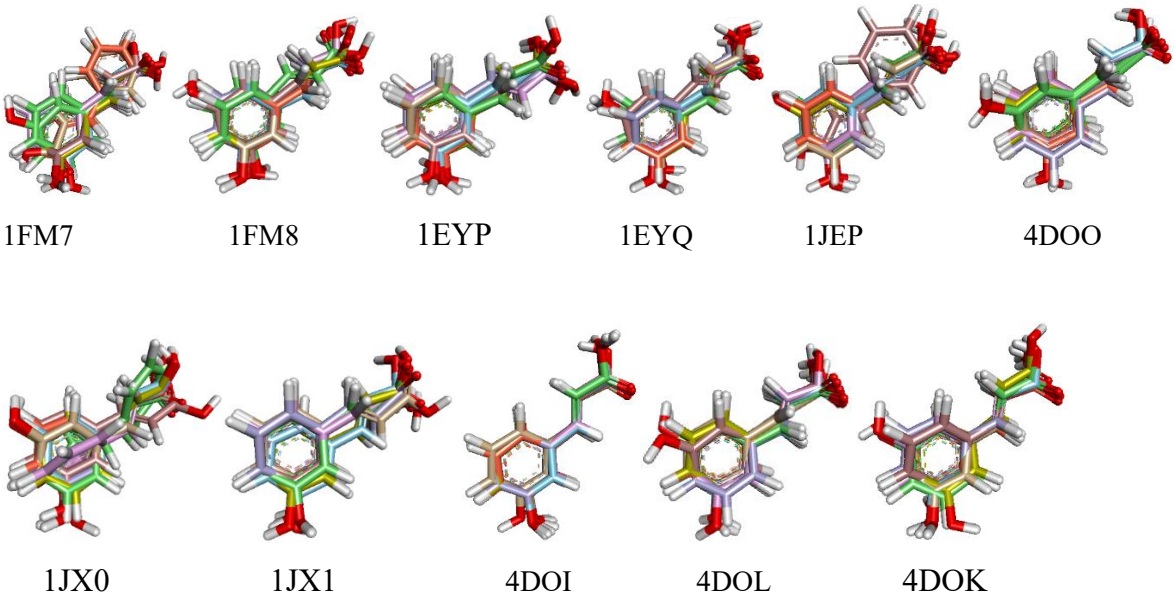


Figure 4.8. Best docked poses of *m*-coumaric acid within the chalcone isomerase enzyme.

Table 4.8: Molecular docking binding affinities and the binding interactions of *m*-coumaric acid with chalcone isomerase enzymes crystal structures from PDB.

CHI crystal structure (PDB code)	Binding Affinity with best docked conformation (kcal/mol)	Types of interaction between receptor and ligand	Residue interacting with ligand and interaction distance
1FM7	-6.8	Pi-Pi Stacked	PHE47 4.97 Å
		Pi-Alkyl	LEU101 4.97 Å
		Pi-Sigma	LEU38 2.94 Å
		Conventional	LYS109 2.71 Å LYS109 2.53 Å GLU105 2.35 Å
1FM8	-6.7	Pi-Cation	LYS109 4.32 Å
		Pi-Alkyl	LEU38 5.28 Å
		Pi-Sigma	LYS109 2.62 Å
		Carbon	TYR106 2.50 Å
		Conventional	GLU105 3.80 Å TYR106 2.68 Å
1EYP	-6.3	Pi-Pi Stacked	PHE47 4.65 Å
		Pi-Alkyl	LEU101 4.40 Å LEU38 4.93 Å
		Conventional	THR48 2.81 Å ARG108 2.52 Å THR48 1.78 Å
1EYQ	-6.8	Pi-Alkyl	LEU38 5.11 Å LEU101 5.09 Å LYS109 4.73 Å
		Carbon	TYR106 2.61 Å
		Conventional	GLU105 2.76 Å
1JEP	-6.7	Pi-Alkyl	LEU38 5.15 Å LEU101 5.10 Å LYS109 4.85 Å
		Conventional	GLY105 2.48 Å
4DOO	-6.5	Pi-Pi Stacked	PHE134 4.78 Å
		Pi-Alkyl	LEU91 5.44 Å ILE146 5.42 Å ILE148 5.27 Å
1JX0	-6.6	Pi-Pi T-shaped	PHE106 5.23 Å
		Pi-Alkyl	ILE50 5.21 Å MET191 5.14 Å VAL110 4.48 Å
		Conventional	ASN113 2.62 Å ARG36 2.41 Å ASN113 2.50 Å ARG36 2.00 Å
1JX1	-6.6	Pi-Pi T-shaped	TYR106 4.71 Å
		Pi-Alkyl	ILE50 5.49 Å ALA48 5.00 Å ARG36 3.75 Å
		Conventional	ARG108 2.84Å

			LYS109 2.73Å GLU105 2.50Å TYR106 2.22Å
4DOI	-5.5	Pi-Alkyl	VAL16 5.31 Å
		Carbon	LEU39 2.87 Å
		Conventional	ASP235 3.04 Å GLY67 2.95 Å PHE40 2.71 Å PRO38 2.44 Å
4DOL	-6.1	Pi-Pi T-shaped	TRY56 5.22 Å PHE113 5.22 Å
		Pi-Alkyl	VAL108 5.11 Å
4DOK	-4.2	Donor-Donor	THR3 1.40 Å
		Pi-Alkyl	ELE17 5.13 Å PRO14 4.54 Å
		Conventional	ELE17 2.52 Å SER2 2.15 Å

Table 4.9: XYZ coordinates of PDB ligand and docked *m*-coumaric acid within chalcone isomerase enzyme crystallographic structures.

CHI crystal structure (PDB code)	Coordinates of the PDB crystal structure ligand and the docked *m*-coumaric acid ligand					
	Ligand coordinates from PDB			*m*-coumaric acid docked coordinates		
	x	y	z	x	y	z
1FM7	26.4094	21.7857	32.9777	29.0694	25.058	29.7929
1FM8	27.2413	22.1962	32.9759	27.3868	22.5755	32.4513
1EYP	27.2701	21.7562	33.2601	29.2493	24.9717	29.4334
1EYQ	26.8902	22.328	33.4699	26.5865	23.0938	31.6679
1JEP	27.534	21.5262	32.7994	26.7344	23.1564	32.098
4DOO	26.4094	21.7857	32.9777	-6.36619	6.66114	-2.76246
1JX0	27.2413	22.1962	32.9759	25.9177	19.5613	34.3322
1JX1	27.2701	21.7562	33.2601	49.9875	41.095	61.3105
4DOI	27.2413	22.1962	32.9759	24.6136	20.9425	40.6463
4DOL	-2.20893	2.51464	-2.86768	20.0127	0.911357	-16.86
4DOK	1.57793	0.132286	-1.99757	20.2759	14.615	25.5241

Table 4.10: Quantitative properties of Tyr106 related to its activity in the active site of CHI.

CHI crystal structure (PDB code)	O(10)···HO Tyr106 Hydrogen bond (Å)	Tyr106 hydrophobicity	Tyr106 pKa	Tyr106 Residue solvent accessibility	Tyr106 % solvent accecibility
1FM7	8.537	-1.3	10	25,76	11,732
1FM8	5.690	-1.3	10	24,384	11,106
1EYP	4.903	-1.3	10	24,159	11,003
1EYQ	4.074	-1.3	10	25,147	11,453
1JEP	4.167	-1.3	10	25,65	11,682
4DOO	N/A	N/A	N/A	N/A	N/A
1JX0	N/A	N/A	N/A	N/A	N/A
1JX1	9.164	-1.3	10	39,095	17,805
4DOI	N/A	N/A	N/A	N/A	N/A
4DOL	N/A	N/A	N/A	N/A	N/A
4DOK	N/A	N/A	N/A	N/A	N/A

***p*-Coumaric acid docked to chalcone isomerase enzymes.**

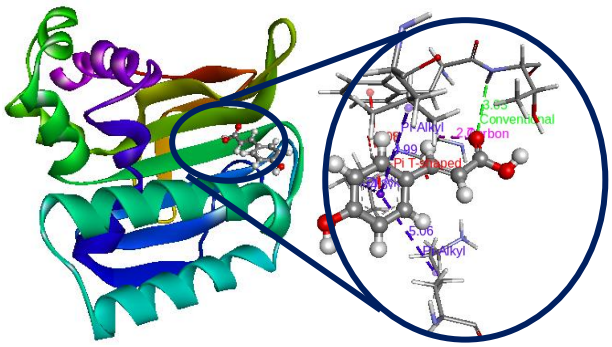


Figure 4.9. 1FM8 with docked *p*-coumaric acid (left) and the pocket site (right) showing receptor-ligand interaction.

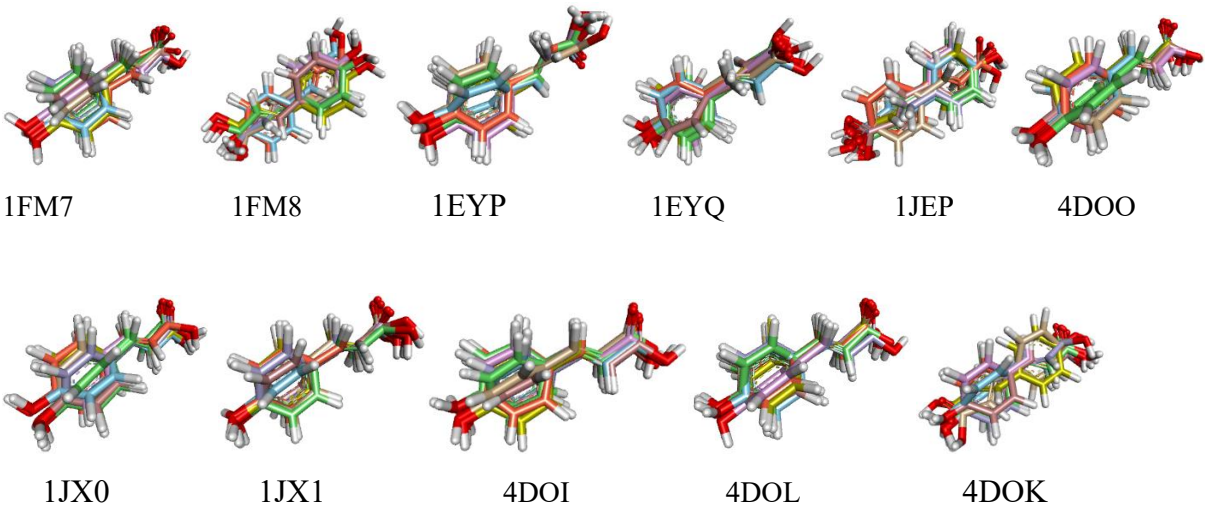


Figure 4.10. Best docked poses of *p*-coumaric acid within the chalcone isomerase enzyme.

Table 4.11: Molecular docking binding affinities and the binding interactions of *p*-coumaric acid with chalcone isomerase enzymes crystal structures from PDB.

CHI crystal structure (PDB code)	Binding Affinity with best docked conformer (kcal/mol)	Types of interaction between receptor and ligand	Residue interacting with ligand and interaction distance
1FM7	-6.5	Pi-Alkyl	LEU101 4.61 Å LYS109 4.79 Å LEU38 5.40 Å
		Conventional	THR48 1.90 Å
1FM8	-6.4	Pi-Pi T-shaped	PHE47 5.08 Å
		Pi-Alkyl	LYS109 5.06 Å LEU101 4.99 Å LEU38 4.37 Å
		Carbon	GLY37 2.74 Å
		Conventional	TYR48 3.05 Å
1EYP	-6.1	Pi-Alkyl	LEU101 4.21 Å LEU38 5.15 Å
		Conventional	ARG36 2.76 Å THR48 1.79 Å
1EYQ	-6.4	Pi-Alkyl	LEU38 5.38 Å LEU101 5.37 Å LYS109 4.75 Å
1JEP	-6.3	Pi-Alkyl	LEU38 5.28 Å LEU101 5.47 Å LYS109 4.83 Å
4DOO	-6.4	Pi-Pi Stacked	PHE134 4.25 Å
		Pi-Alkyl	ILE146 4.83 Å
1JX0	-6.5	Pi-Pi T-shaped	PHE106 5.03 Å
		Pi-Alkyl	MET191 5.45 Å ARG36 5.44 Å ILE50 4.91 Å VAL110 4.82 Å

		Conventional	ASN113 2.52 Å ARG36 2.25 Å ARG36 2.04 Å
1JX1	-6.3	Pi-Pi T-shaped	TYR106 4.54 Å
		Carbon	TYR106 3.02 Å
		Acceptor-Acceptor	GLU105 2.94 Å
		Pi-Alkyl	ALA48 5.04 Å ARG36 3.93 Å
		Conventional	ARG108 2.90 Å
4DOI	-5.4	Pi-Alkyl	VAL16 5.20 Å
		Carbon	LEU39 2.81 Å
		Conventional	ASP234 3.03 Å GLY67 2.87 Å PHE40 2.76 Å ASP235 2.50 Å
4DOL	-5.8	Pi-Alkyl	ILE102 5.31 Å
		Conventional	ARG44 2.40 Å TRY56 2.15 Å
4DOK	-4.1	Pi-Sigma	GLN16 2.95 Å
		Conventional	GLY4 2.48 Å SER2 2.99 Å

Table 4.12: XYZ coordinates of PDB ligand and docked *p*-coumaric acid within chalcone isomerase enzyme crystallographic structures.

CHI crystal structure (PDB code)	Coordinates of the PDB crystal structure ligand and the docked *p*-coumaric acid ligand					
	Ligand coordinates from PDB			*p*-coumaric acid docked coordinates		
	x	y	z	x	y	z
1FM7	26.4094	21.7857	32.9777	27.4446	23.6199	31.2488
1FM8	27.2413	22.1962	32.9759	28.9632	24.4969	30.0677
1EYP	27.2701	21.7562	33.2601	29.1968	25.0901	29.2812
1EYQ	26.8902	22.328	33.4699	26.2894	23.0142	31.8877
1JEP	27.534	21.5262	32.7994	26.7312	23.2007	32.0338
4DOO	26.4094	21.7857	32.9777	-6.22668	7.41136	-1.28452
1JX0	27.2413	22.1962	32.9759	25.637	19.6865	34.1849
1JX1	27.2701	21.7562	33.2601	50.0069	41.1136	61.2007
4DOI	27.2413	22.1962	32.9759	24.6394	21.5026	40.6683
4DOL	-2.20893	2.51464	-2.86768	20.7337	-2.24826	-14.7127
4DOK	1.57793	0.132286	-1.99757	19.2666	17.6626	26.2179

Table 4.13: Quantitative properties of Tyr106 related to its activity in the active site of CHI.

CHI crystal structure (PDB code)	O(10)···HO Tyr106 Hydrogen bond (Å)	Tyr106 hydrophobicity	Tyr106 pKa	Tyr106 Residue solvent accessibility	Tyr106 % solvent accecibility
1FM7	5.126	-1.3	10	25,76	11,732
1FM8	8.149	-1.3	10	24,384	11,106
1EYP	5.408	-1.3	10	24,159	11,003
1EYQ	5.663	-1.3	10	25,147	11,453
1JEP	5.750	-1.3	10	25,65	11,682
4DOO	N/A	N/A	N/A	N/A	N/A
1JX0	N/A	N/A	N/A	N/A	N/A
1JX1	9.224	-1.3	10	39,095	17,805
4DOI	N/A	N/A	N/A	N/A	N/A
4DOL	N/A	N/A	N/A	N/A	N/A
4DOK	N/A	N/A	N/A	N/A	N/A

***p*-Methoxycinnamic acid docked to chalcone isomerase enzymes.**

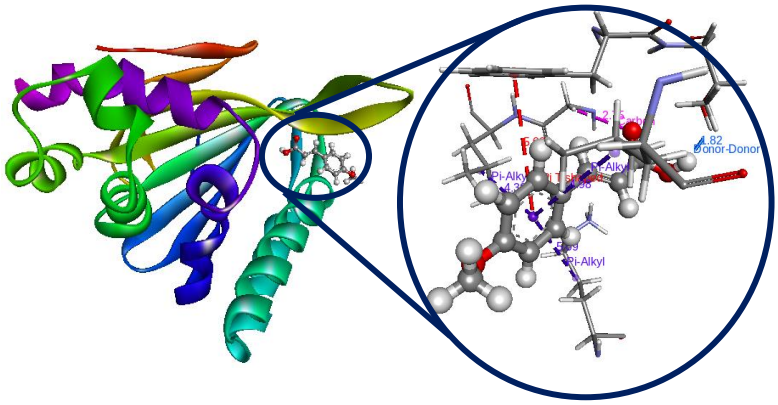


Figure 4.11. 1FM8 with docked *p*-methoxycinnamic acid (left) and the pocket site (right) showing receptor-ligand interaction.

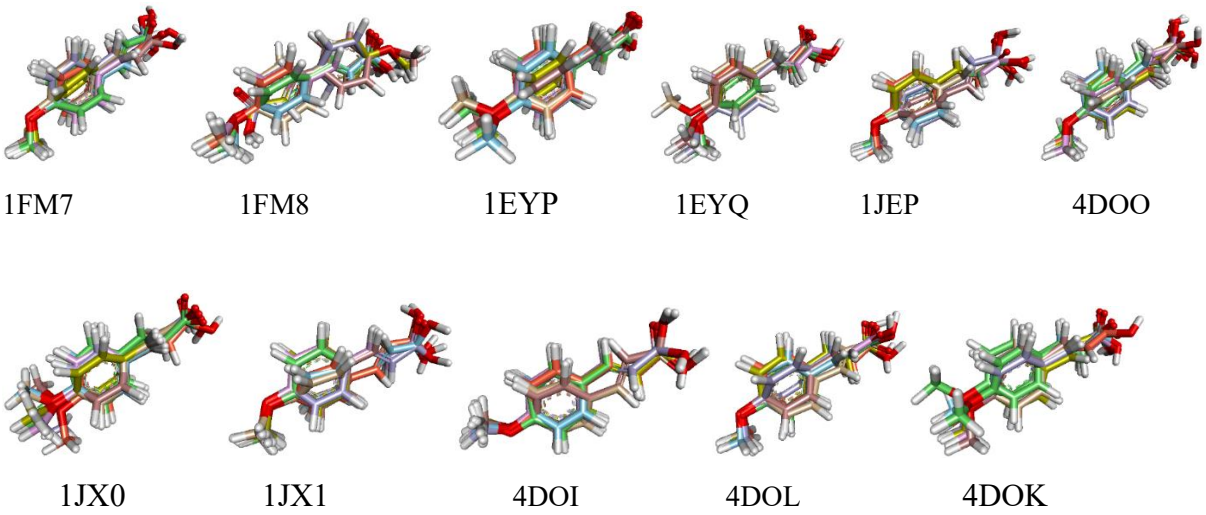


Figure 4.12. Best docked poses of *p*-methoxycinnamic acid within the chalcone isomerase enzyme.

Table 4.14: Molecular docking binding affinities and the binding interactions of *p*-methoxycinnamic acid with chalcone isomerase enzymes crystal structures from PDB.

CHI crystal structure (PDB code)	Binding Affinity with best docked conformation (kcal/mol)	Types of interaction between receptor and ligand	Residue interacting with ligand and interaction distance
1FM7	-6.7	Pi-Alkyl	LEU38 5.42 Å LYS109 4.82 Å LEU101 4.58 Å
		Carbon	GLU105 2.74 Å
		Conventional	THR48 1.80 Å
1FM8	-6.5	Pi-Pi T-shaped	PHE47 5.08 Å
		Donor-Donor	THR48 1.82 Å
		Pi-Alkyl	LYS109 5.09 Å LEU101 4.98 Å LEU38 4.39 Å
		Carbon	GLY37 2.75 Å
1EYP	-6.3	Pi-Pi T-shaped	PHE47 4.74 Å
		Pi-Alkyl	LEU101 4.23 Å LEU38 5.14 Å
		Conventional	THR48 1.77 Å
1EYQ	-6.6	Pi-Alkyl	LEU38 5.30 Å LEU101 5.41 Å LYS109 4.72 Å
1JEP	-6.5	Pi-Alkyl	LEU38 5.32 Å LEU101 5.48 Å LYS109 4.79 Å
4DOO	-6.7	Pi-Pi Stacked	PHE134 4.36 Å
		Carbon	LYS139 2.74 Å
		Pi-Alkyl	LEU140 4.92 Å ILE146 5.42 Å

			VAL103 5.35 Å
1JX0	-6.5	Pi-Pi T-shaped	PHE106 5.15 Å
		Acceptor-Acceptor	ASN113 2.97 Å
		Pi-Alkyl	ILE50 5.01 Å MET191 5.37 Å VAL110 4.74 Å
		Conventional	ARG36 2.43 Å ASN113 2.39 Å ARG36 2.19 Å
1JX1	-6.4	Carbon	TYR106 2.67 Å
		Pi-Alkyl	LYS109 5.23 Å ARG36 4.85 Å
		Conventional	ARG108 2.34 Å ARG108 2.17 Å
4DOI	-5.6	Pi-Alkyl	VAL16 4.91 Å
		Carbon	LEU39 2.82 Å GLU232 2.65 Å
		Conventional	ASP235 3.15 Å GLY67 2.90 Å PHE40 2.73 Å PRO38 2.23 Å
4DOL	-6.0	Pi-Alkyl	ILE102 5.46 Å
		Conventional	ARG44 2.90 Å
4DOK	-3.9	Carbon	ILE17 2.73 Å
		Pi-Alkyl	PRO14 4.37 Å
		Conventional	THR3 1.96 Å GLU4 1.90 Å

Table 4.15: XYZ coordinates of PDB ligand and docked *p*-methoxycinnamic acid within chalcone isomerase enzyme crystallographic structures.

CHI crystal structure (PDB code)	**Coordinates of the PDB crystal structure ligand and the docked p-methoxicinnamic acid ligand**					
	Ligand coordinates from PDB			**p-methoxycinnamic acid docked coordinates**		
	x	**y**	**z**	**x**	**y**	**z**
1FM7	26.4094	21.7857	32.9777	27,7353	23,7174	30,6356
1FM8	27.2413	22.1962	32.9759	29,3419	24,4498	29,3423
1EYP	27.2701	21.7562	33.2601	29,425	25,1147	28,6356
1EYQ	26.8902	22.328	33.4699	26,699	23,1932	31,4006
1JEP	27.534	21.5262	32.7994	27,2842	23,3268	31,5131
4DOO	26.4094	21.7857	32.9777	-5,48369	7,58817	-0,919313
1JX0	27.2413	22.1962	32.9759	25,1822	19,6852	34,2833
1JX1	27.2701	21.7562	33.2601	50,0676	39,2305	60,8118
4DOI	27.2413	22.1962	32.9759	24,0694	22,4209	40,9101
4DOL	-2.20893	2.51464	-2.86768	19,8454	-0,428852	-16,2079
4DOK	1.57793	0.132286	-1.99757	20,2153	13,5883	25,3114

Table 4.16: Quantitative properties of Tyr106 related to its activity in the active site of CHI.

CHI crystal structure (PDB code)	O(10)···HO Tyr106 Hydrogen bond (Å)	Tyr106 hydrophobicity	Tyr106 pKa	Tyr106 Residue solvent accessibility	Tyr106 % solvent accessibility
1FM7	5.068	-1.3	10	25.76	11.732
1FM8	8.088	-1.3	10	24.384	11.106
1EYP	5.079	-1.3	10	24.159	11.003
1EYQ	5.711	-1.3	10	25.147	11.453
1JEP	5.752	-1.3	10	25.65	11.682
4DOO	N/A	N/A	N/A	N/A	N/A
1JX0	N/A	N/A	N/A	N/A	N/A
1JX1	10.646	-1.3	10	39.095	17.805
4DOI	N/A	N/A	N/A	N/A	N/A
4DOL	N/A	N/A	N/A	N/A	N/A
4DOK	N/A	N/A	N/A	N/A	N/A

***p*-Nitrocinnamic acid docked to chalcone isomerase enzymes.**

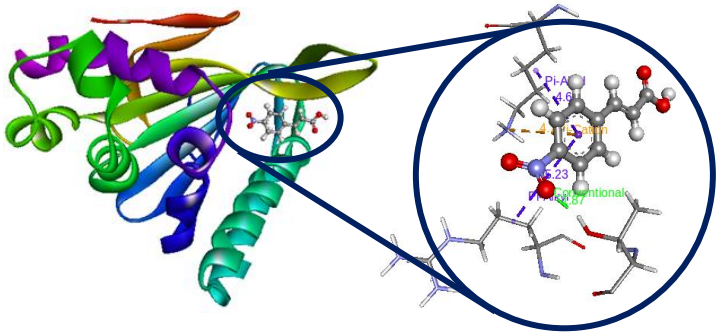


Figure 4.13. 1FM8 with docked *p*-nitrocinnamic acid (left) and the pocket site (right) showing receptor-ligand interaction.

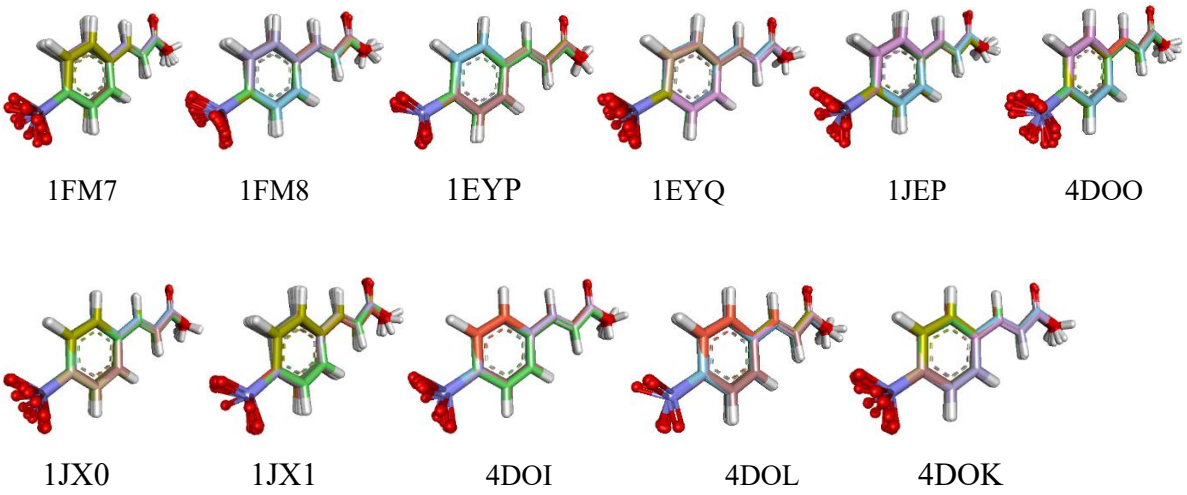


Figure 4.14. Best docked poses of *p*-nitrocinnamic acid within the chalcone isomerase enzyme.

Table 4.17: Molecular docking binding affinities and the binding interactions of *p*-nitrocinnamic acid with chalcone isomerase enzymes crystal structures from PDB.

CHI crystal structure (PDB code)	Binding Affinity with best docked conformation (kcal/mol)	Types of interaction between receptor and ligand	Residue interacting with ligand and interaction distance
1FM7	-7.2	Pi-Cation	LYS109 4.88 Å
		Pi-Alkyl	LEU38 5.45 Å LEU101 5.21 Å LEU109 4.57 Å
		Conventional	THR48 2.31 Å LYS109 2.24 Å
1FM8	-7.0	Pi-Cation	LYS109 4.72 Å
		Pi-Alkyl	ARG36 5.23 Å LYS109 4.62 Å
		Conventional	THR48 1.87 Å
1EYP	-6.7	Carbon	GLY37 2.85 Å
		Pi-Alkyl	LEU101 4.24 Å LEU38 5.80 Å
		Conventional	THR48 2.50 Å
1EYQ	-6.9	Pi-Alkyl	LEU38 4.96 Å LEU101 4.63 Å LYS109 4.48 Å
		Carbon	GLY37 3.09 Å
1JEP	-6.8	Pi-Alkyl	LEU38 5.14 Å LEU101 4.50 Å LYS109 4.57 Å
		Carbon	GLY37 2.77 Å
4DOO	-7.4	Pi-Pi Stacked	PHE134 4.27 Å
		Pi-Alkyl	ILE146 5.34 Å VAL103 5.34 Å

			ILE140 5.07 Å
1JX0	-7.2	Carbon	VAL110 2.68 Å THR190 2.60 Å
		Pi-Alkyl	ARG36 5.03 Å VAL110 4.32 Å
		Conventional	ASN113 2.67 Å ARG36 2.33 Å THR190 2.13 Å
1JX1	-7.1	Amide-Pi Stacked	LYS109 5.41 Å
		Pi-Alkyl	ARG36 4.64 Å VAL110 4.39 Å
		Carbon	THR190 3.01 Å GLY37 2.82 Å VAL110 2.48 Å
		Conventional	THR190 2.25Å ASN113 2.17Å
4DOI	-6.0	Pi-Alkyl	VAL233 5.40 Å
		Carbon	LEU39 2.88 Å
		Conventional	GLY67 2.96 Å PHE40 2.69 Å VAL16 2.41 Å
4DOL	-6.3	Pi-Pi T-shaped	TRY56 4.91 Å
		Carbon	ARG44 3.08 Å
		Pi-Alkyl	ALA116 3.95 Å
		Conventional	ARG44 2.90 Å
4DOK	-4.0	Carbon	ILE17 2.80 Å
		Pi-Alkyl	PRO15 5.39 Å
		Conventional	ELE18 2.42 Å

Table 4.18: XYZ coordinates of PDB ligand and docked *p*-nitrocinnamic acid within chalcone isomerase enzyme crystallographic structures.

CHI crystal structure (PDB code)	Coordinates of the PDB crystal structure ligand and the docked *p*-nitrocinnamic acid ligand					
	Ligand coordinates from PDB			*p*-nitrocinnamic acid docked coordinates		
	x	y	z	x	y	z
1FM7	26.4094	21.7857	32.9777	28.5319	24.9928	30.0486
1FM8	27.2413	22.1962	32.9759	27.8177	23.5213	31
1EYP	27.2701	21.7562	33.2601	29.37	24.9441	29.124
1EYQ	26.8902	22.328	33.4699	28.4761	24.6435	28.9349
1JEP	27.534	21.5262	32.7994	29.1155	24.6333	28.9277
4DOO	26.4094	21.7857	32.9777	-5.78675	7.38627	-1.54446
1JX0	27.2413	22.1962	32.9759	27.6711	21.0923	33.4315
1JX1	27.2701	21.7562	33.2601	49.9961	41.9933	62.546
4DOI	27.2413	22.1962	32.9759	24.5146	22.6352	40.9211
4DOL	-2.20893	2.51464	-2.86768	20.5013	-0.299601	-16.1974
4DOK	1.57793	0.132286	-1.99757	17.8363	18.5923	25.8942

Table 4.19: Quantitative properties of Tyr106 related to its activity in the active site of CHI.

CHI crystal structure (PDB code)	O(10)···HO Tyr106 Hydrogen bond (Å)	Tyr106 Hydrophobicity	Tyr106 pKa	Tyr106 Residue solvent accessibility	Tyr106 % solvent accessibility
1FM7	12.503	-1.3	10	25.76	11.732
1FM8	10.944	-1.3	10	24.384	11.106
1EYP	6.846	-1.3	10	24.159	11.003
1EYQ	12.740	-1.3	10	25.147	11.453
1JEP	12.684	-1.3	10	25.65	11.682
4DOO	N/A	N/A	N/A	N/A	N/A
1JX0	N/A	N/A	N/A	N/A	N/A
1JX1	7.661	-1.3	10	39.095	17.805
4DOI	N/A	N/A	N/A	N/A	N/A
4DOL	N/A	N/A	N/A	N/A	N/A
4DOK	N/A	N/A	N/A	N/A	N/A

4.4.5 Summary of the chapter

1FM7 complexes with the sets of cinnamic acid derivatives exhibited the best binding energies of all CHIs considered in this study. 1FM8 complexes had the second best binding affinity (with the difference of 0.1 kcal/mol from the best binding energy for all coumaric acid derivatives), but also had more conventional hydrogen bonds that could facilitate the direct protonation and deprotonation of *o*-coumaric acid. The CHIs from *A.Thaliana* had the worse binding affinities compared to the *M.Sativa* CHIs. From these results 1FM8 would seem to be an ideal choice for further computational studies, however, a more in-depth understanding of how the CHI interacts with other potential substrates is required, such information was provided by a more detailed screening method - High Throughput Virtual Screening (Chapter 5).

S-*o*-coumaric acid had the best binding affinities to CHIs and the best orientation within the CHI favouring direct protonation and deprotonation with protein residues as compared to the A-*o*-coumaric acid. The O(10)···HO Tyr106 distance were closer for the 1FM8-S-*o*-coumaric

acid complex (3.69 Å) than for the 1FM8-A-*o*-coumaric acid complex (5.63 Å). This shows the ease of direct protonation of S-*o*-coumaric acid complex with Tyr106 of 1FM8. The 1FM8-S-*o*-coumaric acid complex also has a conventional hydrogen bond with a distance 2.56 Å between Gly37 oxygen and the phenol OH of S-*o*-coumaric acid which would allow direct deprotonation of the ligand by protein Gly37. All of these observations lead to S-*o*-coumaric acid being judged the most suitable substrate for the further, ONIOM study on its reactivity within the active site of the 1FM8-enzyme. (Chapter 6)

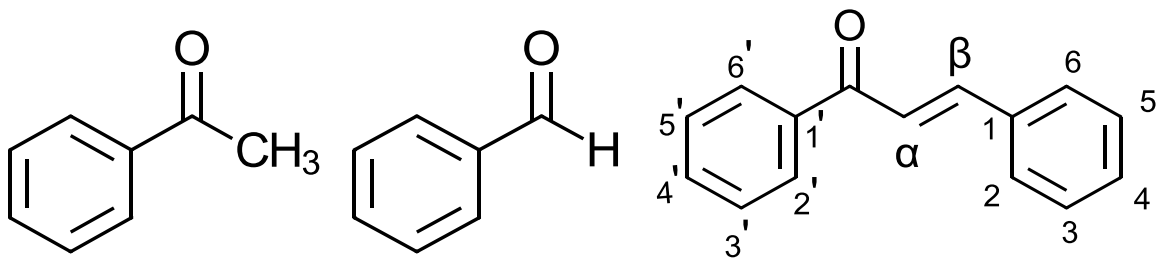
CHAPTER 5
HIGH THROUGHPUT VIRTUAL SCREENING

5.1 Introduction to the chapter

The basis of all bioorganic processes in living organisms is molecular recognition – a process in which biological macromolecules interact with each other or with various other small molecules (228). Protein-ligand recognition plays a pivotal role in most biological processes determining specificity and affinity (229,230). Protein-ligand complexes with lower (i.e. more negative) free-energies of binding (ΔG) result in the formation of stronger complexes (i.e., when a ligand has a low free binding energy (ΔG) its affinity towards the target enzyme is greater than that of a ligand with higher ΔG) (186,229). Protein-ligand binding affinity provides information that can assist in understanding the activities of enzymes at the molecular level. Chapter 4 focused on the analysis of results from docking of cinnamic acid derivatives into chalcone isomerase enzymes (CHI). Docking cinnamic acid derivatives into the active site of each of the CHI enzymes from *M. sativa* was successful. However, the same ligands failed to dock into the active sites of CHIs from *A. thaliana*. High Throughput Virtual Screening (HTVS) is a computational approach which explores at the computational level the properties of libraries of small molecules. In many cases where the property is protein binding, molecular docking is used to follow and identify trends in the binding of receptors of interest. This chapter is concerned with the molecular docking of 8 064 chalcone derivatives to CHI enzymes from *M.sativa* and *A. thaliana* (the 11 CHI crystal structures from the PDB that were considered in docking cinnamic acid derivatives). HTVS of chalcone derivatives to CHI crystal structures is undertaken to facilitate determination of the selectivity of the CHI active sites and to establish if the failure of cinnamic acid derivatives to dock into the active sites of CHI crystal structures from *A. Thaliana* is due to the high selectivity of the active sites or improper description of the active sites.

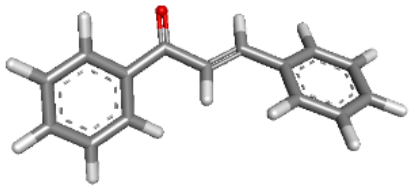
5.2 Receptors and Ligands – objects of HTVS study

The HTVS study utilised the 11 wild and mutated CHI crystal structures from the Protein Data Bank (PDB). These are the CHIs with PDB code: 1FM7, 1FM8, 1JEP, 1EYQ, 1EYP, 4DOI, 4DOK, 4DOL, 4DOO, 1JXO (Y106F mutant), and 1JX1 (T48A mutant). The full names, description and origin of these CHI crystal structures are given in Table 4.1 Chapter 4. Chain A was selected for all crystal structures (of the two chains present). The systematic outline of the preparation of the proteins is outlined in Chapter 2 Section 2.2.2. Chapter 2 Section 2.2.3 shows in a more detailed manner the specifications for the library of chalcone derivatives. Figure 5.1 shows acetophenone, benzaldehyde, and the simple chalcone that may be formed from an Aldol condensation, and the numbering utilized in this study. Figure 5.2 gives some examples of the chalcone derivatives included in this library, together with the label encoding the particular substituted acetophenone and benzaldehyde precursors for the library member. This particular figure only contains the diagonal through the library matrix, i.e. 84 structures (out of a total of 8064 chalcone structures examined) in which the numerical index of the benzaldehyde component matches the index of the acetophenone component.

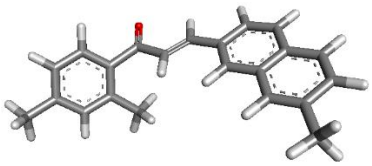


a) b) c)

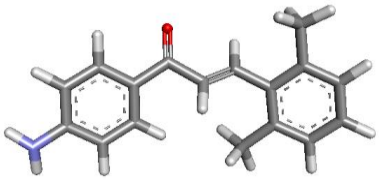
Figure 5.1. Example of chalcone derivatives optimized with AM1 in *vacuo*. a) acetophenone, b) benzaldehyde, c) chalcone



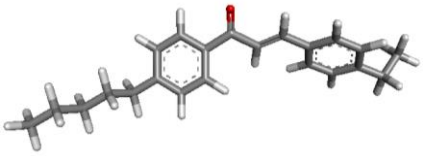
acetoph_1_benzald_1



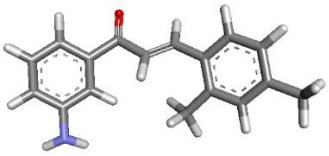
acetoph_2_benzald_2



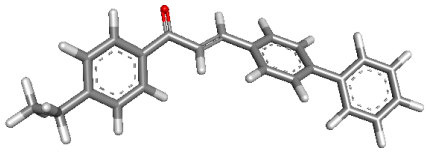
acetoph_3_benzald_3



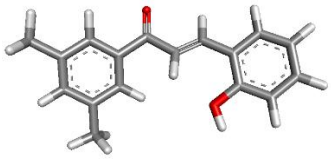
acetoph_4_benzald_4



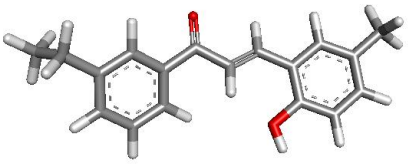
acetoph_5_benzald_5



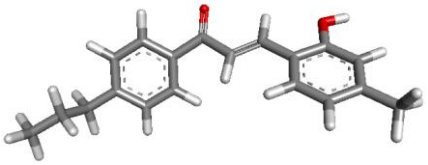
acetoph_6_benzald_6



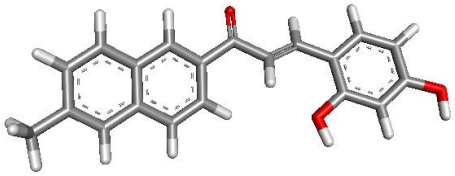
acetoph_7_benzald_7



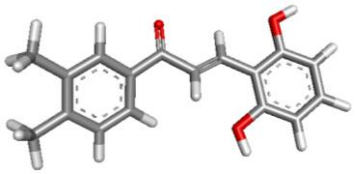
acetoph_8_benzald_8



acetoph_9_benzald_9

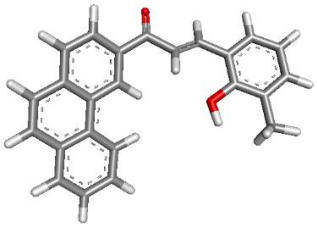


acetoph_10_benzald_10

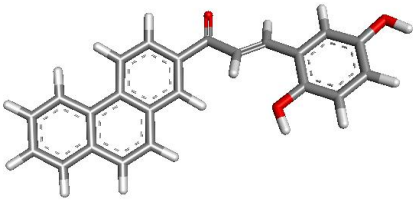


acetoph_11_benzald_11

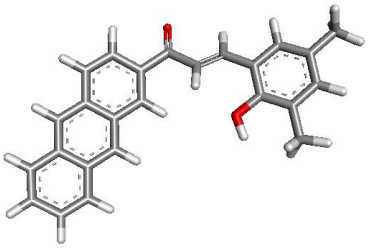
acetoph_12_benzald_12



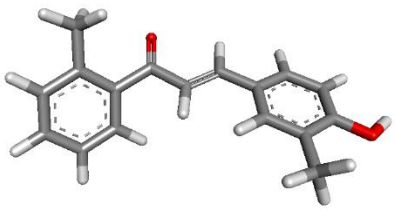
acetoph_13_benzald_13



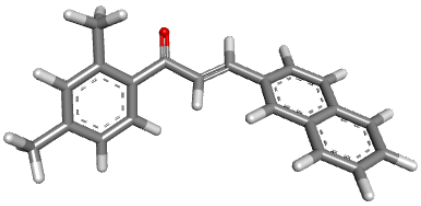
acetoph_14_benzald_14



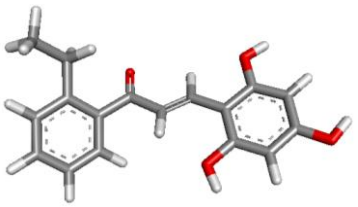
acetoph_15_benzald_15



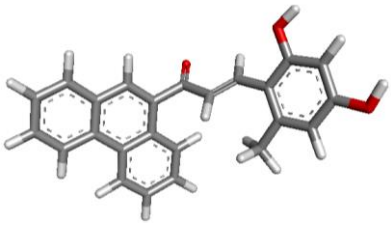
acetoph_16_benzald_16



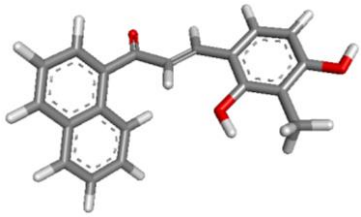
acetoph_17_benzald_17



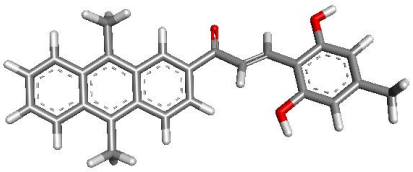
acetoph_18_benzald_18



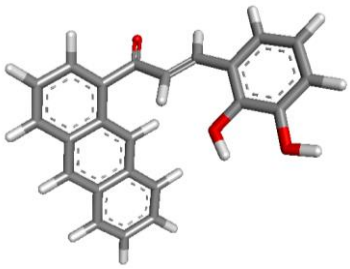
acetoph_19_benzald_19



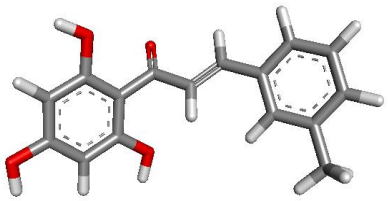
acetoph_20_benzald_20



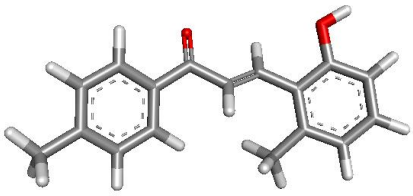
acetoph_21_benzald_21



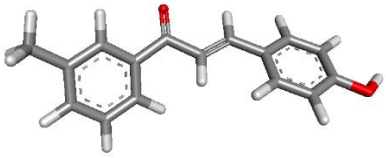
acetoph_22_benzald_22



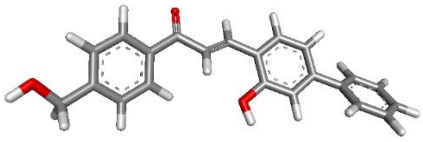
acetoph_23_benzald_23



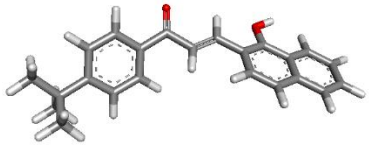
acetoph_24_benzald_24



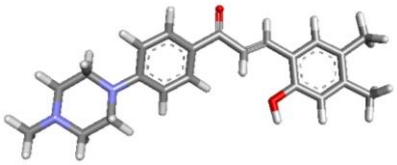
acetoph_25_benzald_25



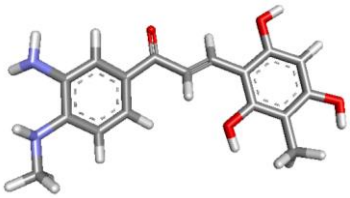
acetoph_26_benzald_26



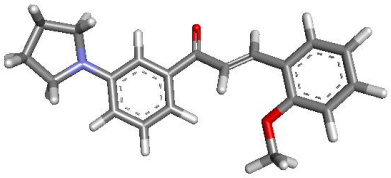
acetoph_27_benzald_27



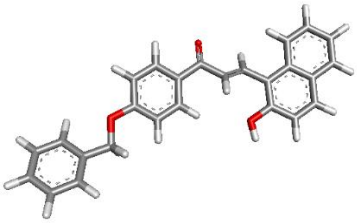
acetoph_28_benzald_28



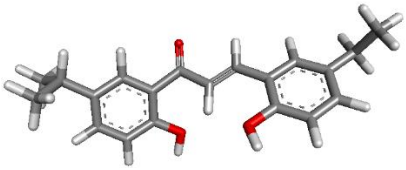
acetoph_29_benzald_29



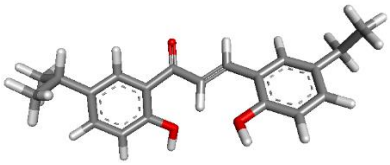
acetoph_30_benzald_30



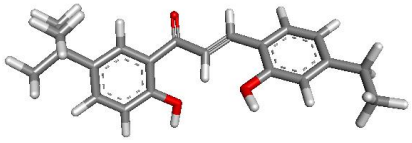
acetoph_31_benzald_31



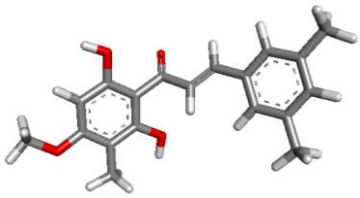
acetoph_31_benzald_31



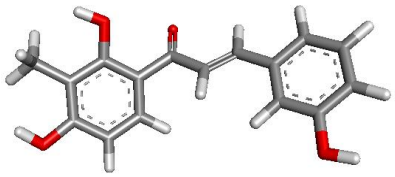
acetoph_32_benzald_32



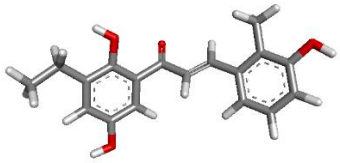
acetoph_33_benzald_33



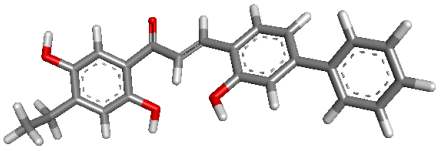
acetoph_34_benzald_34



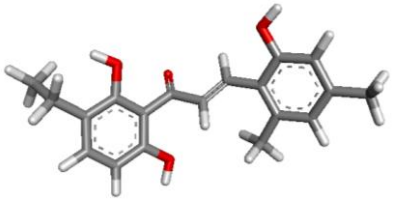
acetoph_35_benzald_35



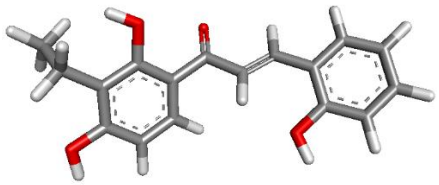
acetoph_36_benzald_36



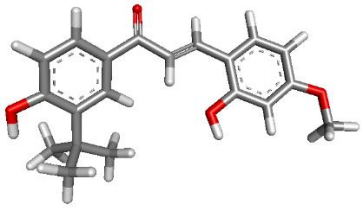
acetoph_37_benzald_37



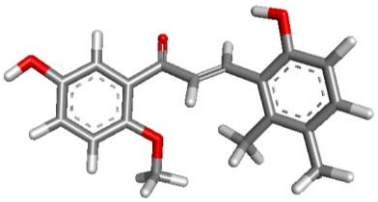
acetoph_38_benzald_38



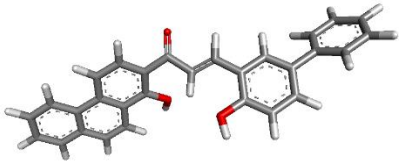
acetoph_39_benzald_39



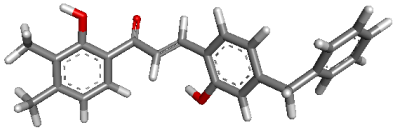
acetoph_40_benzald_40



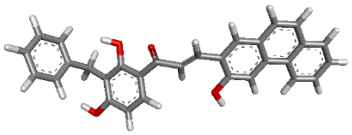
acetoph_41_benzald_41



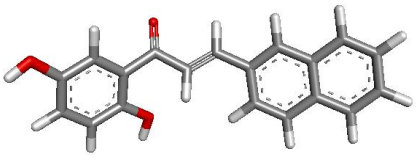
acetoph_42_benzald_42



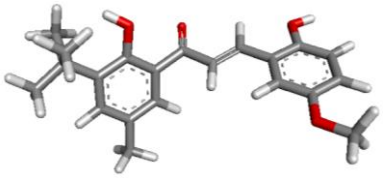
acetoph_43_benzald_43



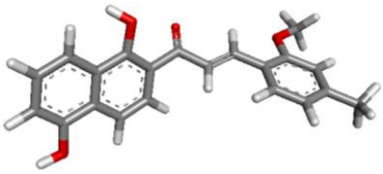
acetoph_44_benzald_44



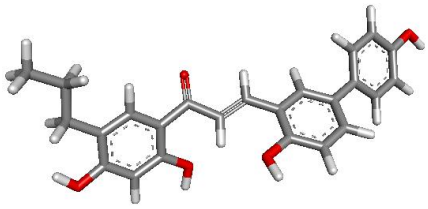
acetoph_45_benzald_45



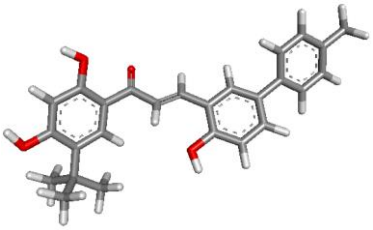
acetoph_46_benzald_46



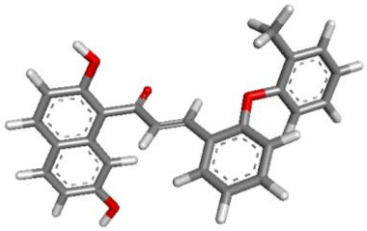
acetoph_47_benzald_47



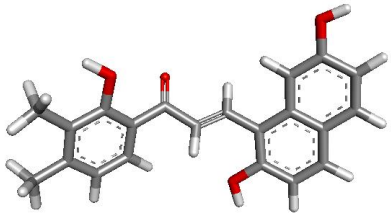
acetoph_48_benzald_48



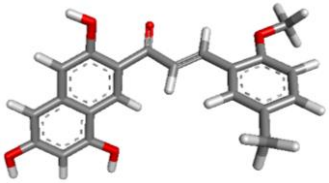
acetoph_49_benzald_49



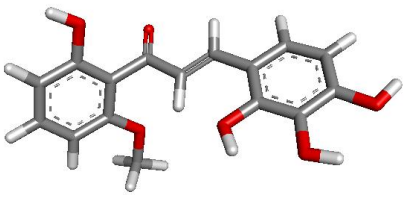
acetoph_50_benzald_50



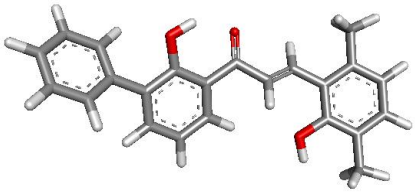
acetoph_51_benzald_51



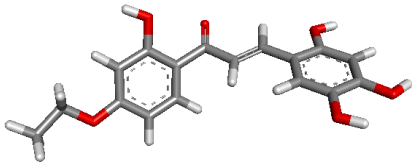
acetoph_52_benzald_52



acetoph_53_benzald_53

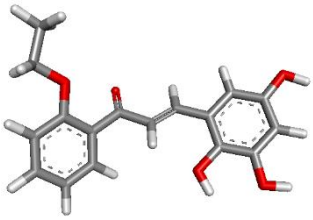


acetoph_54_benzald_54

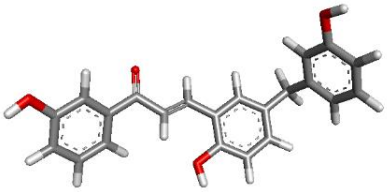


acetoph_55_benzald_55

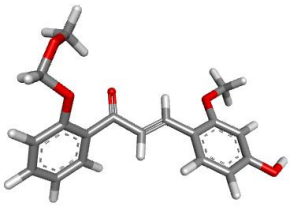
acetoph_56_benzald_56



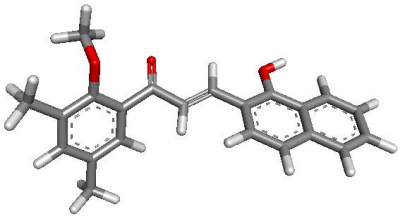
acetoph_57_benzald_57



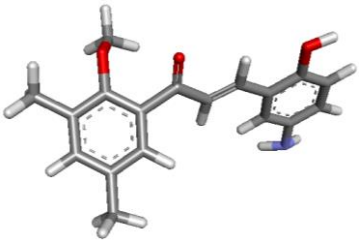
acetoph_58_benzald_58



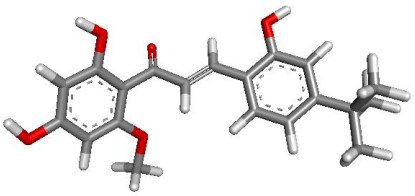
acetoph_59_benzald_59



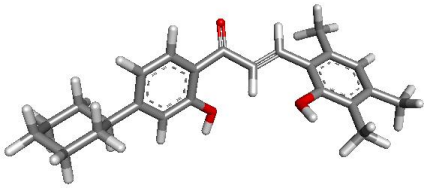
acetoph_60_benzald_60



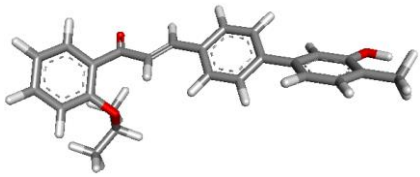
acetoph_61_benzald_61



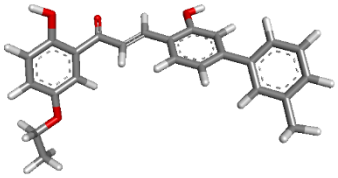
acetoph_62_benzald_62



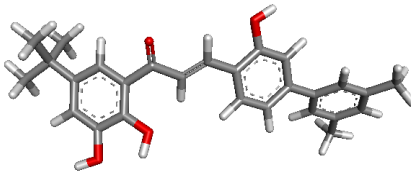
acetoph_63_benzald_63



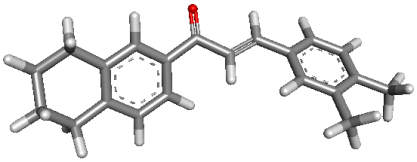
acetoph_64_benzald_64



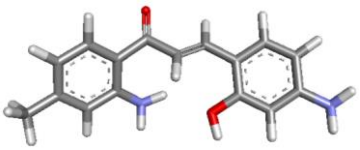
acetoph_65_benzald_65



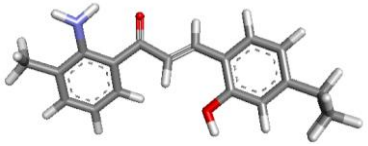
acetoph_66_benzald_66



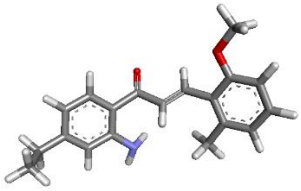
acetoph_67_benzald_67



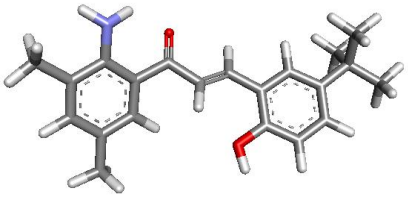
acetoph_68_benzald_68



acetoph_69_benzald_69



acetoph_70_benzald_70



acetoph_71_benzald_71

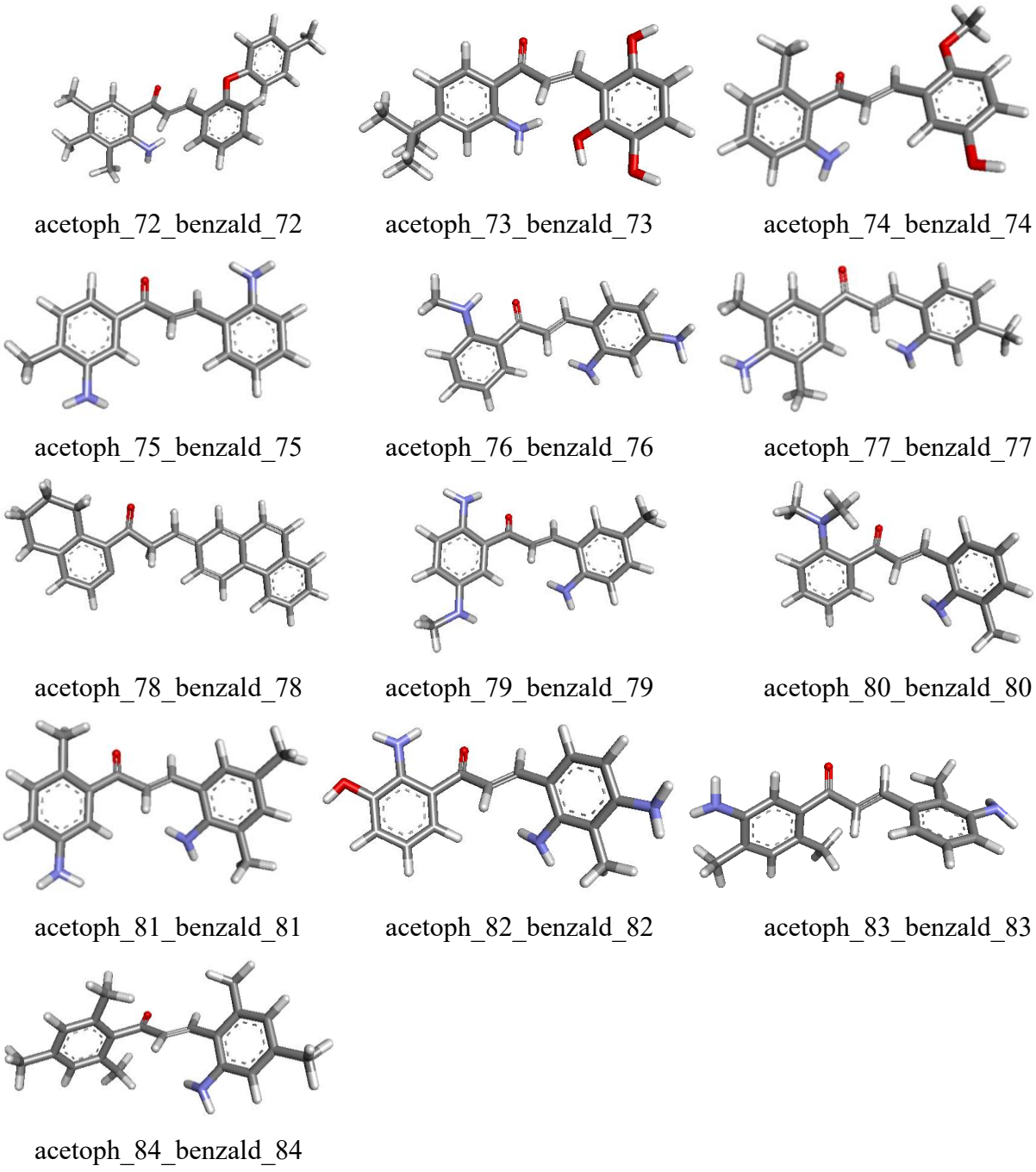


Figure 5.2. Examples of chalcone derivatives optimized with AM1 in *vacuo*. acetoph is an abbreviation of acetophenone and the number after the word acetoph denotes that particular acetophenone moiety. benzald is an abbreviation of benzaldehyde, and the number after the word benzal denotes that particular benzaldehyde moiety. There are 84 acetophenone moieties and 96 benzaldehyde moieties, this figure shows 84 acetophenone and 84 benzaldehyde moieties. For clarity only elements of the library where the numerical index of the acetophenone matches the numerical index of the benzaldehyde are shown.

Because this chapter serves to validate the results obtained in Chapter 4, HTVS was undertaken using the same computational methods as the utilised in docking of cinnamic acid derivatives to CHIs. The same optimization and Empirical Free Energy Scoring Functions were used with Autodock Vina. Selected results were visualized using Discovery Studio Visualizer.

5.3 Results

The results discussed in this section are from the HTVS of 8 064 chalcone derivatives against all 11 CHI enzymes. Due to the bulkiness of the resulting data the different aspects of the analysis are presented in different sections to simplify the discussion. The first section of the results concerns the discussion of the binding affinities, the second section with the analysis of the ligand binding sites, the third on the analysis of the receptor ligand interactions and the final section with the analysis of the interaction of the docked ligands with the TYR106 residue.

5.3.1 Binding affinities

Figure 5.3 shows a heat map of average binding affinities for chalcone derivatives docked to the 11 crystallographic CHI structures. Each value represents the average binding affinity for a set of 96 chalcones (constructed from a particular acetophenone and each of the 96 benzaldeydes) with each of the specified CHI structures. The heatmap showing the individual binding affinities for all ligands and all receptors is given in the Appendix F (Supplementary material on the disc). The y-axis lists the acetophenone moieties (their structural examples are given in Figure 5.2), and the x-axis shows the PDB code of the CHI crystallographic structure considered. Figure 5.4 shows a corresponding heat map of the average CHI binding affinities for sets of chalcones, in which a specified benzaldeyde is combined with each of the 84 acetophenones. The y-axis lists the benzaldehyde moieties (most of which are illustrated in Figure 5.2), and the x-axis again shows the PDB code of the CHI crystallographic structure considered.

	Chalcone isomerase crystal structure PDB Code:										
	1EYP	1EYQ	1FM7	1FM8	1JEP	1JX0	1JX1	4DOI	4DOK	4DOL	4DOO
Acetophenone_1	-8,4	-8,4	-9,1	-9,0	-8,8	-9,6	-9,0	-7,0	-6,6	-8,3	-9,5
Acetophenone_2	-8,7	-8,7	-9,2	-9,2	-8,6	-10,0	-9,1	-6,3	-6,8	-8,2	-9,7
Acetophenone_3	-8,4	-8,4	-9,1	-9,0	-8,8	-9,4	-8,8	-6,3	-6,6	-7,8	-9,3
Acetophenone_4	-8,4	-8,4	-9,0	-8,9	-8,7	-9,5	-8,2	-6,7	-6,5	-7,8	-9,7
Acetophenone_5	-8,4	-8,4	-9,1	-9,0	-8,7	-9,4	-9,0	-6,8	-6,8	-8,0	-9,4
Acetophenone_6	-8,6	-8,6	-9,3	-9,2	-9,0	-10,0	-8,9	-6,4	-6,7	-8,1	-9,8
Acetophenone_7	-8,7	-8,7	-9,0	-8,9	-8,4	-9,6	-9,2	-6,7	-7,0	-8,3	-10,2
Acetophenone_8	-8,8	-8,8	-9,3	-9,1	-8,7	-9,8	-9,5	-6,7	-6,8	-8,4	-9,9
Acetophenone_9	-8,6	-8,6	-9,2	-9,0	-8,9	-9,8	-8,6	-6,4	-6,8	-7,9	-9,7
Acetophenone_10	-10,1	-10,1	-10,5	-10,2	-9,9	-10,9	-10,1	-7,1	-7,7	-8,4	-10,9
Acetophenone_11	-9,1	-9,1	-9,5	-9,3	-8,9	-10,0	-9,3	-6,7	-7,0	-8,3	-10,2
Acetophenone_12	-8,0	-8,1	-8,5	-8,2	-7,9	-8,8	-8,7	-6,7	-6,3	-7,8	-8,6
Acetophenone_13	-10,1	-10,1	-9,4	-9,7	-10,0	-11,4	-9,1	-7,6	-8,2	-9,1	-12,1
Acetophenone_14	-10,0	-10,0	-10,3	-10,3	-9,9	-10,7	-9,2	-7,7	-8,3	-8,3	-11,6
Acetophenone_15	-10,8	-10,8	-11,0	-10,6	-10,6	-11,5	-9,4	-7,7	-8,3	-8,4	-11,8
Acetophenone_16	-8,4	-8,4	-9,0	-8,6	-8,3	-9,6	-9,2	-6,7	-6,6	-8,4	-9,5
Acetophenone_17	-8,7	-8,7	-9,2	-9,0	-8,6	-10,0	-9,1	-6,3	-6,8	-8,2	-9,7
Acetophenone_18	-8,2	-8,2	-8,3	-7,9	-8,0	-9,4	-9,1	-6,5	-6,6	-8,3	-9,4
Acetophenone_19	-10,4	-10,4	-8,9	-9,1	-10,0	-11,1	-10,9	-7,4	-8,2	-9,4	-12,0
Acetophenone_20	-9,4	-9,4	-8,8	-8,7	-9,0	-10,4	-9,8	-6,7	-7,3	-9,0	-10,6
Acetophenone_21	-10,2	-10,2	-9,1	-9,7	-10,4	-10,8	-8,6	-7,7	-8,6	-8,7	-12,4
Acetophenone_22	-10,2	-10,2	-9,1	-9,1	-9,8	-11,1	-10,0	-7,3	-7,8	-9,6	-11,8
Acetophenone_23	-8,1	-8,1	-8,7	-8,3	-8,1	-8,6	-8,4	-6,4	-6,6	-7,4	-8,5
Acetophenone_24	-8,7	-8,7	-9,4	-9,2	-9,0	-9,9	-9,0	-6,4	-6,8	-8,2	-9,8
Acetophenone_25	-8,8	-8,8	-9,4	-9,1	-8,8	-9,8	-9,5	-6,7	-6,8	-8,3	-9,9
Acetophenone_26	-8,4	-8,4	-9,1	-8,9	-8,8	-9,6	-8,7	-6,4	-6,7	-7,9	-9,4
Acetophenone_27	-9,2	-9,2	-9,7	-9,3	-8,9	-9,7	-8,6	-6,5	-7,1	-7,8	-10,0
Acetophenone_28	-9,0	-9,0	-9,6	-9,5	-9,2	-10,0	-7,2	-6,6	-7,3	-6,7	-10,0
Acetophenone_29	-8,5	-8,5	-9,0	-8,8	-8,6	-9,1	-8,9	-6,5	-6,8	-7,6	-9,3
Acetophenone_30	-8,9	-8,9	-8,9	-8,8	-9,0	-9,8	-8,7	-6,9	-7,1	-8,5	-10,1
Acetophenone_31	-9,3	-9,3	-9,9	-9,8	-9,8	-10,0	-8,1	-7,2	-7,6	-8,2	-10,9
Acetophenone_32	-8,6	-8,6	-8,4	-8,3	-8,5	-10,1	-9,3	-6,6	-7,2	-8,4	-9,8
Acetophenone_33	-8,6	-8,6	-8,3	-8,3	-8,4	-9,8	-8,7	-6,6	-7,0	-8,2	-9,9
Acetophenone_34	-8,3	-8,3	-8,6	-7,9	-8,2	-8,5	-8,1	-6,3	-6,6	-7,0	-8,6
Acetophenone_35	-8,4	-8,4	-9,4	-9,0	-8,4	-9,5	-9,2	-6,6	-6,8	-7,8	-9,4
Acetophenone_36	-8,4	-8,4	-8,6	-8,6	-8,0	-9,2	-8,9	-6,6	-6,8	-7,9	-9,1
Acetophenone_37	-8,6	-8,6	-9,0	-8,8	-8,3	-9,5	-8,7	-6,7	-6,8	-7,6	-9,2
Acetophenone_38	-8,4	-8,4	-8,6	-8,2	-8,1	-9,5	-8,7	-6,3	-6,6	-7,9	-8,9
Acetophenone_39	-8,2	-8,2	-8,8	-8,3	-8,0	-9,1	-8,8	-6,4	-6,9	-7,9	-9,1
Acetophenone_40	-8,8	-8,8	-8,5	-8,7	-8,7	-9,0	-8,6	-6,6	-7,1	-8,3	-10,0
Acetophenone_41	-8,1	-8,1	-8,5	-8,1	-7,9	-9,2	-8,7	-6,1	-6,6	-7,7	-9,0
Acetophenone_42	-9,9	-9,9	-9,2	-9,0	-9,7	-10,4	-9,0	-7,7	-8,2	-8,1	-11,6
Acetophenone_43	-8,7	-8,7	-9,0	-8,7	-8,5	-9,7	-9,0	-6,6	-7,1	-8,1	-9,5

Acetophenone_44	-9,0	-9,0	-9,2	-9,1	-9,2	-9,3	-8,2	-6,8	-7,1	-8,4	-10,5
Acetophenone_45	-8,2	-8,2	-8,9	-8,7	-8,3	-9,2	-8,8	-6,6	-6,6	-8,0	-9,1
Acetophenone_46	-8,7	-8,7	-8,5	-8,5	-8,6	-9,2	-8,5	-6,8	-7,2	-8,1	-10,0
Acetophenone_47	-9,1	-9,1	-9,2	-9,0	-8,9	-9,6	-9,0	-7,1	-7,6	-8,0	-10,1
Acetophenone_48	-8,4	-8,4	-8,8	-8,2	-8,3	-9,1	-8,9	-6,4	-7,0	-7,9	-9,7
Acetophenone_49	-8,6	-8,6	-8,1	-8,2	-8,5	-9,0	-8,4	-6,4	-7,1	-7,8	-9,8
Acetophenone_50	-9,0	-9,0	-8,1	-8,1	-8,8	-9,8	-9,4	-6,4	-7,0	-8,6	-10,0
Acetophenone_51	-8,7	-8,7	-9,0	-8,7	-8,5	-9,7	-9,0	-6,6	-7,1	-8,1	-9,5
Acetophenone_52	-9,5	-9,5	-8,7	-8,6	-9,1	-9,7	-9,1	-6,7	-7,5	-7,9	-10,0
Acetophenone_53	-8,0	-8,0	-7,9	-7,5	-7,7	-8,6	-8,3	-6,3	-6,2	-7,9	-8,5
Acetophenone_54	-9,3	-9,3	-9,3	-9,2	-9,4	-10,2	-8,9	-7,0	-7,4	-8,2	-10,7
Acetophenone_55	-8,2	-8,2	-8,6	-8,4	-8,2	-9,3	-8,4	-6,5	-6,8	-7,5	-9,1
Acetophenone_56	-8,5	-8,5	-8,3	-8,6	-8,6	-9,5	-8,4	-6,3	-6,6	-9,1	-10,1
Acetophenone_57	-7,9	-7,9	-7,9	-7,6	-7,8	-8,9	-8,7	-6,3	-6,5	-8,0	-9,1
Acetophenone_58	-8,5	-8,5	-9,2	-9,1	-8,6	-9,4	-9,0	-6,8	-6,7	-8,2	-9,4
Acetophenone_59	-7,7	-7,9	-8,1	-7,7	-7,8	-8,9	-8,7	-6,0	-6,4	-7,9	-8,9
Acetophenone_60	-8,3	-8,3	-8,0	-8,0	-8,0	-9,3	-8,9	-6,3	-7,0	-8,0	-9,5
Acetophenone_61	-8,4	-8,4	-8,0	-8,0	-8,0	-9,3	-8,9	-6,3	-7,0	-7,9	-9,5
Acetophenone_62	-8,1	-8,1	-7,9	-7,6	-7,9	-8,5	-8,2	-6,1	-6,4	-7,3	-8,5
Acetophenone_63	-9,5	-9,5	-10,0	-9,6	-9,7	-10,2	-8,9	-7,1	-7,6	-7,7	-10,6
Acetophenone_64	-8,1	-8,1	-8,1	-7,8	-8,1	-9,2	-8,9	-6,0	-6,7	-8,4	-9,3
Acetophenone_65	-8,1	-8,1	-8,9	-8,8	-8,2	-9,3	-8,7	-6,4	-6,6	-8,1	-9,1
Acetophenone_66	-8,7	-8,7	-8,2	-8,3	-8,6	-9,5	-8,1	-6,4	-7,1	-7,8	-9,9
Acetophenone_67	-9,7	-9,7	-10,0	-9,8	-9,3	-10,5	-9,6	-7,0	-7,5	-8,8	-10,7
Acetophenone_68	-8,5	-8,5	-9,0	-8,9	-8,6	-9,6	-9,0	-6,5	-6,9	-8,0	-9,3
Acetophenone_69	-8,4	-8,4	-9,0	-8,7	-8,2	-9,4	-9,0	-6,5	-6,8	-8,0	-9,4
Acetophenone_70	-8,5	-8,5	-9,1	-8,9	-8,5	-9,7	-8,9	-6,5	-6,9	-7,7	-9,4
Acetophenone_71	-8,6	-8,6	-8,8	-8,6	-8,3	-9,5	-9,0	-6,5	-7,0	-8,0	-9,7
Acetophenone_72	-8,8	-8,8	-8,3	-8,3	-8,5	-9,7	-9,2	-6,6	-7,3	-8,1	-9,8
Acetophenone_73	-9,0	-9,0	-9,6	-9,1	-8,7	-9,1	-8,6	-6,6	-7,2	-7,6	-9,8
Acetophenone_74	-8,3	-8,3	-8,2	-7,9	-7,9	-9,2	-8,7	-6,3	-6,6	-7,9	-8,8
Acetophenone_75	-8,7	-8,7	-9,3	-9,1	-8,5	-9,7	-8,9	-6,6	-7,0	-8,2	-9,6
Acetophenone_76	-8,0	-8,0	-8,2	-7,9	-7,8	-9,1	-8,9	-6,4	-6,7	-7,9	-9,1
Acetophenone_77	-8,6	-8,6	-8,9	-8,6	-8,3	-9,2	-9,1	-6,5	-7,0	-8,0	-9,9
Acetophenone_78	-9,2	-9,2	-8,8	-8,7	-8,8	-10,1	-9,4	-6,5	-7,1	-9,1	-10,4
Acetophenone_79	-8,3	-8,3	-8,9	-8,6	-8,2	-9,3	-9,0	-6,4	-6,7	-8,0	-9,2
Acetophenone_80	-7,8	-7,8	-7,7	-7,6	-7,7	-8,7	-8,9	-5,8	-6,6	-7,8	-8,9
Acetophenone_81	-8,4	-8,4	-8,8	-8,5	-8,2	-9,6	-9,1	-6,3	-6,7	-8,1	-9,4
Acetophenone_82	-8,3	-8,3	-8,9	-8,7	-8,2	-9,1	-8,9	-6,7	-6,8	-7,8	-9,3
Acetophenone_83	-8,7	-8,7	-8,9	-8,5	-8,5	-9,6	-9,1	-6,4	-6,9	-7,8	-9,6
Acetophenone_84	-8,4	-8,4	-8,0	-7,8	-8,0	-9,3	-8,4	-6,3	-6,5	-7,5	-9,4
CHI average	-8,7	-8,7	-8,9	-8,7	-8,6	-9,6	-8,9	-6,6	-7,0	-8,1	-9,8

Figure 5.3. Heat map of average ligand binding affinities (kcal/mol), sorted by the acetophenone. Green hits correspond to good binding affinities, and red corresponds to bad binding affinities.

	Chalcone isomerase crystal structure PDB Code:										
	1EYP	1EYQ	1FM7	1FM8	1JEP	1JX0	1JX1	4DOI	4DOK	4DOL	4DOO
Benzaldehyde_1	-8,6	-8,6	-8,7	-8,7	-8,5	-9,4	-8,7	-6,7	-6,6	-8,3	-9,6
Benzaldehyde_2	-9,8	-9,8	-10,1	-9,8	-9,6	-10,7	-9,7	-7,2	-7,5	-8,2	-11,2
Benzaldehyde_3	-8,3	-8,3	-8,1	-7,9	-8,1	-9,5	-8,8	-6,1	-6,6	-8,6	-9,6
Benzaldehyde_4	-8,8	-8,8	-9,0	-8,9	-8,8	-9,9	-9,1	-6,7	-6,8	-7,6	-9,8
Benzaldehyde_5	-8,9	-8,9	-8,9	-8,8	-8,6	-10,0	-9,3	-6,5	-6,9	-8,2	-10,2
Benzaldehyde_6	-9,9	-9,9	-10,0	-9,8	-9,8	-10,7	-9,3	-7,3	-7,4	-8,2	-10,6
Benzaldehyde_7	-8,6	-8,6	-8,7	-8,6	-8,4	-9,3	-9,0	-6,8	-6,8	-8,3	-9,3
Benzaldehyde_8	-8,7	-8,7	-8,7	-8,5	-8,5	-9,5	-8,9	-6,4	-6,8	-8,4	-9,5
Benzaldehyde_9	-8,7	-8,7	-9,0	-8,7	-8,7	-9,8	-9,3	-6,6	-6,8	-7,9	-9,6
Benzaldehyde_10	-8,5	-8,5	-8,8	-8,5	-8,4	-9,1	-8,8	-6,7	-6,9	-7,7	-9,3
Benzaldehyde_11	-7,9	-7,9	-8,5	-8,3	-8,1	-8,9	-8,8	-6,6	-6,4	-8,1	-8,9
Benzaldehyde_12	-8,9	-8,9	-9,1	-9,0	-8,8	-9,9	-9,2	-6,9	-6,7	-7,9	-9,9
Benzaldehyde_13	-8,4	-8,4	-8,8	-8,6	-8,3	-9,4	-9,1	-6,5	-7,1	-8,4	-9,6
Benzaldehyde_14	-8,3	-8,3	-8,7	-8,6	-8,4	-9,1	-8,9	-6,4	-6,7	-8,1	-9,2
Benzaldehyde_15	-8,4	-8,4	-8,4	-8,3	-8,2	-9,5	-8,6	-6,4	-6,9	-8,0	-9,7
Benzaldehyde_16	-8,8	-8,8	-8,7	-8,8	-8,6	-9,6	-8,7	-6,7	-7,0	-8,0	-9,8
Benzaldehyde_17	-9,7	-9,7	-9,8	-9,7	-9,5	-10,8	-9,7	-7,1	-7,4	-8,3	-10,9
Benzaldehyde_18	-7,9	-7,9	-8,6	-8,5	-8,1	-8,7	-8,7	-6,4	-6,7	-7,4	-8,8
Benzaldehyde_19	-8,2	-8,2	-8,7	-8,6	-8,3	-9,1	-9,0	-6,5	-6,9	-7,8	-9,3
Benzaldehyde_20	-8,2	-8,3	-8,6	-8,4	-8,2	-9,2	-8,8	-6,5	-7,1	-7,8	-9,3
Benzaldehyde_21	-8,2	-8,2	-8,7	-8,5	-8,3	-9,3	-9,2	-6,3	-6,6	-7,7	-9,1
Benzaldehyde_22	-8,4	-8,4	-8,9	-8,6	-8,4	-9,0	-9,1	-6,6	-6,9	-8,2	-9,2
Benzaldehyde_23	-8,9	-8,9	-8,9	-8,9	-8,7	-9,8	-8,9	-6,7	-6,9	-8,5	-10,1
Benzaldehyde_24	-8,1	-8,1	-8,5	-8,4	-8,1	-9,2	-9,0	-6,7	-6,7	-8,5	-9,3
Benzaldehyde_25	-8,5	-8,5	-8,7	-8,6	-8,4	-9,3	-8,6	-6,8	-6,7	-7,7	-9,5
Benzaldehyde_26	-9,8	-9,8	-9,8	-9,6	-9,6	-10,5	-9,4	-7,1	-7,4	-8,3	-10,6
Benzaldehyde_27	-9,2	-9,2	-9,6	-9,1	-9,3	-10,2	-9,7	-7,0	-7,5	-8,1	-10,2
Benzaldehyde_28	-9,0	-9,0	-9,0	-8,6	-8,8	-10,0	-9,1	-6,6	-7,0	-8,0	-9,7
Benzaldehyde_29	-8,0	-8,0	-8,4	-8,2	-8,1	-9,0	-8,6	-6,3	-6,8	-7,6	-8,9
Benzaldehyde_30	-8,2	-8,2	-8,3	-8,0	-7,9	-9,1	-8,6	-6,1	-6,7	-8,4	-9,2
Benzaldehyde_31	-8,8	-8,8	-8,1	-8,1	-8,6	-10,2	-9,1	-6,6	-7,2	-8,9	-10,2
Benzaldehyde_32	-8,6	-8,6	-8,7	-8,4	-8,5	-9,7	-8,8	-6,4	-6,8	-8,2	-9,5
Benzaldehyde_33	-8,7	-8,7	-9,0	-8,7	-8,8	-9,8	-9,2	-6,5	-6,9	-7,6	-9,5
Benzaldehyde_34	-8,5	-8,5	-8,5	-8,9	-8,4	-9,6	-8,5	-6,5	-6,9	-8,1	-10,4
Benzaldehyde_35	-8,4	-8,5	-8,9	-8,8	-8,5	-9,3	-8,9	-6,7	-6,7	-8,2	-9,5
Benzaldehyde_36	-8,4	-8,4	-8,5	-8,3	-8,2	-9,2	-8,9	-6,3	-6,8	-8,3	-9,5
Benzaldehyde_37	-9,8	-9,8	-9,9	-9,6	-9,6	-10,5	-9,4	-7,1	-7,4	-8,3	-10,6
Benzaldehyde_38	-8,5	-8,5	-8,8	-8,7	-8,2	-9,7	-9,4	-6,4	-6,8	-8,0	-9,6
Benzaldehyde_39	-8,6	-8,6	-8,7	-8,6	-8,4	-9,3	-9,0	-6,9	-6,8	-8,3	-9,3
Benzaldehyde_40	-8,5	-8,5	-8,6	-8,5	-8,3	-9,1	-8,8	-6,3	-6,7	-7,3	-9,3
Benzaldehyde_41	-8,2	-8,2	-8,3	-8,1	-8,1	-9,5	-8,9	-6,2	-6,8	-8,5	-9,4
Benzaldehyde_42	-9,3	-9,3	-9,3	-9,2	-9,3	-10,1	-8,6	-7,0	-7,6	-8,1	-10,7
Benzaldehyde_43	-9,3	-9,3	-9,9	-9,6	-9,5	-10,2	-9,0	-7,0	-7,6	-8,5	-10,9
Benzaldehyde_44	-10,0	-10,0	-9,8	-9,6	-9,8	-9,9	-8,9	-7,5	-7,9	-9,2	-11,8
Benzaldehyde_45	-9,6	-9,6	-9,8	-9,7	-9,5	-10,8	-9,7	-7,2	-7,4	-8,3	-10,9
Benzaldehyde_46	-8,2	-8,2	-8,6	-8,4	-8,3	-9,2	-8,7	-6,1	-6,6	-7,7	-9,2

Benzaldehyde_47	-8,3	-8,3	-8,4	-8,0	-8,2	-9,5	-8,9	-6,1	-6,7	-7,9	-9,5
Benzaldehyde_48	-9,1	-9,1	-9,2	-9,3	-9,1	-9,8	-8,0	-6,9	-7,7	-7,7	-10,5
Benzaldehyde_49	-9,5	-9,5	-9,7	-9,5	-9,7	-10,0	-7,9	-7,0	-7,7	-7,9	-10,9
Benzaldehyde_50	-8,9	-8,9	-8,4	-8,2	-8,9	-9,7	-8,4	-6,2	-6,5	-8,9	-10,6
Benzaldehyde_51	-8,9	-8,9	-8,1	-8,1	-8,7	-10,0	-8,9	-6,6	-7,1	-8,5	-10,0
Benzaldehyde_52	-8,3	-8,3	-8,2	-7,9	-8,0	-9,2	-8,7	-6,1	-6,7	-8,3	-9,4
Benzaldehyde_53	-8,2	-8,3	-8,9	-8,7	-8,4	-8,9	-9,0	-6,6	-7,0	-7,6	-9,2
Benzaldehyde_54	-8,3	-8,3	-8,7	-8,6	-8,1	-9,4	-9,2	-6,4	-6,7	-8,7	-9,4
Benzaldehyde_55	-8,2	-8,2	-8,8	-8,6	-8,3	-9,0	-8,8	-6,5	-6,9	-7,6	-9,2
Benzaldehyde_56	-8,6	-8,6	-8,6	-8,5	-8,4	-9,6	-8,9	-6,6	-6,8	-8,7	-9,8
Benzaldehyde_57	-8,2	-8,2	-8,7	-8,4	-8,2	-8,9	-8,6	-6,4	-6,8	-7,6	-9,0
Benzaldehyde_58	-9,2	-9,2	-9,5	-9,3	-9,3	-10,5	-8,8	-7,2	-7,5	-8,4	-10,7
Benzaldehyde_59	-8,1	-8,1	-8,4	-8,1	-8,0	-9,0	-8,5	-6,1	-6,7	-7,6	-9,2
Benzaldehyde_60	-9,2	-9,2	-9,5	-9,1	-9,3	-10,2	-9,6	-7,0	-7,5	-8,0	-10,2
Benzaldehyde_61	-8,2	-8,2	-8,7	-8,5	-8,4	-9,1	-8,8	-6,4	-6,6	-7,9	-9,1
Benzaldehyde_62	-9,1	-9,1	-9,2	-8,8	-8,8	-9,4	-8,7	-6,6	-7,0	-7,7	-9,9
Benzaldehyde_63	-8,7	-8,7	-8,9	-8,7	-8,2	-9,5	-8,9	-6,5	-6,8	-8,2	-9,5
Benzaldehyde_64	-9,9	-9,9	-10,1	-9,9	-10,0	-10,7	-8,5	-7,3	-7,7	-7,9	-10,8
Benzaldehyde_65	-9,9	-9,9	-10,1	-9,9	-9,9	-10,7	-9,0	-7,5	-7,6	-8,1	-11,1
Benzaldehyde_66	-9,9	-9,9	-10,2	-9,9	-10,0	-10,7	-7,0	-7,6	-7,6	-8,0	-11,3
Benzaldehyde_67	-9,2	-9,2	-9,1	-9,2	-8,9	-10,3	-9,0	-6,9	-6,9	-8,1	-10,3
Benzaldehyde_68	-8,5	-8,5	-8,8	-8,6	-8,4	-9,2	-8,8	-6,5	-6,9	-7,5	-9,2
Benzaldehyde_69	-9,0	-9,0	-9,1	-8,9	-8,8	-9,7	-9,1	-6,5	-7,1	-7,8	-9,8
Benzaldehyde_70	-7,9	-7,9	-7,9	-7,4	-7,9	-9,1	-8,3	-6,1	-6,4	-8,2	-9,2
Benzaldehyde_71	-8,5	-8,5	-8,3	-8,3	-8,3	-9,6	-8,5	-6,7	-6,9	-7,8	-9,6
Benzaldehyde_72	-8,9	-8,9	-9,1	-8,8	-9,3	-9,9	-7,5	-6,3	-6,7	-8,8	-10,7
Benzaldehyde_73	-7,9	-7,9	-8,7	-8,6	-8,2	-8,9	-9,1	-6,3	-6,5	-7,9	-8,9
Benzaldehyde_74	-8,1	-8,1	-8,3	-8,1	-8,0	-8,9	-8,8	-6,1	-6,6	-8,0	-9,1
Benzaldehyde_75	-8,6	-8,6	-8,6	-8,4	-8,3	-9,2	-8,9	-6,6	-6,8	-8,2	-9,3
Benzaldehyde_76	-8,5	-8,5	-8,7	-8,5	-8,3	-9,1	-8,8	-6,5	-7,0	-7,5	-9,1
Benzaldehyde_77	-8,8	-8,8	-9,0	-8,8	-8,7	-9,7	-9,3	-6,5	-6,8	-7,8	-9,6
Benzaldehyde_78	-10,2	-10,2	-10,1	-10,3	-10,2	-10,2	-9,4	-7,6	-8,0	-9,1	-11,7
Benzaldehyde_79	-8,7	-8,7	-8,7	-8,4	-8,4	-9,4	-8,9	-6,3	-6,8	-8,3	-9,4
Benzaldehyde_80	-8,3	-8,3	-8,7	-8,5	-8,2	-9,3	-9,1	-6,5	-7,2	-8,3	-9,5
Benzaldehyde_81	-8,3	-8,3	-8,3	-8,1	-8,2	-9,5	-8,7	-6,4	-7,0	-8,0	-9,6
Benzaldehyde_82	-8,2	-8,2	-8,4	-8,3	-8,1	-9,1	-8,6	-6,4	-7,1	-7,6	-9,2
Benzaldehyde_83	-8,4	-8,4	-8,5	-8,3	-8,2	-9,2	-8,8	-6,3	-6,9	-8,3	-9,5
Benzaldehyde_84	-8,5	-8,5	-8,6	-8,5	-8,2	-9,6	-9,5	-6,2	-6,8	-8,0	-9,4
Benzaldehyde_85	-9,2	-9,2	-9,5	-9,0	-9,2	-10,0	-9,5	-6,8	-7,6	-7,9	-10,1
Benzaldehyde_86	-8,4	-8,4	-8,9	-8,8	-8,5	-9,3	-8,9	-6,6	-6,8	-8,1	-9,5
Benzaldehyde_87	-9,3	-9,3	-9,3	-9,1	-9,3	-10,1	-8,4	-7,0	-7,5	-7,8	-10,7
Benzaldehyde_88	-8,8	-8,8	-8,0	-8,1	-8,6	-10,1	-9,0	-6,5	-7,2	-8,7	-10,1
Benzaldehyde_89	-10,6	-10,6	-10,8	-10,4	-10,4	-11,4	-10,3	-7,7	-8,1	-9,1	-11,9
Benzaldehyde_90	-8,5	-8,5	-8,5	-8,3	-8,2	-9,5	-8,8	-6,3	-7,0	-7,8	-9,7
Benzaldehyde_91	-8,1	-8,1	-8,5	-8,2	-8,1	-8,8	-8,4	-6,2	-6,9	-7,4	-8,9
Benzaldehyde_92	-8,8	-8,8	-8,7	-8,5	-8,3	-9,2	-9,3	-7,0	-7,1	-8,4	-9,2
Benzaldehyde_93	-8,2	-8,2	-8,0	-7,7	-7,9	-9,2	-8,7	-6,1	-6,8	-8,3	-9,2
Benzaldehyde_94	-7,9	-7,9	-8,5	-8,4	-8,1	-8,7	-9,2	-6,2	-6,5	-7,7	-8,8
Benzaldehyde_95	-8,1	-8,2	-8,4	-8,1	-7,9	-8,9	-8,7	-6,2	-7,2	-7,7	-9,1
Benzaldehyde_96	-9,4	-9,4	-10,0	-9,7	-9,6	-10,1	-8,5	-7,1	-7,3	-8,1	-10,6

Figure 5.4. Heat map of average ligand binding affinities (kcal/mol), sorted by the benzaldehyde. Green hits correspond to good binding affinities, and red corresponds to bad binding affinities.

In general the results in Figures 5.3 and 5.4 show that the binding affinities with the CHI-chalcone complexes are influenced both by the particular CHI crystal structure used and the substituents on the chalcone derivatives. The average binding affinities with respect to the CHI crystallographic structures, in order of increasing strength (increasingly negative values) are as follows:

4DOO<1JX0<1FM7<1JX1<1EYP<1EYQ<1FM8<1JEP<4DOL<4DOK<4DOI

From the CHIs considered in this study it could not be determined how the CHI-originating organism influences the binding affinity. This is because the best (greatest negative) binding affinity observed across ligands is with 4DOO, a crystal structure of CHI isolated from *A. Thaliana*, whereas the crystal structures of the other *A. Thaliana* CHIs are the ones with the worst (least negative) average binding affinities (4DOL, 4DOK and 4DOI). The PDB crystallographic structure of the mutant enzyme 1JX0 has the second best average binding affinity, but the crystal structure of other mutant 1JX1 ranks fourth worst. The *M. Sativa* PDB crystal structure 1FM7 has the third best average binding affinity, but the other *M. Sativa* crystal structures 5, 6, 7 and 8. This highlights the possible influence of different side-chain conformations in the respective crystal structures – bearing in mind that 4DOI and 4DOK structures contained no ligand in the crystal structure.

The substituents on the chalcone derivatives seem to have a great influence on the binding energies. Additional benzene rings either on the acetophenone moiety or the benzaldehyde moiety seem to enhance the binding affinities. Thus from Figure 5.3, the chalcones with acetophenones 10, 13, 14, 15, 19, 21 and 22 have the best binding affinities (see acetoph_10_benzald_10 in Figure 5.2 as an example), all of these have at least one additional benzene ring fused to the acetophenone moiety. Sometimes, an additional saturated ring fused to the acetophenone moiety also enhances the binding affinity as with the chalcones derived from acetophenones 63 and 67 (see acetoph_67_benzald_67 in Figure 5.2). From Figure 5.4, the chalcones derived from benzaldehydes 2, 17, 44, 60, 78 and 89, which contain an additional

benzene ring (*e.g.* acetoph_60_benzald_60 in Figure 5.2) are examples of chalcones with the best binding affinity. An additional fused benzene ring attached to the benzaldehyde moiety also seems to influence and improve the CHI-chalcone binding affinities, examples are with the chalcones derived from benzaldehydes 6 and 42.

Increases in binding affinity may also be attributed to the presence of additional functional groups, such as ethers, OH, $N(CH_3)_2$, and alkyl chains. Examples of this are chalcones with the acetophenone 18 (derived from 2-ethyl acetophenone), 57 (additional ethoxy), 80 [additional $N(CH_3)_2$], and 45 (two OH groups). Examples of benzaldehyde precursors with functionality that improves binding are benzaldehyde 1 (methyl) and 59 (OH and OMe). Table 5.1 lists the best and worst docked ligand for each CHI crystal structure. When cross-referencing the particular acetophenone and benzaldehyde precursors with Figure 5.2 and the preceding discussion, the trends for good binding already noted are featured in this set of ligands.

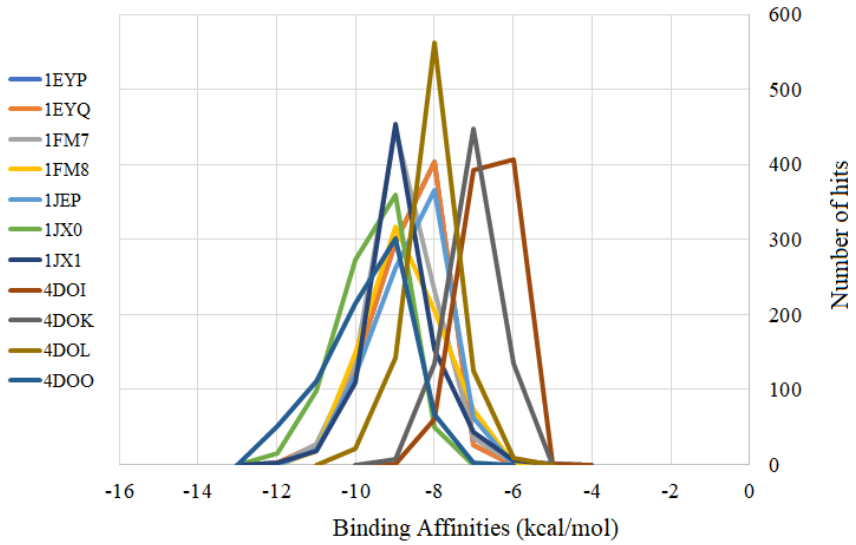
Table 5.1. Ligands with the lowest binding affinity and ligands with the highest binding affinity.

CHI	**Ligands and their binding affinities (kcal/mol)**	
	Ligand with best binding affinity	**Ligand with worst binding affinity**
1EYP	Acetophenone_22_bezaldehyde_89 -13.1	Acetophenone_56_bezaldehyde_83 -6.2
1EYQ	Acetophenone_22_bezaldehyde_89 -13.1	Acetophenone_56_bezaldehyde_83 -6.2
1FM7	Acetophenone_5_bezaldehyde_89 -11.9	Acetophenone_56_bezaldehyde_52 -5.7
1FM8	Acetophenone_10_bezaldehyde_78 -11.8	Acetophenone_18_bezaldehyde_70 -6.0
1JEP	Acetophenone_15_bezaldehyde_44 -13.0	Acetophenone_57_bezaldehyde_30 -6.4
1JX0	Acetophenone_13_bezaldehyde_89 13.5	Acetophenone_56_bezaldehyde_83 -6.8
1JX1	Acetophenone_19_bezaldehyde_60 -12.6	Acetophenone_66_bezaldehyde_72 -4.9
4DOI	Acetophenone_19_bezaldehyde_44 -9.0	Acetophenone_80_bezaldehyde_94 -5.0
4DOK	Acetophenone_21_bezaldehyde_89 -9.7	Acetophenone_78_bezaldehyde_3 -5.5
4DOL	Acetophenone_56_bezaldehyde_42 -11.5	Acetophenone_45_bezaldehyde_59 -5.8
4DOO	Acetophenone_13_bezaldehyde_89 -14.9	Acetophenone_62_bezaldehyde_94 -6.9

The discussion of the averaged heat maps in Figure 5.3 and 5.4 focussed mostly on the trends observed with the ligands. Figures 5.5a and b provide a closer view of the binding affinity trends with respect to the receptors; these figures show the best binding affinity and the number of hits (ligands exhibiting that binding affinity) corresponding to it, the least binding affinity and the number of hits corresponding to it, and the binding affinity corresponding to the most hits with each receptor. The intensity of the hit's colour in each column corresponds to the variable considered in that column.

CHI	Variables					
	best binding affinity (kcal/mol)	Number of hits	worst binding affinity (kcal/mol)	Number of hits	Most occurring binding affinity (kcal/mol)	Number of hits
1eyp	-12	2	-7	26	-8	403
1eyq	-12	2	-7	26	-8	403
1fm7	-11	28	-6	2	-9	448
1fm8	-11	18	-6	1	-9	317
1jep	-11	19	-7	62	-8	366
1jx0	-12	15	-8	50	-9	359
1jx1	-12	2	-5	1	-9	453
4doi	-9	1	-5	1	-6	406
4dok	-9	7	-6	135	-7	447
4dol	-10	22	-6	9	-8	562
4doo	-12	52	-7	2	-9	301

a.

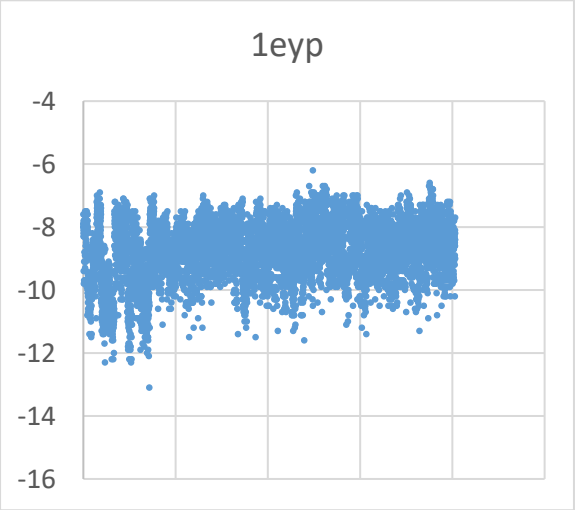


b.

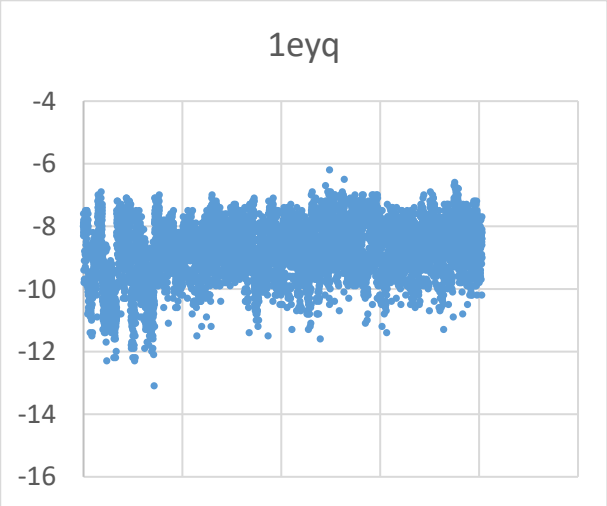
Figure 5.5. a) Heat map (Green hits correspond to good binding affinities and large number of hits, and red corresponds to bad binding affinities and low number of hits) and b) Distribution curve of the range of binding affinities (kcal/mol) for ligands with the CHIs.

Figures 5.5a and b also show that most CHI-chalcone complexes have binding affinities between -6 and -9 kcal/mol. The best CHI-chalcone binding affinity (-12 kcal/mol) corresponding the chalcone complexes with 1EYP, 1EYQ, 1JX0, 1JX1, 4DOO; although these are the best binding affinities, the number of times they are observed with 1EYP and 1EYQ is very small (only 2 hits), and 1JX1 only exhibits this excellent binding with two chalcones from the library. However, 1JX0 exhibits 15 instances of this excellent binding while 4DOO has 52 hits. The best binding affinity with respect to the highest number of hits with CHIs is -9 kcal/mol, and is exhibited by 448 1FM7-chalcone complexes, 317 1FM8-chalcone complexes, 359 1JX0-chalcones complexes, 453 1JX1-chalcones complexes, and 301 1FM8-chalcones complexes. From the CHI structures considered in this study 1FM7 and 1FM8 seem to be ideal choices for further studies because of frequency of hits with a binding affinity of -9 kcal/mol and the fact that their best binding affinities are very good at -11 kcal/mol □□a value corresponding to 18 hits with 1FM8 and 28 hits with 1FM7. The worst binding affinity with both 1FM7 and 1FM8 is -6 kcal/mol corresponding to 2 hits with 1FM7 and 1 hit with 1FM8. 1FM7 and 1FM8 are crystal structures of the identical protein which differ because of the ligand with which they are complexed. 1JX0 and 1JX1 are the second-best structures that can be considered for further studies; they have best binding affinities of -12 kcal/mol, most of their hits have a binding affinity of -9 kcal/mol, while their lowest binding affinities are -8 kcal/mol for 1JX0 and -5 kcal/mol for 1JX1.. Based on the data in Figure 5, and following the same criteria for comparison, 4DOI and 4DOK are the least favourable CHIs. In summary, 1FM7 and 1FM8 generally appear to bind well to all chalcones, whereas1EYP and 1EYQappear to show a high degree of selectivity towards a relatively small number of individual chalcones.

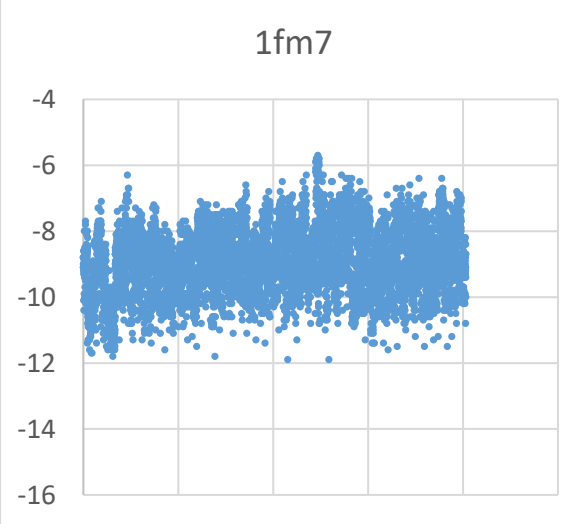
The discussion, thus far, has focussed on the broad patterns provided by the summarised data represented graphically in Figure 5., A detailed illustration of the binding affinities is given in Figure 5.6 which shows the x-y scatters of the binding affinities for each protein receptor across the chalcone library. Although the x-axis, in each case, does not show the ligand numbers, consecutive ligands are related by the ligand numbering established above. One can follow trends in binding which are in agreement with the above discussion. For instance, 1EYP and 1EYQ show selective binding, while the binding to 1FM7 and 1FM8 is generally good across the sets.


1eyp
-4
-6
-8
-10
-12
-14
-16

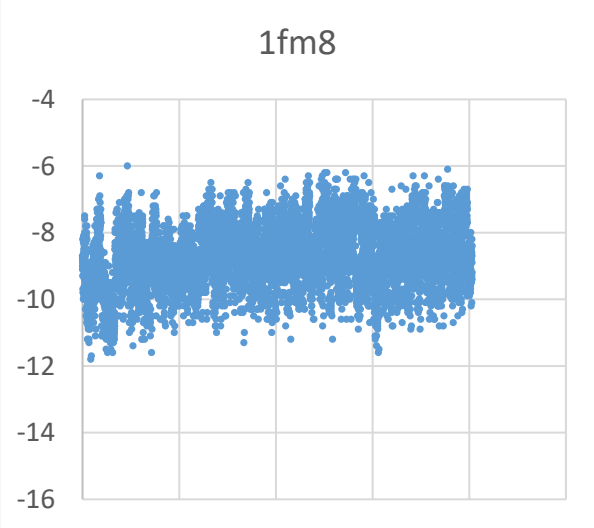
a)


1eyq
-4
-6
-8
-10
-12
-14
-16

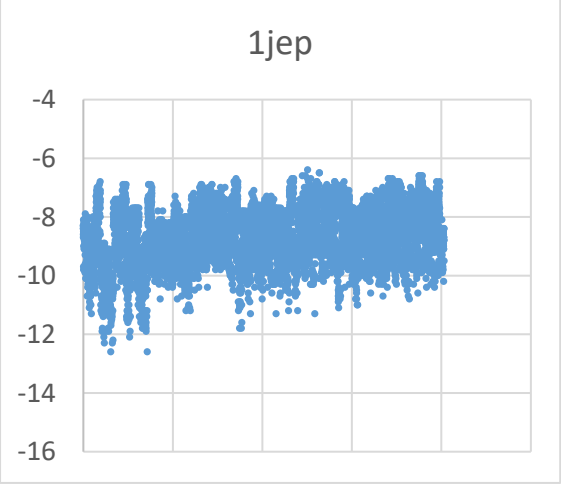
b)


1fm7
-4
-6
-8
-10
-12
-14
-16

c)


1fm8
-4
-6
-8
-10
-12
-14
-16

d)


1jep
-4
-6
-8
-10
-12
-14
-16

e)

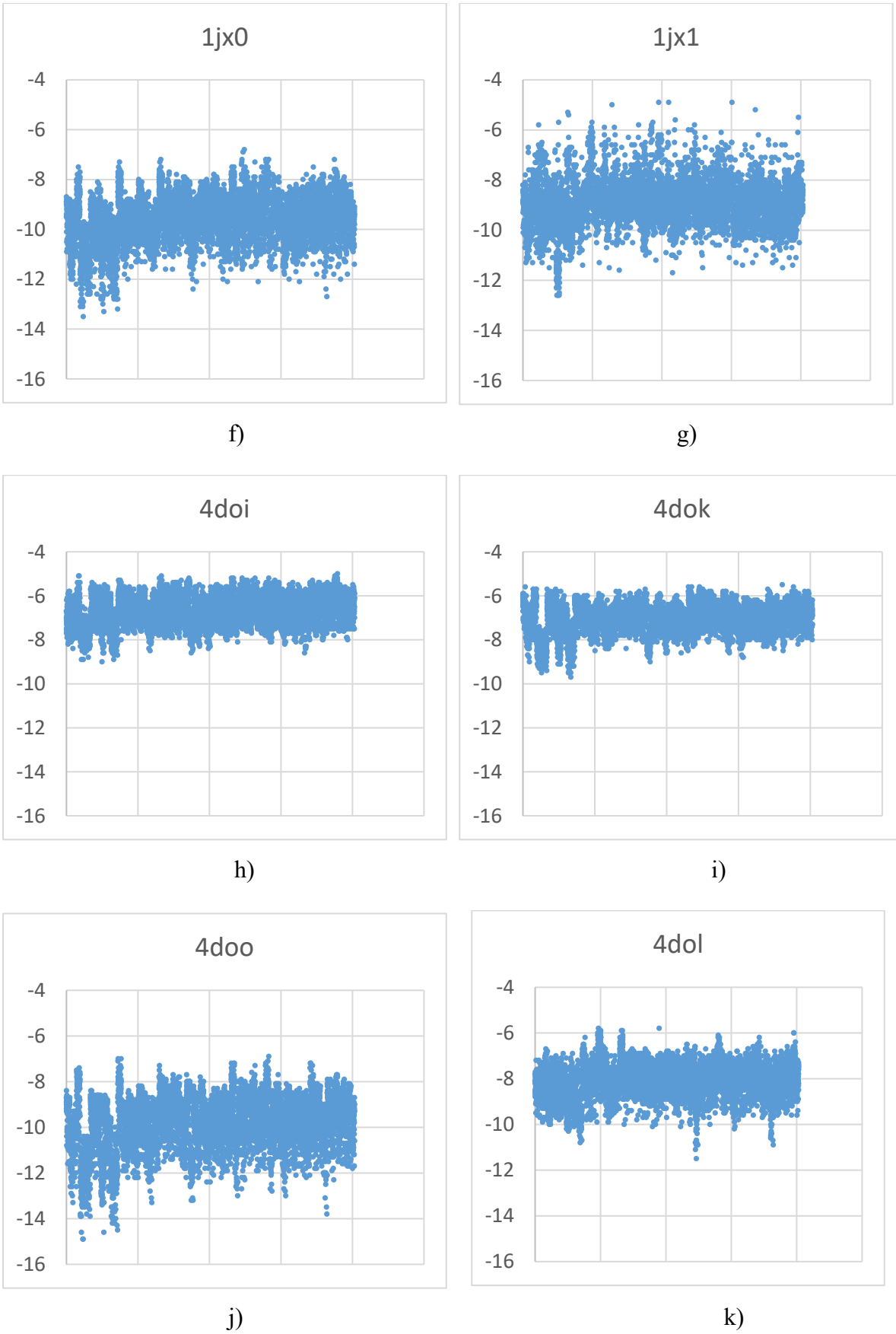


Figure 5.6. The x-y scatter of ligands involved in the CHI-chalcone complexes against their binding affinities (kcal/mol) for each protein.

5.3.2 Receptor-Ligand interactions (RLIs)

Binding affinity refers to the strength of the receptor-ligand binding interactions. Such receptor-ligand interactions are constituted by the non-covalent bonds formed by Van der Waals', hydrophobic, π-, ionic or electrostatic forces holding the receptor-ligand complex together (231). The binding orientation of the ligand and the conformational change(s) (mostly of the ligand) associated with complexation also contribute to the strength of a particular receptor-ligand binding affinity. This section analyses the RLIs and the docking poses of the 11 crystallographic CHI structures obtained from the PDB and the 8 064 chalcone derivatives obtained from HTVS. For simplicity, only the ligand with the greatest negative binding energy (the best docked ligand) and the ligand with the least negative binding energy (the worst docked ligand) for each CHI crystal structure is discussed.

Figure 5.7 shows the poses of the best and worst docked CHI-chalcone complexes for each CHI. The sphere on the left shows the binding pocket of each receptor (the active-site pocket) elucidated by Discovery Studio Visualizer from the receptor coordinates, together with the protein (as cartoon) and the docked ligand (as a ball and stick structure). On the right is shown the exact ligand pose together with the relevant binding-site interactions. The binding affinities of these complexes are available in table 5.1.

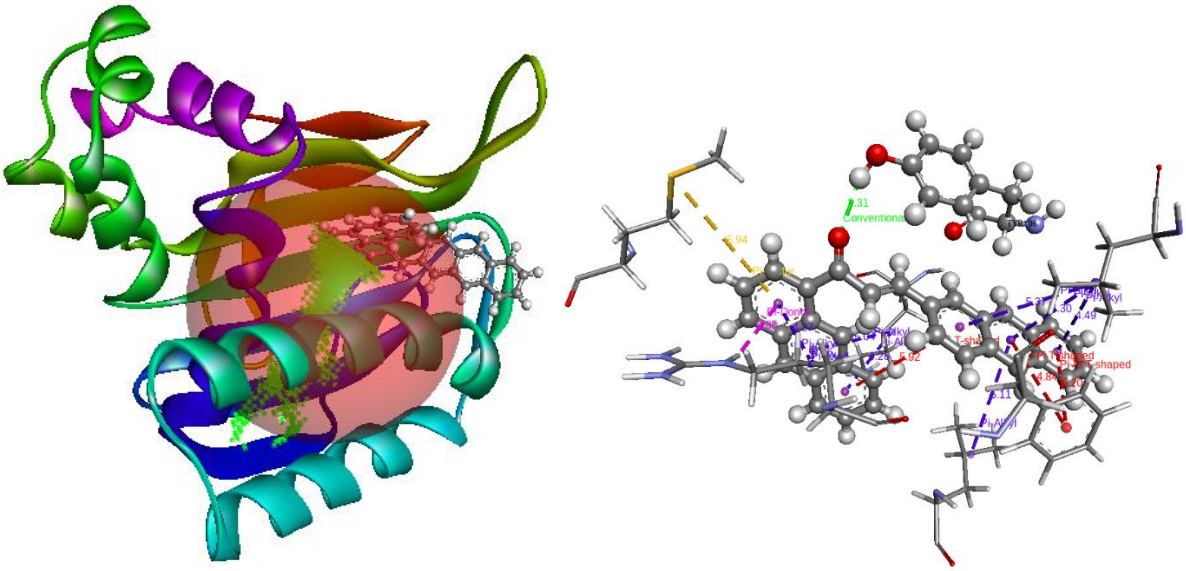


Figure 5.7a. 1EYP_Acetophenone_22_bezaldehyde_89 – ligand with best binding affinity

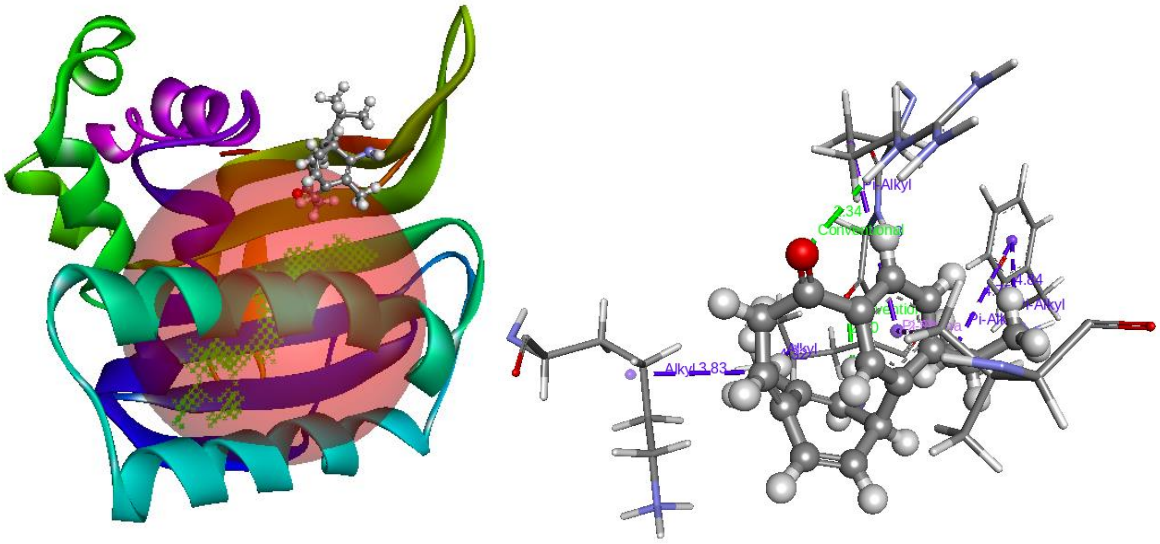


Figure 5.7b. 1EYP_Acetophenone_56_bezaldehyde_83 – ligand with worst binding affinity

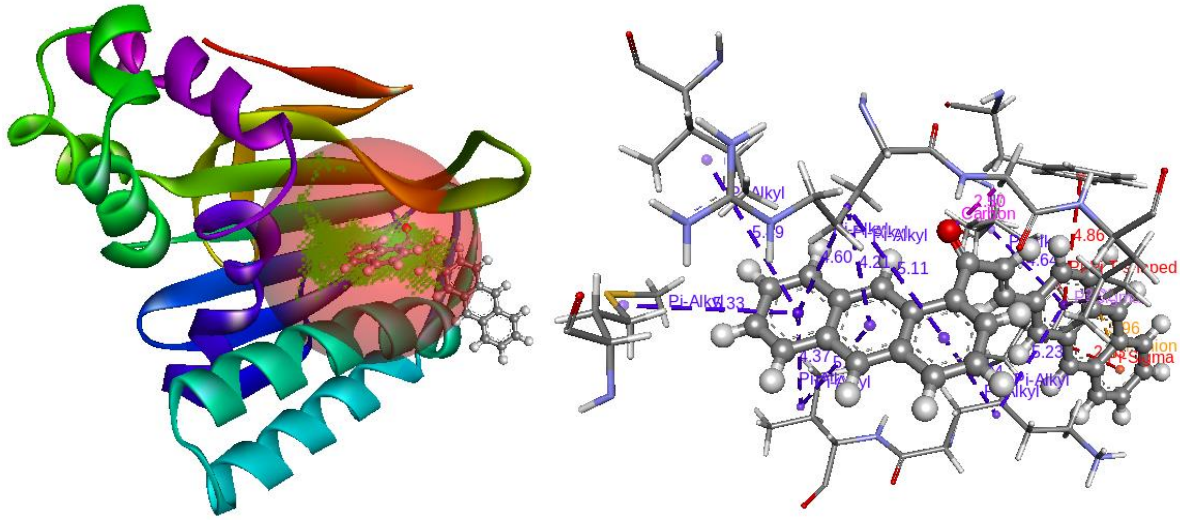


Figure 5.7c. 1EYQ_Acetophenone_22_bezaldehyde_89 – ligand with best binding affinity

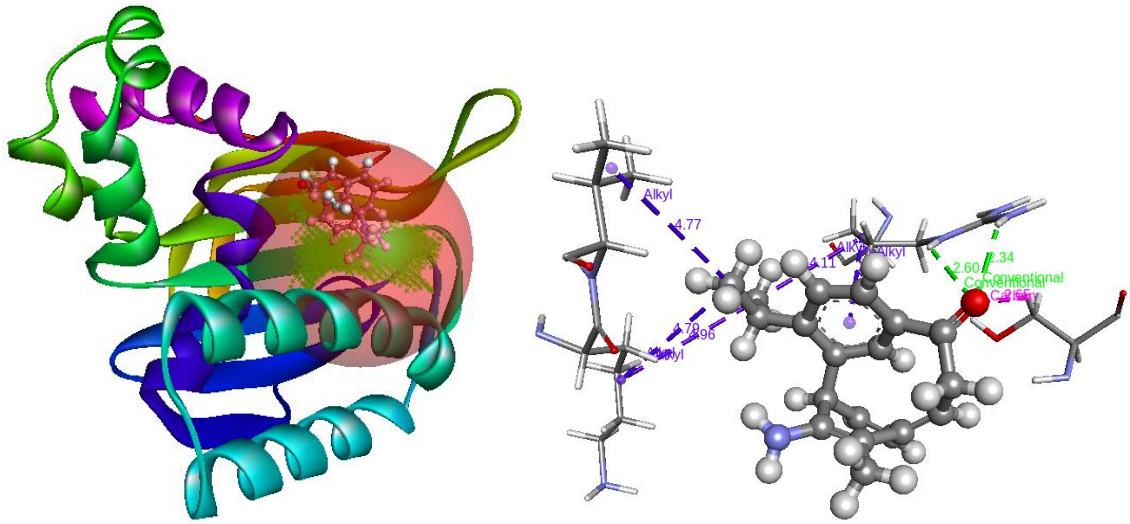


Figure 5.7d. 1EYQ_Acetophenone_56_bezaldehyde_83 – ligand with worst binding affinity

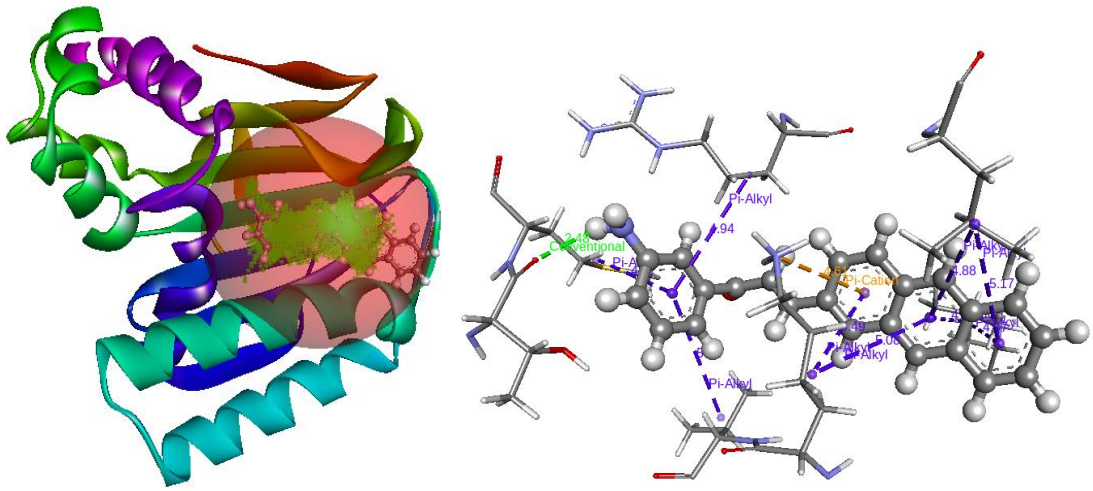


Figure 5.7e. 1FM7_Acetophenone_5_bezaldehyde_89 – ligand with best binding affinity

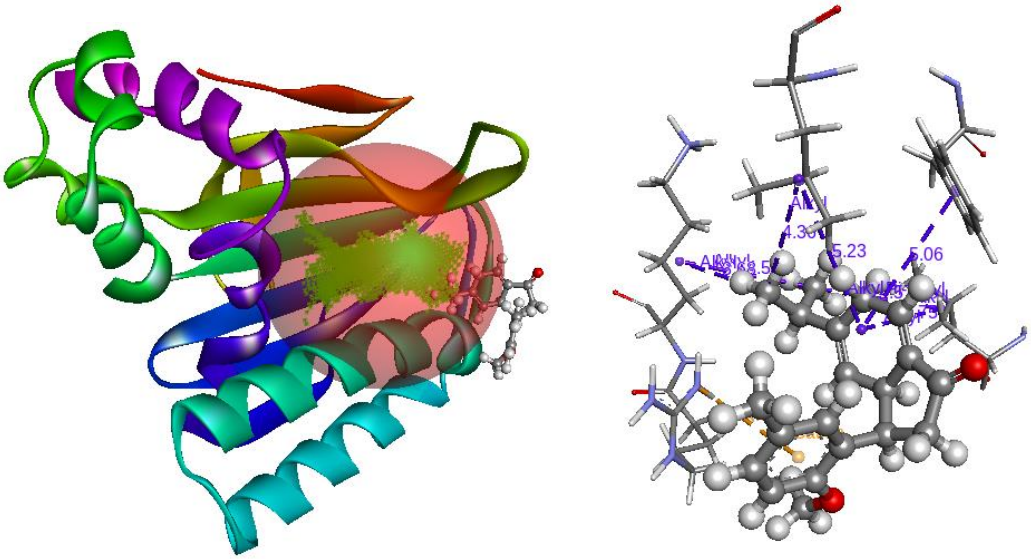


Figure 5.7f. 1FM7_Acetophenone_56_bezaldehyde_52 – ligand with worst binding affinity

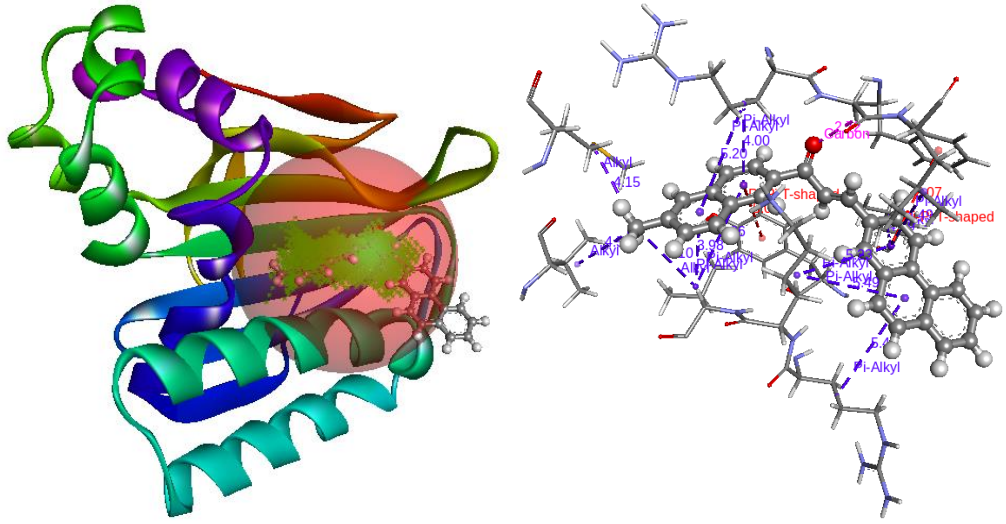


Figure 5.7g. 1FM8_Acetophenone_10_bezaldehyde_78 – ligand with best binding affinity

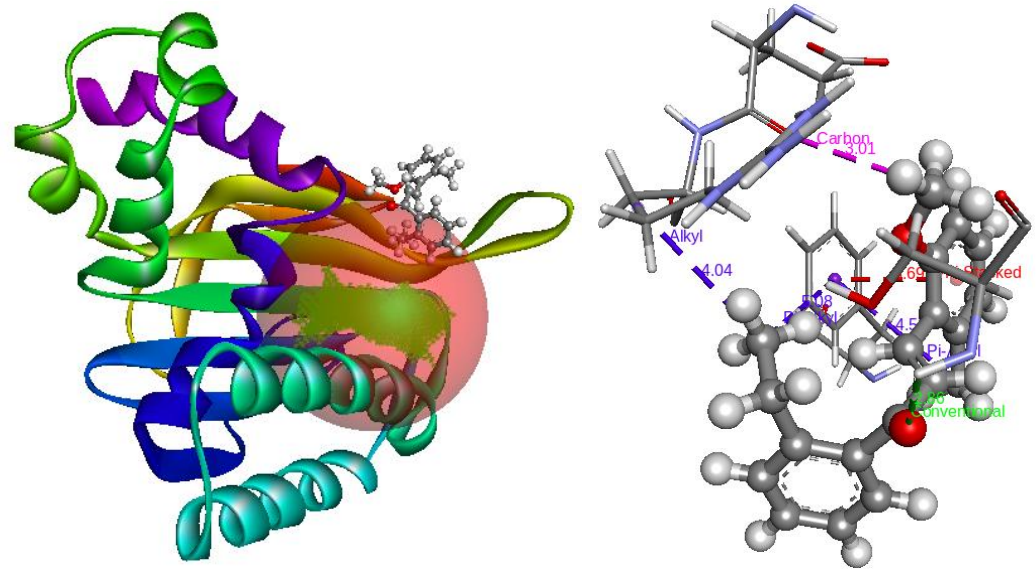


Figure 5.7h. 1FM8_Acetophenone_18_bezaldehyde_70 – ligand with worst binding affinity

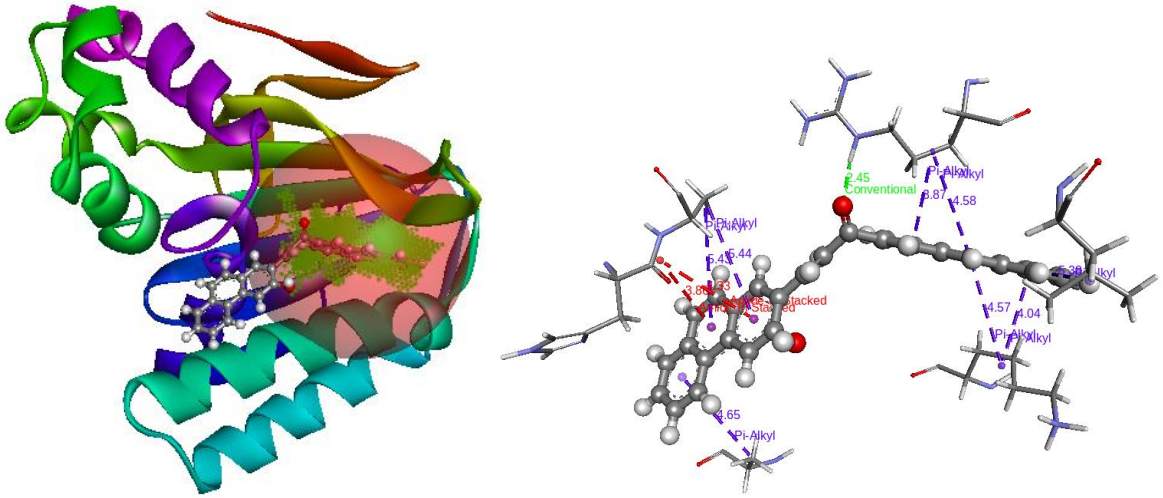


Figure 5.7i. 1JEP_Acetophenone_15_bezaldehyde_44 – ligand with best binding affinity

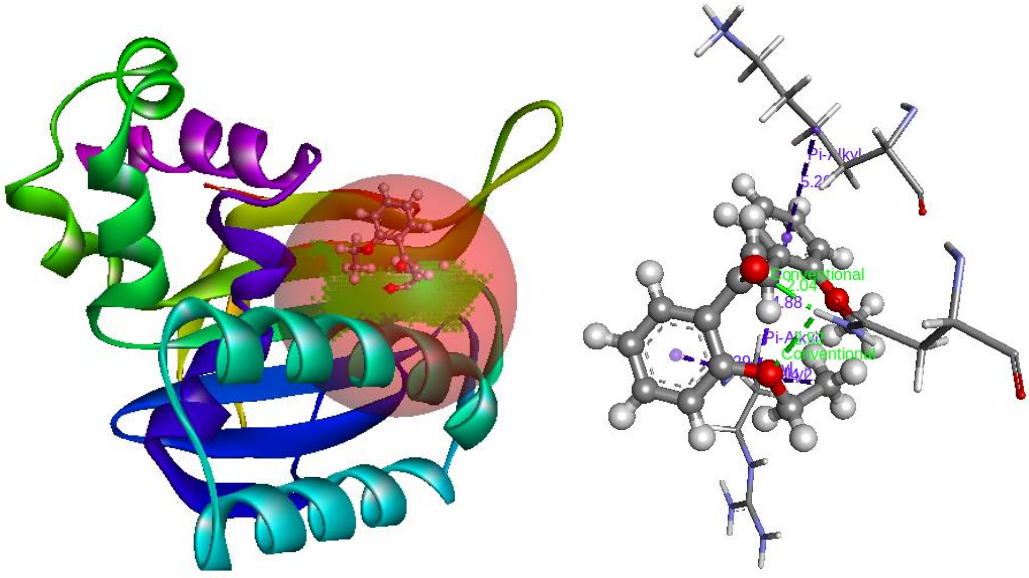


Figure 5.7j. 1JEP_Acetophenone_57_bezaldehyde_30 – ligand with worst binding affinity

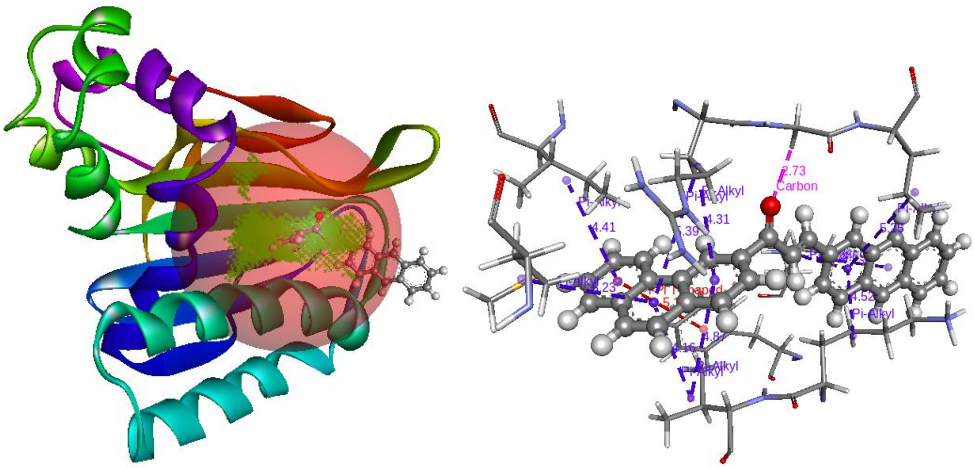


Figure 5.7k. 1JX0_Acetophenone_13_bezaldehyde_89 – ligand with best binding affinity

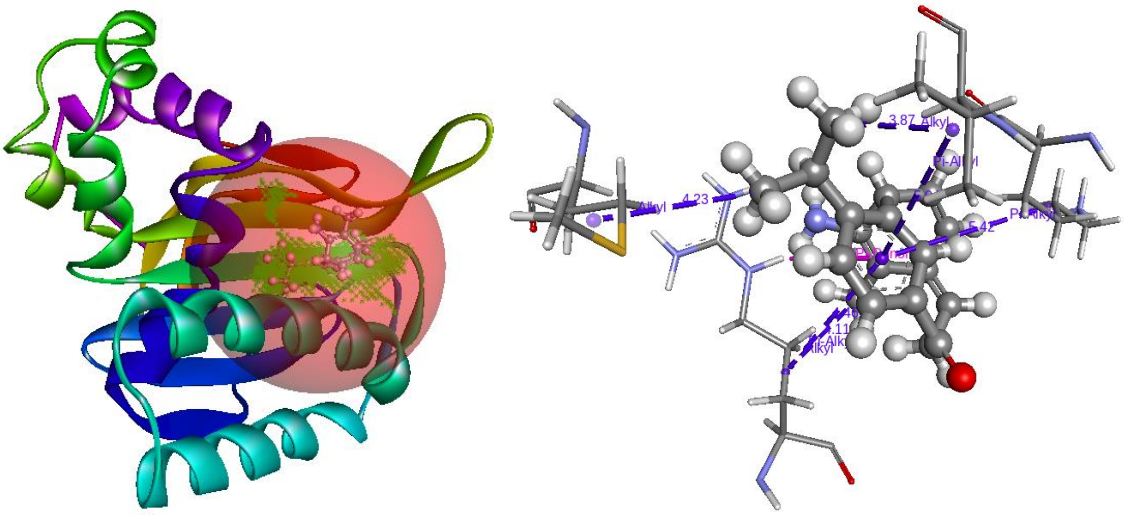


Figure 5.7l. 1JX0_Acetophenone_56_bezaldehyde_83 – ligand with worst binding affinity

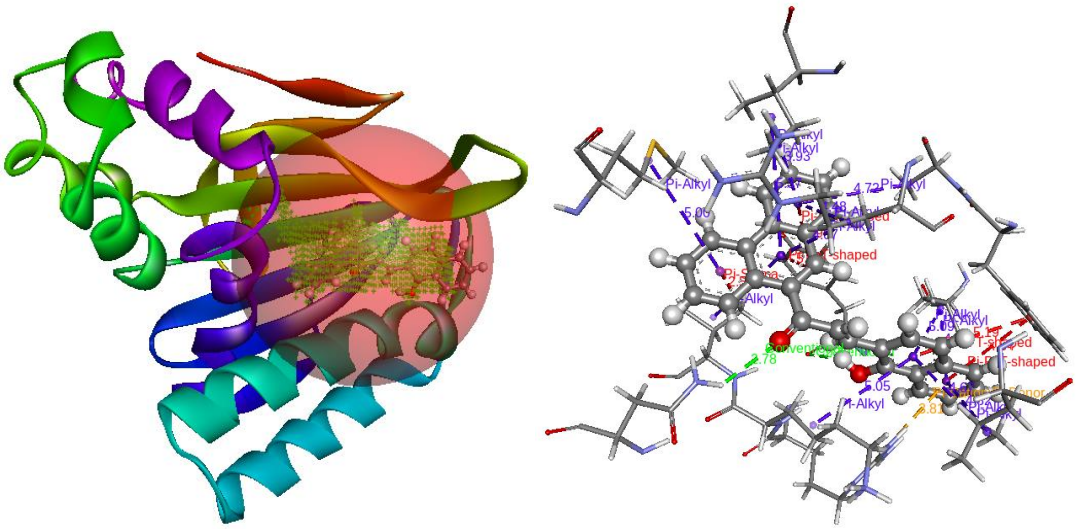


Figure 5.7m. 1JX1_Acetophenone_19_bezaldehyde_60 – ligand with best binding affinity

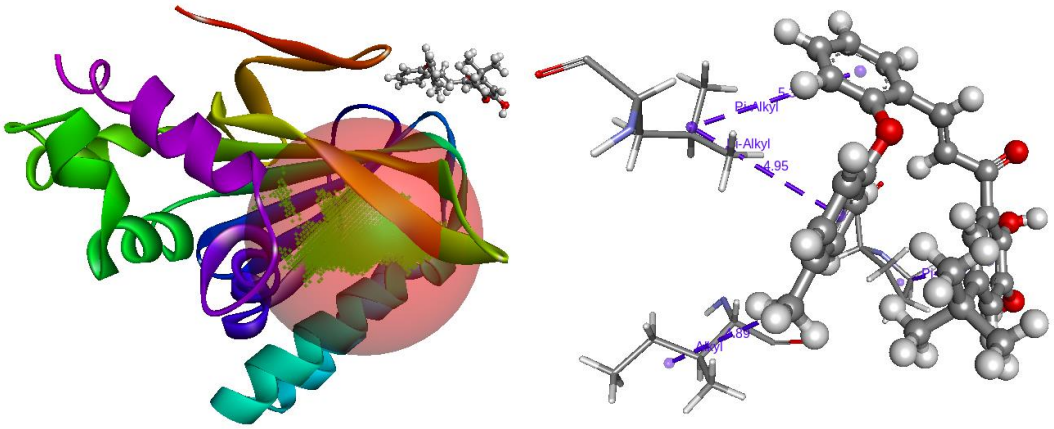


Figure 5.7n. 1JX1_Acetophenone_66_bezaldehyde_72 – ligand with worst binding affinity

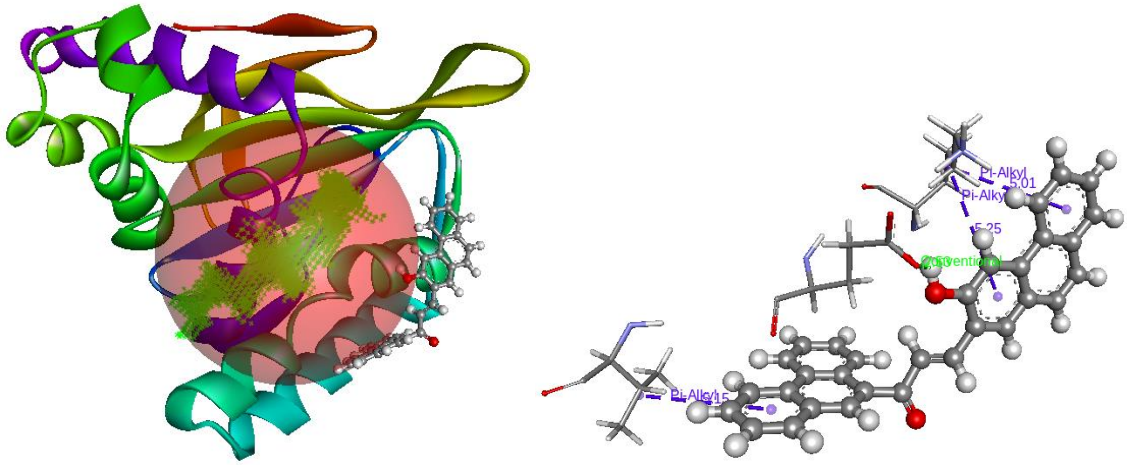


Figure 5.7o. 4DOI_Acetophenone_19_bezaldehyde_44 – ligand with best binding affinity

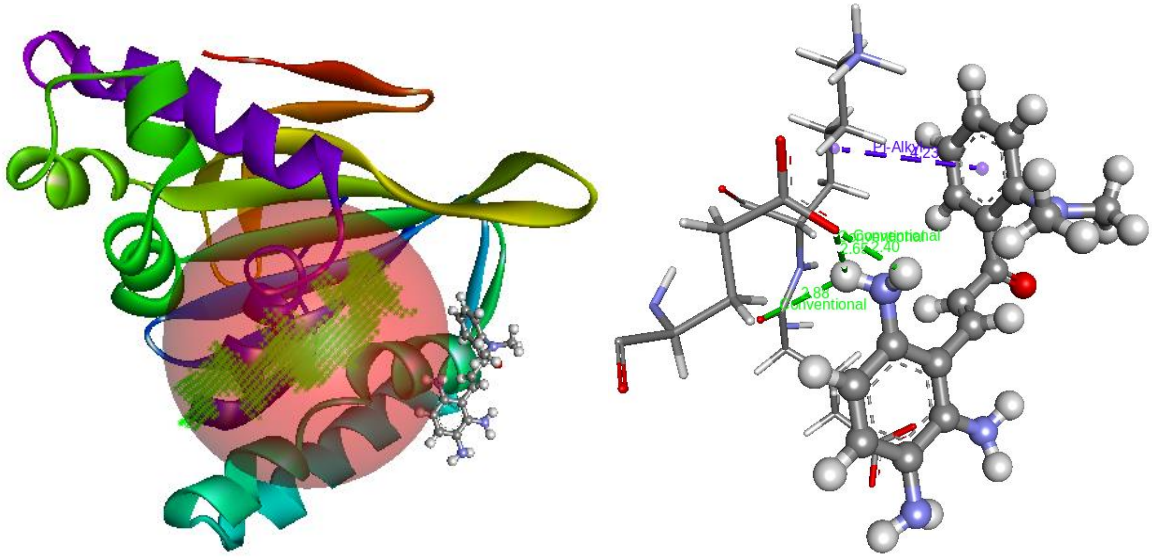


Figure 5.7p. 4DOI_Acetophenone_80_bezaldehyde_94 – ligand with worst binding affinity

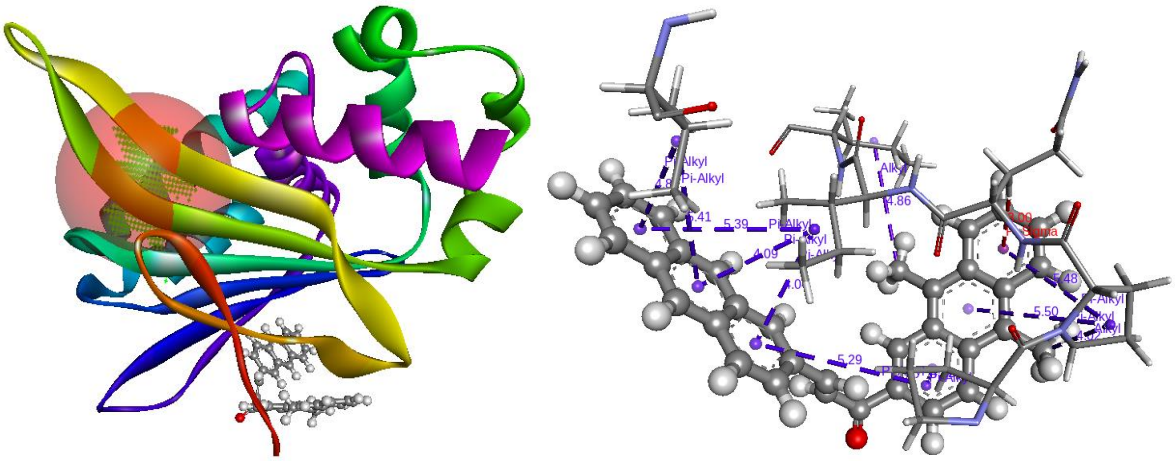


Figure 5.7q. 4DOK_Acetophenone_21_bezaldehyde_89 – ligand with best binding affinity

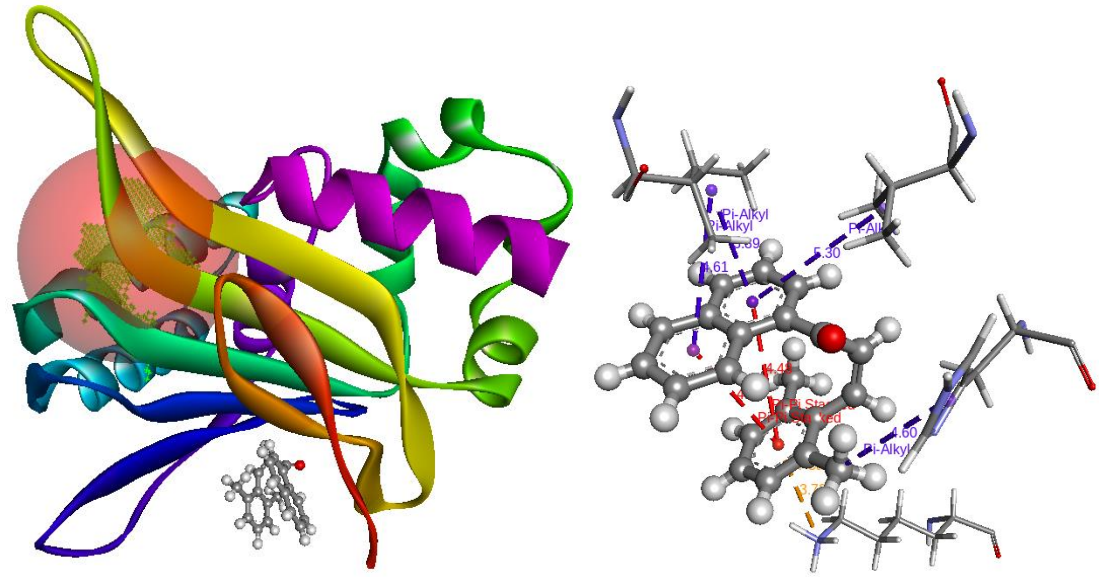


Figure 5.7r. 4DOK_Acetophenone_78_bezaldehyde_3 – ligand with worst binding affinity

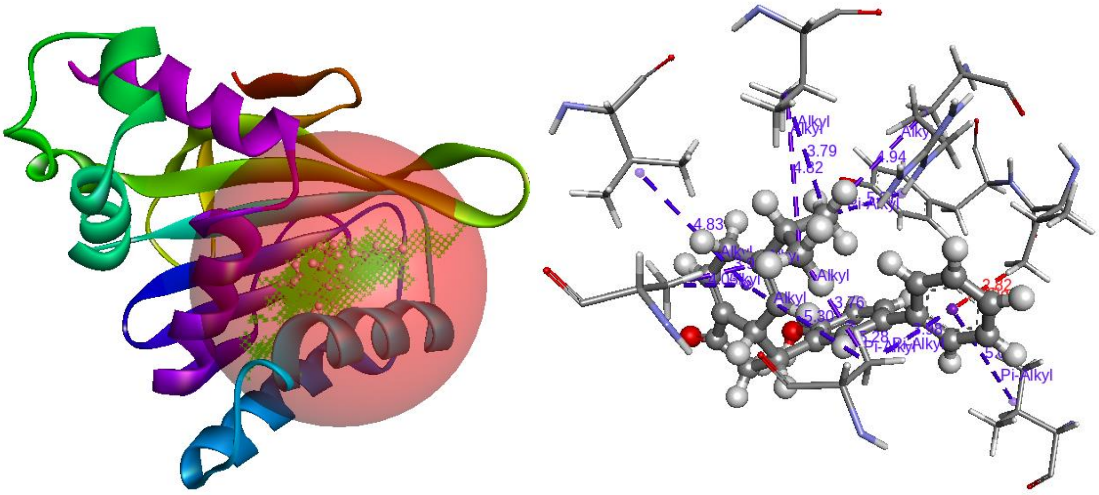


Figure 5.7s. 4DOL_Acetophenone_56_bezaldehyde_42 – ligand with best binding affinity

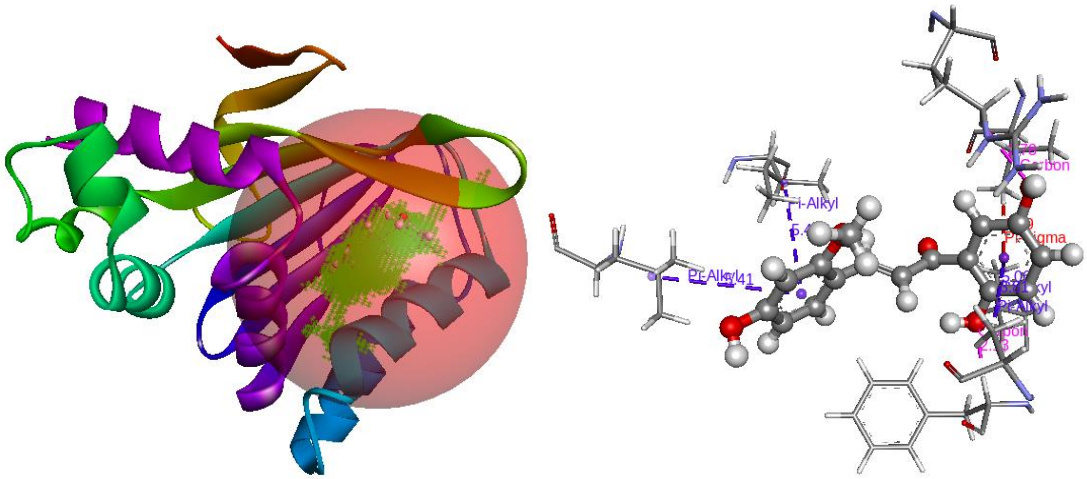


Figure 5.7t. 4DOL_Acetophenone_45_bezaldehyde_59 – ligand with worst binding affinity

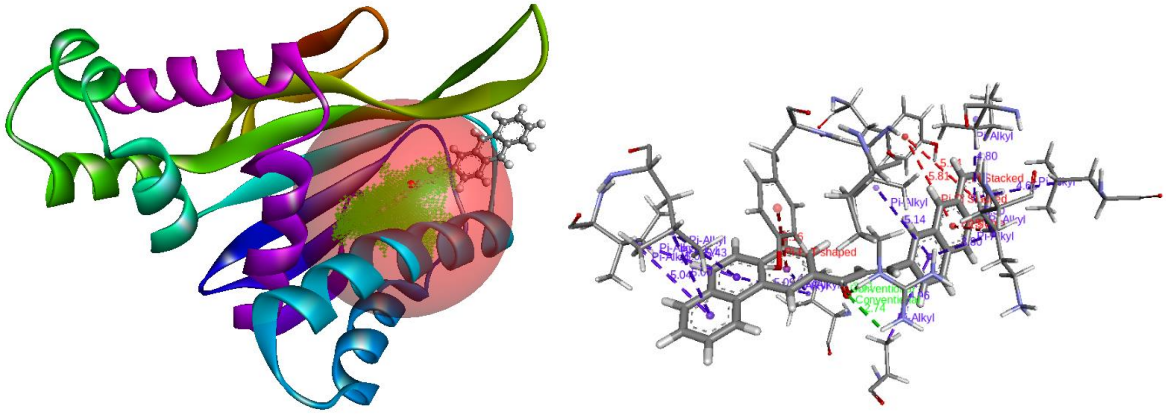


Figure 5.7u. 4DOO_Acetophenone_13_bezaldehyde_89 – ligand with best binding affinity

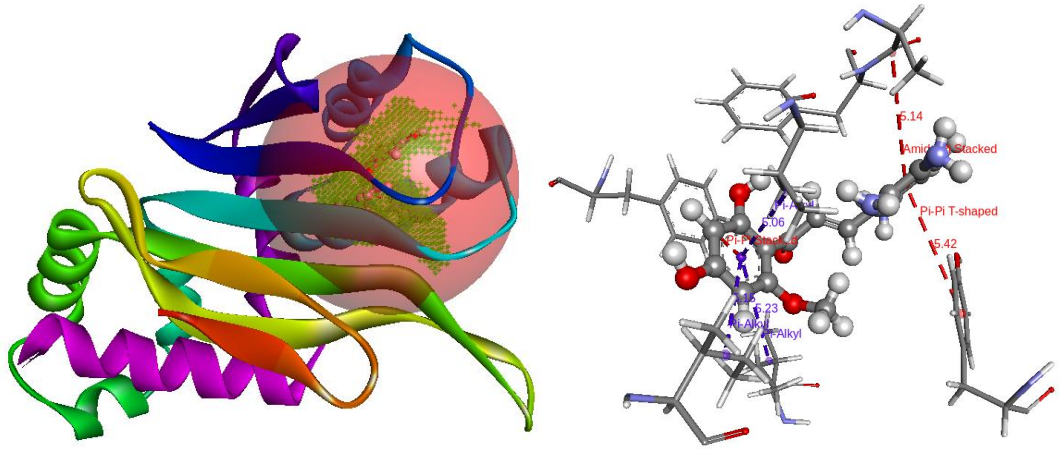


Figure 5.7v. 4DOO_Acetophenone_62_bezaldehyde_94 – ligand with worst binding affinity

Figure 5.7. Structures of the best and worst CHI-ligand complexes for each of the CHIs.

Figure 5.7a shows the 1EYP-Acetophenone_22_benzaldehyde_89 complex, which is the most strongly bound complex within the 1EYP-chalcone library complex set. Figure 5.7b shows the 1EYP-Acetophenone_56_benzaldehyde_83 complex, which corresponds to the most poorly bound ligand within this set. Acetophenone_22_benzaldehyde_89 is a large chalcone that fits well and tightly within the ligand binding site. The ligand in this case is stabilized by many receptor-ligand interactions, several of them being π-alkyl hydrophobic interactions. It seems that the presence of the two additional benzene rings in the benzaldehyde component of the chalcone stabilises the receptor-ligand complex – all three of the benzaldehyde aromatic rings are involved in these hydrophobic π-alkyl interactions supplemented by π-π interactions. Tyr106 of 1EYQ is also involved in a conventional hydrogen-bonding interaction with the carbonyl oxygen of the acetophenone component of the ligand. On the other hand, acetophenone_56_benzaldehyde_83 is a small chalcone and its structure occupies a small portion of the active site pocket. The number of 1EYP-Acetophenone_56_benzaldehyde_83 interactions are almost half (less than) those observed with the preceding 1EYP-Acetophenone_22_benzaldehyde_89 complex. The small number of 1EYP-Acetophenone_56_benzaldehyde_83 interactions explains why Acetophenone_56_benzaldehyde_83 has the worst binding affinity.

Figure 5.8 shows the protein residues and the best and worst ligands in the active site of 1JEP, revealing the spatial filling of the active site by the ligands. The best binding ligand in 1JEP is in favourable proximity to a host of residues around the binding pocket (Figure 5.8a). The distances between the benzaldehyde component of the chalcone and the amide carbonyl carbon for proximate residues are: 3.9 Å from Ala116; 4.6 Å from His173, and 3.9 Å from His195. Other residues surrounding this portion of the chalcone ligand include Ser120, Glu189, and Thr190. Of particular interest on this moiety is the *ortho*-OH substituent on the benzaldehyde component which lies 3.7 Å from Asn113. On the other hand, the portion of the chalcone formed from the acetophenone precursor is surrounded by residues Arg36, Gly37, Leu38, Leu101, Glu105, Tyr106, Lys109 and Val110. Much of this interaction is hydrophobic in nature, including the interaction with Lys109 where the interaction is with the aliphatic portion of the Lys109 sidechain, up to but not including the terminal NH_4^+. The best binding ligand with 1JEP is, in fact, in relatively close proximity to approximately 21 residues of the protein in the active site. The worst binding ligand in 1JEP (Figure 5.8b) is in close proximity to about 15 residues of the protein in the active site. This ligand is: 3.1 Å away from Gly37, 2.0 Å from Leu101 (both involving unfavourable van der Waals' interactions and raising binding energy);

1.9 Å from Tyr106; 3.8 Å form Thr190; and 2.7 Å from Ser198. There is also a hydrogen bond of 2.0 Å between the ligand oxygen of the ligand and Asn113 NH_2. Similar analyses were made for the best and the worst binding ligands for the 11 CHI crystal structures, and the figures showing the ligands and the residues in the active site are given in the Appendix G (Supplementary material in the disc).

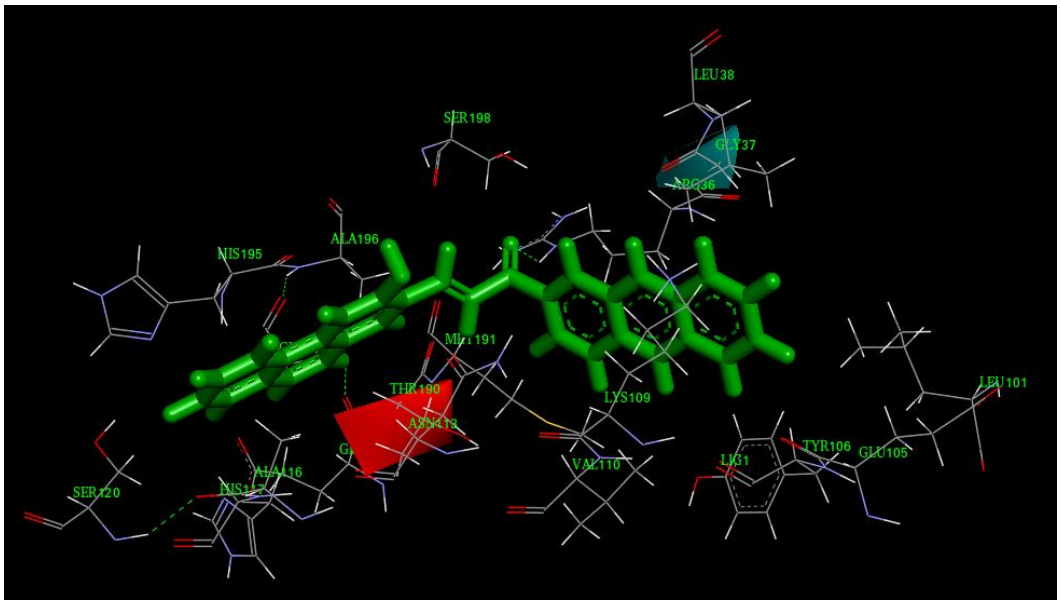


Figure 5.8a. 1jep_Acetophenone_15_bezaldehyde_44 – ligand with best binding affinity

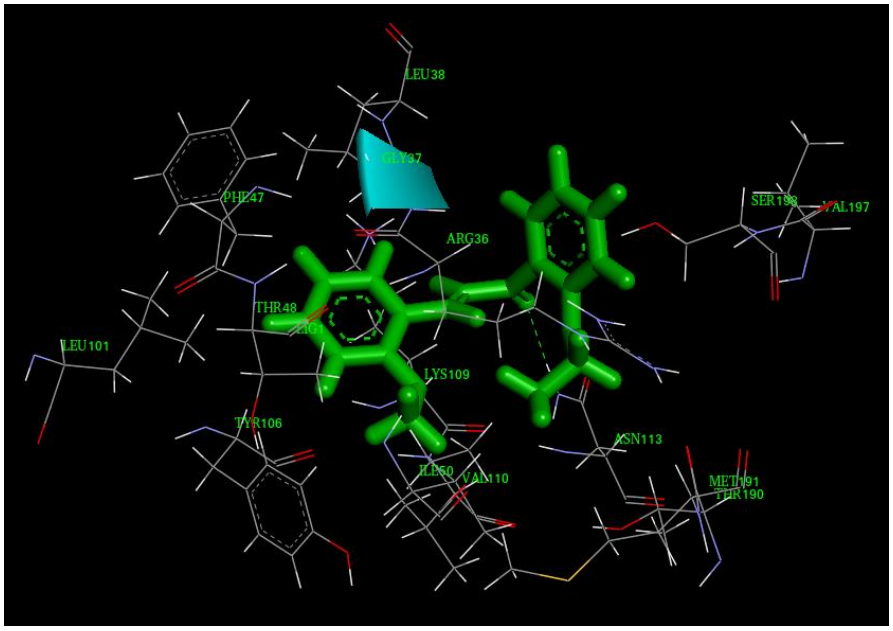


Figure 5.8b. 1jep_Acetophenone_57_bezaldehyde_30 – ligand with worst binding affinity

Figure 5.8. Active sites for 1JEP CHI complexes.

The results of these analyses are all quite similar. For the best CHI-chalcone complexes a large part of the ligand occupies the docking pocket and the ligand is held with multiple receptor-ligand interactions. Most of the best CHI-chalcone complexes are comprised of a benzaldehyde moiety with multiple benzene rings which participate in the hydrophobic π- interactions. On the other hand, the worst binders are the ligands that are not located in the active site, or for which there are a reduced number of these particular interactions with the active site.

CHI-chalcone complexes with better binding affinity seem to be influenced by the size of the chalcone – in particular the substituents originating from the benzaldehyde precursor (the smaller this part of the chalcone, the worse the binding ΔG. The presence of additional OH or esters groups does not significantly improve the binding ΔG (*e.g.* acetoph_53_benzald_53 Figure 5.2).

CHI-chalcone complexes of 1EYP and 1EYQ were identical in terms of binding ΔG across the library. However, although the analysis of the receptor–ligand interactions show the Vina-calculated binding affinities to be the same, the interactions were not always exactly the same. With most complexes, the chalcone exhibiting the best binding affinity to a target docked directly into the active site, while most chalcones exhibiting the worst binding affinity did not dock directly to the active site. For example, the 1JX1-chalcone complex with the worst binding energy (least negative ΔG) docked on the surface of the receptor. Most chalcone ligands failed to dock in the active site of the 4dok and 4doi receptors □even the ligands with the best binding affinity for other CHIs). This can explain why on average the complexes with 4dok and 4doi have the lowest binding ΔG compared to the chalcone complexes with other CHI crystallographic structures.

Conventional interactions with Tyr106 were only observed with 1EYP complexes.

5.3.3 Comparison of HTVS and docking results

Comparing the docking of cinnamic acids (Chapter 4) with the chalcone docking (Chapter 5) to the 11 CHI receptors reveals the effect of ligand size on the interaction with the receptor. HTVS showed that the structural component of chalcones that enhances their binding affinities is the presence of the more than one additional benzene ring. Benzene rings fused to the chalcone moieties (both the acetophenone and benzaldehyde components) gave better binding affinity than phenyl groups linked to either moiety *via* a single bond. The cinnamic acid

derivatives considered in this study have one benzene ring. Generally, the best binding affinities of cinnamic acid derivatives are approximately -7 kcal/mol. This is quite small compared to the best binding affinities of chalcones to the same receptors (the best binding affinity with chalcones is -12 kcal/mol). From the data available we can deduce that to improve the binding affinity of cinnamic acid derivatives, more than one benzene ring should be present. These could be essential in mostly in the synthesis of phenylcoumarins. However, in the context of this study on cinnamic acid derivatives, the lower energy of binding does not imply poor interaction with the receptor, it simply reflects the small size of the ligand.

Other additional groups on the chalcone moieties such as OH, $N(CH_3)_2$, and alkyl chains reduce the binding affinity of chalcones, since they disrupt the hydrophobic interactions characteristic of good binding. This suggests that adding additional OH groups, $N(CH_3)_2$, and alkyl chains to cinnamic acid derivatives could reduce their binding affinities. The π-alkyl interactions seem to have an important role in stabilising the CHI-ligand complexes – this is observed with the chalcone ligands and the cinnamic acid ligands. 1FM7 and 1FM8 exhibited the best binding affinity and protein-ligand interactions with both cinnamic acid derivatives and chalcone derivative.

These results are of complexes where the protein was studied as a rigid system. In nature, however, proteins are not rigid systems and more information can be provided by simulating the protein in a solvated environment where all atoms are free to move in a thermally equilibrated environment. This leads on to Chapter 6 where molecular dynamics informs the evolution and binding over time of these complexes.

CHAPTER 6
MOLECULAR DYNAMICS

6.1 Introduction to the chapter

The theoretical study of the molecular interaction of chalcone isomerase enzymes (CHIs) with *o*-coumaric acid is key to exploring the putative biological function of CHIs in the interconversion of *o*-coumaric acid to coumarin. *In silico* docking permits prediction of the geometry of, and the interactions in, protein-protein or protein-ligand complexes. i.e., it gives us an understanding of how a ligand, typically a substrate or a regulator, binds to a macromolecule – a critical factor in understanding the receptor function. Although *in silico* docking provides a basis for structurally-driven drug design (123,142), most docking methods largely ignore receptor rearrangements caused by ligand docking; this is because available methods which incorporate the flexibility of the receptor during docking are computationally expensive. However, even small receptor conformational changes in the binding site induced by ligand binding can result in dramatic changes in the function of the protein (232). Furthermore, molecular docking does not account explicitly for the rearrangement of solvent molecules within the active site. The docking protocols in this study (Chapter 4) employed these traditional methods of docking a flexible ligand to a rigid receptor and not taking into account environmental conditions such as solvent effects. To remedy these shortcomings, post-docking analysis of the CHI-*o*-coumaric acid complexes was achieved using molecular dynamics simulations in a solvated, neutralized environment, to monitor the receptor-ligand interactions more realistically over a time-frame. Molecular dynamics simulations have developed in recent years into a body of techniques that can be used effectively to understand macromolecular structure-to-function relationships (233,234). This chapter analyses the evolution of the *E*-*o*-coumaric acid ligand in the binding site of CHIs over a set time-frame to determine if the docked *o*-coumaric acid remains in the binding site over time and to monitor other activities in the binding site, such as the motion of protein residues in close proximity to the ligand that have the potential to facilitate the desired $E \rightarrow Z$ isomerization of *o*-coumaric acid. This chapter is concerned with molecular dynamics studies of the docking of *E*-*o*-coumaric acid in CHIs [i.e., the 11 wild and mutated CHI crystal structures from the Protein

Data Bank, PDB codes: 1FM7, 1FM8, 1JEP, 1EYQ, 1EYP, 4DOI, 4DOK, 4DOL, 4DOO, 1JXO (Y106F mutant), and 1JX1 (T48A mutant)]. The full names, description and origin of these CHI X-ray crystallographic structures are given in Table 4.1 Chapter 4. The systematic outline of the preparation of the protein is outlined in Chapter 2 Section 2.2.2.

6.2 Model system and computational methods

The molecular dynamics studies were conducted using two software programs, the Chemistry at HARvard Molecular Mechanics (CHARMM) package (138,204) and the GROningen Machine for Chemical Simulations (GROMACS) package (205,206). Initial work was done using CHARMM in a smaller solvated volume for 30ns, but GROMACS was used for further work involving extension of the simulation times to 100ns within a larger solvated environment. The models that were considered for both CHARMM and GROMACS simulations were the *E-o*-coumaric acid (conformer with the *syn* arrangement) docked to 11 CHI X-ray crystallographic structures in an explicitly solvated water environment. The solvated environment in CHARMM was simulated with rhombic dodecahedron (RHDO) periodic boundary conditions with the crystal dimension of 69.0475; this space was filled with water molecules extracted from an equilibrated water box containing 46656 TIPS water molecules. In GROMACS, solvation was described under cubic periodic boundary conditions, with explicit water solvation; the box and water extending past the extent of the macromolecule for at least a further 3.0 Å in all three dimensions. Visualization and analysis of the results was performed using VMD (209). The outline of the detailed computational approach to the CHARMM and GROMACS simulations is given in Chapter 2 section 2.2.4.

6.3 Results

The results in this chapter are divided into two sections. The first section (Section 6.3) covers the initial results from the use of CHARMM calculations, and the second section (Section 6.4) covers the GROMACS results. In each section, the results are discussed cumulatively for each of the structural properties considered (i.e., Root Mean Square Deviation, Hydrogen-bonding, etc.) for complexes with the 11 CHI crystal structures for CHARRM and 8 CHI crystal structures for GROMACS (7 CHIs from *M. sativa* and 1 CHI from *A. Thaliana*).

6.3.1 CHARMM Results

6.3.1.1 RMSD

Root Mean Square Deviation (RMSD) is commonly used in molecular dynamics because it gives a quantitative measure of the similarity between superimposable atomic coordinates (235,236). RMSD is defined using Equation 6.1.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N_{atoms}} \left(r_i(t_1) - r_i(t_2)\right)^2}{N_{atoms}}} \qquad (6.1)$$

In Equation 6.1, N_{atoms} refers to the number of atoms whose positions are being compared (235,236). The RMSD of the protein in the protein-ligand complexes informs the conformational evolution of the protein over the dynamics time-frame, referenced to the initial protein conformation (235,236). Similarly, the RMSD of the ligand in the protein-ligand complex reflects the conformational evolution of the ligand in the docked site of the protein with respect to the initial docking conformation. Both sets of RMSD data also provide information about the simulation equilibration (the fluctuation of the RMSD towards the end of the simulations must be around some averaged thermal structure) (235,236). The ideal RMSD fluctuations of a thermally equilibrated system is about 1-3 Å for small globular proteins. RMSD analysis was considered in this study because it could provide primary information about the conformational evolution of both the CHI and the bound *o*-coumaric

acid. Since the RMSD provides information relating to equilibration of the system, the RMSD analysis can also be used to determine the thermally averaged structure of a CHI- *o*-coumaric acid complex.

Figure 6.1(a-k) gives the RMSD of the 11 CHI-*o*-coumaric acid complexes over 30 ns simulations; for example, Figure 6.1a shows the RMSD of the 1EYP-*o*-coumaric acid complex. This figure shows that the receptor 1EYP thermally stabilizes within *ca.* 3 ns with an RMSD (from the original protein conformation) of 2-2.5 Å. The RMSD of both 1EYP and *o*-coumaric acid fluctuate within the acceptable literature values of 1-3 Å. The RMSD of the ligand shows that although the ligand is constrained within the binding site, for a short period of time at the start of the simulation, there are abrupt fluctuations (between 0 and 1.5 Å in ligand RMSD values) from about 6 ns through to the end of the simulation. In this case, however, the fluctuations do not signal that the ligand has left the active site. The binding site is quite close to the surface of the protein, and observation of the simulation shows that the ligand is affected by the Periodic Boundary Conditions (PBC), and abrupt changes are due to parts of the ligand crossing this boundary. The 1EYQ-*o*-coumaric acid RMSD (Figure 6.1b) shows that the 1EYQ complex also thermally stabilises around 3 ns, with the overall fluctuation much the same as for the protein in the 1EYP complex. The high ligand RMSD fluctuations are again related to the periodic boundary conditions.

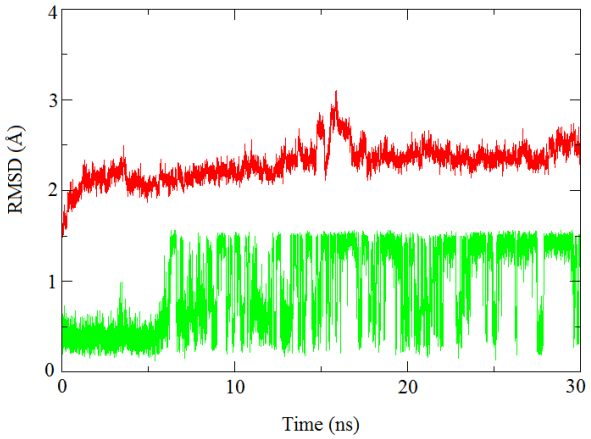
There is an interesting contrast in the behaviour between the related 1FM7 and 1FM8 complexes (Figure 6.1c and d). The 1FM7 complex reaches thermal stability within *ca.* 3 ns, and the ligand's high RMSD fluctuations are not due to PBC issues but to the fact that the ligand is observed to leave the active site during the course of the simulation. Although good binding was observed (the best binding energy of all complexes!) during the docking experiments, the dynamics data suggest that the coumaric acid may not remain in the active site long enough to be affected by 1FM7 catalytic activity. On the other hand, the 1FM8 complex exhibited almost as good binding during molecular docking as the 1FM7 complex but, during molecular dynamics, it takes 7 ns to reach thermal stability. However, after reaching thermal stability, the ligand remains within the active site. Considering that these were the two best candidates for catalytic isomerization of *o*-coumaric acid from the analysis of the docking data, the conclusion is that 1FM7 may not be the best choice for any bioorganic activities involving *o*-coumaric acid, while 1FM8 still remains a viable possibility.

Figures 6.1*e*-g also show differences in the behaviour of *o*-coumaric acid when bound to 1JEP, 1JX0 and 1JX1. Binding duration in the active site) varies. For 1JEP (Figure 6.1f), the ligand remains within the active site for the full duration of the simulation. For 1JX1 (Figure 6.1f), the ligand remains close to the binding site for most of the simulation, only leaving the active site towards the end of the simulation. It may be recalled that the docking of coumaric acid to 1JX1 resulted in a complex with the ligand located only partially within the active site, and the movement of ligand away from the active site is reflected in the increase in the ligand RMSD fluctuations; of course the RMSD is also ultimately affected by the PBC. Figure 6.1g reveals that the 1JX0 complex does not reach thermal stabilisation within the 30 ns of simulation, and this has possible implications with respect to the destabilization of the protein by the ligand.

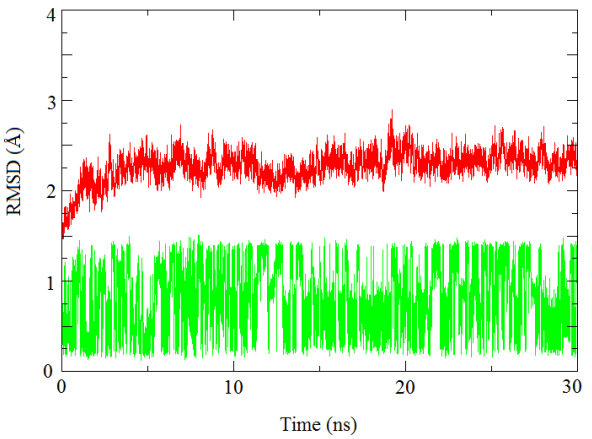
Figures 6.1h-k show the RMSDs during dynamics of the docked complexes of CHIs from *A. thaliana*. The results discussed in Chapter 4 show that these ligands do not dock in the active site. MD simulations of these CHI complexes were undertaken to explore whether dynamic simulation could provide more insight into the region surrounding the active site. The results showed that with 4DOI, 4DOL, and 4DOO the ligand remained within the protein, but no migration towards the active site was observed. The RMSD data in Figure 6.1k, however, shows that the ligand drifts away from its original docked position in the 4DOK active site during the simulation. The evidence from both the molecular docking and the molecular dynamics studies supports the conclusion that cinnamic acids are not accommodated as substrates in the *A. thaliana* CHIs (4DOI, 4DOL,4DOO and 4DOK).

In general, the RMSD results point towards 1FM8 being the most favourable CHI for any possible enzymatic activity with *o*-coumaric acid. There is further evidence to support this conclusion. For example, when comparing the complexes with 1FM8 (Figure 6.1d) and 1JEP (Figure 6.1e), in which the ligand remains bound during simulation, it is clear that the stability of the 1FM8 protein, as judged by the RMSD, is much greater than that for 1JEP which exhibits a slow increase in RMSD during the simulation. Moreover, the ligand RMSD in the 1FM8 complex is stable around 0.5 Å significantly lower than in the case of the 1JEP complex where the ligand RSMSD is stable at around 1.5 Å., This indicates good congruence between the 1FM8 *o*-coumaric acid docking and 1FM8 complex molecular dynamics data, not seen for the 1JEP complex. In contrast, 4DOO (Figure 6.1k) also shows this steady RMSD, but docking is not in the active site. The choice of which protein is to be used for further studies cannot, however, be based solely on the RMSD, other structural properties, such as the hydrogen

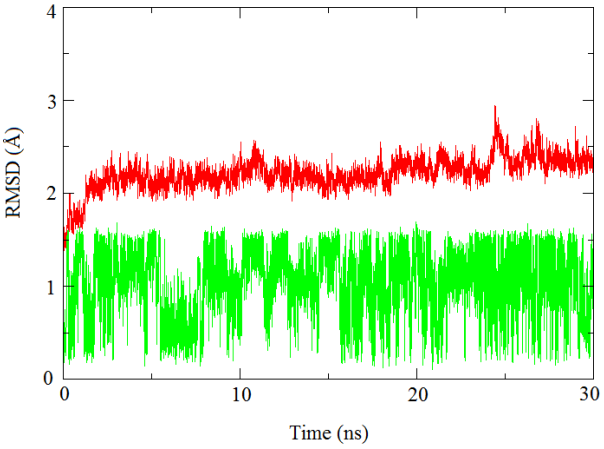
bonding, the position of Tyr106 and other protein-ligand interactions that may influence the enzymatic activities of the CHI must also be taken into account. The evidence at this stage indicated that the 1FM8 structure was the best available model for the proposed theoretical of enzyme-catalysed isomerization of *o*-coumaric acid.



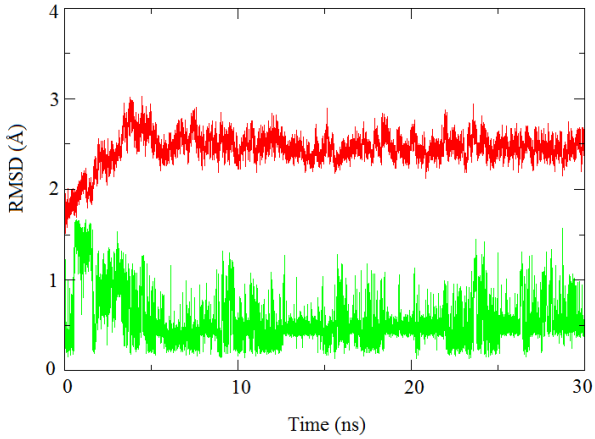
a) 1EYP



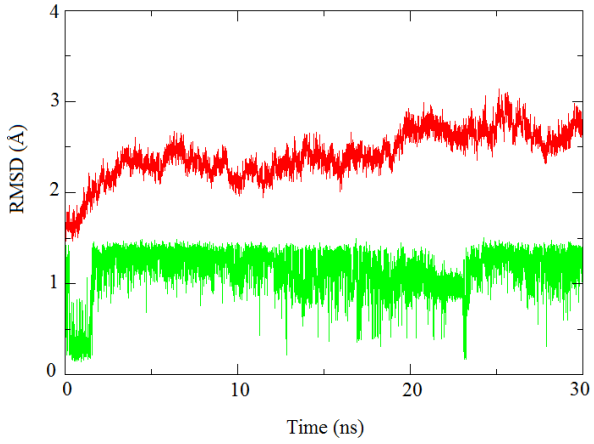
b) 1EYQ



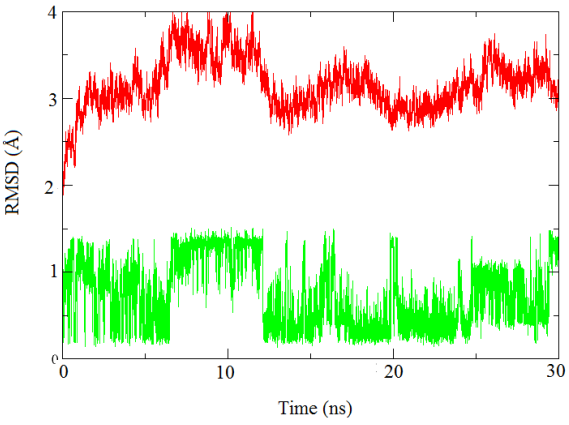
c) 1FM7



d) 1FM8



e) 1JEP



f) 1JX0

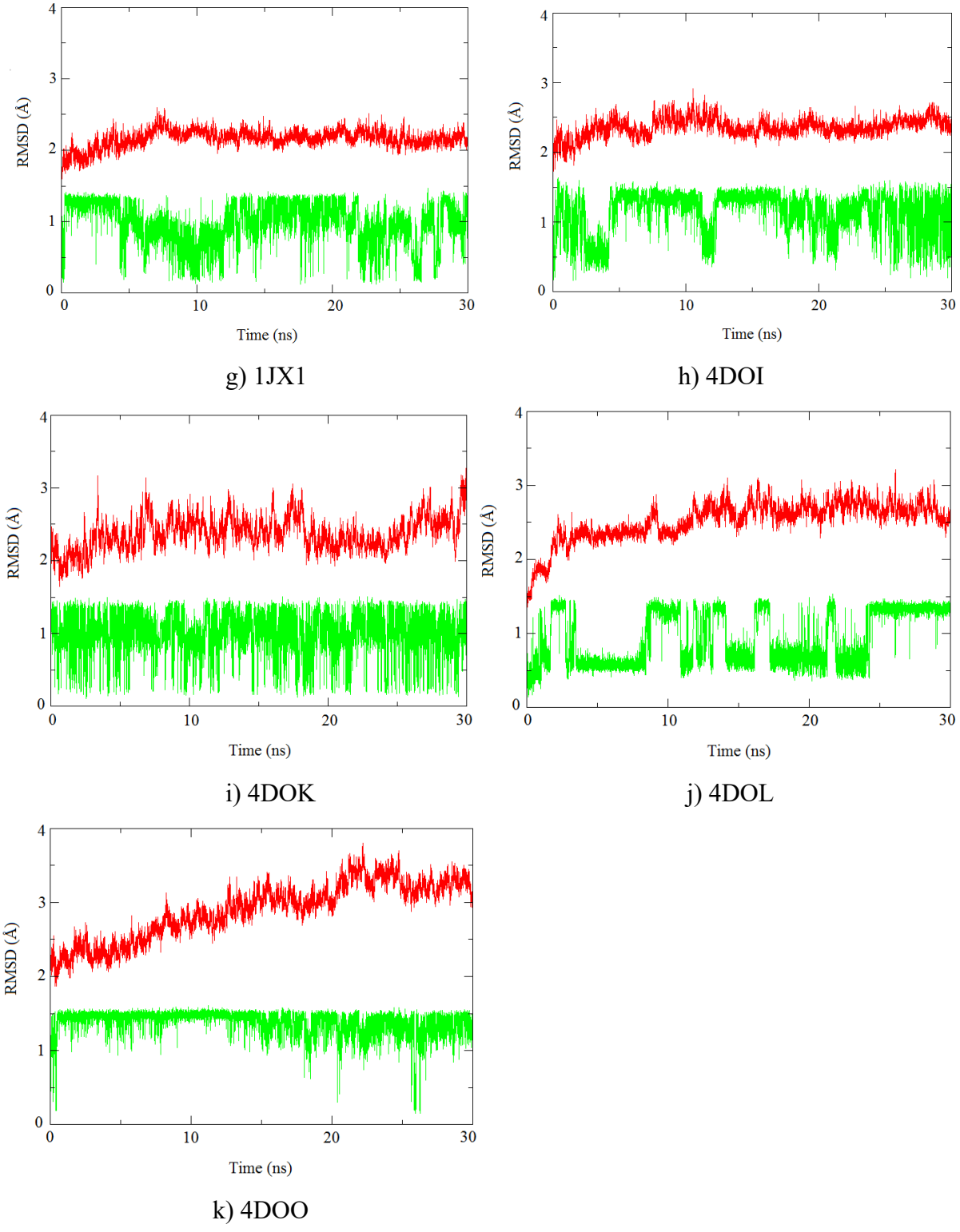


Figure 6.1. Protein and ligand RMSD plots for CHARMM MD studies of 11 CHI-*o*-coumaric acid complexes. Protein RMSD (——) and ligand RMSD (——).

6.3.1.2 Hydrogen-bonding patterns

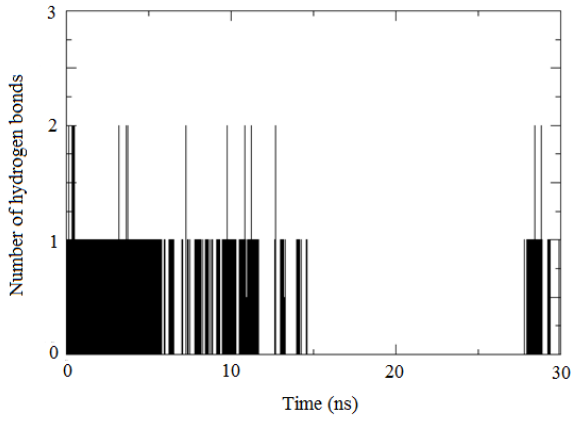
When a hydrogen atom, which is covalently bound to an electronegative polar atom (such as oxygen, nitrogen or any halogen), interacts with an electron pair from another of these electronegative atoms the resulting interaction is referred to as hydrogen-bonding (237–239). When studying the activities of a particular enzyme it is important to take into account the hydrogen-bonding (H-bonding) because of the critical role of these types of bonds in biological systems. Hydrogen-bonds are credited for increasing the probability of binding of a ligand to a receptor, by replacing the hydrogen-bonds already present between the receptor and solvent molecules (237–239). They are also said to modulate molecular function, but there are many cases where the exact manner in which H-bonding modulates molecular function requires further study (237–239).

In this study, H-bonding is proposed to play a critical role in the putative conversion of *o*-coumaric acid to coumarins. The study is guided by a proposed mechanism from the literature in which an enzyme converts *o*-coumaric acid to coumarins, in a process initiated by the protonation and deprotonation of *o*-coumaric acid to achieve an *E*→*Z* isomerization. This situation would only be achieved with precursive hydrogen-bond donation and reception by the *o*-coumaric acid. The previous chapters (and the first part of this chapter) have justified the use of the CHI PDB structure 1FM8 as a suitable enzyme for this purpose. Therefore, this section is concerned with the analysis of H-bonding in CHI-*o*-coumaric acid complexes during dynamics and, in particular, the H-bonding required for the protonation and deprotonation which are proposed to facilitate isomerization, with a particular focus on 1FM8.

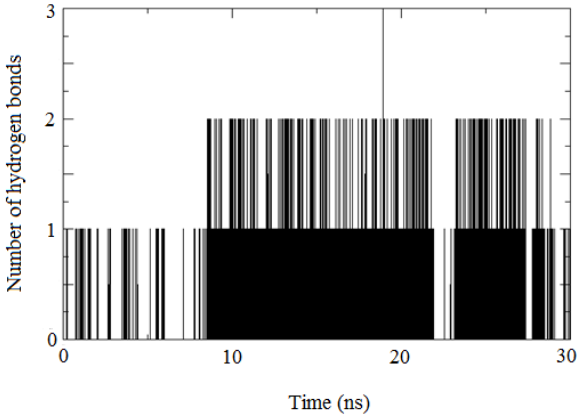
Figures 6.2a-k illustrate the H-bonding tendencies of CHI-*o*-coumaric acid complexes during dynamics simulation. The previous analysis of the RMSD (Section 6.3.1) indicated that most of the complexes thermally equilibrate 3 ns into simulation and, therefore, the H-bonding interactions that are of interest are those that take place after 3 ns (i.e. H-bonds forming within thermally equilibrated structures). It is these H-bonds that are examined in terms of their properties. 1EYP (Figure 6.2a), 1FM7 (Figure 6.2c), and 4DOK (Figure 6.2h) show a flurry of ligand-protein H-bonding interactions between 0 and 10 ns, but this is lost later in the simulation. Migration of the ligand away from the binding sites is the cause of this decrease in H-bonding. 4DOI exhibits sustained protein-ligand H-bonding interactions with as many as three protein-ligand H-bonds for a significant proportion of the simulation, in contrast to the

other systems. The complexes with the other CHIs had at least 1 ligand-protein H-bond per frame above 10 ns of simulation. 1FM8 (Figure 6.2d) and 1JEP (Figure 6.2e) had at least 1 protein-ligand H-bond in more than 95 % of the frames and, in some frames, two or three H-bonding interactions were observed.

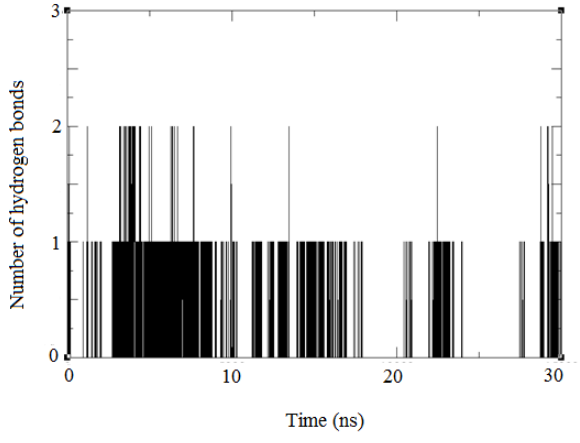
The sustained presence of one or two protein-ligand H-bonds do not exactly translate to the recurrence of the same H-bond, it only refers to the number of H-bonds present per frame. For a closer assessment of which CHI residues are involved in H-bonding with *o*-coumaric acid, a closer look at the active site within a single frame where significant H-bonding is detected provides a better understanding of the exact nature of the H-bonding.



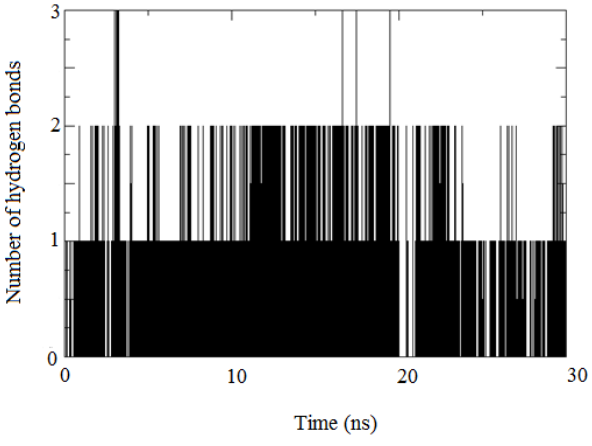
a) 1EYP



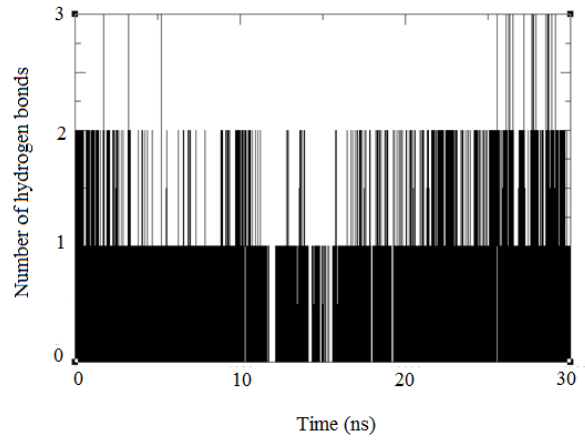
b) 1EYQ



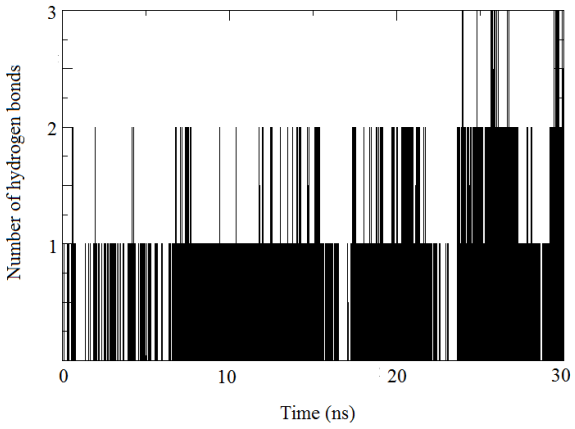
c) 1FM7



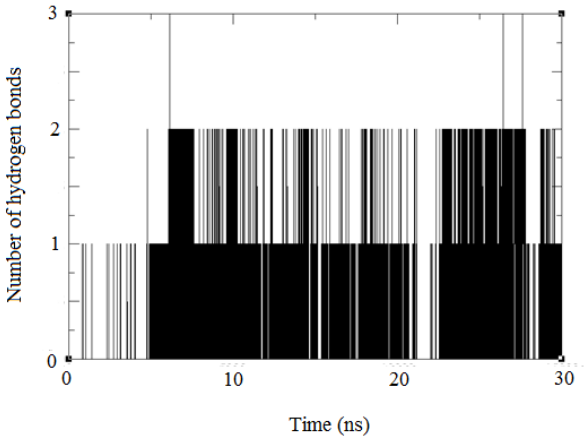
d) 1FM8



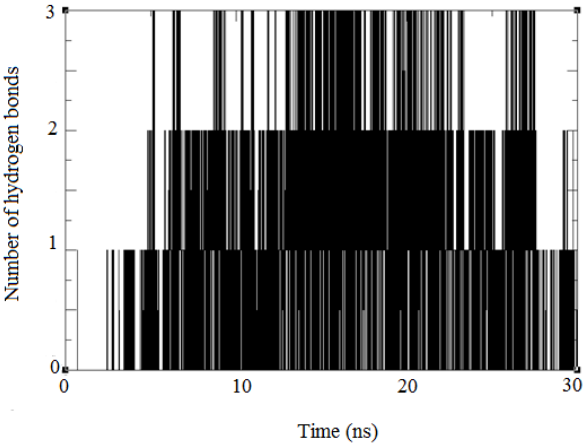
e) 1JEP



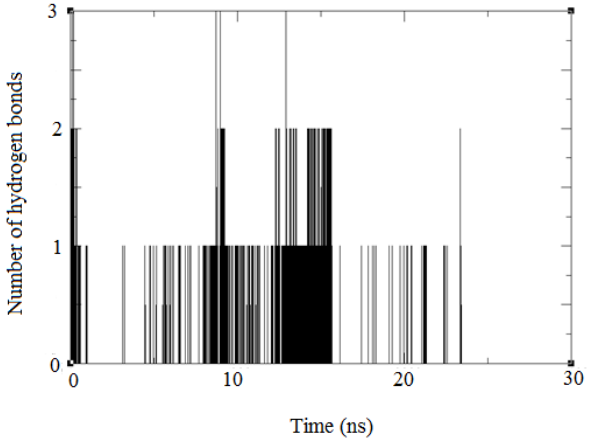
f) 1JX0



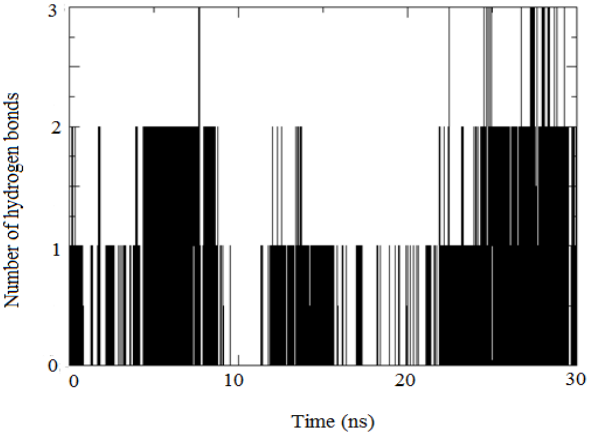
g) 1JX1



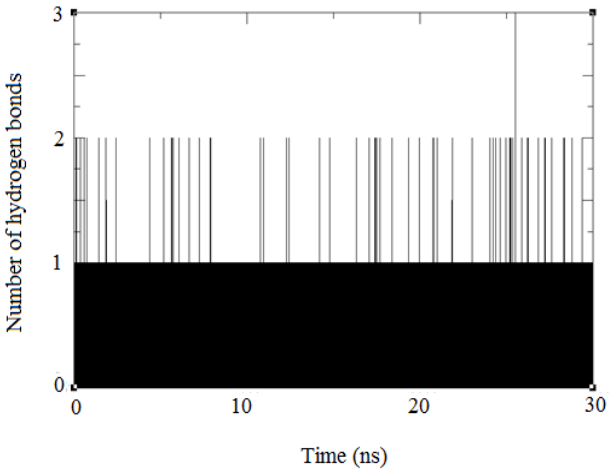
h) 4DOI



i) 4DOK



j) 4DOL

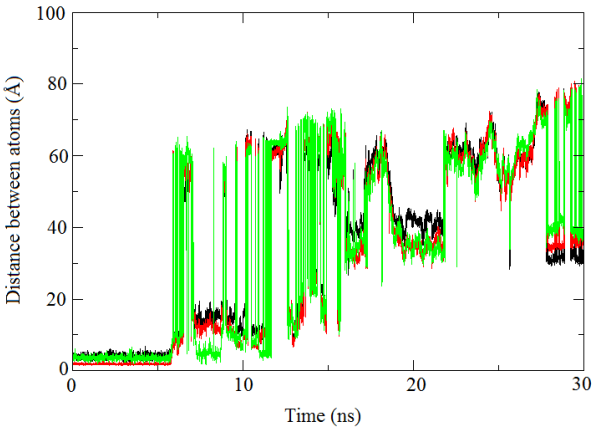


k) 4DOO

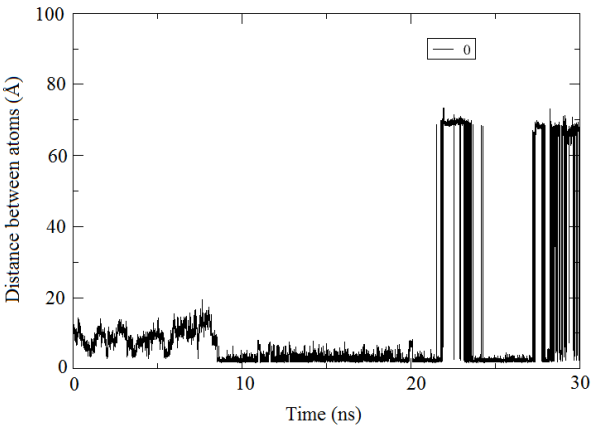
Figure 6.2. H-bonding interactions between CHIs and *o*-coumaric acid during CHARMM Molecular Dynamics Simulations.

6.3.1.3 Hydrogen-bond distance fluctuations

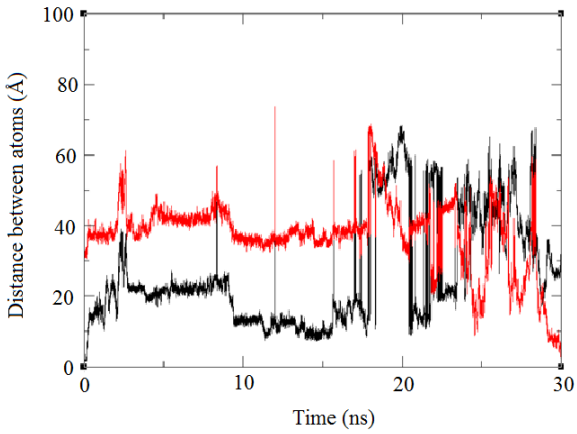
This section is concerned with identifying exactly which protein-ligand H-bonding interactions are being detected during the simulation, to see if the H-bonding fluctuations are consistent with the patterns discussed in section 6.3.2. The protein residue numbering utilized in this section is the relative amino acid numbering given by VMD. (The VMD residue numbers are those assigned in the H-bonding figures, but the residue numbering used in the discussion is the absolute residue numbering utilized in the other chapters. For example, SER4 in 1FM8 is the first residue, which is numbered in the relative VMD system as SER1). Figure 6.3 illustrates the fluctuations of the CHI-*o*-coumaric acid protein-ligand H-bonding interactions, monitored during the 30 ns MD runs. Figure 6.4 shows the arrangement in the active site in the initial frame, and the arrangement at a particular simulation frame where the H-bonding patterns are particularly favourable.



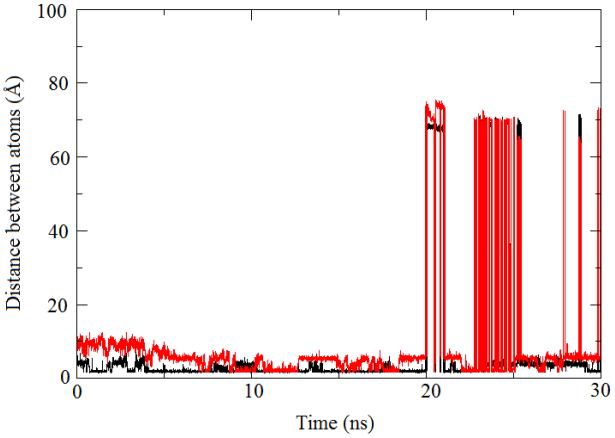
a) 1EYP



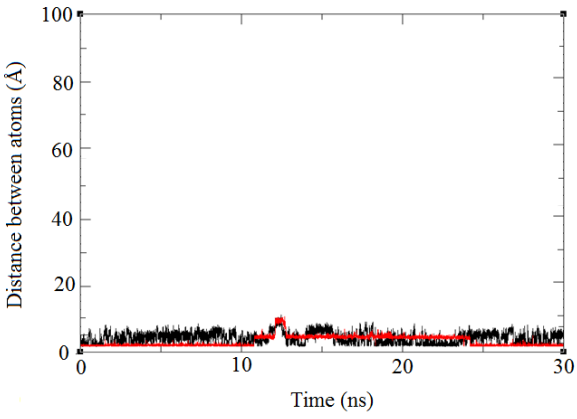
b) 1EYQ



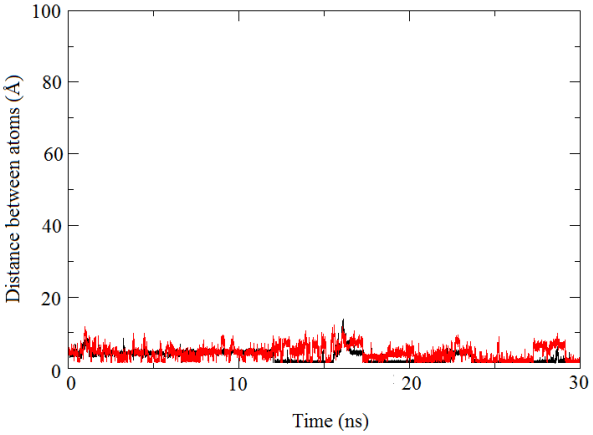
c) 1FM7



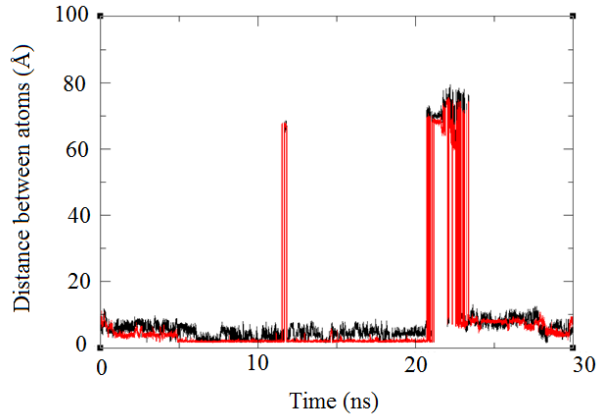
d) 1FM8



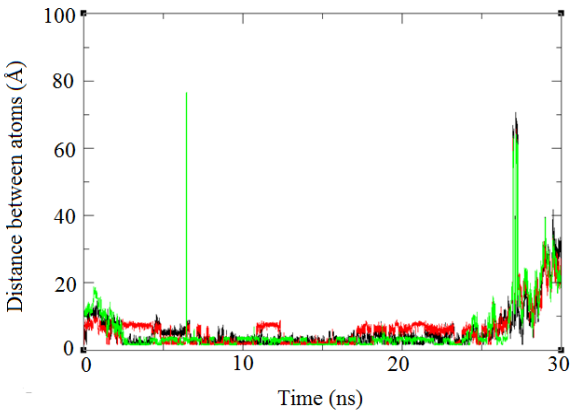
e) 1JEP



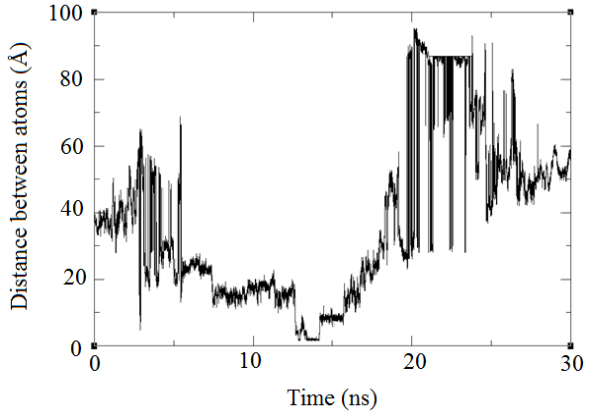
f) 1JX0



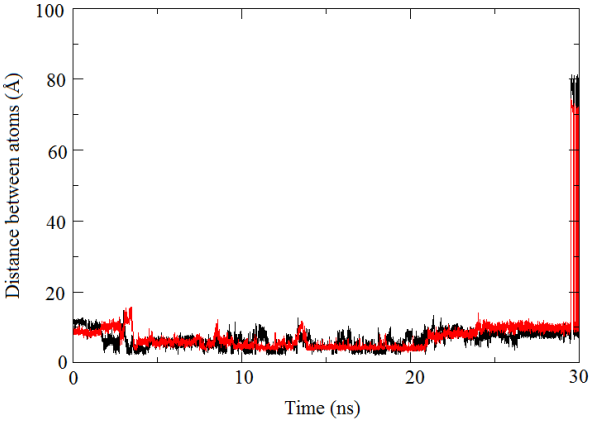
g) JX1



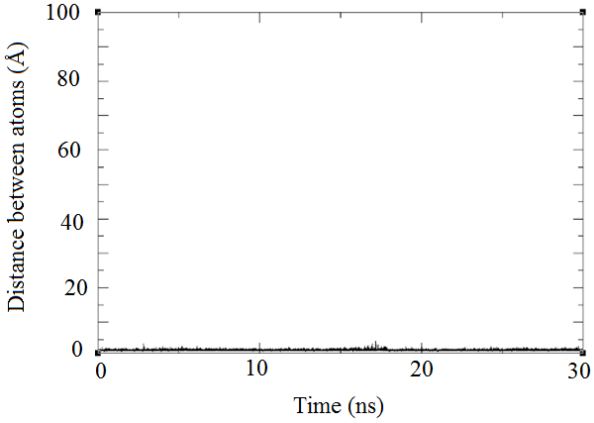
h) 4DOI



i) 4DOK

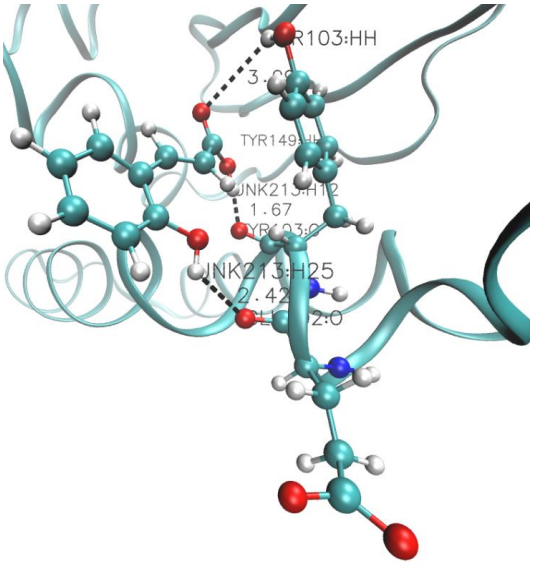


j) 4DOL

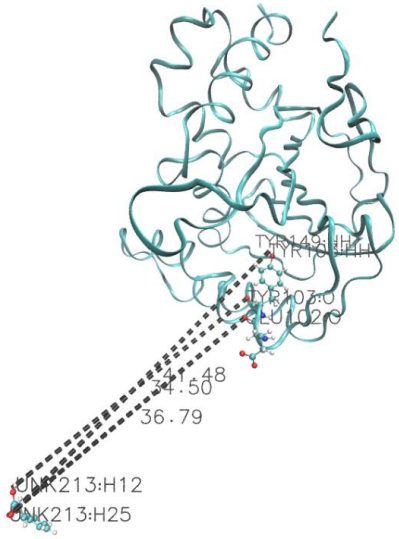


k) 4DOO

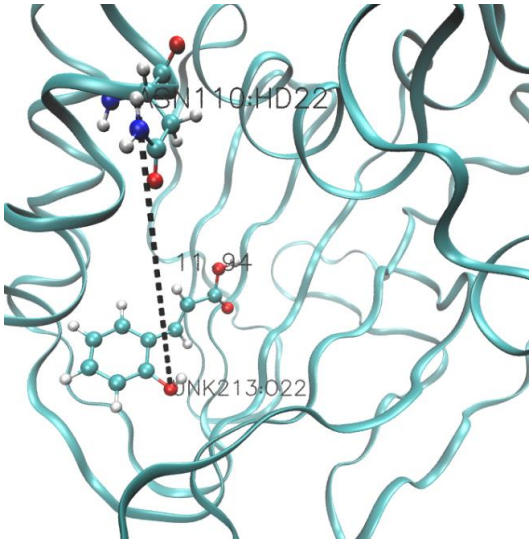
Figure 6.3. Distances between protein and ligand pairs of H-bonded atoms (as shown in the corresponding Figure 6.4 which illustrates exactly the distances measured). Discontinuities are due to Periodic Boundary Conditions (PBC).



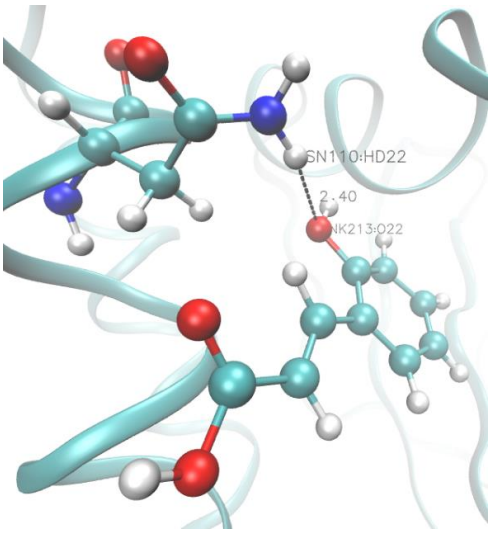
a) 1EYP 1 ns



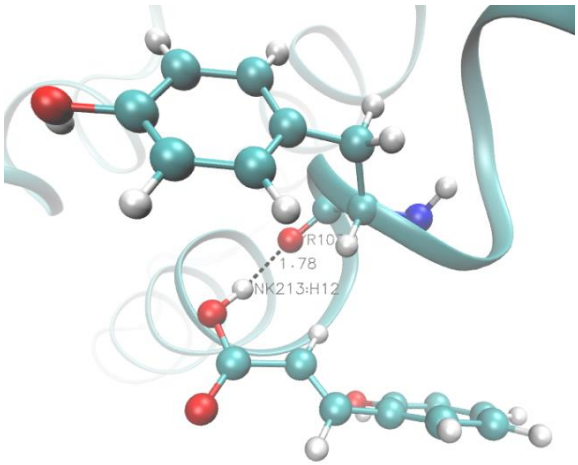
b) 1EYP step 10 ns



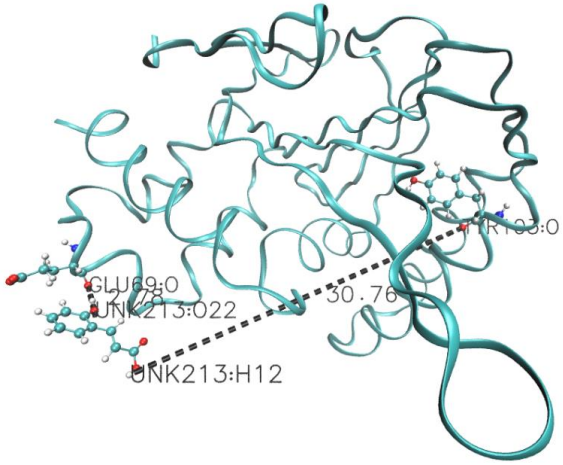
c) 1EYQ 1 ns



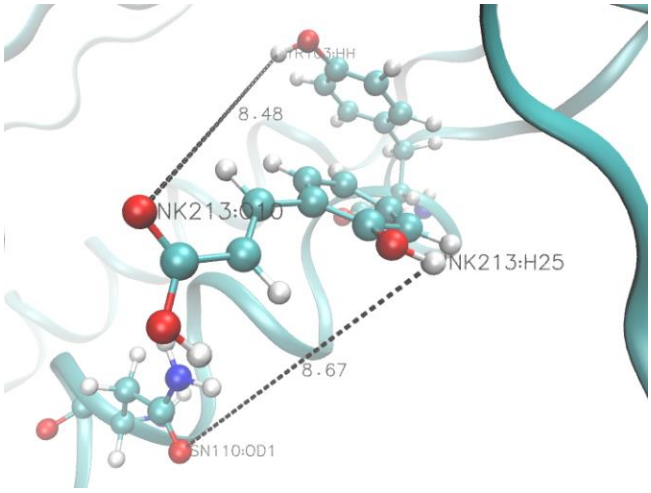
d) 1EYQ 30 ns



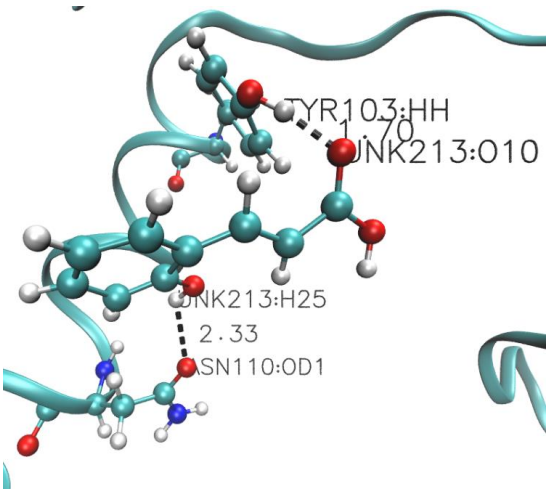
e) 1FM7 1 ns



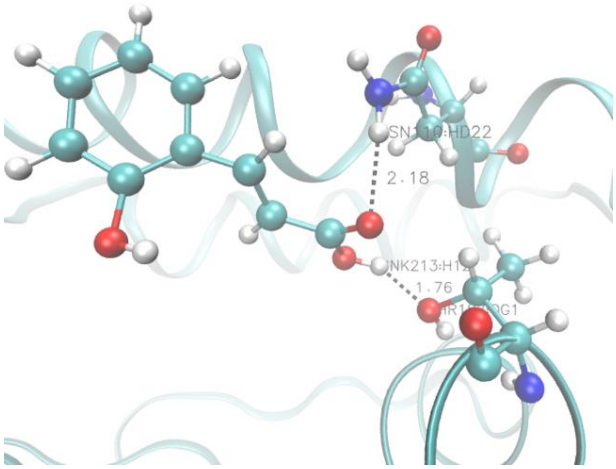
f) 1FM7 30 ns



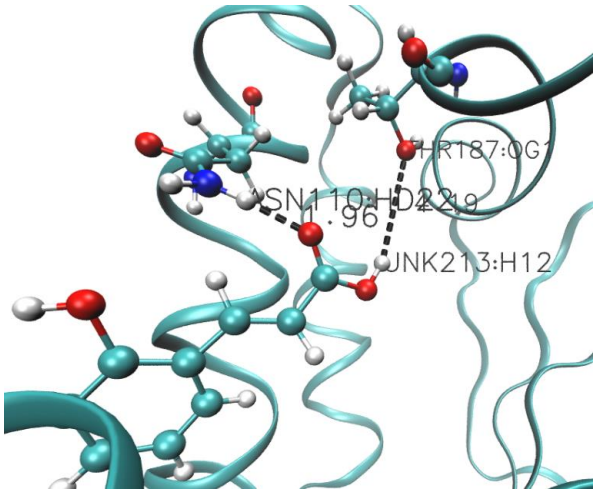
g) 1FM8 1 ns



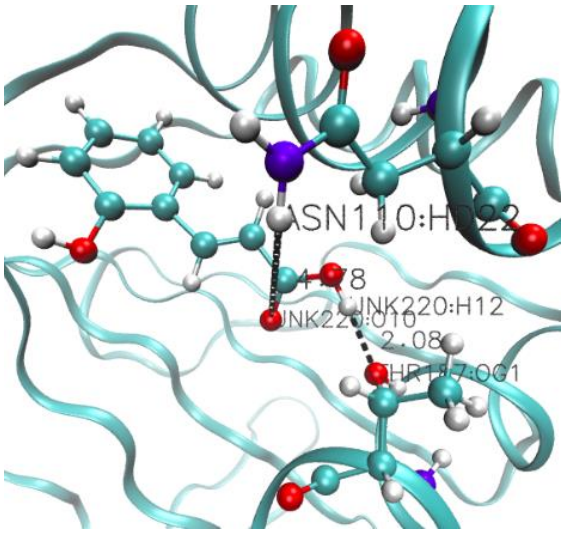
h) 1FM8 17.700 ns



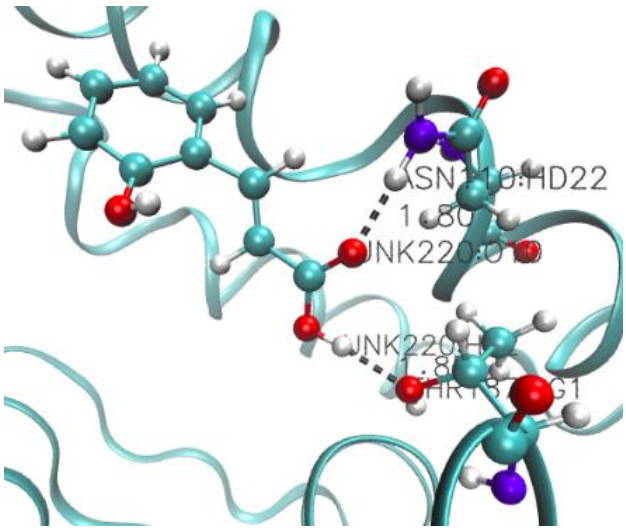
i) 1JEP 1 ns



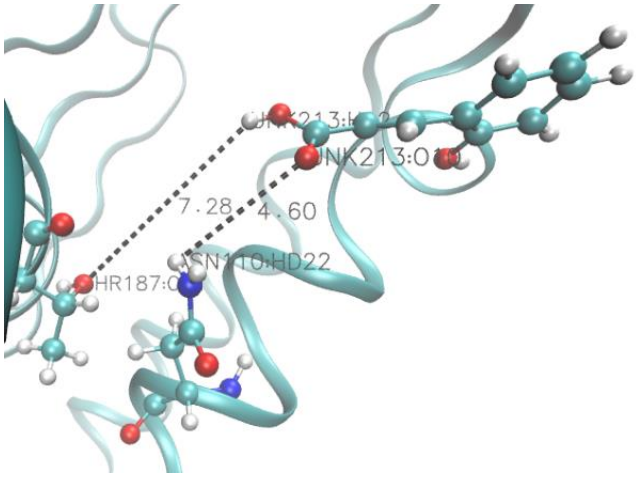
j) 1JEP 10 ns



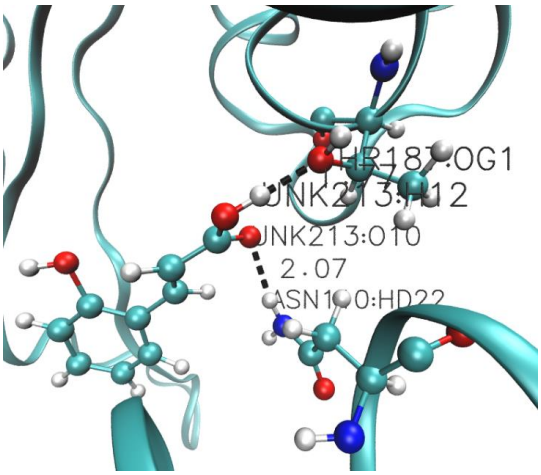
k) 1JX0 1 ns



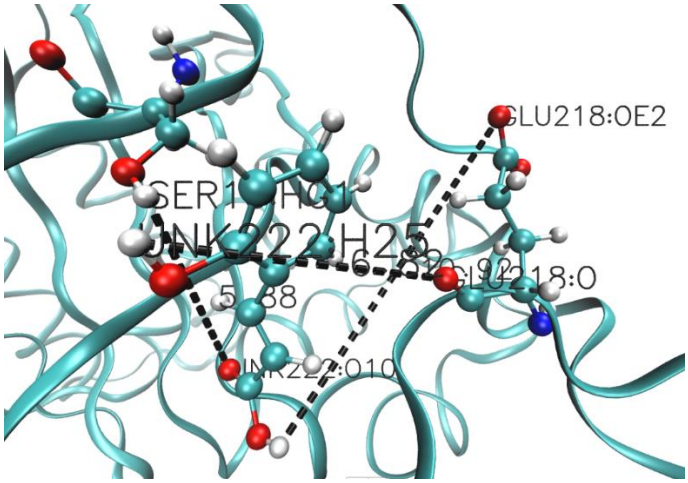
l) 1JX0 30 ns



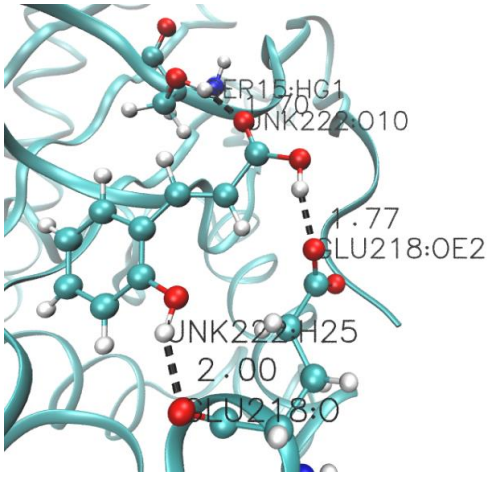
m) 1JX1 1 ns



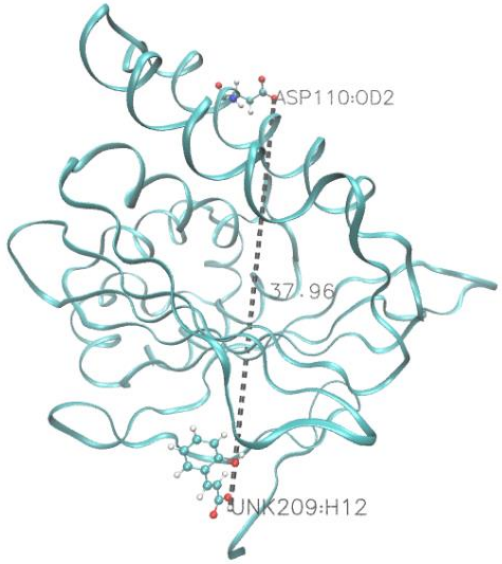
n) 1JX1 7 ns



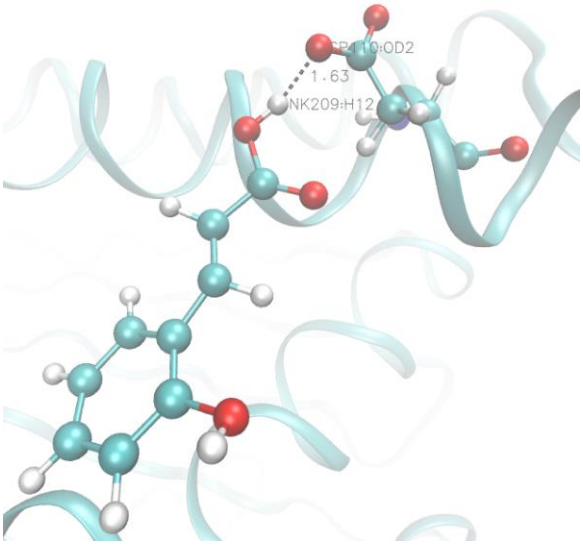
o) 4DOI 1 ns



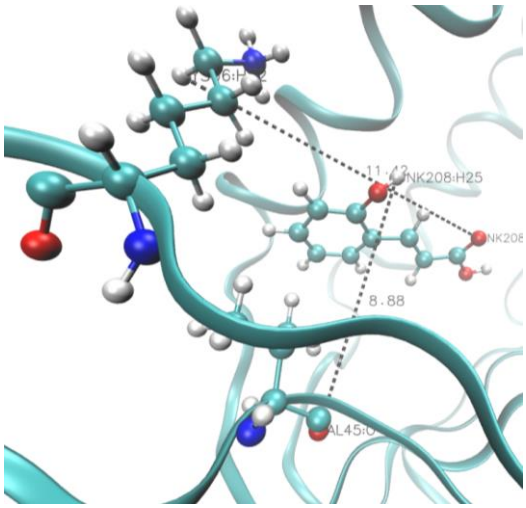
p) 4DOI 14 ns



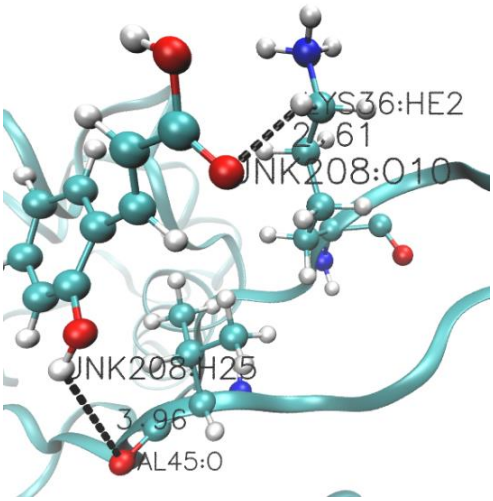
q) 4DOK 1 ns



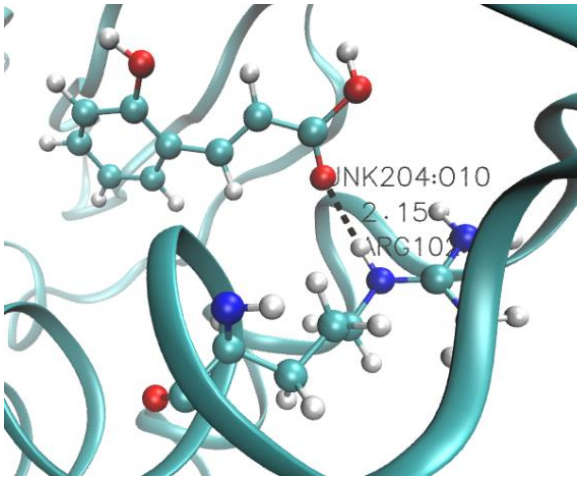
r) 4DOK 14 ns



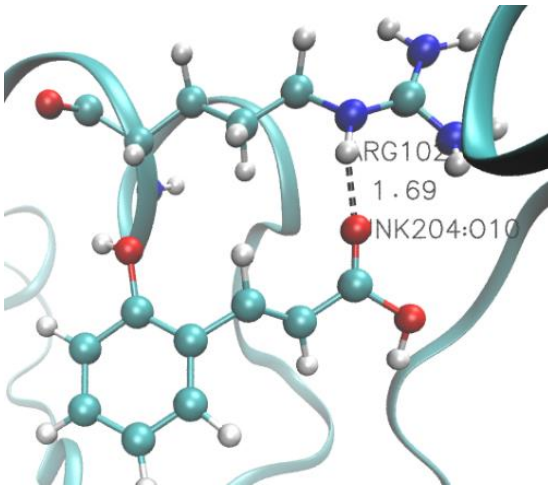
s) 4DOL 1 ns



t) 4DOL 18 ns



u) 4DOO 1 ns



v) 4DOO 7.48 ns

Figure 6.4. H-bond length measurements. The relative VMD residue numbering is used.

The H-bonding patterns of 1EYP-*o*-coumaric acid (Figure 6.2a) showed the consistent appearance of one H-bond between 0 and 10 ns and, occasionally, the simultaneous formation of two H-bonds between ligand and protein. A closer look at the active site (Figure 6.4a) to determine which residues are contributing to the H-bonding patterns observed in figure 6.2a show that, at 1 ns, *o*-coumaric acid is involved in three H-bonding interactions with the CHI; the sp^2 oxygen of the of the alkenoic acid chain of *o*-coumaric acid interacts with Tyr103 OH (3 Å apart), the OH of *o*-coumaric acid interacts with the Tyr103 sp^2 oxygen (1.67 Å apart), and the phenolic OH of *o*-coumaric acid interacts with the Glu202 sp^2 oxygen (2.42 Å apart). Monitoring of these particular H-bonds over a 30 ns MD run (Figure 6.3a) suggests that the contribution to the dominant H-bond observed in Figure 6.2a is through the interaction *o*-

coumaric acid with Tyr103 sp^2 oxygen as this hydrogen bond is generally close to 2 Å for the first 6 ns. A rare event, during this early period, is the observed reduction of the length of the H-bond between the phenolic OH of *o*-coumaric acid and the Glu202 sp^2 oxygen to about 2 Å; this accounts for the few frames showing the presence of two H-bonds in figure 6.2a, where the proximity of the ligand to protein is being measured as an additional H-bond (ligand COOH proton to receptor amide C=O). The abrupt increase of the distance between putative H-bonding centers, identified by the distance graph in Figure 6.3a, can be attributed to the ligand leaving the active site, Figure 6.4b shows the state of the system at 20 ns with the ligand having left the active site, compared with Figure 6.4a which shows the ligand at the start of the dynamics simulation.

Figure 6.2b shows more dominant H-bonding patterns for the 1EYQ- *o*-coumaric acid complex between about 8 ns and 20 ns – a period subsequent to the system reaching thermal equilibrium (Figure 6.1a). A closer look at the active site in Figures 6.4c and d reveals that the dominant H-bonds appearing from 8 ns and 22 ns are partly due to the contribution of the H-bond between the alkenoic acid OH oxygen of *o*-coumaric acid and amino H of the ASN110side chain. When this bond was monitored through to 30 ns (Figure 6.3b), the bond fluctuations showed that this H-bond is formed at around 8 ns and persists to around 22 ns. The fluctuation suggests as lightly predominant but not exclusive interaction of the alkenoic acid with the ASN110 side chain. Certainly, when the ligand leaves the active site after 22 ns, this pattern of hydrogen-bonding is lost (Figure 6.2b)

Figure 6.2c shows that, in the 1FM7- *o*-coumaric acid complex, H-bonds are scarce. A closer look at the active site at the beginning of the simulation (Figure 6.4e) shows an H-bond between the alkenoic OH of *o*-coumaric acid and the sp^2 oxygen of Tyr103; however, once the ligand leaves the active site, distance between the ligand and Tyr103 increases dramatically from 1.78 Å (at 1 ns) to 30 Å at 30 ns (Figure 6.3f. Once the ligand leaves the active site, the H-bonding patterns in Figure 6.2c can be attributed to the interactions of the ligand with numerous non-active site residues of 1FM7. For illustrative purposes, one such an interaction was observed at 30 ns; Figure 6.4f shows the H-bond interaction of the phenyl OH of *o*-coumaric acid with the sp^2 oxygen of GLU69, a residue far from the active site.

Figure 6.3d shows the H-bond fluctuations of the 1FM8- *o*-coumaric acid complex. Figure 6.2d shows the dominant appearance of two H-bonds throughout the simulation (although,

frequently, it was just one). Figure 6.4h shows the H-bond interactions between the alkenoic sp^2 oxygen of *o*-coumaric acid with the OH of Tyr103 and the H-bond interaction of the phenolic OH of *o*-coumaric acid with the sp^2 oxygen of ASN110 at frame 17.700 ns. At the beginning of the simulation, the two distances were too long to be H-bond interactions (Figure 6.4g), but their observed fluctuations (Figure 6.3d) shows that these two H-bonding interactions between 1FM8 and *o*-coumaric acid resulted from thermal equilibration of the system.

Figures 6.3e-g show the H-bond distances of interest for the complexes of 1JEP, 1JX0 and 1JX1 with *o*-coumaric acid. The corresponding H-bond patterns (Figures 6.2e-g) show that all three systems consistently have two H-bonds to the ligand, although those of the 1JEP and 1JX0 complexes are more pronounced from the beginning of the simulation (frame 1) and those of 1JX1 are scarce in the beginning of the simulation. Detailed inspection of these interactions (as was performed with the previous systems) leads to the conclusion that the contributions to the H-bond patterns revealed in Figures 6.2e-g are due to the H-bond interactions between the sp^2 oxygen of *o*-coumaric acid and an amino H of ASN110 and between the alkenoic OH of *o*-coumaric acid with the OH oxygen of THR187 for 1JEP, 1JX0 and 1JX1.

The o-coumaric acid ligand did not dock in the active site of *A. thaliana* CHIs (4DOI, 4DOL, 4DOK, 4DOO), but the H-bonding patterns (Figures 6.2h-k showed that the *o*-coumaric acid complexes with these enzymes do exhibit protein-ligand H-bonding interactions. The corresponding H-bond fluctuations are given in Figures 6.3h-k, while Figures 6.4o-v illustrate the interactions show. The evident non interaction of *o*-coumaric acid with active site residues means that these complexes were not considered for the subsequent mechanistic work.

The results of this section provide further support for the choice of 1FM8 as the most favourable CHI to continue with further studies.

The 1FM8-*o*-coumaric acid dynamics simulation also reveals other H-bonding interactions that are of interest in this study, namely, the H-bond interaction between the alkenoic sp^2 oxygen of *o*-coumaric acid and the OH of Tyr103 (Tyr 106 in literature and other chapters) and the H-bond interaction of the phenolic OH of *o*-coumaric acid with the sp^2 oxygen of ASN110. These H-bonding interactions that may well provide for the protonation and deprotonation of *o*-coumaric acid needed to achieve $E \rightarrow Z$ isomerization.

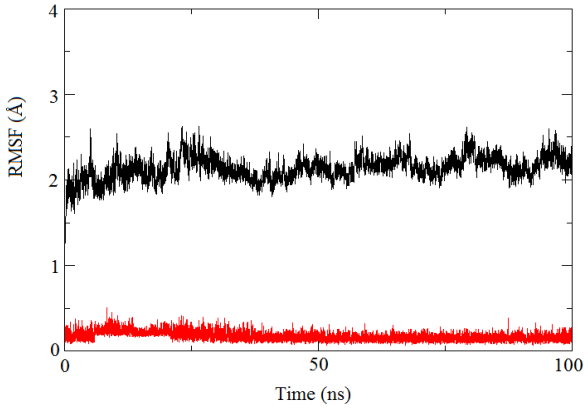
6.3.2 GROMACS results

The CHARMM simulations gave results showing the trends for the 11 CHI-*o*-coumaric acid complexes over a period of 30 ns. The water dodecahedron, although chosen well for efficiency of calculation, was not conducive to ease of analysis. Although the dodecahedron was sufficient to prevent the protein from interacting with itself across boundaries, the translational motion of the protein towards close boundaries of the PBC meant that most of the analysis was affected by the PBC, as is seen in the discontinuities. Due to the complexity of the dodecahedron, it was not easy to correct for this effect. Therefore, the CHI-*o*-coumaric acid complexes were simulated in a larger cubic water box, with boundaries at least 3 Å from the edge of the protein. The software used for this was GROMACS. A different MD simulation programme was chosen to provide diversity and for comparison, to avoid biased results. Some of the proteins (for example 1JXO) did not thermally equilibrate in the 30 ns MD simulation with CHARMM and, therefore, the GROMACS simulations were performed over a longer time-frame – 100 ns. CHIs from *M. Sativa* (PDB codes: 1EYP, 1EYQ, 1FM7, 1FM8, 1JEP, 1JX0, and 1JX1) were the primary focus with GROMACS because the ligand docked in the active site; only one CHI from *A. Thaliana* (4DOK) was considered because *o*-coumaric acid does not dock in the active site of the *A. Thaliana* CHIs. The procedure followed for molecular dynamics simulations using GROMACs has already been described in Chapter 2.

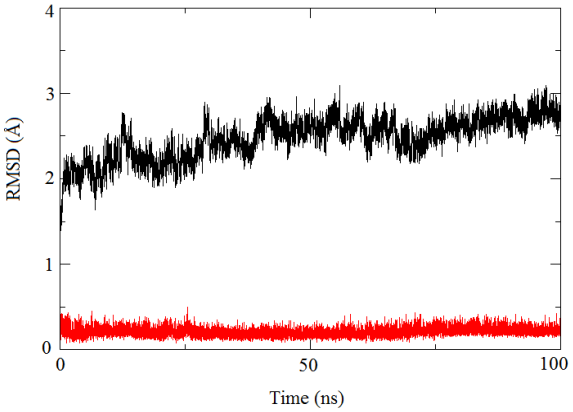
6.3.2.1 RMSD

The RMSD trends observed for the GROMACS simulation (Figure 6.6) are discussed in comparison with the RMSDs of the corresponding simulations with CHARMM. The general trends of the protein RMSDs for 1EYP, 1EYQ, 1FM8, 1FM8, and 1JEP are relatively similar to the trends of the RMSDs obtained with CHARMM. The RMSDs of all the thermally equilibrated proteins fluctuate between 2 and 3 Å with respect to the original conformation. The protein RMSDs of 1JX0 and 1JX1 in the GROMACS simulations are, however, clearly different. The RMSD of 1JX0 with CHARMM fluctuates between 2 and 4 Å, but with GROMACS it fluctuates between 2 and 3 Å (with very different equilibration times). The RMSD of 1JX1 with CHARMM thermally equilibrates around 2 Å (with respect to the original conformation), whereas with GROMACS it thermally equilibrates at about 2.5 Å. The ligand RMSD fluctuates between 0 and 0.5 Å with GROMACS, in contrast to the RMSD of the ligand

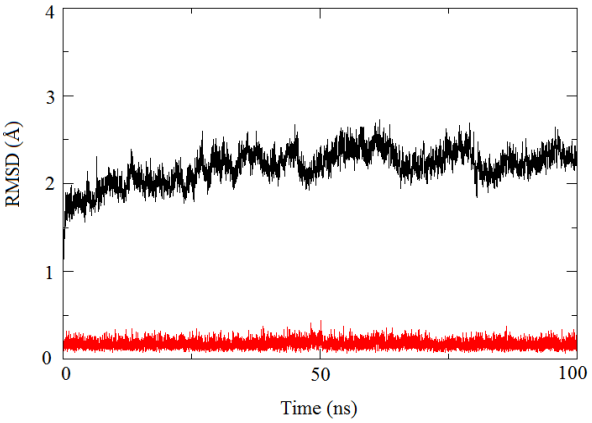
with CHARMM which fluctuated between 0 and 1.5 Å. This points to the ligand remaining in the active sites throughout the GROMACS simulations (Figure 6.6a-h). The placement of the ligand within the active site of the 1JX0 complex matches the CHARMM simulation, but the stability of the ligand in the active site of 1JX1 is different to that observed with CHARMM. However, the stochastic nature of these simulations does not preclude the possibility that the ligand could move from the active sites under the GROMACS calculation conditions.



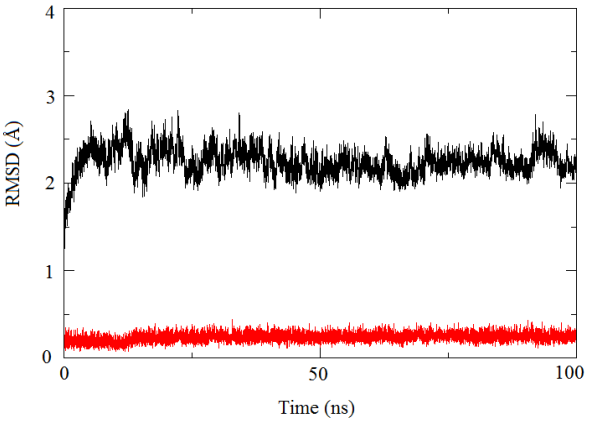
a) 1EYP



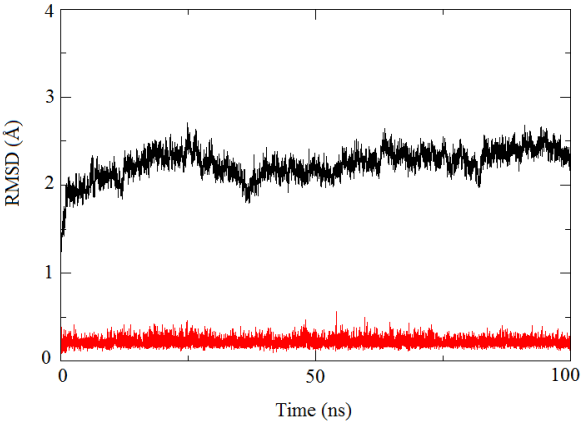
b) 1EYQ



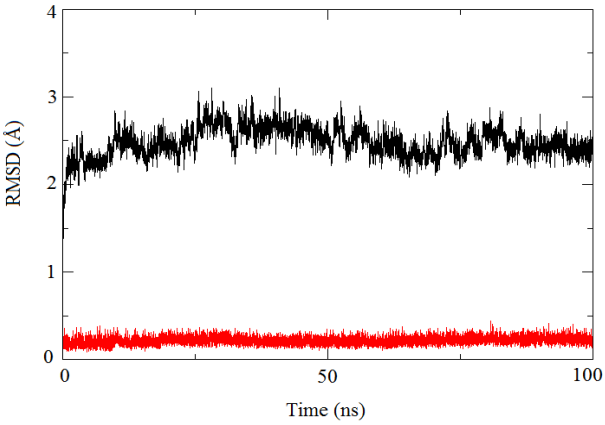
c) 1FM7



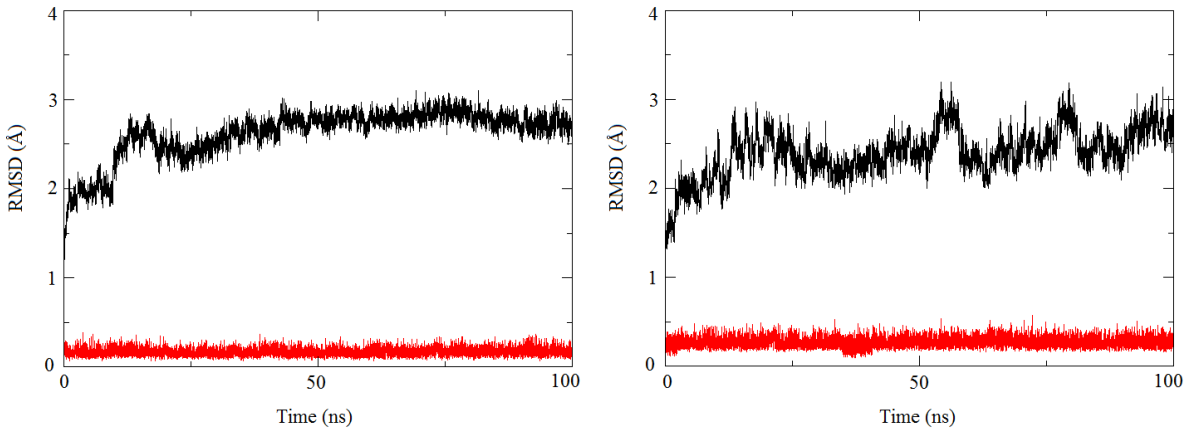
d) 1FM8



e) 1JEP



f) 1JX0



g) 1JX1 h) 4DOK

Figure 6.6. Protein and ligand RMSD plots for GROMAX MD studies of 8 selected CHI-*o*-coumaric acid complexes. . Protein RMSD ( ) and ligand RMSD ( )

6.3.2.2 RMSF

The RMSF reported in this section is per-residue RMSF averaged over time (i.e., the standard deviation of atomic positions supplied over time when fitted to the reference frame supplied with GROMACS (205,206). The RMSF generally gives information about the compactness of the protein by monitoring the flexibility of each residue. All CHIs examined in this section (with the exception of 4DOK) have similar residue RMSF behaviour. Bearing in mind that the differences between the crystallographic structures are, to some extent, due to changes in amino acid sequence (and this will filter through to the protein topology used for dynamics), but there are also significant conformational differences due to the presence of the different crystal structure ligands. The RMSF "fingerprints" of these structures are, however, relatively similar, Most of the RMSF of the residues are within 2 Å, with the exception of the RMSF of some residues around locus 50, and the residue at the terminus of the proteins; these correspond to flexible regions of the protein. While 1FM8 has been shown to have the desired trends making it a putative CHI that could be used in the *E*→*Z* isomerization of *o*-coumaric acid, the RMSF trends do not show any significant trends in the fluctuation of its residues that are unique compared to the other CHIs examined. The residues that are considered for the putative *E*→*Z* isomerization are generally between 100 and 105, and their RMSF values are generally about 2 Å across all the proteins, with the exception of 1JX1 (Figures 6.7a-h). This suggests that the active site geometry across all CHIs remains relatively constant, as expected.

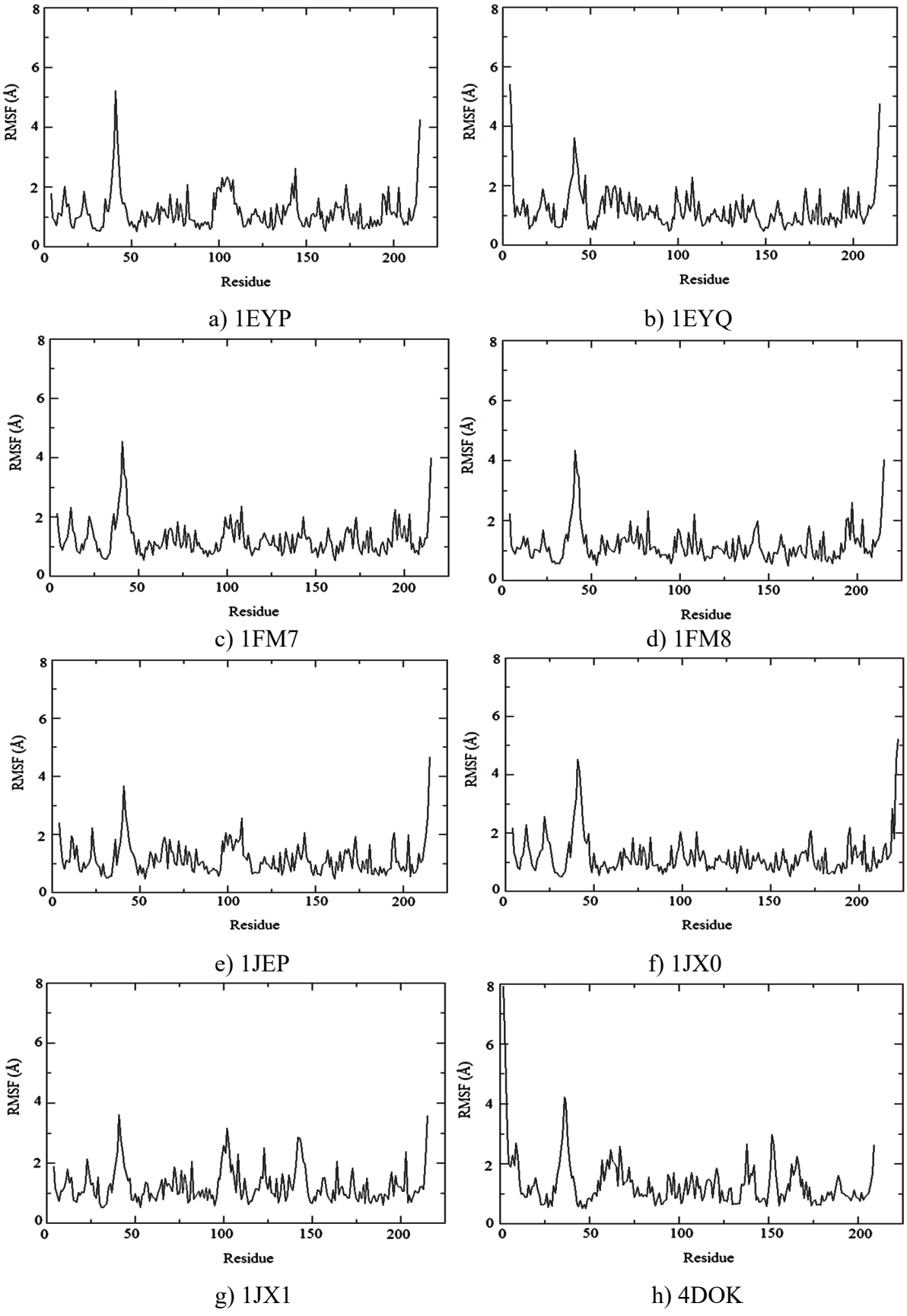
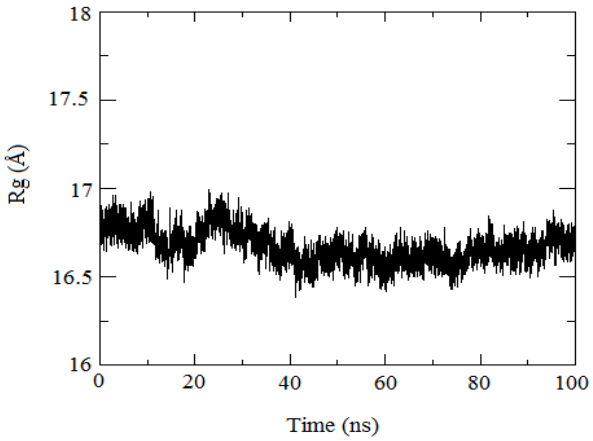


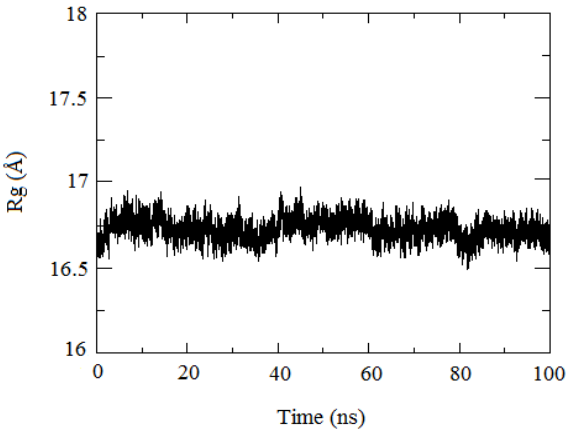
Figure 6.7. Plots of RMSF values per residue for the selected protein structures.

6.3.2.4 Radius of gyration (R_g)

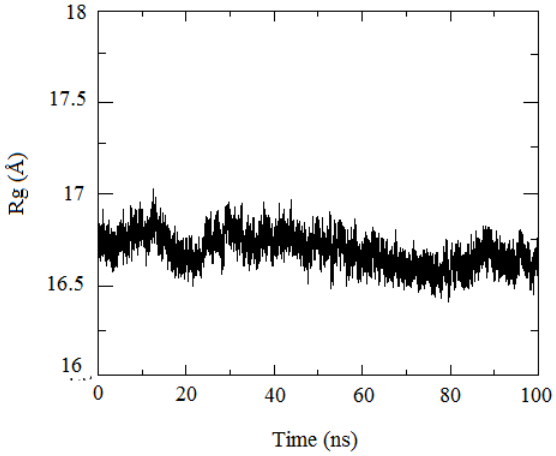
The radius of gyration (R_g) also gives more details about the compactness of the protein because it is the root mean square distance of the protein from its centre of mass or the distance of the protein from the axis of rotation, where the whole mass of the protein is assumed to be concentrated (240,241). A stably folded protein maintains a steady R_g value; an increase in R_g suggests looseness while, conversely, a decrease in R_g suggests a tightness of the protein. In general, the R_g values of all the CHIs considered in the GROMACS simulation fluctuate between 16.25 and 17.25 Å (Figures 6.8a-h). Most of these CHIs have a steady R_g, Figure 6.8 shows that 1FM8, 4DOK and 1JEP have the lowest R_g values compared with the other CHIs and, therefore, they are the most compact (the R_g values fluctuate around 16.5 Å; Figure 6.8 d, e and h). 1JX0 has the highest R_g, (fluctuating around 17 Å; Figure 6.8 f) and, therefore, is considered most loose compared to the other CHIs.



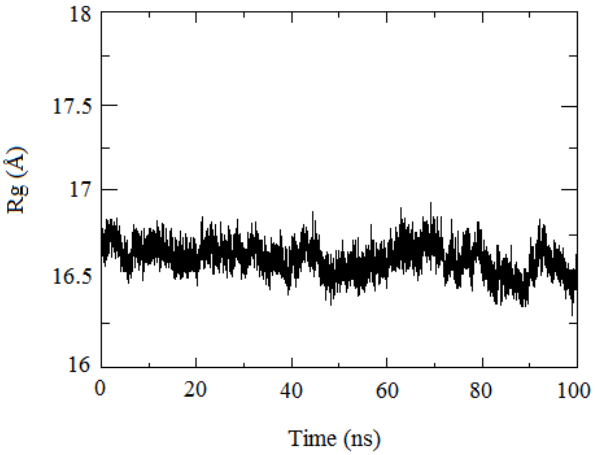
a) 1EYP



b) 1EYQ



c) 1FM7



d) 1FM8

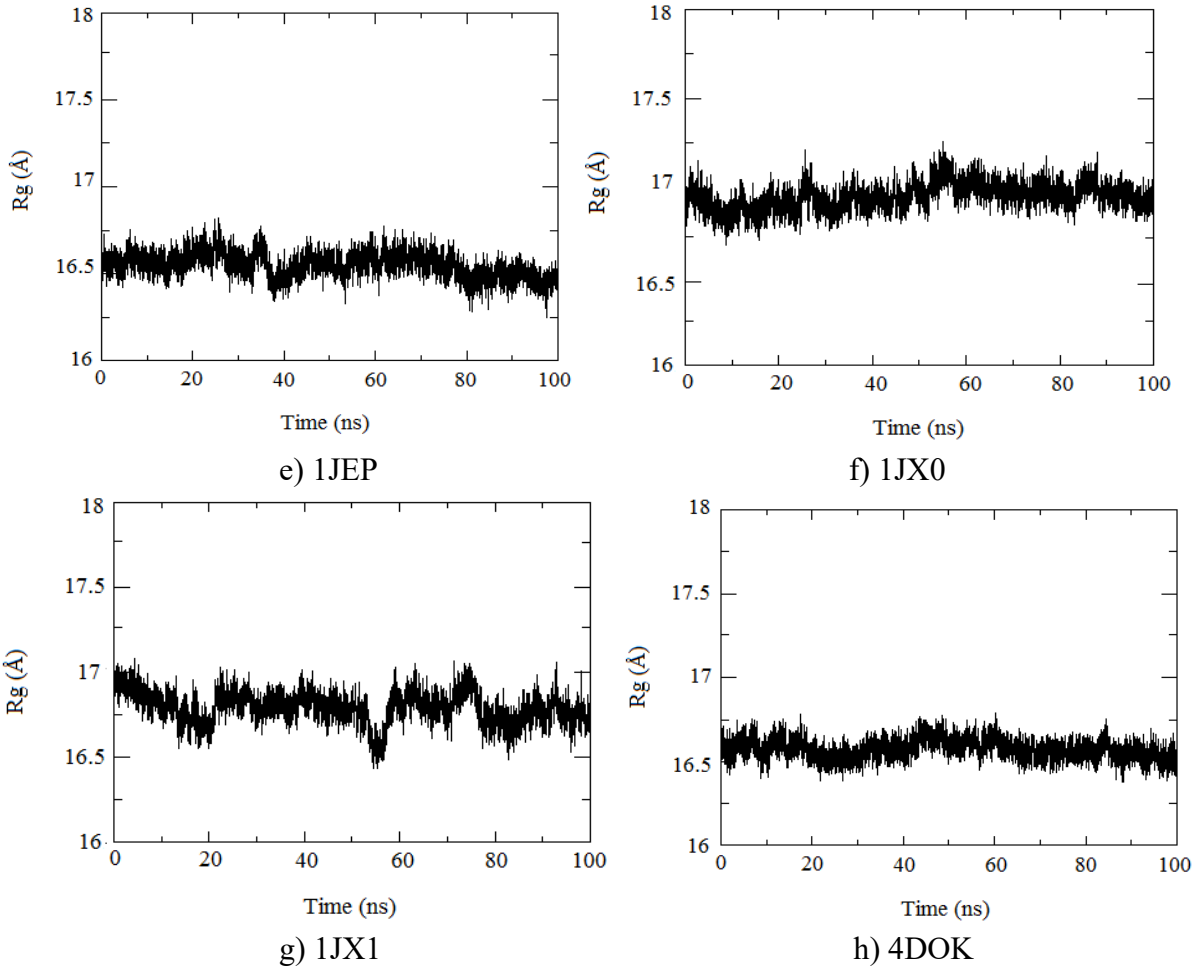
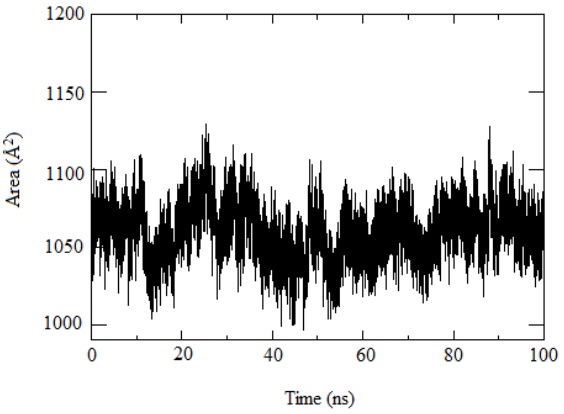


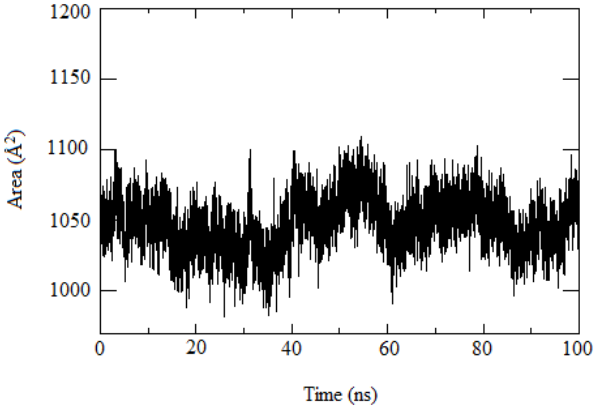
Figure 6.8. Plots of Rg values for the selected protein structures.

6.3.2.5 Solvent Accessibility Surface Area (SASA) protein

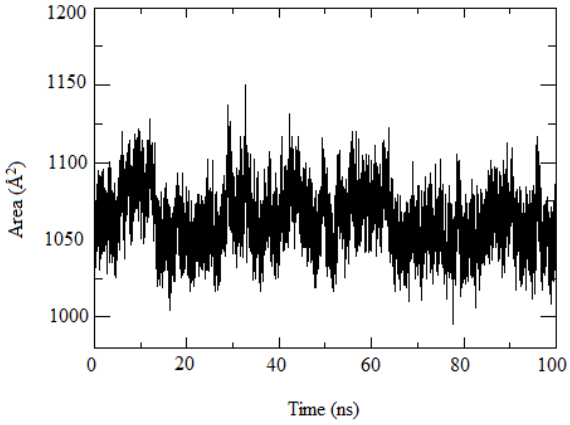
The SASA of protein is a measure of the solvent accessible area which can be used in the determination of the implicit solvent effect of biomolecular structures (180,181,242,243). The SASA can also be used to determine the compactness of the protein, an increase in SASA suggesting a loosening of the protein and a decrease in the SASA suggesting a tightening of the protein (180,181,242,243). The SASA of *M. Sativa* CHIs (Figures 6.7a-e) generally fluctuates between 1000 and 1100 $Å^2$ – the SAS of these proteins reported in Chapter 4 was around 1050 $Å^2$. For 1JX0 and 4DOK the SASA fluctuates between 1025 and 1150 $Å^2$ (Figures 6.7f,h) and for 1JX1 between 950 and 1100 (Figure 6.7g). The CHIs (Figures 6.7a-e) have lower SASA fluctuations than 1JX0 (*M. Sativa* mutant) and 4DOK (Figure 6.7(f and h)) suggesting that the *M. Sativa* CHI are more compact. The SASA results are consistent with the observations made with the RMSF and R_g values in the preceding sections.



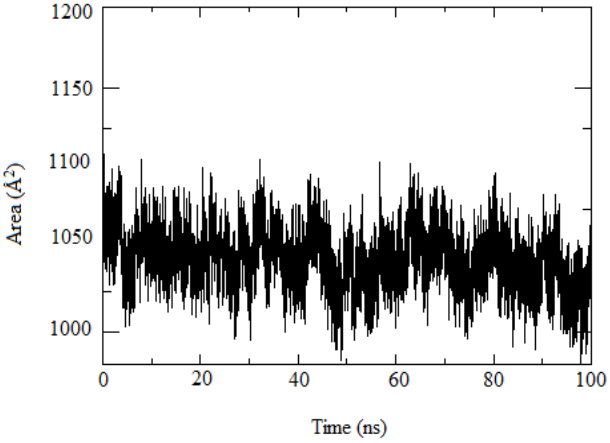
a) 1EYP



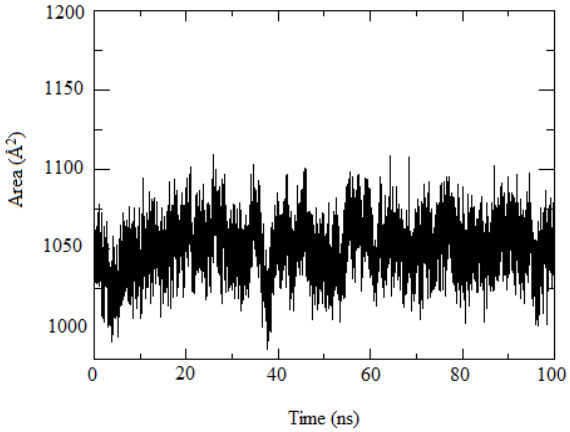
b) 1EYQ



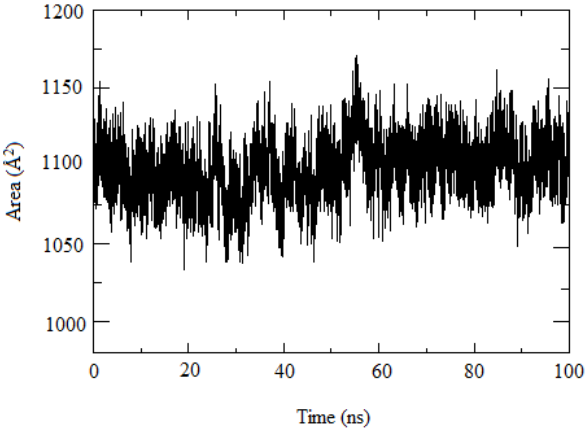
c) 1FM7



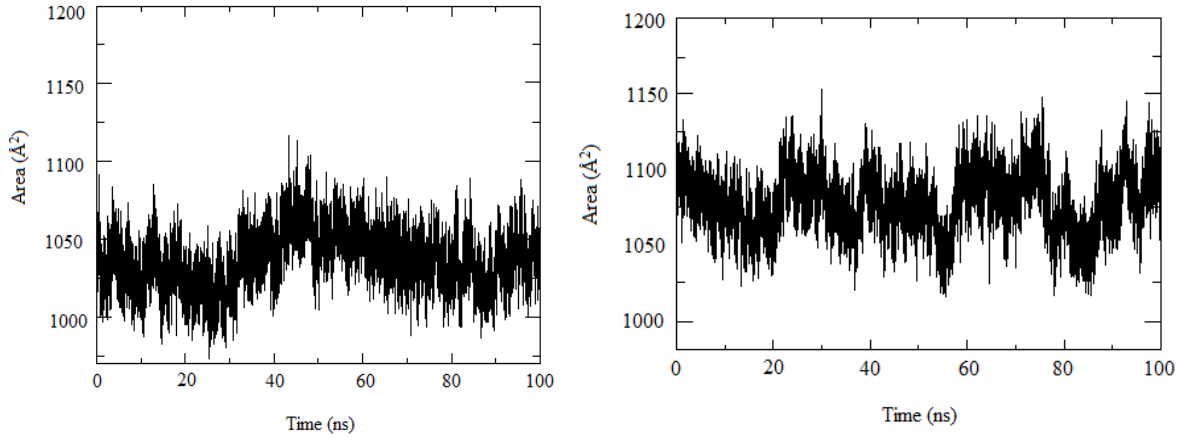
d) 1FM8



e) 1JEP



f) 1JX0



g) 1JX1 h) 4DOK

Figure 6.9. Plots of SASA values for the selected protein structures.

6.3.2.3 Hydrogen bonding

H-bonding patterns in this section are discussed in comparison with the H-bonding patterns of the 30 ns CHARMM simulation data in Section 6.3.2 of this chapter. While the CHARMM simulation (Figure 6.2a) showed the 1EYP-*o*-coumaric acid complex to exhibit sparing protein-ligand H-bonding, the longer GROMACS simulation reveals more predominant H-bonding patterns after about 20 ns (Figure 6.10a). The 1EYQ and 1FM7 complexes (Figures 6.10b, c) also show more predominant H-bonding patterns with the 100 ns GROMACS simulation than with the 30 ns CHARMM simulation (Figures 6.2b, c). Figures 6.10d-g show that the general trends of the H-bonding patterns for the complexes with 1FM8, 1JEP, 1JXO, and 1JX1 with GROMACS are relatively similar to the general trends observed with CHARMM (Figures 6.2d-g). The 1FM8 complex has two predominant H-bonding patterns (Figure 6.10d) with GROMACS simulation similar to those observed with the CHARMM simulations (Figure 6.2d). Figure 6.2h shows that the 4DOK complex has more protein-ligand H-bonds with the GROMACS 100 ns simulation than with CHARMM 30 ns simulation.

Ultimately it is the 1FM8 CHI that is of interest. The predominant protein-ligand H-bonding interactions in the GROMACS simulation were the same as those observed in the previous simulations, *viz.*, H-bond interactions between the alkenoic sp^2 oxygen of *o*-coumaric acid with the OH of Tyr103 (VMD relative numbering – Tyr106 is this residue in literature) of 1FM8 and the H-bond interaction of the phenolic OH of *o*-coumaric acid with sp^2 oxygen of ASN110.

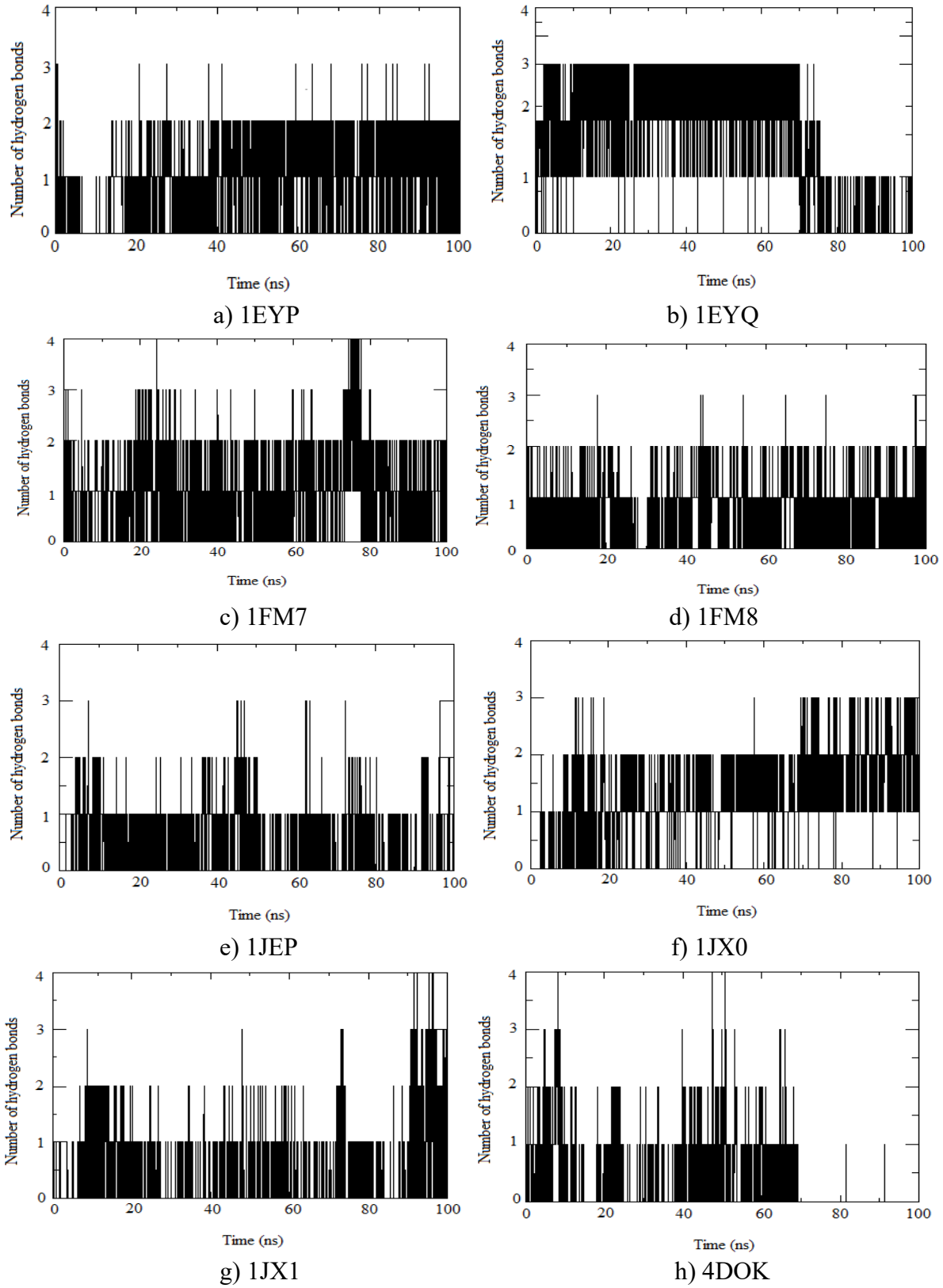


Figure 6.10. H-bonding interactions between selected CHIs and *o*-coumaric acid during GROMAX Molecular Dynamics Simulations.

6.4 Summary of the chapter

Clearly, 1FM8 is the most suitable of all the CHI structures considered in this research for further studies concerning the bioorganic transformation of *o*-coumaric acid to coumarins. This conclusion is based on the following considerations,

- 1FM8 and its ligand *o*-coumaric acid have a steady RMSD within both CHARMM and GROMACS simulations.
- The ligand, *o*-coumaric acid, remains within the active site of 1FM8 throughout these simulations, in contrast to several other cases.
- Hydrogen-bonding patterns of 1FM8 with *o*-coumaric acid show the persistence of two predominant hydrogen-bonds in both the 30 ns CHARMM simulation and the 100 ns GROMACS simulation.
- Analysis of the active site shows H-bonding between the sp^2 oxygen of the alkenoic acid chain of *o*-coumaric acid and Try103 (Tyr106 in literature) and H-bonding of the phenolic OH of *o*-coumaric acid with ASN110. These two protein-ligand H-bonds persist throughout the simulation, fluctuating around 2 Å, suggesting that this 1FM8 complex presents the best option for further transformation of the ligand, appropriate for the biosynthesis of coumarins.

CHAPTER 7
ONIOM QM/MM

7.1 Introduction to the chapter

The preceding chapter elucidated the dynamic evolutions of chalcone isomerase enzyme- *o*-coumaric acid complexes at the molecular mechanical (MM) level, and gave information of which chalcone isomerase enzymes (CHIs) are likely to have the active site characteristics required for the putative catalytic properties. This chapter is concerned with the bioorganic-mediated chemical transformations of *o*-coumaric acid; the first chemical transformation being the *E*→*Z* isomerization of *o*-coumaric acid within the CHI. Because molecular dynamics simulations are conducted at the classical mechanical level by solving Newtonian laws, they are good and computationally affordable approaches for exploring conformational space for large systems such as proteins. Classical dynamics views molecular systems using a ball-spring model, and therefore does not take into account the movement of electrons; for this reason, molecular dynamics approaches based on classical methods do not provide good models for the description of reaction mechanisms (244). Electronic structure methods are better suited for the study of reaction mechanisms because they are based on solutions of the Schrödinger equation (i.e., electronic structure methods are quantum mechanical approaches, and therefore account for orbitals and electrons within these orbitals). The CHI-*o*-coumaric acid complex system is too large to study at a quantum mechanical level but, because our focus is on the reaction mechanism within the active site, molecular mechanical methods are not an ideal. This chapter deals with our employment of a hybrid Quantum mechanical/Molecular Mechanical (QM/MM) method in which the putative centers involved in the *E*→*Z* isomerization of *o*-coumaric acid are defined at the QM level, while the bulk of the system (the remainder of the protein) is defined at the MM level. The results discussed here are from the QM/MM computational approach developed by Morokuma and c*o*-workers – the ONIOM (Our own N-layered Integrated molecular Orbital and molecular Mechanics) model implemented in Gaussian 09 revision E01 (245,246).

7.2 ONIOM layer selection and computational approach

As highlighted earlier, the literature lacks information about enzymes which facilitate the *E*→*Z* isomerization of *o*-coumaric acid in plants. Figure 7.1 (which is the same as Figure 1.5 from Chapter 1, but reproduced here for clarity) illustrates the putative bioorganic transformation of cinnamic acid to coumarins. This study focused on the feasibility of using chalcone isomerase enzymes for the *E*→*Z* isomerization of *o*-coumaric acid.

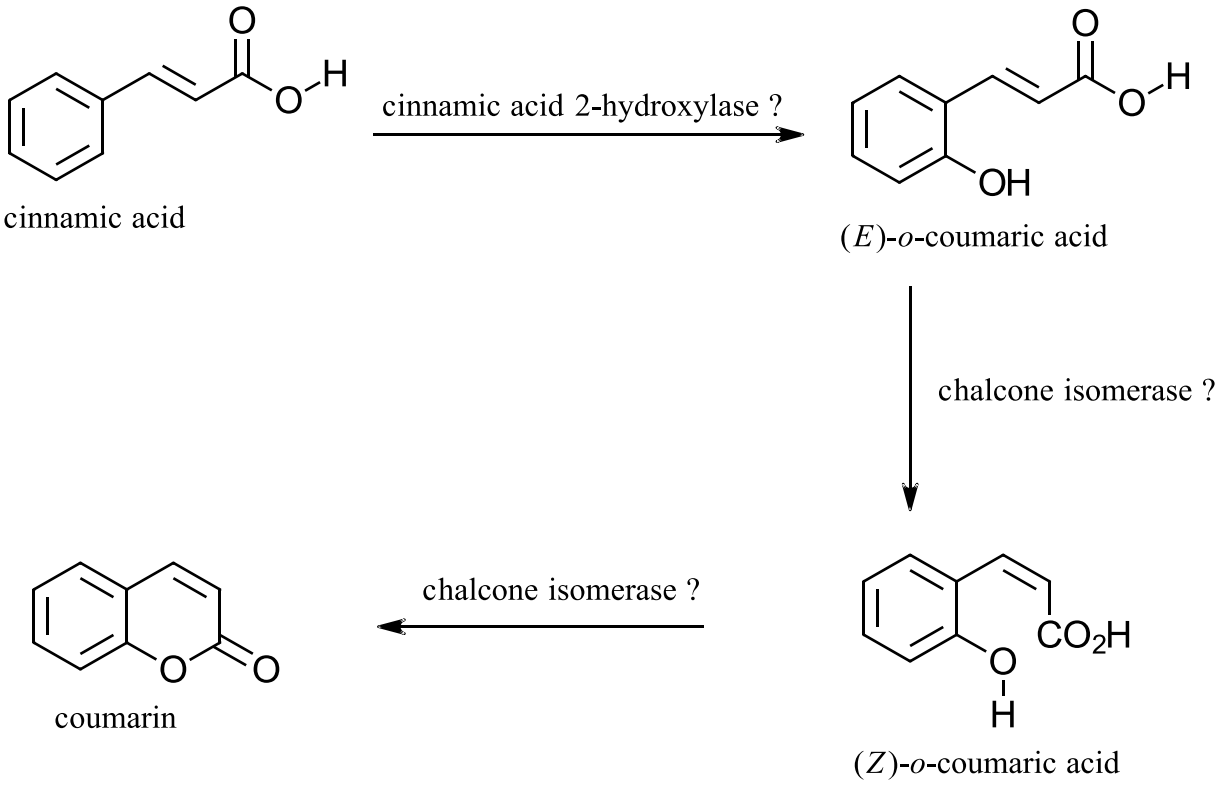


Figure 7.1. Bioorganic transformation of cinnamic acid to coumarins. The question marks indicate that the enzymes are hypothetical.

E→*Z* isomerization of *o*-coumaric acid can be achieved by protonation at O(10) and deprotonation at O(12) or by attack of a nucleophile at the C(7)=C(8) double bond, i.e. conjugate addition (in this case complicated by the proximity of the aromatic system) (247). The study of the different protonation states of cinnamic acid derivatives in Chapter 3 showed

that a feasible approach to achieve the *E*→*Z* isomerization of *o*-coumaric acid involves the protonation and deprotonation of *o*-coumaric acid and the consequent weakening of the C(7)=C(8) double bond.

Figure 7.2 illustrates the putative protonation and deprotonation of *o*-coumaric acid by a CHI and the consequent bond changes that would facilitate *E*→*Z* isomerization. Protein residues, which are in close proximity to the docked *o*-coumaric acid and which have the potential to protonate and deprotonate the ligand, were selected as the "higher layer" to be calculated at the QM level. This was done in two separate calculations; firstly, for comparative purposes at the semi-empirical AM1 level and, secondly, at the hybrid DFT level using the model chemistry B3LYP/6-31G. The bulk of the protein system was selected as the "lower layer" and computed using molecular mechanics, for which the AMBER force field was used. A more detailed outline of the computational approaches is given in Chapter 2 Section 2.2.4.

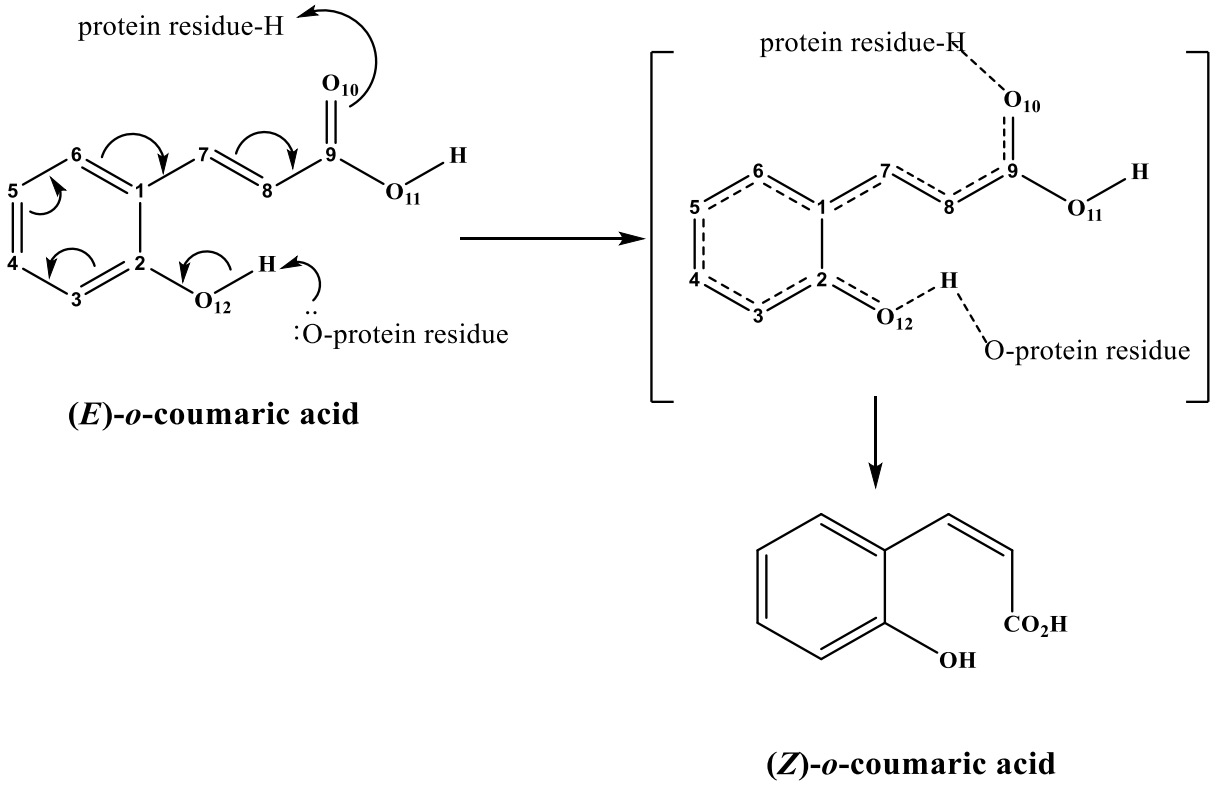


Figure 7.2. Proposed *E*→*Z* isomerization of *o*-coumaric acid.

7.3 Results

The ONIOM calculations were initiated with the optimisation of the CHI-*o*-coumaric acid complexes to find the minimum corresponding to each complex. The geometry optimised complexes were then used for further *in silico* protonation and deprotonation of the ligand *in vacuo*. The protonated and deprotonated ligand was then subjected to a redundant coordinate optimization (a potential energy surface scan) where rotation through 180º around the C(7)=C(8) double bond was effected to achieve the Z isomer. The highest energy point from this scan was then used to do a Berny optimization to a saddle point in order to identify the transition state for the *E*→*Z* isomerization. This transition state was characterized using vibrational analysis, and was subsequently used for an IRC calculation in order to confirm that the transition state was the correct transition state for the *E*→*Z* isomerization.

7.3.1 Receptor-ligand complex optimizations

According to the above procedure, the CHI-*o*-coumaric acid complexes obtained from the docking studies were optimised using Gaussian 09 to analyse the systematic arrangement of the atoms in the geometry optimised active sites. For all systems, the ligand, the protein residues hydrogen-bonded to the ligand, and the protein residues with the potential of protonating and deprotonating the *o*-coumaric acid were calculated as the higher layer. Because Tyr106 is considered to have a significant role in the biological activities of CHIs, it was also included in the higher layer. Table 7.1 shows the resulting ONIOM energy, the C(7)=C(8) bond length, the C(1)-C(7)-C(8)-C(9) dihedral angle, the hydrogen-bond distances between *o*-coumaric acid and the CHI residues, and the Tyr106 distances for each complex. The results show that the ONIOM energy increases in the following order for the CHI complexes with *o*-coumaric acid an indication more of the number of heavy atoms in each of the respective proteins than of relative stability;

1FM7<1EYP<1JX1<1FM8<1JX0<1JEP<1EYQ

The geometry optimised *o*-coumaric acid C(7)=C(8) double bond was close to 1.35 Å in all CHI- *o*-coumaric acid complexes, and the C(1)-C(7)-C(8)-C(9) dihedral angle in the range

174.2° -179.9°. There particular general trend was observed with regard to the hydrogen-bonding; for example, the H-bonding distance between Try106 H-and the phenolic OH of *o*-coumaric acid in 1FM7 decreases after optimization but increases in the 1EYP *o*-coumaric acid complex. The H-bond between Tyr106 and O(12) increases in length after optimization of the 1FM8-*o*-coumaric acid complex but decreases with 1EYQ complex. The arrangement of the species in the active site for all systems are shown in figure 7.3. The results discussed in the previous chapters indicated that 1FM8 is the most promising CHI for further studies involving the *E*→*Z* isomerization of *o*-coumaric acid. However, Figure 7.3b shows that the distance between Tyr106 and O(12) and between Gly37 and the phenolic OH in the 1FM8complex increases after optimization (by a simple rotation of the OH group around the C-O axis); this is not ideal for protonation and deprotonation of the ligand. Therefore, in the further protonation and deprotonation studies involving the 1FM8 complex, the Tyr106 OH is rotated back towards the *o*-coumaric acid O(10) – an arrangement that matches the most dominant conformation obtained from the molecular dynamics simulations (Chapter 6).

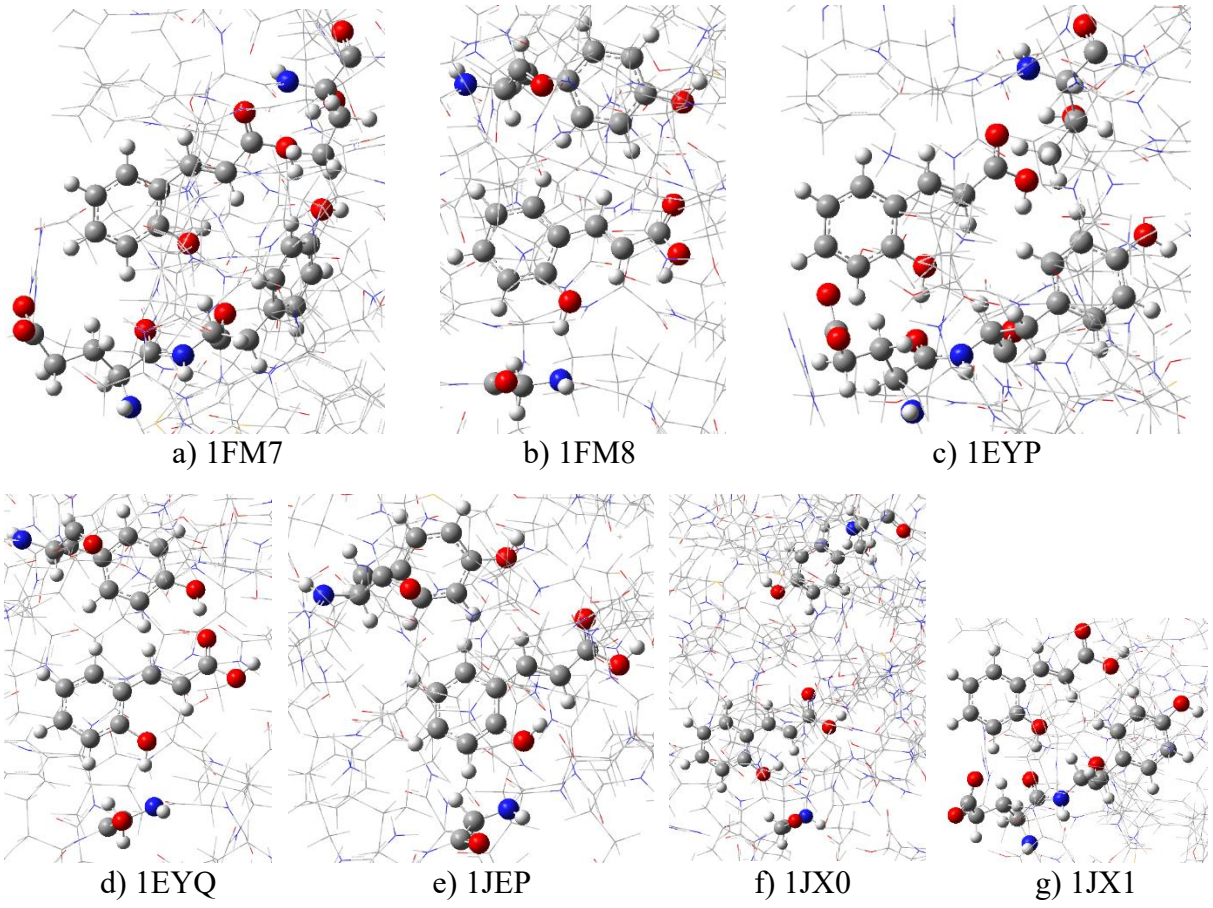


Figure 7.3. ONIOM optimised geometries of the active site of each of the CHI- *o*-coumaric acid complexes.

Table 7.1: Energies and significant dimensions for the CHI-*o*-coumaric acid complexes following ONIOM calculations

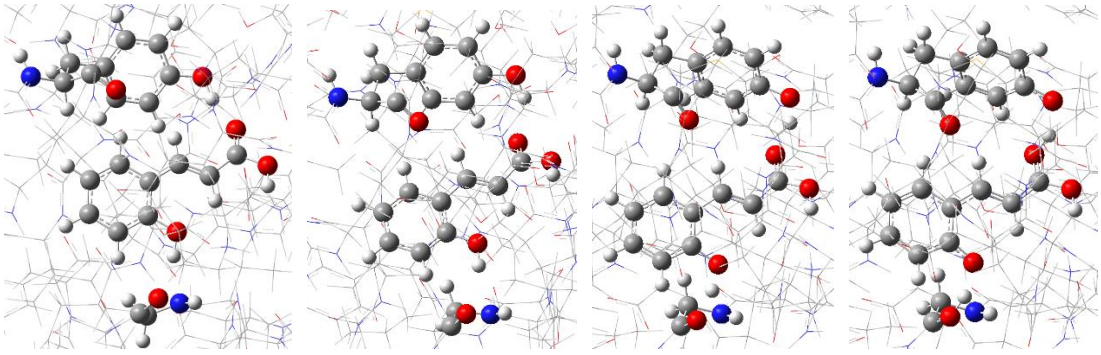
CHI				Properties of hydrogen bonds (distance in Å)		
	ONIOM energy (kca/mol)	C(7)=C(8) (Å)	Dihedral (Å)	Interacting species	Distance before optimiza-tion	Distance after optimiza-tion
1FM7	-1235286.455	1.358	-177.611	Thr48···O(12)	3.153	2.521
				Tyr106···OH	3.218	1.718
				Glu105···OH	3.679	3.610
1FM8	-841093.087	1.351	-178.905	Tyr106···O(12)	3.680	5.629
				Gly37···OH	2.249	3.230
1EYP	-1232336.420	1.354	-178.153	Thr48···O(12)	2.284	1.934
				Tyr106···OH	2.737	3.812
				Gly105···OH	2.567	1.657
1EYQ	-837324.023	1.358	177.341	Gly37···OH	2.717	3.165
				Tyr106···O(12)	3.687	1.749
1JEP	-841022.668	1.356	174.167	Gly37···OH	3.342	5.406
				Tyr106···O(12)	3.754	4.830
1JX0	-841088.259	1.356	-178.327	Tyr152···O(12)	5.436	6.908
				Gly37···OH	3.312	3.452
1JX1	-1007450.588	1.354	-179.9	Tyr106···O(12)	7.421	7.102
				Glu105···OH	1.970	1.717

7.3.2 Protonation and deprotonation

The ONIOM optimised structure of the 1FM8- *E*-*o*-coumaric acid complex, modified as described above, now has a re-formed hydrogen-bonding distance of 2.081 Å between the Tyr106 OH proton and the sp^2 oxygen [O(10)] of the alkenoic acid chain; the other hydrogen-bond between the phenolic proton [i.e. attached to O(12) of *E*-*o*-coumaric acid] and the sp^2 oxygen of Gly37 decreases from 3.230 Å (Table 7.1) to 2.097 Å (Table 7.2). Protonation of *E*-*o*-coumaric acid was performed using a potential energy surface scan, in which the length of the Tyr106 O-H bond of was *increased* from 1 Å (in 10 steps of 0.1 Å), in a direction that *decreased* the length of the H-bond with the sp^2 oxygen of the alkenoic acid chain : this was done simultaneously with the deprotonation of the phenolic OH proton of *E*-*o*-coumaric acid by the sp^2 oxygen of Gly37,by increasing the length of the phenolic O-H bond from 1 Å, in 10

steps with 0.1 Å increments. Ultimately two covalent bonds were broken while forming two new covalent bonds – at the extremities of the potential energy surface scan the Tyr106 H···O(12) and the Gly37 O···H covalent bonds were 0.981 Å [Tyr106 H···O(12)] and 0.997 Å [Gly37 O···H]. The formation of these new covalent bonds (where control of the scan was in breaking bonds) indicates that this method exploring the simultaneous protonation and deprotonation of *E*-*o*-coumaric acid was successful for the 1FM8-*o*-coumaric acid complex under these constrained conditions. The most significant geometries in this scan are shown by Figure 7.4. while the scan for the transformation is illustrated in terms of the energy cost in Figure 7.5. Step 1 is the optimized structure of the complex obtained from the docking study. Step 2 shows the minimum for the *E*- isomer. Step 91 reflects the state of the system in which the protons approximately half-way between the oxygens they are leaving and the oxygens towards which they are migrating. Step 141 is the final structure with the hydrogens having migrated completely to the new sites.

Step 141 not a minimum on the potential energy surface, but represents the correct electronic structure for isomerization to occur. This state of protonation/deprotonation was forced to remain during further redundant coordinate surface scans, by constraining the appropriate Tyr106 H···O(12) and the Gly37 O···H bond distances.



Step 1 Step 2 Step 91 Step 144

Figure 7.4. The ligand and the protein residues that are proposed to be involved in the protonation and deprotonation are represented in stick and ball format, and the rest of the protein is represented as stick. Step 1 is the initial conformer *E*- isomer, Step 2 is the lowest energy conformation for the *E*- isomer, Step 91 represents a state midway through the protonation/deprotonation process, and Step 141 is the protonated and deprotonated *Z*-isomer.

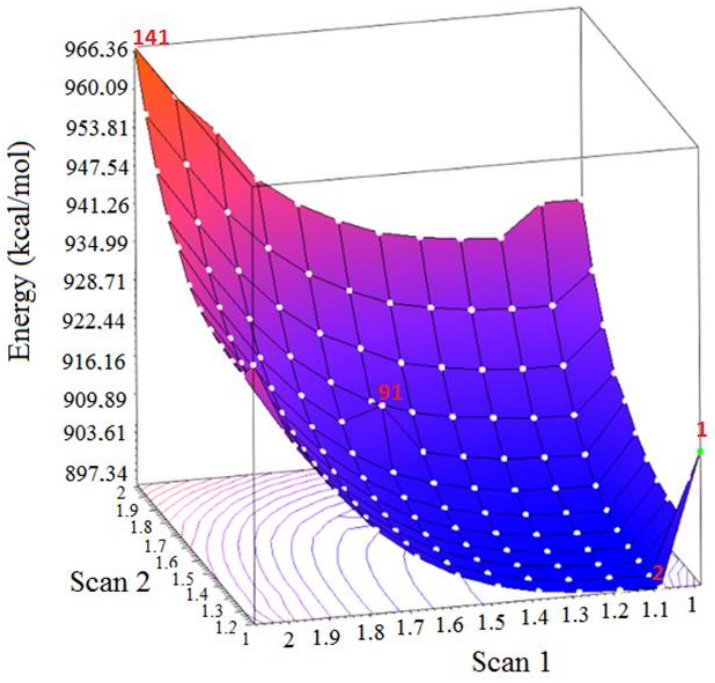


Figure 7.5 Scan grid of the protonation of O(10) with a hydrogen atom from Tyr106 and the deprotonation of O(12) phenolic proton by the oxygen of Gly37.

Table 7.2: Properties of steps in the protonation/deprotonation scan.

	Step 1	Step 2	Step 91	Step 144
E (kcal/mol)	917.409	895.447	913.644	966.364
Tyr106···O(12)	2.081	1.981	1.481	0.981
Gly37···OH	2.097	2.097	1.397	0.997
C(7)=C(8)	1.345	1.344	1.348	1.369
Dihedral	178.635	173.979	175.183	174.693

7.3.3 Rotational scans

The protonation/deprotonation state remained fixed (with the aforementioned constraints in Tyr106 H···O(12) and the Gly37 O···H bond lengths) during the next phase of the study in which the aim was to search for the transition state corresponding to the *E*→*Z* isomerization. The protonated and deprotonated *E*-*o*-coumaric acid (Step 141 from the protonation/deprotonation scan grid) was subjected to internal rotation through 180° along the C(7)-C(8) *partial* double bond by means of a rotational scan to achieve the *E*→*Z* isomerisation.

The protonated and deprotonated *E*-*o*-coumaric acid structure with an energy of 966.354 kcal/mol was the initial complex in the rotational scan, i.e. Step 1 in Table 7.3; the maximum energy reached in this rotational scan (998.357 kcal/mol) was about 30 kcal/mol higher than the energy of the initial complex and corresponds to Step 8 in Table 7.3. After the maximum was reached in the scan, the energy of the resulting *Z*-*o*-coumaric acid structure (Step 13 in Table 7.3) dropped to 961.335 kcal/mol, i.e. about 4 kcal/mol lower than the energy of the starting protonated and deprotonated *E*-*o*-coumaric acid. (,. Table 7.3 also shows the changes in the C(7)-C(8) bond length which increases from 1.372 Å in the initial protonated and deprotonated *E*-*o*-coumaric acid to 1.438 Å in the putative transition state (Step 8), and then decreases to 1.344 Å in the *Z*-isomer (Step 13). These data indicate that the C(7)=C(8) double bond weakens significantly during the surface scan. The C(7)-C(8) dihedral angle of the initial, protonated and deprotonated *E*-*o*-coumaric acid is 174.693º but, as the rotation proceeds, it decreases to -80.307º in the transition state. The final *Z*-*o*-coumaric acid has a C(7)-C(8) dihedral angle of -5.307º. These data confirm that the rotational scan in the search for *E*→*Z* isomerisation stationary states was successful, and the most significant Steps in this process are shown in Figure 7.6, while the details of the scan are shown in Figure 7.7, Step 1 is the structure of the initial protonated/deprotonated receptor-ligand 1FM8 of *E*-*o*-coumaric acid complex; Step 3 is the minimum energy structure of this complex. Step 8 is the putative transition state, while Step 9 is a nascent *Z*-isomer after the putative transition state. Step 13 corresponds to *Z*-*o*-coumaric acid. The lowering of the energy of the Z-isomer at this stage is due to significant stabilization of the complex by a very strong intramolecular hydrogen-bonding interaction between the phenolic O^- and the carboxylic acid proton, together with the two expected, strong hydrogen-bonding interactions between the hydrogens and oxygens of the two cleaved OH bonds (from protonation and deprotonation) (Figure 7.6, Step 9 and 13). The intramolecular hydrogen bonding interaction between the carboxylic acid proton and the phenolic O^- provides further information about the reactivity of the coumaric acid and the capability of the carboxylate anion of deprotonating the phenolic OH in this series of compounds. The highest energy geometry in the rotational scan was then used as a starting point for a transition state optimization, in the search for the transition state for the *E*→*Z* isomerisation.

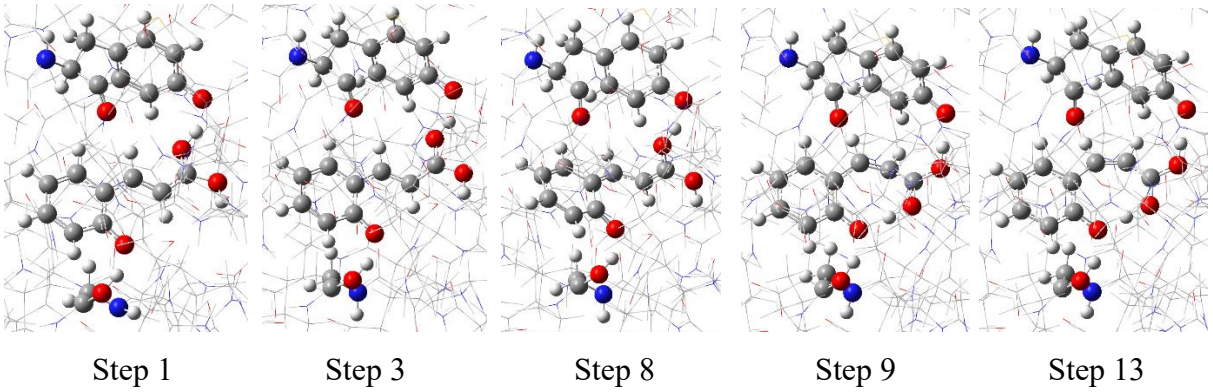


Figure 7.6. The ligand and the protein residues that are expected to be involved in the isomerization process are represented in stick and ball format, and the rest of the protein is represented as stick. Step 1 is the protonated/deprotonated *E*- isomer, Step 3 is the lowest energy conformer for the protonated/deprotonated *E*- isomer, Step 8 is the rotational transition state, Step 9 is a nascent *Z*- isomer conformation, and Step 13 is the lowest energy conformer for the *Z*- isomer.

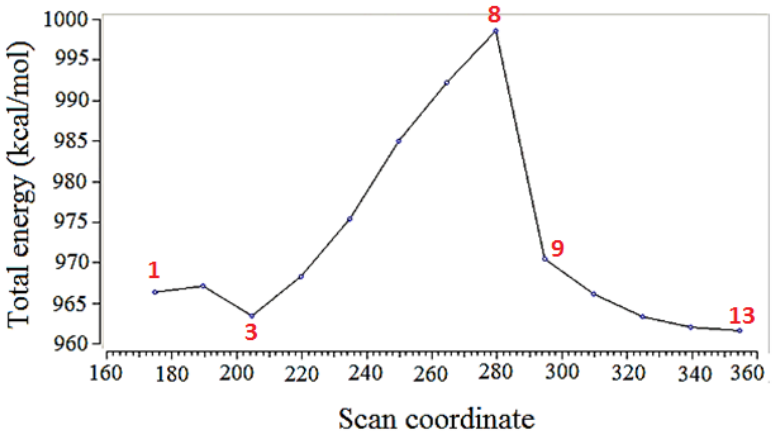


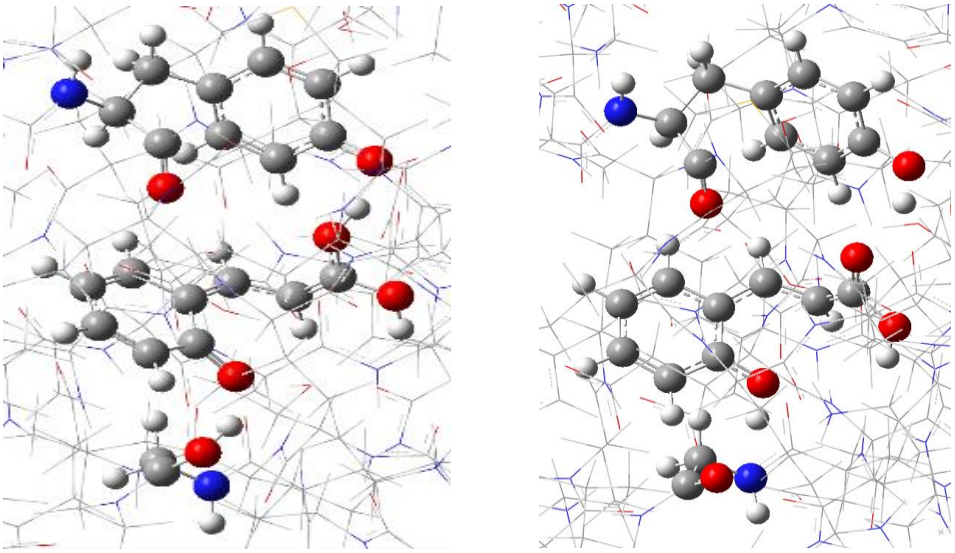
Figure 7.7. Rotational scan of *o*-coumaric acid for the *E*→*Z* isomerisation.

Table 7.3: Properties of steps of protonation.

	Step 1	Step 3	Step 8	Step 9	Step 13
E (kcal/mol)	966.354	963.217	998.357	970.747	961.335
C(7)=C(8)	1.372	1.374	1.438	1.360	1.344
Dihedral	174.693	-155.307	-80.307	-65.307	-5.307

7.3.4 Transition state analysis

All redundant coordinates were removed at this point, and the geometry from Step 8 in the rotational scan was used as input for a transition state optimization (a Berny optimization with re-calculation of force constants at each Step of the optimization procedure). The optimization converged successfully and a subsequent vibrational analysis provided evidence of a single imaginary frequency of -1268.97 cm^{-1}. The removal of all redundant coordinates did allow for significant reversal of the protonation/deprotonation OH geometries but, in the optimized transition state complex, these protons remain strongly hydrogen-bonded to Gly37 and Tyr106; Table 7.4 shows how these bond-lengths have altered during the course of transition state optimization). Figure 7.8 shows the relative positions of the protons on the ligand and the protonating and deprotonating protein residues in the putative, rotational TS and in the optimized TS. The energy of the optimized transition state is *ca*. 45 kcal/mol lower than the transition state from the rotational scan (Table 7.4), and the C(7)=C(8) double bond is slightly shorter for the optimized TS (1.395 Å) than for the rotational TS (1.438 Å). The final dihedral angle for the isomerization TS is -93.693 Å, close to its value from the rotational scan (-80.307°).



TS from rotational scan
998.577 kcal/mol

Optimised TS
945.860 kcal/mol

Imaginary frequency -1268.97 cm^{-1}

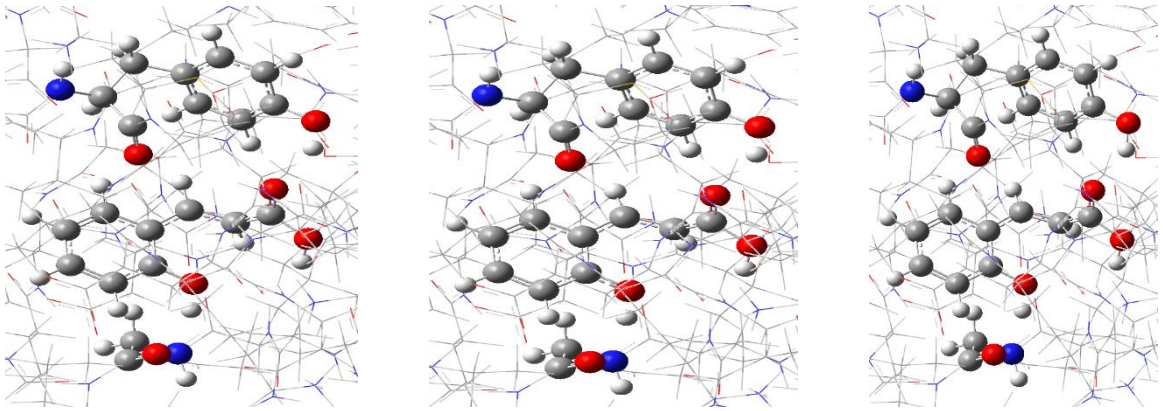
Figure 7.8. Structures of the rotational scan TS and the optimised TS for the *E*→*Z* isomerisation of *o*-coumaric acid.

Table 7.4: Comparative data for the rotational and optimized transition states.

	Rotational TS	True TS
E (kcal/mol)	998.577	945.860
Tyr106 H···O(12)	0.98136	1.903
Gly37 O···H	0.99651	1.925
C(7)=C(8)	1.438	1.395
Dihedral	-80.307	-93.693

7.3.5 Intrinsic reaction coordinate (IRC) study

The optimized transition state was used as a starting point in the IRC calculations to explore the reaction pathway and, hopefully, confirm that this transition state is the link between the *E*- and *Z*- isomers. In the IRC study, 10 Steps were taken towards the *E*- isomer and another 10 Steps towards the *Z*-isomer. Figure 7.9 shows the transition state and the isomers obtained. The name of the isomer is given on the bottom left and the energy is given on the bottom right of each of figure. This confirms that the isomerization of the *E*- and *Z*-isomers is indeed linked by this transition state (see Figure 7.10).



Forming *E*- isomer 937.402 TS 946.072 Forming *Z*- isomer 940.632

Figure 7.9. Structures obtained from the IRC calculations. TS is the transition state structure. Energy is in kcal/mol.

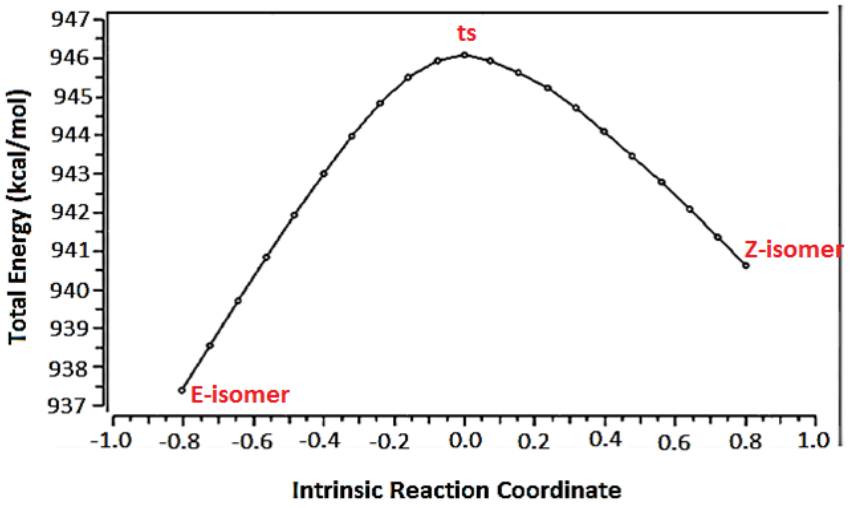


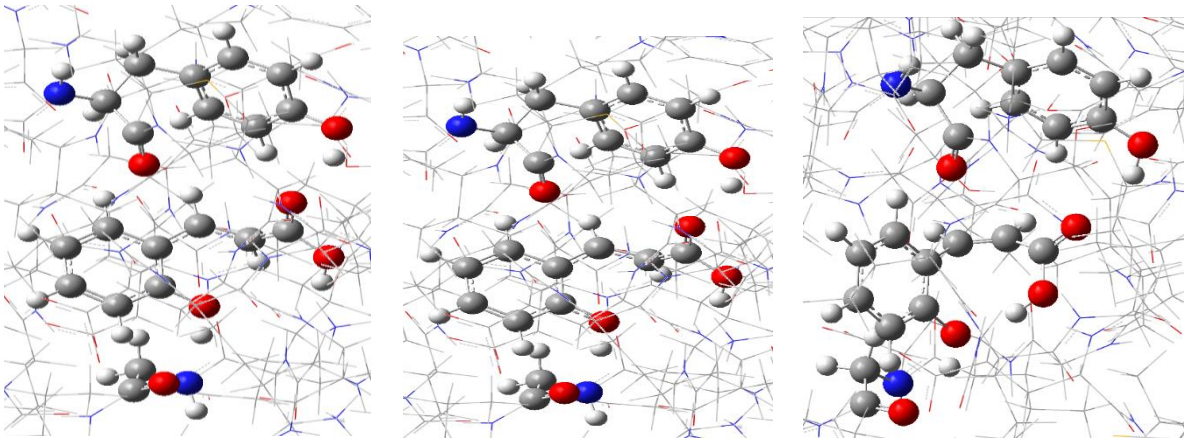
Figure 7.10. Plot of IRC data for the *E*→*Z* isomerisation of *o*-coumaric acid.

Table 7.5: Data from the IRC study.

	Forming *E*- isomer	**TS**	**Forming *Z*- isomer**
E (kcal/mol)	937.402	946.072	940.632
Tyr106 H···O(12)	1.910	1.903	1.920
Gly37 O···H	1.936	1.925	1.923
C(7)=C(8)	1.390	1.395	1.377
Dihedral	-98.970	-93.693	-87.678

7.3.6 *E*- and *Z*-isomers

The *E*- and the *Z*-isomers obtained from the IRC study were not minimum energy structures and, therefore, they were further optimised (with subsequent frequency calculations in order to obtain the zero point energy correction). After this optimization the *E*- isomer had an electronic energy of 892.483 kcal/mol, while the *Z*- isomer had an energy of 906.469 kcal/mol (Figure 7.11). The two minima obtained provided the starting and the final receptor-ligand complexes. Table 7.6 show that the *E*- and the *Z*-isomer are not in a protonated/deprotonated state (the Tyr106 H···O(12) and Gly37 O···H are at around 2 Å, suggesting the presence of strong hydrogen-bonding). The C(7)=C(8) bond lengths for the *E*- and the *Z*- isomer are at 1.33 and 1.34 Å, respectively consistent with the presence of double bonds. The final dihedrals for the *E*-isomer (-179.264) and the *Z*- isomer (-0.454) confirm the *E*→*Z*- isomerization.



E- isomer 892.483 kcal/mol TS 946.072 kcal/mol *Z*- isomer 906.469 kcal/mol

Figure 7.11. Optimized structures of the *E*- and *Z*-isomers and the isomerization transition state.

Table 7.6: Properties of the ONIOM AM1:UFF optimized structures of the *E*- and *Z*-isomers and the transition state complex.

	***E*- isomer**	**TS**	***Z*- isomer**
E (kcal/mol)	892.483	946.072	906,469
Tyr106 H···O(12)	2.024	1.903	2.214
Gly37 O···H	2.013	1.925	1.851
C(7)=C(8)	1.345	1.395	1.333
Dihedral	-179.264	-93.693	-0.454

7.4 QM level calculations at the DFT/6-31G(d) level

This system was too large for successful DFT optimization within the ONIOM framework using the computational resources at the CHPC within the time-frame available for the queuing system. Attempts were made to reproduce the ONIOM AM1:UFF scans at the ONIOM B3LYP/6-31G(d):UFF level, but insufficient computational time was available within the queuing system to afford significant progress. Consequently, single point ONIOM B3LYP/6-31G(d):UFF calculations were performed on stationary points to provide corrected ONIOM B3LYP/6-31G(d):UFF energies for the ONIOM AM1:UFF geometries. The results obtained at the ONIOM B3LYP/6-31G(d):UFF level were in agreement with the results obtained at the

ONIOM AM1:UFF level, and the trends observed are similar to the trends discussed in the preceding sections (Table 7.7). In any case, it would appear that the transformation is plausible for this system (given the simplifications in terms of the model chemistry and the absence of solvent molecules to mediate protonation and deprotonation). However, to complement this information, it was decided that a QM/MM model that adequately treats solvent and its reactive potential would best suit the purpose of this study. Therefore, adaptive partitioning QM/MM-based molecular dynamics was chosen as the next phase in the study of this system, and this is covered in Chapter 8.

Table 7.7: Properties of the ONIOM AM1:UFF optimized structures of the *E*- and *Z*-isomers and the transition state complex, following single-point recalculation of their energies at the ONIOM B3LYP/6-31G(d):UFF level

	***E*- isomer**	**TS**	***Z*- isomer**
E (kcal/mol)	879.462	931.672	892,348
Tyr106···O(12)	2.024	1.903	2.214
Gly37···OH	2.013	1.925	1.851
C(7)=C(8)	1.345	1.395	1.333
Dihedral	-179.264	-93.693	-0.454

7.5 Summary of reaction pathway

The modelling of the *E*→*Z* isomerization of *o*-coumaric acid within 1FM8 chalcone isomerase enzyme structure was successful, having been guided by direct protonation and deprotonation of *o*-coumaric acid mediated by the enzyme residues, Tyr106 and Gly37. Figure 7.12 summarizes the reaction pathway; the path marked in blue illustrates the steps followed in this study to arrive at a plausible transition state, while the green path illustrates the optimized transition state and the *E*- and *Z*-isomers that it connects. The optimized reaction pathway shows that, in the *E*→*Z* isomerization, the involvement of protonating/deprotonating residues is perhaps less dependent on complete transfer of the respective protons, but rather the electronic effects induced by the strong hydrogen-bonding between *o*-coumaric acid and 1FM8. Overall, the results indicate 1FM8 to be a plausible system for effecting this enzymic isomerization.

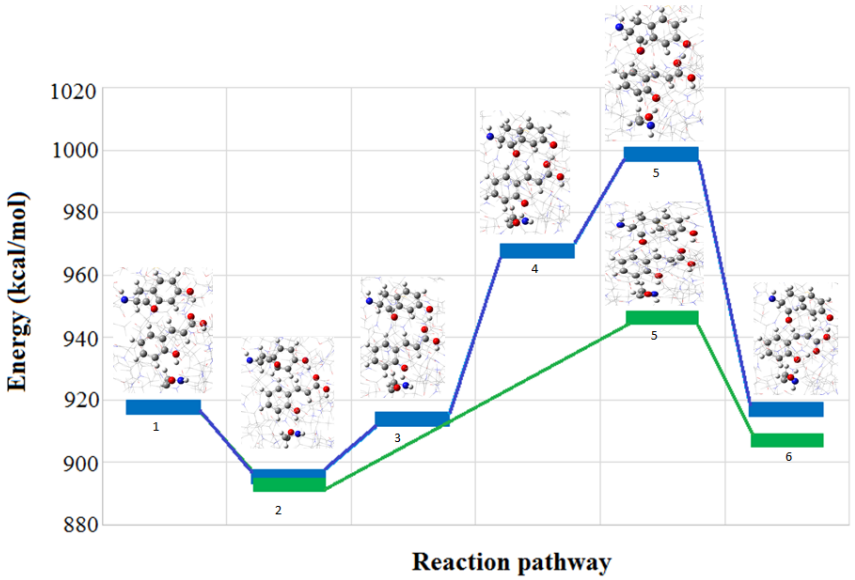


Figure 7.12. The optimized *E*→*Z* isomerisation reaction pathway (green), with energies of the stationary points re-calculated at the ONIOM B3LYP/6-31G(d):UFF level, and the analytical phases (blue) followed computationally in this study. 1 is the receptor-ligand complex from docking. 2 is the *E*-isomer minimum. 3 is a structure midway in the protonation/deprotonation scan 4 is the protonated and deprotonated ligand in the receptor-ligand complex. 5 is the rotational transition state. 6 corresponds to the *Z*-isomer minima.

CHAPTER 8
ADAPTIVE PARTITIONING QM/MM

8.1 Introduction to the chapter

The ONIOM study in Chapter 7 has provided information on a possible reaction pathway involving the bioorganic-mediated protonation and deprotonation of *o*-coumaric acid with chalcone isomerase enzymes (CHIs). However, the reaction pathways described in the ONIOM study were calculated *in vacuo*. This is a natural first step to follow when elucidating reaction mechanism, since it is a simpler representation of the reaction environment with fewer species moving (i.e. only the coordinates of protons involved in the protonation/deprotonation are changing significantly, together with the isomerization rotation). Therefore, the ONIOM model provides clarity on the mechanism, and this clarity is afforded within a reasonable computational time.

However, protonation and deprotonation of the ligand with the ONIOM model was directly affected by protein residues in the active site. Adaptive partitioning QM/MM provides a more realistic approach which takes into account that, in nature, reactions occur in a solvating medium and enables the solvent to be involved in reaction processes. It is essential to consider the interaction of water with the active site, since water molecules could be involved in the protonation and deprotonation process (248). In adaptive partitioning QM/MM, solvent molecules, which are considered to have a direct effect on the active site, are defined at the QM level, while the bulk of the solvent molecules are defined at the MM level together with the bulk of the protein (excluding active site or involved residues). Species in the active site (including the ligand, involved residues and active site water) are calculated at QM level and are denoted as the primary subsystem (PS), while the remaining bulk of the system is denoted as the secondary subsystem (SS). During an adaptive QM/MM simulation, QM water molecules may cross the boundary separating the QM and the MM regions. The danger is that these crossings may, under normal QM/MM simulations, result in abrupt changes in the energy and the forces; therefore, the best adaptive QM/MM simulations are those that circumvent or mediate such abrupt energy changes by including a buffer region between the QM and the MM region. This chapter is concerned with the results of the adaptive QM/MM calculations as

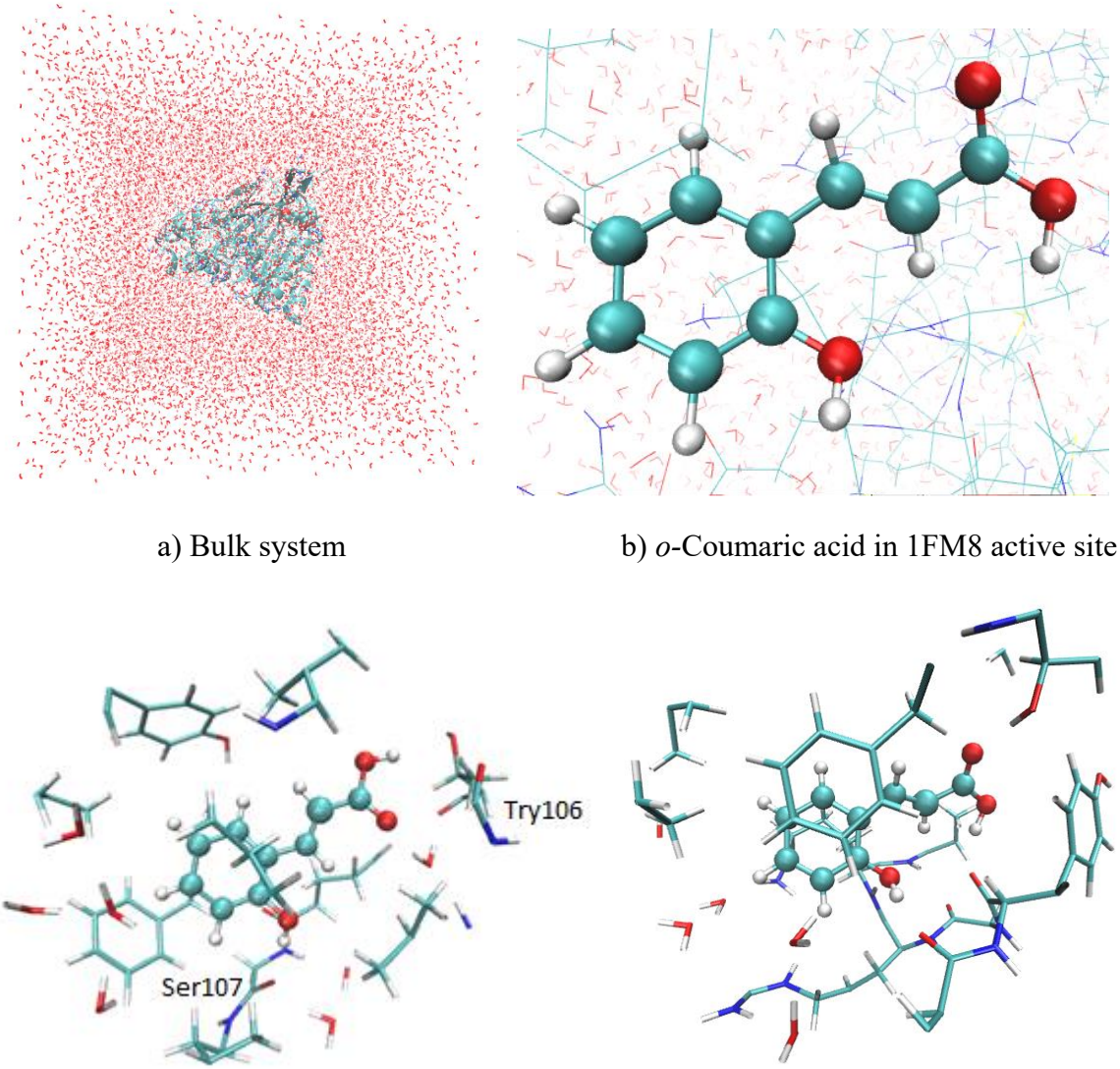
implemented in the CP2K package (191). CP2K uses the electrostatic embedding technique which includes one-electron terms between the PS and the SS in the QM Hamiltonian and which couples the QM charge distribution and the MM charges using a highly efficient multigrid technique (191).

8.2 Computational approach

The adaptive partitioning QM/MM study was carried out on the same receptor and ligand that have been the focus of study in the previous chapter (*viz.*, the 1FM8-*o*-coumaric acid complex). The 1EYQ-*o*-coumaric acid complex was also considered because it was not favourable for the ONIOM study due to the large distances between the proposed putative protonating/deprotonating residues and the ligand and, therefore, might provide more information about the influence of solvent molecules on the active site. The protein crystal structures considered in this chapter were those utilized in previous chapters, and their preparation is discussed in Chapter 2 Sections 2.2.2. The ligand preparation is discussed in Chapter 2 Sections 2.2.5. The PS was computed at the QM level employing DFTB+, while the SS was computed at MM level employing GAFF. A more detailed approach to the setup of the system and the calculation methods is given in Chapter 2 Section 2.2.5.

Both the 1FM8-*o*-coumaric acid complex and the 1EYQ- *o*-coumaric acid complex were each explored in two molecular dynamics studies; the first was a QM/MM-based molecular dynamics study of a solvated system, while the second was an adaptive partitioning QM/MM-based molecular dynamics study. Both simulations on both complexes were performed with CP2K.

Figure 8.1(a) shows an example of the bulk system considered in the chapter, in which each PS visualised comprises only part of the full simulation. Figure 8.1(b) shows the conformation of the ligand in the 1FM8 active site. Figures 8.1(c and d) show the atomic arrangements of the QM regions in the initial adaptive partitioning. The core region is the ligand, and the dynamical QM region is selected by extending the core region by r_{qm}= 5.0-5.8 Å [the bond model species in Figure 8.1(c and d)]. The additional buffer layer was also extended from the dynamical QM region by r_{buffer}= 5.0-5.8 Å.



a) Bulk system

b) *o*-Coumaric acid in 1FM8 active site

c) 1FM8 initial adaptive QM region

d) 1EYQ initial adaptive QM region

Figure 8.1 Model structures considered in this study.

8.3 Results

The discussion focuses on comparative results obtained from the QM/MM dynamics and the adaptive partitioning QM/MM dynamics for the two complexes considered (1FM8-*o*-coumaric acid and 1EYQ-*o*-coumaric acid). Figure 8.2 shows the numbering of *o*-coumaric acid (reproduced from Chapter 1 for clarity, since this numbering of *o*-coumaric acid will be utilised in the analysis of the results in this section).

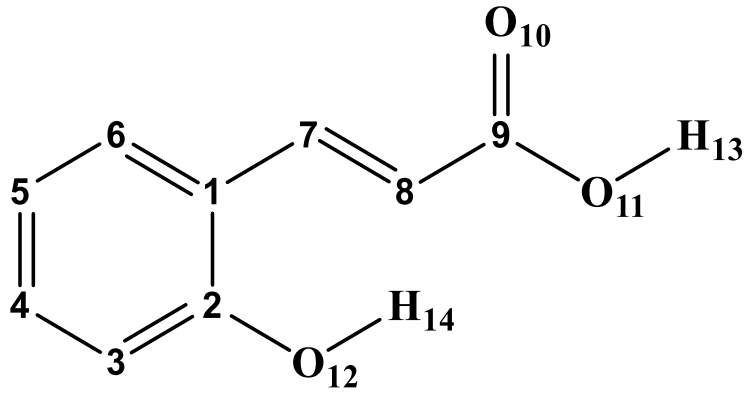


Figure 8.2. Structure of *o*-coumaric acid and the numbering utilized in this study.

8.3.1 Energy profiles

The energy profiles of an adaptive partitioning QM/MM simulation could possibly give the reaction pathway (when a chemical process takes place during the course of the dynamics simulation). Consequently, a first step was the investigation of these profiles, but in neither case was a chemical change observed.

There is, however, an interesting side-effect of the adaptive partitioning procedure when it comes to the change in the total energy during the course of simulation. As more species are accommodated within the active site (in particular the QM/MM buffer region), the differential between the QM energy and the MM energy within the system becomes apparent. The energy profiles of the 1FM8-*o*-coumaric acid and the 1EYQ-*o*-coumaric acid complexes during the

adaptive partitioning QM/MM-based dynamics are given in Figures 8.3(a and b). Figure 8.3(a) shows that the initial energy of the 1FM8-*o*-coumaric acid complex is at -300000 kcal/mol, and Figure 8.3b shows that the initial energy of 1EYQ- *o*-coumaric acid complex is at about -350000 kcal/mol. The energy of the two complexes decrease during the period from 0 fs to 500 fs. The energy of 1EYQ-*o*-coumaric acid complex decreases much faster than the energy of 1FM8-*o*-coumaric acid complex; at 500 fs, the energy of 1EYQ-*o*-coumaric acid complex had dropped to almost -550000 kcal/mol from an initial energy of -350000 kcal/mol (i.e. a decrease of about -200000 kcal/mol) while, over the same period, the energy of the 1FM8-*o*-coumaric acid complex had dropped by about -100000 kcal from -300000 kcal/mol to almost -400000 kcal/mol (. These changes in energy may be attributed to the movement of solvent particularly into the QM/MM buffer region but also into the active site regions during the course of simulation, resulting in apparent non-conservation of total energy for the systems. However, the energy decrease is due to solvation of the active site causing more water to enter the QM region. Since E_QM+E_buffer+E_MM (Energy of the system) is only constant if the atoms in each region do not change; the movement of atoms between the regions is responsible for the observed decreases in energy.

On the other hand, the QM/MM simulations, where both the PS and SS are fixed, do not show this non-conservation of energy during the course of the dynamics [Figures 8.3(c and d)]. These figures illustrate the total energy profiles of the 1FM8-*o*-coumaric acid and 1EYQ-*o*-coumaric acid complexes for the QM/MM calculation with just the ligand at the QM level and for complexes with the ligand and Tyr103, Glu102, and Ser104 at the QM level. The QM/MM energies of both complexes with different PS remains fixed at about -200000 kcal/mol for 5000 fs simulations (the very slight differences in total energy between the two systems relate to atom differences between the two systems).

Thus, the energies of the adaptive partitioning QM/MM systems are generally lower than the energies of the of the QM/MM systems, which remain fixed at *ca*. -200000 kcal/mol, while the energies of the adaptive partitioning systems are generally well below -300000 kcal/mol, again these differences may be attributed to differences in the number of solvent molecules at each of the QM and MM levels.

The fact that isomerization and protonation/deprotonation were not observed to occur during these dynamics simulations may be attributed to two factors. Firstly, given that the calculations

are so much more complex in terms of the QM part of the calculation, the total time for dynamics is limited. Secondly, molecular dynamics will not easily give rise to rare events, or events that take place on time-scales that are orders of magnitude different to the simulation time. It is an aim for future work on this project to use the metadynamics approach, forcing the *E*→*Z* isomerization. However, from the 500 fs simulations we can extract information on the solvent, ligand and active site behaviour in the active site during the course of simulation, which will also inform and guide the set-up for further simulations. Furthermore, the active site can be analysed to determine the bulk motion processes that account for the decrease of the potential energy of the systems observed in Figures 8.3(a and b).

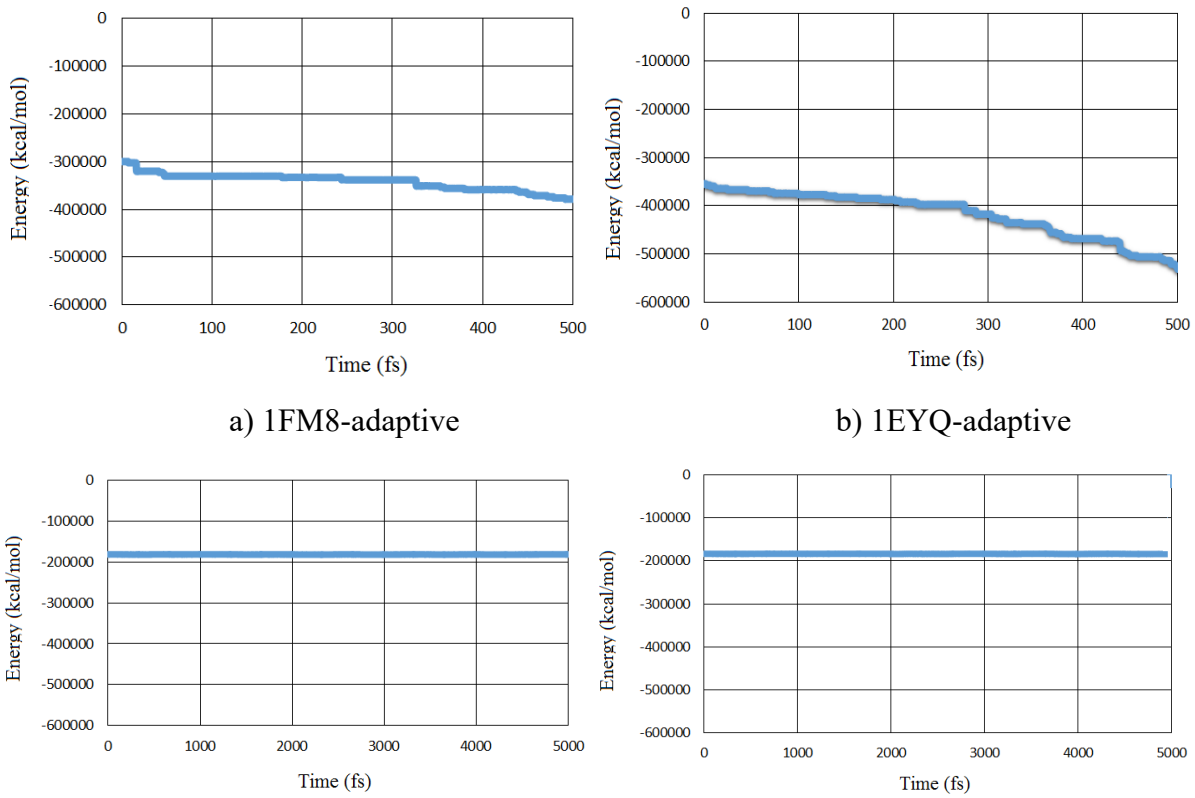


Figure 8.3 Plots of the total QM+MM potential energy during the course of the four dynamics simulations (kcal/mol)

8.3.2 RMSD changes

The changes in the RMSD during the adaptive partitioning QM/MM dynamics simulations of the 1FM8-*o*-coumaric acid and 1EYQ-*o*-coumaric acid complexes are illustrated Figures 8.4 (a and b). These two figures show that 1FM8 and its ligand thermally equilibrate within the 500 fs of the simulation (1FM8 thermally equilibrates with an RMSD of *ca*. 0.75 Å, its *o*-coumaric acid ligand below 0.5 Å). Compare this with the much longer, but simpler QM/MM calculation for 5000fs, where the equilibration takes place in a similar time-frame [Figure 8.4(c)]. 1EYQ and its ligand, however, do not reach thermal equilibration within this 500 fs time frame, i.e., the RMSD of 1EYQ and its *o*-coumaric acid ligand continue to increase throughout the simulation and, at the end of the 500 fs simulation, the RMSD of 1EYQ is at 1 Å and that of its ligand is at *ca*. 0.8 Å). The chemical properties of the 1FM8-*o*-coumaric acid complex that are amenable to analysis depend on having a thermally equilibrated system, and, therefore, detailed analysis was only conducted using the later parts of the simulation. Although the 1EYQ-*o*-coumaric acid system did not thermally equilibrate within 500 fs, the results from the later part of this simulation were still useful for comparison with those of 1FM8-*o*-coumaric acid complex.

Figures 8.4 (c and d) shows the RMSD fluctuations of the simpler QM/MM simulations for the 1FM8- *o*-coumaric acid and the 1EYQ- *o*-coumaric acid complexes over 5000 fs simulations. These figures show that 1FM8 and 1EYQ are both thermally equilibrated at 100 fs to 5000 fs. These observations are consistent with those from the shorter adaptive partitioning simulations. The ligand in FM8 is more stably organized than the ligand in 1EYQ (the RMSD of the 1FM8 ligand fluctuates less than the RMSD of the ligand in 1EYQ). These RMSD observations were also confirmed by visualisation of the simulations using VMD.

The adaptive partitioning QM/MM RMSD and the QM/MM RMSD results show that *o*-coumaric acid is more stable when docked to 1FM8 than when docked to 1EYQ; this is consistent with the observations in Chapters 4 and5.

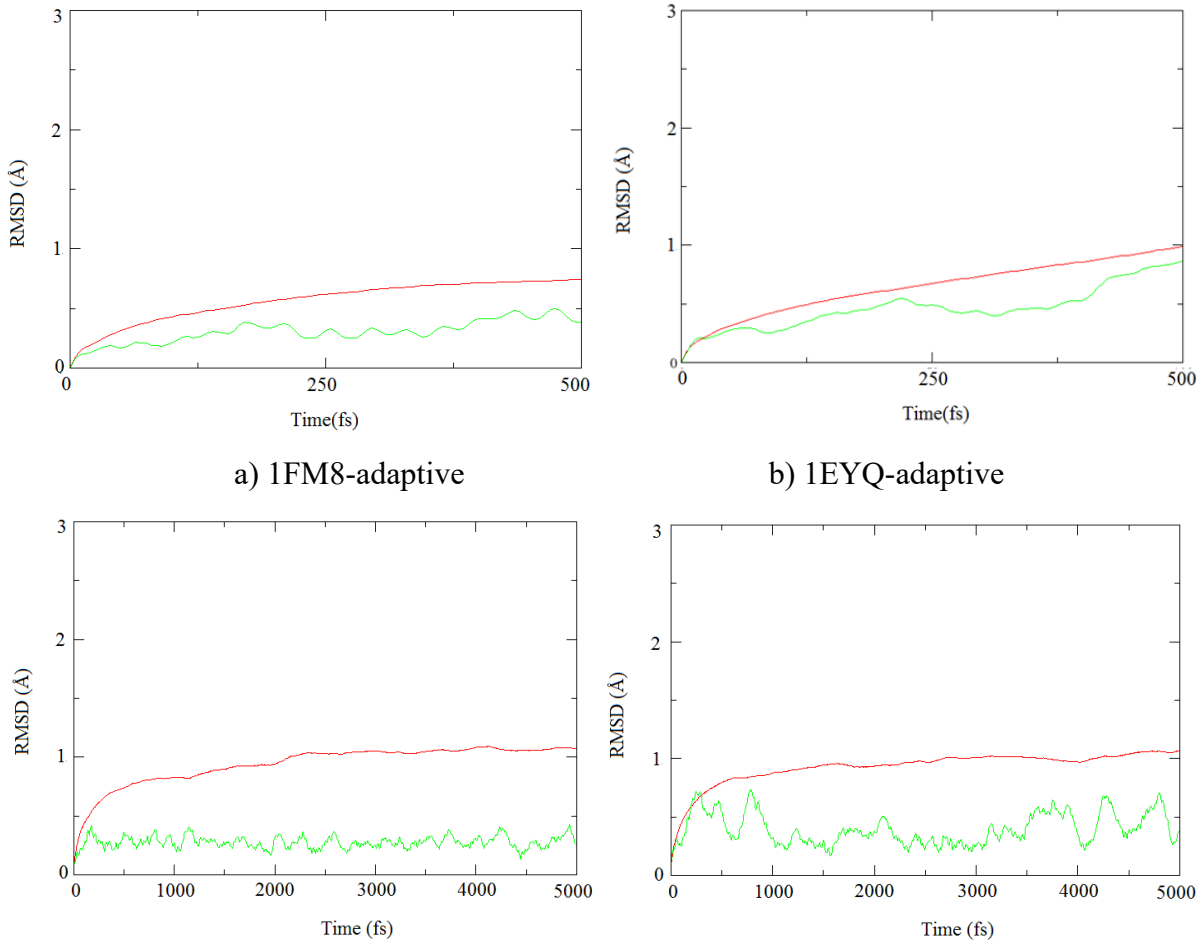


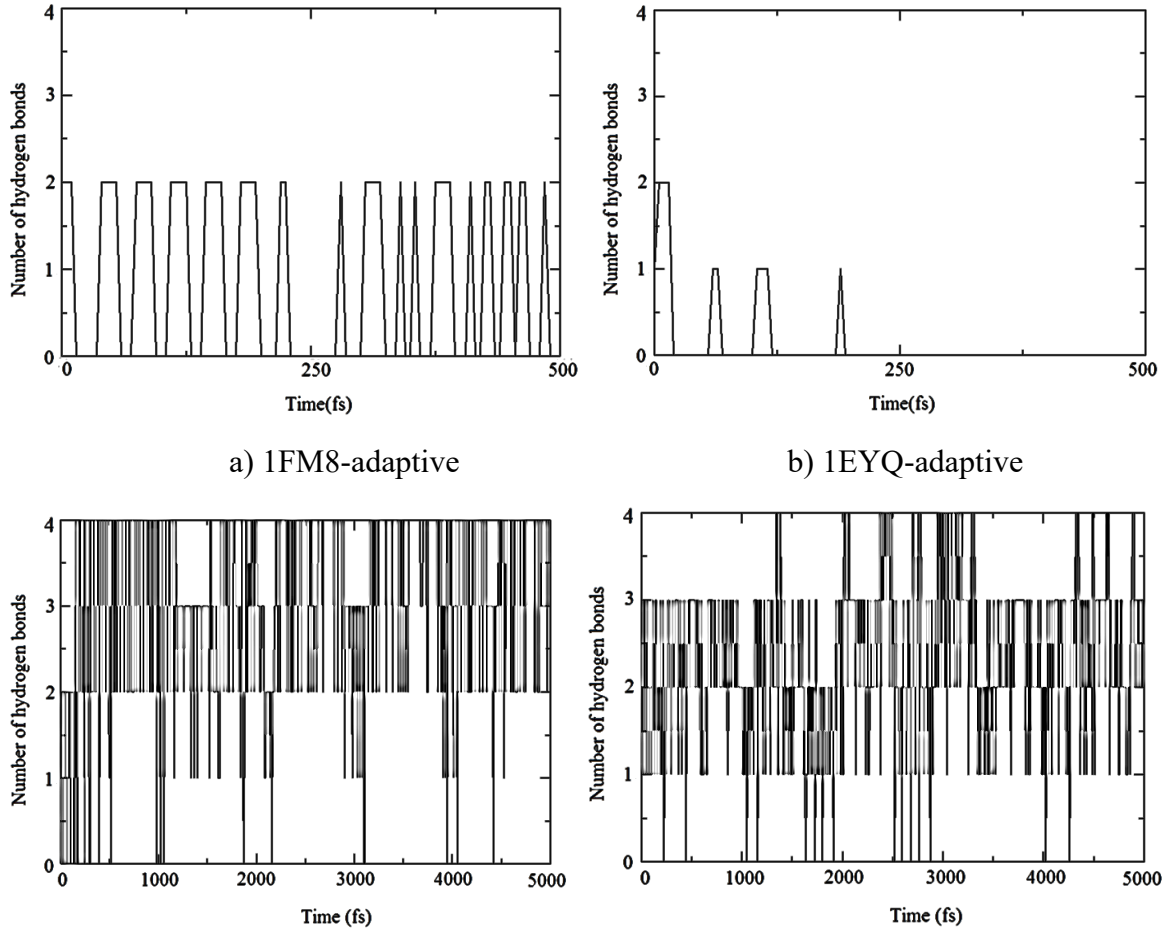
Figure 8.4 RMSD changes during QM/MM and adaptive partitioning QM/MM dynamics simulations. Protein RMSD (——) and ligand RMSD (——). The adaptive partitioning QM/MM RMSD (Figure a and b) is from 0 to 500 fs and corresponds to a highly unequilibrated portion of dynamics, compared to the QM/MM RMSD (figure c and d) which is from 0 to 5000 fs and therefore more equilibrated. The adaptive partitioning simulations are more computationally expensive because of water molecules crossing boundaries, hence less data.

8.3.3 Protein-ligand hydrogen-bonds

Given the increase in accuracy of the dynamics calculation at the orbital level, hydrogen-bonding patterns were of particular interest. The hydrogen-bonding patterns of the 1FM8-*o*-coumaric acid and 1EYQ-*o*-coumaric acid complexes from adaptive partitioning QM/MM [Figures 8.5(a and b)] show that *o*-coumaric acid has two persistent-hydrogen bonds with 1FM8 throughout the 500 fs MD simulations, while *o*-coumaric acid docked to 1EYQ exhibits far less by way of protein-ligand hydrogen-bonding (no H-bonding patterns are observed beyond 250 fs of the simulation). This is consistent with the less stable nature of the *o*-coumaric acid ligand within the active site of 1EYQ (the RMSD of *o*-coumaric acid docked to 1EYQ shows high fluctuations, as discussed in the preceding section). This lack of hydrogen-bonding also suggests the absence of strong interactions between the ligand and 1EYQ. The combination of both the unstable nature of the ligand within the complex and the lack of strong interactions between the ligand and the protein argue against the capacity of 1EYQ to process isomerization. The persisting appearance of hydrogen-bonds between the ligand and 1FM8, on the other hand, suggests that *o*-coumaric acid interacts well (stably and strongly) with 1FM8 in the adaptive QM region, and is thus well bound to 1FM8.

Figures 8.5(c and d) show the hydrogen-bonding of *o*-coumaric acid with 1FM8 and 1EYQ from the QM/MM simulations. These figures would suggest that *o*-coumaric acid has significant H-bonding with both 1FM8 and 1EYQ in a non-adaptive system, but the adaptive portioning data affords a very different perspective.

It is also interesting to note the differences in H-bonding patterns between the QM/MM and adaptive partitioning QM/MM systems. Compare Figure 8.5a with 8.5c or Figure 8.5b with 8.5d. In both adaptive partitioning cases, fewer protein-ligand hydrogen-bonds occur, and it should be noted that there are fewer restraints on solvation of the active site at the adaptive partitioning level. The adaptive partitioning dynamics is predominated not by direct protein ligand H-bonding, but rather by water-mediated H-bonding patterns. In other words, H-bonding of *o*-coumaric acid with the CHI becomes controlled by active-site water molecules.



a) 1FM8-adaptive b) 1EYQ-adaptive

c) 1FM8-QM/MM d) 1EYQ-QM/MM

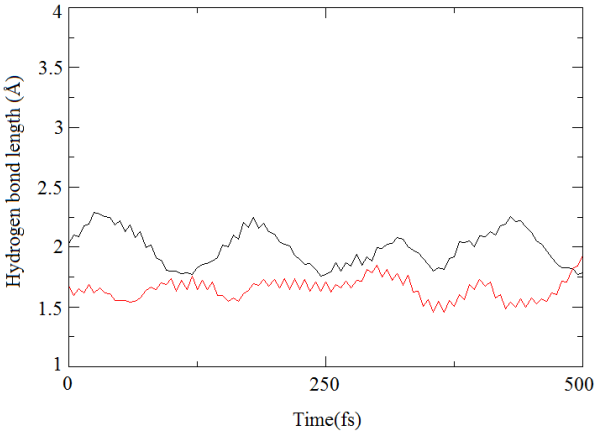
Figure 8.5 Protein-ligand hydrogen-bonds from QM/MM and adaptive partitioning QM/MM dynamics simulations

8.3.4 Hydrogen-bonding with water molecules

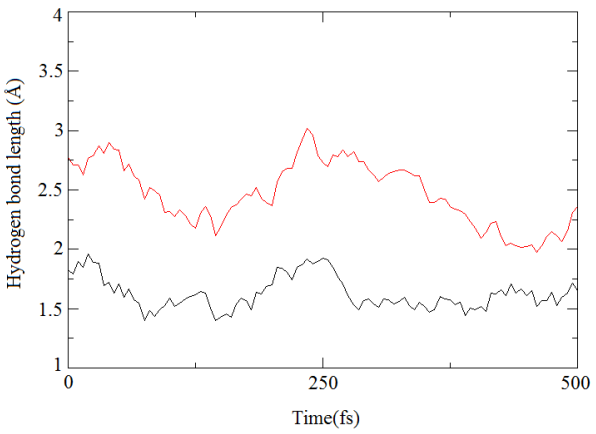
The literature suggests that the enzymatic activities of chalcone isomerase enzymes are water mediated. In general, the aqueous environment is important for ligand mobility to the active site but, for catalytic activity, the involvement of the water varies in its precise nature. This section analyses the interaction of water molecules with the 1FM8-*o*-coumaric acid and 1EYQ-*o*-coumaric acid complexes at the respective active sites. The results show that the adaptive partitioning QM/MM simulated 1FM8-*o*-coumaric acid complex has significantly more interactions with solvent within the active site than the 1EYQ complex.

Figures 8.6(c and d) show the interaction of the 1FM8 active site and *o*-coumaric acid with water molecules. Figure 8.6c shows that the hydrogen-bonding between O(10) of *o*-coumaric acid and Tyr106 (Tyr103) observed in previous simulations has now been bridged with a water molecule. This suggests that the H-bonding between O(10) of *o*-coumaric acid and water is more favourable than the H-bonding between O(10) of *o*-coumaric acid and Tyr106, and this eases the structural proximity requirement of Tyr106 to the *o*-coumaric acid for enzymatic activity. The fluctuation of the H-bonding of the water molecule with O(10) of *o*-coumaric acid and Tyr106 are illustrated in figure 8.6a – both of these H-bonds persist throughout the 500 fs simulation, although the O(10)···H_2O H-bond is generally stronger than the H_2O···Tyr106 (O(10)···H_2O. The former H-bond fluctuates between 1.5 and 2 Å throughout a 500 fs simulation, while the latter H_2O···Tyr106 H-bond fluctuates between 1.5 and 2.5 Å). Figure 8.6c also shows the interaction of H(14) of *o*-coumaric acid with water molecules, H(14) favours H-bonding with Ser107 (Ser104) over H-bonding with water molecules (H(14)···Ser107 H-bond fluctuates between 1.25 and 2 Å, and the H(14)···H_2O H-bond fluctuates between 2 and 3 Å, Figure 8.6b). Figure 8.6d shows the H-bonding of the 1FM8 active site at the final frame of simulation at 500 fs.

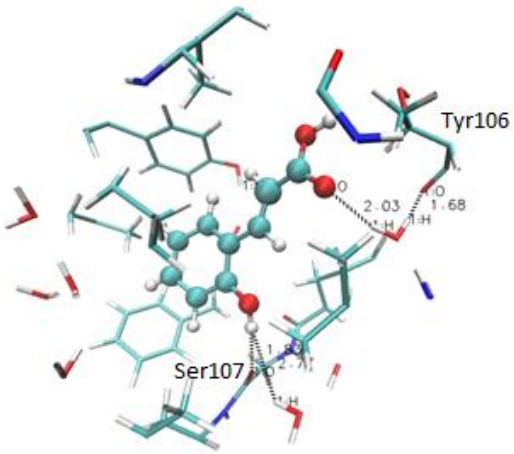
In an adaptive partitioning QM/MM simulations the QM region is continuously updated mostly with some water molecules transiting between the QM region and the buffer region. Figure 8.6e shows the updated QM region at 500 fs; this figure shows that although the H-bonding of the water molecule bridged between O(10) of *o*-coumaric acid and Tyr106 is maintained, the Tyr106 oxygen atom is not part of the dynamic QM region at the end of the simulation but forms part of the buffer region (Figure 8.6e). The 1EYQ-*o*-coumaric acid complex, on the other hand, does not form significant H-bonding with water molecules in the QM region in the 500 fs dynamics simulation; Figure 8.6f shows the PS of 1EYQ-*o*-coumaric acid complex at 500 fs.



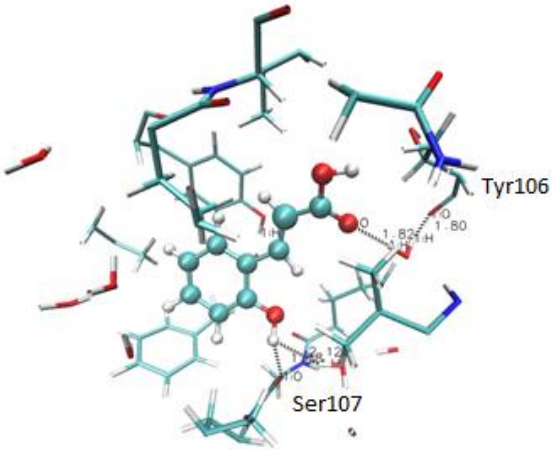
a) O(10)···H_2O (—), H_2O···1FM8 (—)



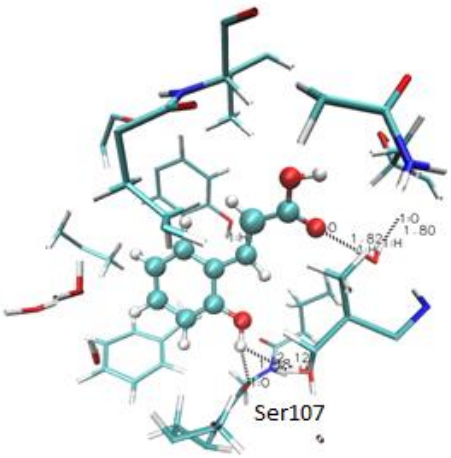
b) H(14)···H_2O(—), H(14)···1FM8 (—)



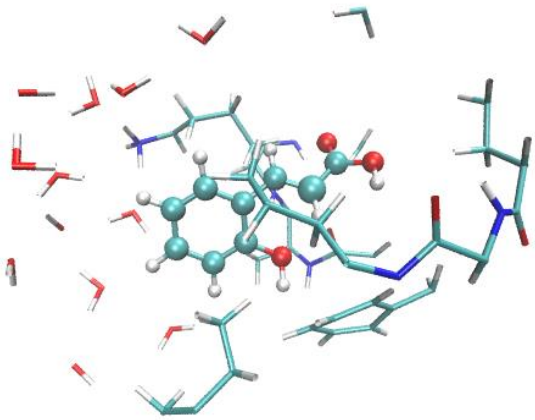
c) 1FM8-initial H-bonding



d) 1FM8-final H-bonding



e) 1FM8-final step QM



f) 1EYQ-final step QM

Figure 8.6 Hydrogen-bonding. 8.6(a and b) illustrate the fluctuation of the H-bonding of the water molecule with *o*-coumaric acid and protein residues. 8.6(c and d) show the interaction of the 1FM8 active site and *o*-coumaric acid with water molecules. 8.6(e and f) show the QM region at the end of the simulations.

8.3.5 The C(7)=C(8) double bond

Chapter 3 provided information on the general trends of the C(7)=C(8) double bond in the simultaneously protonated and deprotonated system; when *o*-coumaric acid is protonated at O(10) and deprotonated at O(12), the C(7)=C(8) double bond generally had a bond length of about 1.38 Å (the length of a single bond from the literature is generally about 1.48 Å). In this section, the C(7)=C(8) double bond was monitored during the adaptive partitioning QM/MM and the QM/MM simulations, to determine whether its length changed due to the free involvement of active site residues and water in a partially protonated and deprotonated state of the system. Chapter 7 showed that the C(7)=C(8) double bond of *o*-coumaric acid, simultaneously protonated and deprotonated in the active site of a CHI, has a bond length of *ca*. 1.37 Å. The C(7)=C(8) double bond in the $E \rightarrow Z$ transition state is 1.395 Å, and monitoring the C(7)=C(8) double bond provides information on the possibility of the protonated and deprotonated state of the system undergoing $E \rightarrow Z$ isomerization.

The C(7)=C(8) double bond of *o*-coumaric acid in the active site of both 1FM8 and 1EYQ fluctuates between 1.3 and 1.4 Å in the adaptive partitioning QM/MM simulations [Figure 8.7(a and b)]. Chapter 3 showed that the general C(7)=C(8) double-bond length of neutral *o*-coumaric acid is about 1.35 Å. Variation of the C(7)=C(8) double-bond length above 1.35 Å (Figure 8.7a) suggests the possibility of this simultaneous partial protonation and deprotonation *o*-coumaric acid within the active site. H-bonding data in the preceding section showed that strong H-bonding contributed to the formation of partial covalent bonds between the species involved in the putative protonation and deprotonation. Given the lengthening of this bond during adaptive partitioning QM/MM simulations, there is the strong possibility that the electronic environment (including solvent and active site residues, contained in the QM region) contributes to the putative partial protonation and deprotonation of *o*-coumaric acid. As suggested by the ONIOM results in Chapter 7, only a partial protonation/deprotonation is necessary to effect the isomerization of *o*-coumaric acid. Figures 8.7(c and d) show the C(7)=C(8) double-bond length fluctuations from the QM/MM simulations. The variation in the C(7)=C(8) double-bond length also suggests the occurrence of a partially protonated and deprotonated *o*-coumaric acid (there are clearly occurrences of bond-length fluctuation above 1.35 Å and, in some instances, the C(7)=C(8) double-bond length reaches 1.4 Å).

However, similar C(7)=C(8) double-bond length fluctuations are also evident in the 1EYQ complex, which has been included as a negative control. It may be that the time-length of the adaptive partitioning molecular dynamics simulation, which, due to computational cost, allows for only 500fs, is too short to detect fluctuations which may occur during time-scales orders of magnitude longer.

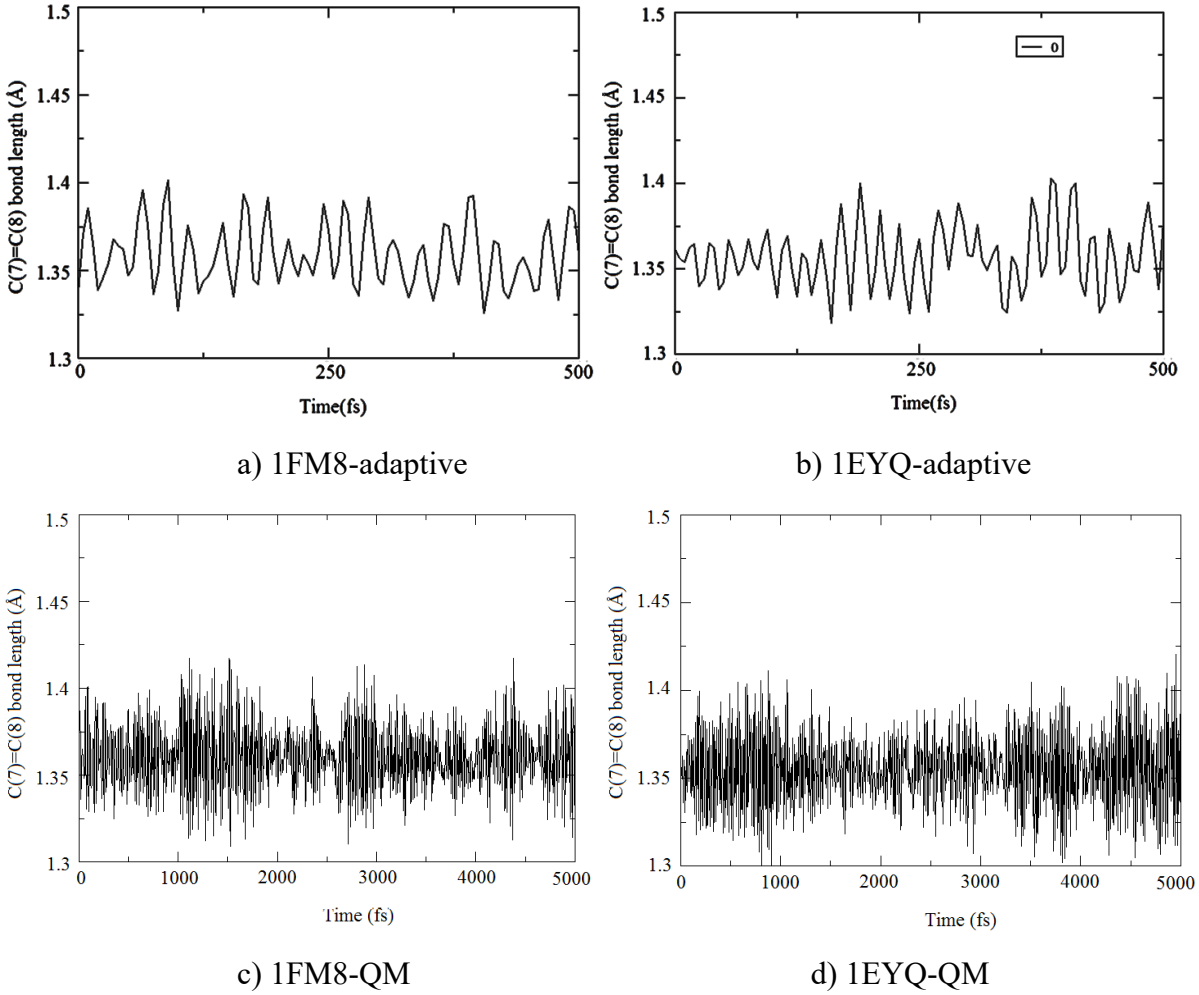


Figure 8.7 Variation of the C(7)=C(8) double-bond length during dynamics simulations. The scan for the adaptive partitioning QM/MM (a and b) is from 0 to 500 fs and the scan for the QM/MM (c and d) is from 0 to 5000 fs.

8.3.6 C(1)-C(7)-C(8)-C(9) dihedral

Fluctuations in the C(1)-C(7)-C(8)-C(9) dihedral angle of *o*-coumaric acid docked to 1FM8 and 1EYQ do not show any clear trends towards *E*→*Z* isomerization for both types of simulation considered. This is consistent with the energy profiles in Section 8.3.1 as significant changes in the C(1)-C(7)-C(8)-C(9) dihedral angle would have resulted in significant energy changes. The dihedral angel fluctuations in *o*-coumaric acid do not show substantial changes of the *E*-*o*-coumaric acid C(1)-C(7)-C(8)-C(9) angle with potential of interconverting to *Z*-*o*-coumaric acid. These simulations were carried out at 300 K on CP2K, perhaps simulations at a slightly high temperatures would favour the decrease or increase of the dihedrals (depending on the direction of rotation of the dihedral), however, increments of the temperature should be cautious so as to avoid denaturing of the protein. One other option, as mentioned previously is to carry out a metadynamics simulation focused on the rotation of the C(1)-C(7)-C(8)-C(9) dihedral angle to achieve the *E*→*Z* isomerization.

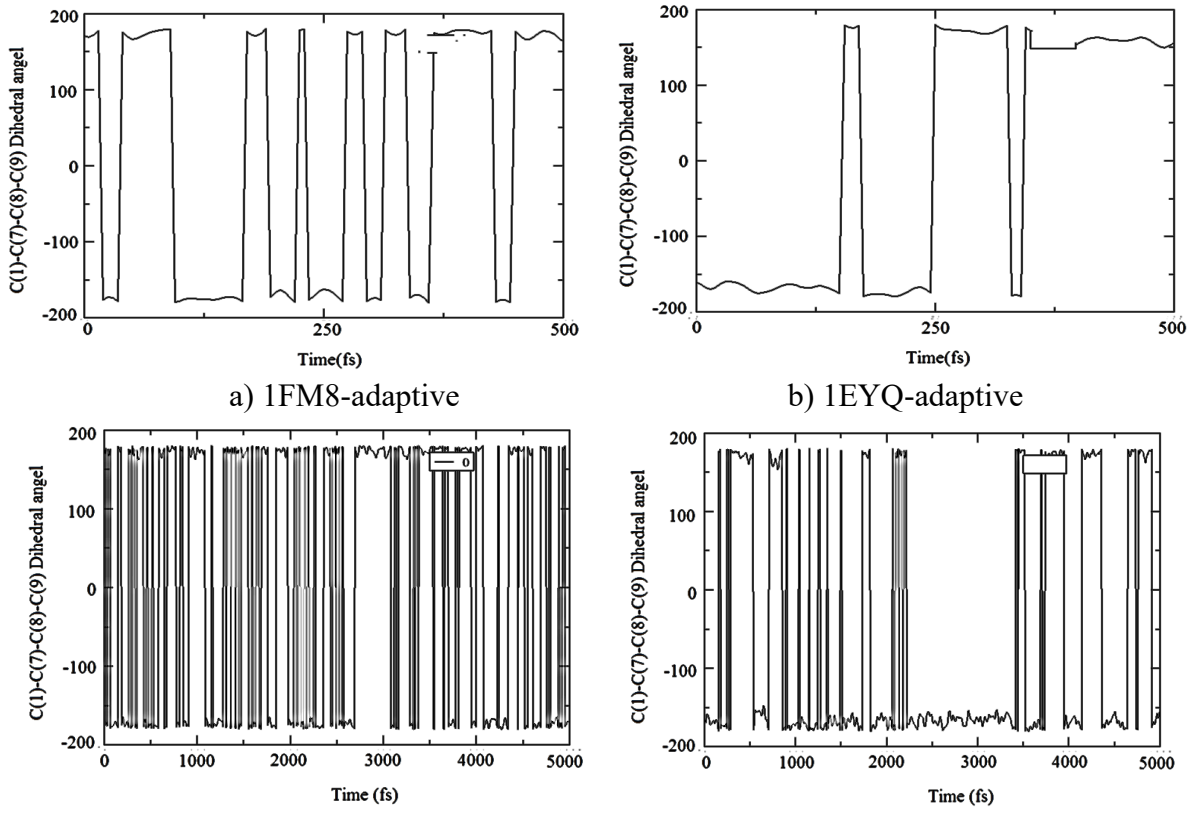


Figure 8.8 C(1)-C(7)-C(8)-C(9) dihedral bond fluctuations during dynamics simulations. The change in the dihedral angle is not due to the *E*→*Z* isomerization, rather due to discontinuity of measurement of C(9) between 180° and -180° with respect to the benzene ring.

8.4 Summary of the chapter

The RMSD of the simulations with CP2K showed that *o*-coumaric acid is more stable in the active site of 1FM8 than in the active site of 1EYQ □□an observation which is consistent with the results from the previous Chapters. Moreover, the adaptive partitioning QM/MM shows that the sp^2 oxygen of the alkenoic acid moiety favours H-bonding with water molecules over H-bonding with protein residues, and that the water mediates interaction between the substrate and the protein *only in the 1FM8 case*. This is consistent with the previous studies of the enzymatic activities of CHIs from literature (104,116,249); it also suggests that the putative CHI interconversion of *E*-*o*-coumaric acid to *Z*-*o*-coumaric acid may be a water-mediated process with indirect action by specific residues of 1FM8, and maintains the possibility of 1FM8 being a potential isomerase for this substrate.

CHAPTER 9
SUMMARY AND CONCLUSIONS

9.1 Introduction to the chapter

This chapter provides an overview of this study, starting from the study of the population analysis of cinnamic acid derivatives through to the adaptive partitioning QM/MM study of the chalcone isomerase enzyme-*o*-coumaric acid complexes. The main objective of this research project has been the application of high-level theoretical methods to explore the molecular basis of the enzyme-catalysed E → Z isomersation *o*-coumaric acid, within the context of three critical phases in the biosynthesis of coumarin derivatives, *viz*., of:

i) *E* → *Z* isomersation of the (*E*)-*o*-coumaric acid substrate;
ii) cyclisation (lactonisation) of the (*Z*)-*o*-coumaric acid to the hemi-acetal intermediate; and
iii) dehydration of the hemi-acetal intermediate to afford the coumarin derivative.

Before embarking on a full computational study of the bioorganic E → Z isomersation of (*E*)-*o*-coumaric acid, it was essential to establish in detail the chemical properties of (*E*)-*o*-coumaric acid as a substrate, given the local environments in which it has to reside within the active site of an enzyme. On the assumption that the E → Z isomerisation step may be mediated by concomitant protonation and deprotonation of the substrate within the enzyme active site, the behaviour of the different protonation/deprotonation states of (*E*)-*o*-coumaric acid and their reactivities were therefore explored in the first step of the study. This involved conformational population analysis of the different protonation/deprotonation states of (*E*)-*o*-coumaric acid and the results were discussed in Chapter 2.

Since no enzymes have, as yet, been identified as being responsible for any of the critical phases (i to iii above) in the biosynthesis of coumarin derivatives, *in silico* molecular docking studies were undertaken to establish the capacity of various, known chalcone isomerase enzymes (CHIs) to accommodate *o*-coumaric acid in their active sites. The nature of the interactions of

o-coumaric acid with the active sites were explored after docking – a process which included searching for protein residues in the active sites that were in close proximity to the ligand and which might thus be capable of facilitating the E→Z isomerization of (*E*)-*o*-coumaric acid. Since CHIs are the enzymes associated with the interconversion of chalcones to flavones, a thorough study of the general receptor-ligand interactions between expected chalcone substrates and CHIs was undertaken using high throughput virtual screening. Molecular dynamics simulations were then performed to study the conformational evolution of the *E*-*o*-coumaric acid ligand within the binding site of CHIs over a set time-frame and to determine if the docked *o*-coumaric acid remained in the binding site over time. The dynamics simulations were also used to explore other activities and processes in the binding site, including the types of receptor-ligand interaction that persist over time. This was followed by an ONIOM study to explore the putative protonation and deprotonation of (*E*)-*o*-coumaric acid by CHI residues, and the subsequent E→Z isomerization. In the final step in the study, the complexity of the simulation was increased by employing an adaptive partitioning QM/MM approach to obtain a more realistic simulation which involved studying the system in a water-solvated medium at a highly accurate level.

9.2 Conformational and protonation state population analysis of cinnamic acid derivatives

The chapter on the conformational population analysis of cinnamic acid derivatives aimed at studying the different protonation states of several cinnamic acid derivatives, with particular focus on (*E*)-*o*-coumaric acid. The results showed the following general trends in the relative energy with respect to protonation state:

Corrected relative energies trends:

protonated systems < neutral systems < simultaneously protonated and deprotonated structures < deprotonated systems

The energies of the simultaneously protonated and deprotonated system were generally about 20 kcal/mol higher than the energies of the neutral structures. The observed trends in the C(7)=C(8) double-bond length are as follows:

Derivatives protonated at O(10) and deprotonated at O(12) generally had a bond length of around 1.38 Å. This bond length is second closest to that resulting from protonation at either C(7) or C(8) where the bond length is generally around 1.47 Å. The other protonation states had a C(7)=C(8) double-bond length less than 1.38 Å.

The general trends for the frontier orbitals and their energy differences are as follows:

The LUMO energies of the deprotonated systems are generally higher, by more than a 100 kcal/mol (over 150 kcal/mol for systems deprotonated at two positions), than the LUMO energies of the neutral system. The LUMO energies of the protonated systems are lower (by as much as 100 kcal/mol less in some cases) than the energies of a neutral system. Simultaneous protonation and deprotonation of cinnamic acid derivatives does not cause a significant change in the LUMO energy when compared to the energy changes resulting from either protonating or deprotonating a cinnamic acid derivative.

The HOMO-LUMO energy gap results showed that, generally, the simultaneously protonated and deprotonated systems had the smallest HOMO-LUMO energy gaps, compared to the other protonation states.

Based on these general observations, a mechanism that involves simultaneous protonation and deprotonation of coumaric acid was considered most likely to induce chemical change, more specifically because a simultaneously protonated deprotonated system has the lowest HOMO-LUMO energy gap.

9.3 *In silico* molecular docking

The structures of (*E*)-coumaric acid conformers obtained from the population analysis were docked into eleven wild and mutant CHIs. The results showed that the "A" and "S" conformers of (*E*)-*o*-coumaric acid maintain their "A" and "S" conformation after docking. During docking in an ideal ensemble, where both the "S" and the "A" conformers of *o*-coumaric acid are highly populated, the receptor might not be highly selective of conformations, since the binding affinities of both the "A" and the "S" conformers are relatively close. The "S" conformer appears to have stronger protein-ligand interactions with the CHI crystal structures than the "A" conformer and, in principle, this is good news because the "S" conformer is the global minimum obtained from the QM study.

Docking results also showed that the ligands in the CHIs are mostly held by hydrophobic π-π or π-alkyl interactions.

The binding of (*E*)-*o*-Coumaric acid to 1FM8 gave the best binding energy of all the CHIs considered in this study (7.1 kcal/mol). There were also more conventional hydrogen bonds present in this particular binding that could facilitate the protonation and deprotonation of *o*-coumaric acid. From these results, 1FM8 was identified as a potential CHI for further computational studies. However, a more in-depth understanding of how the various CHI crystal structures interacted with expected substrates was required, and the next chapter involved High Throughput Virtual Screening of the CHIs with a series of chalcone derivatives.

9.4 High throughput virtual screening

High throughput virtual screening (HTVS) showed that the structural feature of chalcones that enhances their binding affinities is the presence of additional aromatic rings in the chalcone system, which facilitate hydrophobic interactions with the receptor. Benzene rings fused to the chalcone moieties (the acetophenone-derived or benzaldehyde-derived portions of the chalcone) gave better binding affinity than phenyl substituents linked to either moiety *via* a single bond. However, the cinnamic acid derivatives considered in other chapters in this study

have only one aromatic moiety in the form of a single benzene ring. The best binding affinities of these cinnamic acid derivatives is generally about -7 kcal/mol, not nearly as strong as the best binding affinities of chalcones to the same receptors (the best binding affinity with chalcones to CHIs is -12 kcal/mol). From the data available we can deduce that to improve the binding affinity of cinnamic acid derivatives, benzannulated analogues could be considered to be better ligands, and these CHIs might well be better suited to mediate the synthesis of benzannulated coumarins. However, that would miss the point of attempting to identify CHIs capable of mediating the synthesis of coumarin from (*E*)-*o*-coumaric acid!

Other additional polar substituents on the chalcones, such as OH, $N(CH_3)_2$ and alkyl chains, reduce the binding affinity of these chalcones. This suggests that adding such substituents to cinnamic acid derivatives could well reduce their binding affinities.
The π-alkyl hydrophobic interactions seem to have an important role in stabilising the CHI-ligand complexes – this is observed with the chalcone ligand and the cinnamic acid ligands. Complexes with 1FM7 and 1FM8 exhibited the best binding affinities and protein-ligand interactions with both the cinnamic acid derivatives and the chalcone derivatives.

9.5 Molecular dynamics

Molecular dynamics results were in agreement with the results from the preceding study in that. 1FM8 was again found to be the most suitable of all the CHIs considered for further studies concerning the bioorganic transformation of *o*-coumaric acid to coumarins. Analysis of the MD results showed that the 1FM8-*o*-coumaric acid complex has a steady RMSD with molecular dynamics performed using both CHARMM and GROMACS. During most of the simulation, the *o*-coumaric acid remained docked to the active site of 1FM8, and the receptor ligand hydrogen-bonding patterns were persistent for a total of two hydrogen bonds for both the 30 ns CHARMM simulation and the 100 ns GROMACS simulation. Analysis of the active site shows that the H-bonding of the sp^2 oxygen of the alkenoic acid chain of *o*-coumaric acid and Tyr103 (VMD relative numbering, Tyr106 in literature) and the H-bonding of the phenolic OH of *o*-coumaric acid with ASN110 persisted throughout the simulation, fluctuating in both cases at a distance of *ca*. 2 Å. This suggests that this particular complex is set up in an optimal manner for simultaneous protonation and deprotonation of the substrate.

9.6 ONIOM study

Because the computational studies leading to the ONIOM study showed that 1FM8 is more favourable than the other CHIs for the isomerization of (*E*)-*o*-coumaric acid, the direct protonation and deprotonation of the acid by protein residues in the 1FM8 protein-(*E*)-*o*-coumaric acid complex was explored at the ONIOM mixed QM/MM level. The protonation and deprotonation was successful and was followed by the *E*→*Z* isomerization. The highest energy point in a rotational search mimicking the *E*→*Z* isomerization rotation was used as input for a transition state search which successfully generated a complex with a single imaginary frequency. This transition state was then used in an IRC analysis to explore the reaction pathway. The resulting pathway shows that the *E*→*Z* isomerization does not result from complete protonation and deprotonation of o-coumaric acid, but is facilitated by strong and varying hydrogen-bonding interactions between o-coumaric acid and 1FM8.

9.7 Adaptive partitioning QM/MM molecular dynamics

Although the ONIOM study showed that 1FM8 is theoretically capable of facilitating the *E*→*Z* isomerization of o-coumaric acid and, hopefully, the subsequent lactonization, the ONIOM study was performed *in vacuo*. Therefore, to take into account the involvement of water in the chemical transformations taking place in the active site of the 1FM8- o-coumaric acid complex, the system was further studied in a water-simulated environment using both QM/MM based dynamics and the adaptive partitioning QM/MM based dynamics as implemented in in the CP2K package. The 1EYQ-*o*-coumaric acid complex was also simulated for comparative purposes, but failed show the characteristics required for the interconversion of (*E*)-*o*-coumaric acid to (*Z*)-*o*-coumaric acid. The results for the 1FM8 simulation showed that the RMSD stabilized quickly, compared to the 1EYQ simulation; a result consistent with the results from the previous Chapters. Adaptive partitioning QM/MM based dynamics showed that the sp^2 oxygen of the alkenoic acid chain favours H-bonding with water molecules over H-bonding with protein residues, and this is consistent with the previous studies of the enzymatic activities of CHIs in the literature (104,116,249); it also suggests that the putative CHI interconversion

from (*E*)-*o*-coumaric acid to (*Z*)-*o*-coumaric acid is a water-mediated process within the enzyme active site. Adaptive partitioning QM/MM-based molecular dynamics also showed that the protonation and deprotonation of *o*-coumaric acid involves partial proton transfer (i.e. strong H-bonding) rather than the complete migration of a proton form one oxygen to another.

9.8 Overall conclusions

For any further experimental studies involving bioorganic interconversion of (*E*)-*o*-coumaric acid derivatives to coumarins, 1FM8 (the CHI from *M. Sativa*) is recommended. 1FM8 exhibited: i) the best binding affinity to (*E*)-*o*-coumaric acid in molecular docking studies; ii) the desired H-bonding interactions that were sustained in the thermally equilibrated system during the molecular dynamics simulation; iii) direct protonation and deprotonation of o-coumaric acid by protein active site residues; in the ONIOM study; iv) strong water-mediated H-bonding between the putative protonating and deprotonating species in the adaptive partitioning study.

Further study will involve the investigation of the lactonization of the Z-*o*-coumaric acid within the active site of this enzyme.

9.9 Future studies

Further computational studies are expected to involve the investigation of the lactonization of the (*Z*)-*o*-coumaric acid within the active site of the 1FM8 enzyme. Figure 9.1 shows the orientation of (*Z*)-*o*-coumaric aid in the active site of 1FM8 and the two protein residues with the potential of facilitating the lactonization. *In silico* results will also be complemented with the isolation of 1FM8 for biochemical and, if appropriate, protein-NMR kinetic studies. However, given that isomerization leads to this complex with the required protein ligand interactions for this lactonization to occur, we have a proof of concept for the biosynthesis of coumarins from coumaric acid mediated by 1FM8. Figure 9.2 provides a proposed mechanism for this final step.

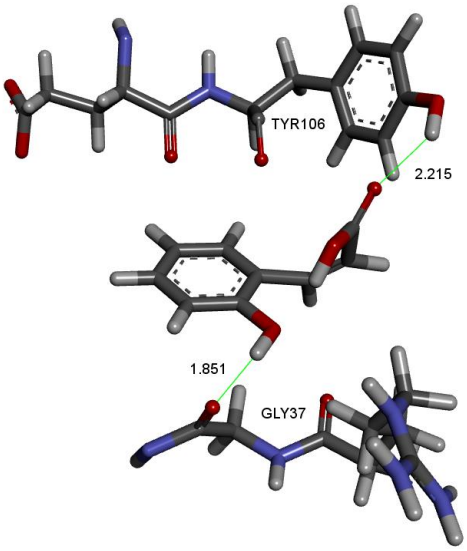


Figure 9.1. (*Z*)-*o*-coumaric acid and the 1FM8 residues with the potential of facilitating the lactonization. Part of the active site of the (Z)-o-coumaric acid- 1FM8 complex optimised from the IRC calculations in reported in Chapter 7.

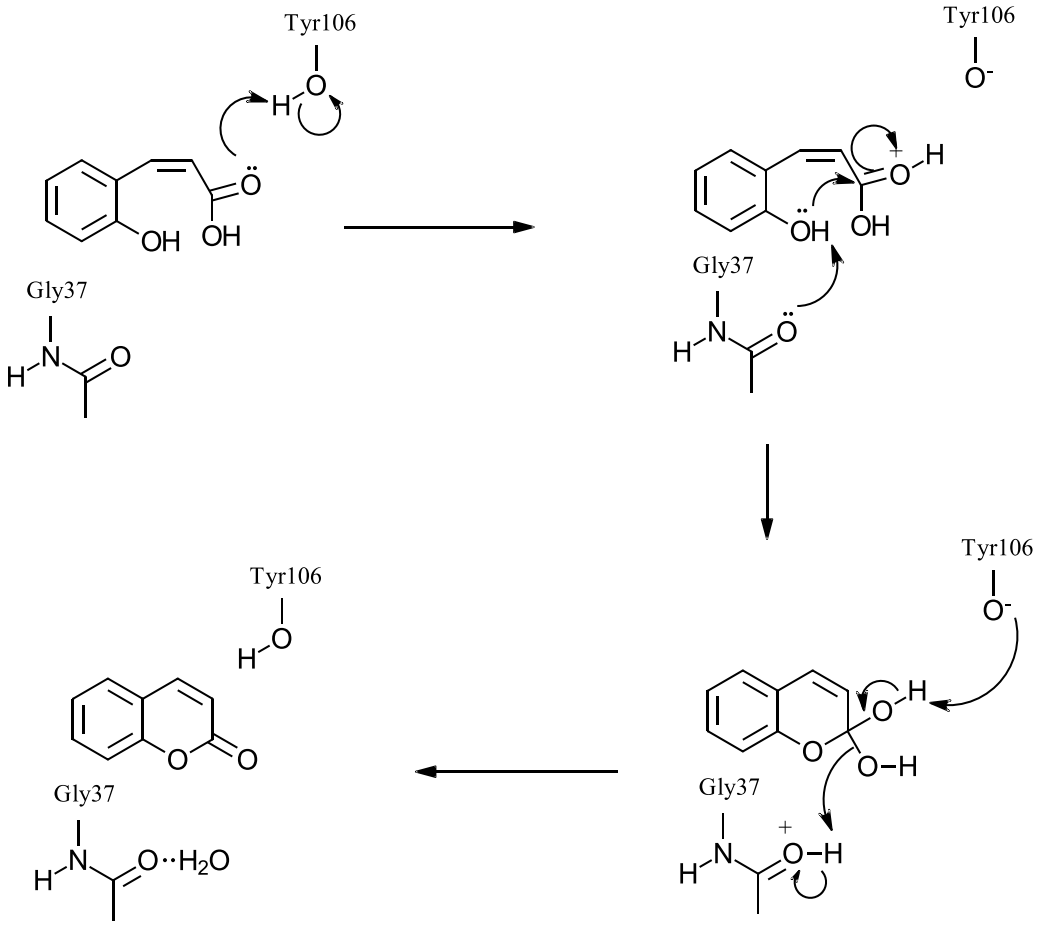


Figure 9.2. Schematic representation of the lactonization of (*Z*)-*o*-coumaric acid by 1FM8 protein residues Tyr106 and Gly37.

References

1. Venugopala KN, Rashmi V, Odhav B. Review on natural coumarin lead compounds for their pharmacological activity. Biomed Res Int. 2013;2013(Table 1).
2. Lake BG. Coumarin metabolism, toxicity and carcinogenicity: relevance for human risk assessment. Food Chem Toxicol. 1999;37:423–53.
3. Bourgaud F, Hehn A, Larbat R, Doerper S, Gontier E, Kellner S, et al. Biosynthesis of coumarins in plants: A major pathway still to be unravelled for cytochrome P450 enzymes. Phytochem Rev. 2006;5(2–3):293–308.
4. Marumoto S, Miyazawa M. Structure-activity relationships for naturally occurring coumarins as -secretase inhibitor. Bioorganic Med Chem [Internet]. 2012;20(2):784–8. Available from: http://dx.doi.org/10.1016/j.bmc.2011.12.002
5. Jain PK, Joshi H. Coumarin: Chemical and pharmacological profile. J Appl Pharm Sci. 2012;2(6):236–40.
6. Lacy A, O'Kennedy R. Studies on coumarins and coumarin-related compounds to determine their therapeutic role in the treatment of cancer. Curr Pharm Des. 2004;10(30):3797–811.
7. Abdel-Wahab BF, Mohamed HA, Farhat AA. Ethyl coumarin-3-carboxylate: Synthesis and chemical properties. Org Commun. 2014;7(1):1–27.
8. A. Shaabani, R. Ghadari AR and A. R. Coumarin Synthesis via Knoevenagel Condensation Reaction in 1,1,3,3-N,N,N',N'- Tetramethylguanidinium Trifluoroacetate Ionic Liquid. J iran Chem Soc. 2009;6(4):710–4.
9. Kalkhambkar RG, Kulkarni GM, Kamanavalli CM, Premkumar N, Asdaq SMB, Sun CM. Synthesis and biological activities of some new fluorinated coumarins and 1-aza coumarins. Eur J Med Chem. 2008;43(10):2178–88. Available from: http://dx.doi.org/10.1016/j.ejmech.2007.08.007
10. De Souza SM, Delle Monache F, Smânia A. Antibacterial activity of coumarins. Zeitschrift fur Naturforsch - Sect C J Biosci. 2005;60(9–10):693–700.
11. WHO. WHO | Cancer. Fact sheet N°297. 2013. Available from: http://www.who.int/mediacentre/factsheets/fs297/en/index.html
12. Miyake Y, Murakami A, Sugiyama Y, Isobe M, Koshimizu K, Ohigashi H. Identification of coumarins from lemon fruit (Citrus limon) as inhibitors of in vitro tumor promotion and superoxide and nitric oxide generation. J Agric Food Chem.

1999;47(8):3151–7.

13. Thati B, Noble A, Creaven BS, Walsh M, McCann M, Devereux M, et al. Role of cell cycle events and apoptosis in mediating the anti-cancer activity of a silver(I) complex of 4-hydroxy-3-nitro-coumarin-bis(phenanthroline) in human malignant cancer cells. Eur J Pharmacol. 2009;602(2–3):203–14. Available from: http://dx.doi.org/10.1016/j.ejphar.2008.11.020
14. Ivancheva S, Nikolova M, Tsvetkova R, Academy B. Pharmacological activities and biologically active compounds of Bulgarian medicinal plants. Res Signpost. 2006;661(2):87–103.
15. Sashidhara K V., Kumar A, Kumar M, Sarkar J, Sinha S. Synthesis and in vitro evaluation of novel coumarin–chalcone hybrids as potential anticancer agents. Bioorg Med Chem Lett. 2010;20(24):7205–11. Available from: http://dx.doi.org/10.1016/j.bmcl.2010.10.116
16. Lee S, Lee YS, Jung SH, Shin KH, Kim BK, Kang SS. Antioxidant activities of decursinol angelate and decursin from Angelica gigas roots. Nat Prod Sci. 2003;9(3):170–3.
17. Yu J, Wang L, Walzem RL, Miller EG, Pike LM, Patil BS. Antioxidant activity of citrus limonoids, flavonoids, and coumarins. J Agric Food Chem. 2005;53(6):2009–14.
18. Lin HC, Tsai SH, Chen CS, Chang YC, Lee CM, Lai ZY, et al. Structure-activity relationship of coumarin derivatives on xanthine oxidase-inhibiting and free radical-scavenging activities. Biochem Pharmacol. 2008;75(6):1416–25.
19. Witaicenis A, Seito LN, Da Silveira Chagas A, De Almeida LD, Luchini AC, Rodrigues-Orsi P, et al. Antioxidant and intestinal anti-inflammatory effects of plant-derived coumarin derivatives. Phytomedicine. 2014;21(3):240–6. Available from: http://dx.doi.org/10.1016/j.phymed.2013.09.001
20. Chaudhary G, Goyal S, Poonia P. Lawsonia inermis Linnaeus : A Phytopharmacological Review. Int J Pharm Sci Drug Res. 2010;2(2):91–8.
21. Melagraki G, Afantitis A, Igglessi-Markopoulou O, Detsi A, Koufaki M, Kontogiorgis C, et al. Synthesis and evaluation of the antioxidant and anti-inflammatory activity of novel coumarin-3-aminoamides and their alpha-lipoic acid adducts. Eur J Med Chem. 2009;44(7):3020–6. Available from: http://dx.doi.org/10.1016/j.ejmech.2008.12.027
22. Stanchev S, Hadjimitova V, Traykov T, Boyanov T, Manolov I. Investigation of the antioxidant properties of some new 4-hydroxycoumarin derivatives. Eur J Med Chem. 2009;44(7):3077–82. Available from: http://dx.doi.org/10.1016/j.ejmech.2008.07.007

23. Roussaki M, Kontogiorgis CA, Hadjipavlou-Litina D, Hamilakis S, Detsi A. A novel synthesis of 3-aryl coumarins and evaluation of their antioxidant and lipoxygenase inhibitory activity. Bioorganic Med Chem Lett. 2010;20(13):3889–92. Available from: http://dx.doi.org/10.1016/j.bmcl.2010.05.022
24. Cai Y, Luo Q, Sun M, Corke H. Antioxidant activity and phenolic compounds of 112 traditional Chinese medicinal plants associated with anticancer. Life Sci. 2004;74(17):2157–84.
25. Rauha J-PP, Remes S, Heinonen M, Hopia A, Kähkönen M, Kujala T, et al. Antimicrobial effects of Finnish plant extracts containing flavoniods and other phenolic compounds. Int J Food Microbiol. 2000;56(1):3–12. Available from: file://g/Library/Reference Manager/Papers/1564. International Journal of Food Microbiology. 2000, 56, 3-12.pdf%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/10857921
26. Ojala T, Remes S, Haansuu P, Vuorela H, Hiltunen R, Haahtela K, et al. Antimicrobial activity of some coumarin containing herbal plants growing in Finland. J Ethnopharmacol. 2000;73(1–2):299–305.
27. Manolov I, Maichle-Moessmer C, Danchev N. Synthesis, structure, toxicological and pharmacological investigations of 4-hydroxycoumarin derivatives. Eur J Med Chem. 2006;41(7):882–90. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0223523406000973
28. Stanchev S, Momekov G, Jensen F, Manolov I. Synthesis, computational study and cytotoxic activity of new 4-hydroxycoumarin derivatives. Eur J Med Chem. 2008;43(4):694–706.
29. Hirsh J, Fuster V. AHA Medical / Scientific Statement Special Report Guide to Anticoagulant Therapy Part 2 : Oral Anticoagulants. 1984;
30. AIDS in the World - Jonathan M.
31. WHO. World Health Statistics 2010 - World Health Organization - Google Books. Available from: https://books.google.co.ke/books?hl=en&lr=&id=Z69vxfRfFIsC&oi=fnd&pg=PA1&dq=mdgs+and+health&ots=cGQIeoHJ4w&sig=3NOEoFfxobVt2jM2QB7GlWRHedc&redir_esc=y#v=onepage&q=mdgs and health&f=false
32. Singh IP, Bharate SB, Bhutani KK. Anti-HIV natural products. Vol. 89, Current Science. 2005. p. 269–90.
33. Yu D, Suzuki M, Xie L, Morris-Natschke SL, Lee KH. Recent progress in the

development of coumarin derivatives as potent anti-HIV agents. Med Res Rev. 2003;23(3):322–45.

34. Olomola TO, Klein R, Lobb KA, Sayed Y, Kaye PT. Towards the synthesis of coumarin derivatives as potential dual-action HIV-1 protease and reverse transcriptase inhibitors. Tetrahedron Lett. 2010;51(48):6325–8. Available from: http://dx.doi.org/10.1016/j.tetlet.2010.09.121
35. Matthee G, Wright AD, Konig GM. HIV reverse transcriptase inhibitors of natural origin. Planta Med. 1999;65(6):493–506.
36. Kostova I, Raleva S, Genova P, Argirova R. Structure-activity relationships of synthetic coumarins as HIV-1 inhibitors. Bioinorg Chem Appl. 2006;2006:1–9.
37. Ng T. B, Huang B, Fong W. P, Yeung H. W. Anti-human immunodeficiency virus (anti-HIV) natural products with special emphasis on HIV reverse transcriptase inhibitors. Life Sci. 1997;61(10):933–49. Available from: http://www.sciencedirect.com/science/article/pii/S0024320597002452%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/9296332
38. El-gewely R. Biotechnology Annual Review. Review Literature And Arts Of The Americas. 2005. p. 452.
39. Karami B, Kiani M. Synthesis of the Coumarins via Pechmann Method in the Presence of Environmentally Friendly $Y(NO_3)_3 \times 6H_2O$. J Chinese Chem Soc. 2014;61(2):213–6. Available from: http://doi.wiley.com/10.1002/jccs.201200610
40. Chavan OS, Baseer MA. Comparative study of various synthetic methods of 7-hydroxy-4-methyl coumarins via Pechmann reaction. Der Chem Sin. 2014;5(5):67–70.
41. Olomola TO, Klein R, Caira MR, Kaye PT. Elucidating Latent Mechanistic Complexity in Competing Acid-Catalyzed Reactions of Salicylaldehyde-Derived Baylis-Hillman Adducts. J Org Chem. 2016;81(1):109–20.
42. Faridoon, Olomola TO, Tukulula M, Klein R, Kaye PT. Strong base- or acid-mediated chemoselectivity shifts in the synthesis of 2H-chromene or coumarin derivatives from common Baylis-Hillman adducts. Tetrahedron. 2015;71(29):4868–73. Available from: http://dx.doi.org/10.1016/j.tet.2015.04.104
43. Faridoon, Olomola TO, Klein R, Kaye PT. Application of Baylis-Hillman methodology in the direct construction of chromone derivatives. Tetrahedron. 2016;72(3):392–5. Available from: http://dx.doi.org/10.1016/j.tet.2015.11.039
44. Olomola TO, Mosebi S, Klein R, Traut-Johnstone T, Coates J, Hewer R, et al. Novel furocoumarins as potential HIV-1 integrase inhibitors. Bioorg Chem. 2014;57:1–4.

Available from: http://dx.doi.org/10.1016/j.bioorg.2014.07.008

45. Olasupo I, Rose NR, Klein R, Adams LA, Familoni OB, Kaye PT. Evaluation of Baylis–Hillman Routes to 3-(Aminomethyl)coumarin Derivatives. Synth Commun. 2014;44(2):251–8. Available from: http://www.tandfonline.com/doi/abs/10.1080/00397911.2013.803575
46. Martins D, Nunez CV. Secondary metabolites from Rubiaceae species. Molecules. 2015;20(7):13422–95.
47. Alhassan AM, Abdullahi MI, Uba A, Umar A. Prenylation of aromatic secondary metabolites: A new frontier for development of novel drugs. Trop J Pharm Res. 2014;13(2):307–14.
48. Liu T, Yao R, Zhao Y, Xu S, Huang C, Luo J, et al. Cloning, Functional Characterization and Site-Directed Mutagenesis of 4-Coumarate: Coenzyme A Ligase (4CL) Involved in Coumarin Biosynthesis in Peucedanum praeruptorum Dunn. Front Plant Sci. 2017;8(January):1–14. Available from: http://journal.frontiersin.org/article/10.3389/fpls.2017.00004/full
49. Vialart G, Hehn A, Olry A, Ito K, Krieger C, Larbat R, et al. A 2-oxoglutarate-dependent dioxygenase from Ruta graveolens L. Exhibits p-coumaroyl CoA 2-hydroxylase activity (C2H): A missing step in the synthesis of umbelliferone in plants. Plant J. 2012;70(3):460–70.
50. Larbat R, Kellner S, Specker S, Hehn A, Gontier E, Hans J, et al. Molecular cloning and functional characterization of psoralen synthase, the first committed monooxygenase of furanocoumarin biosynthesis. J Biol Chem. 2007;282(1):542–54.
51. Larbat R, Hehn A, Hans J, Schneider S, Jugde H, Schneider B, et al. Isolation and functional characterization of CYP71AJ4 encoding for the first P450 monooxygenase of angular furanocoumarin biosynthesis. J Biol Chem. 2009;284(8):4776–85.
52. Hehmann M, Lukačin R, Ekiert H, Matern U. Furanocoumarin biosynthesis in Ammi majus L. Cloning of bergaptol O-methyltransferase. Eur J Biochem. 2004;271(5):932–40.
53. Bartsch S, Bornscheuer UT. Mutational analysis of phenylalanine ammonia lyase to improve reactions rates for various substrates. Protein Eng Des Sel. 2010;23(12):929–33.
54. Appert C, Logemann E, Hahlbrock K, Schmid J, Amrhein N. Structural and catalytic properties of the four phenylalanine ammonia-lyase isoenzymes from parsley (Petrosselinum crispum Nym.). Eur J Biochem. 1994;225(1):491–8.

55. Report S. IMMOBILIZED NATIVE AND MUTANT AMMONIA LYASES FOR THE SYNTHESIS OF ALPHA AND BETA PHENYLALANINE ANALOGUES. 2016;(263).
56. Hyun MW, Yun YH, Kim JY, Kim SH. Fungal and plant phenylalanine ammonia-lyase. Mycobiology. 2011;39(4):257–65.
57. Roots S, Pinus L, Ronald P, Soderhall K, Yl PS, Tris VES. Ammonia Lyase and Peroxidase Activity in Mycorrhizal and Short Roots of Scots. New Phytol. 2010;101(3):487–94.
58. JOOS H ???J, HAHLBROCK K. Phenylalanine ammonia???lyase in potato (Solanum tuberosum L.): Genomic complexity, structural comparison of two selected genes and modes of expression. Eur J Biochem. 1992;204(2):621–9.
59. Lim H-W, Park S-S, Lim C-J. Purification and Properties of Phenylalanine Ammonia-Lyase from Leaf Mustard. Mol Cells. 1997;7(6):715–20.
60. Sangsil P, Nualsri C, Woraathasin N, Nakkanong K. Characterization of the phenylalanine ammonia lyase gene from the rubber tree (Hevea brasiliensis Müll . Arg .) and differential response during Rigidoporus microporus infection. 2016;56(4).
61. Zhu Q, Xie X, Lin H, Sui S, Shen R, Yang Z, et al. Isolation and functional characterization of a phenylalanine ammonia-lyase gene (SsPAL1) from coleus (Solenostemon scutellarioides (L.) Codd). Molecules. 2015;20(9):16833–51.
62. Kumar A, Ellis BE. The phenylalanine ammonia-lyase gene family in raspberry. Structure, expression, and evolution. Plant Physiol. 2001;127(1):230–9.
63. Burns RG, Dick RP. Enzymes in the environment: activity, ecology, and applications. 2002.
64. Moffitt MC, Louie G V, Bowman ME, Pence J, Joseph P, Moore BS. NIH Public Access. 2008;46(4):1004–12.
65. Xiang L, Moore BS. Biochemical Characterization of a Prokaryotic Phenylalanine Ammonia Lyase Biochemical Characterization of a Prokaryotic Phenylalanine Ammonia Lyase. J Bacteriol. 2005;187(12):4286–9.
66. Xiang L, Moore BS. Inactivation, complementation, and heterologous expression of encP, a novel bacterial phenylalanine ammonia-lyase gene. J Biol Chem. 2002;277(36):32505–9.
67. Al Hafid N, Christodoulou J. Phenylketonuria: a review of current and future treatments. Transl Pediatr. 2015;4(4):304–17. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26835392%5Cnhttp://www.pubmedcentral.nih.g

ov/articlerender.fcgi?artid=PMC4728993

68. Strisciuglio P, Concolino D. New Strategies for the Treatment of Phenylketonuria (PKU). Metabolites. 2014;4(4):1007–17. Available from: http://www.mdpi.com/2218-1989/4/4/1007/

69. Sarkissian CN, Shao Z, Blain F, Peevers R, Su H, Heft R, et al. A different approach to treatment of phenylketonuria: phenylalanine degradation with recombinant phenylalanine ammonia lyase. Proc Natl Acad Sci U S A. 1999;96(5):2339–44.

70. Kong J-Q. Phenylalanine ammonia-lyase, a key component used for phenylpropanoids production by metabolic engineering. RSC Adv. 2015;5(77):62587–603. Available from: http://xlink.rsc.org/?DOI=C5RA08196C

71. Das SK, Menezes ME, Bhatia S, Wang X-Y, Emdad L, Sarkar D, et al. Gene Therapies for Cancer: Strategies, Challenges and Successes. J Cell Physiol. 2015;230(2):259–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25196387%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4363073

72. Lv M, Kong H, Liu H, Lu Y, Zhang C, Liu J, et al. Induction of phenylalanine ammonia-lyase (PAL) in insect damaged and neighboring undamaged cotton and maize seedlings. Int J Pest Manag. 2016;0874(April):1–6. Available from: https://www.tandfonline.com/doi/full/10.1080/09670874.2016.1255804

73. Sánchez-Murcia PA, Bueren-Calabuig JA, Camacho-Artacho M, Cortés-Cabrera Á, Gago F. Stepwise Simulation of 3,5-Dihydro-5-methylidene-4 *H*- imidazol-4-one (MIO) Biogenesis in Histidine Ammonia-lyase. Biochemistry [Internet]. 2016;55(41):5854–64. Available from: http://pubs.acs.org/doi/abs/10.1021/acs.biochem.6b00744

74. Jhin C, Hwang KT. Identification of lettuce phenylalanine ammonia-lyase inhibitors based on in silico virtual screening followed by in vitro evaluation. Food Sci Biotechnol. 2015;24(5):1617–21.

75. Zhang F, Huang N, Zhou L, Cui W, Liu Z, Zhu L, et al. Modulating the pH Activity Profiles of Phenylalanine Ammonia Lyase from Anabaena variabilis by Modification of Center-Near Surface Residues. Appl Biochem Biotechnol [Internet]. 2017; Available from: http://link.springer.com/10.1007/s12010-017-2458-8

76. Hamberger B, Hahlbrock K. The 4-coumarate:CoA ligase gene family in Arabidopsis thaliana comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proc Natl Acad Sci U S A . 2004;101(7):2209–14. Available from:

http://www.pnas.org/content/101/7/2209.short

77. Li Y, Kim JI, Pysh L, Chapple C. Four isoforms of Arabidopsis thaliana 4-coumarate: CoA ligase (4CL) have overlapping yet distinct roles in phenylpropanoid metabolism. Plant Physiol [Internet]. 2015;169(December):pp.00838.2015. Available from: http://www.plantphysiol.org/lookup/doi/10.1104/pp.15.00838
78. Ehlting J, Büttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E. Three 4-coumarate:coenzyme A ligases in arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms. Plant J. 1999;19(1):9–20.
79. Petersen M, Hans J, Matern U. Biosynthesis of Phenylpropanoids and Related Compounds. Vol. 40, Biochemistry of Plant Secondary Metabolism: Second Edition. 2010. 182-257 p.
80. Jiang H, Wood K V, Morgan J a. Metabolic Engineering of the Phenylpropanoid Pathway in Saccharomyces cerevisiae Metabolic Engineering of the Phenylpropanoid Pathway in Saccharomyces cerevisiae. Appl Environ Microbiol. 2005;71(6):2962–9.
81. Faraji M, Fonseca LL, Escamilla-Treviño L, Dixon RA, Voit EO. Computational inference of the structure and regulation of the lignin pathway in Panicum virgatum. Biotechnol Biofuels. 2015;8:151. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26388938%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgiartid=PMC4574612
82. Gulick AM, Lu X, Dunaway-Mariano D. Crystal structure of 4-chlorobenzoate:CoA ligase/synthetase in the unliganded and aryl substrate-bound statest. Biochemistry. 2004;43(27):8670–9.
83. Gupta OP, Nigam D, Dahuja A, Kumar S, Vinutha T, Sachdev A, et al. Regulation of Isoflavone Biosynthesis by miRNAs in Two Contrasting Soybean Genotypes at Different Seed Developmental Stages. Front Plant Sci. 2017;8(April). Available from: http://journal.frontiersin.org/article/10.3389/fpls.2017.00567/full
84. Hu Y, Gai Y, Yin L, Wang X, Feng C, Feng L, et al. Crystal structures of a Populus tomentosa 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. Plant Cell. 2010;22(9):3093–104.
85. Gómez-Robledo H-B, Cruz-Sosa F, Bernabé-Antonio A, Guerrero-Analco A, Olivares-Romero JL, Alonso-Sánchez A, et al. Identification of candidate genes related to calanolide biosynthesis by transcriptome sequencing of Calophyllum brasiliense (Calophyllaceae). BMC Plant Biol. 2016;16(1):177. Available from: http://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-016-0862-9

86. Płaŝek A, Dubert F. Physiological Role of Phenolics. 2008;23–44.
87. Al-Snafi AE. Chemical constituents and pharmacological activities of Ammi majus and Ammi visnaga. A review. Int J Pharm Ind. 2013;3(3):257–65.
88. Bartnik M, Mazurek AK. Isolation of Methoxyfuranocoumarins From *Ammi majus* by Centrifugal Partition Chromatography. J Chromatogr Sci. 2015;54(1):bmv098. Available from: http://chromsci.oxfordjournals.org/lookup/doi/10.1093/chromsci/bmv098
89. Sbai H, Saad I, Ghezal N, Greca M Della, Haouala R. Bioactive compounds isolated from Petroselinum crispum L. leaves using bioguided fractionation. Ind Crops Prod. 2016;89:207–14. Available from: http://dx.doi.org/10.1016/j.indcrop.2016.05.020
90. Munakata R, Olry A, Karamat F, Courdavault V, Sugiyama A, Date Y, et al. Molecular evolution of parsnip (Pastinaca sativa) membrane-bound prenyltransferases for linear and/or angular furanocoumarin biosynthesis. New Phytol. 2016;211(1):332–44.
91. Nebert DW, Wikvall K, Miller WL. Human cytochromes P450 in health and disease. Philos Trans R Soc B Biol Sci. 2013;368(1612):20120431/1-20120431/21.
92. Kumar S, Jin M, Weemhoff J. Cytochrome P450-Mediated Phytoremediation using Transgenic Plants: A Need for Engineered Cytochrome P450 Enzymes. J Pet Environ Eng. 2012;29(6):997–1003.
93. Dueholm B, Krieger C, Drew D, Olry A, Kamo T, Taboureau O, et al. Evolution of substrate recognition sites (SRSs) in cytochromes P450 from Apiaceae exemplified by the CYP71AJ subfamily. BMC Evol Biol [Internet]. 2015;15:122. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4482195&tool=pmcentrez&rendertype=abstract
94. Ishikawa A, Kuma T, Sasaki H, Sasaki N, Ozeki Y, Kobayashi N, et al. Constitutive expression of bergaptol O-methyltransferase in Glehnia littoralis cell cultures. Plant Cell Rep. 2009;28(2):257–65.
95. Jahnen W, Hahlbrock K. Cellular localization of nonhost resistance reactions of parsley (Petroselinum crispum) to fungal infection. Planta. 1988;173(2):197–204.
96. Terato M, Ishikawa A, Yamada K, Ozeki Y, Kitamura Y. Increased furanocoumarin production by Glehnia littoralis roots induced via Agrobacterium rhizogenes infection. Plant Biotechnol. 2011;28(3):317–21.
97. Altenberg B, Greulich KO. Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. Genomics. 2004;84(6):1014–20.
98. White D. Central Metabolic Pathways. Physiol Biochem Procaryotes. 2006;180–211.

99. Martínez Cuesta S, Rahman SA, Thornton JM. Exploring the chemistry and evolution of the isomerases. Proc Natl Acad Sci. 2016;113(7):201509494.
100. Hoh C. Enzyme Classification. 2000. p. 31–56.
101. Committee N. The Enzyme List Class 5 — Isomerases. Enzyme. 2010;1–58.
102. Word K. Purification and Characterization. 1993;33(5):0–4.
103. He F, Pan Y. Purification and characterization of chalcone isomerase from fresh-cut Chinese water-chestnut. LWT - Food Sci Technol. 2017;79:402–9. Available from: http://dx.doi.org/10.1016/j.lwt.2017.01.034
104. Jez JM, Noel JP. Reaction mechanism of chalcone isomerase: pH dependence, diffusion control, and product binding differences. J Biol Chem. 2002;277(2):1361–9.
105. Dastmalchi M, Dhaubhadel S. Soybean chalcone isomerase: evolution of the fold, and the differential expression and localization of the gene family. Planta. 2015;241(2):507–23.
106. Pelletier MK, Winkel-Shirley B. Analysis of flavanone 3-hydroxylase in Arabidopsis seedlings. Plant Physiol. 1996;111(1):339–345.
107. Mehdy MC, Lamb CJ. Chalcone isomerase cDNA cloning and mRNA induction by fungal elicitor, wounding and infection. EMBO J. 1987;6(6):1527–33.
108. Kang JH, McRoberts J, Shi F, Moreno JE, Jones AD, Howe GA. The flavonoid biosynthetic enzyme chalcone isomerase modulates terpenoid production in glandular trichomes of tomato. Plant Physiol. 2014;164(3):1161–74.
109. Hur S, Bruice TC. Enzymes do what is expected (chalcone isomerase versus chorismate mutase). J Am Chem Soc. 2003;125(6):1472–3.
110. Gensheimer M, Mushegian A. Chalcone isomerase family and fold: no longer unique to plants. Protein Sci. 2004;13(2):540–4.
111. Lim W, Li J. Co-expression of onion chalcone isomerase in Del/Ros1-expressing tomato enhances anthocyanin and flavonol production. Plant Cell, Tissue Organ Cult. 2017;128(1):113–24.
112. Jiang W, Yin Q, Wu R, Zheng G, Liu J, Dixon RA, et al. Role of a chalcone isomerase-like protein in flavonoid biosynthesis in Arabidopsis thaliana. J Exp Bot. 2015;66(22):7165–79.
113. Gall M, Thomsen M, Peters C, Pavlidis I V., Jonczyk P, Grunert PP, et al. Enzymatic conversion of flavonoids using bacterial chalcone isomerase and enoate reductase. Angew Chemie - Int Ed. 2014;53(5):1439–42.
114. Shoeva OY, Khlestkina EK, Berges H, Salina EA. The homoeologous genes encoding

chalcone-flavanone isomerase in Triticum aestivum L.: Structural characterization and expression in different parts of wheat plant. Gene. 2014;538(2):334–41.

115. Zhou L, Wang Y, Ren L, Shi Q, Zheng B, Miao K, et al. Overexpression of Ps-CHI1, a homologue of the chalcone isomerase gene from tree peony (Paeonia suffruticosa), reduces the intensity of flower pigmentation in transgenic tobacco. Plant Cell Tissue Organ Cult. 2014;116(3):285–95.
116. Ngaki MN, Louie G V., Philippe RN, Manning G, Pojer F, Bowman ME, et al. Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. Nature. 2012;485(7399):530–3. Available from: http://dx.doi.org/10.1038/nature11009
117. Lewars E. Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics. 2003;471.
118. Jensen F. Introduction to Computational Chemistry. Angewandte Chemie International Edition. 2007. 624 p.
119. Dykstra CE, Frenking G, Kim KS, Scuseria GE. Theory and Applications of Computational Chemistry: The First Forty Years. 2011.
120. Computational Chemistry_ Introduction to the Theory and Applications of .
121. Prize CN. Impact of Theoretical Chemistry on Chemical and Biological Sciences. 2014;(April):347–67.
122. Shoichet BK. Virtual screening of chemical libraries. Nature. 2004;432(7019):862–5.
123. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational Methods in Drug Discovery. Pharmacol Rev. 2013;66(1):334–95.
124. Jörg Hüser. High throughput-screening in drug discovery. 2006. 343 p.
125. Ganesan A, Coote ML, Barakat K. Molecular dynamics-driven drug discovery: Leaping forward with confidence. Drug Discov Today. 2016;22(2):249–69.
126. Liao C, Sitzmann M, Pugliese A, Nicklaus MC. Software and resources for computational medicinal chemistry. Future Med Chem. 2011;3(8):1057–85.
127. Dastmalchi S. Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery. Vol. i. 2016. p. 456.
128. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Vol. 20, Molecules. 2015. 13384-13421 p.
129. Katsila T, Spyroulias GA, Patrinos GP, Matsoukas M-T. Computational approaches in target identification and drug discovery. Comput Struct Biotechnol J. 2016;14:177–84.
130. Wadood a, Ahmed N, Shah L, Ahmad a, Hassan H, Shams S. In-silico drug design :

An approach which revolutionarised the drug discovery process. OA Drug Des Deliv. 2013;1(1):1–4.

131. Allen M. Introduction to molecular dynamics simulation. Comput Soft Matter From Synth Polym2004;23(2):1–28.
132. Mackerell AD, Bashford D, Bellott M, Evanseck JD, Fischer S, Gao J, et al. Potential Energy Function. 2009. p. 1–7.
133. Vanommeslaeghe K, Guvench O, MacKerell AD, Jr. Molecular mechanics. Curr Pharm Des. 2014;20(20):3281–92.
134. Adcock SA, McCammon JA. Molecular dynamics: Survey of methods for simulating the activity of proteins. Chem Rev. 2006;106(5):1589–615.
135. González MA. Force fields and molecular dynamics simulations. Collect SFN. 2011;12:169–200.
136. Fallis A. A molecular dynamics primer. J Chem Inf Model. 2013;53(9):1689–99.
137. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun. 1995;91(1–3):1–41.
138. Gumbart J, Aksimentiev A, Tajkhorshid E, Wang Y, Schulten K. Molecular dynamics simulations of proteins in lipid bilayers. J Comput Chem. 2007;15(4):80–90.
139. Scott WRP, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, et al. The GROMOS Biomolecular Simulation Program Package. J Phys Chem A. 1999;103:3596–607.
140. Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des. 2011;7(2):146–57.
141. Bartuzi D, Kaczor A, Targowska-Duda K, Matosiuk D. Recent Advances and Applications of Molecular Docking to G Protein-Coupled Receptors. Molecules. 2017;22(3):340.
142. de Ruyck J, Brysbaert G, Blossey R, Lensink MF. Molecular docking as a popular tool in drug design, an in silico travel. Adv Appl Bioinforma Chem. 2016;9(1):1–11.
143. Ruppert J, Welch W, Jain AN. Automatic identification and representation of protein binding sites for molecular docking. Protein Sci. 1997;6(3):524–33.
144. Levitt DG, Banaszak LJ. POCKET: A computer graphies method for identifying and displaying protein cavities and their surrounding amino acids. Vol. 10, Journal of

Molecular Graphics. 1992. p. 229–34.
145. Laskowski RA. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. Vol. 13, Journal of Molecular Graphics. 1995. p. 323–30.
146. Brady GP, Stouten PFW, Company DP, Road HC. Fast Prediction and Visualization of Protein Binding Pockets with. 2000;(Mi):1–43.
147. Mezei M. A new method for mapping macromolecular topography. Vol. 21, Journal of Molecular Graphics and Modelling. 2003. p. 463–72.
148. Map S, Site S, Us C. AutoDock AutoDock. 2007. p. 3–6.
149. AutoDock Vina - molecular docking and virtual screening program.
150. Trott O, Olson A. NIH Public Access. J Comput Chem. 2010;31(2):455–61.
151. Hossen T, Sahu K. Elucidating the H-Bonding Environment of Coumarin 102 in a Phenol–Cyclohexane Mixture by Molecular Dynamics Simulation: Implications for H-Bond-Guided Photoinduced Electron Transfer. J Phys Chem A [Internet]. 2017;121(3):616–22. Available from: http://pubs.acs.org/doi/abs/10.1021/acs.jpca.6b12134
152. Yao Y, Zhang H, Li ZS. Stereoselectivity of chalcone isomerase with chalcone derivatives: A computational study. J Mol Model. 2013;19(11):4753–61.
153. Bharatham K, Bharatham N, Park KH, Lee KW. Binding mode analyses and pharmacophore model development for sulfonamide chalcone derivatives, a new class of ??-glucosidase inhibitors. J Mol Graph Model. 2008;26(8):1202–12.
154. Shankaraiah N, Siraj KP, Nekkanti S, Srinivasulu V, Sharma P, Senwar KR, et al. DNA-binding affinity and anticancer activity of ??-carboline-chalcone conjugates as potential DNA intercalators: Molecular modelling and synthesis. Bioorg Chem. 2015;59:130–9.
155. Yeggoni DP, Gokara M, Manidhar DM, Rachamallu A, Nakka S, Reddy CS, et al. Binding and Molecular Dynamics Studies of 7 - Hydroxycoumarin Derivatives with Human Serum Albumin and Its Pharmacological Importance. 2014;
156. Wiley_ Computational Organic Chemistry, 2nd Edition - Steven M.
157. Atkins PW, Friedman R. Molecular Quantum Mechanics. Quantum [Internet]. 2011;134:588. Available from: http://www.citeulike.org/user/atreju/article/3390344
158. Griffiths DJ. Introduction to Quantum Mechanics. Physics Today. 1995. p. 409. Available from: http://pdfserv.aip.org/PHTOAD/vol_48/iss_3/94_1.pdf
159. Gordon MS. an Introduction To Quantum Chemistry.

160. Morin D. Introduction to quantum mechanics. 1920;1–20.
161. Atkin P, Paula J. Physical chemistry. W.H. Freeman. 2006. 1072 p.
162. Sherrill CD. An Introduction to Hartree-Fock Molecular Orbital Theory What Problem Are We Solving ? Motivation and the Hartree Product. Sch Chem Biochem Georg Inst Technol. 2000;1–8.
163. Young D. Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems. Vol. 9, Computers and Chemistry. 2004. p. 408.
164. Ramachandran KI, Deepa G, Namboori K. Computational chemistry and molecular modeling: Principles and applications. Computational Chemistry and Molecular Modeling: Principles and Applications. 2008.
165. Jensen F. Atomic orbital basis sets. 2013;3(June):273–95.
166. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria GE., Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennuci B., Petersson GA. et al. Gaussian 09. 2009.
167. The ORCA program system - Neese - 2011 - Wiley Interdisciplinary Reviews_ Computational Molecular Science - Wiley Online Library.
168. Samundeeswari S, Kulkarni MV, Yenagi J, Tonannavar J. Dual Fluorescence and Solvatochromic Study on 3-Acyl Coumarins. J Fluoresc. 2017; Available from: http://link.springer.com/10.1007/s10895-017-2052-z
169. SambathKumar K, Nithiyanantham S. Synthesis, characterization and theoretical properties of coumarin NLO single crystal by DFT method. J Mater Sci Mater Electron [Internet]. 2017;28(9):6529–43. Available from: http://link.springer.com/10.1007/s10854-017-6342-7
170. Lonsdale R, Harvey JN, Mulholland AJ. A practical guide to modelling enzyme-catalysed reactions. Chem Soc Rev. 2012;41(8):3025–38.
171. Acevedo O. Simulating chemical reactions in ionic liquids using QM/MM methodology. J Phys Chem A. 2014;118(50):11653–66.
172. Ghysels A, Woodcock HL, Larkin JD, Miller BT, Shao Y, Kong J, et al. Efficient calculation of QM/MM frequencies with the mobile Block Hessian. J Chem Theory Comput. 2011;7(2):496–514.
173. Ghosh D. Multiscale modelling : hybrid quantum mechanics / molecular mechanics as an example and some recent developments. 2017;112(7).
174. Reiher M. Atomistic Approaches in Modern Biology From Quantum Chemistry to Molecular Simulations (Topics in Current Chemistry). Quantum. 2007. p. 373.
175. Lin H, Zhang Y, Pezeshki S, Wang B, Truhlar DG. Manual. 2015;2.

176. Lai R. Development and Application of Combined Quantum Mechanical and Molecular Mechanical Methods. 2014;
177. Ruiz-Pernía JJ, Silla E, Tuñón I. Comparative computational analysis of different active site conformations and substrates in a chalcone isomerase catalyzed reaction. J Phys Chem B. 2006;110(41):20686–92.
178. Ruiz-perni JJ, Moliner V, Hynes JT, Roca M, Ruiz-pernı JJ. Dynamic Effects on Reaction Rates in a Michael Addition Catalyzed by Chalcone Isomerase . Beyond the Frozen Environment Approach Dynamic Effects on Reaction Rates in a Michael Addition Catalyzed by Chalcone Isomerase . Beyond the Frozen Environment Approa. 2008;7477–88.
179. Roux B, Simonson T. Implicit solvent models. Biophys Chem. 1999;78(1–2):1–20.
180. Marenich A V., Cramer CJ, Truhlar DG. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J Phys Chem B. 2009;113(18):6378–96.
181. Feig M, Brooks CL. Recent advances in the development and application of implicit solvent models in biomolecule simulations. Curr Opin Struct Biol. 2004;14(2):217–24.
182. Cramer CJ, Truhlar DG. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. Chem Rev. 1999;99(8):2161–200.
183. Tomasi J, Persico M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. Chem Rev. 1994;94(7):2027–94.
184. Gallicchio E, Levy RM. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. J Comput Chem. 2004;25(4):479–99.
185. Mennucci B, Cancès E, Tomasi J. Evaluation of Solvent Effects in Isotropic and Anisotropic Dielectrics and in Ionic Solutions with a Unified Integral Equation Method: Theoretical Bases, Computational Implementation, and Numerical Applications. J Phys Chem B. 1997;101(49):10506–17.
186. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. Med Res Rev. 2006;26(5):531–68.
187. Zheng M, Kuriappan JA, Waller MP. Toward more efficient density-based adaptive QM/MM methods. 2017;(November 2016):1–7.
188. Pezeshki S, Lin H. Adaptive-Partitioning QM/MM for Molecular Dynamics Simulations: 4. Proton Hopping in Bulk Water. J Chem Theory Comput. 2015;11(6):2398–411.

189. Nielsen SO, Bulo RE, Moore PB, Ensing B. Recent progress in adaptive multiscale molecular dynamics simulations of soft matter. Phys Chem Chem Phys. 2010;12(39):12401–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20734007
190. Bulo RE, Michel C, Fleurat-Lessard P, Sautet P. Multiscale Modeling of Chemistry in Water: Are We There Yet? J Chem Theory Comput. 2013;9(12):5567–77.
191. Iannuzzi M, Schiffmann F, Vandevondele J. CP 2 K : atomistic simulations of condensed matter systems. 2014;4(February):15–25.
192. Pezeshki S, Davis C, Heyden A, Lin H. Adaptive-partitioning QM/MM dynamics simulations: 3. Solvent molecules entering and leaving protein binding sites. J Chem Theory Comput. 2014;10(11):4765–76.
193. Bodill T, Conibear AC, Mutorwa MKM, Goble JL, Blatch GL, Lobb KA, et al. Exploring DOXP-reductoisomerase binding limits using phosphonated N-aryl and N-heteroarylcarboxamides as DXR inhibitors. Bioorganic Med Chem. 2013;21(14):4332–41.
194. Conibear AC, Lobb KA, Kaye PT. 31P NMR kinetic study of the tandem cleavage of phosphonate esters by bromotrimethylsilane. Tetrahedron. 2010;66(43):8446–9.
195. Bodill T, Conibear AC, Blatch GL, Lobb KA, Kaye PT. Synthesis and evaluation of phosphonated N-heteroarylcarboxamides as DOXP-reductoisomerase (DXR) inhibitors. Bioorganic Med Chem. 2011;19(3):1321–7.
196. Head-Gordon M. Quantum Chemistry and Molecular Processes. J Phys Chem. 1996;100(95):13213–25.
197. Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys. 2010;132(15).
198. Harrison NM. An Introduction to Density Functional Theory. Technology. 1995;2(1):1–26. Available from: http://link.aip.org/link/AJPIAS/v68/i1/p69/s1&Agg=doi
199. Dennington R.D., Keith T.A., Millan J., Neilsen A.B., Holder A.J. HJ. GaussView. Pittsburgh: Pittsburgh; 2007.
200. RCSB PDB Homepage. https://www.rcsb.org/
201. BIOVIA DS. BIOVIA Workbook. San Diego: Dassault Systèmes; 2016.
202. Trott O, Olson AJ. Software News and Update AutoDock Vina : Improving the Speed and Accuracy of Docking with a New Scoring Function , Efficient Optimization , and

Multithreading. 2009;
203. Acros. https://www.**acros**.com/
204. Brooks BR, Iii CLB, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM : The Biomolecular Simulation Program. 2009;
205. Hess B, Uppsala S-, Lindahl E. GROMACS 4 : Algorithms for Highly Efficient , Load-Balanced , and Scalable Molecular Simulation. 2008;435–47.
206. Berendsen HJC, Spoel D Van Der, Drunen R Van. GROMACS : A message-passing parallel molecular dynamics implementation. 1995;91:43–56.
207. Vanommeslaeghe, K., and Mackerell, A. D. (2012). Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing, (1).
208. Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2018). H++ : a server for estimating p K a s and adding missing hydrogens to macromolecules, *33*(May), 368–371.
209. Humphrey W, Dalke A, Schulten K. VMD : Visual Molecular Dynamics. 1996;7855(December 1995):33–8.
210. Pronk S, Apostolov R, Shirts MR, Smith JC, Kasson PM, Spoel D Van Der, et al. Structural bioinformatics molecular simulation toolkit ′ rd Pa. 2018;29(7):845–54.
211. Sousa AW, Vranken WF. Open Access ACPYPE - AnteChamber PYthon Parser interfacE. 2012;1–8.
212. Karelson M, Lobanov VS, Katritzky AR. Quantum-Chemical Descriptors in QSAR / QSPR Studies. 1996
213. Klein ML. Conformations of Flexible Molecules in Fluid Phases. 1992;88(13):1701–5.
214. Haven N. Commentationes The C-Potential Surface for Predicting C o n f o r m a t i o n s o f M o l e c u l e s in Solution. 1974;284:279–84.
215. Perczel A, Angyin JG, Kaj M, Viviani W, Rivail J, Marcoccia J, et al. Peptide Models . 1 . Topology of Selected Peptide Conformational Potential Energy Surfaces (Glycine and Alanine Derivatives). 1991;6256–65.
216. Goldberg DE. A Note on Boltzmann Tournament Selection for Genetic Algorithms and Population-Oriented Simulated Annealing. 1990;4:445–60.
217. Zhou HX, Szabo A. Theory and simulation of the time-dependent rate coefficients of diffusion-influenced reactions. Biophys J. 1996;71(5):2440–57.
218. Smith GD, Jaffe RL. Quantum Chemistry Study of Conformational Energies and Rotational Energy Barriers in n -Alkanes. 1996;3654(96):18718–24.
219. Bernardi RC, Melo MCR, Schulten K. Biochimica et Biophysica Acta Enhanced

sampling techniques in molecular dynamics simulations of biological systems. BBA - Gen Subj. 2015;1850(5):872–7.

220. Frisch (Gaussian, Inc.), Michael J. Trucks (Gaussian, Inc.), G. W. Schlegel (Gaussian, Inc.), H. Bernhard (Gaussian I. Gaussian 09 Citation. 2009.
221. Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Montgomery 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. PJA. Gaussian. Pittsburgh, PA: Gaussian, Inc; 2009.
222. Liptak MD, Shields GC, York N. Accurate p K a Calculations for Carboxylic Acids Using Complete Basis Set and Gaussian-n Models Combined with CPCM Continuum Solvation Methods. 2001;(16):7314–9.
223. Baldridge KK, Battersby TR, Vernonclark R, Siegel JS. Does π - σ - π Through-Bond Coupling Significantly Increase C - C Bond Lengths ? 1997;7863(96):7048–54.
224. Chaitanya KVSS. Lauglin wave function, Berry Phase and Quantization.
225. Gómez-jeria JS. A New Set of Local Reactivity Indices within the Hartree- Fock-Roothaan and Density Functional Theory Frameworks. 2013;1(1):25–55.
226. Amsom M, Society A, St R, Lipscomb II. The Protein D a t a Bank : A Computer-based Archival File for Macromolecular Structures The Protein Data Bank is a computer-based archival file for macromolecular (a) Scope. 1977;535–42.
227. Tool AI. 24.5. The Protein Data Bank, 1999–. 2006;675–6.
228. Willett P, Glen C. Molecular Recognition of Receptor Sites using a Genetic Algorithm with a Description of Desolvation. 1995;43–53.
229. Mckie VA, Gilbert HJ, Charnock SJ, Bolam DN, Nurizzo D, Davies GJ. Promiscuity in ligand-binding : The three-dimensional structure of a Piromyces carbohydrate-binding module , CBM29-2 , in complex with cello- and mannohexaose. 2002;99(22):14077–82.
230. Simpson PJ, Xie H, Bolam DN, Gilbert HJ, Williamson MP. The Structural Basis for the Ligand Specificity of Family 2 Carbohydrate-binding Modules. 2000;275(52):41137–42.
231. Online VA, Peng W, Ding F, Peng Y, Sun Y. Molecular BioSystems Molecular recognition of malachite green by hemoglobin and their specific interactions : insights from in silico docking and molecular spectroscopy. 2014;138–48.
232. Koide A, Abbatiello S, Rothgery L, Koide S. Probing protein conformational changes in living cells by using designer binding proteins : Application to the estrogen receptor.

2002;99(3):1253–8.
233. Press D. Molecular dynamics simulations : advances and applications. 2015;37–47.
234. Board B, Castillo JG, Romero J, Fries RA, Macy GA, Nerenberg PS. Posters : Molecular Dynamics I. Biophysj. 2018;114(3):337a.
235. Lim C. How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations. 2017;(1).
236. Mutt E, Sowdhamini R. Molecular Dynamics Simulations and Structural Analysis to Decipher Functional Impact of a Twenty Residue Insert in the Ternary Complex of Mus musculus TdT Isoform. 2016;(D):1–24.
237. Mu K, Hobza P. Noncovalent Interactions : A Challenge for Experiment and Theory. 2000;
238. Reed AE, Curtiss LA, Weinhold F. Intermolecular Interactions from a Natural Bond Orbital , Donor-Acceptor Viewpoint. 1988;899–926.
239. Paton RS, Goodman JM. Hydrogen Bonding and π -Stacking : How Reliable are Force Fields ? A Critical Evaluation of Force Field Descriptions of Nonbonded Interactions. 2009;944–55.
240. Lobanov MY, Bogatyreva NS, Galzitskaya O V. STRUCTURAL − FUNCTIONAL ANALYSIS OF BIOPOLYMERS AND THEIR COMPLEXES Radius of Gyration as an Indicator of Protein Structure Compactness. 2008;42(4):701–6.
241. Cele FN, Kumalo H, Soliman MES. Mechanism of Inhibition of Hsp90 Dimerization by Gyrase B Inhibitor Coumermycin A1 (CA1) Revealed by Molecular Dynamics Simulations and Thermodynamic Calculations. Cell Biochem Biophys. 2016;74(3):353–63.
242. Richmond T. Solvent Accessible Surface Area and Excluded Volume in Proteins. 1984;i:63–89.
243. Marsh JA. Buried and Accessible Surface Area Control Intrinsic Protein Flexibility. J Mol Biol. 2013;425(17):3250–63.
244. Access O. Quantum Mechanical Modeling: A Tool for the Understanding of Enzyme Reactions. 2013;662–702.
245. Morokuma K. ONIOM and Its Applications to Material Chemistry and Catalyses. 2003;24(6).
246. V EU. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. 2006;815–26.
247. Jensen KH, Sigman MS, Sigman MS. Mechanistic approaches to palladium-catalyzed

alkene difunctionalization reactions. 2008;4083–8.
248. Ball P. Water as an Active Constituent in Cell Biology. 2008;4–6.
249. Jez JM, Bowman ME, Noel JP. Role of hydrogen bonds in the reaction mechanism of chalcone isomerase. Biochemistry. 2002;41(16):5168–76.