Caractérisation structurale et biochimique d'enzymes impliquées dans la biosynthèse des acides chlorogéniques

Laura Amandine Lallemand

UNIVERSITÉ DE GRENOBLE

Thesis

Submitted to obtain the

PhD DEGREE OF GRENOBLE UNIVERSITY

Speciality : **Structural Biology and Nanobiology**

Ministerial decree: 7th August 2006

Presented by

Laura Amandine LALLEMAND

Thesis directed by **Sean Mc SWEENEY** and

co-directed by **Andrew Mc CARTHY**

prepared in the **Structural Biology Group of ESRF**

(European Synchrotron Radiation Facility)

in the **Doctoral School of Chemistry and Life Sciences**

Structural and biochemical characterisation of enzymes involved in chlorogenic acid biosynthesis

Nicotiana tabacum 4-coumarate CoA ligase (4CL)

Coffea canephora hydroxycinnamoyl-CoA quinate/ shikimate hydroxycinnamoyltranferases (HCT, HQT)

Dissertation defended on **March, 21st 2011**,

in front of a jury composed of :

Dr. Nushin AGHAJARI
Reviewer
Dr. Anne IMBERTY
President
Pr. Catherine LAPIERRE
Reviewer
Dr. Andrew Mc CARTHY
Thesis director
Dr. James Mc CARTHY
Member
Dr. Sean Mc SWEENEY
Thesis director

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biologie structurale et nanobiologie**

Arrêté ministériel : 7 août 2006

Présentée par

Laura Amandine LALLEMAND

Thèse dirigée par **Sean McSWEENEY** et

co-dirigée par **Andrew McCARTHY**

préparée au sein du **Groupe de Biologie Structurale de l'ESRF (Installation Européenne de Rayonnement Synchrotron)**

dans **l'École Doctorale Chimie et Sciences du Vivant**

Caractérisation structurale et biochimique d'enzymes impliquées dans la biosynthèse des acides chlorogéniques

Nicotiana tabacum 4-coumarate CoA ligase (4CL)

Coffea canephora hydroxycinnamoyl-CoA quinate/shikimate hydroxycinnamoyltranférases (HCT, HQT)

Thèse soutenue publiquement le **21 Mars 2011**,

devant le jury composé de :

Mme, Nushin, AGHAJARI
Rapporteur
Mme, Anne, IMBERTY
Président du Jury
Mme, Catherine, LAPIERRE
Rapporteur
M., Andrew, McCARTHY
Directeur de thèse
M., James, McCARTHY
Membre
M., Sean, McSWEENEY
Directeur de thèse

Acknowledgments

This project was supported by a doctoral fellowship co-financed by Nestlé Research and the European Synchrotron Research Facility (ESRF). Most of the work was carried out in the Partnership for Structural Biology building and at the synchrotron facility in Grenoble.

I am grateful to my supervisors Sean McSweeney, for accepting me as a PhD student in his group, and Andrew McCarthy, for his scientific support and guidance throughout this project. I express all my gratitude to Gordon Leonard for his valuable comments on this manuscript. I thank all the members of the ESRF Structural Biology Group and especially Virginie Carbonell, who contributed to the structural studies as a Master student.

I also thank the European Molecular Biology Laboratory (EMBL)/ ESRF Joint Structural Biology Group for access and support at the synchrotron beamlines. I thank the High Throughput Crystallisation Laboratory (HTX) of the EMBL-Grenoble Outstation and the Quality Control platform of the Institut de Biologie Structurale for their technical assistance. I thank the EMBL-Heidelberg Protein Production and Purification core facility for their help concerning the pET-28M-SUMO3 vector.

I gratefully acknowledge James McCarthy for his collaboration and Emilie Devienne, who contributed to the biochemical studies as an undergraduate student. I thank all the Gene Discovery and the Plant Chemistry and Functionality groups of Nestlé R&D centre in Tours, and especially Maud Lepelley, Victoria Caillet and Stéphane Michaux for their help and kindness during my visits. Finally I thank Karin Kraehenbuehl for carrying out the MS analysis and for providing the 5-FQA sample.

Contents

Abstract

Chlorogenic acids (CGAs) are esters formed between cinnamic acid derivatives and quinic acid. These secondary metabolites play a major role in the plant's response to various biotic and abiotic stresses, and are important intermediates for lignin biosynthesis in higher plants. The green coffee bean is particularly rich in CGAs and one of the most important dietary sources of this group of antioxidants. Hydroxycinnamoyl-CoA thioesters are the precursors of CGAs and other phenylpropanoid compounds. These activated intermediates are synthesised from the corresponding hydroxycinnamic acids and coenzyme A (CoA) by 4-coumarate CoA ligase (4CL, EC 6.2.1.12), which belongs to the adenylate-forming enzyme superfamily. *Nicotiana tabacum* 4CL isoform 2 (*Nt*4CL2) was produced for the enzymatic synthesis of hydroxycinnamoyl-CoA thioesters not available commercially. The crystal structure of *Nt*4CL2 was solved by molecular replacement (MR) in two different conformational states, one representative of the apo-enzyme, the other in a ternary complex with CoA and adenosine monophosphate. Structural and docking analyses enabled the identification of several residues important for catalysis and substrate recognition in 4CLs. Two *Coffea canephora* hydroxycinnamoyl-CoA shikimate/ quinate hydroxycinnamoyl-transferases (*Cc*HCT/ *Cc*HQT, EC 2.3.1.99) were over-expressed in *E. coli* and purified to homogeneity. Both belong to the clade V of the acyl-CoA-dependent acyltransferase (BAHD) superfamily. X-ray diffraction analysis of native *Cc*HCT crystals to 3.0 Å resolution resulted in a structure solution by MR. Crystals of a trypsin-proteolysis resistant mutant (K210A/K217A) HCT diffracted to 1.7 Å resolution. *Cc*HCT crystal structures showed some major differences in the active site, mainly concerning the conformation of the Val31-Pro37 loop and the catalytic His153 side-chain. Docking and site-directed mutagenesis were carried out to identify potential residues involved in catalysis and substrate binding. *Cc*HCT was used as a template for homology-modelling of *Cc*HQT, with which it shares 64 % identity at the amino acid level. The differences in the amino acid residues lining the two active sites provide some clues to their alternative substrate specificities. High-performance liquid chromatography was used to analyse the *in vitro* activity of the recombinant proteins. *Cc*HCT is capable of synthesizing the most abundant coffee CGA compound, 5-*O*-caffeoylquinic acid, but also the diester 3,5-*O*-dicaffeoylquinic acid, which is a major component of the immature coffee grain. The combined approach of structural biology, mutagenesis and enzymology provides new molecular insights into the role of these key enzymes in CGA biosynthesis.

Résumé

Les acides chlorogéniques sont des esters formés d'un ou plusieurs dérivés de l'acide *t*-cinnamique conjugués à l'acide quinique. Ces métabolites secondaires sont impliqués dans les mécanismes de défense contre certains stress biotiques et abiotiques, et représentent des intermédiaires majeurs dans la biosynthèse des lignines chez les végétaux supérieurs. Le grain de café est exceptionnellement riche en acides chlorogéniques et représente une source majeure de cette classe d'antioxydants. Dans la voie des phénylpropanoides, les précurseurs des acides chlorogéniques et d'autres composés phénoliques sont des thioesters de coenzyme A (CoA). Ces intermédiaires activés sont synthétisés à partir d'un acide hydroxycinnamique (HCA) et de CoA par la 4-coumarate CoA ligase (4CL, EC 6.2.1.12) qui appartient à la superfamille des enzymes formant des adénylates. L'isoforme 2 de 4CL isolée de *Nicotiana tabacum* (*Nt*4CL2) a été produite chez *E. coli* et purifiée pour la synthèse de thioesters de CoA qui ne sont pas disponibles commercialement. La structure cristalline de *Nt*4CL2 a été résolue par remplacement moléculaire sous forme native et en complexe avec CoA et l'adénosine monophosphate. L'analyse structurale et la modélisation par docking ont permis d'identifier des résidus potentiellement importants pour la catalyse et la liaison au substrat. Deux hydroxycinnamoyl-CoA shikimate/ quinate hydroxycinnamoyltransférases de *Coffea canephora* (*Cc*HCT/ *Cc*HQT, EC 2.3.1.99) ont été clonées, sur-exprimées dans *E. coli* et purifiées jusqu'à homogénéité. *Cc*HCT et *Cc*HQT appartiennent au clade V de la superfamille des acyltransférases acyl-CoA-dépendantes, appelée BAHD. La structure tridimensionnelle de *Cc*HCT a été résolue par analyse au rayonnement synchrotron de cristaux à une résolution maximale de 3.0 Å. Un mutant (K210A/K217A) de *Cc*HCT, résistant à la trypsine, a permis d'obtenir des cristaux diffractant à une résolution plus élevée (1.7 Å). Ces structures diffèrent dans la région du site actif principalement au niveau de la conformation de boucle Val31-Pro37 et du résidu catalytique His153. La modélisation par docking et la mutagénèse dirigée ont permis d'identifier des résidus potentiellement impliqués dans la catalyse de la réaction et l'interaction enzyme-substrat. Un modèle a été proposé pour *Cc*HQT par homologie de séquence avec *Cc*HCT, les deux protéines partageant 64 % d'identité. Les différences concernant les résidus présents au sein des deux sites actifs donnent des indices sur les bases moléculaires déterminant la spécificité de substrats au sein de cette famille d'enzymes. La chromatographie liquide a été utilisée pour analyser l'activité enzymatique. *Cc*HCT est capable de synthétiser l'acide chlorogénique le plus abondant, l'acide 5-*O*-caféoylquinique, mais aussi le diester 3,5-*O*-dicaféoylquinique qui est un composant majeur du grain de café avant maturation. L'approche combinée de la biologie structurale, de la mutagénèse et de l'enzymologie permet de mieux comprendre le rôle de ces enzymes impliquées dans la biosynthèse des acides chlorogéniques.

Abbreviations

4CL	4-coumarate CoA ligase
ACS	acetyl-CoA synthase
AMP	adenosine monophosphate
At	*Arabidopsis thaliana*
ATP	adenosine triphosphate
AU	asymmetric unit
BME	β-mercaptoethanol
bp	base pair
C3H	coumarate 3-hydroxylase
C4H	cinnamate 4-hydroxylase
CAD	cinnamyl alcohol reductase
CAT	chloramphenicol acetyltransferase
CBL	chlorobenzoate CoA-ligase
Cc	*Coffea canephora*
CCR	cinnamoyl-CoA reductase
cDNA	complementary deoxyribonucleic acid
CGA	chlorogenic acid
CoA	coenzyme A
CQA	caffeoylquinic acid
CrAT	carnitine acetyltransferase
CSA	caffeoylshikimic acid
CV	column volume
DAHP	3-dehydroxy-D-arabino-heptulose-7 phosphate
DAT	deacetylvindoline 4-*O*-acetyltransferase
DLS	dynamic light scattering
dNTP	deoxynucleotide triphosphate
DON	deoxynivalenol
DTT	dithiothreitol
DW	dry weight
EPSP	5-enolpyruvate shikimate 3-phosphate
EST	expressed sequenced tag
FPLC	fast protein liquid chromatography
FQA	feruloylquinic acid
GST	glutathione S-transferase
HCA	hydroxycinnamic acid
HCBT	hydroxycinnamoyl/benzoyl-CoA anthranilate N-hydroxycinnamoyl/benzoyltransferase
HCT	hydroxycinnamoyl-CoA shikimate/ quinate hydroxycinnamoyltransferase
HPLC	high-performance liquid chromatography
HQT	hydroxycinnamoyl-CoA quinate/shikimate hydroxycinnamoyltransferase
HTX	high-throughput crystallography
IPTG	isopropyl β-D-1-thiogalactopyranoside
K-mutant	K210A/K217A mutant (*Cc*HCT)
kDa	kilo Dalton (10^3 Da)
LB	Luria-Bertani
MaT	malonyltransferase
MES	2-(N-morpholine)-ethane sulphonic acid
MR	molecular replacement
MS	mass spectrometry
MW	molecular weight
MX	macromolecular crystallography
NRPS	non ribosomal peptide synthase
Nt	*Nicotiana tabacum*
NTA	nitriloacetic acid
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PAL	phenylalanine ammonia lyase
PapA5	polyketide-associated protein A5
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PEP	phosphoenolpyruvate
PMSF	phenylmethanesulfonyl fluoride
PPi	pyrophosphate
RMSD	root mean square deviation
RT	retention time
SBP	substrate binding pocket
SDS	sodium dodecylsulfate
SM	secondary metabolites
SUMO	small ubiquitin-related modifier
TRI101	trichothecene 3-*O*-acetyltransferase
UV	ultraviolet
VS	vinorine synthase

Chapter 1. Introduction

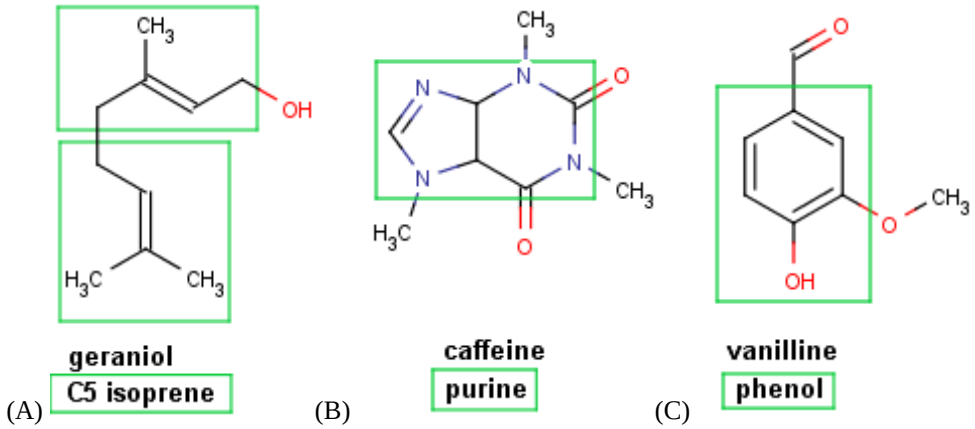
1.1. Structure and function of plant secondary metabolites

1.1.1. Generalities

A significant amount of a plant's energy and carbon resource is dedicated to secondary (also called specialised) metabolism, which results in the synthesis of a myriad of low molecular weight (MW) compounds. In contrast to primary metabolites, secondary metabolites (SM) are not universally distributed and are not directly involved in basic cellular functions such as growth, photosynthesis or reproduction. However, SM are known to mediate plant-environment interactions and therefore play a major role in the plant's survival and adaptation to environmental changes (Wink, 2003). To date, more than 200,000 different SM have been identified (Hartmann, 2007). Many of these compounds are derived from a limited number of carbon scaffolds often synthesised through specific metabolic pathways. Three major classes of SM are generally considered (Figure 1): the terpenes, built from 5-carbon isoprene units; the nitrogen-containing alkaloids; and the phenolics, which consist of a benzenoid ring bearing at least one hydroxyl substituent. These carbon skeletons can be subsequently modified by substrate and/ or regio-specific biosynthetic enzymes (e.g. hydroxylases, methyltransferases, glycosylases etc.), resulting in a high structural diversity.

Figure 1: Examples of chemical structures of the main classes of secondary metabolites

The basic units of terpenoid (A), alkaloid (B) and phenolic compounds (C) are highlighted in green.



1.1.2. Terpenes

Terpenes constitute the most abundant and structurally diverse group of SM with more than 15,500 compounds (Wink, 2003). Some plant species contain as many as 100 different terpenes as components of complex oils, resins or floral scent, such as geraniol (Figure 1A). Certain specialised groups of terpenes have well-characterised physiological roles, for example sterols as membrane components or phytohormones, and carotenoids as photosynthetic pigments and antioxidants. Volatile and non-volatile terpenes are implicated in the attraction of both pollinators and predators of herbivores, and in direct defence against microbes and insects (Pichersky *et al.*, 2002; Dudareva *et al.*, 2004). Further information on the biosynthesis and function of terpenes can be found in (Cheng, 2007; Gershenzon *et al.*, 2007).

1.1.3. Alkaloids

Over 20,000 different alkaloid compounds are known (Wink, 2003; Tholl, 2006). The skeleton of most of them is derived from amino acids, although moieties from other pathways, such as terpenoids, are often combined with these. A number of these nitrogen-containing compounds possess a heterocyclic ring. Although the precise physiological role of alkaloids is still unclear, most of their known functions are related to plant defence. These include protection against pathogenic organisms and herbivores, and regulation of plant growth. As an example, *Coffea* species are characterised by a high content of caffeine (Figure 1B) and related purine alkaloids (Noirot *et al.*, 2004). Two main hypotheses have been proposed concerning their roles (Ashihara *et al.*, 2008). The chemical defence theory argues that high concentrations of caffeine may protect young tissues from pathogens and herbivores. The allelopathic function theory proposes that caffeine in seed coats and falling leaves is released into the soil to inhibit germination of seeds around the parent plants.

1.1.4. Phenolics

Phenolic compounds (Figure 1C) account for more than 6,000 different molecules (Wink, 2003). These have a wide array of important functions in plants, including pigmentation for the attraction of insect pollinators (anthocyanins), protection against solar radiation damage (hydroxycinnamic acids and flavones), defence against pathogens (stilbenes), signalling (flavonoids), structural support and water transport (polymeric lignins) (Douglas, 1996; Iwashina, 2003; Lattanzio *et al.*, 2008). The structure, function and biosynthesis of phenylpropanoids are described in more detail in the following section.

1.2. Structural diversity of the phenolic compounds

1.2.1. Classification

Phenolic compounds can be divided into two main groups (Table 1). The flavonoids are characterised by a C_6-C_3-C_6 structure and are classified as a function of the type of heterocycle (e.g. flavonols, flavones, flavanones, flavanols, anthocyanidins and isoflavones) they contain. Non-flavonoids are varied and include simple phenols, hydroxybenzoic acids, hydroxycinnamic acids (HCAs), coumarins and tannins, among others.

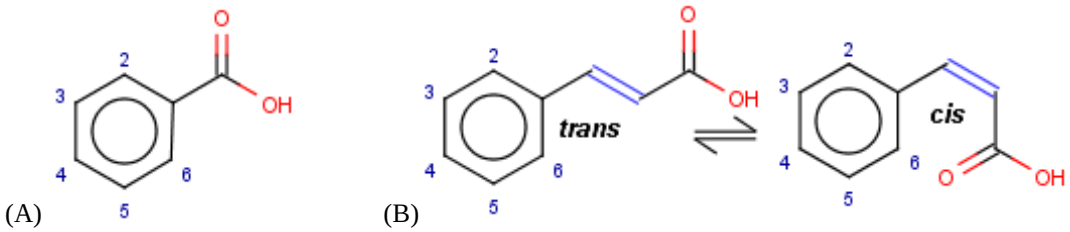
Table 1: Classification of the phenolic compounds

	Carbon skeleton	*Compound class*
Non-flavonoids	C_6	simple phenols
	C_6-C_1	hydroxybenzoic acids
	C_6-C_2	acetophenones, phenylacetates
	C_6-C_3	hydroxycinnamic acids, phenylpropenes, coumarins, chromones
	C_6-C_4	naphtoquinones
	C_6-C_1-C_6	xanthones
	C_6-C_2-C_6	stilbenes, anthraquinones
	$(C_6$-$C_3)_2$	lignans
	$(C_6)_n$	catechol melanins
	$(C_6$-$C_1)_n$:Glucose	hydrolysable tannins
	$(C_6$-$C_3)_n$	lignins
Flavonoids	C_6-C_3-C_6	flavonols
	C_6-C_3-C_6	flavones
	C_6-C_3-C_6	flavanones
	C_6-C_3-C_6	flavanols
	C_6-C_3-C_6	anthocyanidins
	C_6-C_3-C_6	isoflavones
	$(C_6$-C_3-$C_6)_2$	biflavonoids
	$(C_6$-C_3-$C_6)_n$	condensed tannins (catechin polymers, proanthocyanidins)

1.2.2. Structure of the hydroxycinnamic acids (HCAs)

HCAs are characterised by a unique C_6-C_3 chemical structure. They possess a phenol group and an additional carboxylic acid function, characteristic of the phenolic acids, which comprise the benzoic and cinnamic acid derivatives (Figure 2). Both *cis*- and *trans*-cinnamic acid derivatives (Figure 2B) are common in nature, the latter being naturally derived from the biosynthetic precursor *trans*-cinnamic acid. Isomerisation to *cis* derivatives may occur during extraction, processing or exposure to light (Kahnt, 1967). The simple benzoic and cinnamic acid derivatives differ in the degree of hydroxylation and/ or methoxylation of the aromatic ring (Figure 2). Coumaric, caffeic, ferulic and sinapic acids are amongst the most widely distributed phenolic compounds in plants. Three position isomers exist for coumaric acid: 2-, 3- and 4-coumaric acids. The latter is the most abundant in Nature and will be referred to as coumaric acid in the rest of the document. HCAs are rarely found as free acids in unprocessed plant material and are identified in plant extracts mainly after degradation of the soluble and insoluble-bound derivatives (Clifford, 2000).

Figure 2: Chemical structures of common benzoic (A) and cinnamic (B) acid derivatives



Benzoic acids (C_6-C_1)	*Aromatic substitution*	Cinnamic acids (C_6-C_3)
salicylic acid	2-OH	2-coumaric acid
4-hydroxybenzoic acid	4-OH	(4-)coumaric acid
vanillic acid	3-OCH_3, 4-OH	ferulic acid
syringic acid	3,5-di-OCH_3, 4-OH	sinapic acid
-	3-OCH_3, 4-OH, 5-OH	5-hydroxyferulic acid
protocatechuic acid	3,4-di-OH	caffeic acid
gentisic acid	2,5-di-OH	-
gallic acid	3,4,5-tri-OH	-

1.2.3. Central role of HCAs and their corresponding CoA thioesters

All major groups of plant phenolics, except gallic acid and its derivatives, are synthesised by further metabolism of the simple HCAs. In many plants, hydroxycinnamoyl-CoA thioesters are the most common activated intermediates for phenylpropanoid biosynthesis (Ulbrich *et al.*, 1979). They are capable of entering various downstream pathways by taking part in different types of side-chain reactions. These include: side-chain elongation (condensation) with malonyl-CoA molecules leading to flavonoids, stilbenes, styrylpyrones, benzophenones, lignans and chain-elongated HCAs; NADPH-dependent reduction leading to dihydrocinnamic acids, hydroxycinnamyl alcohols (e.g. monolignols, precursors for lignins) and hydroxycinnamyl aldehydes; oxidative reactions resulting in side-chain shortening leading to benzoic acids; 2-hydroxylation and lactonisation leading to hydroxycoumarins; conjugation with hydroxyacids or amino compounds leading to ester (e.g. CGAs) and amide conjugates, respectively. A schematic illustration of the reactions from which varied phenolic compounds are derived is presented in Figure 3.

1.2.4. Structure of the chlorogenic acids (CGAs)

Hydroxycinnamic acids are commonly found as soluble conjugates, esters, amides or glycosides, within the cytoplasm of plant cells. Among these conjugates, CGAs have been shown to account for up to 90 % of the total phenolic fraction of some plant species (see below). The generic name "CGA" used to refer to the single compound 5-*O*-caffeoylquinic acid (5-CQA), which was first detected in green coffee beans by Robiquet and Boutron in 1837 (Sondheimer, 1964). The definition was later extended to include all esters of quinic acid with a cinnamic acid derivative (Clifford, 2000). Quinic acid (1,3,4,5-tetrahydroxycyclohexane-1-carboxylic acid) is an alicyclic acid containing 4 readily accessible hydroxyl groups (Figure 4). CGAs are often grouped with esters of shikimic acid (3,4,5-trihydroxy-1-cyclohexene-1-carboxylic acid), which differ from quinic acid by the presence of a double bond at C-1 and only three readily accessible hydroxyl groups (Figure 4). CGAs are generally classified depending on the identity, number and position of the acyl group (Clifford, 1999; Clifford, 2000). The common CGA groups are caffeoylquinic acids (CQAs), coumaroylquinic acids, feruloylquinic acids (FQAs) and dicaffeoylquinic acids (diCQAs). The chemical structure of major CGAs is shown in Figure 4.

Figure 3: Side-chain reactions involving hydroxycinnamoyl-CoA thioesters

Adapted from (Strack *et al.*, 1997).

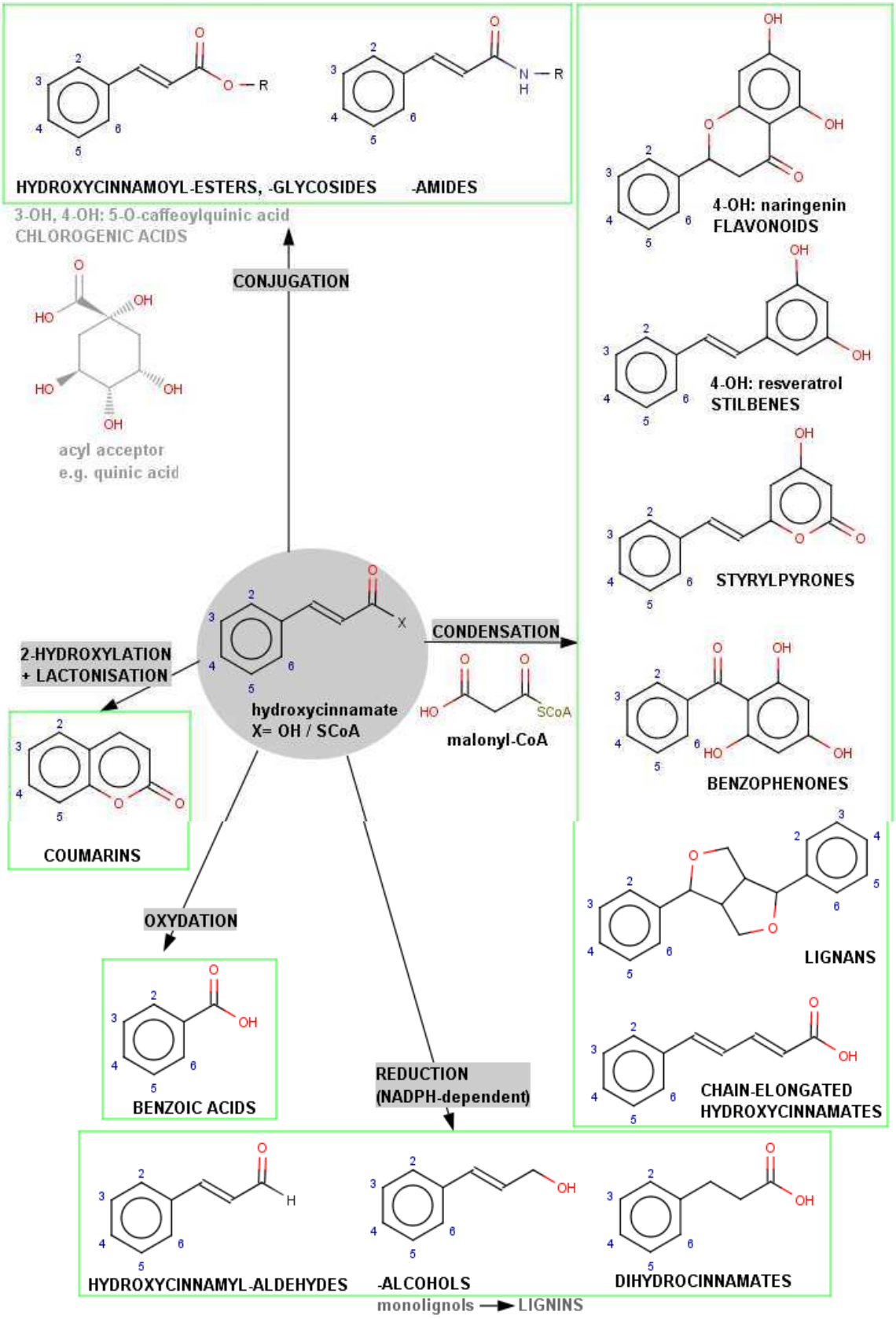
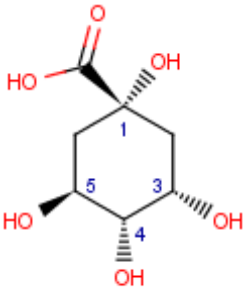
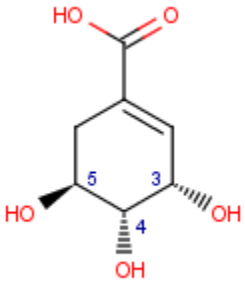


Figure 4: Chemical structures of quinic and shikimic acid acceptor molecules and major CGAs

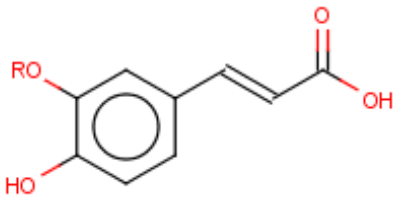
Compounds are named with the preferred IUPAC numbering (IUPAC, 1976). The ester linkages are circled.



quinic acid (QA)



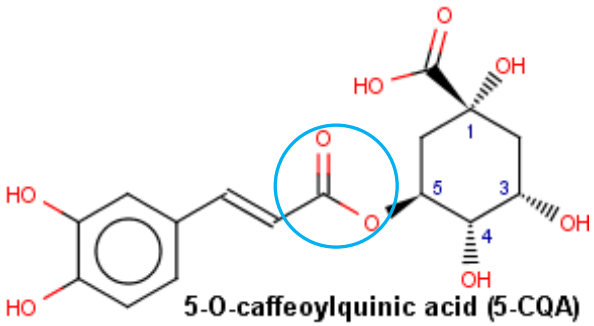
shikimic acid



R= H : caffeic acid (C in Table)
R= Me : ferulic acid (F in Table)

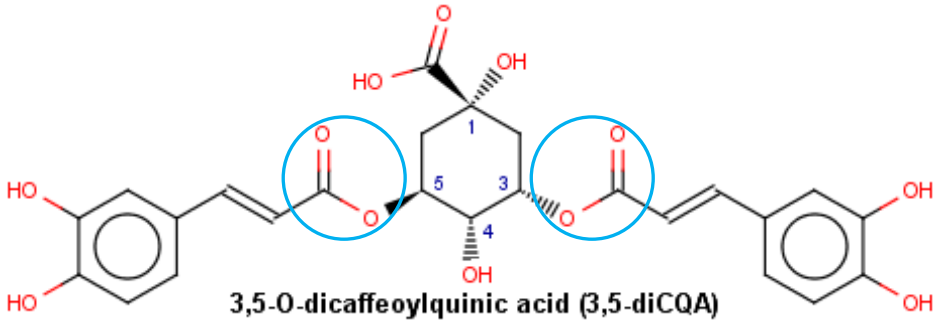
Name	*Abbreviation*	*Substitution*			
		C-1	C-3	C-4	C-5
quinic acid	QA	OH	OH	OH	OH
3-*O*-caffeoylquinic acid	3-CQA	OH	C	OH	OH
4-*O*-caffeoylquinic acid	4-CQA	OH	OH	C	OH
5-*O*-caffeoylquinic acid (1)	5-CQA	OH	OH	OH	C
1,3-di-*O*-caffeoylquinic acid	1,3-diCQA	C	C	OH	OH
1,5-di-*O*-caffeoylquinic acid	1,5-diCQA	C	OH	OH	C
3,4-di-*O*-caffeoylquinic acid	3,4-diCQA	OH	C	C	OH
3,5-di-*O*-caffeoylquinic acid (2)	3,5-diCQA	OH	C	OH	C
4,5-di-*O*-caffeoylquinic acid	4,5-diCQA	OH	OH	C	C
5-*O*-feruloylquinic acid	5-FQA	OH	OH	OH	F

(1)



5-O-caffeoylquinic acid (5-CQA)

(2)



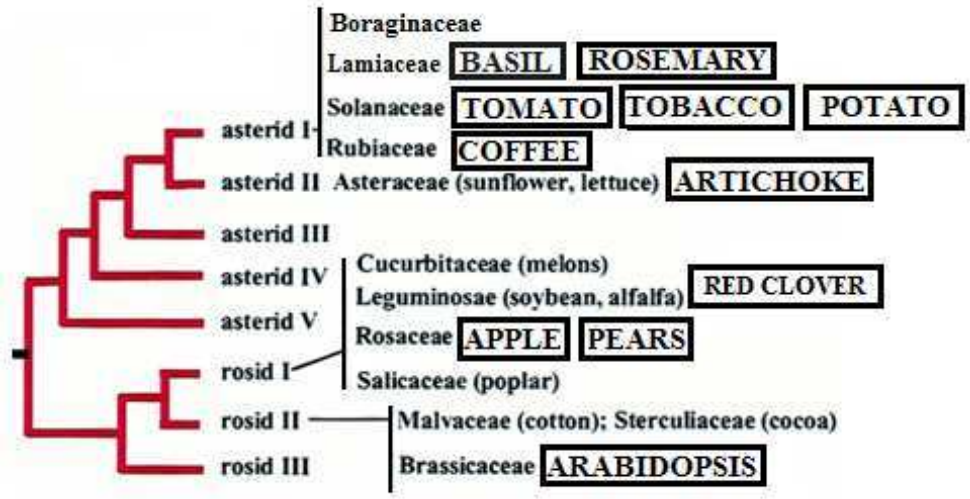
3,5-O-dicaffeoylquinic acid (3,5-diCQA)

1.2.5. Distribution of CGAs in the plant kingdom

CGA derivatives are abundant and diverse in the plant kingdom (Clifford, 1999; Clifford, 2000). They have been reported in several orders of mono- and dicotyledonous angiosperm species (Petersen *et al.*, 2009). The phylogenetically related *Rubiaceae* (e.g. plums, coffee), *Solanaceae* (e.g. tomato, tobacco, potato, eggplant), *Asteraceae* (e.g. artichoke, sunflower, chicory) and the more distant *Rosaceae* (e.g. apple, pears, peach) families possess most of the CGA-accumulating species (Figure 5). Monoesters of caffeic acid are relatively widespread, with 5-CQA being by far the most predominant isomer in green coffee beans (Clifford, 1979; Aerts *et al.*, 1994), potato tubers (Griffiths *et al.*, 1997), tomato leaves (Niggeweg *et al.*, 2004) and tobacco leaves (Hoffmann *et al.*, 2004). Some plant families such as *Solanaceae* and *Rubiaceae* also abound in diesters, such as diCQAs, diFQAs and mixed diesters (e.g. caffeoylferuloylquinic acids). Many plants, including coffee, produce CGAs derivatives esterified at C-3, C-4 and C-5, but not at C-1 of the quinic acid moiety. This is in contrast with artichoke and some other *Asteraceae* plants, which abound in 1,3-diCQA (cynarin) and 1,5-diCQA, and also produce tri- and tetracaffeoylquinic acids.

Figure 5: Phylogenetic tree of Asterids and Rosids

Adapted from (Ku *et al.*, 2000)



Coffee plants contain the most diverse and highest amounts of CGAs reported thus far. *C. arabica* and *C. canephora* are the two *Coffea* species used extensively in the beverages known commercially as Arabica and Robusta coffee respectively. A total of 45 different CGAs have been identified in green Arabica coffee beans (Clifford *et al.*, 2006), while 69 CGAs have been reported in green Robusta coffee beans (Jaiswal, 2010). Quantitatively, the major coffee CGAs are the monocaffeoylquinic acids (3-CQA, 4-CQA and 5-CQA), the dicaffeoylquinic acids (3,4-diCQA, 3,5-diCQA and 4,5-diCQA) and the monoferuloylquinic acids (mainly 5-FQA, Figure 4). 5-CQA is the most abundant soluble ester, representing 45 to 50 % of the total CGAs in *C. canephora* seeds (Koshiro *et al.*, 2007). Other CGA compounds, such as coumaroylquinic and sinapoylquinic acids, as well as mixed diesters (e.g. caffeoylferuloylquinic acids) and 3,4,5-trimethoxycinnamoylquinic acid, can be found in low levels in Robusta coffee (Clifford, 2000; Jaiswal, 2010). *C. canephora* produces substantially more CGAs than *C. Arabica*. The CGA level in green coffee grains varies between 7.9 and 14.4 % dry weight (DW) in *C. canephora* and 3.4 to 4.8 % DW in *C. arabica* (Ky *et al.*, 2001). (Lepelley *et al.*, 2007) reported lower but still relatively high CGA levels in *C. canephora* (6.6 % to 7.5 % DW). The content of diCQAs in *C. canephora* is also much higher than in *C. arabica*. CGA pools can vary with the coffee grain development (De Castro *et al.*, 2006; Lepelley *et al.*, 2007). While 5-CQA is the predominant molecule at all stages, 3-CQA and 4-CQA levels increase with seed growth. The 5-FQA level is relatively low in the early stage of grain development but rises 5 to 10 fold as the development progresses and reaches 22 % of total CGAs in ripe *C. canephora*. 3,5-diCQA is the major diCQA, especially in the early stages of fruit growth, but its level declines significantly during the development process.

Quinic acid is an abundant metabolite in young coffee grains, representing between 6 and 16 % DW. So, supply of this precursor may not restrict CGA biosynthesis as the grain develops. However, towards the end of grain development, quinic acid levels decrease below 1 % DW (Rogers *et al.*, 1999; Lepelley *et al.*, 2007). Considerable amounts of inositol (3-4 % DW) are found in young coffee grains and other organic acids, such as citric and malic acids, are dominant in the mature coffee grain (Rogers *et al.*, 1999). In Robusta coffee, caffeic and coumaric acids have also been found as conjugates of tryptophan (Murata *et al.*, 1995).

1.2.6. Related esters

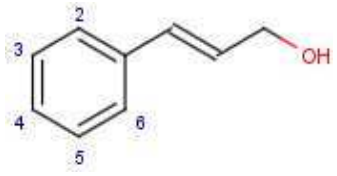
Shikimate esters (also named dactylifric acids) have been reported in *Palmae* plants (e.g. palm, date, endive) (Maier *et al.*, 1964; Harborne *et al.*, 1974; Goupy *et al.*, 1990; Ziouti *et al.*, 1996). The shikimic acid precursor (Figure 4) fails to accumulate to measurable levels in most of the important agronomic crops, although it is found in star anise (*Illicium verum*) and various species of evergreen trees. An ester of caffeic acid and 3,4-dihydroxyphenyllactic acid, commonly known as rosmarinic acid, is one of the active components of several medicinal plants and predominant in some *Boraginaceae* and *Lamiaceae* species, such as basil (*Coleus blumei*) and rosemary (*Rosmarinus officinalis*) (Petersen *et al.*, 2003; Abdullah *et al.*, 2008; Petersen *et al.*, 2009). HCAs may also be conjugated to other acyl acceptors such as aliphatic acids (e.g. acetic, citric, malic, glucaric, tartric acids), sugar alcohols (e.g. arabinose, galactose, glucose, rhamnose), amino compounds (e.g. tryptophan, tyramine, octopamine, anthranilate, agmatine, choline, spermidine) and lipids (Clifford, 2000). The leaves of red clover (*Trifolium pratense*) accumulate phaselic acid (2-*O*-caffeoyl-L-malic acid), which prevents the breakdown of forage proteins during storage (Sullivan, 2009). *Arabidopsis thaliana* and other members of the *Brassicaceae*, such as radish (*Raphanus sativus*), accumulate hydroxycinnamoyl esters formed with aliphatic acids. These include sinapoyl-malic acid, a leaf-specific ester, as well as sinapoyl-choline, a seed-specific ester, and their common biosynthetic precursor sinapoyl-glucose (Chapple *et al.*, 1992).

1.3. The biological role of CGAs in plants

1.3.1. Intermediates in lignin biosynthesis

HCAs occur in the insoluble-bound form in the secondary cell walls of plants as a result of their incorporation into complex lignin polymer substructures. Lignin is the principal component of cell walls in higher plants and one of the most abundant biopolymers on Earth, second only to cellulose. It is composed of cross-linked phenylpropanoid units derived from three main monomers: *p*-hydroxyphenyl, guaiacyl and syringyl lignin subunits (Campbell *et al.*, 1996; Boerjan *et al.*, 2003). The chemical structures of the major lignin subunits and their monolignol precursors are represented in Figure 6.

Figure 6: The major monolignols and lignin subunits



Aromatic substitution	*Monolignol*	*Lignin subunit*	*Short name*
4-OH	coumaryl alcohol	*p*-hydroxyphenyl	H-type
3-OCH_3, 4-OH	coniferyl alcohol	guaiacyl	G-type
3,5-di-OCH_3, 4-OH	sinapyl alcohol	syringyl	S-type

Coumaroyl-CoA is the biosynthetic precursor of coumaryl alcohol, and caffeoyl-CoA is the precursor of the coniferyl and sinapyl monolignols (Hoffmann *et al.*, 2004; Besseau *et al.*, 2007). CGA accumulation in coffee seeds is utilised for the deposition of lignin in the cell wall during germination (Aerts *et al.*, 1994; Mondolot *et al.*, 2006). These observations support the importance of CGAs as intermediates in the biosynthesis of the S- and G-type lignins. Cell-wall bound phenolics provide a plant with the ability to transport water and solutes through the vascular system and resist desiccation. In addition, lignins are crucial for the structural integrity of the cell wall and participate in the elaboration of an upright and rigid plant body (Abdulrazzak *et al.*, 2006). Lignins also play a defensive role against herbivores and pathogens (Boerjan *et al.*, 2003).

1.3.2. Defence compounds

CGA biosynthesis is regulated by developmental processes in specific tissues and can be activated in response to a variety of environmental stresses (Hahlbrock *et al.*, 1989; Dixon *et al.*, 1995; Abdulrazzak *et al.*, 2006). CGAs have been shown to accumulate in response to pathogenic attack, representing a chemical defence mechanism (Ruelas *et al.*, 2006). CGA synthesis is also induced by wounding or elicitor treatment, as observed in the potato tuber (Matsuda *et al.*, 2005). Elevated CGA levels in tomato plants have been shown to increase their resistance against bacterial infection (Niggeweg *et al.*, 2004). CGAs also protect plants from abiotic stresses such as UV-damage and oxidative pressure (Burchard *et al.*, 2000). This is because phenolic compounds absorb light between 280 and 340 nm wavelength, thus filtering much of the damaging UV-B radiation (Caldwell *et al.*, 1983). Evidence for this comes from the presence of high levels of hydroxycinnamoyl esters, which constitute a UV screen in the leaves of tobacco and *A. thaliana* plants (Cerovic *et al.*, 2002). Increased UV-protection, which is influenced by the total phenolic level as well as the specific composition

profile, is also associated with elevated CGA accumulation in tomato leaves (Niggeweg *et al.*, 2004; Cle *et al.*, 2008). CGAs are also found to be associated with chloroplasts in the young leaves of coffee plants (Ranjeva *et al.*, 1977), again supporting their protective role against light-induced damage. CGAs therefore fulfil multiple physiological roles in plants, such as defence compounds against various biotic and abiotic stresses, and as building blocks for lignins. These observations support the crucial role of the phenylpropanoid pathway in plant evolution, allowing them to adapt from an aquatic to a dry land environment (Douglas, 1996; Waters, 2003).

1.4. The impact of CGAs in humans

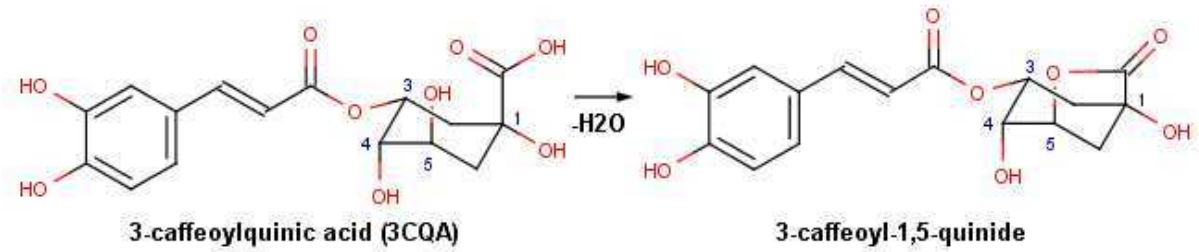
1.4.1. On the food properties

Besides the importance for the plant itself, HCAs and CGAs have attracted focus because they can influence food properties such as colour, taste and aroma. In general, phenolic compounds are considered to negatively influence food selection as they impart a bitter flavour (Drewnowski, 1997). CGAs, especially the diesters of caffeic acid, constitute a class of astringent compounds (Ohiokpehai *et al.*, 1983). In coffee beverages, (Farah *et al.*, 2006) reported that 3,4-diCQA levels in green and roasted coffee strongly correlate with satisfactory sensory attributes, but that higher levels of CQAs and 5-FQA are detrimental to beverage quality. A higher CGA content could be responsible for the perceived inferior quality of brews made with Robusta, which are occasionally described as bitter (Bertrand *et al.*, 2003).

During food processing, naturally occurring phenolic compounds can undergo thermally-induced reactions that have a significant impact on both food aroma and taste (Clifford, 1985; Jiang *et al.*, 2010). CGAs are degraded to chlorogenic acid lactones (CGLs) by the loss of a water molecule from the quinic acid moiety of CGA and the formation of an intramolecular ester bond during coffee roasting (Figure 7). Only CGAs that lack a substituent at C-5 of the quinic acid moiety are able to form a caffeoyl-1,5-quinide. CGLs considerably contribute to coffee bitterness, an attribute that is not generally appealing to the consumer (Farah *et al.*, 2005; Frank *et al.*, 2006). During coffee roasting, the diCQAs can also be hydrolysed and produce caffeic acid and a CQA moiety (Leloup *et al.*, 1995). At later stages of roasting, the thermal degradation of HCAs leads to major aroma compounds (e.g. caffeic acid → catechol) (Leloup *et al.*, 1995).

Additionally, CGAs are known to be good substrates for polyphenol oxidase, which is involved in the browning of fruits and vegetables after harvesting (Maier *et al.*, 1964; Murata *et al.*, 2002). The ability of the CGAs to form complexes with ferric ions during the cooking process has also been shown to be the primary cause in the development of the after-cooking blackening observed in potatoes (Griffiths *et al.*, 1992).

Figure 7: Thermally-induced formation of a chlorogenic acid lactone (CGL)



1.4.2. On health

CGAs and other HCA derivatives are significant components of diets rich in plant-based food such as cereals, legumes, oilseeds, fruits, vegetables and beverages. An increasing number of epidemiological studies have shown that people who consume higher quantities of plant-based food appear to have lower risks for significant health problems, such as cardiovascular diseases and certain cancers (Clifford, 2004; Finley, 2005). This health-promoting effect has been associated with the presence of phenolic phytonutrients. When compared to other beverages, coffee represents the richest source of CGAs and, interestingly, it has the highest antioxidant activity (Clifford, 1999; Wang *et al.*, 2009). The presence of conjugated ring structures and hydroxyl groups allows CGAs to actively scavenge free radicals generated in the aqueous phase (Rice-Evans *et al.*, 1996). This ability can prevent oxidative damage and lipid peroxidation in living tissues. While caffeic acid and 5-CQA have similar antioxidant capacities *in vitro* (Rice-Evans *et al.*, 1996), the number and nature of substitutions on the aromatic ring can influence the antioxidant properties. For instance, it has been shown *in vitro* that hydroxyl groups are more electron donating than methoxyl groups and that dihydroxyl-substituted caffeic acid shows a higher antioxidant activity than mono-substituted coumaric acid (Foley *et al.*, 1999; Shahidi *et al.*, 2010). Further studies demonstrate that CGAs can exert their health benefits through various other mechanisms, such as metal chelation and modification of redox potentials (Rice-Evans *et al.*, 1996).

CGAs have attracted much attention due to their exceptional antioxidant properties, as well as their high bioavailability and absorption in humans (Nardini *et al.*, 2002). Following coffee intake, CGAs are absorbed directly by the small intestine or hydrolysed by the large intestine microflora to release caffeic acid (Nardini *et al.*, 2002; Stalmach *et al.*, 2009; Stalmach *et al.*, 2010). This degradation product, which has same antioxidant capacities as 5-CQA *in vitro* (see above), is also absorbed and relatively stable in the gut (Scalbert *et al.*, 2002). CGAs have been attributed many other health benefits including anti-inflammatory, antibacterial, antiproliferative and anticarcinogenic properties (Rice-Evans *et al.*, 1997). CGAs have also been associated with reduced hepatic glucogenolysis and glucose absorption (Johnston *et al.*, 2003), inhibition of HIV-1 integrase (McDougall *et al.*, 1998), caffeine antagonistic cerebral effects (de Paulis *et al.*, 2004), reduced incidence of atherosclerosis, type 2 diabetes and various types of cancer (McCarty, 2005; Natella *et al.*, 2007). CGAs have been proposed to play a role in the prevention of neuro-degenerative diseases (Shahidi *et al.*, 2010).

Modulating CGA levels in plants therefore could have a major impact in nutrition and health. The growing recognition of the beneficial effects of CGAs in human health has supported the development of new nutraceuticals based on CGA-rich plant extracts (e.g. Svetol®). CGAs and their degradation products also appear to be an essential criterion for the quality of agro-industrial products. For example, genetic markers could be developed to select coffee varieties with desired CGA content and profile to modulate the beverage organoleptic and bioactive properties. As intermediates of lignins, the biosynthesis of CGAs is also crucial for alternate energy supplies and in the pulping industry as it affects the efficiency of bioconversion and processing of lignocellulosic raw material (Baucher *et al.*, 2003; Chen *et al.*, 2007). Depending on the desired effect, reducing or increasing the expression of a particular gene/ enzyme in the phenylpropanoid pathway can have major impacts on the different substrate pools (Niggeweg *et al.*, 2004; Dixon, 2005; Ververidis *et al.*, 2007). The following part of this introductory chapter is dedicated to the presentation of the shikimic acid and general phenylpropanoid pathways leading to the key CGA precursors, and the genes/ enzymes directly associated with CGA biosynthesis and modification.

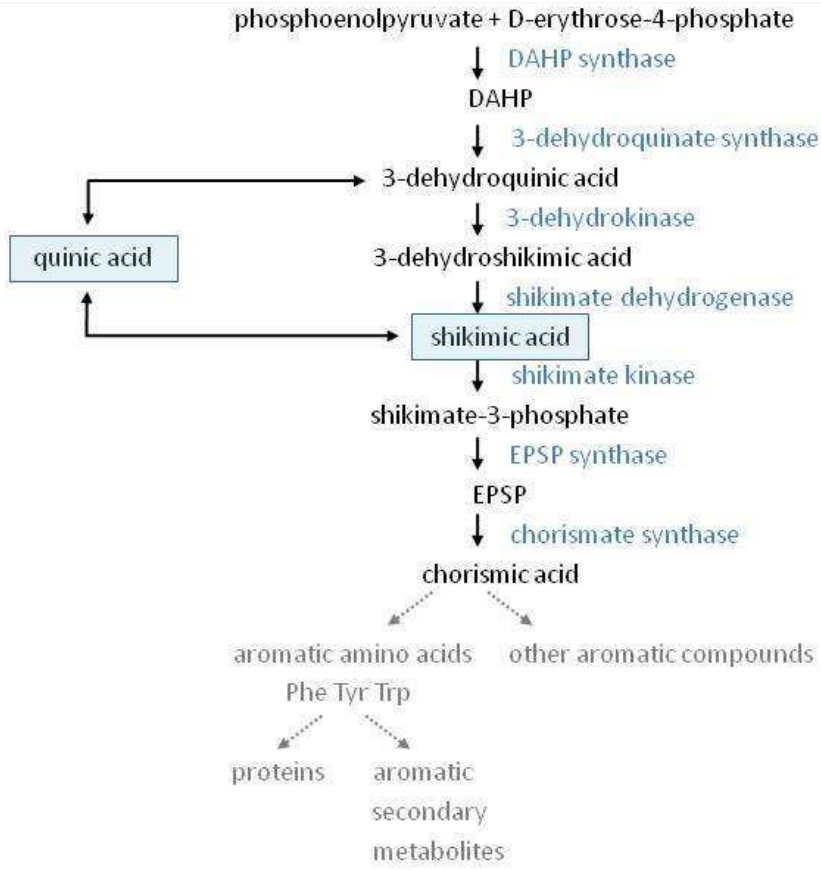
1.5. Provider pathways of CGA precursors

1.5.1. The shikimic acid pathway

The shikimic acid pathway, also known as the common aromatic biosynthetic pathway, is essential for the production of all the key aromatic compounds. These include the three aromatic amino acids found in proteins (L-Phe, L-Tyr and L-Trp). The shikimic acid pathway is ubiquitously employed by plants and micro-organisms. Mammals lack this pathway and derived compounds, so they depend on consumption for their supply. A schematic view of the shikimic acid pathway is presented in Figure 8.

Figure 8: Shikimic acid biosynthetic pathway and quinic acid cycle, adapted from (Herrmann *et al.*, 1999)

DAHP: 3-dehydroxy-D-arabino-heptulose-7-phosphate; EPSP: 5-enolpyruvate shikimate 3-phosphate.



The first reaction involves the aldol-like condensation of the carbohydrate precursors, phosphoenolpyruvate and D-erythrose 4-phosphate, giving 3-dehydroxy-D-arabino-heptulose-7-phosphate (DAHP). This reaction is catalysed by the enzyme DAHP synthase. 3-dehydroquinate synthase subsequently converts DAHP into 3-dehydroquinic acid by a

sequence of reactions requiring an oxidation, a β-elimination, a reduction and an intramolecular aldol condensation. The dehydration of 3-dehydroquinic acid to 3-dehydroshikimic acid is catalysed by 3-dehydroquinase. The hydration of 3-dehydroshikimic acid to shikimic acid is catalysed by shikimate dehydrogenase. Phosphorylation of shikimic acid to shikimate 3-phosphate is brought about by shikimate kinase in the presence of ATP. The enol ether 5-enolpyruvate shikimate 3-phosphate (EPSP) is produced via the condensation of shikimate 3-phosphate with phosphoenolpyruvate by EPSP synthase. The elimination of phosphoric acid from EPSP is catalysed by the enzyme chorismate synthase yielding chorismate. Several pathways for the conversion of chorismic acid into the aromatic amino acids L-phenylalanine and L-tyrosine are known to exist depending on the organism.

The shikimic acid pathway provides the aromatic amino acid phenylalanine, which is the major precursor for HCAs and their derivatives, including CGAs, lignins, flavonoids and stilbenes. In higher plants, the carbon of the shikimic acid primary metabolite can also be diverted and stored in the form of quinic acid, which is another major precursor of CGAs (Herrmann, 1995). Quinic acid can be synthesised from 3-dehydroquinic acid or from shikimic acid by the action of specific enzymes (Herrmann, 1995). Quinic acid can also be converted back into intermediates of the shikimic acid pathway either by oxidation to 3-dehydroquinate or by dehydration directly to shikimic acid. Characterisation of the enzymes of the quinic acid cycle as anabolic, catabolic, or both, remains to be made (Herrmann, 1995).

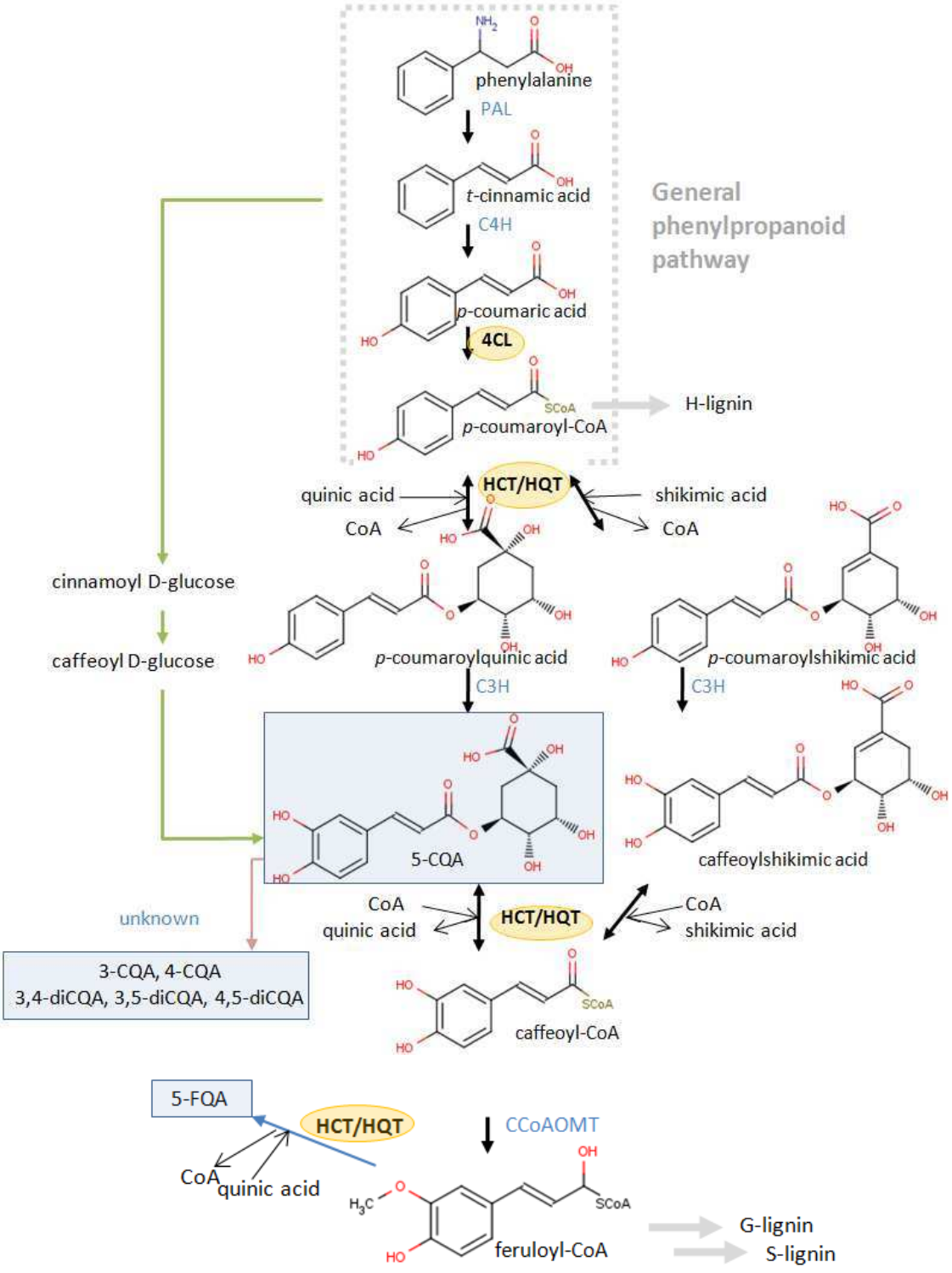
1.5.2. The general phenylpropanoid pathway

1.5.2.1. Overview

The general phenylpropanoid pathway represents a major branch-point between primary and secondary metabolism in plants (Dixon *et al.*, 1995; Rohde *et al.*, 2004). The pathway, represented in Figure 9, starts with the amino acid Phe and leads to coumaroyl-CoA thioester in three steps catalysed by highly conserved enzymes (Dixon *et al.*, 1995). Deamination by phenylalanine ammonia lyase (PAL) produces *t*-cinnamic acid. This is followed by the introduction of a hydroxyl group at the C-4 position of the phenyl ring by cinnamate 4-hydroxylase (C4H), producing 4-coumaric acid. The carboxylate substrate is then activated by the formation of a thioester bond with CoA catalysed by 4-coumarate CoA ligase (4CL).

Figure 9: Current view of the phenylpropanoid pathway leading to CGAs

Adapted from (Hoffmann *et al.*, 2004) and (Niggeweg *et al.*, 2004).
PAL: phenylalanine ammonia-lyase; C4H: cinnamate 4-hydroxylase; 4CL: 4-coumarate CoA ligase; HCT: hydroxycinnamoyl-CoA shikimate/ quinate hydroxycinnamoyltransferase; HQT: hydroxycinnamoyl-CoA quinate / shikimate hydroxycinnamoyltransferase; C3H: coumaroyl ester 3-hydroxylase; CCoAOMT: caffeoyl-CoA O-methyltransferase. The main CGA compounds of coffee are highlighted in blue.



1.5.2.2. Phenylalanine ammonia lyase (PAL)

Phenylalanine ammonia lyase (PAL, EC 4.3.1.5) catalyzes the stereospecific elimination of ammonia from L-Phe. The reaction leads to *t*-cinnamic acid, which provides the characteristic C_6-C_3 skeleton of the phenylpropanoid compounds. Four different PAL genes have been characterised in *A. thaliana* (Raes *et al.*, 2003). The first crystal structures of PALs from *Rhodosporidium toruloides* and *Petroselinum crispum* (parsley) were solved in 2004 (Calabrese *et al.*, 2004; Ritter *et al.*, 2004). PALs are tetramers that are largely made up of α-helices. The PAL active site reveals an unusual electrophilic prosthetic group, 4-methylidene-5-one, which acts as a covalent cofactor required for the non-oxidative elimination of ammonia (Ritter *et al.*, 2004).

1.5.2.3. Cinnamate 4-hydroxylase (C4H)

The second enzyme in the general phenylpropanoid pathway catalyses the hydroxylation at the C-4 position of the phenylalanine aromatic ring to form 4-coumaric acid. Cinnamate 4-hydroxylase (C4H, EC 1.14.13.11) is a member of the cytochrome P450 hydroxylase family. P450s, so-named because of the characteristic absorption of their catalytic iron-containing heme cofactors, catalyse irreversible reactions and are hypothesised to serve as membrane-associated anchor sites for putative multiprotein complexes (Winkel-Shirley, 1999; Ehlting *et al.*, 2006). cDNAs encoding C4H were isolated from different plants and called CYP73 based on standard P450 enzyme nomenclature. They have highly similar sequences and are specific for *t*-cinnamic acid. The first C4H enzyme was purified from Jerusalem artichoke and a number of genomic DNA sequences have been isolated from diverse plant species. In *Arabidopsis thaliana*, only one gene has been found, whereas in some other plants, two or more C4H genes have been characterised. No plant representative of this enzyme has so far been structurally characterised.

1.5.3. 4-Coumarate CoA ligase (4CL)

1.5.3.1. Isolation and characterisation of 4CL genes/ enzymes

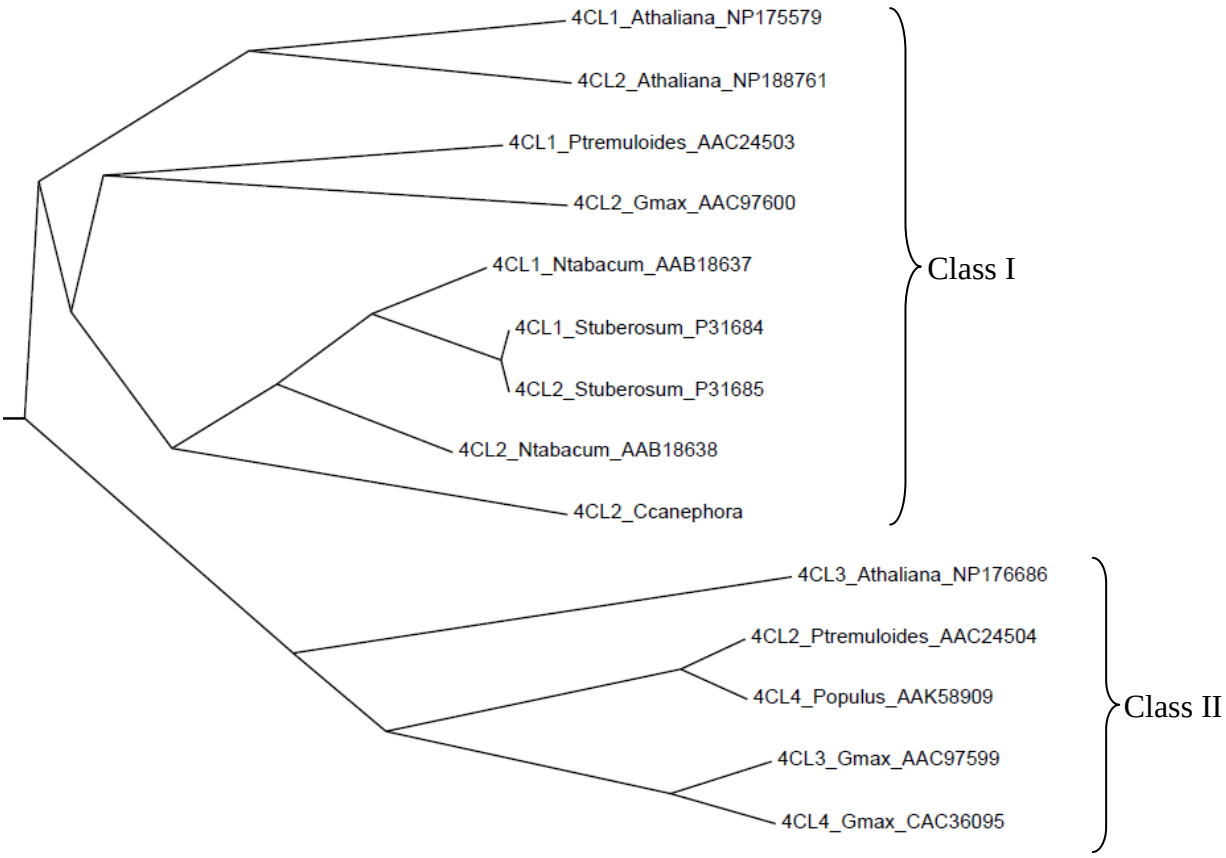
The final enzyme of the general phenylpropanoid pathway is 4-coumarate CoA ligase (4CL, EC 6.2.1.12), which catalyses the conversion of 4-coumaric acid to coumaroyl-CoA. 4CL enzymes are encoded by gene families in all plants examined to date. They are expressed differentially depending on the tissue type and development stage. In some plants, the genes encode identical, or very similar, proteins (e.g. parsley, loblolly pine and potato). In other species, structurally divergent isoforms have been identified (e.g. tobacco, arabidopsis, aspen,

hybrid poplar and soybean). These isoenzymes generally display broad but distinct substrate specificity and/ or tissue distribution (Lee *et al.*, 1996; Ehlting *et al.*, 1999; Costa *et al.*, 2005).

Phylogenetic comparisons indicate that 4CLs can be classified into two major clusters (Figure 10). Class I proteins (e.g. *At*4CL1, *At*4CL2, *Pt*4CL1) are more related to each other than to class II proteins (e.g. *At*4CL3, *Pt*4CL2) from the same plant (Hu *et al.*, 1998). Class II 4CLs are generally associated with flavonoid biosynthesis, whereas class I 4CLs are more closely related with the biosynthesis of lignins and other phenylpropanoids (Hu *et al.*, 1998; Ehlting *et al.*, 1999; Cukovic *et al.*, 2001).

Figure 10: Cladogram of the amino acid sequences of 4CL enzymes

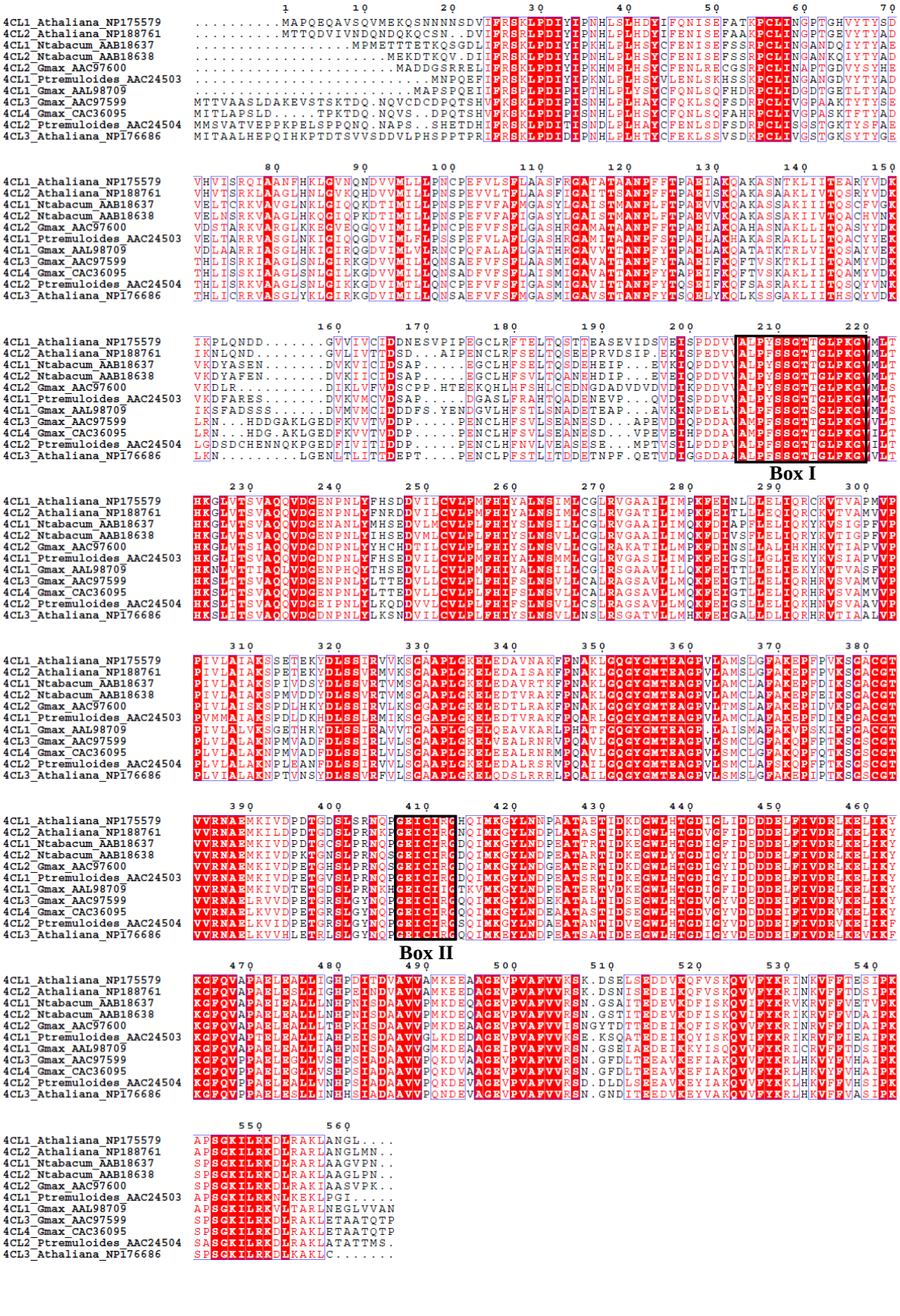
The cladogram was drawn from a ClustalW alignment with the Phylodendron server (http://iubio.bio.indiana.edu/treeapp/treeprint-form.html). The enzyme abbreviations, species and accession numbers are all included in the names.



Two divergent members classified as 4CL1 and 4CL2 have been identified in *Nicotiana tabacum* (common tobacco); they utilise similarly coumaric, caffeic and ferulic acids as substrates (Lee *et al.*, 1996). A partial cDNA encoding 4CL1 and a complete cDNA encoding 4CL2 have been isolated from *Coffea canephora* (Lepelley *et al.*, 2007). In *Glycine max* (soybean), three structurally and functionally distinct 4CL cDNAs encoding 4CL1, 4CL2, and 4CL3 were isolated and have clear differences in expression patterns and enzymatic properties (Lindermayr *et al.*, 2002). A number of "4CL-like" genes have been identified in the *A. thaliana* genome, only four of which encode proteins with hydroxycinnamate CoA ligase activity *in vitro* (Costa *et al.*, 2005). *At*4CL1, *At*4CL2 and *At*4CL3 were found to be capable of forming CoA esters with coumaric and caffeic acids. *At*4CL1 and *At*4CL3 were able to convert ferulic acid to the corresponding CoA thioester, whereas *At*4CL2 was not (Ehlting *et al.*, 1999). *At*4CL4 exhibits the rare property of efficiently activating sinapic acid, besides the usual substrates (Hamberger *et al.*, 2004). Biochemical analysis showed that some "4CL-like" *Arabidopsis thaliana* genes have the capacity to activate different long-chain fatty acid derivatives, including the jasmonic acid precursors (Schneider *et al.*, 2005; Souza *et al.*, 2008). The precise function of different 4CL isoforms is largely unknown and a number of 4CL-like genes remain to be characterised. 4CL enzymes share 56 to 93 % sequence identity at the amino acid level and are characterised by several conserved sequence motifs (Figure 11). The Box I motif is rich in Gly, Ser and Thr, and constitutes a nucleotide-binding signature (Stuible *et al.*, 2000). Another conserved sequence motif, Box II (GEICI[R/I]G), was proposed to be involved in maintaining the structural integrity of the enzyme (Stuible *et al.*, 2001).

Figure 11: Multiple sequence alignment of 4CL enzymes from different plant species

The alignment was generated by ClustalW and ESPript. The short names, species and accession numbers are all included in the names. Black boxes highlight highly conserved sequence motifs (Box I and II).



1.5.3.2. The adenylate-forming enzyme superfamily

From sequence comparisons, 4CLs belong to a superfamily of prokaryotic and eukaryotic adenylate-forming enzymes. Members of this superfamily transfer a wide variety of acyl, aryl and aminoacyl moieties to the phosphopantetheine tail of CoA or carrier proteins via an adenylate intermediate. Depending on the type of carboxylate-containing substrate, the superfamily has been functionally divided into three subclasses: the acyl-/ aryl-CoA synthetases (including 4CLs, medium and long-chain fatty acid CoA ligases and acetate CoA ligases), the adenylation domains of non-ribosomal polypeptide synthetases (NRPS) (such as the phenylalanine-activating domain (PheA) of gramicidin S-synthetase) and the luciferase oxidoreductases. The sequence identity of proteins from different subfamilies are rather low, but they share conserved peptide motifs that justify their classification in a superfamily (Fulda *et al*., 1994). 4CL enzymes form a plant-specific group most closely related to the plant luciferases.

Members of the adenylate-forming enzyme superfamily share a common mechanism that proceeds in two partial reactions involving several coordinated substrate binding and product release steps (Gulick, 2009). The first half reaction is activation of the carboxylate substrate by adenylation with Mg-ATP producing pyrophosphate (PPi) as a by-product (Figure 12A). The second partial reaction involves the nucleophilic attack of the CoA thiol at the carbonyl of the adenylate intermediate to form the CoA thioester and adenosine monophosphate (AMP) products (Figure 12B).

Figure 12: Two-step CoA ligation reaction catalysed by adenylate-forming enzymes

(A) Adenylation step: X + Mg-ATP → X-AMP + Mg-PPi

(B) Thioesterification step: X-AMP + Y → X-Y + AMP

Enzyme	*Substrate X*	*Substrate Y*
4CL	(hydroxy)cinnamic acids	CoA
luciferase	luciferin	oxygen
PheA	phenylalanine	amino acid

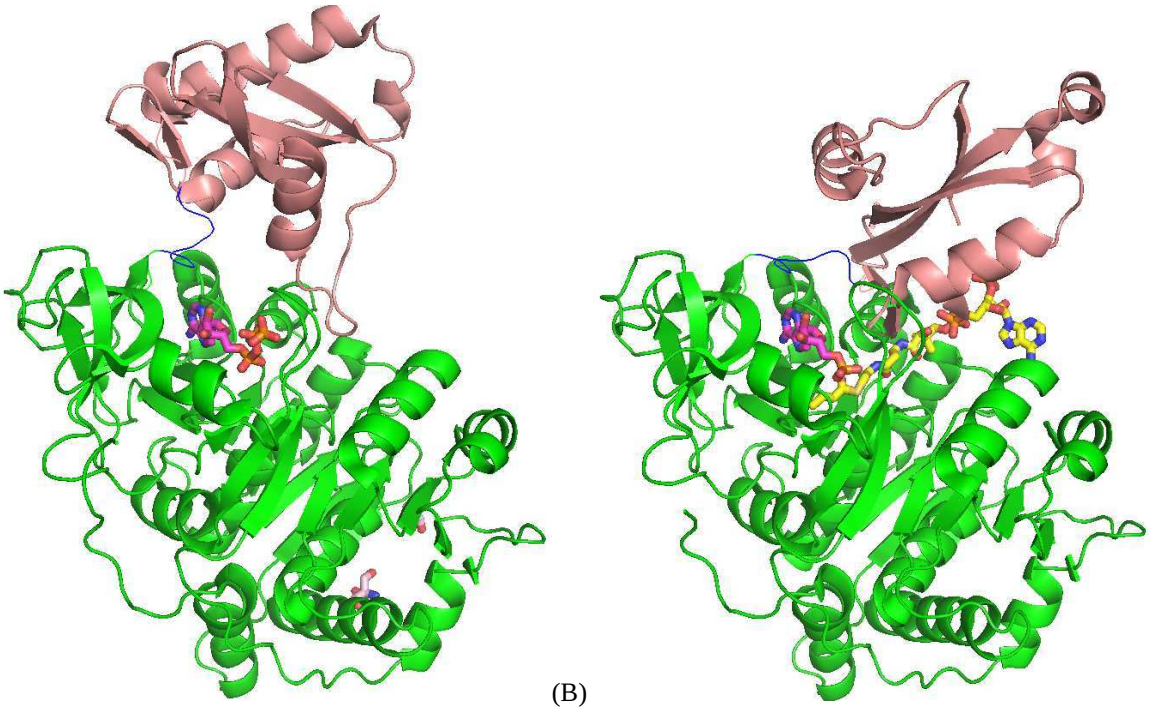
1.5.3.3. Crystal structures

Crystal structures for a dozen members of the adenylate-forming enzyme superfamily are available in the Protein Data Bank (PDB) (Berman *et al.*, 2002). The first crystal structure to be solved was firefly luciferase, which catalyses the bioluminescence reaction in which luciferin, Mg-ATP and molecular oxygen yield the electronically excited oxyluciferin (Conti *et al.*, 1996; Nakatsu *et al.*, 2006). The bacterial phenylalanine-activating domain (PheA) of gramicidin S-synthetase was co-crystallised with ATP and phenylalanine (Conti *et al.*, 1997).

Despite their low sequence identities, adenylate-forming enzymes adopt a similar two-domain fold composed of a large N-terminal domain and a small C-terminal domain connected by a short hinge region. The crystal structures of the apo and complex forms of human medium-chain acyl-CoA synthetase (ACSM2A) revealed a substantial rearrangement between the N- and C-terminal domains driven by the identity of the bound ligand in the active site (Kochan *et al*., 2009) (Figure 13). This domain alternation was shown to induce the selective catalysis of the adenylation or thioesterification steps (Reger *et al.*, 2007; Reger *et al.*, 2008; Kochan *et al.*, 2009).

Figure 13: Crystal structures of human acetyl-CoA synthase (ACSM2A)

(A) ACSM2A-ATP (PDB: 3C5E) representative of the adenylate-forming conformation. (B) ACSM2A-butyryl-CoA-AMP (PDB: 3EQ6) in the thioester-forming conformation. The N-terminal domain (residues 33-459) is coloured in green, the C-terminal domain (466-569) in salmon and the linker region in blue. AMP/ATP and butyryl-CoA are shown respectively with magenta and yellow carbon atoms.



(A) (B)

1.5.3.4. A branch point of the general phenylpropanoid pathway

4CLs catalyse the formation of hydroxycinnamoyl-CoA thioesters from the corresponding HCA and CoA in the presence of Mg-ATP. These activated compounds are central substrates for subsequent condensation, reduction and transfer reactions leading to phenolic secondary metabolites. 4CL plays a crucial role in channelling the carbon flow into diverse branches leading to various end products. Coumaroyl-CoA is the common substrate for chalcone synthase, stilbene synthase and hydroxycinnamoyl-CoA shikimate/ quinate hydroxycinnamoyltransferases (HCT/ HQT). 4CL thus controls the metabolic flux entering the biosynthetic routes leading either to flavonoids and stilbenes or to CGAs and monolignols. chalcone and stilbene synthases are two closely related polyketide synthases catalysing the first step in flavonoid and stilbene biosynthesis respectively. chalcone synthase catalyses the condensation of three molecules of malonyl-CoA and one molecule of coumaroyl-CoA to give naringenin chalcone, which is the first intermediate in flavonoid biosynthesis (Winkel-Shirley, 2001). Stilbene synthase uses the same substrates and an alternate cyclisation pathway to form the phytoalexin resveratrol. The crystal structures of chalcone and stilbene synthases provided a molecular understanding of the reactions leading to chalcone and stilbene synthesis respectively (Ferrer *et al.*, 1999; Shomura *et al.*, 2005). In addition, coumaroyl-CoA, feruloyl-CoA and sinapoyl-CoA are the main substrates of cinnamoyl-CoA reductase. Cinnamoyl-CoA reductase catalyses the reduction of hydroxycinnamoyl-CoA thioesters to hydroxycinnamyl aldehydes, which are subsequently reduced into guaiacyl and syringyl monolignols by cinnamyl alcohol dehydrogenase during the lignification process. Hydroxycinnamoyl-CoA thioesters are also substrates for the hydroxycinnamoyl-CoA shikimate/ quinate hydroxycinnamoyltransferases involved in CGA biosynthesis (see below).

1.6. Genes/ enzymes of the core pathway for CGA biosynthesis

1.6.1. Coumaroyl ester 3-hydroxylase (C3H)

Free HCAs have long been thought to be key intermediates in the phenylpropanoid biosynthesis pathway. However, it is now clear that many enzymatic conversions occur at the ester level. The conversion of coumaroyl to caffeoyl moieties is catalysed by coumaroyl ester 3-hydroxylase (C3H, EC 1.14.14.1) (Mahesh *et al.*, 2007), which belongs to the CYP98 family of cytochrome P450 mono-oxygenases. C3H is responsible for the 3-hydroxylation of the aromatic ring of phenolic substrates, but this only occurs when it is esterified to either quinic or shikimic acid (Schoch *et al.*, 2001; Franke *et al.*, 2002; Nair *et al.*, 2002;

Abdulrazzak *et al.*, 2006). Three C3H genes have been identified in *A. thaliana* (Schoch *et al.*, 2001). Two CYP98 genes were characterised from *C. canephora*, but only one of them, CYP98A35, was shown to be able to hydroxylate coumaroylquinic acid to form the caffeoyl ester (Mahesh *et al.*, 2007). (Lepelley *et al.*, 2007) isolated and characterised full-length cDNAs encoding C3H1 from *C. canephora*. Coumaroylshikimic acid is the preferred substrate for all of the recombinant CYP98 proteins tested so far (Schoch *et al.*, 2001; Mahesh *et al.*, 2007; Sullivan *et al.*, 2010).

1.6.2. Hydroxycinnamoyltransferases (HCT, HQT)

1.6.2.1. Possible routes leading to CGA biosynthesis

Hydroxycinnamoyl-CoA shikimate/ quinate hydroxycinnamoyltransferases (HCT/ HQT, EC 2.3.1.99) catalyse the esterification of coumaroyl-CoA, supplied by the general phenylpropanoid pathway, with quinic and shikimic acids. The combined activities of HCT/ HQT and C3H can lead to the accumulation of chlorogenic (quinate esters) and dactylifric (shikimate esters) acids. HCT and HQT can also act after the 3-hydroxylation step by catalysing the reverse acyltransfer reaction leading to caffeoyl-CoA, a major biosynthetic precursor for S- and G-lignin subunits, therefore contributing to CGA turnover (Comino *et al.*, 2007). An alternative route, where 5-CQA would be synthesised directly from caffeoyl-CoA via HCT/ HQT catalysis has been proposed, but experimental evidence has not been clearly demonstrated for this yet. An alternative pathway catalysed by a hydroxycinnamoyl-glucose quinate hydroxycinnamoyltransferase has been proposed in sweet potato (*Ipomoea batatas*), where caffeoyl-glucose is the precursor of caffeoylquinic acid (Villegas *et al.*, 1986; Moriguchi *et al.*, 1988) (Figure 9). However, the biosynthesis of other mono- and diesters is still unclear. An enzyme catalysing the synthesis of 3,5-diCQA from two 5-CQA molecules have been described (Villegas *et al.*, 1987), but this has not been supported by further studies.

1.6.2.2. Isolation of hydroxycinnamoyltransferase genes/ enzymes

HCT/ HQT are responsible for the last step of CGA biosynthesis. These enzymes can shuttle hydroxycinnamoyl moieties between CoA-thioesters and shikimate/ quinate-esterified forms. cDNAs encoding HCT and HQT enzymes have been isolated and characterised from several plant species: tobacco (Hoffmann *et al.*, 2003; Hoffmann *et al.*, 2004; Niggeweg *et al.*, 2004), coffee (Lepelley *et al.*, 2007), artichoke (Comino *et al.*, 2007; Comino *et al.*, 2009; Sonnante *et al.*, 2010), tomato (Niggeweg *et al.*, 2004; Cle *et al.*, 2008) and red clover (Hoffmann *et al.*, 2003; Niggeweg *et al.*, 2004; Sullivan, 2009). Some of them possess both,

such as coffee, artichoke and tomato, with HCT and HQT showing a higher affinity for shikimic and quinic acid, respectively. No *hqt* gene was identified in *A. thaliana* (Hoffmann *et al.*, 2005). HCT/ HQT enzymes of all plant species share between 30 and 96 % sequence identity at the amino acid level.

1.6.2.3. Potential physiological role of HCT and HQT

Histochemical and RNA expression studies indicate a difference in tissue localisation: highest levels of *hct* and *hqt* are detected in the stems and leaves of tobacco and tomato plants, respectively (Hoffmann *et al.*, 2004; Niggeweg *et al.*, 2004). The results of both transient and stable transformation of tomato plants indicate that tomato HQT is the enzyme of greatest importance for CGAs biosynthesis (Niggeweg *et al.*, 2004). Similar assays in tobacco plants showed that the artichoke *hqt* gene induced a higher production of CGA and cynarin (Sonnante *et al.*, 2010). In coffee, the higher transcript levels observed in the early part of grain development and in the leaf imply that HQT may exert a stronger influence on CGA biosynthesis and accumulation in these tissues. In contrast, *hqt* silencing caused up to a 98 % reduction of CGA levels in the leaves of tomato without changing the levels of other polyphenols (Niggeweg *et al.*, 2004). Moreover, over-expression of *hqt* resulted in an increase in CGA levels of up to 85 % in the leaves. All these studies are consistent with HQT playing a major role in CGA biosynthesis. HQT homologues are only found in CGA-accumulating species (e.g. coffee, tomato and artichoke), so this may explain why the *Arabidopsis* model species does not accumulate CGAs. *Hqt* genes are therefore key targets for engineering plants with increased CGAs for improved antioxidant and microbial properties.

The expression levels of HCT, particularly in branch tissue, suggest that it plays an important role in lignin formation (Lepelley *et al.*, 2007). *Hct* down-regulation in *A. thaliana* impacts on their lignin content and composition (Hoffmann *et al.*, 2004). In HCT-deficient plants, the metabolic flux deriving from phenylalanine was redirected towards flavonoid synthesis through chalcone synthase (Besseau *et al.*, 2007). The more prevalent occurrence of HCT is correlated with the presence of lignin in all vascular plants. *Hct* gene is therefore a key target for altering the phenylpropanoid pathway in lignified plants.

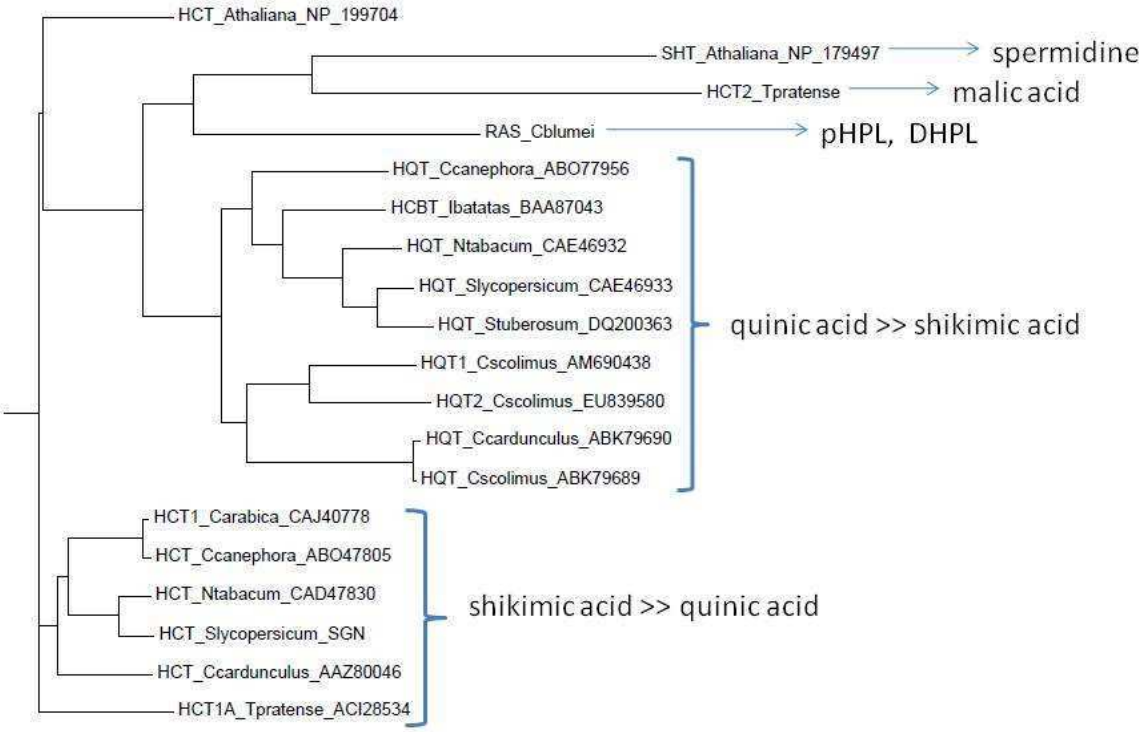
1.6.2.4. Substrate specificity and sequence comparisons

Hydroxycinnamoyltransferases can generally accommodate various substituted cinnamic acid derivatives, most commonly coumaroyl-, caffeoyl- and feruloyl-CoA, while sinapoyl-CoA is less common. In contrast, they usually show a narrow specificity for the acyl acceptor molecule (Figure 14). Tobacco HCT has been tested with a range of alternative

acceptors (anthranilate, glucose, malate, tyramine, spermidine, spermine, putrescine and agmatine), but is only specific for quinic and shikimic acids (Hoffmann *et al.*, 2003). The globe artichoke HCT has been shown to also accept 3-hydroxyanthranilic acid, but not anthranilic or 5-hydroxyanthranilic acids (Moglia *et al.*, 2010). This study suggests that HCT may be able to transfer HCAs not only to alcohol groups, but to amino moieties as well. Other hydroxycinnamoyltransferases with different acyl acceptor specificities have been isolated from various plant species. These include the bifunctional hydroxycinnamoyl/ benzoyl-CoA anthranilate N-hydroxycinnamoyl/benzoyltransferase (HCBT) involved in phytoalexin biosynthesis from *Dianthus caryophyllus* (Yang *et al.*, 1997) and *Ipomoea batatas* (Villegas *et al.*, 1986). In *Coleus blumei*, a hydroxycinnamoyl-CoA hydroxyphenyllactate hydroxycinnamoyltransferase, better known as rosmarinic acid synthase (RS), has been characterised (Berger *et al.*, 2006; Petersen *et al.*, 2009). Hydroxycinnamoylspermidine synthase (SHT), which is implicated in pollen development, was identified in *A. thaliana* (Grienenberger *et al.*, 2009; Luo *et al.*, 2009). Lastly, a hydroxycinnamoyl-CoA malate hydroxycinnamoyltransferase involved in phaselic acid (caffeoylmalic acid) biosynthesis was isolated from red clover (Sullivan, 2009).

Figure 14: Cladogram of the amino acid sequences of various plant hydroxycinnamoyltransferases

Abbreviations, species and accession numbers are included in the names. Enzyme activities are assigned towards their main acyl acceptor substrates according to published data (Petersen, 2009).
pHPL: 4-hydrophenyllactic acid; DHPL: 3,4-dihydroxyphenyllactic acid.



1.6.2.5. Plant acyltransferase superfamily

In plants, acylation is implicated in the formation and modification of a variety of secondary metabolites that have varied roles in plant metabolism, development and disease resistance. This is particularly true for phenolics, where the addition of acyl substituents can alter the stability and solubility of the compounds or protect them from enzymatic degradation. A large family of enzymes catalysing acyl-CoA-dependent acyltransfer reactions (EC 2.3.1) has been identified (D'Auria, 2006). This superfamily was named BAHD from the first four enzymes isolated: benzyl alcohol *O*-acetyltransferase (BEAT), anthocyanin-*O*-hydroxycinnamoyltransferase (AHCT), anthranilate-*N*-hydroxycinnamoyl/ benzoyltransferase (HCBT) and deacetylvindoline 4-*O*-acetyltransferase (DAT) (St-Pierre *et al.*, 2000). To date, over 50 BAHD members associated with secondary metabolism have been characterised on the basis of genetic mutant screening and biochemical experiments. These acyltransferases catalyse the transfer of either aliphatic or aromatic acyl moieties from a CoA thioester to an oxygen or nitrogen-containing acceptor molecule, thus producing diverse natural compounds such as CGAs, phytoalexins, anthocyanins and volatile esters.

The low sequence identity and the catalytic versatility of the BAHD acyltransferases make functional predictions difficult from their primary sequence alone. Phylogenetic analysis supports the classification of these members into distinct clades (Figure 15), whose members display some similarity in their substrate preferences. HCT and HQT hydroxycinnamoyltransferases belong to clade V, which also contains enzymes involved in volatile and paclitaxel biosynthesis (D'Auria, 2006). In addition to those in higher plants, a few eukaryotic fungal acyltransferases have also been identified (e.g. *Fusarium* trichothecene 3- and 15-*O*-acetyltransferase) and belong to clade VI. Multiple sequence alignments reveal two highly conserved regions in members of the BAHD superfamily: Box I (HX_3D) and Box II (DFGWG) (Figure 16).

Figure 15: Cladogram of functionally characterised BAHD members

Protein names are abbreviated and shown with their origin species and accession numbers.

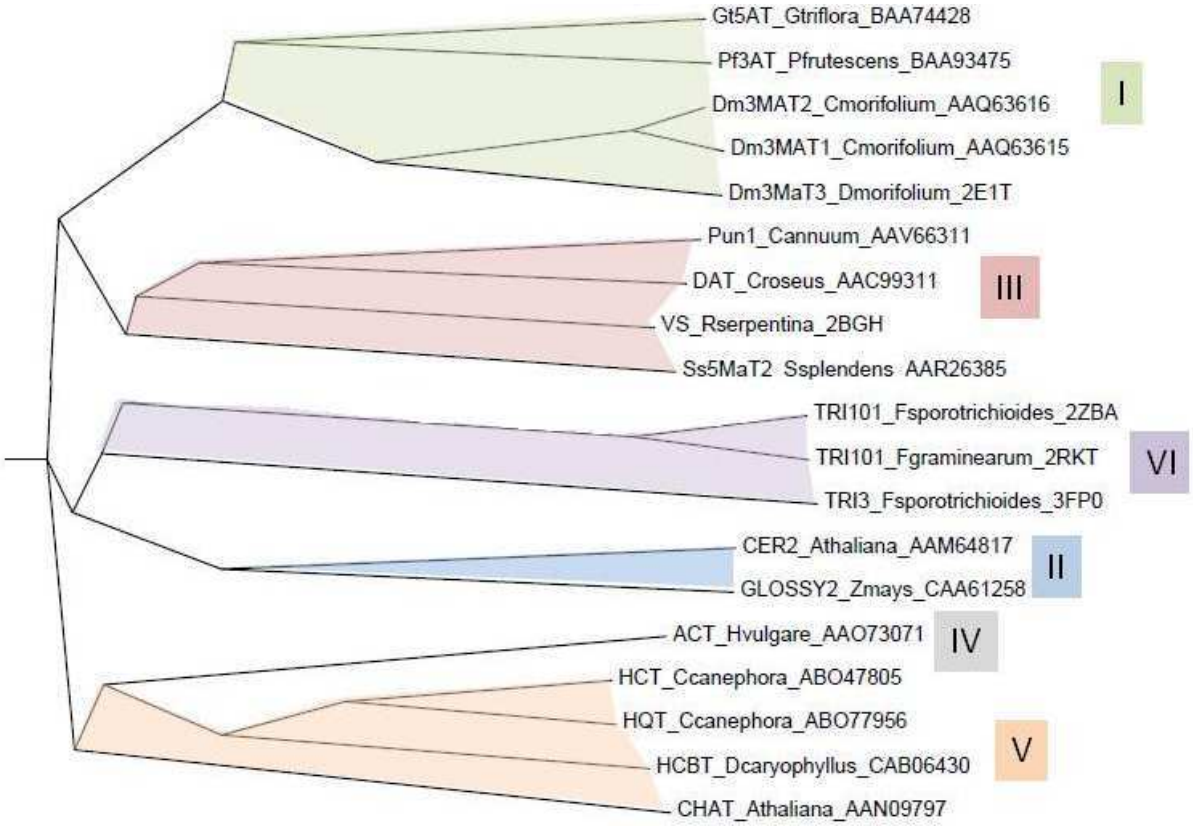
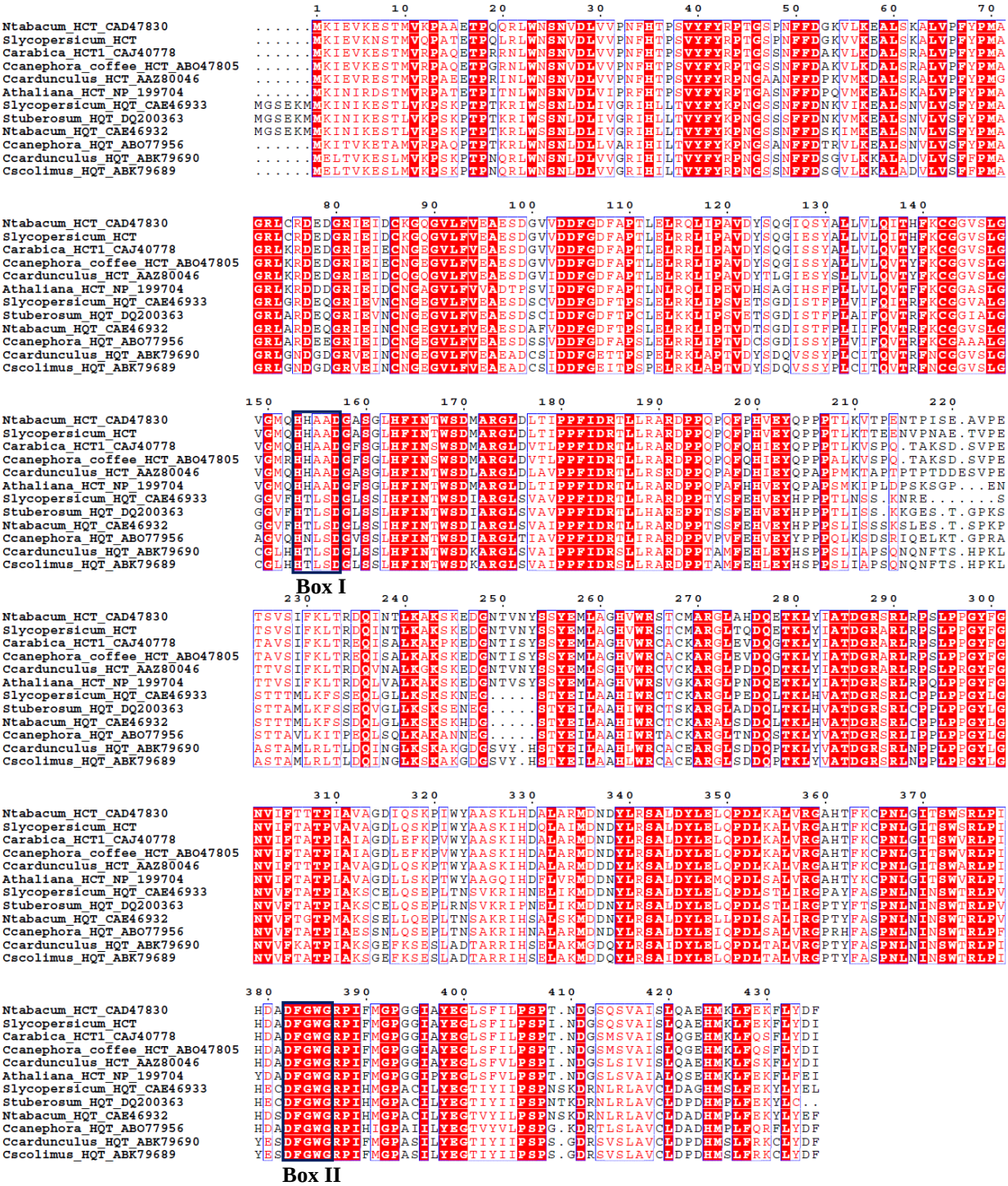


Figure 16: Multiple sequence alignments of HCTs and HQTs

Protein names are abbreviated and shown with their origin species and accession numbers. Black boxes highlight the conserved motifs (Box I and II), which are characteristic of the BAHD superfamily.



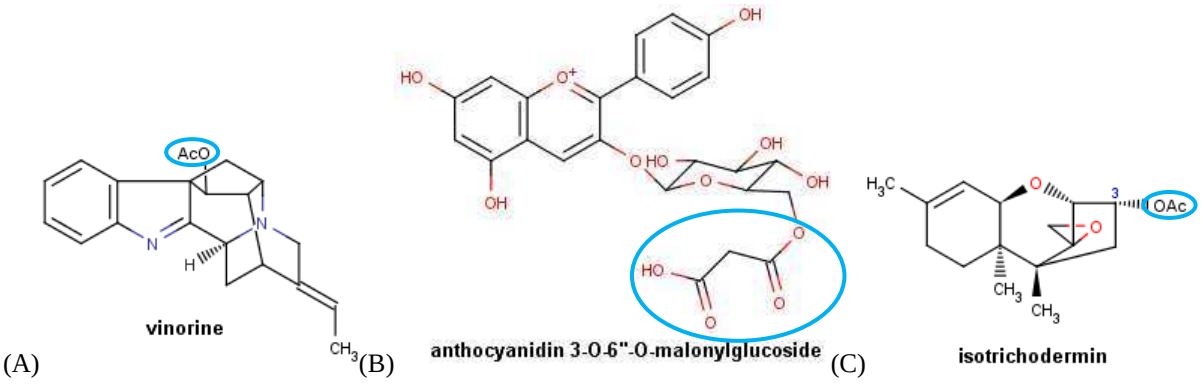
1.6.2.6. Representative crystal structures of BAHD acyltransferases

The first crystal structure of a BAHD member to be solved was vinorine synthase (VS) (Ma *et al.*, 2005). Vinorine synthase (EC 2.3.1.160) is involved in the biosynthesis of ajmaline, an antiarrhythmic alkaloid drug from the Indian plant *Rauvolfia serpentina* (Ma *et al.*, 2004; Ma *et al.*, 2005). The crystal structure of *Dendranthema morifolium* anthocyanin malonyltransferase *Dm*3MaT3 (referred to MaT in the rest of the document) was later solved in complex with malonyl-CoA (Unno *et al.*, 2007). This aliphatic anthocyanin acyltransferase (EC 2.3.1.172) specifically utilises malonyl-CoA for the synthesis of anthocyanins that are responsible for the red colour of chrysanthemum petals. The crystal structure of trichothecene 3-*O*-acetyltransferase (TRI101) from *Fusarium sporotrichioides* and *Fusarium graminearum* was subsequently determined in complex with ethyl-CoA, with CoA and deoxynivalenol (DON), and with CoA and T-2 mycotoxin (Garvey *et al.*, 2008). TRI101 catalyses the transfer of the acetyl group from acetyl-CoA to the C-3 hydroxyl of several trichothecene mycotoxins as part of a detoxification process. More recently, the crystal structure of 15-*O*-trichothecene acetyltransferase (TRI3) from *Fusarium sporotrichioides* was solved in complex with 15-decalonectrin (Garvey *et al.*, 2009).

VS, MaT and TRI101 share respectively 20, 18 and 13 % sequence identity at the amino acid level with *C. canephora* HCT. VS, MaT and TRI101 catalyse the transfer of aliphatic moieties such as acetyl and malonyl groups onto different acceptor substrates (Figure 17). They belong to different clades of the BAHD superfamily: the malonyltransferase clade I for MaT; the clade III for VS; and the clade VI for TRI101 and TRI3 (Figure 15).

Figure 17: Chemical structures of the main products synthesised by VS, MaT and TRI101

The acyl moieties, acetyl for VS (A) and TRI101 (C), and malonyl-CoA for MaT (B), are highlighted in blue.

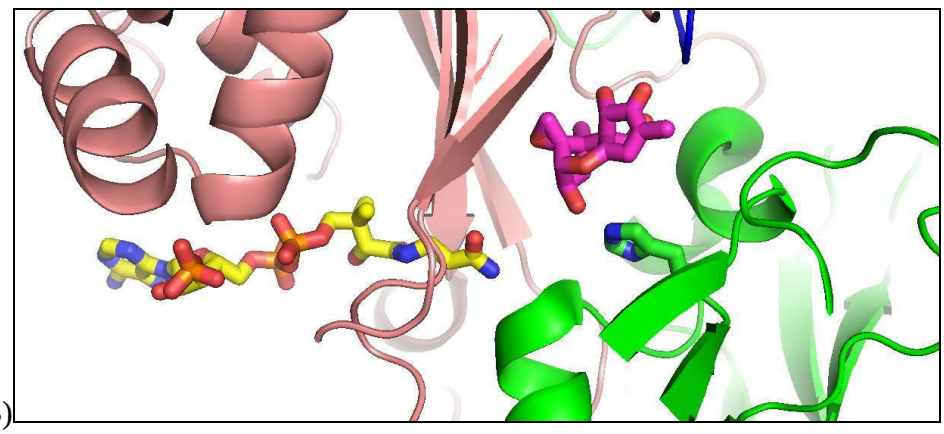


The tertiary structures of BAHDs consists of two nearly equal-sized domains connected by a large cross-over loop (Figure 18A). In the crystal structures solved, the conserved HX_3D motif (Box I in Figure 16) that has been shown to be important for catalysis (Ma *et al.*, 2005; Unno *et al.*, 2007; Garvey *et al.*, 2008) contains a catalytic histidine, which is accessible from both sides of the solvent channel at the domain interface (Figure 18B). The DFGWG motif (Box II in Figure 16) located near the carboxyl terminus plays a structural rather than a catalytic role (D'Auria, 2006). The crystal structures of the ternary complexes of TRI101 (Garvey *et al.*, 2008) reveal the CoA and acceptor substrate binding pockets, accessible from the front and back faces of the protein respectively (Figure 18B).

Figure 18: Crystal structure of *Fg*TRI101 (PDB: 3B2S) in a ternary complex with CoA and DON

(A) The N- (2-214) and C- (227-451) terminal domains are coloured green and salmon respectively and the flexible loop Pro215-Pro226 is in blue. (B) Active site region. Bound CoA (yellow) and DON (magenta), and the catalytic His156 (green C atoms) are represented as sticks.

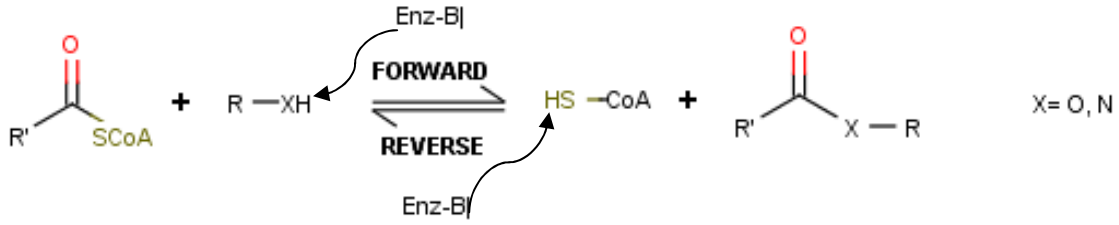
(A)



(B)

The acyltransfer reaction proceeds via a general base mechanism, where the enzyme, the acyl donor and the acyl acceptor form a ternary complex before catalysis can occur (Garvey *et al.*, 2008). The proposed mechanism is as follows: the catalytic His residue of the enzyme deprotonates a hydroxyl/ amine group of the acyl acceptor, promoting a nucleophilic attack on the carbonyl of the acyl-CoA thioester, which forms an ester/ amide product (Figure 19, forward direction). BAHD acyltransferases can also catalyse the transfer of the acyl moiety of back to CoA (Figure 19, reverse direction).

Figure 19: General base mechanism of the acyltransfer reaction catalysed by BAHD enzymes



1.6.2.7. Structural relationships between BAHD enzymes and other CoA-dependent acyltransferases

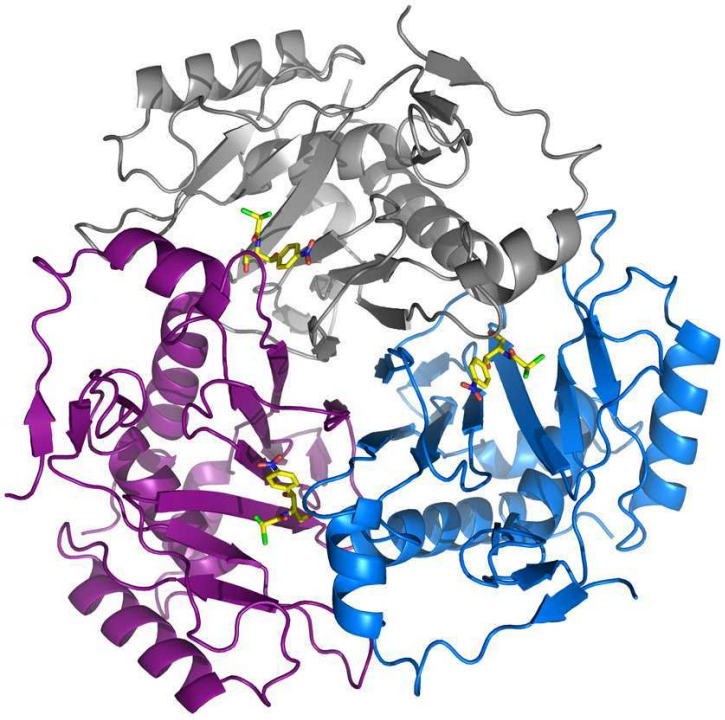
The BAHD superfamily shares some similarities with the chloramphenicol acetyl-CoA transferase (CAT) family. *Escherichia coli* (*E. coli*) CAT was the first CoA-dependent acyltransferase to be structurally determined (Leslie, 1988). This enzyme is responsible for high-level bacterial resistance to the antibiotic chloramphenicol. CAT catalyses the transfer of an acetyl group from acetyl-CoA to the primary hydroxyl (3-*O*) of chloramphenicol. CAT is trimeric and has a distinctive protein fold stabilised by a β-pleated sheet extending from one monomer to the next (Figure 20A). The active site is located at the interfaces between two subunits and accommodates both CoA and chloramphenicol via clefts on opposite sides. The majority of the residues forming the chloramphenicol-binding site belong to one subunit, while the catalytic His is provided by the adjacent monomer.

Other CoA-dependent acyltransferases are monomeric. These include polyketide-associated protein A5 (PapA5) (Buglino *et al.*, 2004) and carnitine acetyltransferase (CrAT), which have a two-domain structure sharing an unexpected structural similarity to CAT (Figure 20B). CrAT catalyses the transfer between carnitine and acetyl-CoA (Jogl *et al.*, 2003). The active site of *Mycobacterium tuberculosis* CrAT is located at the domain interface and contains a His residue appropriately positioned to act as a general base in the reaction

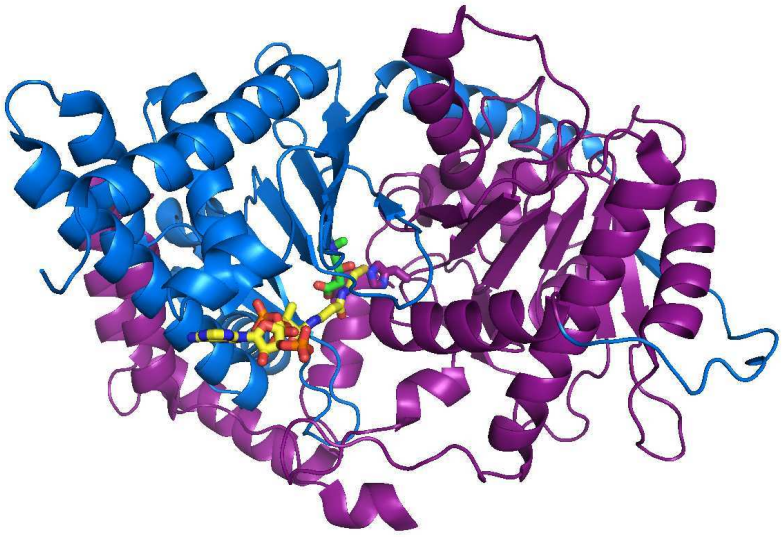
(Figure 20B). This structure and active site arrangement of CAT family members is very similar to that observed in the BAHD superfamily (Figure 18), indicating a possible evolutionary relationship (St-Pierre *et al.*, 2000).

Figure 20: Crystal structures of chloramphenicol acetyltransferase (CAT) and carnitine acetyltransferase (CrAT)

(A) Trimeric structure of CAT (PDB: 3CLA) in complex with chloramphenicol (yellow sticks). (B) CrAT bound to CoA and carnitine (PDB: 1NDF) is shown in the same orientation as TRI101-CoA-DON in Figure 18. The N- (1-363) and C- (364-625) terminal domains are coloured in purple and marine respectively. (C) Active site of CrAT with bound CoA (yellow) and carnitine (green). His343 is shown as purple sticks.

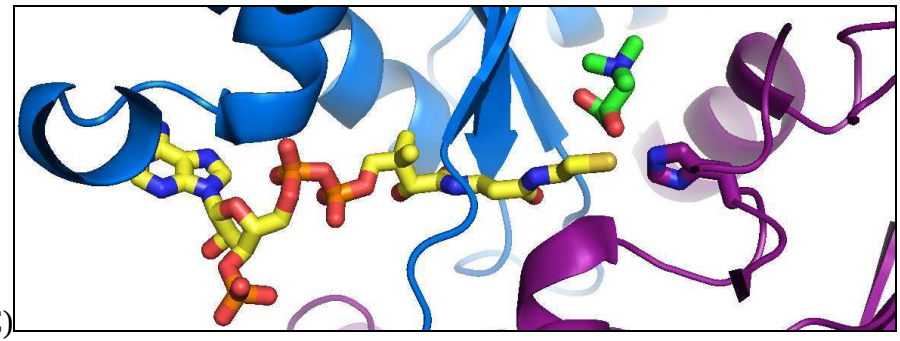


(A)



(B)

Figure 20 (continued)

(C) 

Chapter 2. Materials and methods

2.1. Cloning, over-expression and purification

2.1.1. Preliminary data

A protein that is difficult to isolate in sufficient quantities from its natural source can be produced using recombinant expression in bacterial, yeast, insect or mammalian cells (Baneyx, 1999; Kost *et al.*, 1999; Malissard *et al.*, 1999; Hunt, 2005; Baldi *et al.*, 2007). Expression in *E. coli* is the easiest, quickest and cheapest method currently available. However, many eukaryotic proteins do not fold properly in prokaryotic systems and form insoluble aggregates or inclusion bodies. Sometimes it is possible to improve the solubility of such proteins by decreasing expression temperature, using a fusion partner/ polypeptide tag (Esposito *et al.*, 2006) or, as a last resort, refolding (Li *et al.*, 2004). Taking advantage of cleavable polypeptide/ protein tags, simplified affinity purification procedures can then be applied to obtained sufficient amounts of pure protein for subsequent characterisation.

Cornell University and Nestlé S.A. released 47,000 expressed sequence tag (EST) sequences from *Coffea canephora* (Lin *et al.*, 2005). The EST databank (http://www.sgn.cornell.edu) was used for a BLAST search (Basic Alignment Search Tool, http://www.ncbi.nlm.nih.gov/blast/) against *Nicotiana tabacum hct* (AJ507825) and *hqt* (AJ582651) gene sequences (Hoffmann *et al.*, 2003; Hoffmann *et al.*, 2004). This search resulted in two hits: the unigenes CGN-U123197 and CGN-U125212, corresponding to the partial open reading frames (ORF) for putative *Coffea canephora* hydroxycinnamoyltransferases. The full length cDNAs were isolated by 5'-RACE-PCR for rapid amplification of cDNA ends followed by the cloning of *Cchct* and *Cchqt* genes (Lepelley *et al.*, 2007). A multiple sequence alignment was generated with ClustalW (Larkin *et al.*, 2007) using acyltransferase nucleotide sequences from different plant species. Sequence comparisons confirmed that one of the cloned sequences lies on a branch containing *hqt* gene sequences, the other in a branch clustering *hct* gene sequences from different plant species (Figure 14). *Cchct* and *Cchqt* exhibit 66 % sequence identity at the nucleotide level. The translated sequences contain both the HX_3D and DFGWG motifs, which are characteristic of BAHD acyltransferase amino acid sequences.

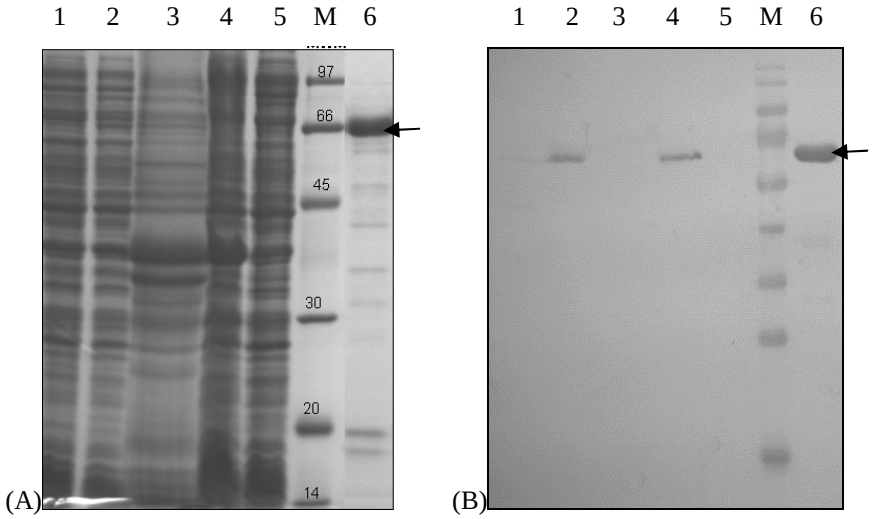
2.1.2. Production of recombinant 4CL2

2.1.2.1. Coffea canephora 4-coumarate CoA ligase isoform 2

A plasmid provided by Nestlé contained the *Cc4cl2* gene (1626 bp) inserted into the inverted pET28a(+) vector with BamHI (5') and HindIII (3') restriction sites. This plasmid allows for expression of *Cc*4CL2 with a N-terminal hexahistidine (His_6) tag. BL21 (DE3) cells were transformed with the plasmid for standard expression tests at 37 and 20 °C. A band corresponding to the expected size of His_6-*Cc*4CL2 with a 24 amino acid linker (62.6 kDa) was detected in the cell extracts using SDS-PAGE and western-blot analysis against the polyhistidine tag (Figure 21).

Figure 21: SDS-PAGE (A) and western-blot (B) analysis of *Cc*4CL2 following over-expression and purification

(A) Lane 1: not induced; lane 2: induced; lane 3: insoluble cell extract; lane 4: soluble cell extract; lane 5: flow through; lane 6: elution fraction; M: MW marker (kDa).



Expression was scaled-up in flasks containing 1 L of Luria-Bertani (LB) medium supplemented with 30 µg/mL kanamycin, which were inoculated with a 1/100 (v/v) overnight starter culture. Cells were grown at 37 °C until the optical density (OD) at λ= 600 nm reached 0.6. After cooling to room temperature, over-expression was induced with 1 mM IPTG and cells were grown overnight at 20 °C. The cells were harvested by centrifugation at 5,000 x g and resuspended in a solution composed of 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 2.5 mM $MgCl_2$, 10 % glycerol and 5 mM β-mercaptoethanol (BME) (buffer A). The cells then were flash-frozen and stored at -80 °C until purification. EDTA-free protease inhibitors tablets, lysozyme and DNaseI were added prior to cell lysis using a French press (2 cycles at 10 kPSI). The lysate was centrifuged for 30 min at 50,000 x g and 4 °C and the supernatant

loaded on to a 5 mL HisTrap column (GE Healthcare) using an Äkta Prime system. After washing the column with 10 CV of buffer A and 5 CV of 5 % buffer B (buffer A supplemented with 500 mM imidazole), an elution gradient of 5-100 % buffer B was applied. 1 mL fractions were collected and analysed on 12 % SDS-PAGE gel stained with Coomassie blue. A SDS-PAGE band corresponding to the size of His_6-*Cc*4CL2 was observed in the elution fractions collected from the affinity column (Figure 21A). Over-expression of the recombinant protein was confirmed by western-blot analysis against the hexahistidine tag. The fractions containing the protein were pooled and dialysed against 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2.5 mM $MgCl_2$ and 1 mM DTT. A tag cleavage test was carried out with thrombin protease (Sigma) after dialysing the protein into a buffer composed of 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 % glycerol, 2.5 mM $MgCl_2$, 2.5 mM $CaCl_2$ and 5 mM BME. Cleavage was successful at both room temperature and at 4 °C, but some precipitation occurred. The cleaved *Cc*4CL2 was recovered with subtractive affinity chromatography. Dynamic Light Scattering (DLS) analysis performed at 20 °C indicated that coffee 4CL2 with or without the His_6-tag was highly polydisperse. The over-expression procedure was repeated with Rosetta 2 (DE3) cells (Novagen) in order to overcome the codon bias for expression in *E. coli*. However, the resulting protein was still aggregated.

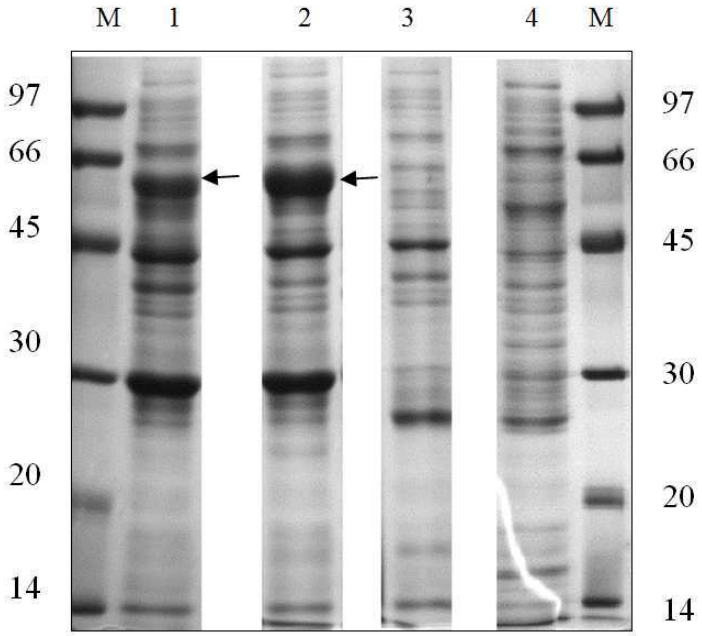
2.1.2.2. Nicotiana tobacco 4-coumarate CoA ligase isoform 2

Because *Cc*4CL2 required more exploratory work to obtain a fully soluble, monodisperse protein, a synthetic gene encoding *Nicotiana tabacum* 4CL2 (GenBank accession number: U50846) was purchased from Geneart (Regensburg, Germany). This protein had previously been used for the enzymatic synthesis of various hydroxycinnamoyl-CoA thioesters (Beuerle *et al.*, 2002). The codon-optimised gene (1653 bp) was subcloned into the pET21d vector (Novagen) digested with BamHI (5') and XhoI (3') for expression with an uncleavable C-terminal His_6-tag and a two-amino acid (Lys-Glu) linker (60.5 kDa).

Several *E. coli* strains, namely BL21 (DE3), BL21* (DE3), BL21 (DE3) pLysS and BL21* (DE3) pLysS, were transformed with the pET21d_*Nt4cl2* plasmid for small-scale expression tests. The soluble extracts were analysed by SDS-PAGE (Figure 22). The highest soluble expression level was obtained with BL21* (DE3) cells, which were subsequently used for large scale production and purification of the recombinant *Nt*4CL2-His_6. Here, 4 L of LB medium containing 100 mg/L ampicillin were inoculated with a 1/100 (v/v) overnight starter culture and grown at 37 °C until the OD at λ= 600 nm reached 0.6. The temperature was lowered to 20 °C before inducing the expression with 1 mM IPTG, and then the cells were

incubated overnight. After centrifugation for 15 min at 5,000 x g, the supernatant was discarded and the bacterial pellet resuspended on ice in a lysis buffer composed of 50 mM bis-Tris pH 7.0, 10 % glycerol and 10 mM NaCl according to the previously described procedure (Beuerle *et al.*, 2002). In contrast with the authors however, a later comparison with a standard buffering agent (50 mM Tris-HCl pH 8.0) did not show any difference in the protein stability, so this standard buffer was subsequently used.

Figure 22: SDS-PAGE analysis of the soluble fraction of the cell lysates following *Nt*4CL2 over-expression
Lane 1: BL21 (DE3); lane 2: BL21* (DE3); lane 3: BL21 (DE3) pLysS; lane 4: BL21* (DE3) pLysS; M: MW marker (kDa).

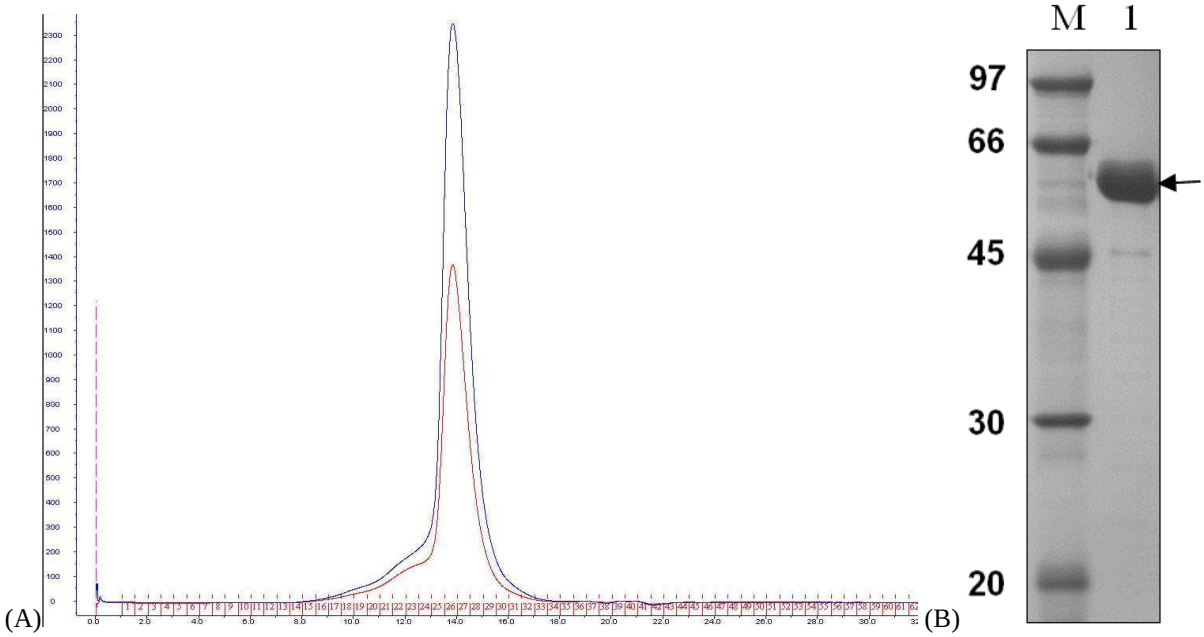


Cell lysis was achieved using one cycle at 2.0 kPSI in the One Shot cell disrupter (Constant Systems). The lysate was centrifuged for 30 min at 50,000 x g and 4 °C. The supernatant was then loaded on a 2 mL nickel-nitriloacetic acid (NTA) gravity column (Qiagen). After washing with 10 CV of lysis buffer (see above) and 10 CV of 20 mM Tris-HCl pH 8.0 and 500 mM NaCl, the protein was eluted with 20 mM Tris-HCl pH 8.0, 500 mM NaCl and 500 mM imidazole. The fractions containing the target protein were pooled and dialysed overnight against 20 mM Tris-HCl pH 8.0 and 50 mM NaCl. The protein was further purified by anion-exchange chromatography (MonoQ GL 5/50, GE Healthcare) using a gradient from 50 mM to 1 M NaCl on a Äkta Explorer system at 4 °C. The fractions containing the recombinant protein were pooled, concentrated and loaded on a HiLoad 16/60 Superdex 200 (GE Healthcare) for size-exclusion chromatography in 20 mM Tris-HCl pH 7.5

and 50 mM NaCl (Figure 23A). The purest fractions from the size-exclusion chromatography were pooled and concentrated to 15 mg/mL in 10 mM Tris-HCl pH 7.5 and 50 mM NaCl (Figure 23B). The protein was stable in a limited proteolysis experiment on ice with trypsin 1/1000 (w/w) for at least 1 h. Aliquots were flash-frozen and stored at -80 °C for biochemical assays (2.3.3.1). The remaining protein solution was incubated with 2 mM ATP, 2 mM $MgCl_2$, 1 mM coumaric acid and 1 mM CoA for subsequent crystallisation trials (2.4.2.1).

Figure 23: SDS-PAGE analysis of *Nt*4CL2 following purification by size-exclusion chromatography

(A) Chromatogram. (B) 12 % SDS-PAGE lane 1: purified *Nt*4CL2; M: MW marker (kDa).



2.1.3. Production of recombinant HCT and HQT

2.1.3.1. HCT and HQT GST-fusion as a starting material

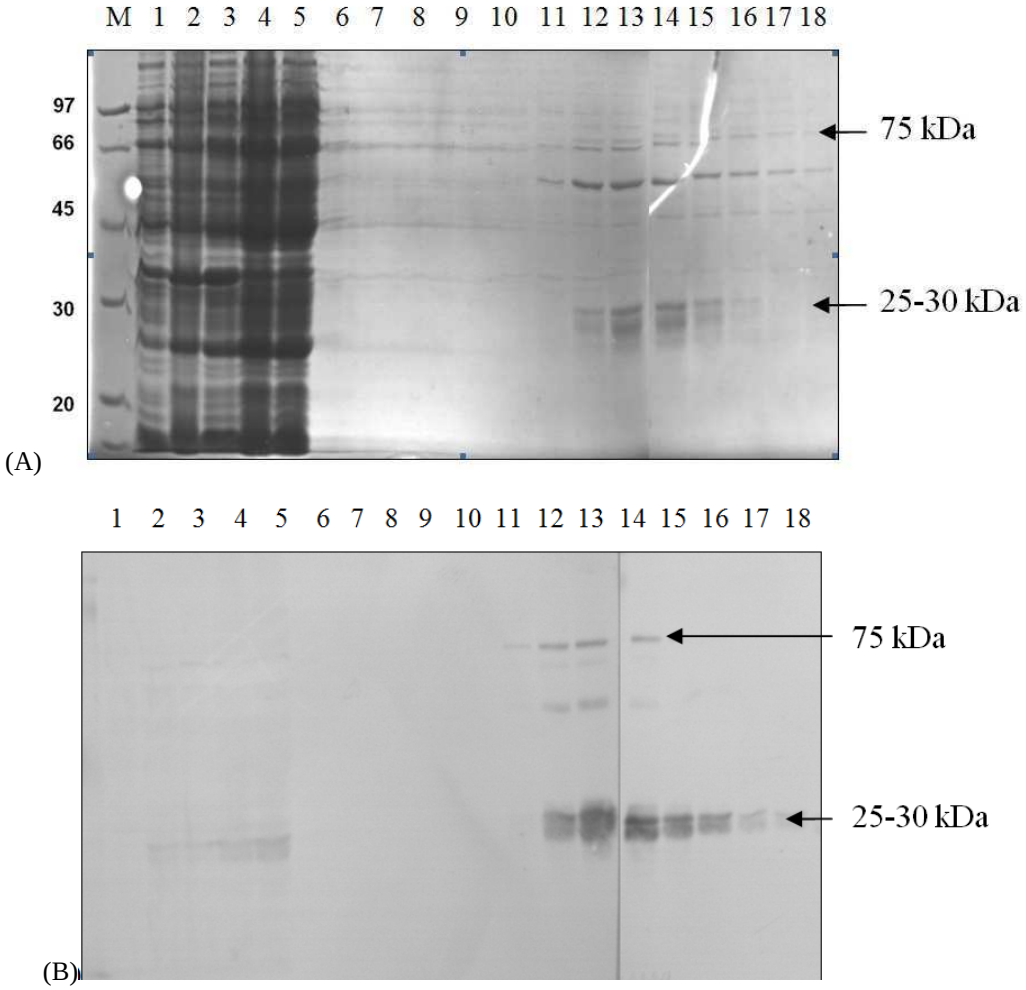
The cDNAs encoding *Coffea canephora* HCT and HQT (GenBank accession numbers: EF137954 and EF153931 respectively) were provided by Nestlé R&D centre in Tours. They were inserted in the pGTPc103a vector (GTP Technologies), which produces a recombinant protein in fusion with N-terminal glutathione S-transferase (GST). BL21 (DE3) cells were transformed with pGTPc103a_*Cchct* and pGTPc103a_*Cchqt* plasmids and plated on LB-agar medium containing 30 µg/mL kanamycin. Cultures of 100 mL LB medium were grown until the OD at λ= 600 nm reached 0.6 and expression was induced with 1 mM IPTG. Cells were grown for 3 h at 37 °C or overnight at 20 °C. No band corresponding to the size of the full-length GST-HCT and GST-HQT proteins (75.7 and 75.3 kDa respectively) could be

detected on a 12 % SDS-PAGE gel stained with Coomassie blue, where the non-induced and induced cell fractions were loaded.

Like most eukaryotic genes, coffee *hct* and *hqt* display an unfavourable codon distribution in *E. coli* that can limit expression yields (Kane, 1995). Consequently, expression in a strain that co-expresses the transfer RNAs for these rare codons was considered. Rosetta 2(DE3) cells (Novagen) were transformed with each plasmid and plated on LB-agar medium containing 30 µg/mL kanamycin and 34 µg/mL chloramphenicol. The full-length GST-HCT and GST-HQT proteins could not be detected on SDS-PAGE, while only a faint band was detected on western-blot together with N-terminal fragments of 25-30 kDa. An affinity chromatography purification step was carried out for GST-HQT over-expressed in Rosetta 2(DE3) at 37 °C for 3 h. The bacterial pellet was resuspended in phosphate buffered saline (PBS) buffer pH 7.4 containing 1 mM DTT. The cells were subjected to three cycles of freezing in liquid nitrogen and thawing in a 25 °C water-bath. The cell lysate was centrifuged for 30 min at 50,000 x g and 4 °C. The supernatant was loaded onto a 5 mL Glutathione Sepharose 4 Fast Flow column (GE Healthcare) equilibrated with PBS buffer pH 7.4. After washing with 5 CV of buffer, the protein was eluted in 50 mM Tris-HCl pH 8.0, 10 mM reduced glutathione and 1 mM DTT. The fractions were analysed by SDS-PAGE with Coomassie blue staining. The elution fractions contained several bands, among which was a band at 75 kDa, corresponding to the expected size of GST-HQT. Western-blot analysis confirmed the production of GST-HQT by detecting the 75 kDa polypeptide, as well as a band at 60 kDa and smeared bands around 25-30 kDa, probably corresponding to truncated N-terminal fragments including the GST fusion (Figure 24).

Figure 24: Analysis of GST-HQT following expression and affinity chromatography

(A) 12 % SDS-PAGE. (B) Western-blot against the GST-fusion. M: MW marker (kDa); lanes 1-5: not induced, induced, lysate, soluble extracts 1 and 2; lanes 5-18: wash and elution fractions from the affinity column.



2.1.3.2. HCT and HQT sub-cloning for expression with a hexahistidine-tag

To try to improve the expression levels of the full-length proteins, both *Cchct* and *Cchqt* genes were sub-cloned into the pPROEX_HTb vector (Life Technologies) for expression with a cleavable His_6-tag at the N-terminus. The target nucleotides were amplified by polymerase chain reaction (PCR) using specifically designed oligonucleotide primers flanked with the appropriate cloning sites (Table 2). The reaction mixtures were made in a final volume of 50 µL as follows: 1 µL of plasmid DNA template, 5 µL of PCR buffer (10x), 1 µL of forward and reverse primers (10 pmol/µL), 200 µL dNTP mix (10 mM) and 0.3 µL (2.5 u) of High Fidelity PCR enzyme mix (Fermentas) or Pfu DNA polymerase (Stratagene). A standard PCR program was run with an annealing temperature of 55 °C and an elongation time of 4 min. The PCR products were analysed using 1 % agarose gel electrophoresis stained with SYBR Safe (Figure 25). The PCR products, obtained with Pfu DNA polymerase,

were double digested with the appropriate restriction enzymes and purified by gel extraction (Qiaquick kit, Qiagen). The DNA fragments were ligated into the linearised and dephosphorylated pPROEX_HTb vector. TOP10 cells (Invitrogen) were transformed for plasmid DNA preparation and the insert sequence was verified by DNA sequencing (Genecore, Germany).

Figure 25: Agarose gel analysis of PCR products from *Cchct* and *Cchqt* amplification

(A) Amplification of *Cchct*. (B) Amplification of *Cchqt*. Lane 1, 2, 3, 4: High Fidelity enzyme mix; lane 5, 6, 7, 8: Pfu polymerase; M: MW marker.

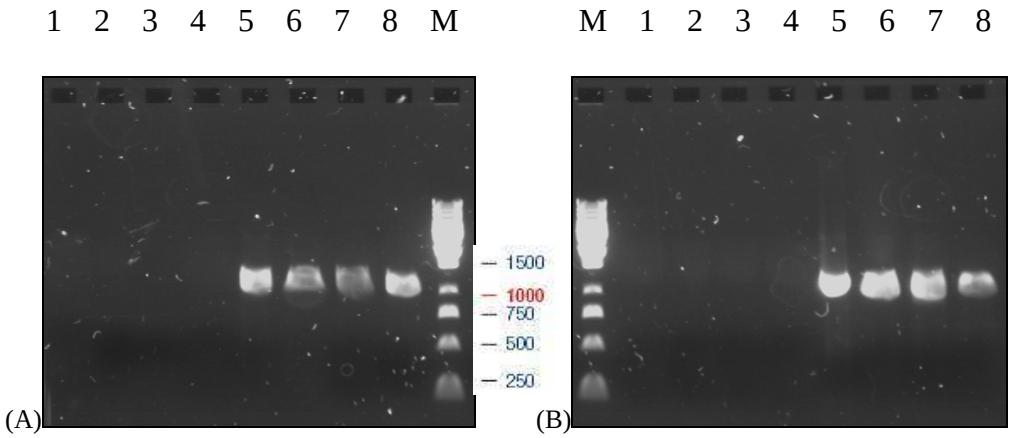


Table 2: Oligonucleotide primers used to amplify *Cchct* and *Cchqt*

The recognition site of the restriction enzyme (in brackets) is underlined in the primer sequence.

Amplicon	*Cchct*
Template	pGTPc103a_*Cchct*
Primer 5'	cagcactagggatccatgaaaatcgaggtgaaggaatcg (BamHI)
Primer 3'	tcgtcgatcactcgagtcaaatgtcatacaagaaactctggaa (XhoI)
Amplicon	*Cchqt*
Template	pGTPc103a_*Cchqt*
Primer 5'	cagcactagggatccatgaagataaccgtgaaggaaaca (BamHI)
Primer 3'	tcgtcgatcatctagatcagaaatcgtacaggaacctttg (XbaI)

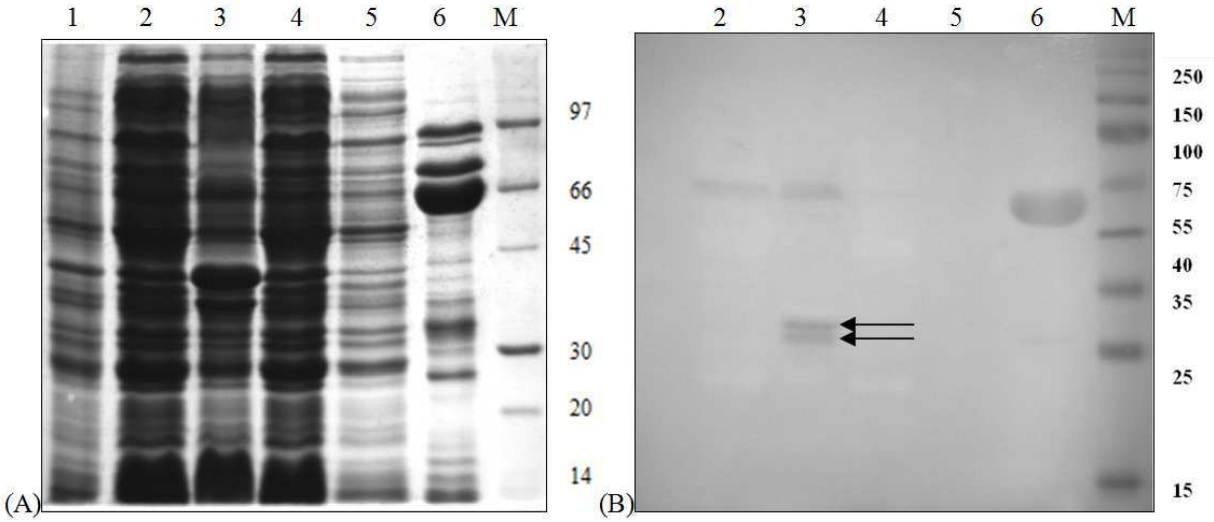
2.1.3.3. Expression and buffer tests for hexahistidine-tagged HCT

Several *E. coli* strains were transformed with the new pPROEX_HTb_*hct* construct and small-scale expression tests performed at 37 °C and 20 °C. The bacterial pellets were resuspended in BugBuster protein extraction reagent (Novagen) for chemical lysis. The soluble cell extracts were loaded onto single-use His-Trap Spin columns (GE Healthcare)

containing 100 µL Ni-NTA agarose beads for affinity purification. The fractions were eluted with buffer containing 500 mM imidazole and analysed by SDS-PAGE and western-blot against the His_6-tag. HCT expression levels were highest in BL21* (DE3) pLysS cells grown overnight at 20 °C. Two truncated N-terminal fragments (~25-30 kDa) remained in the insoluble fraction of the cell lysate (Figure 26).

Figure 26: SDS-PAGE and western-blot of His_6-*Cc*HCT over-expression in BL21* (DE3) pLysS cells

(A) 12 % SDS-PAGE. (B) Western-blot against the His_6-tag. Lane 1: non-induced cells; 2: soluble cell extract; 3: insoluble cell extract; 4: flow through; 5: wash; 6: elution; M: MW marker.



In order to optimise the purification conditions for HCT, 100 mL LB cultures of BL21* (DE3) pLysS were prepared. Expression was induced when the OD reached 0.6/0.9 with 0.5/1 mM IPTG and cells incubated overnight at 20 °C. The pellets were resuspended in different lysis buffers (50 mM Tris-HCl pH 7.4, 300/500 mM NaCl, 20 mM imidazole, 2 mM BME, 0/10 % glycerol, 0/1 % Tween 20) supplemented with lysozyme and DNaseI. The cells were lysed with 4 cycles of freeze and thaw using liquid nitrogen and a water-bath at 25 °C. A similar purification step using His-Trap Spin columns was then performed. No major difference was detected for induction at 0.5 or 1 mM IPTG and at OD at λ= 600 nm of 0.6 or 0.9. The best buffer composition was 50 mM Tris-HCl pH 7.4, 500 mM NaCl, 20 mM imidazole, 2 mM BME and 10 % glycerol.

2.1.3.4. HCT codon-optimised genes for expression in E. coli

A synthetic cDNA with improved codon usage was later obtained from Geneart (Regensburg, Germany) in order to allow a high and stable expression rate of the heterologous protein. The optimised genes for *Cchct* and *Cchqt* were cloned in the pPROEX_HTb

expression vector. Similar protein expression levels were obtained with the wild-type and the codon-optimised *hct* genes. The synthetic gene was used in all the following experiments.

2.1.3.5. Large scale expression and purification of HCT

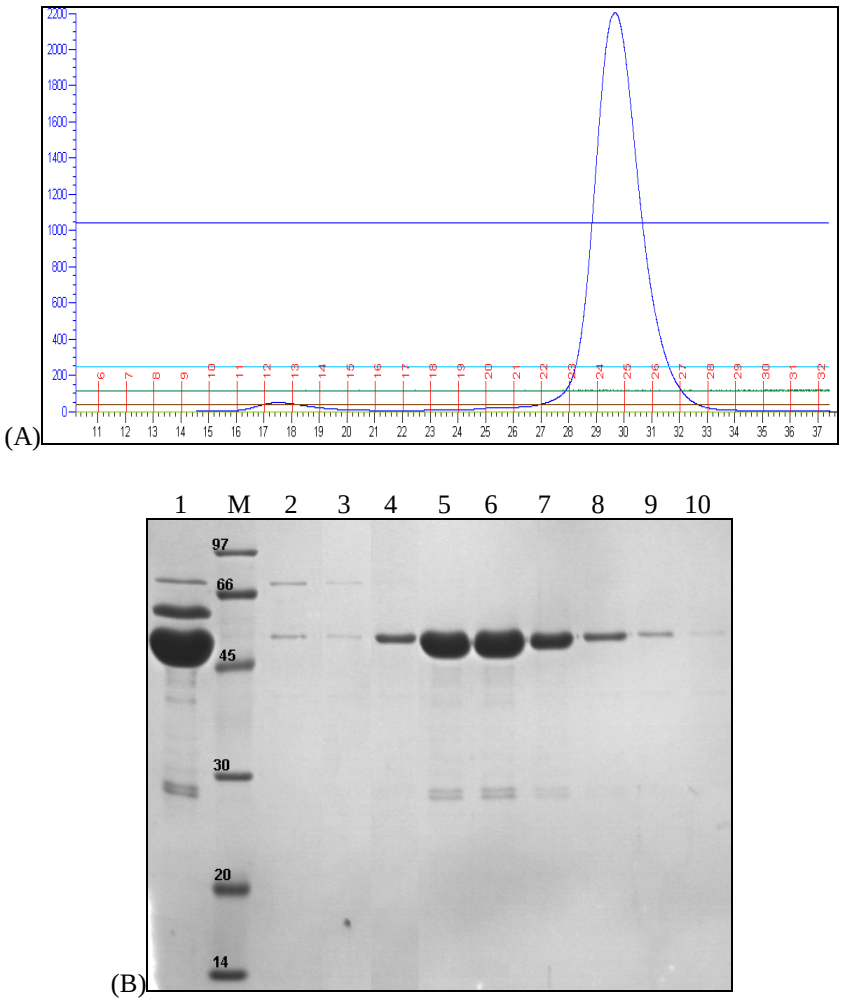
The synthetic pPROEX_HTb_*hct* construct was used to over-express *Cc*HCT with an N-terminal His_6-tag in BL21* (DE3) pLysS strain of *E. coli*. A single isolated bacterial colony grown on a freshly streaked plate of LB-agar medium containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol was used to inoculate a starter culture for overnight growth at 37 °C. Cultures of 1 L LB medium were then inoculated with a 1/100 (v/v) starter culture and incubated at 37 °C until the OD at λ= 600 nm reached 0.6. The flasks were cooled in ice-water and 1 mM IPTG was added for overnight expression at 20 °C. The cultures were centrifuged for 20 min at 5,000 x g and 4 °C and the bacterial pellet produced from a 1 L culture was resuspended in 25 mL buffer composed of 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 10 % glycerol, 20 mM imidazole and 5 mM BME (buffer A). The cell stock was flash-frozen and stored at -80 °C.

All subsequent operations were carried out at 4 °C. Prior to cell lysis, lysozyme (~200 µg/mL), DNaseI (~20 µg/mL) and EDTA-free Complete Protease Inhibitor Cocktail Tablets (Roche) were added to the thawing cells and stirred for 15 min. The resuspended cells were ruptured by passage through a French pressure cell (2 cycles at 10,000 PSI). The lysate was then centrifuged at 50,000 x g for 30 min to remove the cell debris. A 5 mL His-Trap HP column (GE Healthcare) was equilibrated with 10 % (v/v) buffer B (buffer A supplemented with 500 mM imidazole). The soluble fraction of the cell lysate was loaded and a gradient from 10 to 100 % (v/v) buffer B imidazole was applied. The His_6-tagged protein eluted with 250 mM imidazole. The fractions containing the recombinant protein were pooled, incubated with 1/100 (w/w) Tobacco Etch Virus (TEV) protease, and dialysed overnight in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 % glycerol, 1 mM DTT and 0.5 mM EDTA. The TEV cleavage leaves only three residues (Gly-Ala-Met) attached to the target protein. Cleaved HCT was recovered by subtractive chromatography using the same nickel column. The protein sample was concentrated to 500 µL and injected on a Superdex 200 10/300 GL size-exclusion column (GE Healthcare) with 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 5 mM BME. HCT eluted at 14.5 mL, which corresponds to the expected size for a monomer of 50 kDa, according to the column calibration curve. The peak fractions were collected and analysed on a 12 % SDS-PAGE with Coomassie blue staining (Figure 27). Only the purest

fractions were pooled, concentrated by centrifugation and flash-frozen at -80 °C. A mean yield of 4.5 mg/L culture was obtained.

Figure 27: Chromatogram and SDS-PAGE analysis of *Cc*HCT from size-exclusion chromatography

(A) Chromatogram for the size-exclusion step. (B) 12 % SDS-PAGE. Lane 1: injected sample, M: MW marker (kDa); lanes 2-10: collected fractions n° 21-29.



2.1.3.6. Degradation and limited proteolysis of HCT

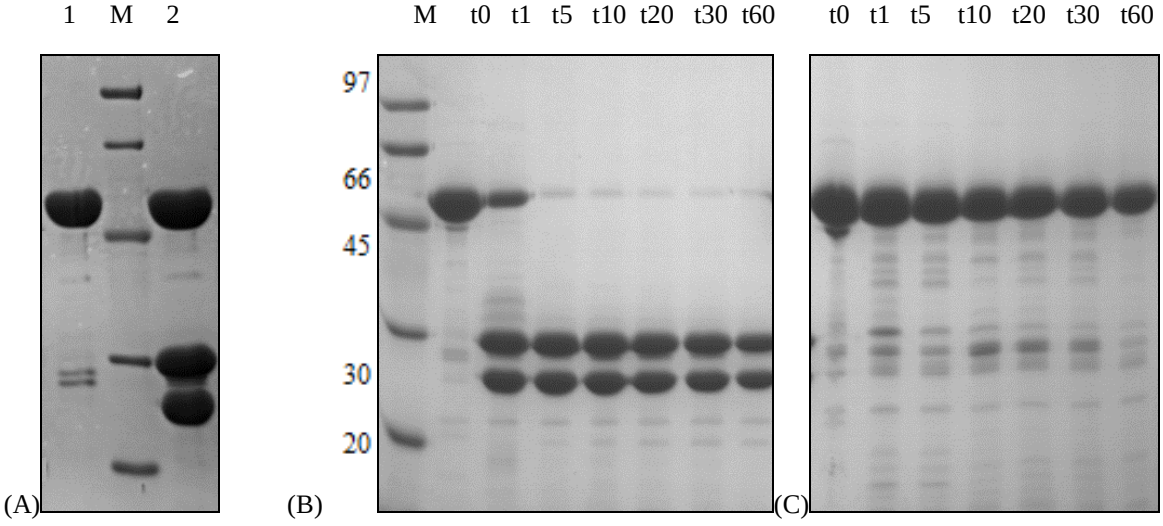
A comparison of protein samples produced from batch to batch and stored at different temperatures (-80, -20 and 4 °C) was carried out. It was observed that the purified protein sample usually contained two additional bands around 25-30 kDa that were amplified with time (Figure 28A). This indicates that the native enzyme is partially degraded into two major fragments by an endogenous endopeptidase activity, probably originating from the crude *E. coli* extract. Size-exclusion chromatography of the degraded sample showed only one symmetric peak of elution corresponding to a full-length protein size (data not shown). This

indicates that the fragments remained physically associated throughout the purification procedure. N-terminal sequencing of *Cc*HCT fragments after storage resulted with a limited confidence level in the sequence DSVPE, which maps on Ser218.

The degradation pattern was reproduced by incubating freshly purified His_6-*Cc*HCT with several proteases. Reaction samples were taken at several time points and supplemented with 1 mM phenylmethanesulfonyl fluoride (PMSF) to stop the reaction. 12 % SDS-PAGE and western-blot (anti-His_5) analysis revealed two bands of approximately 25 kDa appearing after a few minutes incubation with trypsin, thermolysin or subtilisin. The band of the full-length protein was barely visible after 10 min incubation (Figure 28B). The protein was intact in the reactant mixture where no protease was added. The fragment bands were extracted from the polyacrylamide gel and subjected to N-terminal sequencing. This resulted in the identification of the sequence SDSVPETA, which maps after Lys210 that is predicted to be part of the inter-domain loop region.

Figure 28: SDS-PAGE analysis of HCT proteolysis

(A) Lane 1: purified *Cc*HCT stored at -80 °C. M: MW marker (kDa); lane 2: same sample stored at 4 °C for several weeks. (B) Limited proteolysis of with trypsin at a 1/500 (w/w) dilution on ice. M: MW marker; t0-t60: incubation time in min. (C) Limited proteolysis of the methylated derivative of *Cc*HCT.



2.1.3.7. Chemical modification of CcHCT by reductive methylation

The reductive methylation of solvent-accessible lysine residues is a simple and inexpensive method to aid in the crystallisation of some proteins. The protocol used was adapted from (Walter *et al.*, 2006) and applied to *Cc*HCT, which contains 20 potential methylation sites (19 Lys and the N-terminal amino group). The reaction was performed at 4

°C with 30 mg of pure protein at 0.65 mg/mL in 50 mM sodium HEPES pH 7.5, 250 mM NaCl and 5 mM BME. 20 µL of freshly prepared 1 M dimethylamine-borane complex (ABC) (Fluka) and 40 µL of 1 M formaldehyde per mg of protein were added to the protein sample and incubated for 1 h. The procedure was repeated once before leaving overnight under gentle agitation. The precipitated material formed during the methylation reaction was removed prior to concentration of the remaining protein sample (76 %). The buffer was exchanged to 20 mM Tris-HCl pH 7.5, 200 mM NaCl and 5 mM BME for size-exclusion chromatography using a Superdex 200.

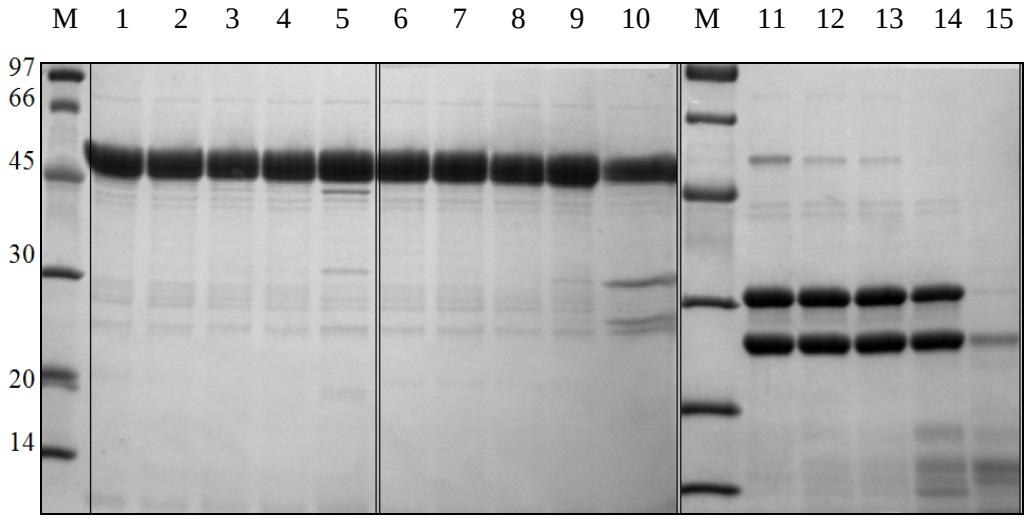
Limited proteolysis was then carried out on the native and methylated HCTs. The proteins were incubated at 30 °C with trypsin at a 1/100 (w/w) ratio in a buffer composed of 20 mM Tris-HCl pH 7.5. The reaction was monitored by taking aliquots from t= 0 to 120 min and stopped by the addition of 1 mM PMSF prior to loading the samples on a 12 % SDS-PAGE gel. As expected, the results showed that, contrary to native *Cc*HCT, the methylated derivative was resistant to trypsin proteolysis (Figure 28C). From these observations and the N-terminal sequencing results (2.1.3.6), it was postulated that the solvent-accessible Lys residues part of the predicted inter-domain loop could have an influence on the stability of *Cc*HCT. Once cleaved, the loose parts may hinder the crystal formation. The two charged amino acids, Lys210 and Lys217, which are characterised by a high conformational energy, were consequently mutated to Ala to engineer a trypsin-resistant protein (2.1.3.8).

2.1.3.8. K210A/K217A (K-mutant) CcHCT

K-mutant *Cc*HCT was produced using expression and purification procedures similar to that for the native protein (2.1.3.5). Limited proteolysis with trypsin and chymotrypsin had no or little effect on the K-mutant HCT, while subtilisin could still cut the protein into two stable fragments (Figure 29).

Figure 29: SDS-PAGE analysis of the K-mutant HCT following limited proteolysis

M: MW marker (kDa), lanes 1-5: trypsin digestion, lanes 6-10: chymotrypsin digestion, lanes 11-15: subtilisin digestion, respectively at t= 1, 5, 15, 60 min and overnight incubation with 1/1000 (w/w) protease at 4 °C.



2.1.3.9. *His$_6$-HQT expression*

A similar expression protocol was applied to pPROEX_HTb_*Cchqt*. No band corresponding to the correct size of the His$_6$-*Cc*HQT could be detected in the elution fractions loaded on polyacrylamide gel. However, multiple bands at lower molecular weight (25-30 kDa) were observed in the insoluble and soluble fractions, similar the expression pattern for GST-HQT (Figure 24). These bands probably correspond to N-terminal truncated fragments of *Cc*HQT. No further work was carried out with this expression vector as no satisfactory expression conditions were found. A synthetic *Cchqt* gene (Geneart) was also tested, but did not result in any improvement in the expression level of the full-length *Cc*HQT.

2.1.3.10. *Cchqt subcloning in pET-28M-SUMO3*

As *Cc*HQT was difficult to over-express with previous systems, the synthetic *Cchqt* gene was sub-cloned into pET-28M-SUMO3 expression vector (EMBL Heidelberg) using the BamHI (5') and XhoI (3') restriction sites. The new construct was designed to enhance the expression level and solubility of *Cc*HQT with a N-terminal His$_6$-SUMO3 (small ubiquitin-related modifier) fusion tag (~100 amino acid residues, 11.5 kDa) that has been shown to provide a chaperoning effect (Malakhov *et al.*, 2004). The fusion tag is cleaved off with 1/100 (w/w) SenP2 protease incubated overnight at 4 °C using the elution fraction from the first affinity chromatography step. The robust and specific SenP2 protease is known to recognise the 3-dimensional structure of human SUMO3 rather than a short, but specific, amino acid

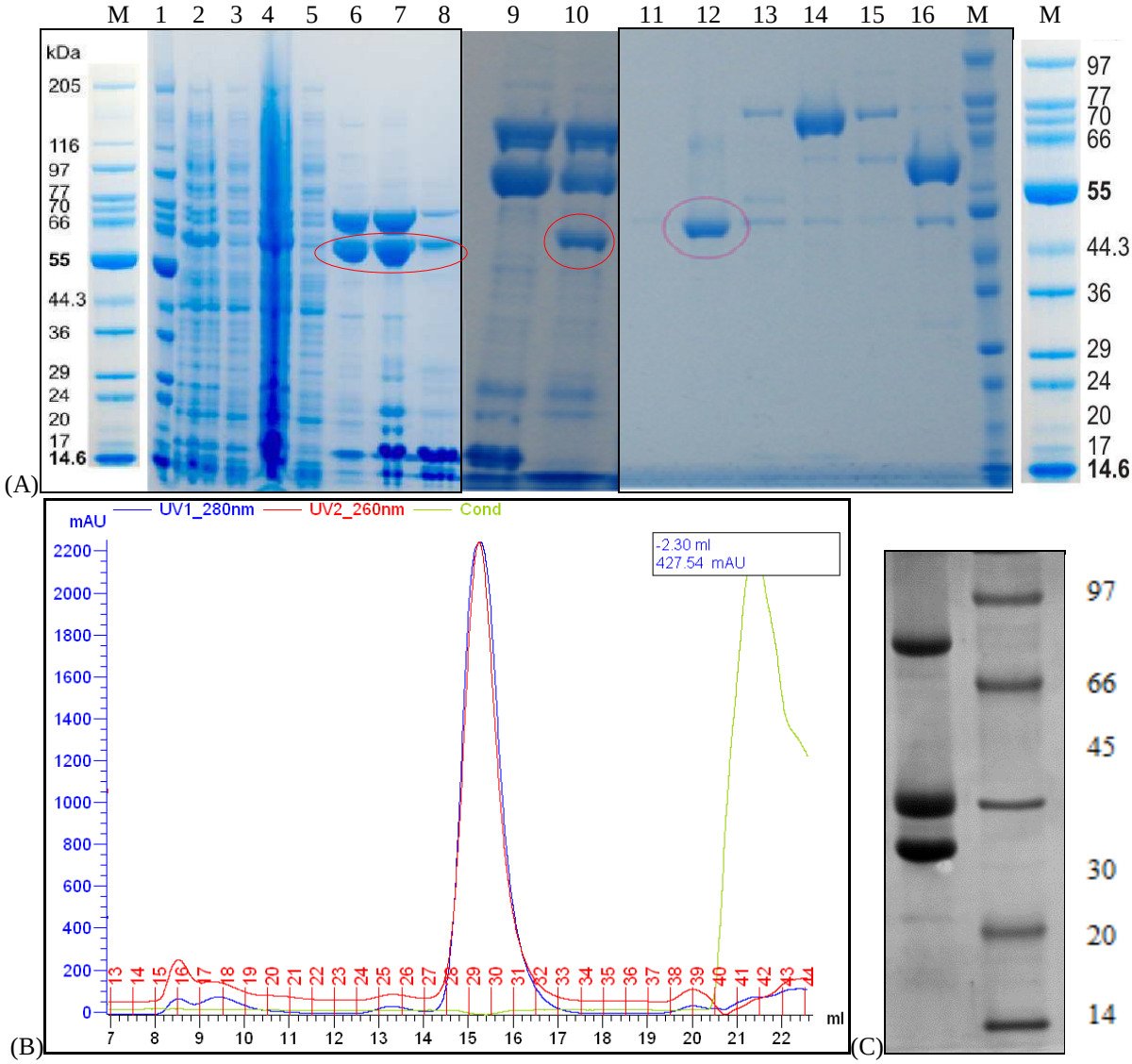
sequence (Reverter *et al.*, 2006). This procedure introduces only one residue (Ser) at the N-terminus of the expressed protein.

2.1.3.11. Expression and purification of His_6-SUMO3-CcHQT

Following the recommended procedure of the Protein Expression and Purification core facility of EMBL-Heidelberg, BL21 (DE3) cells were transformed with pET-28M-SUMO3-*Cchqt* and grown at 37 °C in LB medium containing 30 µg/mL kanamycin and 34 µg/mL chloramphenicol to an OD at λ= 600 nm of 0.5. Expression was induced with 0.2 mM IPTG and cells grown overnight at 18 °C. The cells were harvested by centrifugation and flash-frozen in liquid nitrogen. The bacterial pellets were thawed and resuspended to a final volume of 25 mL/L culture in a lysis buffer composed of 50 mM Tris-HCl pH 8.0, 250 mM NaCl, 20 mM imidazole, Complete Protease Inhibitor, DNaseI and lysozyme. The cells were maintained on ice and ruptured by sonication (Vibra Cell, Bioblock Scientific) with 6 x 25 sec pulses at 60 % intensity and 30 sec breaks. The cell lysate was cleared by centrifugation and the supernatant loaded on to a 5 mL Ni-NTA column (Qiagen). After washing to eliminate unbound contaminants, His_6-SUMO3-HQT was eluted using a gradient of 20 to 300 mM imidazole in the buffer. A protein of 70 kDa that co-elutes with *Cc*HQT from the affinity column may correspond from size comparison to *E. coli* chaperone DnaK (Figure 30A). The purest fractions were dialysed against 50 mM Tris-HCl pH 8.0 and 250 mM NaCl and digested with 1/100 (w/w) SenP2 overnight at 4 °C. The following day, the digestion mixture was centrifuged and loaded again on the Ni-NTA column. The cleaved *Cc*HQT was found in the flow through and remained stable in solution, while the His_6-SUMO3 fusion was retained on the column. The sample buffer was exchanged for 50 mM Tris-HCl pH 8.0 and 100 mM NaCl, filtered and loaded on a 5 mL HiTrap Q anion-exchange column using an Äkta purifier system. *Cc*HQT was eluted in the flow through (Figure 30B), while the contaminants remained bound to the MonoQ column. Finally, HQT was analysed by size-exclusion chromatography and eluted as a monomer. The mean yield was 1 mg/L culture of pure *Cc*HQT. The protein was also sensitive to proteolysis and showed a degradation profile similar to that of *Cc*HCT (Figure 30C). N-terminal sequencing of truncated fragments from HQT degradation gave two sequences: TGPRA and GPRAS, consistent with a cleavage at Lys219 and Thr220. A similar mutant as for HCT, K210A/K219A HQT, was produced and purified using a similar protocol to the native HQT.

Figure 30: SDS-PAGE analysis of His_6-SUMO3-*Cc*HQT during the purification procedure

(A) Lanes 1-4: cell lysate; lanes 5-8: fractions from the first affinity column, lanes 9-10: before and after SenP2 digestion, lanes 11-16: fractions from the ion-exchange column, M: MW marker (kDa). The band corresponding to His_6-SUMO3-HQT/ cleaved HQT is circled in red. (B) size-exclusion chromatogram. (C) Degradation pattern of *Cc*HQT after storage at 4° C for several weeks.



2.1.3.12. HCT and HQT site-directed mutagenesis

The ability to change specific residues or regions of proteins by site-directed mutagenesis allows a better understanding of structure-function relationships in proteins. Prior to structure determination of the native HCT, several residues potentially involved in catalysis and/ or substrate binding were selected from sequence and structure comparisons with VS, MaT and TRI101. HCT mutants were designed and generated by site-directed mutagenesis on the *Cchct* gene template to produce H35A, H153A ($\underline{H}X_3D$), H154N ($H\underline{X}X_2D$), D157A ($HX_3\underline{D}$), Y252A, R374E and H154N/A155L/A156S ($H\underline{X}_3D$) mutant HCTs.

The QuikChange site-directed mutagenesis kit (Stratagene) was used to generate single site mutations in the codon-optimised *hct*. The double mutants were produced by using the template containing the first mutation and primers corresponding to the second mutation to be introduced. The gene encoding the triple H154N/A155L/A156S mutant HCT was constructed in a single step as the target bases were adjacent. According to the manufacturer's specifications, the reaction mixture was composed of 0.25 µL DNA template, 2.5 µL 10x DNA polymerase buffer, 0.5 µL dNTP, 1 µL forward (F) and 1 µL reverse (R) mutagenic primers at 15 pmol/µL (Table 3), 0.5 µL Pfu Turbo pdymerase and completed to 25 µL with milliQ water. After denaturing at 95 °C for 1 min, amplification consisted of 18 cycles of 1 min at 95 °C, 1 min at 60 °C for annealing and 8 min at 68 °C for elongation (1 min/kb of plasmid length). An additional final step of elongation was carried out at 68 °C for 8 min. The PCR products were digested with 0.5 µL DpnI for 1 h at 37 °C. XL-1 Blue super competent cells (Stratagene) were transformed with 5 µL of the digest. The cells were plated on LB-agar containing the appropriate antibiotic to select colonies containing the mutant plasmid. A few clones were selected on each plate for plasmid DNA preparation and sequencing (Genecore, EMBL-Heidelberg) in order to identify one that contained the desired mutation. Each mutant was then over-expressed and purified as previously described for the wild-type enzymes. All mutant proteins were successfully produced and purified to homogeneity.

Table 3: Sequences of the oligonucleotide primers used for the production of mutant HCTs

Mutation	*Pair of mutagenic primers*
c103g/a104c (H35A)	F: atctggtggtgccgaactttgctaccccgagcg R: cgctcggggtagcaaagttcggcaccaccagat
c457g/a458c (H153A)	F: ggtgtgggcatgcgtgctcatgcggcgga R: tccgccgcatgagcacgcatgcccacacc
c460a (H154N)	F: ggtgtgggcatgcgtcataatgcggcgga R: tccgccgcattatgacgcatgcccacacc
a470c (D157A)	F: catcatgcggcggctggctttagcggc R: gccgctaaagccagccgccgcatgatg
t754g/a755c (Y252A)	F: aagatggcaacaccattagcgctagcagctatgaaatgctgg R: ccagcatttcatagctgctagcgctaatggtgttgccatctt
c1120g/g1121a/t1122a (R374E)	F: gcattaccagctgggtggaactgccgatccatgatgc R: gcatcatggatcggcagttccacccagctggtaatgc
c460a/g463c/c464t/g466a/c467g/g468c (H154N/A155L/A156S)	F: gggtgtgggcatgcgtcataatctgagcgatggctttagcggcctgcat R: atgcaggccgctaaagccatcgctcagattatgacgcatgcccacaccc

2.2. Biophysical characterisation

2.2.1. Thermofluor assays

Fluorescence-based thermal shift (Thermofluor) assays, a procedure first described by (Pantoliano *et al.*, 2001), have been used by a number of laboratories as a high-throughput way to measure protein stability. Because the thermal stability of a protein increases with the stabilising effect of a buffer or the binding of a ligand, Thermofluor assays can be used to identify optimal protein buffer solutions, to screen for potential ligand-induced stabilisation and therefore to increase the success rate in crystallisation experiments (Ericsson *et al.*, 2006).

The fluorescent dye fluoroprobe SYPRO orange (Molecular Probes) was added to the protein in the presence of various buffer solutions to monitor the protein unfolding process as a function of temperature. As a protein unfolds, hydrophobic residues that are buried in the native protein become exposed to the solvent. The fluoroprobe is normally quenched in an aqueous environment but, as the dye binds to hydrophobic sites, the fluorescence intensity increases on protein unfolding. The melting temperature (Tm) is obtained by calculating the midpoint temperature from the typical two-state transition curve (Figure 31B). The experimental error of Tm determination is in the range of 0.2-0.5 °C and a significant shift is considered when Tm increases by more than 2 °C (Ericsson *et al.*, 2006).

Thermofluor assays were carried out on freshly purified protein samples prior to crystallisation trials. The final volume of 50 µL per well was composed of 25 µL 2x buffer solution containing the test buffer, additives and/ or ligands or only water as a control, 10 µL 10 to 50 µM protein, 13.75 µL milliQ H_2O and 1.25 µL SYPRO orange 200x (5x final). A real-time quantitative MX3005P PCR instrument (Stratagene) was used with standard 96-well PCR plates and optical seals. The plate was heated from 25 to 89 °C with a heating rate of 1 °C/min and the fluorescence intensity was measured using a SYBR filter (excitation/ emission wavelengths: 492/516 nm). The raw data were analysed using Microsoft Excel by calculating the temperature value corresponding to the average of the maximum and minimum fluorescence intensity values.

2.2.1.1. Cc4CL2 and Nt4CL2

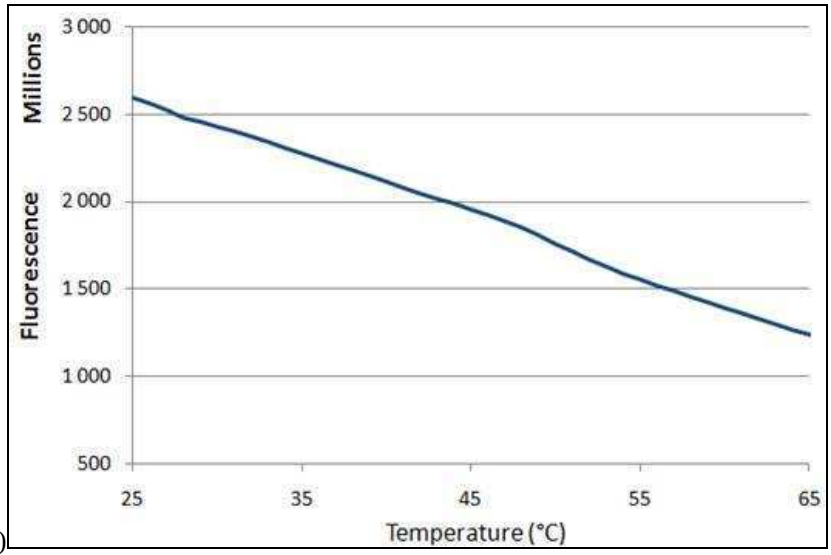
The stability of coffee and tobacco 4CL2 samples was verified using Thermofluor assays. *Cc*4CL2 (2.1.2.1) did not present a clear unfolding transition (Figure 31A). A buffer and ligand screening was performed to improve the stability in solution on a purified fraction

of *Cc*4CL2, but no improvement was seen. In the case of *Nt*4CL2, the melting temperature determined in the storage buffer was 42 °C (Figure 31B).

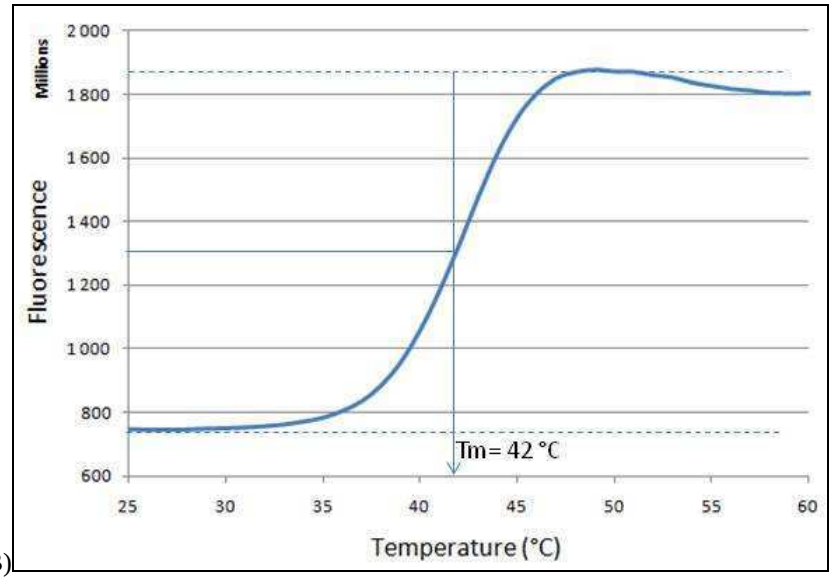
Figure 31: Thermofluor analysis of coffee and tobacco 4CL2 prior to crystallisation screening

(A) *Cc*4CL2. (B) *Nt*4CL2.

(A)



(B)



2.2.1.2. CcHCT

Native *Cc*HCT was diluted to 0.76 mg/mL in order to decrease the storage buffer and salt concentrations by a factor of ten. Several buffer types at a concentration of 0.1 M and pH values were screened: citric acid pH 5.0, MES pH 5.0/ 5.5/ 5.8/ 6.0/ 6.2/ 6.5, sodium citrate pH 5.5/ 6.0, sodium phosphate pH 6.6/ 7.0/ 7.6, potassium phosphate pH 6.0/ 7.0, sodium

HEPES pH 7.0/ 8.0, ammonium acetate pH 7.3, Tris-HCl pH 7.5/ 8.0/ 8.5, tricine pH 8.0, imidazole pH 8.0, bicine pH 8.0/ 8.5/ 9.0. Different NaCl concentrations were also tested: 0/ 0.1/ 0.2/ 0.5 M. According to the Tm values measured, HCT is more stable in buffers with a pH higher than 7.0 and sodium phosphate stabilises HCT the most compared to the other buffers tested (Figure 32). Phosphate buffers are generally avoided in crystallisation trials as they may lead to the formation of salt crystals. Therefore it was only used as a buffering agent for the biochemical characterisation (2.3). The methylated HCT (2.1.3.7) showed a similar buffer preference, but with lower Tm values (ΔTm~ 2 °C) probably due to a decrease in the protein solubility.

The influence of adding of ligands on the Tm was also evaluated. 1 mM ligand (5-CQA, 5-FQA, CoA, quinic, shikimic, coumaric, caffeic or ferulic acid) was added individually or in combination to a solution containing 0.5 mM HCT prior to the Thermofluor assay. Four different types of buffer (0.1 M HEPES pH 7.5; 0.1 M Tris-HCl pH 7.5; 0.1 M sodium phosphate pH 7.5; 0.1 M MES pH 6.0) were also tested. The most significant stabilising effect for HCT was observed in the presence of 1 mM CoA alone or in combination with 5-CQA or quinic acid (Figure 33). Therefore the cofactor was used as an additive in the crystallisation trials (2.4.3.1).

Figure 32: Melting temperatures from *Cc*HCT Thermofluor assays in different buffer solutions (0.1 M)

The melting temperature (Tm) is indicated after the buffer description for each plot.

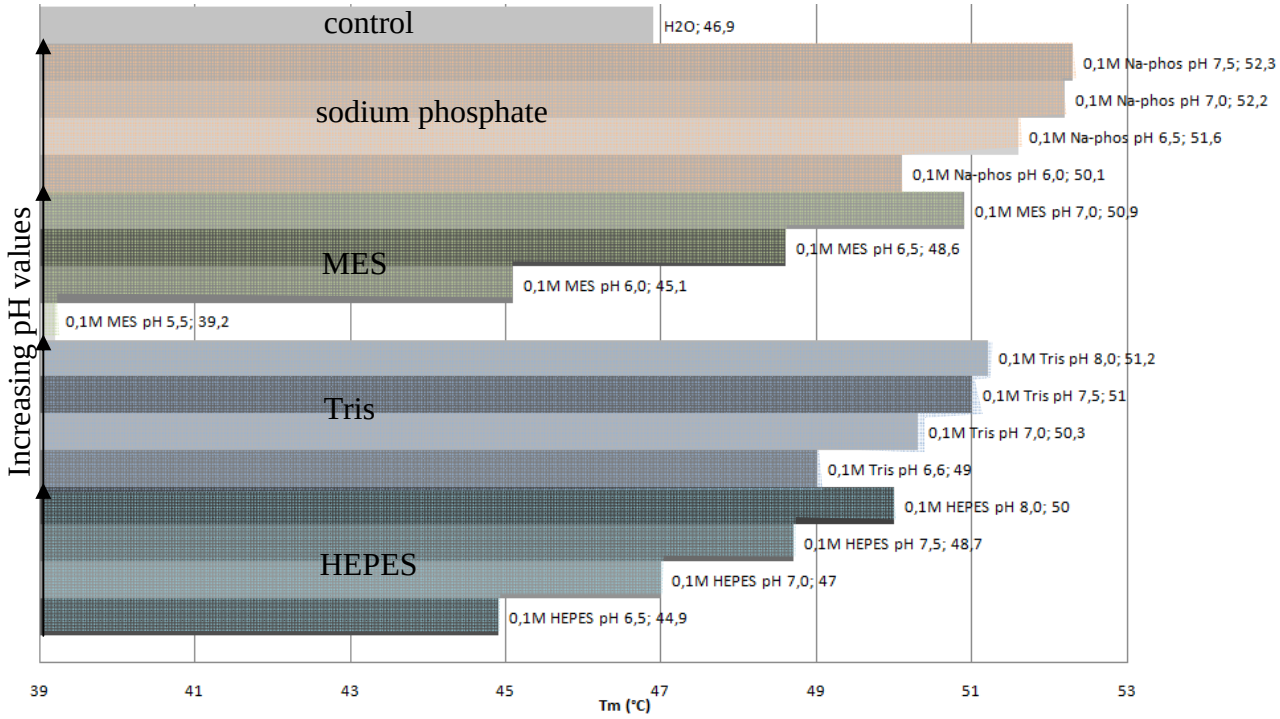
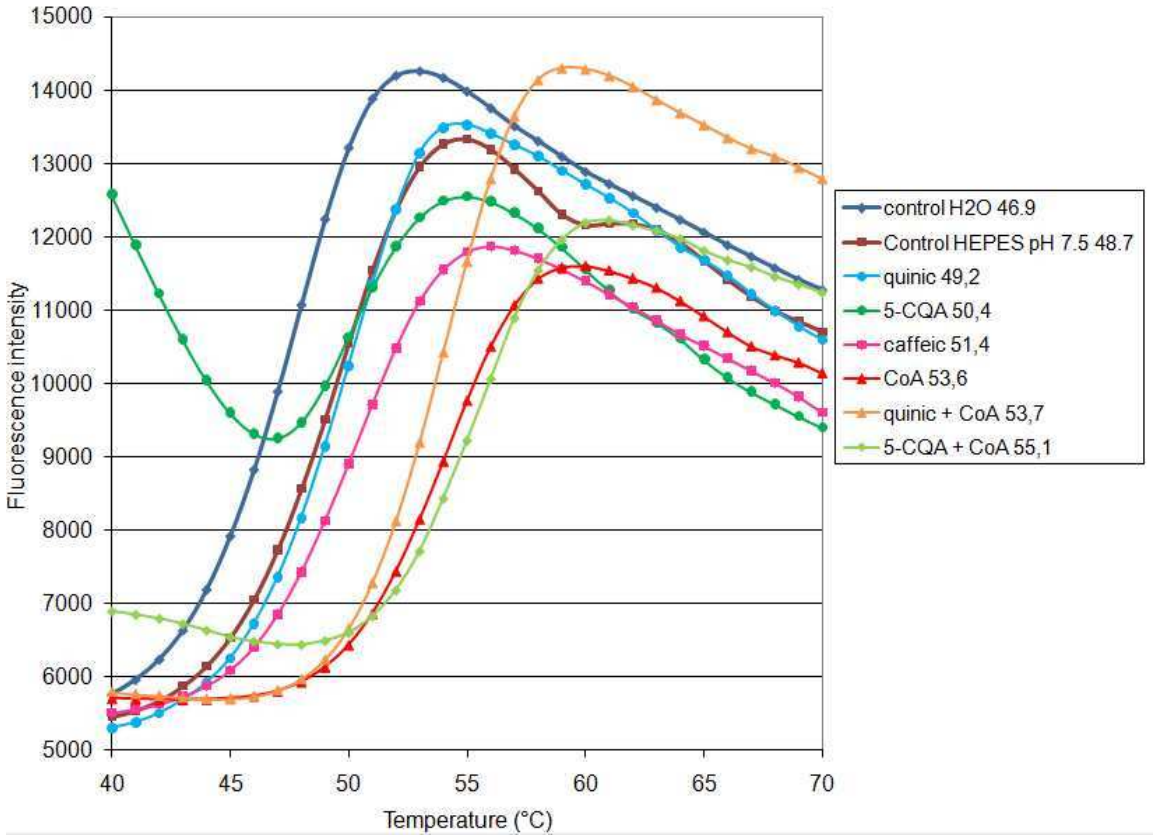


Figure 33: Melting temperatures from *Cc*HCT Thermofluor assays with various ligands in 0.1 M HEPES pH 7.5



2.2.1.3. HQT

*Cc*HQT was used in a Thermofluor assay prior to crystallisation trials and showed a relatively high Tm value (49 °C), indicating that the protein was stable in the final buffer composed of 10 mM Tris-HCl pH 7.5 and 50 mM NaCl. The protein was incubated with 1 mM CoA overnight and showed a Tm of 55 °C, suggesting that the cofactor significantly stabilises the enzyme.

2.2.2. Isothermal calorimetry

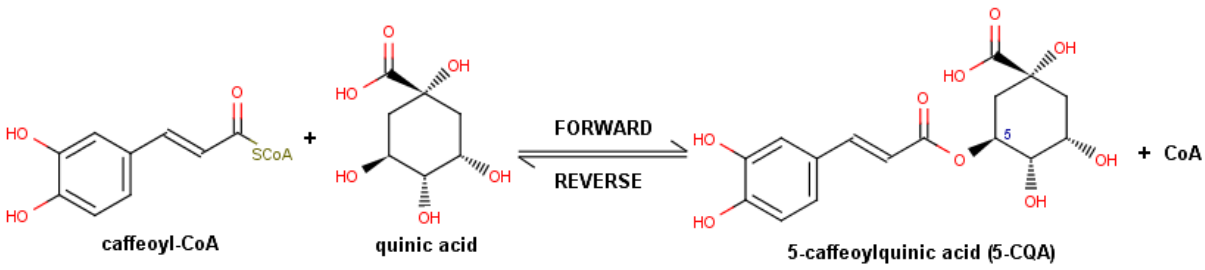
Isothermal calorimetry (ITC) is a routine biophysical technique used to determine the thermodynamics of protein-ligand interactions (Freyer *et al.*, 2008; Liang, 2008). Native *Cc*HCT was thawed and dialysed overnight into a buffer composed of 10 mM sodium HEPES pH 6.8 and 100 mM NaCl. The control and ITC experiments were carried out at 25 °C with the MicroCal VP-ITC system. The protein concentration was adjusted to 20 µM in the cell. CoA (alternatively 5-CQA) was dissolved at 200 µM in the dialysis buffer and loaded in the syringe for 14 µL injections. In both cases, no binding interaction was observed, even with a two-times higher protein concentration. It is clear that this technique is unsuitable for measuring the binding interactions of HCT.

2.3. Biochemical characterisation

2.3.1. Introduction

Nicotiana tabacum 4-coumarate CoA ligase (*Nt*4CL2) and *Coffea canephora* hydroxycinnamoyl-CoA shikimate/ quinate hydroxycinnamoyltransferases (*Cc*HCT/ *Cc*HQT) were over-expressed in *E. coli* and purified to homogeneity using several chromatographic steps. The following section presents the experimental conditions used to investigate the *in vitro* activity of these enzymes, which are essential components of the biochemical pathway leading to CGAs and other phenolic compounds in plants. *Nt*4CL2 is involved in the synthesis of hydroxycinnamoyl-CoA thioesters, while *Cc*HCT and *Cc*HQT catalyse the transfer of hydroxycinnamoyl moieties between CoA and quinic or shikimic acid. Mutants, notably targeting the conserved HX_3D motif (0), were also assayed. The acyltransfer leading to CGA biosynthesis is designated as the "forward" reaction, while the "reverse" reaction involves the conversion of a CGA molecule to the corresponding hydroxycinnamoyl-CoA thioester (Figure 34). The reverse reactions using 5-CQA/ 5-FQA and CoA as substrates were studied first because the CoA thioesters were not commercially available. In order to study the forward reaction catalysed by *Cc*HCT and *Cc*HQT, these substrates were synthesised enzymatically using *Nt*4CL2 and purified based on a previously published procedure (Beuerle *et al.*, 2002).

Figure 34: Biosynthesis and turnover reaction of the major CGA compound, 5-CQA



2.3.2. The high-performance liquid chromatography system

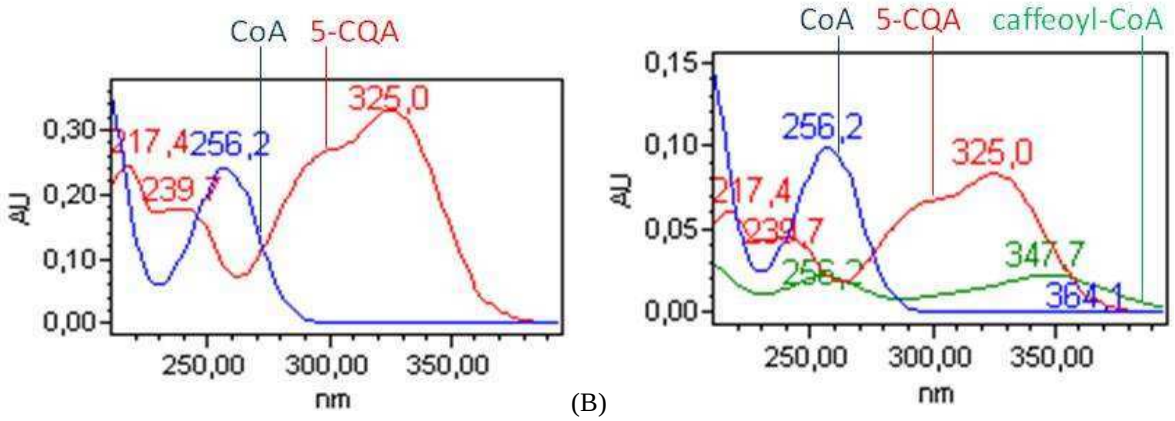
2.3.2.1. Spectrophotometry versus HPLC

To study an enzyme-catalysed reaction, one needs to follow substrate depletion and product formation. Spectrophotometric methods are commonly used for this purpose because they are simple and accurate. However, obvious requirement is that either a substrate or a

product in the reaction must absorb light in a spectral region where other substrates or products do not. In the case of the reaction under study, CoA thioesters and quinate/ shikimate esters have distinct maximum absorbance values, but present overlapping absorption spectra (Figure 35). Compared to spectrophotometry, liquid chromatography is more informative about the chemical entities present in a reaction mixture. It is also more accurate as the peak area of the absorbing species is directly proportional to its concentration by the Beer-Lambert law. Therefore, high-performance liquid chromatography (HPLC) was preferred for the studies carried out here. Quinic acid is not detectable in the range of 210 to 400 nm wavelength, while shikimic acid can be measured at λ= 210 nm due to the double bond within the cyclohexene ring. High-performance anion-exchange chromatography coupled to pulsed electrochemical detection (HPAE-PED) is the usual technique for the analysis of these acids (Rogers *et al.*, 1999), but it was not available for this study.

Figure 35: Overlap of the absorption spectra of CoA, 5-CQA and the reaction product caffeoyl-CoA

Reactions were set up using 5 mM CoA and 5 mM 5-CQA in 0.1 M sodium phosphate pH 6.0. The reaction was started by adding 0.002 µM HQT and incubated at 34 °C. Samples taken at t= 0 (A) and 30 min (B) were analysed by HPLC with photodiode array (PDA) detector.



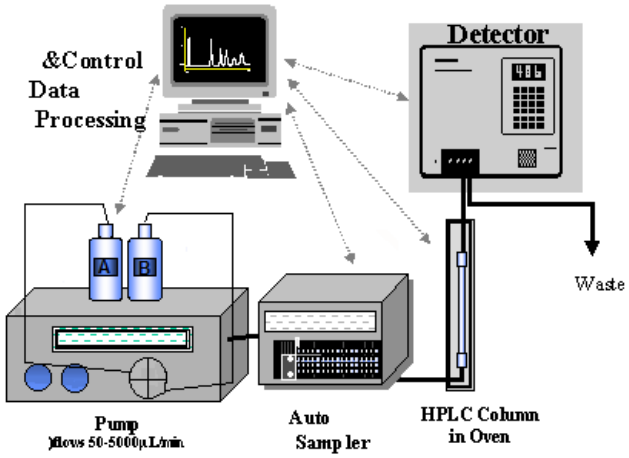
2.3.2.2. Description of the HPLC system

Chromatography separates a mixture of analytes on the basis of their relative affinity for the mobile and stationary phases. In HPLC, the separation speed is increased by forcing the liquid solvent through the stationary phase of the column under high pressure. Reversed-phase HPLC, where a polar mobile phase and a non-polar stationary phase are used, is the most frequently used liquid chromatography technique. In this study, the analytes were resolved by their relative hydrophobicity on a C_{18} bonded-phase column composed of octadecyl hydrocarbon chains anchored to non-polar silica gel particles. A Novapak column

(4 µm, 4.6x 250 mm, Waters) or a Nucleosil column (5 µm, 4x 250 mm, Macherey-Nagel) were used in the experiments, depending on their availability.

Before injection, samples were filtered using a 1 ml syringe and a 13 mm Acrodisc filter with a 0.2 µm membrane (Pall Corporation). A pre-column was necessary to eliminate compounds that irreversibly bind the C_{18} column and filter any precipitation that may occur between the time samples are loaded for analysis and their actual injection time by an automatic sampler. Samples were injected into the mobile phase (injection volume of 10 µL, unless otherwise specified) with a flow rate of 0.8 mL/min. The mobile phase was composed of two solvents (A and B) whose respective contributions varied according to the programmed elution gradient. The percentage of organic solvent in the mobile phase was incrementally increased to elute the compounds off the column, the more hydrophobic component eluting last (Table 4). The mobile phase was acidified with 0.1 % phosphoric acid (0.1 % formic acid in case of a coupling with a mass spectrometer) in order to better separate the ionic and ionisable compounds. As illustrated in Figure 36, the HPLC system used comprised an autosampler (Waters 717), high precision pumps (Waters 600E System Controller), an oven where the column was kept at 30 °C and a Waters photodiode array (PDA) UV detector, scanning wavelengths in the range 210 to 400 nm wavelength. When the HPLC-PDA was not available, a simpler system equipped with a dual absorbance (DA) detector was used to record the absorbance at two particular wavelengths and did not have a column incubator. Data processing was performed using the Waters Empower software. Peak assignments were made by comparison with available standards using the peak retention time (RT) and with UV absorbance spectra recorded with the PDA detector.

Figure 36: Components of the HPLC system



2.3.2.3. *From an acetonitrile to a methanol solvent system*

The elution gradient used was initially based on acetonitrile and a method developed at Nestlé to analyse coffee CGA extracts (Table 4). Due to a global acetonitrile shortage in the first half of 2009 (Lowe, 2009), an alternative methanol-based method was designed. As the two solvents differ in their relative eluotropic strength, the elution gradient was modified accordingly to maintain a good separation of coffee CGA isomers (Table 4).

Table 4: Mobile phase gradients used for HPLC analysis of coffee CGAs

Mobile phase	Acetonitrile		Methanol	
Composition (%)	A:	B:	A:	B:
	8 % CH_3CN	50 % CH_3CN	20 % CH_3OH	60 % CH_3OH
Time (min)	0.1 % H_3PO_4	0.1 % H_3PO_4	0.1 % H_3PO_4	0.1 % H_3PO_4
0	93	7	93	7
2	93	7	93	7
20	90	10	90	10
31	65	35	65	35
40	30	70	30	70
42	30	70	20	80
44	93	7	93	7
55	93	7	93	7

2.3.2.4. *Substrates and chemicals*

Only the 5-CQA and rosmarinic acid were available as commercial standards. 5-FQA was provided in limited amounts by Nestlé. A kit of mono- and diCQAs purified from plant extracts was eventually obtained in 2009 from Biopurify (Chengdu, China). All other chemicals were obtained in a lyophilised form and stored as specified in Table 5. MilliQ-purified water was used for all reagent dilutions and buffer preparation. Stock solutions were freshly prepared and stored on ice for a maximum of one day or frozen at -80 °C prior to use. CoA and 5-CQA stock solutions were prepared at 100 mM and 20 mM respectively. The HPLC grade solvents (methanol, acetonitrile) as well as the additives (phosphoric and formic acids) were filtered before use. The reaction buffer was prepared using monobasic (0.5 M NaH_2PO_4 pH 4.0) and dibasic (0.5 M Na_2HPO_4 pH 9.0) solutions mixed to obtain the appropriate pH value and stored at room temperature.

Table 5: Information on the substrate compounds used in the biochemical assays

Compound	*Formula*	*MW (g/mol)*	*Storage (°C)*	*Min. purity (%)*	*Brand*
5-CQA	$C_{16}H_{18}O_9$	354.31	20	99.1 98	Biopurify Fluka
3-CQA	$C_{16}H_{18}O_9$	354.31	4	99.4	Biopurify
4-CQA	$C_{16}H_{18}O_9$	354.31	4	99.6	Biopurify
5-FQA	$C_{17}H_{20}O_9$	368	4	98	gift
3,5-diCQA	$C_{25}H_{24}O_{12}$	516.46	4	99.3	Biopurify
3,4-diCQA	$C_{25}H_{24}O_{12}$	516.46	4	99.4	Biopurify
4,5-diCQA	$C_{25}H_{24}O_{12}$	516.46	4	99.1	Biopurify
caffeic acid	$C_9H_8O_4$	180.16	20	98	Sigma
ferulic acid	$C_{10}H_{10}O_4$	194.19	20	99	Sigma
t-cinnamic acid	$C_9H_8O_2$	148.16	20	99	Sigma
coumaric acid	$C_9H_8O_3$	164.16	20	98	Sigma
sinapic acid	$C_{11}H_{12}O_5$	224.21	20	98	Sigma
coenzyme A	$C_{21}H_{36}N_7O_{16}P_3S$	767	-20	85 96	Sigma BioChemika
quinic acid	$C_7H_{12}O_6$	192	20	98	Fluka
shikimic acid	$C_7H_{10}O_5$	174	20	99	Sigma
rosmarinic acid	$C_{18}H_{16}O_8$	360.3	20	97	Aldrich
spermidine	$C_7H_{19}N_3$	145.25	20	99.5	Sigma
myo-inositol	$C_6H_{12}O_6$	180.2	20	99.5	Sigma
pHPL	$C_9H_{10}O_4$	182.2	20	98	Aldrich

2.3.2.5. HPLC analysis of CGA and HCA standards

The retention times (RT) and absorbance spectra of the HCAs, which serve as substrates for 4CL2, were determined (Table 6). Caffeic, coumaric, ferulic and sinapic acids were injected individually and analysed by HPLC with the methanol or acetonitrile method.

Table 6: Typical retention times and absorbance maxima of the HCA standards

Compound	*RT (min)*		*Max. abs. (nm)*
	methanol	*acetonitrile*	
Coumaric acid	31.9	28.9	308.3
Caffeic acid	18.3	16.7	322.6
Ferulic acid	37.8	35.2	322.6
Sinapic acid	40.2	38.2	322.6

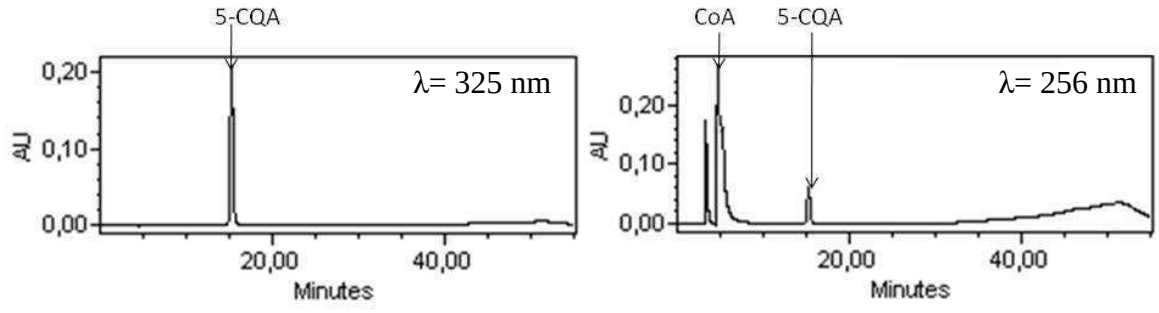
CoA eluted close to the void volume and was best measured at λ= 256 nm (Figure 37A). The CGA elution sequence, which was driven by an increase in hydrophobicity, did not change in the chromatographic systems used (3-CQA< 4-CQA< 5-CQA< 5-FQA< 3,4-diCQA< 3,5-diCQA< 4,5-diCQA) (Figure 37B). The UV spectrum of each CQA isomer is very similar (Figure 37C). RTs for each compound could vary depending on the mobile phase (acetonitrile/ methanol), HPLC device (photodiode array or dual absorbance detector) or column (Nucleosil/ Novapak) used (Table 7). As conditions were different from one Nestlé visit to another, pure CGAs were systematically injected as standards either individually or in a mixture before each reaction series. As an alternative to pure commercial standards, methanolic extracts of green coffee beans proved to be a convenient source of the major coffee CGAs. Typically, the beans were frozen in liquid nitrogen and ground. Phenolic compound were then extracted with 70 % methanol for 1 h at 40 °C in shaking flasks. The methanolic extract was filtered (0.2 µm) and a 10-fold dilution was analysed by HPLC-PDA (Figure 37D).

Table 7: Typical retention times and absorbance maxima of the CGA standards

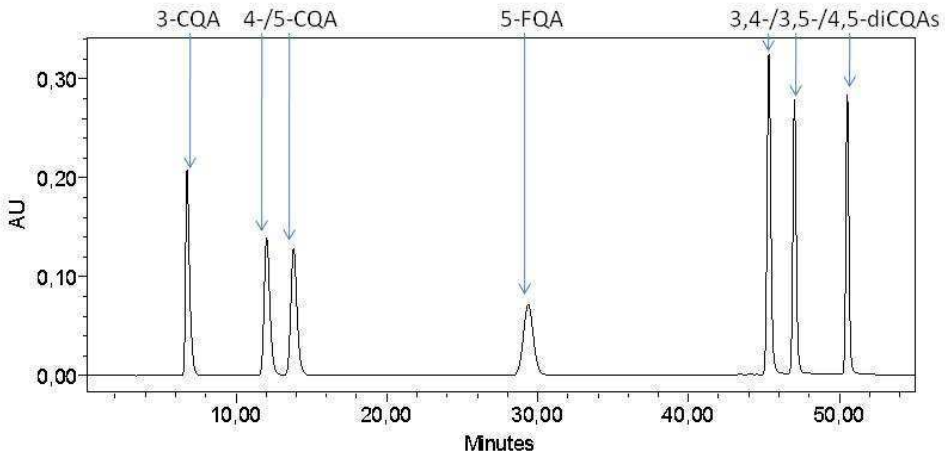
Compound	*RT (min)*		*Maximum absorbance (nm)*
	methanol	*acetonitrile*	
CoA	4.7	3.9	256.0
3-CQA	6.8	6.8	325.0
4-CQA	12.0	11.9	326.2
5-CQA	13.8	13.0	325.0
5-FQA	29.3	27.1	325.0
3,4-diCQA	45.6	42.1	325.0
3,5-diCQA	47.2	43.1	328.6
4,5-diCQA	50.6	44.5	328.6

Figure 37: HPLC profiles of CoA and CGA standards

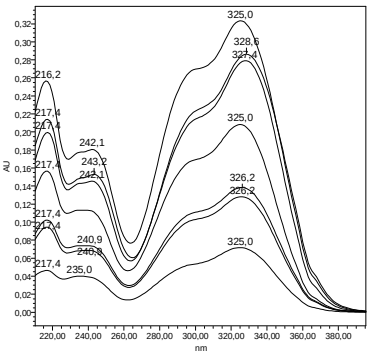
(A) CoA and 5-CQA standard solutions analysed with the acetonitrile method. The chromatograms were extracted at λ= 325 nm (left) and 256 nm (right). (B) 0.2 mM solutions of 3-, 4- and 5-CQA, 5-FQA and 3,4-, 3,5- and 4,5-diCQA analysed with the methanol method and measured at λ= 325 nm (sample 883). (C) Overlap of CGA absorbance spectra. (D) Coffee extract analysed with the acetonitrile method.



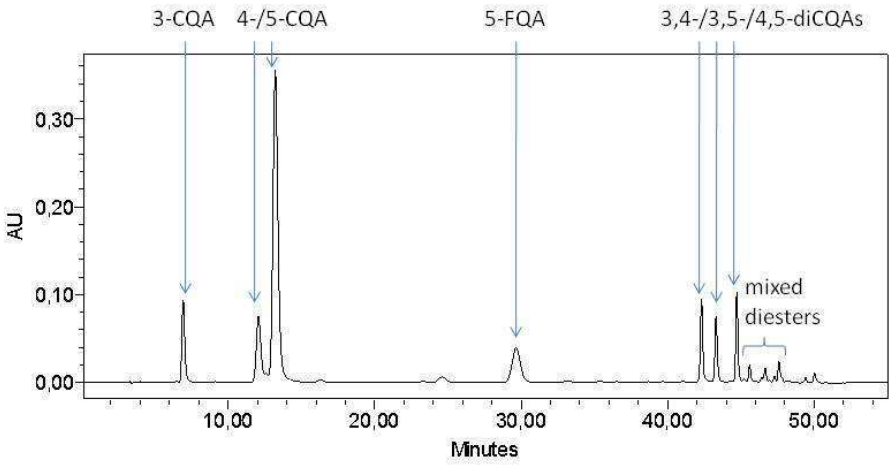
(A)



(B)



(C)



(D)

A calibration of peak area against concentration was obtained for 5-CQA at λ= 325 nm. Linearity was respected in the range 0 to 10 mM (data not shown). To quantify CGAs in a coffee methanolic extract, a reference solution of 5-CQA at a known concentration was analysed as an external standard before the sample of interest was tested. The concentration of CGAs in the sample was deduced from the corresponding peak area at λ= 325 nm and the values for the 5-CQA standard (Equation 1). The relative response factor (RRF) was taken into account as a correction for molecular weight differences with the 5-CQA standard (Lepelley *et al.*, 2007).

Equation 1: Quantification of CGA compounds in a sample with 5-CQA as an external standard

A: absorbance; C: concentration; RRF: relative response factor: RRF= 1 for the mono-CQAs, 0.96 for 5-FQA and 1.36 for the diCQAs.

$$A_{std}/C_{std} = A_{5\text{-}CQA}/C_{5\text{-}CQA} = A_{CGA}/C_{CGA} * RRF$$

2.3.2.6. Stability of CGA solutions

The stability of CGA compounds was verified by heating 10 mM 5-CQA/ 5-FQA in 0.1 M sodium phosphate pH 6.0/ 7.0 at 90 °C. Samples taken at t= 0, 30 and 60 min were analysed by HPLC-DA. Similar experiments were performed with 1 mM 3-/ 4-/ 5-CQA, 3,4-/ 3,5-/ 4,5-diCQA in 0.1 M sodium phosphate pH 6.5 incubated overnight (~15 h) at 34/ 25 °C. Samples (1369-1386) were analysed using HPLC-PDA with the acetonitrile method.

2.3.2.7. General preparation of the enzymatic reactions

Stock solutions of substrates and buffer were mixed with MilliQ water to reach the desired concentration and final volume. Reactions were started by adding the enzyme (or water for the control reaction) and solutions incubated in a 34 °C water-bath during sampling, unless otherwise stated. Typically, a 40 µL aliquot was taken and diluted with 210 µL HPLC solvent A. After filtration, the automated system injected a 10 µL sample into the HPLC system. The following sections describe the enzymatic reaction details. The figure legends in the Results section of this thesis also contain reaction details and sample numbers. All chromatograms reported represent, as a function of time, the absorbance value measured at λ= 325 nm, except where otherwise stated.

2.3.3. Enzymatic synthesis and purification of the CoA thioesters

2.3.3.1. Coffee and tobacco 4-coumarate CoA ligases

In order to assay the activity of *Cc*4CL2, a pure fraction of the recombinant protein (5 µM) was incubated with 0.2 mM coumaric acid, its major biological substrate, in the presence of 1 mM CoA, 5 mM $MgCl_2$ and 5 mM ATP in 0.1 M sodium phosphate pH 7.5. Samples (876-879) taken at t= 0, 15 min and after overnight incubation at 34 °C were analysed using HPLC-PDA with the methanol method.

The activity of *Nt*4CL2 was tested with cinnamic acid and its common derivatives coumaric, caffeic, ferulic and sinapic acids. Reaction mixtures containing 1 mM (hydroxy)cinnamic acid, 5 mM CoA, 2.5 mM $MgCl_2$ and 2.5 mM ATP in 0.1 M Tris-HCl pH 7.5 were prepared in a 400 µL final volume. Reactions were started by adding 1.6 µM recombinant enzyme. Samples (590-603) taken at t= 0, 2 and 20 min were analysed using HPLC-PDA with the methanol method. Similar reaction mixtures were made up in three different buffers: 0.1 M sodium phosphate pH 6.0/ 7.5 and 0.1 M Tris-HCl pH 7.5. Samples (634-683) taken at t= 0, 5, 25 min and after overnight incubation were analysed using HPLC-PDA with the methanol method.

As no commercial CoA thioester was available, coumaroyl-CoA, caffeoyl-CoA and feruloyl-CoA were synthesised enzymatically as follows. 10 mL reaction mixtures composed of 0.8 mM coumaric/ caffeic/ ferulic acid, 1 mM CoA, 2 mM $MgCl_2$ and 2 mM ATP were prepared in 0.1 M sodium phosphate pH 6.0. Reactions were started by adding 0.05 µM *Nt*4CL2. Samples (699-715) taken at t= 0, 10, 60 and 150 min were filtered through a 50 kDa cut-off membrane to remove the enzyme. The filtrates were analysed using HPLC-PDA with the methanol method. The reaction was complete after 150 min and all subsequent reactions to produce the hydroxycinnamoyl-CoA thioesters were stopped at this time. The final concentration of the CoA thioester was deduced from the initial amount of HCA supplied in the reaction.

2.3.3.2. Purification of CoA thioesters

Successive injections of reaction volumes of up to 150 µL, corresponding to the maximum loop capacity, were carried out to purify caffeoyl-CoA by preparative HPLC with either the acetonitrile or methanol method. The reaction product was subsequently collected manually as it eluted off the column.

A simpler method to purify a CoA thioester from a reaction mixture was later developed. This was based on a previously described procedure (Beuerle *et al.*, 2002). For

this, a 1 mL solid-phase extraction (SPE) Sep-Pak C_{18} cartridge (Waters) was activated with 10 column volumes (CV) of methanol, washed with 10 CV milliQ H_2O and equilibrated with 10 CV of 4 % ammonium acetate buffer. Ammonium acetate was added to the reaction mixture to a final concentration of 4 %, before loading onto the cartridge using a syringe. The column was subsequently washed with 10 CV of 4 % ammonium acetate and 10 CV of milliQ water. The CoA thioester was then eluted with several 500 µL fractions of 100 % methanol, which were analysed by HPLC.

The fractions containing the CoA thioester collected from preparative HPLC or eluted from the SPE column were evaporated to near dryness using a vacuum concentrator in 1.5 mL Eppendorf tubes. The residue was resuspended in milliQ-purified water before further HPLC analysis or use in enzymatic reactions.

2.3.4. Comparison of the HCT and HQT-catalysed forward reactions

2.3.4.1. HCT and HQT acyl acceptor specificity

In the forward direction, HCT and HQT catalyse the transfer of hydroxycinnamoyl moieties from CoA to quinic or shikimic acid. To verify their substrate preference, reaction mixtures containing 5 mM quinic/ shikimic acid or both and 0.2 mM caffeoyl-CoA in 0.1 M sodium phosphate pH 6.5 were prepared in a final volume of 100 µL. Reactions were started by adding 1 µM HCT/ HQT. Samples (1023-1052) taken at t= 0, 5 and 60 min were analysed using HPLC-PDA with the methanol method.

In order to test the reactivity of HCT/ HQT towards other acyl acceptor molecules, reactions containing 10 mM quinic/ shikimic/ 4-hydroxyphenyllactic acid/ spermidine or *myo*-inositol and 0.2 mM caffeoyl-CoA in 0.1 M sodium phosphate pH 6.5 were prepared in a 100 µL final volume. Reactions were started by adding 0.5 µM HCT/ HQT. Samples (1146-1163) taken at t= 0, 15 min and after overnight incubation were analysed using HPLC-PDA with the methanol method.

2.3.4.2. HCT and HQT phenolic donor specificity

In order to compare the phenolic donor preference of HCT and HQT in the forward reaction, the following assays were set up. Reaction mixtures containing 0.2 mM caffeoyl-CoA/ coumaroyl-CoA/ feruloyl-CoA and 10 mM quinic/ shikimic acid in 0.1 M sodium phosphate pH 6.0 were prepared in a 500 µL final volume. Reactions were started by adding 0.4/ 0.004 µM HCT/ HQT. Samples (716-727, 764-775 and 800-811) taken at t= 0, 10 and 75 min were analysed with the methanol method.

2.3.5. Comparison of the HCT and HQT-catalysed reverse reactions

2.3.5.1. Characterisation of the products obtained in the reverse reaction

The activity of recombinant HCT and HQT was first tested towards the CGAs available, 5-CQA and 5-FQA, with CoA as the cofactor. Reaction mixtures containing 3.75 mM 5-CQA/ 5-FQA and 5 mM CoA in 0.1 M sodium phosphate pH 6.0 were prepared in a final volume of 100 µL. Reactions were started by adding 1 µM HCT or 1/ 0.1/ 0.01 µM HQT. Samples (395-428) taken at t= 0, 5, 15, 30 and 60 min were analysed using HPLC-PDA with the methanol method.

To confirm the formation of caffeoyl- and feruloyl-CoA, liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis was performed. For this, reaction mixtures made up of 2 mM 5-CQA/ 5-FQA and 2 mM CoA in 0.1 M sodium phosphate buffer pH 6.0 were prepared to a final volume of 500 µL. Reactions were started by adding 0.04 µM HQT and incubated at 25 °C. Samples taken at t= 0 and 45 min for 5-CQA (sample 3) and t= 0 and 60 min for 5-FQA (sample 4) were analysed using HPLC-PDA with the acetonitrile method. These samples were also sent to Karin Kraehenbuehl (Nestlé Research Centre, Lausanne), who performed the LC-MS/MS analysis. The negative detection mode was adopted because phenolic acids ionise better in this mode.

The Biopurify kit enabled the testing of the activity of HCT and HQT towards other CGA monoesters (3-CQA and 4-CQA) and diesters (3,4-diCQA, 3,5-diCQA and 4,5-diCQA). Reaction mixtures comprised 1 mM CGA and 1/ 5 mM CoA in 0.1 M sodium phosphate pH 6.0/ 7.5. Reactions were started by adding 0.4 µM HCT/ HQT. Samples (604-633 and 684-698) were taken at t= 0, 30 min and after overnight incubation and analysed using HPLC-PDA with the methanol method.

Rosmarinic acid is an ester of caffeic acid and 3,4-dihydroxyphenyllactic acid, similar in nature to CGAs. The reactivity of HCT and HQT was therefore tested towards this potential substrate. A reaction mixture composed of 1 mM rosmarinic acid and 1 mM CoA in 0.1 M sodium phosphate pH 6.5 was prepared in a 100 µL final volume. Reactions were started by adding 1 µM HCT/ HQT. Samples (978-987) were taken at t= 0 and after overnight incubation and analysed using HPLC-PDA with the methanol method.

2.3.5.2. Temperature and pH influence

Initial characterisation of the optimal pH and temperature conditions was carried out with a reaction mixture composed of 2 mM 5-CQA, 2 mM CoA and 1 µM HCT/ HQT in 0.1 M sodium phosphate buffer. The temperature (22, 34 and 42 °C at pH 6.0) and pH values

(4.6, 6.0 and 7.6 at 34 °C) were varied. Samples (36-47 and 49-58) were analysed by HPLC-PDA with the methanol method.

2.3.5.3. Influence of enzyme, 5-CQA, CoA and quinic acid concentrations

To study the influence of 5-CQA concentration, reaction mixtures containing 0.02 to 10 mM 5-CQA and 2 mM CoA in 0.1 M sodium phosphate pH 7.0 were prepared in a 500 µL final volume. Reactions were started by adding 0.1 µM HQT and incubated at 30 °C. Samples taken at t= 0 to 30 min were analysed using HPLC-DA with the acetonitrile method.

To determine the influence of enzyme concentration, reaction mixtures containing 2 mM 5-CQA and 5 mM CoA in 0.1 M sodium phosphate pH 6.0 were prepared in a 500 µL final volume. Reactions were started by adding various concentrations of HCT/ HQT and incubated at 30 °C. Samples taken at t= 0 to 60 min were analysed using HPLC-DA with the acetonitrile method.

The influence of quinic acid concentration was studied using reaction mixtures containing 5 mM CoA, 5 mM 5-CQA and 0/ 1/ 5/ 20 mM quinic acid in 0.1 M sodium phosphate pH 6.0 in a 100 µL final volume. Reactions were started by adding 1 µM HQT. Samples (488-503) taken at t= 0, 5 and 30 min for were analysed using HPLC-PDA with the methanol method.

To study the influence of CoA, varied concentrations of CoA (0.125, 0.25, 0.5, 1 and 5 mM) were added to 1 mM 5-CQA in 0.1 M sodium phosphate pH 6.5 to a 500 µL final volume. Reactions were started by adding 0.5 µM HCT/ HQT. Samples (1101-1145 and 1164-1168) taken at t= 0, 3, 30, 180 min and after overnight incubation were analysed using HPLC-PDA with the acetonitrile method.

2.3.6. Comparison of the activity of native and mutant HCTs

2.3.6.1. Activity assay in the forward and reverse reactions

In order to determine the influence of point mutations on HCT enzymatic activity, 0.1 mM caffeoyl-CoA, 1 mM quinic/ shikimic acid or both were mixed in 0.1 M sodium phosphate pH 6.5 to a 100 µL final volume. Reactions were started by adding 5 µM native, K210A/K217A (K), H153A, H35A, D157A, Y252A, R374E, H154N and H154N/A155L/A156S mutant HCTs. Samples (1321-1350) taken at t= 0 and 210 min were analysed using HPLC-PDA with the methanol method.

2.3.6.2. Additional analysis of the H154N and H154N/A155L/A156S mutant HCTs

In order to study in more detail the activity of the mutants targeting the His154 residue, reaction mixtures containing 5 mM 5-CQA and 5 mM CoA in 0.1 M sodium phosphate pH 6.0 were prepared in a 100 µL final volume. Reactions were started by adding 1 µM native, H154N or H154N/A155L/A156S mutant HCTs. Samples (326-349, 464-487, 956-977 and 1023-1052) were taken at various time points and analysed using HPLC-DA/ PDA with the methanol or acetonitrile method.

Caffeoyl-CoA (0.01-0.2 mM) was similarly incubated with 1 mM 5-CQA and the native or H154N mutant HCTs. Samples (1169-1193) taken at t= 0, 15 min and after overnight incubation were analysed using HPLC-PDA with the acetonitrile method.

In addition to the assays above, the H154N and H154N/A155L/A156S mutant HCTs were incubated with various acyl donor and acceptor moieties. In this case, reaction mixtures consisting of 10 mM 5-CQA/ 5-FQA and 0.2 mM CoA/ caffeoyl-CoA/ feruloyl-CoA/ coumaroyl-CoA in 0.1 M sodium phosphate pH 6.5 were prepared in a 200 µL final volume. Reactions were started by adding 0.5 µM native or mutant HCT. Samples (1194-1273) taken at t= 0, 60 min and after overnight incubation were analysed using HPLC-PDA with the acetonitrile method.

2.4. Structural characterisation

2.4.1. Introduction

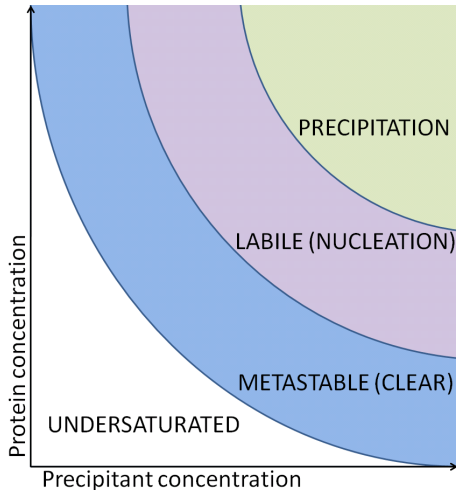
X-ray crystallography is the most powerful method for determining the 3D structure of biological macromolecules (Ilari *et al.*, 2008). However, its success is dependent on the growth of diffraction-quality single crystals. The homogeneity, stability and solubility of a protein are strongly correlated with its probability to crystallise. Optimisation of these properties can thus improve success rates. A number of variables affect crystallisation potential, including pH, ionic strength, additives, substrates, precipitants and the protein concentration. Temperature is also a significant variable in crystallisation experiments as it can influence nucleation and crystal growth by manipulating the solubility of the protein in the crystallisation solution. Thus, a vast number of conditions are usually screened before a promising crystallisation condition can be identified.

The screening phase can be accelerated by high-throughput techniques using robotic systems, but the crystallisation of macromolecules still remains a trial-and-error process. The EMBL-Grenoble HTX laboratory (https://embl.fr/htxlab/) provides for the extensive

screening of crystallisation conditions with limited amounts of sample. Typically, 100 nL of protein solution at the desired concentration and 100 nL of crystallisation solution are dispensed, using a Cartesian robot, into 96-well Greiner Crystal Quick plates, requiring about 100 µL of protein to screen 576 conditions.

The vapour diffusion technique is the most commonly used method for protein crystallisation (McPherson, 1982). In this method, the protein solution and crystallisation solution (containing precipitants, buffers and additives) are mixed and the resulting drop is equilibrated against a reservoir containing the crystallisation solution. Crystal growth can occur during the water evaporation-mediated equilibration between the protein-containing drop and the reservoir. A crystallisation phase diagram (Figure 38) is characterised by: a stable zone (undersaturated); a metastable zone (supersaturated), where nucleation cannot occur but crystals can grow from seeds; a labile zone (supersaturated), where spontaneous nucleation and crystal growth can occur; and a precipitation zone. Seeding procedures with preformed crystalline material usually allows larger single crystals to grow in the metastable zone. By performing serial dilutions from a concentrated seed stock, the number of crystals grown in the drop can also be controlled.

Figure 38: Schematic representation of a protein solid-liquid phase diagram



The high quality X-rays produced by synchrotron-radiation sources, such as the ESRF, in combination with the automation available on synchrotron macromolecular crystallography (MX) beamlines (http://www.esrf.eu/UserAndScience/Experiments/MX) enables the optimal X-ray data to be collected on relatively small crystals. However, such high intensity sources often result in radiation damage of the sample (Ravelli *et al.*, 2006). One way to mitigate this

is cryo-cooling, so most MX experiments are now carried out at cryogenic temperatures (Garman *et al.*, 2007).

Different methods exist to solve a crystal structure. The most common is molecular replacement (MR), which requires a model structure with a homologous sequence or a very similar fold to be available (Evans *et al.*, 2008). MR involves two fundamental steps: a rotation function and a translation function. If a homologous structure is not available, an experimental phasing method must be used. This generally requires the attachment of heavy atoms to the protein molecules in the crystal either via co-crystallisation, soaking or substitution of methionine residues by selenomethionine. Isomorphous replacement (IR) (Blundell and Johnson, 1976) or single/ multiple-wavelength anomalous diffraction (SAD/ MAD) phasing methods can then be applied (Hendrickson *et al.*, 1989). After a first set of protein phases is obtained and an electron density map is calculated by the Fourier summation. The map quality can be improved through density modification techniques (e.g. solvent flattening and non-crystallographic averaging) (Carter *et al.*, 1997; Terwilliger, 1999; Terwilliger, 2000). The next step is the interpretation of the resulting map in terms of the polypeptide chain. The resulting model is then refined in order to minimise the difference between the observed structure factor (F_{obs}) and those calculated from the model (F_{calc}).

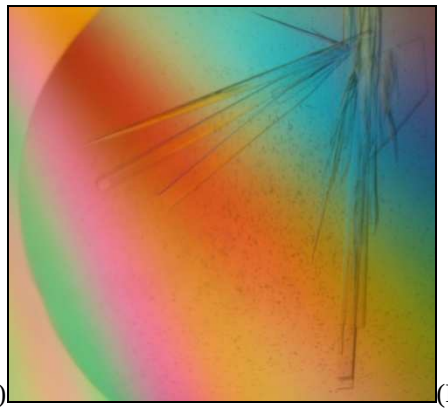
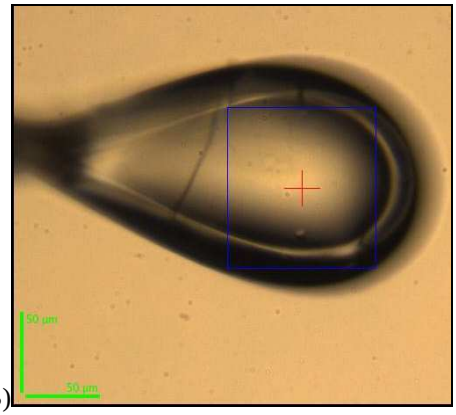
2.4.2. Tobacco 4CL2 crystallisation and structure determination

2.4.2.1. Crystallisation of Nt4CL2

An initial crystallisation screening was perfomed at 20 °C using the Cartesian robot of EMBL-Grenoble HTX laboratory with 100 nL protein solution at 5, 10 or 20 mg/mL plus 100 nL crystallisation solution per drop, and 100 µL reservoir solution. The 576 different crystallisation solutions corresponded to Hampton Research Crystal Screen I & II, Crystal Screen Lite & PEG/Ion, MembFac & Natrix, QuickScreen & Grid screens Ammonium Sulfate, Sodium Malonate, Sodium Formate, Grid screens PEG 6K, PEG/LiCl, MPD, Mme and Index Screen. A number of crystallisation hits were obtained with different crystallisation solutions. The morphology of the crystals was always similar (thin plates) but these varied in their multiplicity. The best diffracting crystals were obtained with condition n° 28 of Crystal Screen Lite (Hampton) composed of 0.2 M sodium acetate trihydrate, 0.1 M sodium cacodylate pH 6.5 and 15 % (w/v) PEG 8000. These crystals were reproduced manually in 2 µL drops with an optimal protein concentration of 12 mg/mL (Figure 39A). Several ligands (Mg-ATP, CoA and coumaric acid) individually or in combination were also incubated with the protein before setting up the crystallisation drops.

Figure 39: Crystals of tobacco 4CL isoform 2

(A) Crystallisation drop; (B) Crystal mounted in a nylon loop for data collection.

(A)  (B) 

2.4.2.2. *Data collection and X-ray analysis of apo and complex Nt4CL2 crystals*

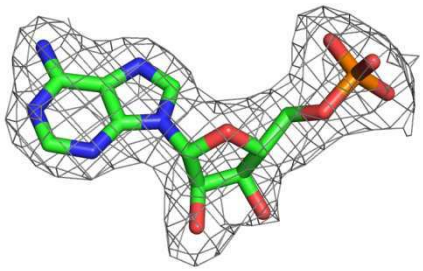
Large *Nt*4CL2 crystals were obtained by co-crystallisation with 2 mM ATP, 2 mM $MgCl_2$, 1 mM CoA and 1 mM coumaric acid. Prior to flash-freezing, the crystals were transferred to a reservoir solution containing 20 % glycerol, 2 mM ATP, 2 mM $MgCl_2$, 1 mM CoA and 1 mM coumaric acid and mounted in a nylon loop (Figure 39B). The best crystal diffracted to 2.5 Å and a complete diffraction data set was collected on ESRF beamline ID14-4 (McCarthy *et al.*, 2009). All data were processed and scaled with the XDS Suite (Kabsch, 2010). Data collection statistics are reported in Table 8. The crystal belongs to the orthorhombic space group ($P2_12_12_1$), with two molecules in the asymmetric unit (AU). A partial solution was obtained with the automated MR system BALBES (Long *et al.*, 2008) using the N-terminal domain of firefly luciferase (PDB: 3IEP, $P4_1$ form) as a search model. Firefly luciferase and *Nt*4CL2 share a 20 % identity at the amino acid level. The partial solution consisting of the N-terminal domain of *Nt*4CL2 was improved with several rounds of manual building in COOT (Emsley *et al.*, 2004) and refinement using REFMAC5 (Murshudov *et al.*, 1997). However, no electron density for the C-terminal domain was visible.

An additive screen (Hampton Research) was carried out by Virginie Carbonell with the same protein concentration range and crystallisation conditions. The addition of 0.1 M strontium chloride improved the crystal morphology and quality, as they diffracted to a higher resolution compared to the initial crystals (Table 8). Crystals of apo-*Nt*4CL2 obtained with strontium belong to the same space group but have a smaller unit cell and contain only one molecule in the AU. The previously built N-terminal domain of *Nt*4CL2 (see above) and the

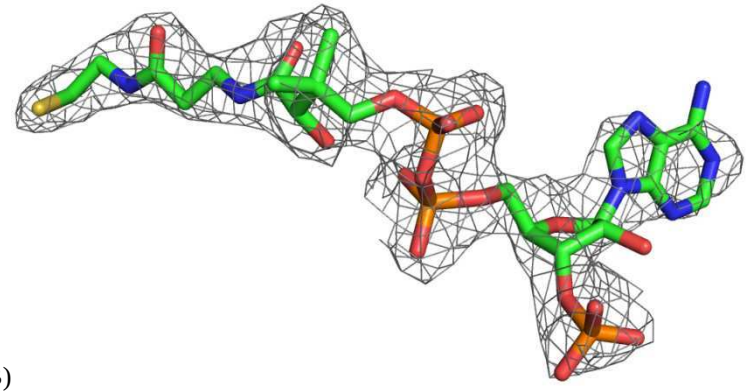
C-terminal domain of firefly luciferase (PDB: 1LCI, $P4_12_12$ form) were used in PHASER (McCoy *et al.*, 2007) as search models for structure determination by MR. The resulting model was improved using COOT and REFMAC5 and refined to a final crystallographic R_{cryst} and R_{free} of 21 and 24 % respectively with good stereochemical geometry as defined by Molprobity (Table 8). Only Ile11, which is poorly ordered in the final maps, is an outlier in the Ramachandran plots.

The refined structure of the C-terminal domain of apo-*Nt*4CL2 was subsequently used as a search model in PHASER for its placement in the crystal form obtained by co-crystallisation with substrates and cofactors (see above). Iterative rounds of manual building and refinement including NCS restraints with REFMAC5 resulted in a model with R_{cryst} and R_{free} of 22.2 and 27.8 % respectively. Again, the final model has good stereochemical geometry, except for Ile11. Table 8 contains a summary of refinement statistics. Calculation of difference Fourier maps (F_{obs}-F_{calc}) showed discernable density for CoA and AMP in both molecules of the AU, so they were included in the model (Figure 40). It is most likely that ATP, which was added in both crystallisation and cryo-protection solutions, was degraded to AMP and PPi, and only AMP was bound in the crystal.

Figure 40: Electron density for AMP (A) and CoA (B), contoured at 1.0 sigma, in the *Nt*4CL2 ternary complex crystal structure



(A)



(B)

Table 8: Data collection, processing and refinement statistics for *Nt*4CL2

Crystal name	*Nt4CL2-CoA-AMP*	*apo-Nt4CL2*
Data collection		
X-ray source	ID14-4	ID29
Wavelength (Å)	0.9393	0.9796
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell dimensions (Å)	*a*= 92.53 *b*= 110.68 *c*= 112.99	*a*= 54.36 *b*= 102.55 *c*= 110.40
Monomers/AU	2	1
Matthews coefficient ($Å^3Da^{-1}$)	2.39	2.54
Solvent content (%)	48.6	51.7
Resolution range (Å)	50-2.5 (2.7-2.5)	50-1.8 (2.0-1.8)
Observed reflections (I/σ(I)>0)	162500	215460
Unique reflections	40412	52430
Completeness (%)	99.0 (98.6)	96.1 (95.9)
R_{merge} (%)	9.5 (65.5)	6.1 (45.0)
<I/σ(I)>	11.2 (2.2)	13.6 (3.5)
Wilson B-value ($Å^2$)	52.7	27.3
Model quality indicators		
R_{cryst} (%)	22.2	21.1
R_{free} (%)	27.8	23.4
Mean B value ($Å^2$)	38.5	27.6
RMSD bonds (Å)/ angles (°)	0.017/ 1.8	0.024/ 2.0
Refined model composition		
Amino acids	1028	532
CoA	2	-
AMP	2	-
Glycerol	-	1
Ramachandran analysis		
Allowed regions (%)	95.5	96.9
Generously allowed regions (%)	4.2	2.9
Disallowed regions (%)	0.3	0.2

Values in parentheses are for the outermost resolution shell.

$R_{merge} = \sum_h \sum_i \left| I_{h,i} - \langle I \rangle_h \right| / \sum \left| \langle I \rangle_h \right|$ was calculated for the whole data set.

$R_{cryst} = \sum \left| |F_{obs}| - |F_{calc}| \right| / \sum |F_{obs}|$

R_{free} was calculated as for R_{cryst} with 5 % of the data omitted from the refinement process.

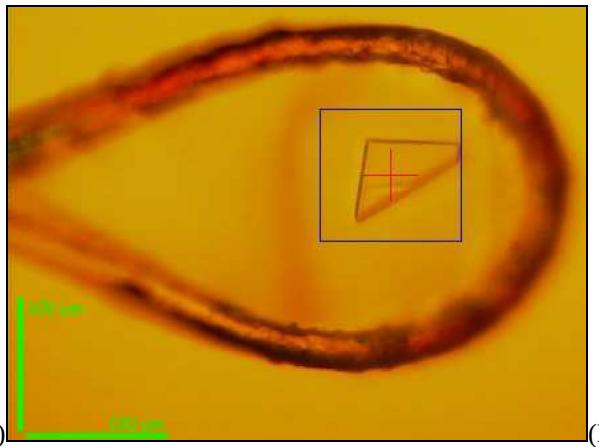
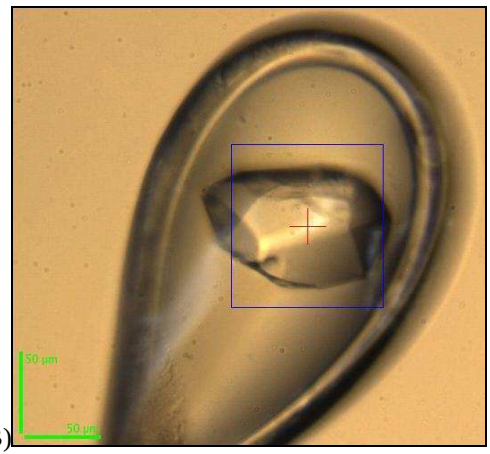
2.4.3. Coffee HCT crystallisation and structure determination

2.4.3.1. Crystallisation of native CcHCT

*Cc*HCT was purified to homogeneity, concentrated and submitted to the EMBL-Grenoble HTX laboratory for crystallisation trials. Preliminary investigations were performed, in which the temperature was varied (20/ 4 °C), the protein concentration was varied (5 to 50 mg/mL) and substrates (5-CQA, CoA, shikimic acid, caffeic acid) were added or not. Microcrystals were obtained in several conditions, including: 15 mg/mL HCT with 1.5 M sodium malonate pH 7.0; 8 mg/mL HCT with 0.056 M sodium phosphate and 1.344 M potassium phosphate pH 8.2; 25 mg/mL HCT with 0.1 M magnesium formate and 15 % PEG 3350. Unfortunately, crystal size was not sufficient to perform diffraction tests and the conditions could never be optimised. Modified HCT samples were also tried. For example, *in situ* proteolysis was attempted with 1/1000 (w/w) trypsin added to the reservoir solution before setting up the drop. The limited proteolysis with trypsin was also scaled-up (2.1.3.6) and submitted for crystallisation trials. The purest fractions of methylated HCT (2.1.3.7) were pooled and sent for crystallisation trials at 18 and 9 mg/mL. Unfortunately, none of these crystallisation trials were ever successful.

Crystals also appeared after several weeks at 20 °C in a 200 nL drop with HCT at 20 mg/mL in condition n° 62 of Crystal Screen Classics Suite (Qiagen), containing 1.6 M magnesium sulfate in 0.1 M MES pH 6.5. They were large enough for X-ray measurements and so were fished directly from the Greiner plates with a nylon loop and transferred to a mother liquor containing the crystallisation solution supplemented with 20 % glycerol prior to being flash-frozen in liquid nitrogen (Figure 41A). Despite many efforts, these crystals could not be reproduced with subsequent protein batches.

Figure 41: Native (A) and K-mutant (B) *Cc*HCT crystals mounted in a nylon loop

(A)  (B) 

2.4.3.2. Native CcHCT data collection and X-ray analysis

Data sets to a maximum resolution of 3.0 Å were collected from three similar crystals at the highly automated microfocus beamline ID23-2 (Flot *et al.*, 2010). The crystals belong to space group $P4_22_12$ and contain two molecules per AU. All data were processed and scaled using the XDS Suite (Kabsch, 2010). A summary of data collection statistics is given in Table 9. The native HCT crystal structure was solved by MR using PHASER (McCoy *et al.*, 2007) and polyalanine VS (PDB: 2BGH) as a search model. RESOLVE (Terwilliger, 2003) was used to improve these initial phases. Several rounds of manual building in COOT (Emsley *et al.*, 2004) and restrained refinement with REFMAC5 (Murshudov *et al.*, 1997) were subsequently carried out. Non-crystallographic symmetry (NCS) restraints (Kleywegt, 1996), as well as bulk solvent modelling and TLS refinement were applied during refinement. The native HCT model was refined to a final R_{cryst} and R_{free} of 19 and 25 % respectively. Table 9 contains a summary of refinement statistics.

2.4.3.3. Crystallisation of the K-mutant HCT

Crystallisation experiments with 10 and 20 mg/mL of K-mutant HCT resulted in a single hit out of 576 conditions tested. This condition, corresponding to solution n° 6 of Crystal Screen Lite kit (Hampton Research), consisted of 0.2 M magnesium chloride hexahydrate, 0.1 M Tris-HCl pH 8.5 and 15 % (w/v) PEG 4000. The initial crystals obtained with 10 mg/mL protein were fished and cryo-protected with a mother liquor containing 20 % glycerol. However, the diffraction patterns contained more than one lattice. The size and diffraction quality of the crystals was successfully improved using micro-seeding techniques resulting in 3D crystals (Figure 41B). Here, protein crystals were crushed in an Eppendorf tube containing a small bead and used in fresh crystallisation drop setups to obtain larger single crystals. The K-mutant HCT was incubated with 1 to 5 mM CoA, 1 to 5 mM shikimic acid or 1 mM 5-CQA for co-crystallisation experiments.

2.4.3.4. Data collection and X-ray analysis of the K-mutant HCT crystals

Several data sets were collected from crystals of K-mutant HCT. The best diffracted to a maximum resolution of 1.7 Å and belong to the $P2_12_12_1$ space group with two molecules per AU. Occasionally, an alternative space group, $C222_1$, was observed in crystals grown in the same condition from robot plates. These crystals diffracted to 2.5 Å resolution. All data were processed and scaled using the XDS suite (Kabsch, 2010). A summary of data collection statistics is given in Table 9. The structures of both crystal forms were solved by MR using PHASER (McCoy *et al.*, 2007) and a partially refined native HCT structure (see 2.4.3.2) as a

search model. Multiple rounds of manual building (COOT) and refinement (REFMAC5) resulted in a R_{cryst} of 17 % and R_{free} of 20 % for form 1 ($P2_12_12_1$), and 20 and 26 % for form 2 ($C222_1$), with good geometry as defined by Molprobity (Davis *et al.*, 2007; Chen *et al.*, 2010). Refinement statistics are summarised in Table 9.

Table 9: Data collection, processing and refinement statistics for the native and K-mutant HCTs

R_{cryst} and R_{free} values were calculated as defined in Table 8.

Crystal name	*native HCT*	*K-mutant form 1*	*K-mutant form 2*
Data collection			
X-ray source	ID23-2	ID14-4	ID14-4
Wavelength (Å)	0.873	0.939	0.976
Space group	$P4_22_12$	$P2_12_12_1$	$C222_1$
Unit-cell dimensions (Å)	*a*= *b*= 116.1 *c*= 158.9	*a*= 63.55 *b*= 116.48 *c*= 118.04	*a*= 76.55 *b*= 96.28 *c*= 118.76
Molecules/AU	2	2	1
Matthews coefficient ($Å^3Da^{-1}$)	2.79	2.26	2.26
Solvent content (%)	55.9	45.7	45.5
Resolution range (Å)	45–3.0 (3.2-3.0)	50-1.7 (1.8-1.7)	50-2.5 (2.6-2.5)
Observed reflections ($I/\sigma(I)>0$)	72116	674662	55469
Unique reflections	22073	92179	14082
Completeness (%)	99.0 (98.4)	95.0 (83.9)	90. 5 (79.2)
R_{merge} (%)	16.6 (56.6)	9.6 (53.3)	7.1 (42.7)
$<I/\sigma(I)>$	8.5 (3.1)	12.7 (2.8)	16.8 (3.5)
Wilson-B value ($Å^2$)	44.7	23.3	47.2
Model quality indicators			
R_{cryst} (%)	0.19	0.17	0.20
R_{free} (%)	0.25	0.20	0.26
Mean B value ($Å^2$)	11.6	19.8	34.0
RMSD bonds (Å)/ angles (°)	0.014/1.5	0.026/2.0	0.015/1.5
Refined model composition			
Amino acids	844	837	419
Water (Cl^-)	29	634 (2)	27
SO_4^{2-}	5	-	-
glycerol	-	3	-
Ramachandran analysis			
Allowed regions (%)	96.5	98.3	95.4
Generously allowed regions (%)	3.3	1.7	4.6
Disallowed regions (%)	0.2	-	-

2.4.4. Homology-modelling of *CcHQT*

Multiple crystallisation attempts using varied concentrations of HQT (8, 15 and 24 mg/mL) at 20 °C with or without substrates were not successful. The K210A/K219A mutant HQT was also screened unsuccessfully. A sequence alignment generated using ClustalW revealed two differences in the polypeptide chain length of *Cc*HCT and *Cc*HQT: insertion of Arg223 in HQT and deletion of NTISY 248-252 from HCT. These residues are located in the predicted linker region connecting the two domains of BAHD enzymes. Consequently, segments 1-210 and 253-434 of the native HCT structure (chain A) were used as templates to model HQT N- and C-terminal domains using the automated modelling program Swiss-Model (Arnold *et al.*, 2006).

2.5. Docking calculations

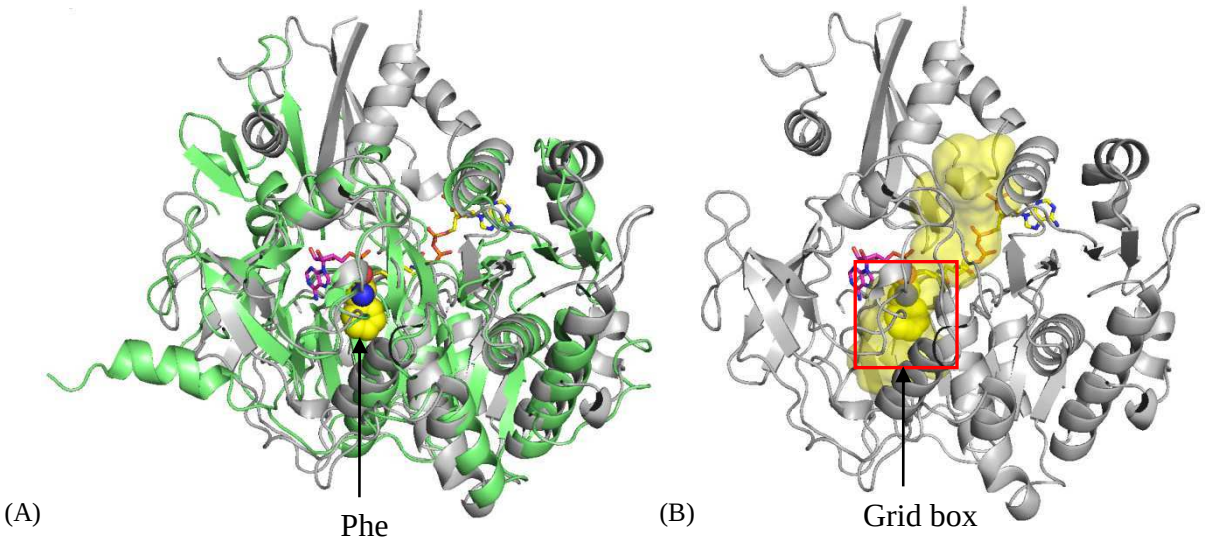
2.5.1. Docking of hydroxycinnamic acids in 4CL2-CoA-AMP

2.5.1.1. Prediction of the phenolic substrate binding site

Despite many attempts, no crystal structure of *Nt*4CL2 in complex with coumaric acid could be obtained. Docking analyses were therefore carried out to identify residues potentially involved in HCA binding. A pocket detection algorithm, Pocket-Finder (Hendlich *et al.*, 1997), was used to predict the substrate binding pocket (SBP) in the *Nt*4CL2-CoA-AMP ternary complex crystal structure. The 1.6 Å probe radius used found a major site of volume 966 $Å^3$ that was too large to accommodate only HCA substrates (Figure 42). The structure of PheA-ATP-phenylalanine (PDB: 1AMU) was therefore used to locate the SBP in the *Nt*4CL2-CoA-AMP crystal structure based on the phenylalanine binding site. PheA is a well-suited template because of the structural similarity between phenylalanine and cinnamic acid derivatives. The SBP could be clearly defined as part of the N-terminal domain of *Nt*4CL2 with CoA thiol and AMP phosphate group at the entry of the pocket (Figure 42A). *Nt*4CL2 residues within 4 Å of the predicted SBP are: Gln213, Leu221, Ile238, Tyr239, Ser243, Val244, Pro279, Met306-Ala309, Gly330-Thr336, Pro340, Val341 and Phe348.

Figure 42: Superimposition of *Nt*4CL2-CoA-AMP with PheA bound to phenylalanine for the localisation of the SBP by Pocket Finder

(A) *Nt*4CL2-CoA-AMP is coloured in grey with CoA in magenta. The predicted SBP is shown as a yellow surface. Phenylalanine extracted from PheA (PDB: 1AMU) is represented as yellow spheres. (B) Definition of the search space in *Nt*4CL2-CoA-AMP in AutoDock Tools.



2.5.1.2. Docking calculations

The search space for the docking calculations with CGAs was centred on the predicted SBP in the structure of *Nt*4CL2-CoA-AMP chain A with a box size of 12 x 12 x 14 $Å^3$ (Figure 42B). Coordinate files and dictionaries of the potential ligands (cinnamic, coumaric, caffeic, ferulic and sinapic acids) were generated using the PRODRG2 Server (http://davapc1.bioch.dundee.ac.uk/prodrg/) (Schuttelkopf *et al.*, 2004) including all hydrogen atoms. AutoDock Tools (Goodsell *et al.*, 1996) was then used to prepare the receptor and ligand PDBQT files for AutoDock Vina (Trott *et al.*, 2010). The exhaustiveness parameter was set to 16 (twice the default value) to maximise the probability of finding the global scoring function minimum. The docking analyses resulted in a list of the ligands ordered by binding affinity from lowest to highest. The result file was split into individual PDB files using the vina_split algorithm. The results of the calculations are shown in Table 10, in which, for each ligand docked, the highest and lowest binding affinity values are given.

Table 10: Binding affinity of docked ligands in the *Nt*4CL2-CoA-AMP receptor as calculated by AutoDock Vina

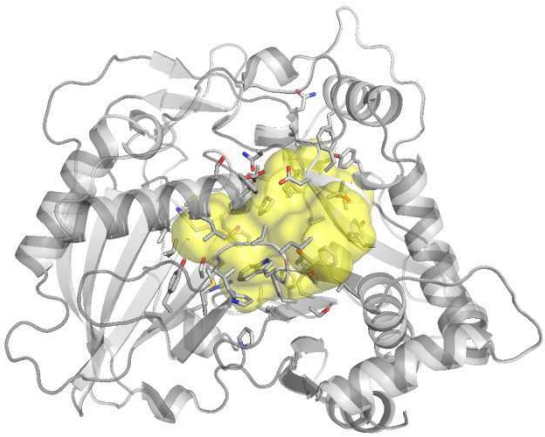
	Binding affinity values (kcal/mol)		
Ligand	*highest*	*lowest*	*# solutions*
Cinnamic acid	-6.8	-6.0	10/10
Coumaric acid	-7.0	-6.5	10/10
Caffeic acid	-7.4	-5.8	10/10
Ferulic acid	-7.0	-5.5	9/10
Sinapic acid	-6.2	-3.6	8/10

2.5.2. Docking of substrates in HCT

2.5.2.1. Prediction of the substrate binding site in HCT

Pocket-Finder (Hendlich *et al.*, 1997) was used to predict the location of the SBP in the native *Cc*HCT crystal structure chain A, which most likely adopts a functional conformation (see Results section). The 1.6 Å probe radius used found a major site (volume~ 1850 $Å^3$) corresponding to the solvent channel located at the domain interface (Figure 43). The residues lining the solvent channel in HCT were identified. They belong to the following sequence regions: Asn26-Tyr40, Pro110, Arg115, Val149-Ile165, Tyr203, Tyr255, Tyr281-Asp285, Arg289, Asn301-Thr305, Leu331, Leu346, Leu355-Phe362, Gly368-Ile377, Met390-Tyr397 and Leu400-Phe402. A number of these residues are conserved between the HCT and HQT enzymes, while some are not. These differences may contribute to the different substrate specificities of HCT and HQT.

Figure 43: Prediction of the substrate binding pocket in the HCT crystal structure by Pocket-Finder



2.5.2.2. AutoDock calculations for HCT

AutoDock Vina (Trott *et al.*, 2010) was used for docking various biological ligands in HCT. A box centred on the macromolecule with a size of 20 x 20 x 20 $Å^3$ was set as the search space, so that it covered the predicted SBP (Figure 43). A list of potential ligands was established comprising 5-CQA, 5-FQA, 5-*O*-caffeoylshikimic acid (5-CSA), 3,5-diCQA, CoA, caffeoyl-CoA, quinic, shikimic, caffeic and ferulic acids. The same procedure as for *Nt*4CL2 was followed for the docking calculations. Preliminary results for CoA showed that the cofactor does not adopt an extended conformation when docked alone in the HCT receptor. As the CoA binding site is well-conserved in MaT and TRI101 crystal structures, CoA bereft of the $(CH_2)_2$-SH group was extracted from TRI101-CoA-T2 (PDB: 2RKV) and placed into the structure of HCT. This HCT-CoA model was then used as a receptor for subsequent docking of the other ligands (Table 11). Once a preferential binding site for quinic/ shikimic acid was identified, HCT-shikimic acid was set as a receptor. Caffeic acid and caffeoyl-CoA were subsequently docked to refine the CoA and acyl binding sites. The binding affinity values for caffeic acid were between -6.4 and -5.2 kcal/mol, and for caffeoyl-CoA between -8.3 and -6.3 kcal/mol.

Table 11: Binding affinity of 10 docked ligands in the HCT-CoA receptor as calculated by AutoDock Vina

	Binding affinity values (kcal/mol)	
Ligand	*Highest*	*Lowest*
Quinic acid	-5.7	-4.3
Shikimic acid	-5.5	-4.6
Caffeic acid	-6.1	-4.9
Ferulic acid	-6.0	-5.1
5-CQA	-7.9	-6.8
5-CSA	-8.2	-7.1
5-FQA	-7.8	-6.9
3,5-diCQA	-9.7	-7.8

Chapter 3. Results

3.1. Introduction

The production and purification of sufficient quantities of *Cc*HCT, *Cc*HQT, *Cc*4CL2 and *Nt*4CL2 allowed for a detailed characterisation of these acyltransferases and CoA ligases. Here, some results of biochemical assays performed on *Cc*HCT, *Cc*HQT, *Nt*4CL2 and some *Cc*HCT mutants are presented. Crystallisation trials resulted in crystals of the native and K-mutant HCTs and native *Nt*4CL2. These crystal structures were then solved by MR. *Nt*4CL2 was crystallised in two forms, an apo form and a ternary complex with CoA and AMP bound. Docking of HCAs was carried out in the predicted SBP of the *Nt*4CL2-CoA-AMP ternary complex crystal structure in order to identify the potential residues involved in catalysis and/ or substrate binding. The overall structure of *Cc*HCT is described and the native and double mutant structures compared. This is followed by a comparison with the crystal structures of other BAHD acyltransferases. As crystals of *Cc*HCT in complex with its biological substrates could not be obtained, the results of ligand docking experiments using AutoDock Vina are described. A homology-model based on the crystal structure of native *Cc*HCT was constructed and used to compare *Cc*HCT and *Cc*HQT active sites.

3.2. Protein over-expression, purification and characterisation

3.2.1. Coffee and tobacco 4-coumarate CoA ligases

During these investigations of CGA biosynthesis in coffee, the need to generate hydroxycinnamoyl-CoA thioesters was encountered. In a number of biological systems, CoA activation can facilitate the transfer of acylated moieties by acyltransferases. Due to their limited stability, hydroxycinnamoyl-CoA thioesters are not commercially available. Initially, caffeoyl-CoA was synthesised from 5-CQA and CoA using *Cc*HQT in the reverse reaction. But due to the reversibility of the acyltransfer reaction, only a small amount of CoA thioester was produced. As this procedure was material- and time-consuming, using a 4-coumarate CoA ligase (EC 6.2.1.12) that efficiently catalysed the formation of hydroxycinnamoyl-CoA thioesters was considered. The complete cDNA encoding *Cc*4CL2 was obtained from Nestlé R&D Tours and the recombinant protein produced and purified. It was most likely incorrectly folded as aggregation occurred, and it was inactive towards the biological substrate, coumaric acid. A synthetic gene encoding *Nt*4CL2 was then ordered (Beuerle *et al.*, 2002). This protein

was purified using several chromatographic steps with typical yields of 10 mg/L culture and used for further biochemical and structural studies.

3.2.2. Coffee HCT and HQT hydroxycinnamoyltransferases

Over-expression of these target proteins required the sub-cloning from the pGTP vector into pPROEX_HTb for *hct* and pET-28M-SUMO3 for *hqt*, as well as the screening of a number of different *E. coli* host strains. High yields of N-terminal His_6-tagged *Cc*HCT were obtained with BL21* (DE3) pLysS induced at 20 °C (4.5 mg/L culture). *Coffea canephora* cDNAs contain a high number of rare *E. coli* codons. However, a comparison between the wild-type *Cchct* gene and a codon-optimised gene indicated that the protein expression levels were comparable. The synthetic gene was used in all subsequent protein production, for which an optimised protocol is described in section 2.1.3.5. *Cc*HQT was expressed in BL21 (DE3) cells in fusion with a His_6-SUMO partner, but lower yields were obtained (~1 mg/L culture). HCT and HQT were purified by immobilised metal affinity chromatography (IMAC). The fusion partner or His_6-tag were subsequently cleaved using specific proteases. *Cc*HCT (residues 1-434) plus a five amino acid N-terminal extension (GAMGS) and HQT (1-430) plus a serine introduced at its N-terminus were recovered by subtractive chromatography. Ion-exchange and size-exclusion chromatography were carried out in order to remove all remaining contaminants. Both *Cc*HCT and *Cc*HQT exist as monomers of 48 kDa in solution and can be concentrated up to 50 mg/mL.

The native enzymes were partially degraded into two proteolytic fragments of ~25 kDa that appeared during storage at 4 °C and amounts of which increased with time. This may be due to protease contamination from the *E. coli* extract. Limited proteolysis experiments also clearly showed two similar-sized fragments on SDS-PAGE although the fragments remain physically associated throughout size-exclusion chromatography. Mass spectrometry and N-terminal sequencing analysis showed that the protease cleavage site was located in the predicted cross-over loop region near two lysine residues. In order to increase the stability of the recombinant *Cc*HCT and *Cc*HQT, trypsin-proteolysis resistant mutants, K210A/K217A *Cc*HCT and K210A/K219A *Cc*HQT, were produced.

3.3. Biochemical analysis

3.3.1. Introduction

High-performance liquid chromatography (HPLC) is the preferred technique for the characterisation and analysis of plant secondary metabolites. HPLC was therefore used here for the biochemical characterisation of the enzymes produced. The experimental set up used is described in section 2.3.2. A shortage of acetonitrile occurred during the first half of 2009, so methanol was used to replace acetonitrile in the mobile phase and a new gradient elution protocol was developed to optimise CGA separation. A preparative HPLC procedure was developed for the purification of caffeoyl-CoA from enzymatic reactions. However, a simpler SPE method was ultimately used for the purification of coumaroyl-CoA, caffeoyl-CoA and feruloyl-CoA esters, which serve as substrates for HCT and HQT in the forward reactions leading to CGA biosynthesis.

3.3.2. Enzymatic synthesis and purification of CoA thioesters

3.3.2.1. Coffee and tobacco 4-coumarate CoA ligases

Enzymatic synthesis is an efficient way of producing large amounts of CoA-activated HCAs that are not commercially available. The well-characterised *Nt*4CL2 was produced using a published procedure (Beuerle *et al.*, 2002). With all substrates used, apart from sinapic acid, there is a high conversion of HCA to the relevant CoA thioester. The retention time and maximum absorbance values in absorbance spectra are presented in Table 12. Figure 44 shows the activity of purified recombinant 4CL2 towards cinnamic, coumaric, caffeic, ferulic and sinapic acids in the presence of CoA and Mg-ATP required for catalysis.

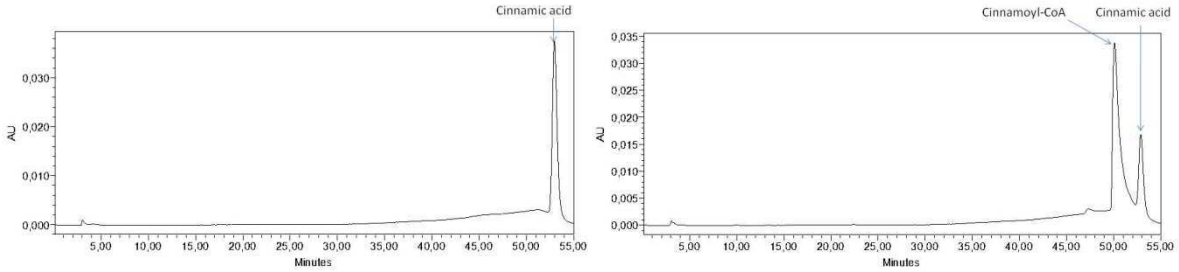
Table 12: Retention times and absorbance maxima of CoA thioester products

Compound	*RT (min)*		*Max. abs. (nm)*
	methanol	*acetonitrile*	
Cinnamoyl-CoA	50.1	-	306.0
Coumaroyl-CoA	43.1	42.4	333.3
Caffeoyl-CoA	34.4	34.6	348.9
Feruloyl-CoA	44.4	44.1	347.7
Sinapoyl-CoA	45.3	45.4	351.3

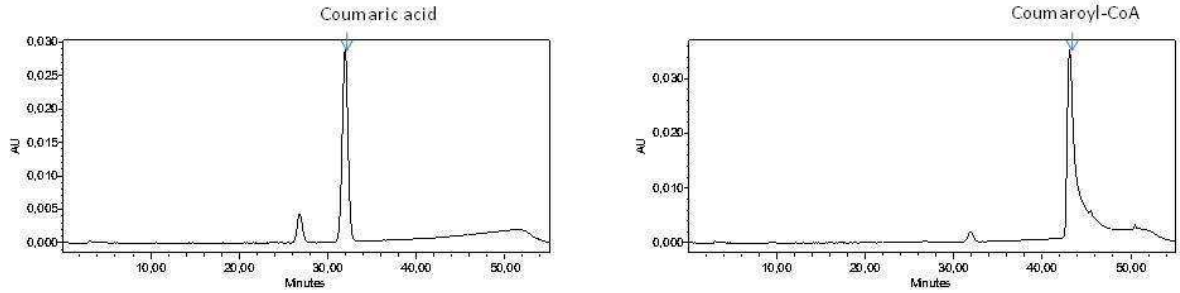
Figure 44: HPLC analysis of reactions comprising *Nt*4CL2 incubated with (hydroxy)cinnamic acids and CoA

Reactions were set up with 0.1 M Tris-HCl pH 7.5, 5 mM CoA and 1 mM of the various (hydroxy)cinnamic acids noted in each panel. Reactions were started by adding 1.6 µM of enzyme. Samples taken at the times indicated were analysed using HPLC-PDA with the methanol method. All chromatograms were extracted at λ= 325 nm except for cinnamic acid (A), which was measured at λ= 300 nm.

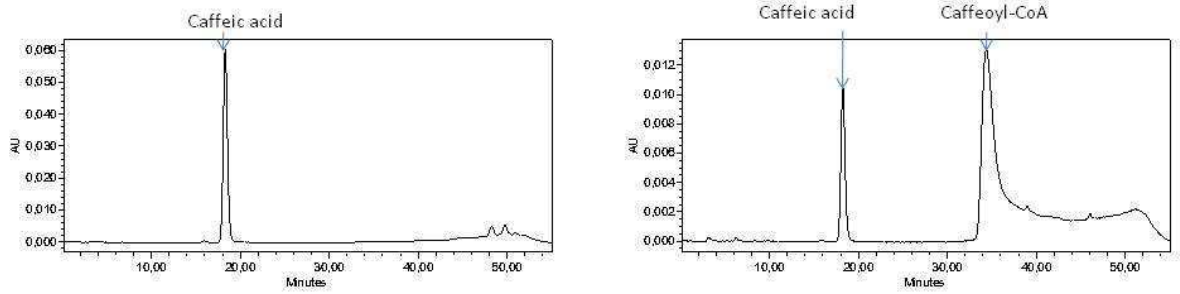
(A) Cinnamic acid t= 0 min (sample 598) t= 20 min (sample 600)



(B) Coumaric acid t= 0 min (sample 593) t= 20 min (sample 594)



(C) Caffeic acid t= 0 min (sample 590) t= 20 min (sample 592)



(D) Ferulic acid t= 0 min (sample 595) t= 20 min (sample 597)

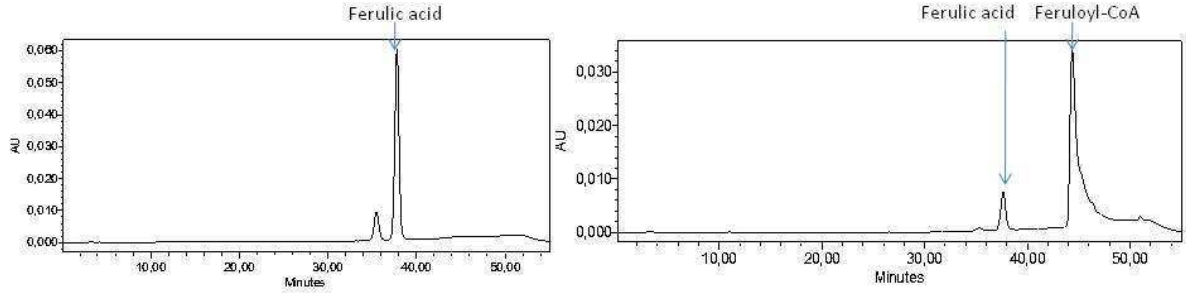
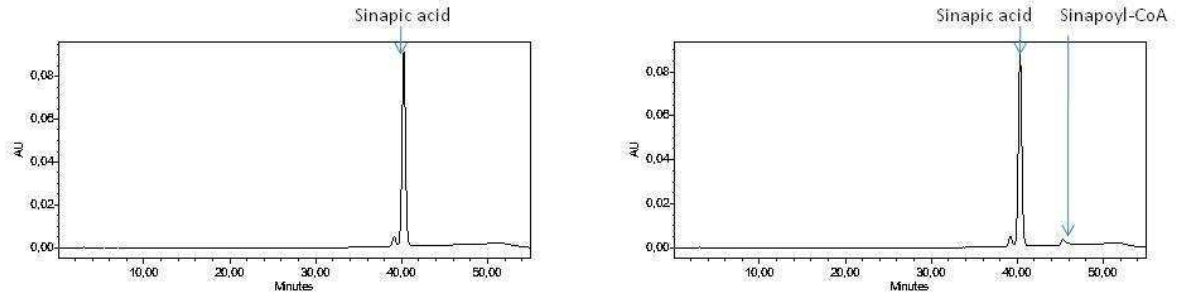


Figure 44 (continued)

(E) Sinapic acid t= 0 min (sample 601) t= 20 min (sample 603)



Hydroxycinnamoyl-CoA thioesters were characterised by their UV spectra (Figure 45). Later experiments show that, with the exception of sinapic acid, the conversion by *Nt*4CL2 of these substrates reached over 80 % after 60 min (Figure 46). During optimisation of the reaction conditions, no major difference was observed in product levels obtained at pH 6.0 or 7.5, or with sodium phosphate compared to Tris-HCl (data not shown). Therefore all subsequent reactions were carried out in 0.1 M sodium phosphate pH 6.0. The CoA esters could be produced almost free of their acid precursors by the addition of more enzyme. This further enabled the quantification of the CoA thioester product, assuming that one mole of HCA supplied in the reaction generated one mole CoA thioester.

Figure 45: Absorption spectra of the CoA thioester products measured using HPLC-PDA

(A) cinnamoyl-CoA; (B) coumaroyl-CoA; (C) caffeoyl-CoA; (D) feruloyl-CoA and (E) sinapoyl-CoA.

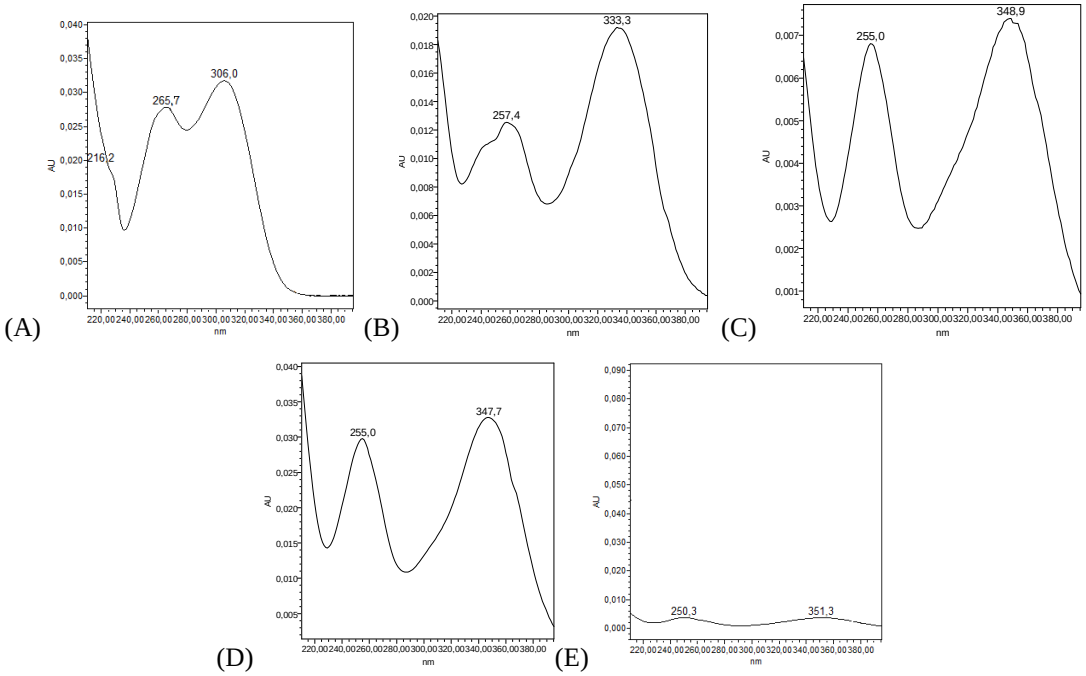
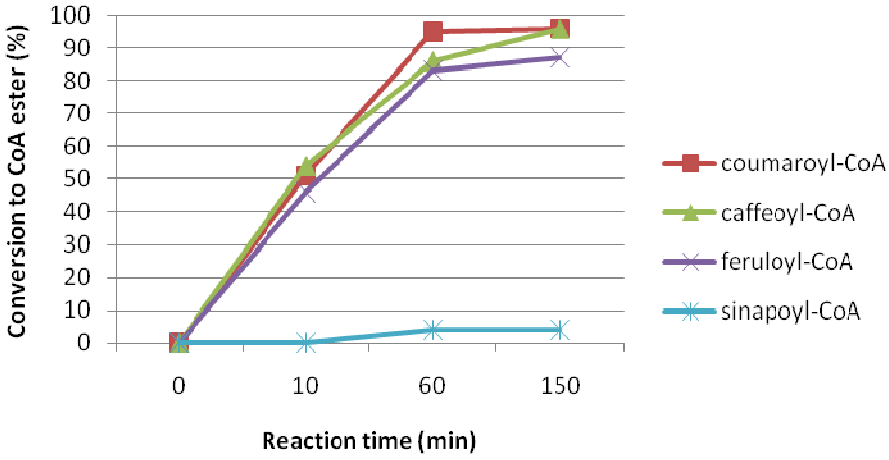


Figure 46: Levels of CoA thioester production by recombinant *Nt*4CL2 with various HCAs and CoA in the presence of Mg-ATP

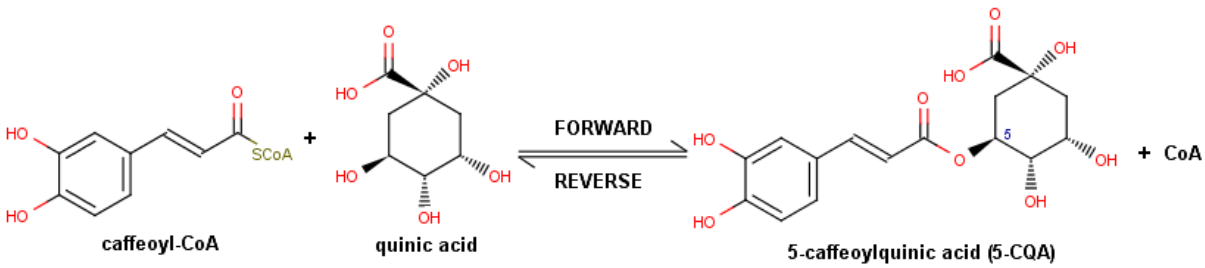
At t= 0, 0.05 µM *Nt*4CL2 was added to a 10 mL reaction mixture containing 0.8 mM coumaric/ caffeic or ferulic acid, 1 mM CoA, 2 mM $MgCl_2$ and 2 mM ATP in 0.1 M sodium phosphate pH 6.0. Samples (699-715) were taken at t= 0, 10, 60 and 150 min and analysed using HPLC-PDA with the methanol method. The conversion to CoA thioester in percentage was calculated from the HCA precursor peak area diminution.



3.3.2.2. Purification of CoA thioesters by preparative HPLC or SPE

The coumaroyl-, caffeoyl- and feruloyl-CoA thioesters produced were purified by preparative HPLC or solid phase extraction (SPE). Initially, preparative HPLC was used for purification. A 100-150 µL volume of reaction mixture was injected and the fractions containing the CoA ester were collected manually. This purification method was abandoned because it was time-consuming and the HPLC profiles obtained indicated a partial degradation of caffeoyl-CoA to caffeic acid. This was most likely due to the increase in relative amounts of phosphoric acid during concentration. A simpler Sep-Pak purification procedure was eventually devised. The enzyme was removed from the reaction mixture by filtration through a 50 kDa membrane before loading onto a Sep-Pak column. The CoA thioesters were bound to the C_{18} column using a 4 % ammonium acetate buffer, washed with milliQ water and eluted with methanol. The elution fractions were collected and analysed by HPLC. The fractions containing the CoA ester were lyophilised and resuspended in milliQ water before use.

3.3.3. Comparison of the HCT and HQT-catalysed forward reactions



3.3.3.1. Test of the acyl acceptor specificity

Coumaroyl-, caffeoyl- and feruloyl-CoA thioesters were produced using *Nt*4CL2 and purified almost free from their acid precursors. The reactivity of *Cc*HCT and *Cc*HQT towards these activated substrates was then tested. HCT and HQT transfer the acyl moiety between CoA and acceptor molecules such as quinic and shikimic acids. The precursor molecule for the major CGA compound, 5-CQA, is caffeoyl-CoA, and this compound may also be an important precursor in mono- and dicaffeoylquinic acid biosynthesis. Coumaroyl-CoA is the precursor for coumaroylquinic acids that are present only in trace amounts in coffee, while feruloyl-CoA is probably the precursor for 5-feruloylquinic acid (5-FQA). However, currently there is no experimental evidence in the literature demonstrating which enzyme is responsible for the synthesis of coumaroylquinic or feruloylquinic acids in plants.

3.3.3.2. Activity of CcHCT and CcHQT towards caffeoyl-CoA

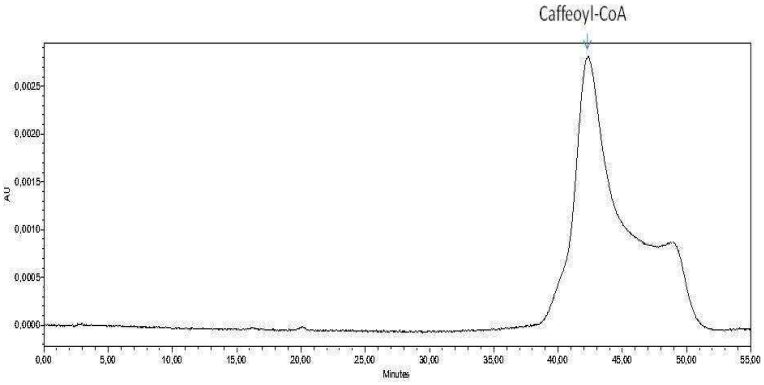
When *Cc*HCT was incubated with caffeoyl-CoA and quinic acid, 5-CQA (RT= 17.6 min) was formed. 76 % of the caffeoyl-CoA substrate had been consumed after 60 min respectively (Figure 47B). With shikimic acid as the acyl acceptor, all the caffeoyl-CoA present was transformed after only 5 min (Figure 47C) to a product (RT= 35.4 min) with a typical CGA absorbance spectrum (Figure 47D). The UV spectrum showed a maximum at λ= 322.6 nm and is similar to that of 5-CQA. As caffeoyl-CoA was used as a precursor, and the absorption spectrum of the novel compound shared some properties with that of 5-CQA, it was hypothesised that the novel compound resulted from a coupling of the caffeoyl moiety with shikimic acid via an ester linkage, presumably caffeoylshikimic acid (CSA). After 60 min, a major peak representing the same ester and two other peaks, which may correspond to the two other CSA isomers (RT= 25.9 and 32.2 min) were present (Figure 47D). When both quinic and shikimic acids were supplied to *Cc*HCT at the same concentration, a caffeoylshikimic acid isomer was formed after 5 min (Figure 47E); after 60 min, small peaks

corresponding to the two other CSA isomers were detected as well as a small peak for 5-CQA (Figure 47F). This result clearly demonstrates that *Cc*HCT strongly favours shikimic over quinic acid as the acyl acceptor molecule and is capable of producing more than one CSA isomer.

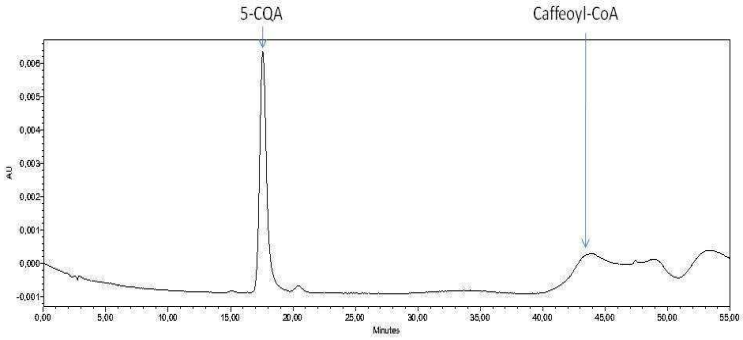
Figure 47: HPLC profiles of HCT reactions with caffeoyl-CoA and quinic/ shikimic acid

Reactions were set up with 0.1 M sodium phosphate pH 6.5, 0.2 mM caffeoyl-CoA and 5 mM quinic/ shikimic acid or both. Reactions were started by adding 1 µM enzyme. Samples taken at t= 5 and 60 min were analysed using HPLC-PDA with the methanol method. An overlap of the absorbance spectra for the putative CSA isomers is indicated in panel D (sample 1042).

(A) caffeoyl-CoA, t= 0 min (sample 1023)



(B) Quinic acid, HCT t= 60 min (sample 1041)



(C) Shikimic acid, HCT t= 5 min (sample 1027)

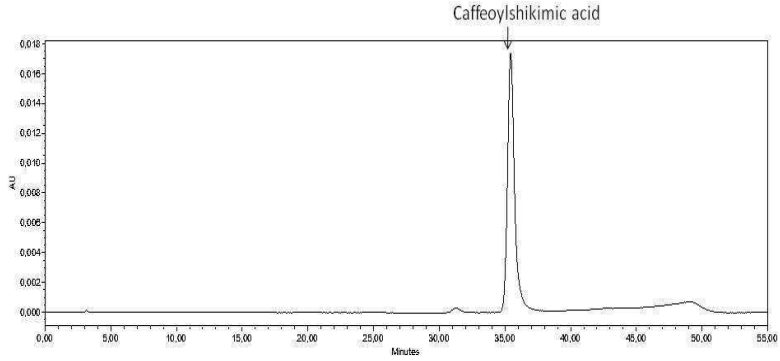
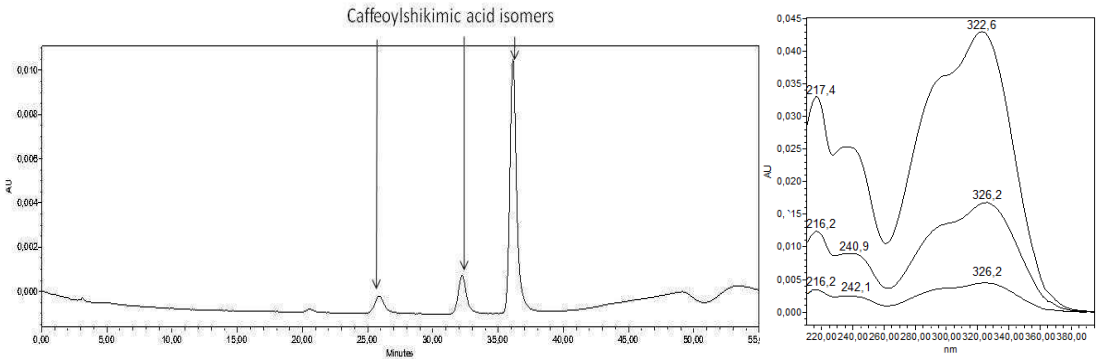
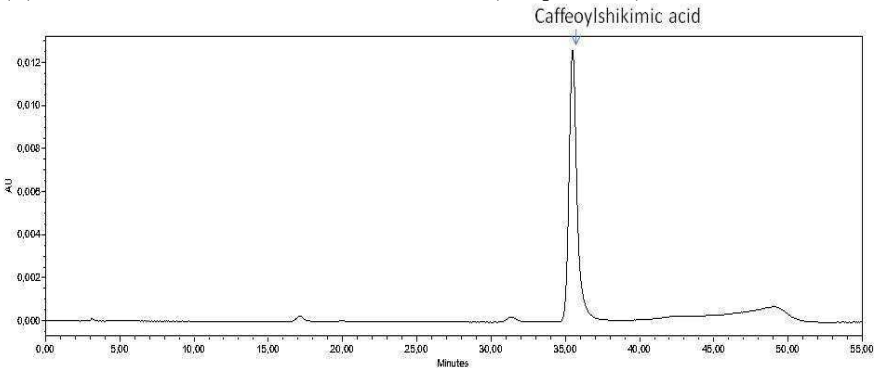


Figure 47 (continued)

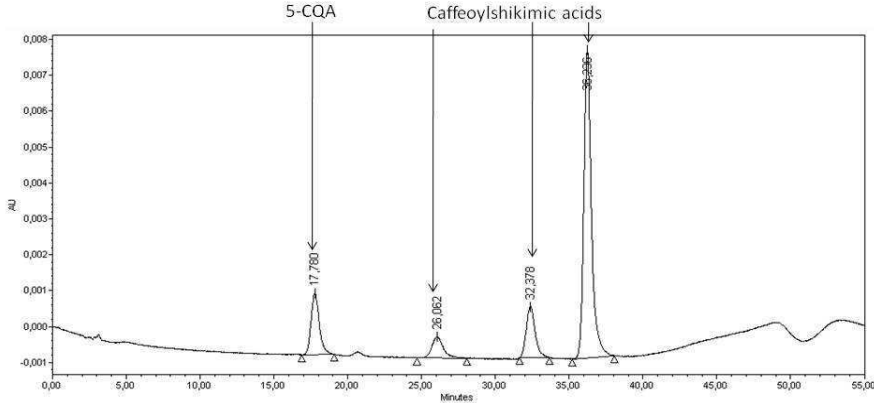
(D) Shikimic acid, HCT t= 60 min (sample 1042)



(E) Quinic and shikimic acids, HCT t= 5 min (sample 1028)



(F) Quinic and shikimic acids, HCT t= 60 min (sample 1043)



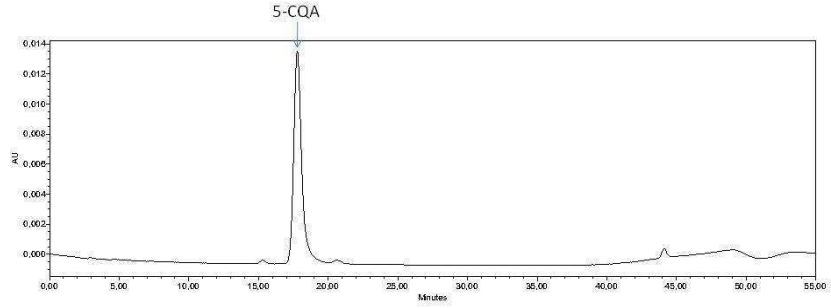
*Cc*HQT was incubated with caffeoyl-CoA and quinic or shikimic acid (Figure 48). With quinic acid as a substrate, caffeoyl-CoA was converted to 5-CQA (RT= 17.5 min). As for *Cc*HCT, incubation with caffeoyl-CoA and shikimic acid resulted in the accumulation of a CSA (RT= 35.5 min) (Figure 48B). Conversely, when both quinic and shikimic acids were supplied to *Cc*HQT, the enzyme mainly formed 5-CQA and only a small amount of CSA after t= 5 min (Figure 48C). After 60 min, 5-CQA and CSA are present in much closer ratios, 57 and 43 % respectively (Figure 48D). This indicates that *Cc*HQT prefers quinic acid, but is also capable of forming CSA.

The activity of *Cc*HCT and *Cc*HQT towards other potential substrates (4-hydroxyphenyllactic acid, spermidine and *myo*-inositol) was also assayed. These acceptor molecules were incubated with caffeoyl-CoA and the enzymes, however no activity was observed (data not shown). *Cc*HCT and *Cc*HQT are therefore active only towards the alicyclic quinic and shikimic acids as previously mentioned (1.6.2.4).

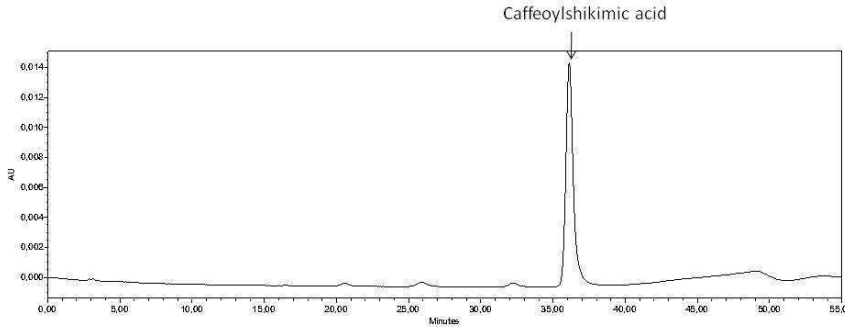
Figure 48: HPLC profiles of HQT reactions with caffeoyl-CoA and quinic/ shikimic acid

Reactions were set up with 0.1 M sodium phosphate pH 6.5, 0.2 mM caffeoyl-CoA and 5 mM quinic/ shikimic acid or both. Reactions were started by adding 1 µM enzyme. Samples taken at t= 5 and 60 min were analysed using HPLC-PDA with the methanol method. See Figure 47A for the control reaction.

(A) Quinic acid, HQT t= 60 min (sample 1044)



(B) Shikimic acid, HQT t= 60 min (sample 1045)



(C) Quinic and shikimic acids, HQT t= 5 min (sample 1031)

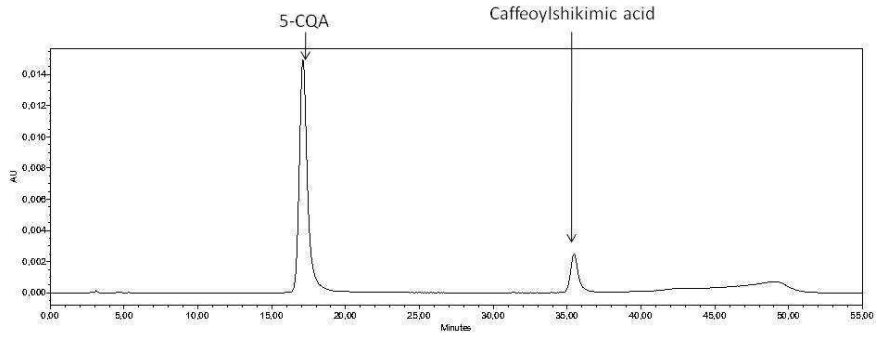
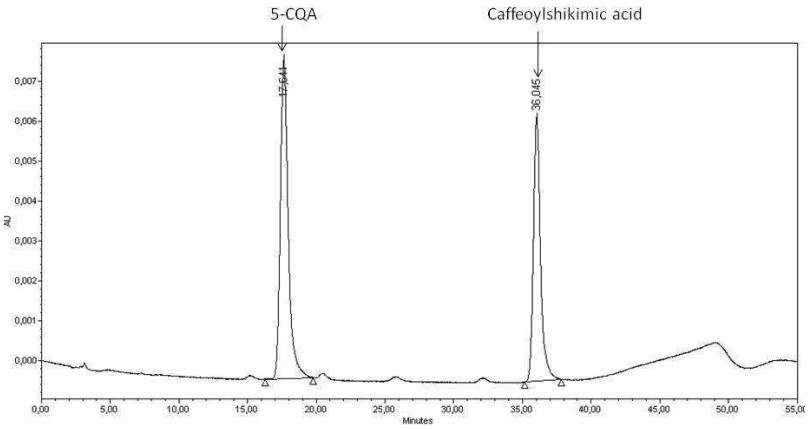


Figure 48 (continued)

(D) Quinic and shikimic acids, HQT t= 60 min (sample 1046)



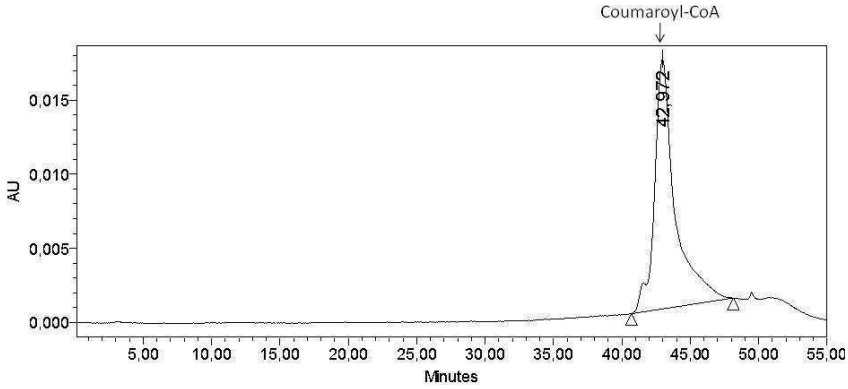
3.3.3.3. Activity of HCT and HQT towards coumaroyl-CoA

The incubation of coumaroyl-CoA and quinic acid with *Cc*HCT resulted in a peak (RT= 27.2 min) presumably corresponding to a coumaroylquinic acid (Figure 49B). With shikimic acid, three peaks were detected at RT= 33.2, 38.5 and 42.1 min, presumably representing 3-, 4- and 5-coumaroylshikimic acids (Figure 49C). When both acyl acceptors were present, a mixture of the three coumaroylshikimic acids plus a small amount of coumaroylquinic acid was observed (Figure 49D). So with coumaroyl-CoA, CcHCT appears to prefer shikimic acid as an acyl acceptor and produces more than one coumaroylshikimic acid isomer.

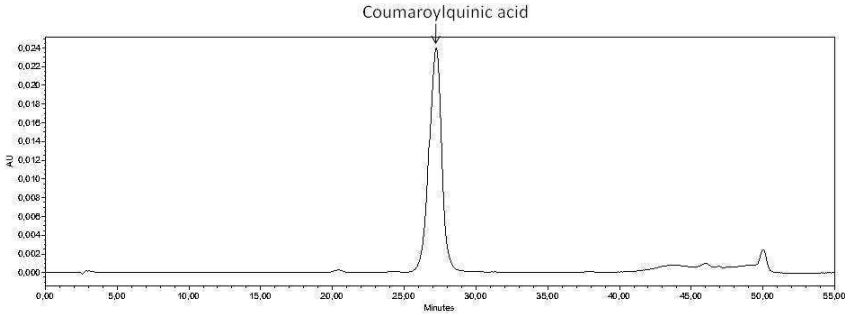
Figure 49: HPLC profiles of HCT reactions with coumaroyl-CoA and quinic/ shikimic acid

Reaction were set up with 0.2 mM coumaroyl-CoA and 10 mM quinic/ shikimic acid/ both in 0.1 M sodium phosphate pH 6.0. Reactions were started by adding 0.4 µM enzyme. Samples taken at t= 10 min were analysed using HPLC-PDA with the methanol method.

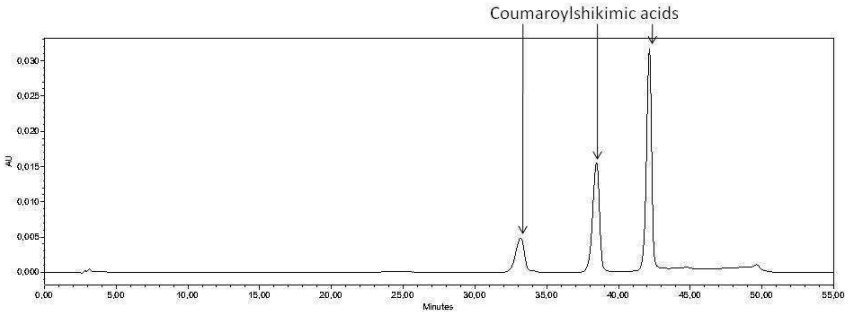
(A) coumaroyl-CoA, t= 0



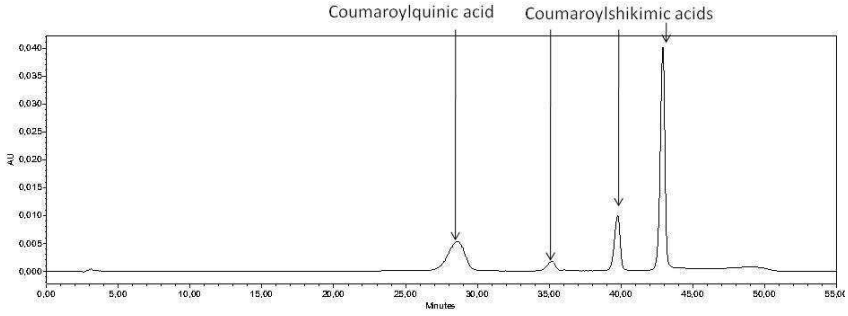
(B) Quinic acid, t= 10 min (sample 764)



(C) Shikimic acid, t= 10 min (sample 765)



(D) Quinic and shikimic acids, t= 10 min (sample 766)

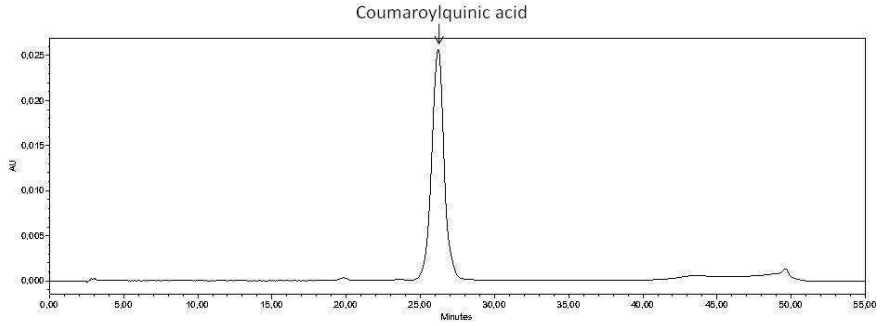


*Cc*HQT formed a coumaroylquinic acid from coumaroyl-CoA and quinic acid (Figure 50A). Coumaroyl-CoA and shikimic acid incubated with *Cc*HQT resulted in only one coumaroylshikimic acid isomer at RT= 41.9 min (Figure 50B). When both quinic and shikimic acids were supplied, HQT synthesised a mixture of presumably coumaroylquinic and coumaroylshikimic acid in the ratio of 85 and 15 % respectively at t= 10 min (Figure 50C). This ratio shifted to 78 and 22 % at t= 75 min (data not shown). As seen when using caffeoyl-CoA as the forward substrate, it is again noted that *Cc*HQT prefers quinic acid and produces only one coumaroylshikimic acid isomer.

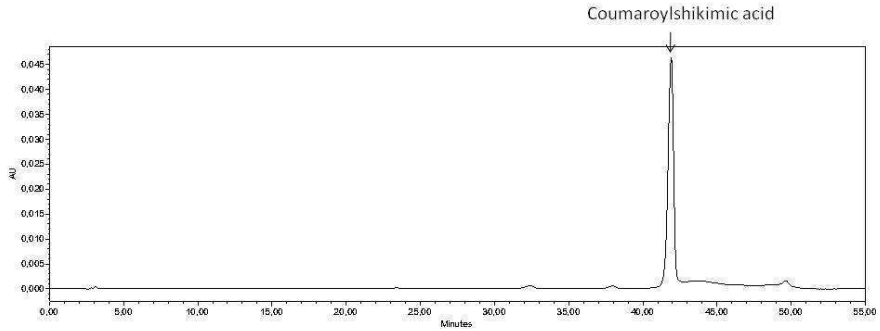
Figure 50: HPLC profiles of HQT reactions with coumaroyl-CoA and quinic/ shikimic acid

The same conditions as Figure 49 were applied.

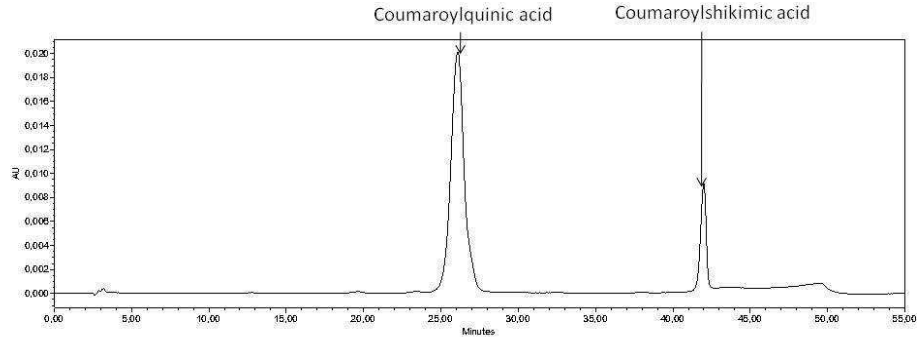
(A) Quinic acid, t= 10 min (sample 767)



(B) Shikimic acid, t= 10 min (sample 768)



(C) Quinic and shikimic acids, t= 10 min (sample 769)



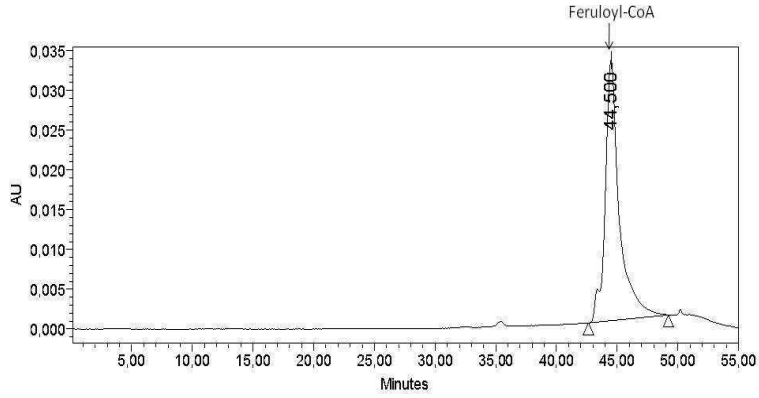
3.3.3.4. Activity of CcHCT and CcHQT towards feruloyl-CoA

Here, *Cc*HCT can synthesise a shikimate ester, presumably a feruloylshikimic acid (RT= 45.2 min) from feruloyl-CoA and shikimic acid and the reaction was complete after 2 min. The new compound has a typical CGA absorbance spectrum (Figure 51B). However, only a small peak of 5-FQA (RT= 30.5 min) is formed in the presence of quinic acid at t= 20 min (Figure 51C). When both acyl acceptors were supplied, *Cc*HCT catalyses the synthesis of a feruloylshikimic acid corresponding to the RT for the 5-acyl isomer (Figure 51D). This demonstrates the preference of CcHCT for shikimic acid and its much lower efficiency towards 5-FQA formation.

Figure 51: HPLC profile of HCT reactions with feruloyl-CoA and quinic/ shikimic acid

Reaction mixtures containing 0.2 mM feruloyl-CoA, 10 mM quinic or shikimic acid or both and 0.1M sodium phosphate pH 6.0 were prepared in a 100 µL final volume. The reaction was started by adding 0.01 µM HCT. Samples taken at t= 20 min were analysed using HPLC-PDA with the methanol method. Panel B shows the absorption spectrum of the putative feruloylshikimic acid peak.

(A) feruloyl-CoA t= 0



(B) Shikimic acid t= 2 min (sample 717)

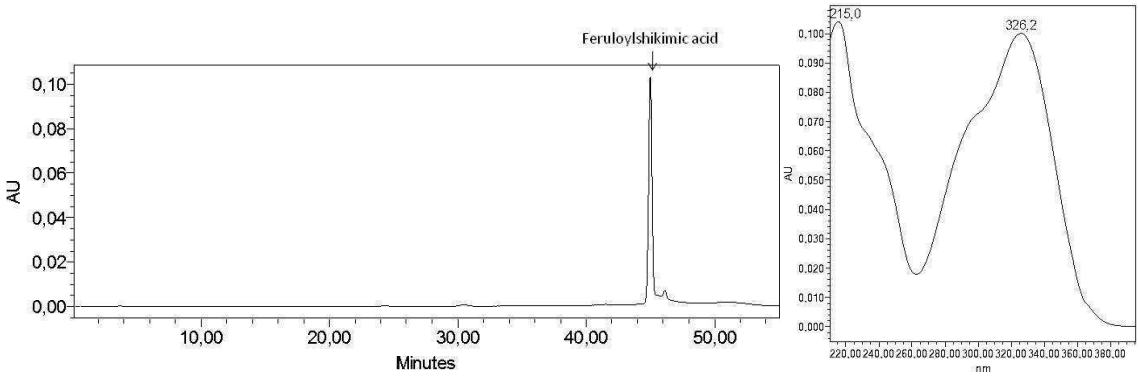
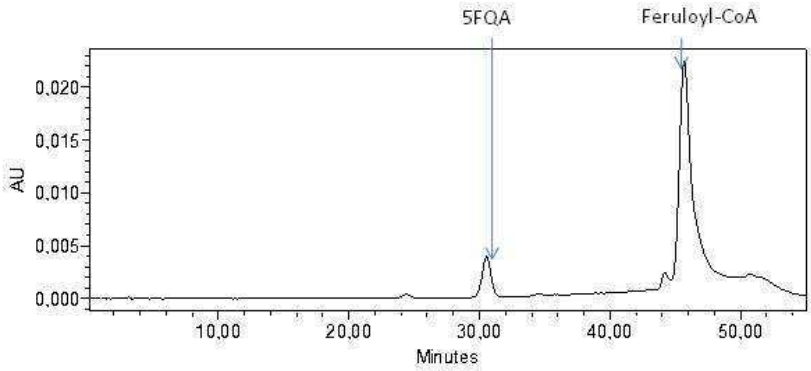
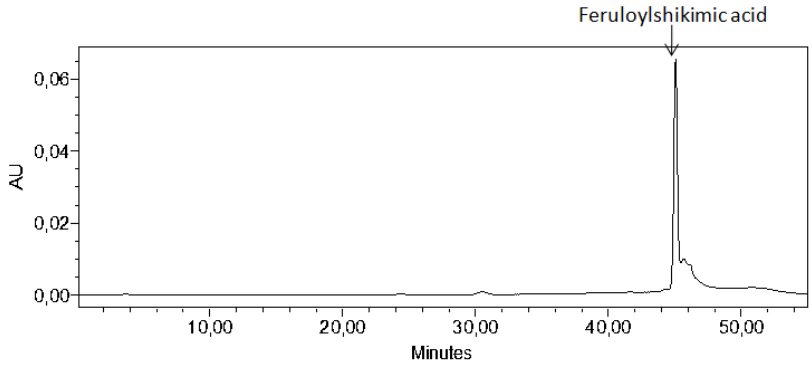


Figure 51 (continued)

(C) Quinic acid t= 20 min (sample 722)



(D) Quinic and shikimic acids t= 2 min (sample 718)



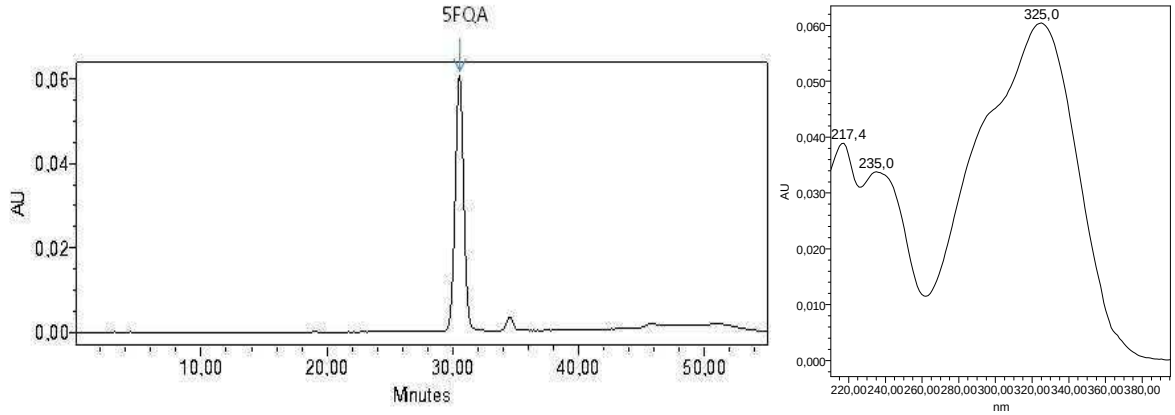
The incubation of feruloyl-CoA with HQT and quinic acid leads to 5-FQA (RT= 30.5 min) (Figure 52A). While with shikimic acid it leads to a feruloylshikimic acid (RT= 45.1 min) (Figure 52B). The conversion was complete after t= 2 min and 20 min respectively. As expected, when both quinic acid and shikimic acid were supplied, HQT preferentially synthesised 5-FQA (Figure 52C).

In summary, in the presence of the same CoA thioester (caffeoyl-/ coumaroyl-/ feruloyl-CoA), *Cc*HCT clearly prefers shikimic over quinic acid. *Cc*HCT is also much less efficient with feruloyl-CoA in the presence of quinic acid. In contrast, *Cc*HQT is able to use all three CoA thioesters equally with both quinic and shikimic acids. It is also interesting to note that HCT can form three probable shikimate esters of caffeic and coumaric acids, while with feruloyl-CoA only a single shikimate ester is formed. In contrast, *Cc*HQT forms only a single shikimate ester with all CoA thioesters tested.

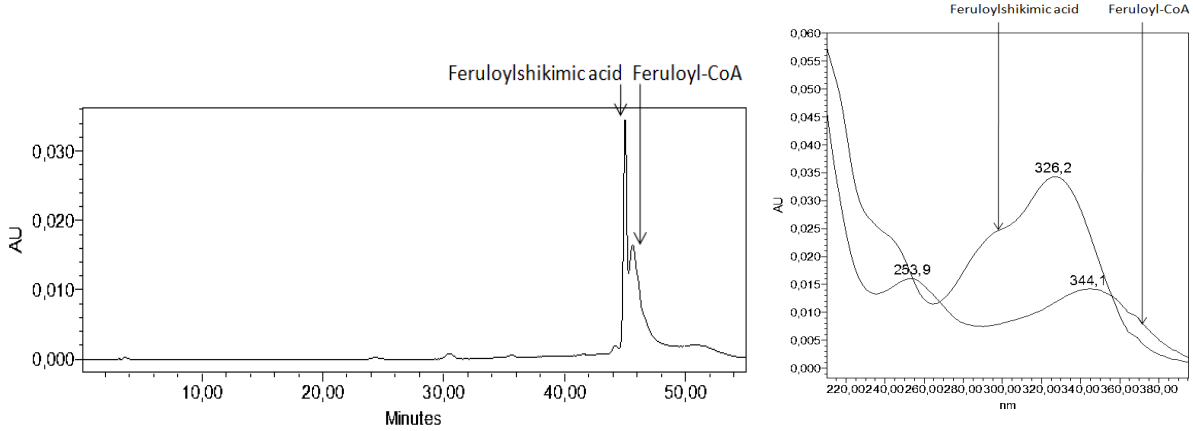
Figure 52: HPLC profiles of HQT reactions with feruloyl-CoA and quinic/ shikimic acid

The same conditions as presented in Figure 51 were applied.

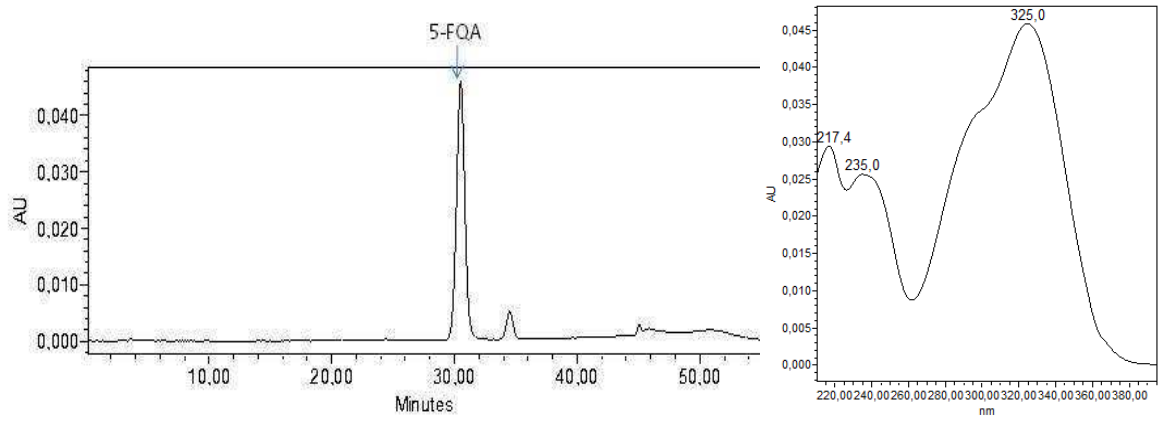
(A) Quinic acid t= 2 min (sample 719)



(B) Shikimic acid t= 2 min (sample 720)



(C) Quinic and shikimic acids t= 2 min (sample 721)



3.3.4. Comparison of the HCT and HQT-catalysed reverse reactions

3.3.4.1. Stability of chlorogenic acids

CGAs are known to be unstable at basic pH and stable at acidic pH (Friedman, 2000). This was verified by heating buffered solutions of 5-CQA and 5-FQA to 90 °C. At pH 6.0, significant amounts of 3- and 4-CQA were produced from the 5-CQA isomer. At neutral pH, 5-CQA was converted to approximately equivalent levels of 3- and 4-CQA isomers, resulting in an equimolar mixture of the three caffeoylquinic acid isomers (Table 13). No major difference was observed between 30 min and 60 min incubation times (data not shown). 5-FQA behaved very similarly to 5-CQA by partial isomerisation to 3- and 4-FQA (data not shown). Further experiments performed at 34 and 25 °C on the mono- and diCQAs from Biopurify showed that, even though the enzymatic reactions were carried out at 34 °C and pH 6.5, 5-CQA partially isomerised into 4-CQA and to a lesser extent into 3-CQA (Figure 53C and C'). In a 4-CQA solution, the acyl migrated to the 3-hydroxyl position and to a lesser extent to the 5-hydroxyl position (Figure 53B and B'). 3-CQA isomerised into 4-CQA and, to a lesser extent, into 5-CQA (Figure 53A and A'). Similarly, 3,5-diCQA isomerised into 3,4- and 4,5-diCQA when incubated overnight at 25 °C and this isomerisation increased when the sample was kept at 34 °C (data not shown).

Table 13: Effect of temperature and pH on the isomerisation of 5-CQA

10 mM 5-CQA in 0.1 M sodium phosphate ph 6.0 or 7.0.

pH	*Treatment*	*Isomeric ratio (%)*		
		5-CQA	4-CQA	3-CQA
6.0	90 °C for 30 min	65.0	19.6	15.4
7.0	90 °C for 30 min	32.4	31.3	36.3

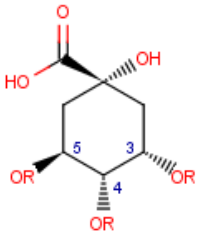
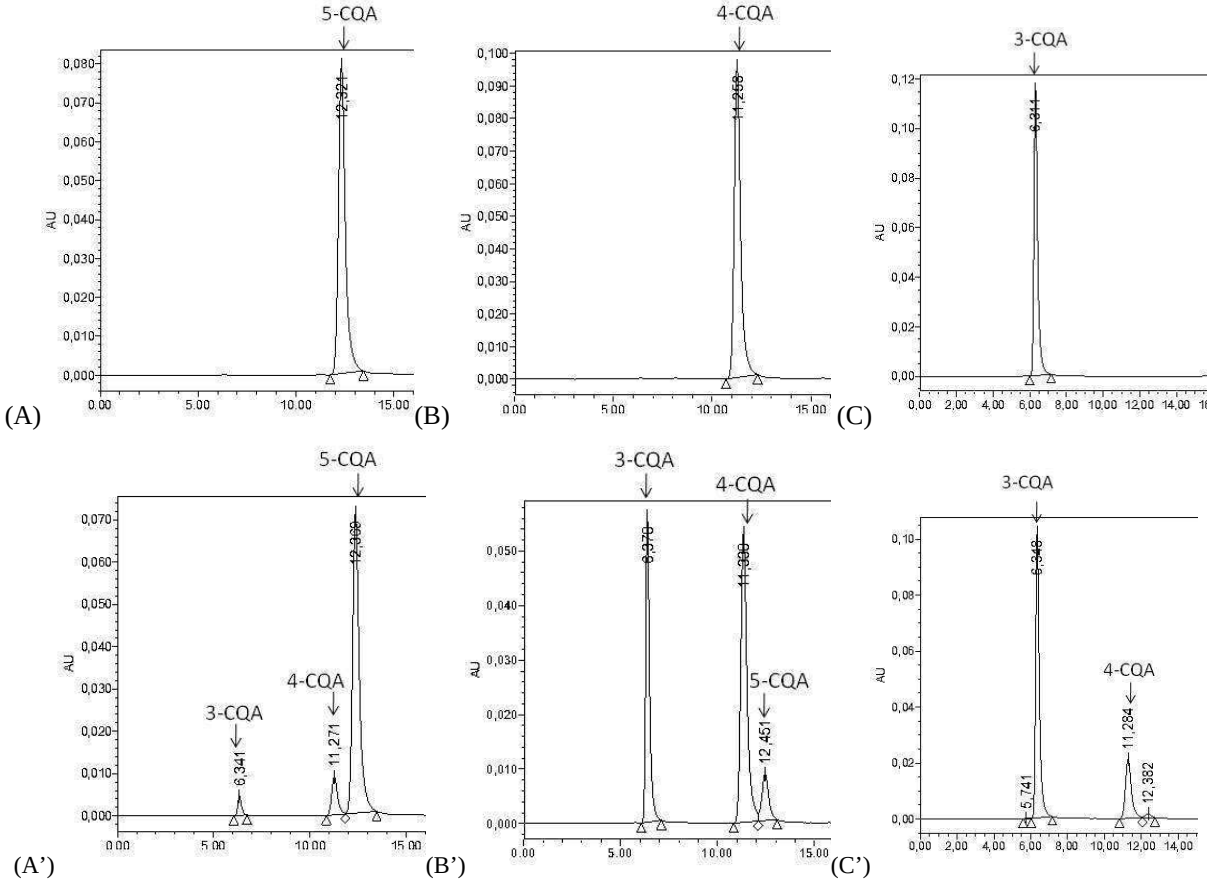


Figure 53: HPLC profiles of CQA standards upon heating

Samples containing 1 mM CQA isomer and 0.1M sodium phosphate pH 6.5 were incubated overnight and analysed using HPLC-PDA with the acetonitrile method.

(A) 5-CQA t= 0 (sample 1371); (B) 4-CQA t= 0 (sample 1370); (C) 3-CQA t= 0 (sample 1369); (A') 5-CQA overnight at 34 °C (sample 1377); (B') 4-CQA overnight at 34 °C (sample 1376); (C') 3-CQA overnight at 34 °C (sample 1375).



This isomerisation most probably occurs via the acyl transfer of the hydroxycinnamoyl moiety between the three readily accessible hydroxyl positions of quinic acid (hydroxyl at C-1 is not involved). In some of the enzymatic reactions that will be described later, a significant amount of 3- and 4-CQA isomers was derived from solutions containing 5-CQA, especially after overnight incubation. It is assumed that this occurred not from enzymatic synthesis but from chemical isomerisation.

3.3.4.2. Characterisation of the reverse reaction products formed by HCT and HQT

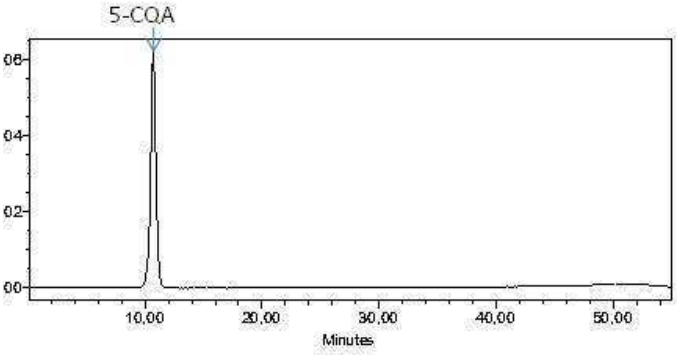
When CoA was added as a substrate with 5-CQA in the presence of *Cc*HCT or *Cc*HQT, there was the emergence of a peak at RT= 27.5 min showing a characteristic CoA thioester UV spectrum (Figure 54B and C), presumably corresponding to caffeoyl-CoA. However, under identical reaction conditions and enzyme concentrations, *Cc*HCT (Figure 54B) produces significantly less caffeoyl-CoA than HQT (Figure 54C). A 100-fold dilution of

HQT can synthesise the same amount of caffeoyl-CoA as HCT at a 1 µM concentration (Figure 54B and D). This is further evident that CcHQT has a preference for reactions involving quinate moieties than does CcHCT.

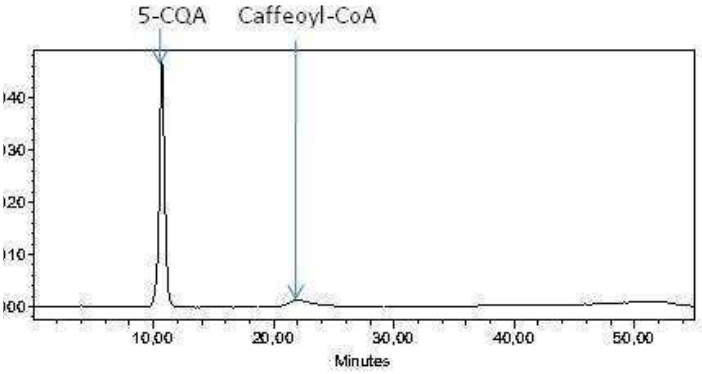
Figure 54: HPLC profiles of HCT and HQT incubated with 5-CQA and CoA

Reactions were set up with 0.1 M sodium phosphate pH 6.0, 3.75 mM 5-CQA, 5 mM CoA, and milliQ. Reactions were started by adding 1 µM HCT or 1.0/ 0.01 µM HQT to the reaction mixture. Samples taken at t= 5, 15, 30 and 60 min were analysed using HPLC-PDA with the methanol method. The retention times here are different from earlier experiments due to the different chromatographic conditions.

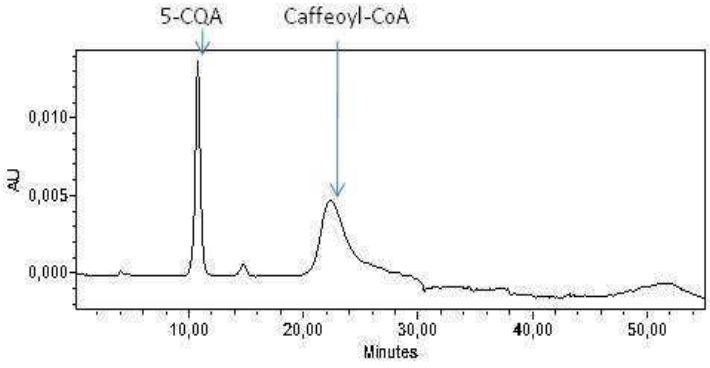
(A) t= 0 (sample 427)



(B) 1 µM HCT t= 60 min (sample 425)



(C) 1 µM HQT t= 60 min (sample 419)



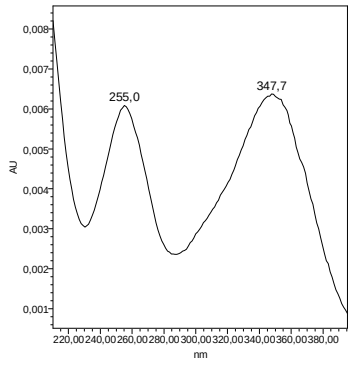


Figure 54 (continued)

(D) 0.01 µM HQT t= 60 min (sample 423)

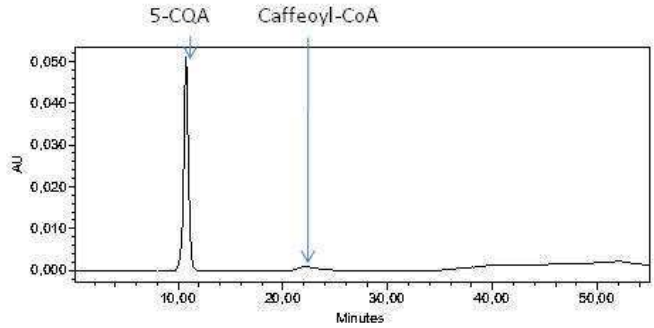
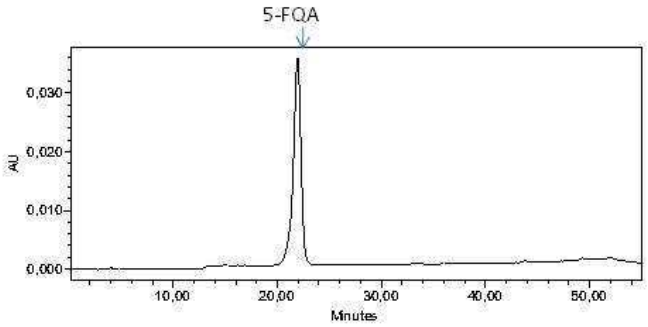


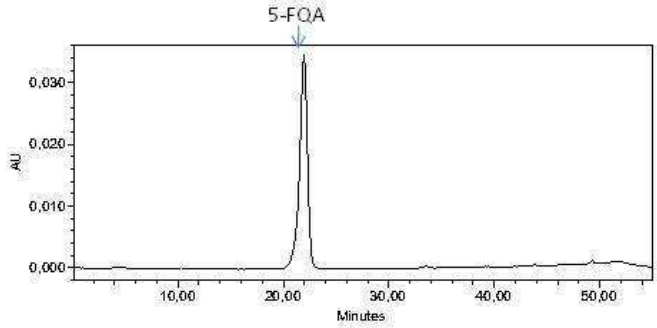
Figure 55: HPLC profiles of HCT and HQT incubated with 5-FQA and CoA

Reactions were set up with 0.1 M sodium phosphate pH 6.0, 3.75 mM 5-FQA, 5 mM CoA and milliQ water in a final volume of 100 µL. Reactions were started by adding 1 µM HCT/ HQT. Samples (395-428) taken at t= 0, 5, 15, 30 and 60 min were analysed using HPLC-PDA with the methanol method.

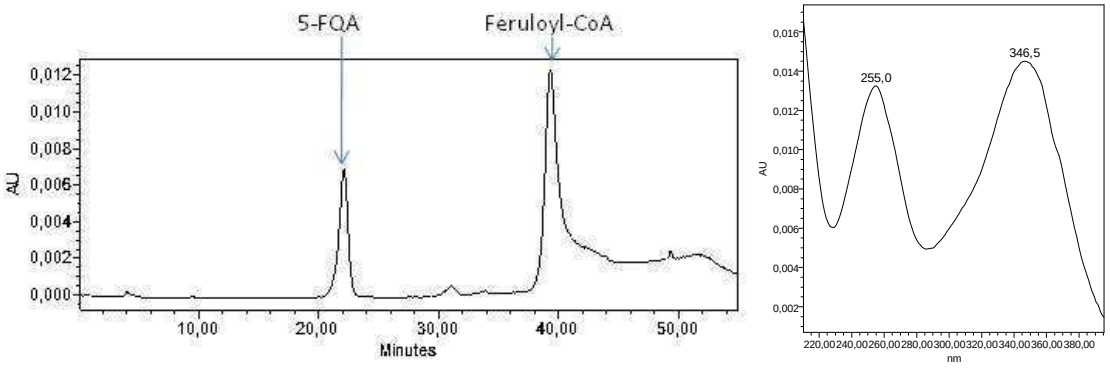
(A) t= 0 (sample 428)



(B) 1 µM HCT t= 60 min (sample 426)



(C) 1 µM HQT t= 60 min (sample 420)



Further evidence of this is that no product formation was observed when *Cc*HCT was incubated with 5-FQA and CoA under the conditions tested (Figure 55A and B) and that, in contrast, *Cc*HQT was with the same substrates able to synthesise a product at RT= 39.8 min, which probably corresponds to feruloyl-CoA, as it shows a UV spectrum characteristic of a CoA thioester (Figure 55C).

The main reaction products when using 5-CQA or 5-FQA as substrates were presumed to be CoA thioesters. The absorption spectra of the compounds contained in the HPLC peaks measured using HPLC-PDA present two maxima around 256 and 346 nm, the latter being characteristic of the presence of a thioester bond. The compounds were subjected to a LC-MS/MS analysis in order to confirm their identity. An m/z value of 928 for the reaction product synthesised by *Cc*HQT from 5-CQA and CoA corresponds to caffeoyl-CoA (MW= 930 g/mol). An m/z value of 942 for the reaction product synthesised by HQT from 5-FQA and CoA corresponds to feruloyl-CoA (MW= 944 g/mol).

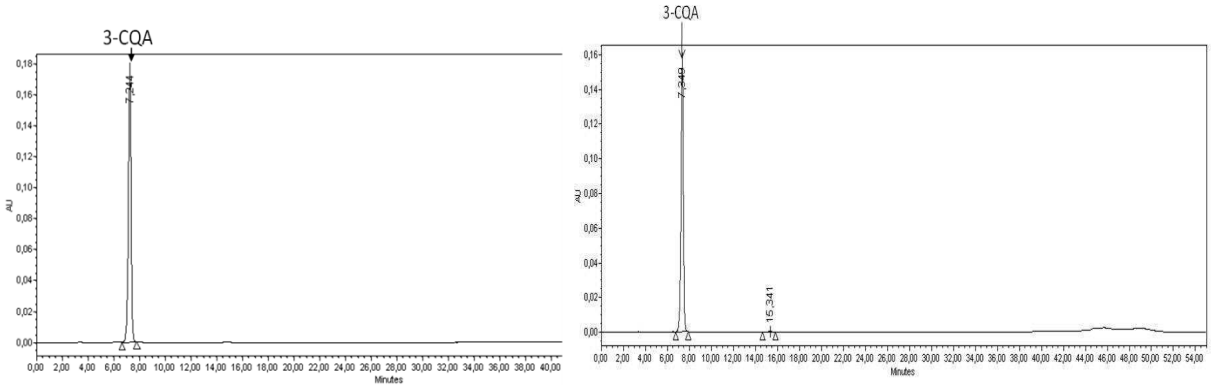
3.3.4.3. Substrate specificity of HCT and HQT in the reverse reaction

For preliminary experiments testing *Cc*HCT and *Cc*HQT specificity (data not shown), various CGA isomers were generated by heating solutions of 10 mM 5-CQA/ 5-FQA buffered at pH 7.0 (2.3.2.6). Only the 5-CQA peak was clearly decreased by conversion to caffeoyl-CoA with *Cc*HCT and *Cc*HQT in the presence of CoA. No activity towards 3-CQA and 4-CQA was detected. Similarly, with a FQA mixture only the 5-FQA isomer peak was decreased in the presence of *Cc*HQT. As expected (3.3.4.2), HCT was less efficient towards this substrate. When the CQA kit was obtained from Biopurify, reaction mixtures were prepared with each monoester individually and CoA. Again, only 5-CQA was converted by *Cc*HCT and *Cc*HQT to caffeoyl-CoA in the presence of CoA (Figure 56), confirming the results obtained with the CQA isomer mixture. Rosmarinic acid was also incubated overnight with CoA, but no reaction product was observed with *Cc*HCT or *Cc*HQT (data not shown).

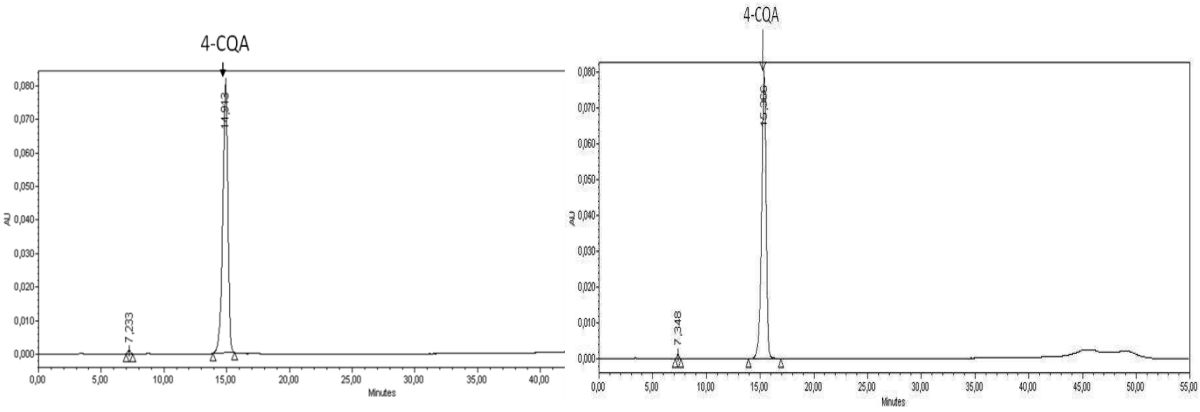
Figure 56: HPLC profiles of *Cc*HQT incubated with 3-, 4- and 5-CQA isomers and CoA

Reactions were set up with 1 mM CQA and 5 mM CoA in 0.1 M sodium phosphate pH 7.5. Reactions were started by adding 0.4 µM HQT. Samples taken at t= 0 and 30 min were analysed using HPLC-PDA with the methanol method.

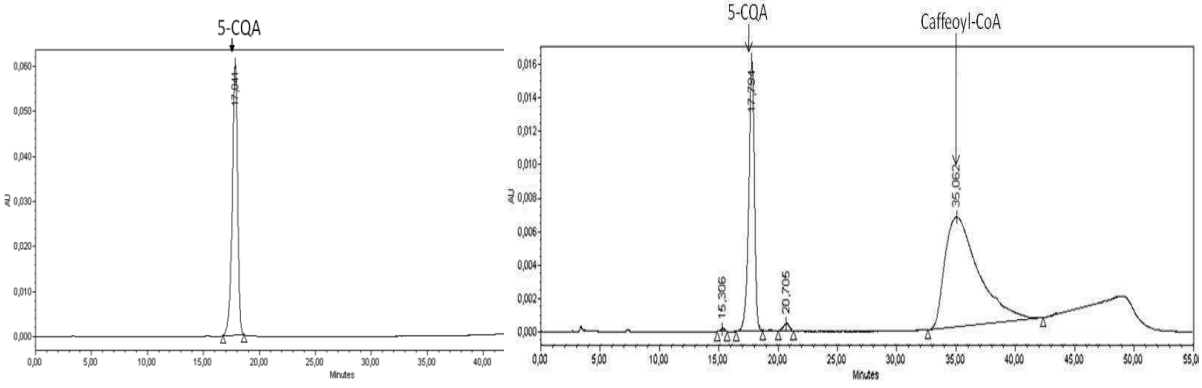
(A) 3-CQA t= 0 t= 30 min (sample 612)



B) 4-CQA t= 0 t= 30 min (sample 611)



(A) 5-CQA t= 0 t= 30 min (sample 610)

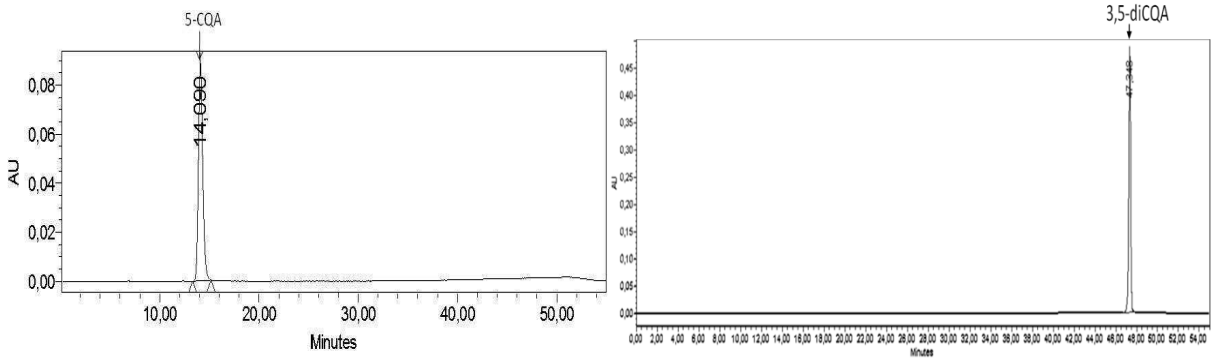


*Cc*HCT and *Cc*HQT were also incubated with 3,5-diCQA and CoA. In the control with no enzyme, 3,5-diCQA isomerises into 3,4-diCQA and possibly 4,5-diCQA, however, this peak could not be resolved from the 3,5-diCQA peak at RT= 47.2 min (Figure 57B). In the presence of *Cc*HCT, 5-CQA and caffeoyl-CoA are detected as indicated from the 5-CQA standard and the absorbance spectra (Figure 57A and C). In the presence of *Cc*HQT, no reaction product was observed after overnight incubation (data not shown).

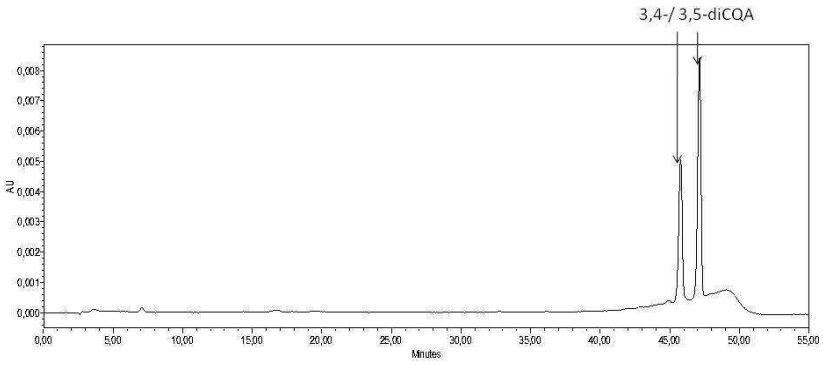
Figure 57: HPLC profiles of HCT incubated with 3,5-diCQA and CoA

Reactions were set up containing 1 mM 3,5-diCQA and 5 mM CoA in 0.1 M sodium phosphate pH 6.0 to a final volume of 200 µL. Reactions were started by adding 1 µM HCT/ HQT/ water as a control.

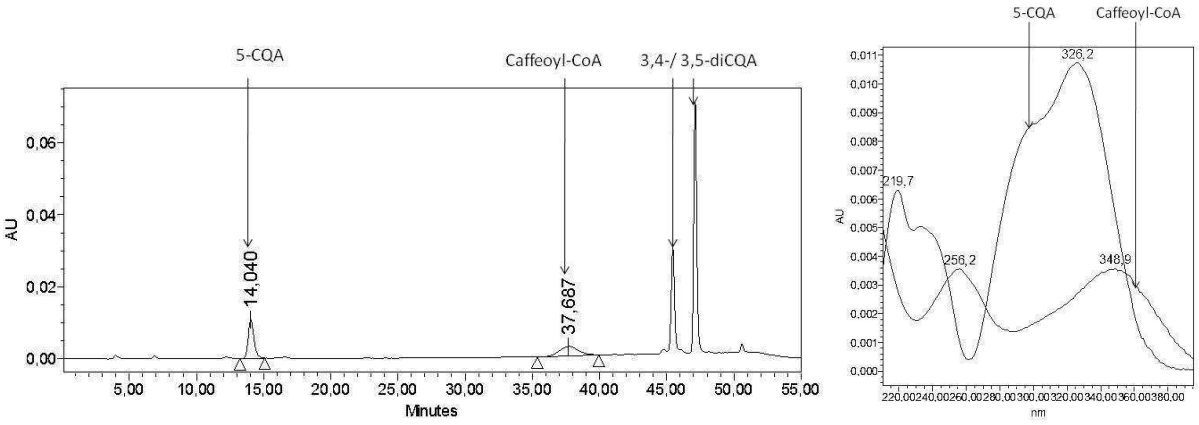
(A) Standards: 5-CQA 3,5-diCQA



(B) Control with no enzyme overnight (sample 882)



(C) HCT overnight (sample 880)



3.3.4.4. Other experiments investigating the enzymatic properties of HCT/ HQT

Because of both a lack of time and limited access to a HPLC system for analysis, a much smaller amount of work was carried out regarding more detailed studies of *Cc*HCT and *Cc*HQT reaction properties. Temperature variation (22 °C, 34 °C and 42 °C) did not significantly alter the activity of *Cc*HCT and *Cc*HQT in the reverse reaction with 5-CQA and CoA. All subsequent enzymatic reactions were therefore carried out in a 34 °C water-bath.

In order to address the effects of pH on the reverse reactions of *Cc*HCT and *Cc*HQT, several experiments were carried out. In the first set of reactions performed, *Cc*HQT and *Cc*HCT were not active at pH 4.6 (Figure 58). The results from a more extensive analysis on the effects of pH on the reverse reaction confirmed that the *Cc*HQT optimal pH value is close to 6.0, whereas *Cc*HCT seemed to be more active towards 5-CQA at pH 8.0 (Figure 59). pH values higher than pH 8.0 have not yet been tested with either enzyme because CGA compounds are not stable at this pH. These results also highlight the remarkable difference in the catalysis efficiency of *Cc*HCT and *Cc*HQT with similar levels of 5-CQA and CoA substrates and confirm the preference of *Cc*HQT for quinate substrates.

Figure 58: Influence of pH on caffeoyl-CoA production by HCT and HQT

Reactions were set up with 1 µM enzyme, 2 mM CoA and 2 mM 5-CQA in 0.1 M sodium phosphate at pH 4.6/ 6.0/ 7.6. Reactions were stopped at t= 300 min.

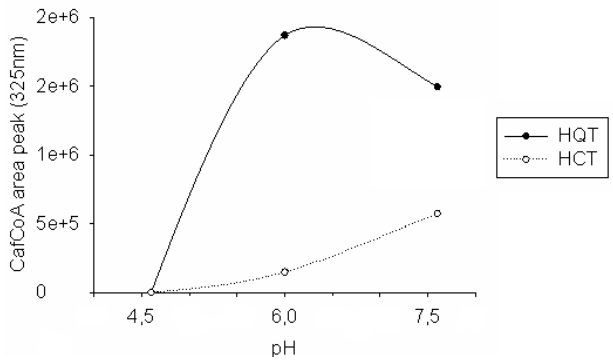
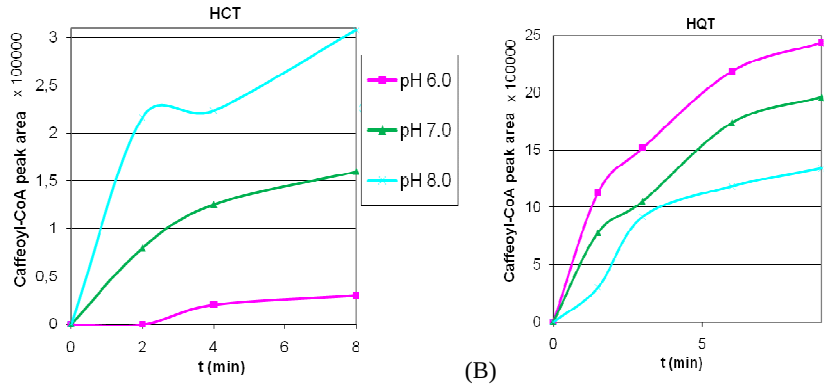


Figure 59: Influence of pH on HCT and HQT activity towards 5-CQA and CoA

Reactions were set up with 2 mM CoA and 2 mM 5-CQA in 0.1 M sodium phosphate at pH 6.0, 7.0 and 8.0. Reactions were started by adding 1 µM HCT (A) or 0.1 µM HQT (B).



(A) (B)

3.3.4.5. Effect of different enzyme and substrate concentrations on product formation

As seen in 3.3.4.2, the production of the same amount of caffeoyl-CoA (at constant 5-CQA and CoA concentrations) requires very different amounts of HCT and HQT. The reaction steady-state persists for many minutes with HCT in the range of concentrations that were used (0.1-0.8 µM) (Figure 60A). Under these conditions, the Michaelis-Menten law may be applicable. The reaction velocity as a function of substrate and enzyme concentrations can yield the Michaelis and catalytic constants. However, lower enzyme concentrations are required in the case of HQT (Figure 60B).

When serial dilutions of 5-CQA were incubated with CoA and *Cc*HQT, a maximal conversion to caffeoyl-CoA production is reached very rapidly, which corresponds to the equilibrium attained between substrates and products (Figure 61).

To study the potential effect of quinic acid concentration on the *Cc*HQT-catalysed reverse reaction, initial quinic acid concentrations were varied from 0 to 20 mM in the presence of constant levels of 5-CQA (5 mM) and CoA (5 mM). The results show that increasing quinic acid concentration produces a significant inhibitory effect on the reverse reaction (Figure 62). It would be interesting in future experiments to compare this with *Cc*HCT.

The effect of varying the CoA concentration in the reverse reaction was also explored. As expected, increasing the CoA concentration while maintaining 5-CQA levels constant pushes the reverse reaction towards the formation of more caffeoyl-CoA (Figure 63).

Figure 60: Caffeoyl-CoA formation as a function of HCT/ HQT concentration

Reactions were set up with 5 mM CoA and 2 mM 5-CQA in 0.1 M sodium phosphate pH 6.0. The reaction was started by adding either HCT (A) or HQT (B). Enzyme concentration units are nM.

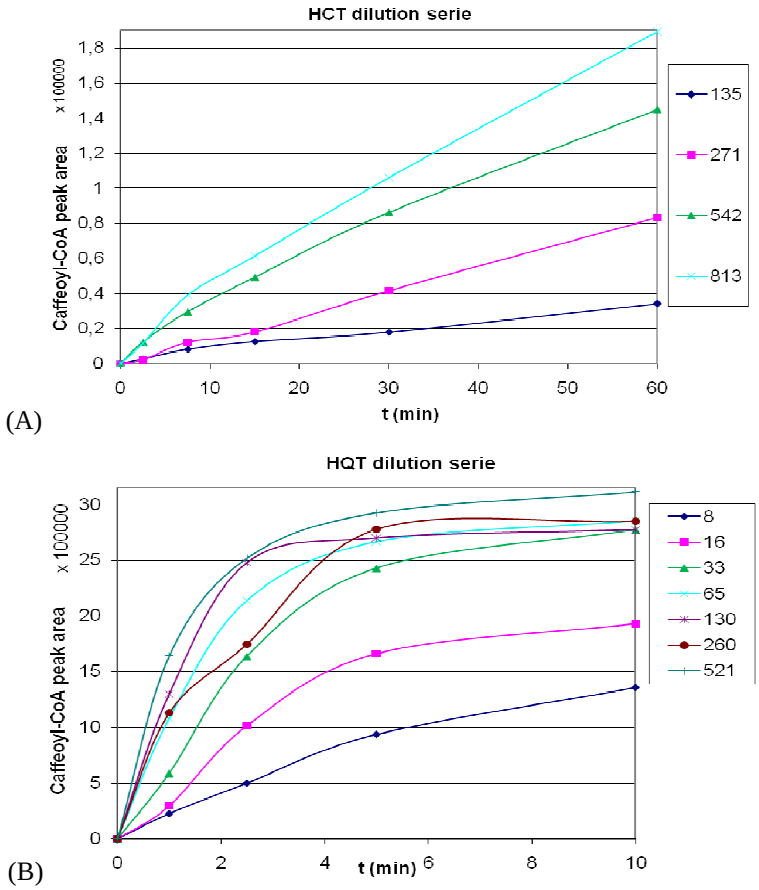


Figure 61: Influence of 5-CQA concentration on caffeoyl-CoA formation catalysed by HQT

Reactions were set up with 0.08-10 mM 5-CQA and 2 mM CoA in 0.1 M sodium phosphate pH 6.0. Reactions were started by adding 0.1 µM HQT.

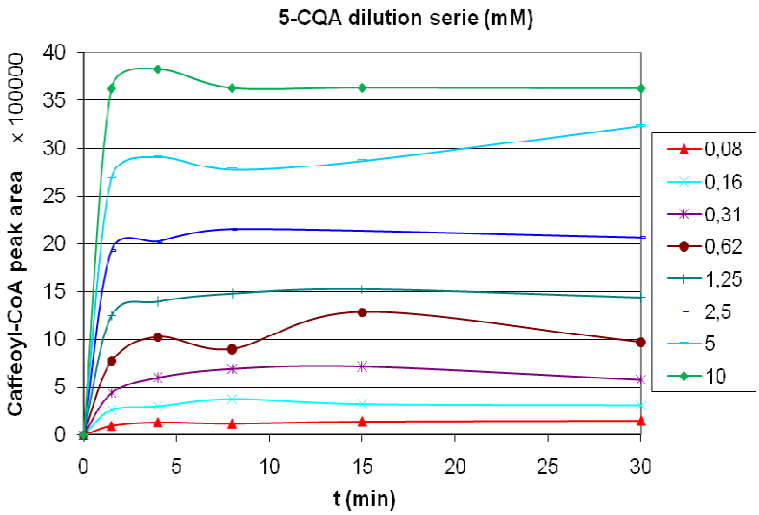


Figure 62: Inhibitory effect of quinic acid on the level of caffeoyl-CoA formation by HQT

Reactions were set up with 5 mM 5-CQA, 5 mM CoA and 0-20 mM quinic acid in 0.1 M sodium phosphate pH 6.0. Reactions were started by adding 1 µM HQT. Samples taken at t= 0, 5 and 30 min (488-503) were analysed using HPLC-PDA with the methanol method.

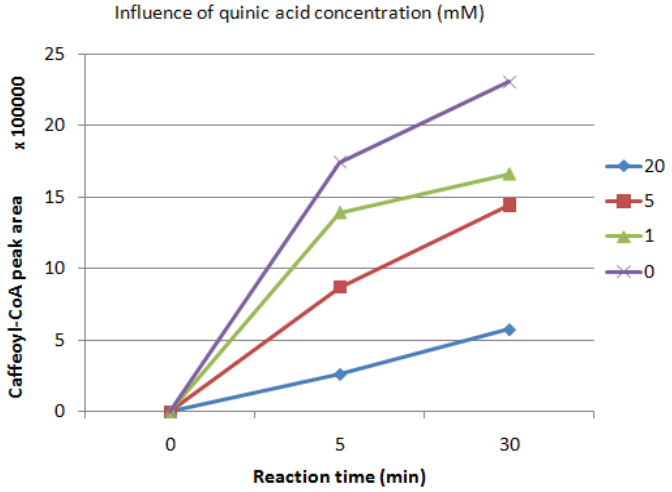
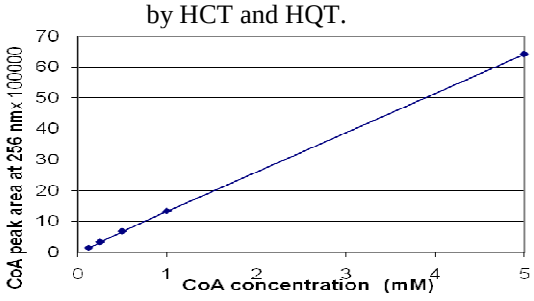


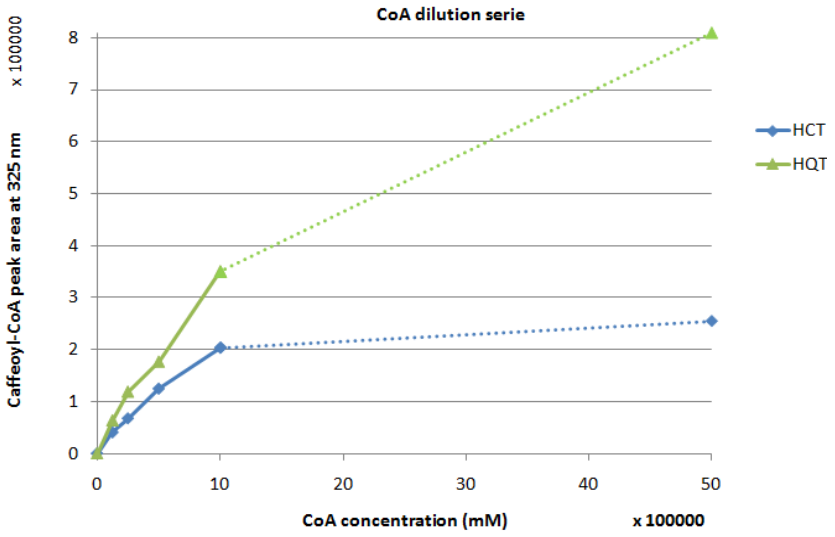
Figure 63: Influence of CoA concentration on the level of caffeoyl-CoA formed by HCT and HQT

Reactions were set up with 0.125-5 mM CoA and 1 mM 5-CQA in 0.1 M sodium phosphate pH 6.5. Reactions were started by adding 5 µM HCT or 0.5 µM HQT. Samples (1101-1145) taken at t= 0 and 180 min were analysed using HPLC-PDA with the acetonitrile method.
(A) CoA concentration added; (B) Effect of CoA concentration on the level of caffeoyl-CoA formation catalysed by HCT and HQT.

(A)



(B)



Thus, caffeoyl-CoA formation depends on the initial concentration of 5-CQA, CoA and quinic acid. It should be noted that the production of caffeoyl-CoA by *Cc*HQT was initially used for the production of this compound. However, as this reaction reaches equilibrium well below the 100 % conversion rate due to the forward reaction, it is not efficient. The preliminary data on the properties of *Cc*HCT and *Cc*HQT enzymes presented above will serve as a strong foundation for future experiments directed at a detailed kinetic analysis.

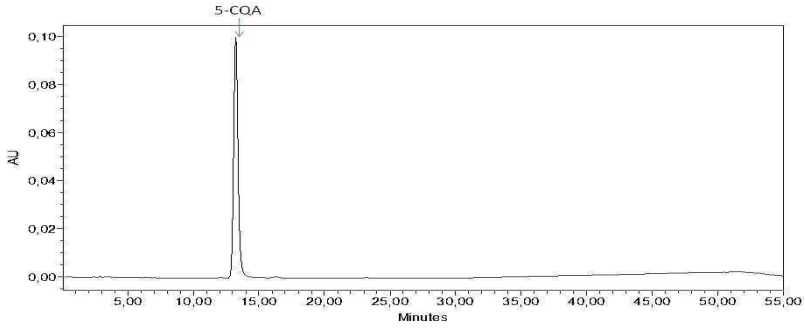
3.3.4.6. Evidence for diCQA formation in the reverse reaction

In reactions containing 5-CQA and CoA catalysed by *Cc*HCT, unknown peaks were observed at later retention times (Figure 64C). To confirm the identity of these peaks, CGA standards were run, confirming their assignments as 3,4-, 3,5- and 4,5-diCQA. DiCQA formation was observed with *Cc*HCT when 5-CQA and CoA were supplied in a 1:1 molar ratio but not when it was 1:10 (Figure 64C and D). DiCQAs were not observed when using *Cc*HQT (Figure 64E and F).

Figure 64: HPLC profiles of HCT/ HQT incubated with 5-CQA and CoA at two different molar ratios

Reactions were set up with 1 mM 5-CQA and either 1 or 10 mM CoA in 0.1 M sodium phosphate pH 6.5. Reactions were started by adding 0.5 µM enzyme. Samples taken at t= 180 min were analysed using HPLC-PDA with the acetonitrile method.

(A) t= 0



(B) Control 5-CQA and CoA no enzyme overnight

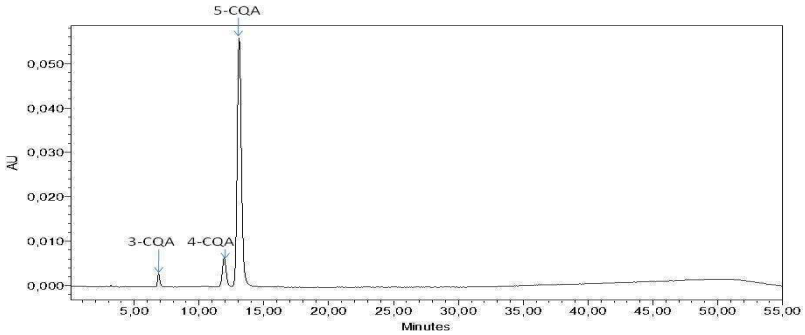
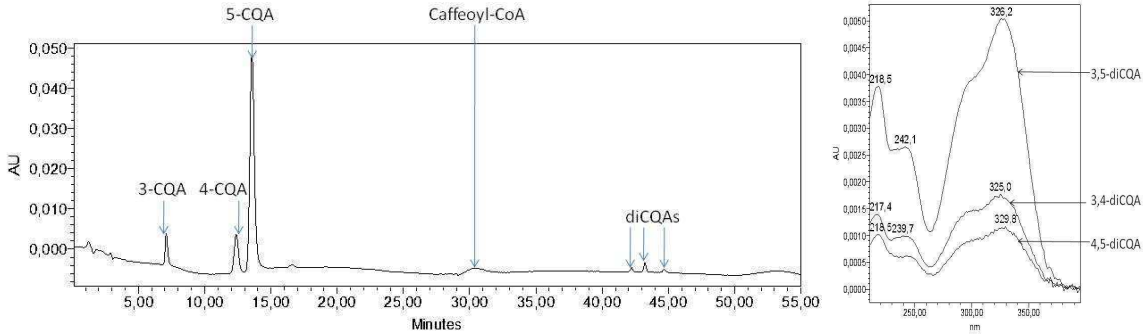
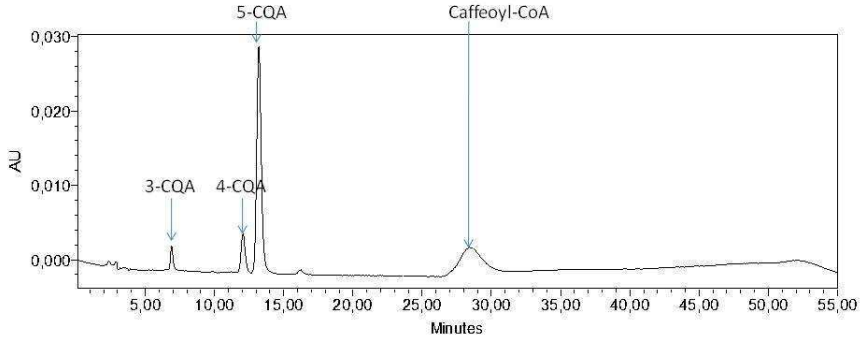


Figure 64 (continued)

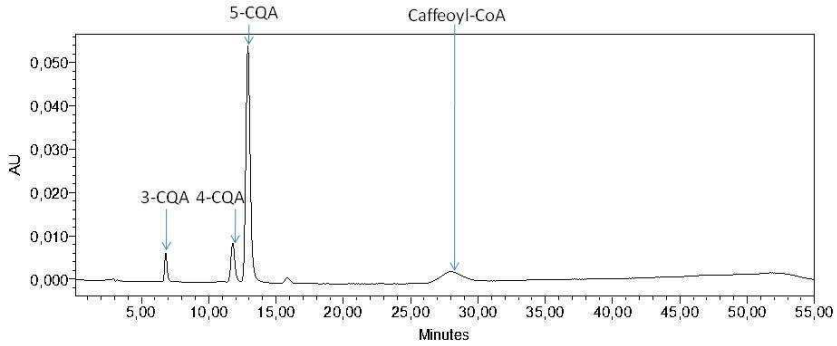
(C) HCT 1/1 5-CQA/CoA ratio (sample 1085) and overlap of absorbance spectra of the diCQA peaks



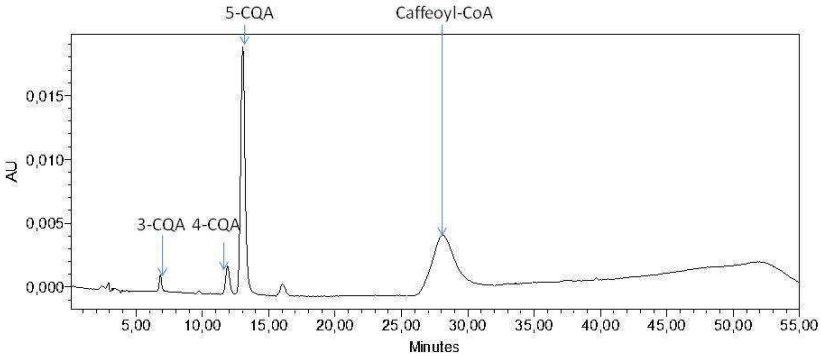
(D) HCT 1/10 5-CQA/CoA ratio (sample 1086)



(E) HQT 1/1 5-CQA/CoA ratio (sample 1089)



(F) HQT 1/10 5-CQA/CoA ratio (sample 1090)



3.3.5. Comparison of the activity of native and mutant HCTs

3.3.5.1. Comparison of native and mutant HCT in the forward and reverse reactions

In order to determine the importance of specific residues on the enzymatic activity of *Cc*HCT, point amino acid mutations were introduced. The activities of the mutant HCTs (K, H35A, H153A, H154N, D157A, Y252A and R374E) were compared to that of the native *Cc*HCT. All mutants were incubated with 1 mM CoA and 1 mM 5-CQA (reverse reaction), or with 0.2 mM caffeoyl-CoA and 1 mM quinic/ shikimic acid (forward reaction). A summary of the results when either 5-CQA and CoA or caffeoyl-CoA and quinic acid were used is presented in Figure 65.

In the forward reaction, both the native and the K-mutant HCTs used caffeoyl-CoA and quinic acid to form 5-CQA and, to a lesser extent, free caffeic acid. For both proteins there appears to be a complete conversion of the caffeoyl-CoA substrate. When using caffeoyl-CoA and shikimic acid as substrates, the K-mutant HCT catalysed the formation of all three caffeoylshikimic acids produced by native *Cc*HCT (data not shown). In the reverse reaction using 5-CQA and CoA, both the native and K-mutant HCTs produced low levels of caffeoyl-CoA and free caffeic acid, indicating only a partial conversion of 5-CQA to the thioester product.

In the forward reaction with caffeoyl-CoA and quinic acid, the H35A mutant showed a relatively weak catalytic activity, with only a partial conversion of caffeoyl-CoA into 5-CQA, and caffeic acid. With shikimic acid, the H35A mutant were also weakly catalytically active (data not shown). The H35A mutant showed no detectable activity in the reverse reaction.

In the forward reaction, the mutants D157A and R374E only produced caffeic acid in the presence of caffeoyl-CoA and quinic acid. The Y252A mutant produced caffeic acid and a small amount of 5-CQA. In comparison to the H35A mutant, these mutants fully transformed the caffeoyl-CoA substrate provided into caffeic acid. In the presence of caffeoyl-CoA and shikimic acid, the D157A mutant HCT exclusively formed caffeic acid (data not shown). The Y252A and R374E mutants were not tested with caffeoyl-CoA and shikimic acid. In the reverse reaction, D157A, Y252A and R374E mutants showed no activity towards 5-CQA and CoA as no peak of caffeoyl-CoA was detected. It is noted however that Y252A did produce caffeic acid. These results suggest that these mutants are not effective in the reverse reaction involving the quinate ester. Most probably, the Y252A mutant is capable of making caffeoyl-CoA, which is subsequently degraded into caffeic acid via a putative lyase activity.

In accordance with its proposed role in catalysis, the H153A mutant was inactive in the forward reaction involving caffeoyl-CoA and quinic or shikimic acid. As expected, this mutant is also inactive in the reverse direction with 5-CQA and CoA. No caffeic acid was detected. Most interesting are the H154N and H154N/A155L/A156S mutant HCTs, which formed 5-CQA and diCQAs from caffeoyl-CoA and quinic acid with a major peak corresponding to the 3,5-diester. The triple mutant also formed a small amount of free caffeic acid. In the reverse direction with 5-CQA and CoA, the two mutants formed a low level of caffeoyl-CoA and a relatively significant amount of diCQAs, again suggesting the importance of the 5-CQA/CoA molar ratio. The major diCQA peak corresponds to the 3,5-diester from which low amounts of the other two peaks, the 3,4- and 4,5-diCQA isomers, were presumably derived by chemical isomerisation. Again, the triple mutant also produced a small amount of free caffeic acid.

Figure 65: Comparison of the activities of native and mutant HCTs in the forward and reverse reactions

Reaction mixtures are described below. Reactions were started by adding 5 µM enzyme. Samples taken at t= 210 min were analysed using HPLC-PDA with the acetonitrile method.

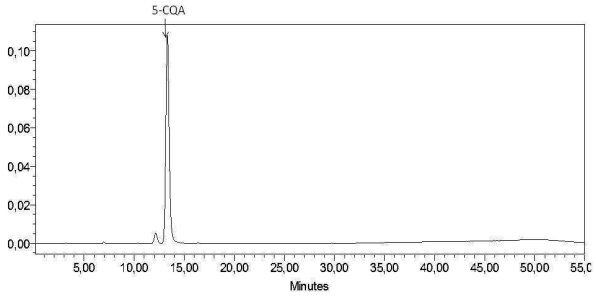
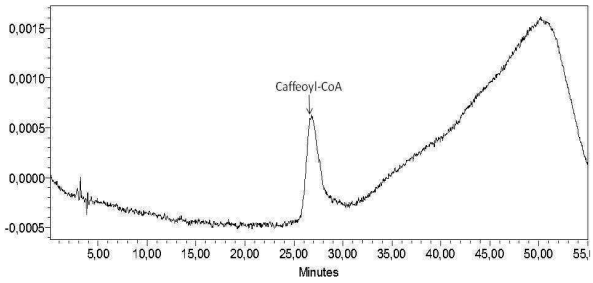
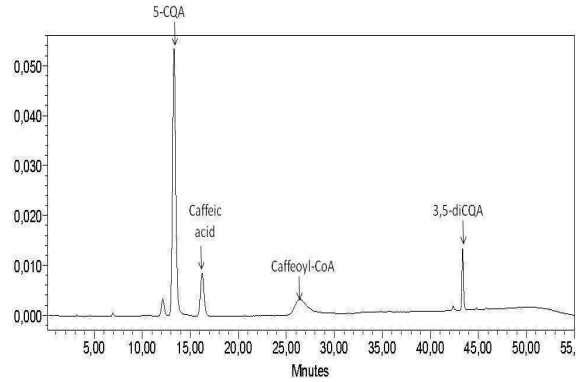
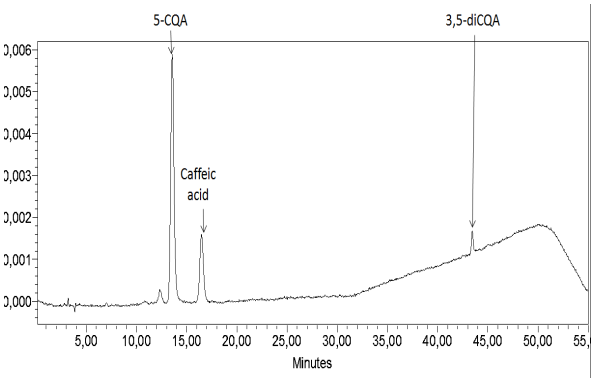
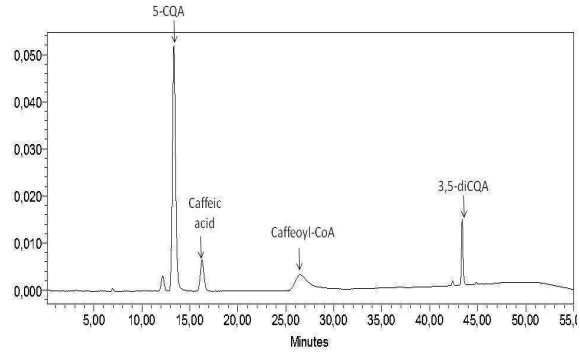
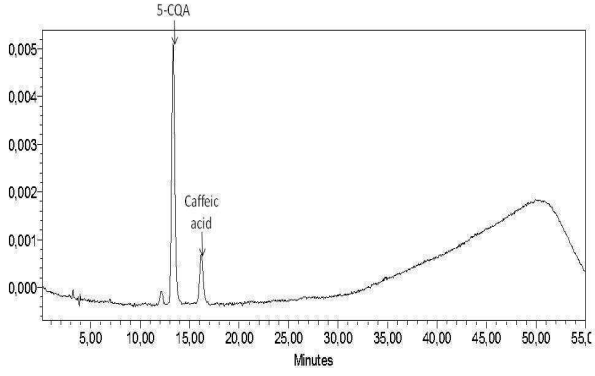
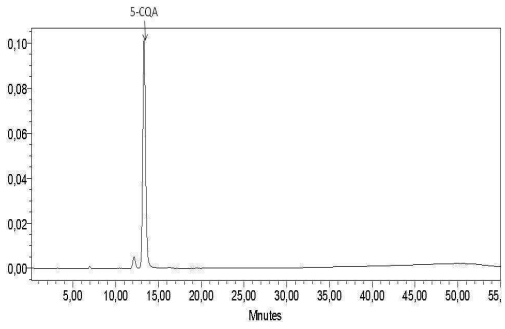
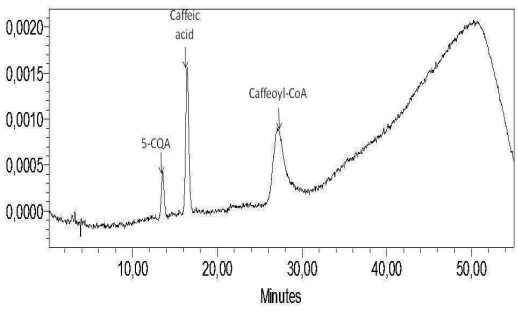
Reverse reaction	Forward reaction
Reactions were set up with 1 mM 5-CQA and 1 mM CoA in 0.1 M sodium phosphate pH 6.5.	Reactions were set up with 0.1 mM caffeoyl-CoA and 1 mM quinic acid in 0.1 M sodium phosphate pH 6.5.
Control with no enzyme (sample1321)	Control with no enzyme (sample 1337)
	
native HCT (sample 1322)	native HCT (sample 1338)
	
K-mutant (sample 1323)	K-mutant (sample 1339)
	
H35A mutant (sample 1326)	H35A mutant (sample 1342)
	

Figure 65 (continued)

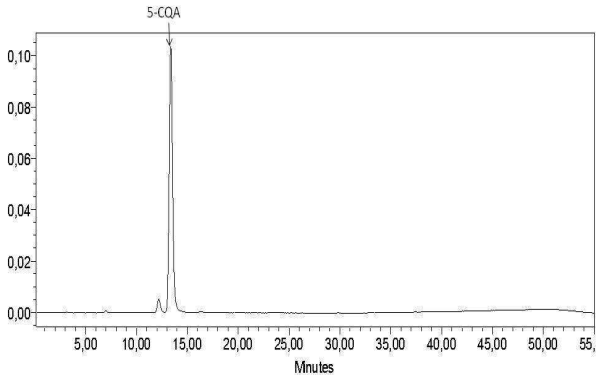
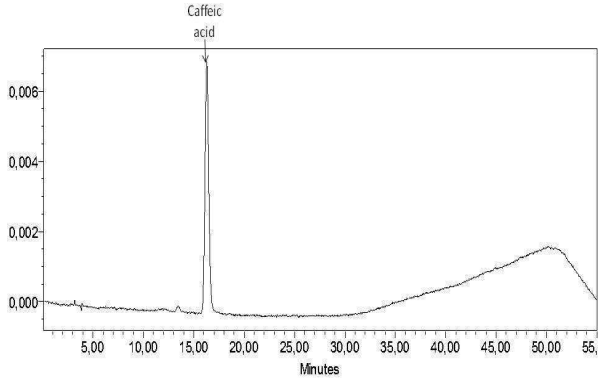
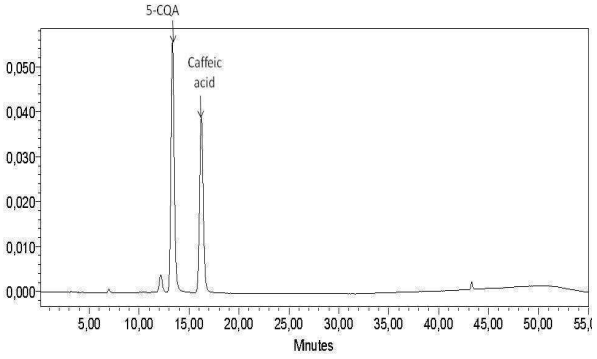
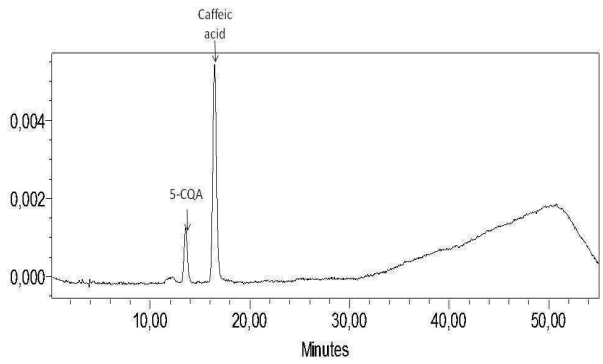
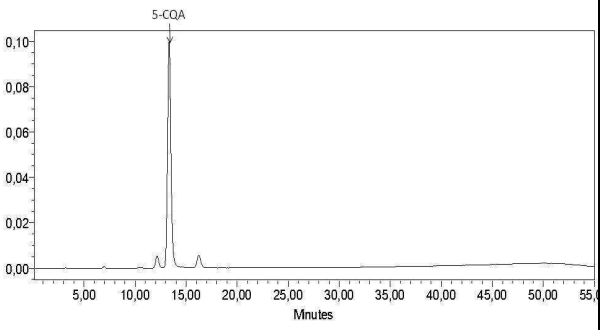
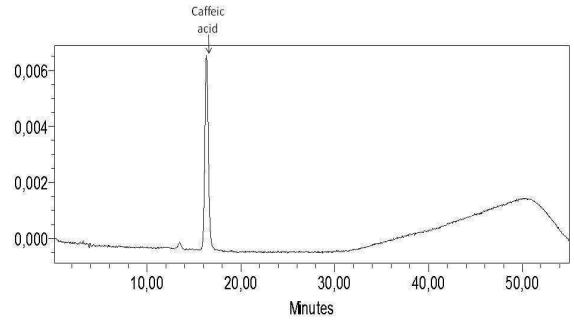
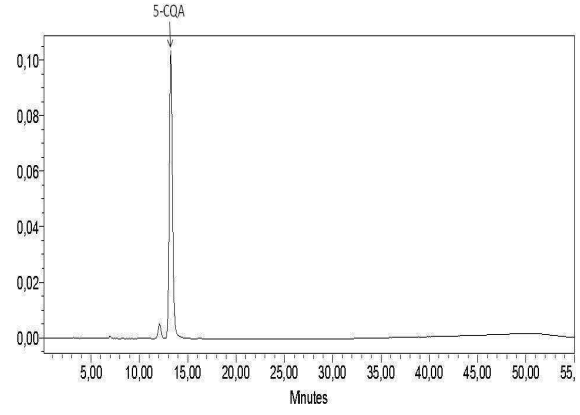
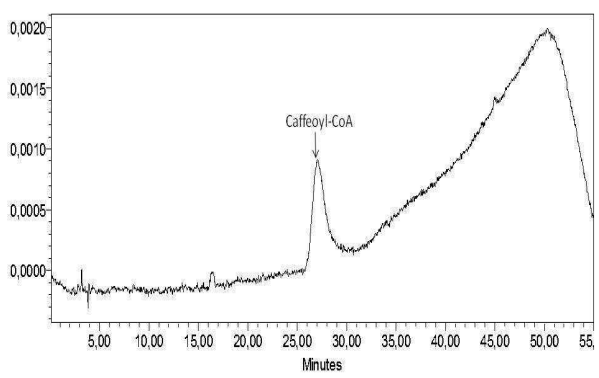
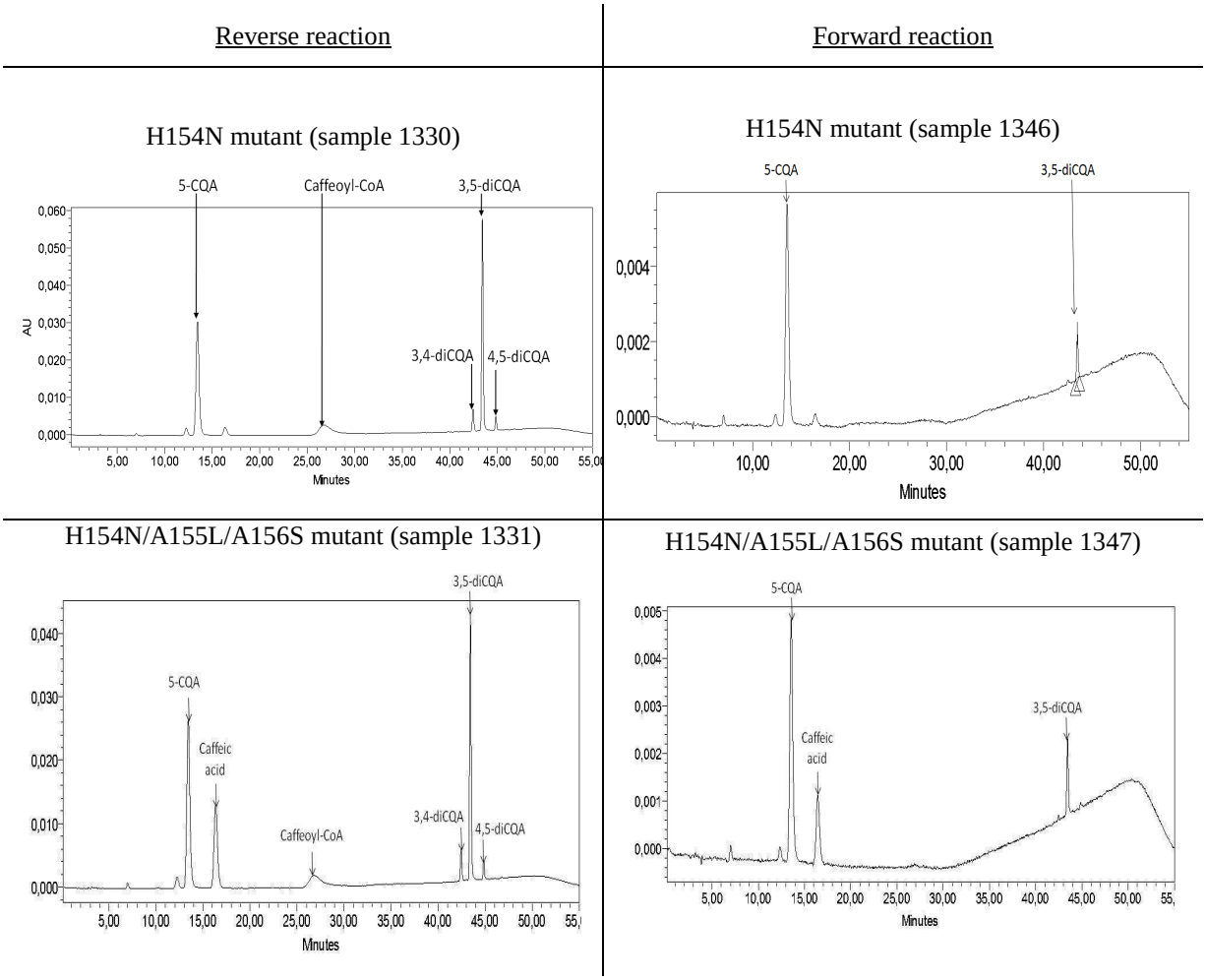
Reverse reaction	Forward reaction
D157A mutant (sample 1325)	D157A mutant (sample 1341)
	
Y252A mutant (sample 1327)	Y252A mutant (sample 1343)
	
R374E mutant (sample 1328)	R374E mutant (sample 1344)
	
H153A mutant (sample 1324)	H153A mutant (sample 1340)
	

Figure 65 (continued)



3.3.5.2. Additional analysis of the H154N and H154N/A155L/A156S mutant HCTs in the reverse reaction

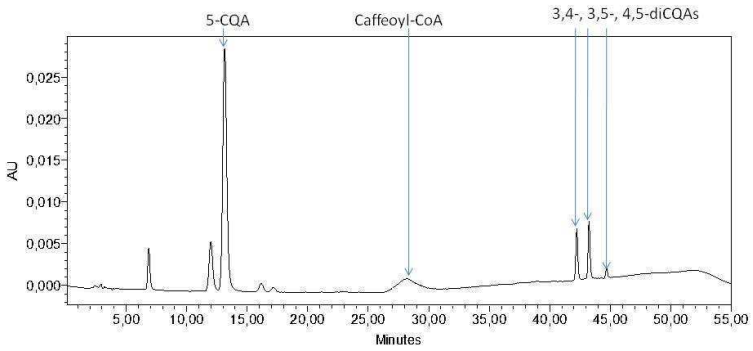
The results presented in Figure 65 suggest that the mutants targeting His154 favour the synthesis of diCQAs when compared to native *Cc*HCT. The two mutants targeting the His154 residue were thus further studied by doing a new comparison between the native HCT and the H154N and H154N/A155L/A156S mutant HCTs. In the forward reaction, caffeoyl-CoA and quinic or shikimic acid or both were supplied. The results showed that these mutants have comparable activities to the native enzyme, namely they both continue to show a clear preference for shikimic over quinic acid with caffeoyl-CoA as the second substrate. Caffeic acid formation was observed in the triple mutant only. In the forward reaction with caffeoyl-CoA and shikimic acid, a major peak was detected that is thought to correspond to a caffeoylshikimic acid isomer and two other smaller peaks that are thought to be the two other CSA isomers were seen (data not shown). Again, this result is similar to results obtained with

native *Cc*HCT (Figure 47). The effect of different 5-CQA/CoA molar ratios (1/1 and 1/10) supplied in the reaction was also tested. This experiment confirmed that both mutants again produced diCQAs. The data in Figure 66 also shows that adding more CoA slightly increased the amount caffeoyl-CoA produced, and that the level of diCQA decreases.

Figure 66: HPLC profiles of H154N and H154N/A155L/A156S mutant HCTs incubated with 5-CQA and CoA at 1/1 and 1/10 molar ratios

Reactions were set up with 1 mM 5-CQA and 1 or 10 mM CoA in 0.1 M sodium phosphate pH 6.5. Reactions were started by adding 0.5 µM enzyme. Samples taken at t= 180 min were analysed using HPLC-PDA with the acetonitrile method. The t= 0 is the same as in Figure 64.

(A) H154N 1/1 5-CQA/CoA (sample 1093)



(B) H154N 1/10 5-CQA/CoA (sample 1094)

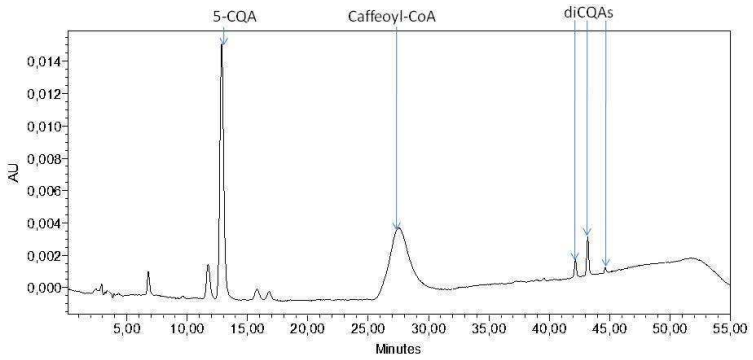
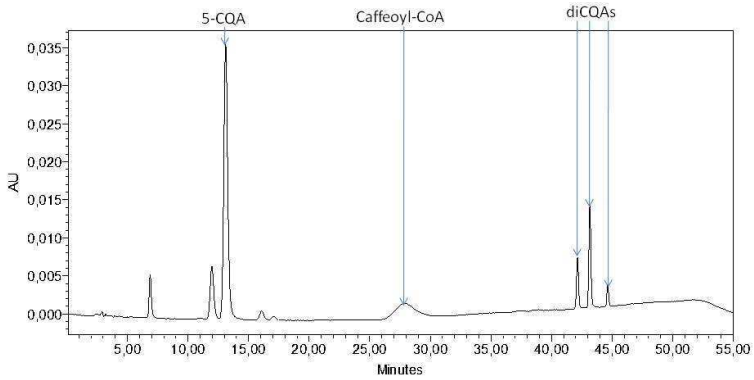
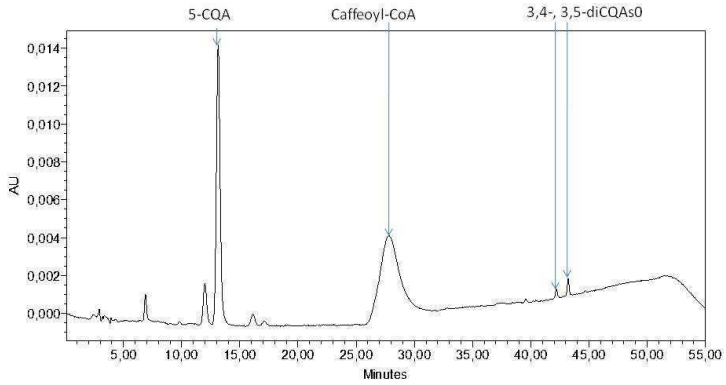


Figure 66 (continued)

(C) H154N/A155L/A156S 1/1 5-CQA/CoA (sample 1097)



(D) H154N/A155L/A156S 1/10 5-CQA/CoA (sample 1098)



3.3.5.3. Study of the formation of mixed diesters

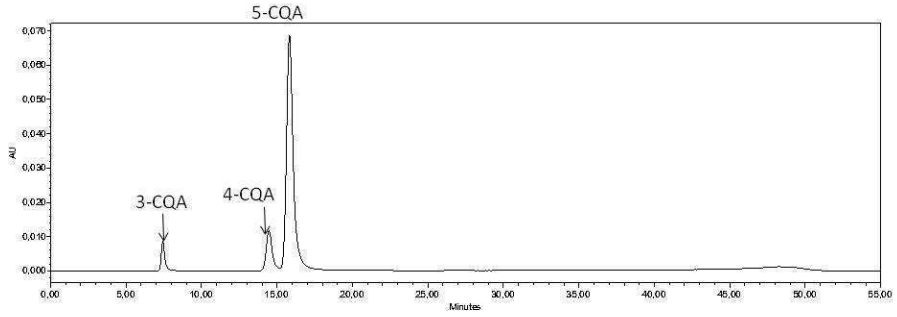
In order to determine if the native and H154N mutant HCTs could form mixed diesters of quinic acid (e.g. caffeoylferuloylquinic acids found in coffee extracts), the enzymes were incubated with 5-CQA/ 5-FQA and CoA/ caffeoyl-/ feruloyl-/ coumaroyl-CoA thioesters. As in the previous experiment, the H154N mutant HCT formed higher levels of these diesters. The results concerning the reactions containing 5-CQA and CoA or diverse acyl-CoA thioesters are presented in Figure 67. The control with no enzyme (Figure 67A) and all enzymatic reactions showed a partial isomerisation of 5-CQA (RT= 15.9 min) to the 3- and 4-CQA isomers (RT= 7.4 and 14.5 min) after overnight incubation in 0.1 M sodium phosphate pH 6.5. In the presence of 10 mM 5-CQA and 0.5 mM CoA (Figure 67B), as well as 10 mM 5-CQA and 0.5 mM caffeoyl-CoA (Figure 67C), three peaks corresponding to 3,4-, 3,5- and 4,5-diCQAs (RT= 41.4, 42.4 and 43.6 min respectively) were formed with native HCT. Interestingly, when 5-CQA and feruloyl-CoA are incubated with H154N mutant HCT, all three diCQAs are synthesised. Three additional peaks formed at RT= 43.9, 45.0 and 45.9 min may correspond to mixed caffeoylferuloylquinic acids. When 5-CQA and coumaroyl-CoA are

incubated with the H154N mutant HCT, the three diCQAs are formed (RT= 41.4, 42.4 and 43.6 min), plus two additional peaks at RT= 44.6 and 45.7 min that may correspond to mixed caffeoylcoumaroylquinic acids (Figure 67E).

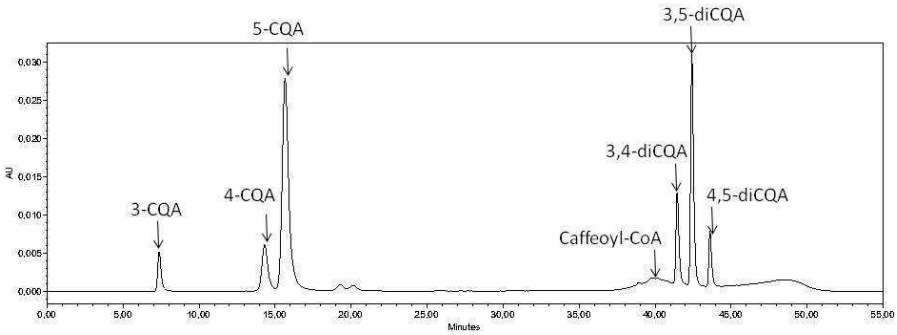
Figure 67: HPLC profiles of reaction products generated by the H154N mutant HCT from 5-CQA and various acyl donors

Reactions were set up with 10 mM 5-CQA and either 0.5 mM CoA/ caffeoyl-/ feruloyl-/ coumaroyl-CoA in 0.1 M sodium phosphate pH 6.5. The reaction was started by adding 0.5 µM enzyme. Samples were analysed after overnight incubation using HPLC-PDA with the acetonitrile method.

(A) Control with no enzyme 5-CQA + CoA (sample 1234)



(B) 5-CQA + CoA → caffeoyl-CoA + diCQAs (sample 1258)



(C) 5-CQA+ caffeoyl-CoA → diCQAs (sample 1259)

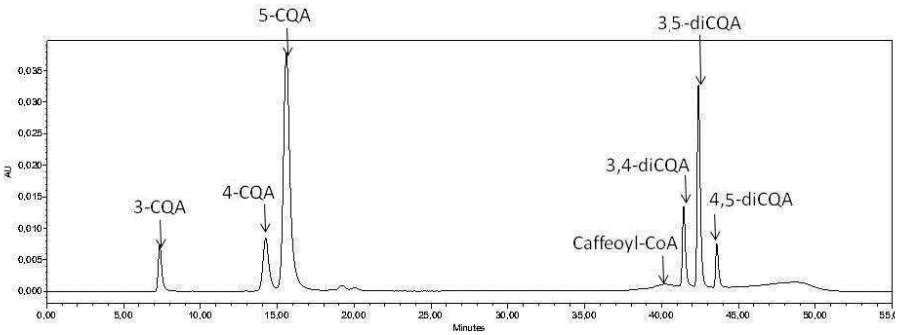
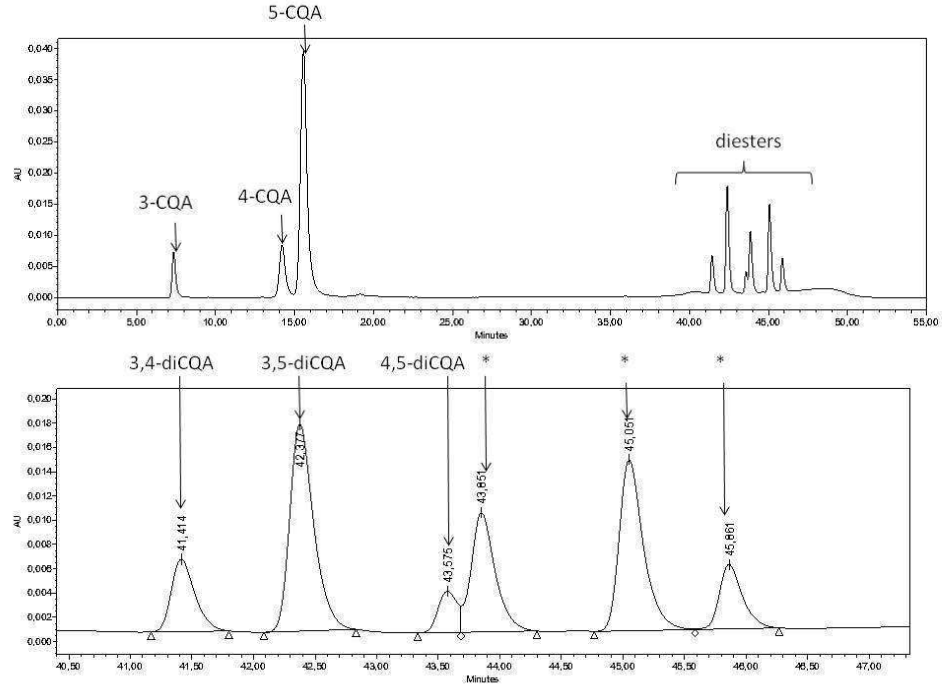
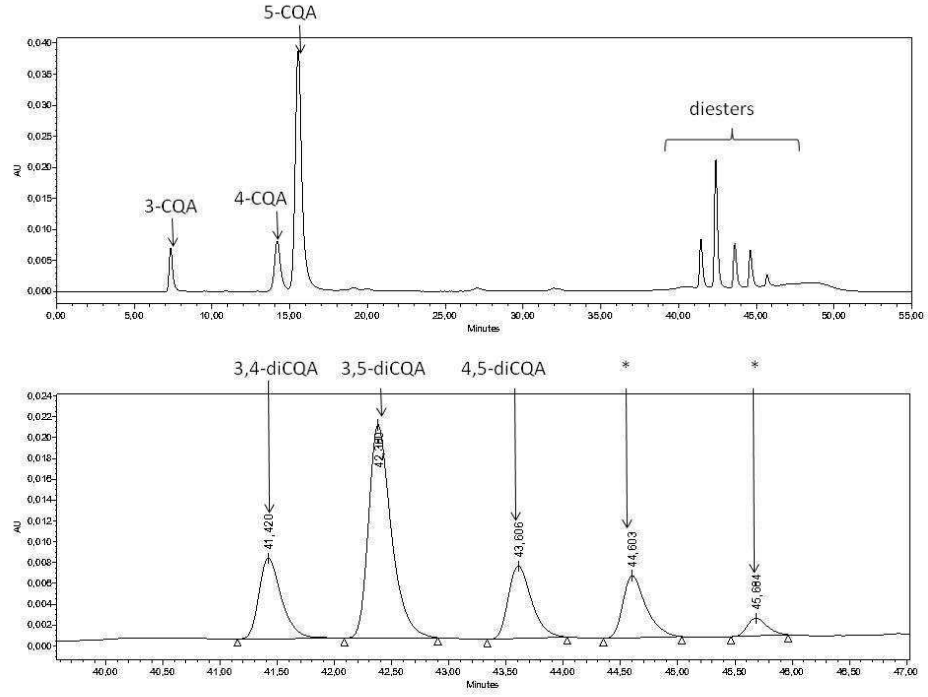


Figure 67 (continued)

(D) 5-CQA + feruloyl-CoA → diCQAs + mixed caffeoylferuloylquinic acids (*) (sample 1260)



(E) 5-CQA + coumaroyl-CoA → diCQAs + mixed coumaroylcaffeoylquinic acids (*) (sample 1261)

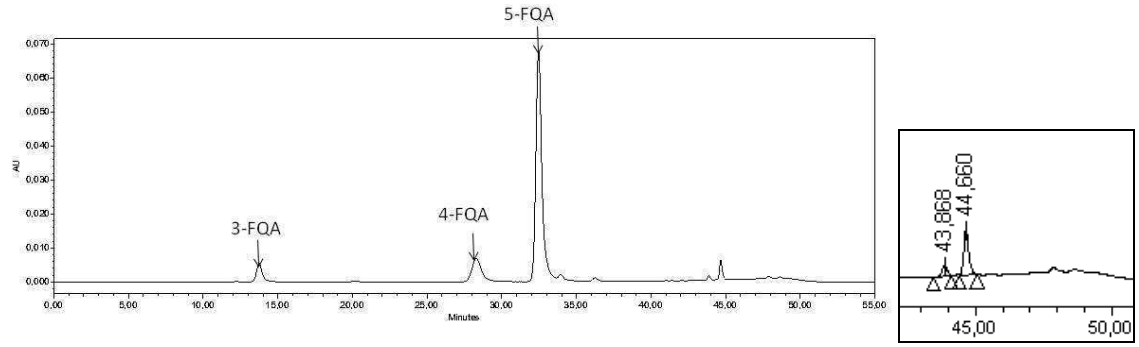


Similarly, the H154N mutant HCT was incubated with 5-FQA and CoA or diverse acyl-CoA thioesters (Figure 68). The control with no enzyme (Figure 68A) and all enzymatic reactions showed a partial isomerisation of 5-FQA (RT= 32.3 min) to the 3- and 4-FQA isomers (RT= 13.7 and 28.2 min) after overnight incubation in 0.1 M sodium phosphate pH 6.5. Additionally, 4 unknown peaks (RT= 43.7, 44.5, 47.7 and 48.4 min) were observed at later retention times, but these showed a maximum absorbance value at 256 nm, so they were not CGA molecules. Consequently, these peaks are not described in the following enzymatic reactions. With 5-FQA and either CoA or feruloyl-CoA, a new peak appeared at RT= 46.3 min that may correspond to a diferuloylquinic acid (Figure 68B and D). When 5-FQA and caffeoyl-CoA were incubated, 4 new peaks at RT= 44.2, 45.3, 46.1 and 46.3 min were observed. These were presumed to be the same diFQA (RT= 46.3 min) and mixed caffeoylferuloylquinic acids. When 5-FQA and coumaroyl-CoA were incubated, the same diferuloylquinic acid (RT= 46.3 min) and presumably mixed coumaroylferuloylquinic acids were obtained (RT= 45.4 and 47.3 min).

Figure 68: HPLC profiles of the reaction products synthesised by H154N mutant HCT from 5-FQA and various acyl donors

Reactions were set up with 10 mM 5-FQA and either 0.5 mM CoA/ caffeoyl-/ feruloyl-/ coumaroyl-CoA in 0.1 M sodium phosphate pH 6.5. The reaction was started by adding 0.5 µM enzyme. Samples were analysed after overnight incubation using HPLC-PDA with the acetonitrile method.

(A) Control with no enzyme 5-FQA + CoA (sample 1238)



Control with no enzyme 5-FQA + coumaroyl-CoA (sample 1241)

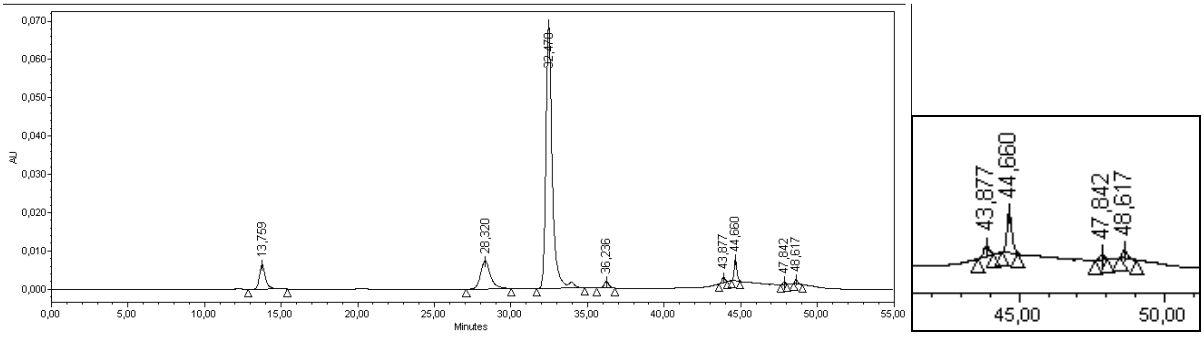
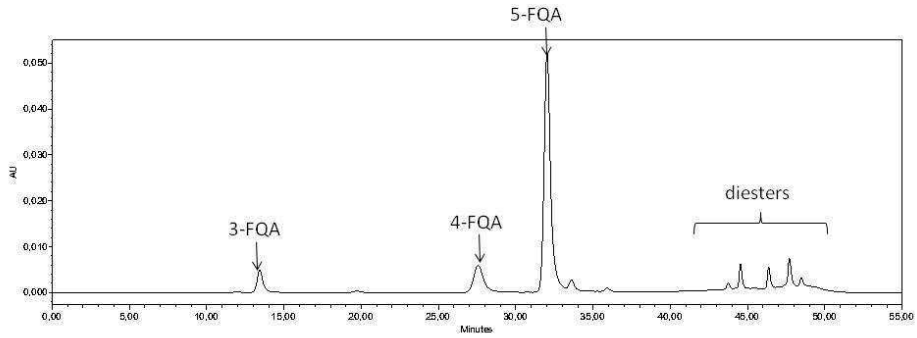
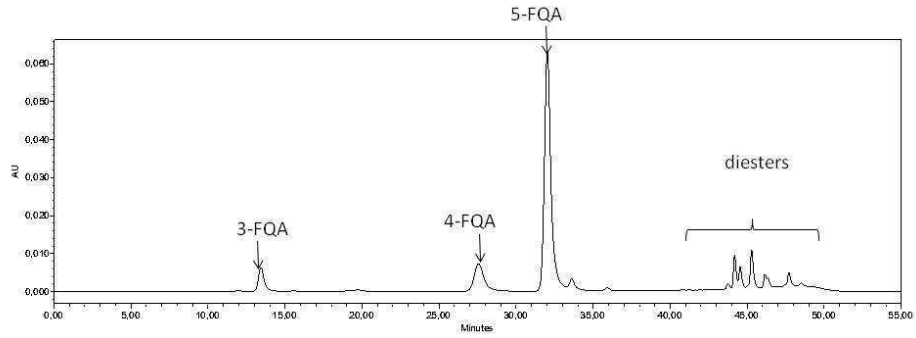


Figure 68 (continued)

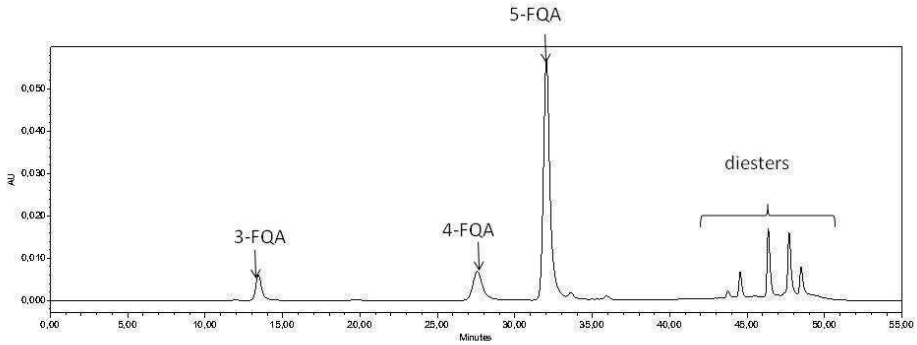
(B) 5-FQA + CoA → new peaks = diFQAs? (sample 1262)



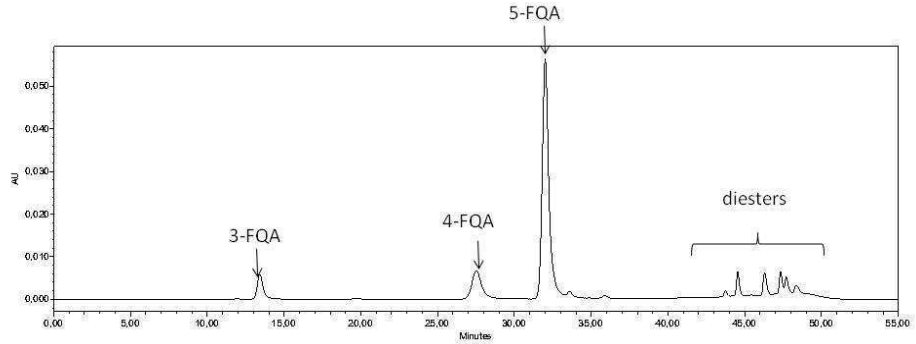
(C) 5-FQA + caffeoyl-CoA → new peaks = caffeoylferuloylquinic acids? (sample 1263)



(D) 5-FQA + feruloyl-CoA → new peaks = diFQAs? (sample 1264)



(E) 5-FQA + coumaroyl-CoA → new peaks = coumaroylferuloylquinic acids? (sample 1265)



In summary, the H154N and H154N/A155L/A156S mutant HCTs seem to favour diCQA formation. When feruloyl- or coumaroyl-CoA thioesters were supplied with 5-CQA/ 5-FQA, new peaks with a typical quinate ester absorbance spectrum were detected at late

retention times. They were presumed to be new esters of quinic acid. New peaks were observed and putatively attributed to either diferuloylquinic acids (diFQAs) (RT= 45.5, 46.3 and 47.5 min) or mixed diesters, feruloylcaffeoylquinic (RT= 44.2, 45.3 and 46.1 min) and feruloylcoumaroylquinic acids (RT= 45.4, 46.3 and 47.3 min), their identity should to be confirmed by mass spectrometry analysis. These assays show that these enzymes may be involved in the biosynthesis of not only caffeoylquinic, feruloylquinic and coumaroylquinic acids but also mixed diesters that are present in coffee (e.g. caffeoylferuloylquinic acids). The identity of the new products that were formed must be confirmed by comparison with authentic standards or by mass spectrometry analysis.

3.4. X-ray diffraction analysis

3.4.1. Tobacco 4CL2 crystal structures

3.4.1.1. Crystallisation of Nt4CL2

*Nt*4CL2 crystallised both in its apo-form and in a ternary complex with CoA and AMP. Similar crystallisation conditions were used for both, except that strontium chloride was used as an additive to obtain the higher resolution apo-*Nt*4CL2 crystal structure. However, no specific binding site for the strontium cation was observed. The two crystal forms belong to the same orthorhombic space group ($P2_12_12_1$), but they have different cell parameters and number of molecules per AU, one for the apo-enzyme and two for the ternary complex. In the crystal structure of the *Nt*4CL2-CoA-AMP ternary complex no electron density was observed for coumaric acid, although it was added in the crystallisation and cryo-protectant solutions.

3.4.1.2. Overall structure of Nt4CL2

*Nt*4CL2 exhibits an overall structure (Figure 69A) similar to those of other adenylate-forming enzymes and comprises two domains: a large N-terminal domain (residues 1-434) and a smaller C-terminal domain (residues 438-542). The N-terminal domain can be subdivided into three smaller subdomains each with a α/β topology, which are conserved in all adenylate-forming enzymes for which crystal structures are known (Conti *et al.*, 1996; Schmelz *et al.*, 2009). The C-terminal domain, which was only partially built in both the apo and ternary complex structures, forms a cap on top of the N-terminal core domain. A small linker region (Arg435-Leu436-Lys437) connects the two domains and contains an arginine residue that is invariant among adenylate-forming enzymes.

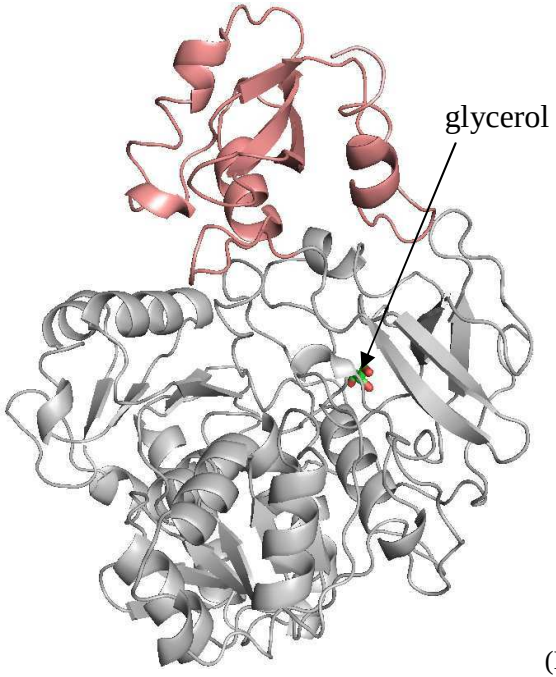
3.4.1.3. Comparison between the apo and ternary complex crystal structures

The crystal structures of apo-*Nt*4CL2 and Chain A of the ternary complex (Figure 69A and B) share a RMSD of 0.57 Å for 427 equivalent Cα atoms encompassing most of the N-terminal domain. In the crystal structure of the ternary complex, there is a large relative movement (relative to the apo crystal structure) of the C-terminal domain, which rotates as a rigid body in relation to the N-terminal domain (Figure 69C). The apo-enzyme thus adopts an "open" conformation, where the inter-domain region buries 797 $Å^2$ (Figure 69A). While the *Nt*4CL2-CoA-AMP ternary complex adopts a "closed" conformation, where the domain interface buries 953 $Å^2$ (Figure 69B). There are different direct inter-domain contacts in the apo and ternary complex structures: 5 H-bonds and an additional salt bridge (Lys375/Asp531) in the apo-enzyme; 9 H-bonds and two additional salt bridges (Glu115/Lys508 and Asp434/Arg533) in the complex form. The closed form of *Nt*4CL2 appears to be required for binding of substrate. Upon domain closure, the side-chain of Phe506, which belongs to the C-terminal cap domain and is solvent-exposed in apo-*Nt*4CL2, is positioned such that it forms a weak interaction with the CoA ribose moiety in the ternary complex structure (Figure 69E). The Lys441-Gly44 loop belonging also to the C-terminal domain becomes involved in CoA binding (Figure 69E).

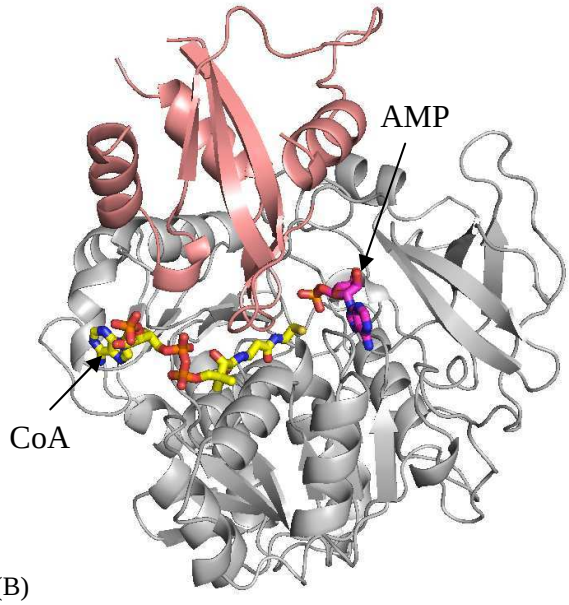
A few regions of the N-terminal domain also undergo significant movements when comparing the apo and ternary complex structures (Figure 69D). These include the Ser189-Leu195 loop and Ser307-Ala309 residues, the linker region and the side-chain of Tyr239. All these regions are located near the domain interface. In the ternary complex form, Tyr239 is buried inside a pocket, while it is solvent-exposed in apo-*Nt*4CL2 (Figure 69C and D). This movement is related to the displacement of the Ser307-Ala309 loop nearby. The Ser189-Leu195 loop also changes its conformation at the domain interface. This loop belongs to the Box I sequence motif, which is conserved among the adenylate-forming superfamily. This consensus sequence is reminiscent of the phosphate-binding "P-loop" found in many ATP-binding proteins (Saraste *et al.*, 1990). However, residues of the Ser189-Leu195 loop do not directly interact with the AMP nucleotide in the *Nt*4CL2-CoA-AMP ternary complex crystal structure. This is probably because this loop is involved in binding the β- and γ-phosphates of ATP, as observed in the crystal structure of ACSM2A-ATP (PDB: 3C5E), where it changes its conformation upon ATP binding (Kochan *et al.*, 2009).

Figure 69: Crystal structures of 4CL2-apo and ternary complex with CoA and AMP bound

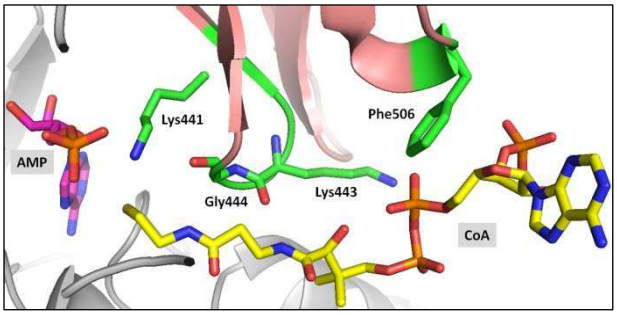
(A) Apo-enzyme structure with glycerol bound. The N-terminal (1-434) and C-terminal (438-538) domains are coloured in grey and salmon respectively. (B) The structure of the *Nt*4CL2-CoA-AMP ternary complex. CoA is in yellow and AMP in magenta. The colour scheme for the protein is as in (A). (C) Active site region of *Nt*4CL2-CoA-AMP with residues of the C-terminal domain interacting with CoA and AMP highlighted in green. (D) Superposition of the apo (salmon) and ternary complex (grey) structures. Tyr239, Phe506, the Ser307-Ala309 and Ser189-Leu195 loops are highlighted (apo, cyan; ternary complex, green). (E) Close-up of the superimposition of the N-terminal domains of apo and ternary complex crystal structures. The colour scheme is as in (D).



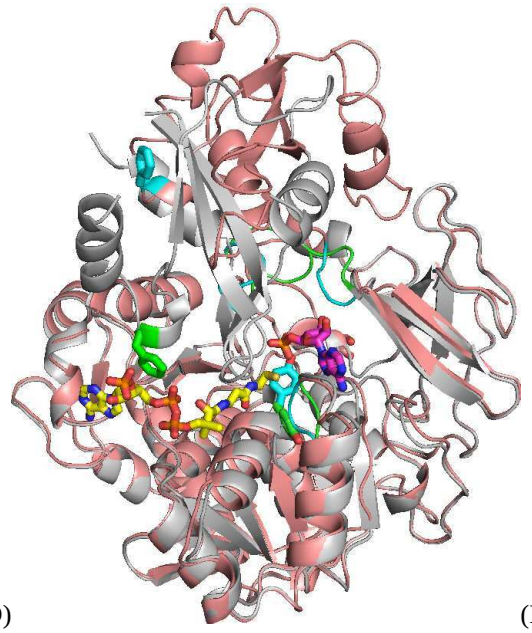
(A)



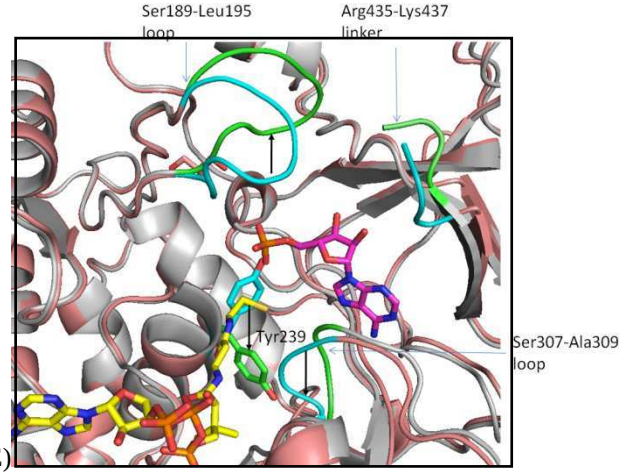
(B)



(C)



(D)



(E)

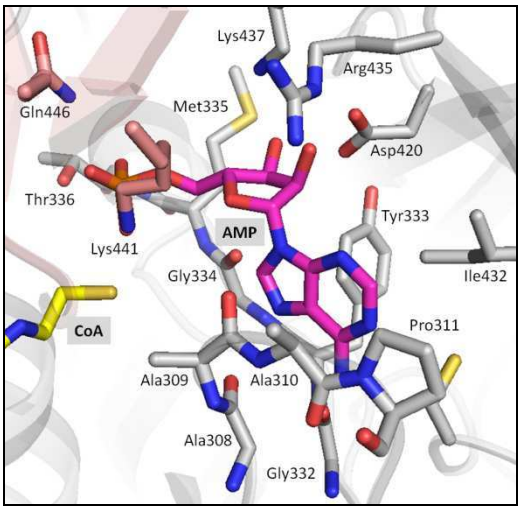
3.4.1.4. CoA and AMP binding sites

The ternary complex crystal structure revealed clear electron density for AMP and CoA (Figure 40). The binding sites for the nucleotide, which originated from the probable degradation of ATP in the crystal, and CoA, that is a substrate in the thioesterification step, are both located at the interface of the N- and C-terminal domains of the ternary complex (Figure 69B). This is in agreement with the previously described crystal structures of adenylate-forming enzymes (Schmelz *et al.*, 2009).

The residues lining the AMP binding pocket are: Gly308, Ala309, Ala310, Pro311, Gly332, Tyr333, Gly334, Met335, Thr336, Cys360, Asp420, Arg435, Lys437, Lys441 and Gln446 (Figure 70). The adenylate moiety is sandwiched between Tyr333 and Ile432 on one side and Ala309-Ala310 on the other side. Tyr333 is conserved in a number of adenylate-forming enzymes and is replaced by a phenylalanine in others. This aromatic residue is part of a glycine-rich loop (GQGYGMTE 330-337 in *Nt*4CL2), which acts as a nucleotide binding signature. The adenylate moiety forms additional H-bonds with the Gly308 and Gly332 carboxyl groups. The AMP ribose moiety is H-bonded to Asp420, Arg435 and Lys437 while its phosphate group interacts with the side-chains of Thr336 (O3P), Lys441 (O1P) and Gln446 (O1P) via H-bonding. Thr336 (part of the Box I motif) and Lys441 are conserved in 4CLs and luciferases, but is replaced by an Ala in CBL. Lys441 is substituted by Asn in ACSM2A and by Ile in CBL. Gln446 is generally not conserved in other adenylate-forming enzymes.

Figure 70: Representation of the AMP binding site in the *Nt*4CL2-AMP-CoA ternary complex crystal structure

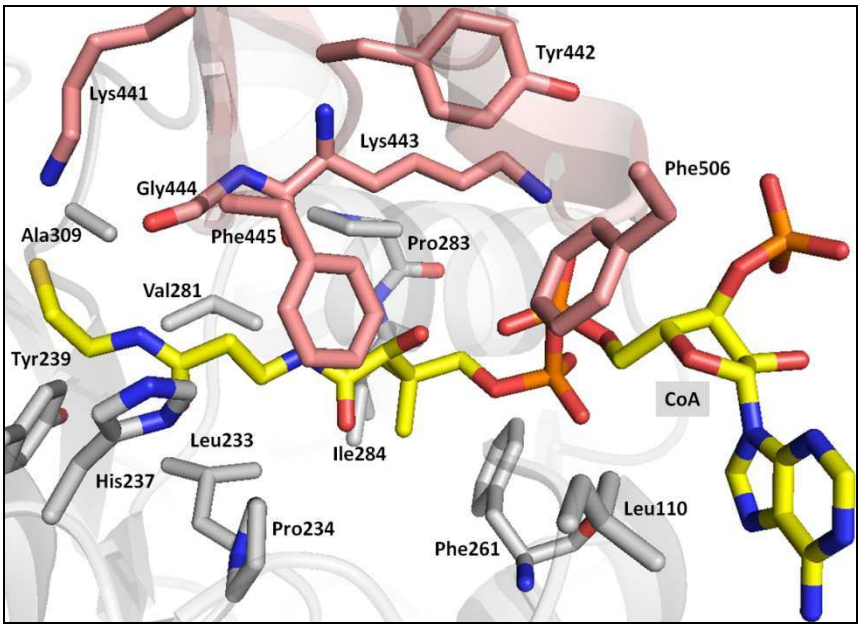
AMP is coloured magenta, CoA yellow, N-terminal residues grey and C-terminal residues salmon.



Residues within 5 Å of the CoA ligand include: Leu110, Leu233, Pro234, His237, Tyr239, Phe261, Val281, Pro283, Ile284, Ala309, Lys441, Tyr442, Lys443, Gly444, Phe445 and Phe506 (Figure 71). The adenosine moiety of CoA is solvent-exposed and makes only a weak van der Waals interaction with Leu110. Intriguingly, in the crystal, the CoA adenosine moiety also mediates a potential stacking interaction with another CoA molecule due to the crystal symmetry. In the CoA binding pocket, the CoA ribose interacts with the side chain of Phe506 and H-bonds are formed between the amide groups of the CoA pantetheine moiety and the main-chain carbonyl oxygen atoms of Lys443 and Gly444. These two residues are strictly conserved among the adenylate-forming enzyme superfamily and belong to the Lys441-Phe445 loop, which completes the CoA binding site upon closure of the C-terminal domain when compared to the apo-enzyme structure. Tyr442 is within H-bonding distance of the CoA pyrophosphate, as is the carbonyl group of Phe261. Finally, Phe445 interacts with the carbon structure between the two amide groups of CoA (Figure 71). Phe261 and Phe445 are conserved in all 4CLs (although it is replaced by an equivalent Tyr in *At*4CL4) and luciferases, but not in other adenylate-forming enzymes. Within the CoA molecule, H-bonds are formed between the oxygen of the pyrophosphate and hydroxyl group of the pantetheine moiety.

Figure 71: Representation of the CoA binding sites in the *Nt*4CL2-AMP-CoA ternary complex crystal structure

CoA is coloured in yellow, N-terminal residues in grey and C-terminal residues in salmon.



3.4.1.5. Putative catalytic residues

An analysis of the *Nt*4CL2-CoA-AMP ternary complex crystal structure suggests that Lys441 may be involved in catalysing the thioesterification step. The side-chain of Lys441 points towards both the CoA thiol and the AMP phosphate groups (Figure 70 and 71). This charged residue may increase the electrophilicity of the carboxyl group of the substrate-adenylate intermediate, favouring the nucleophilic attack of the CoA thiol moiety. The Lys441-Phe445 loop also the conserved Gly444. Interactions between this residue and CoA are important in optimally orienting the thiol group in the active site (Figure 71). The side-chains of Thr336 and Gln446, which interact via H-bonding with O3P and O1P oxygen atoms of the AMP phosphate group may be involved in correctly orienting this moiety in the active site.

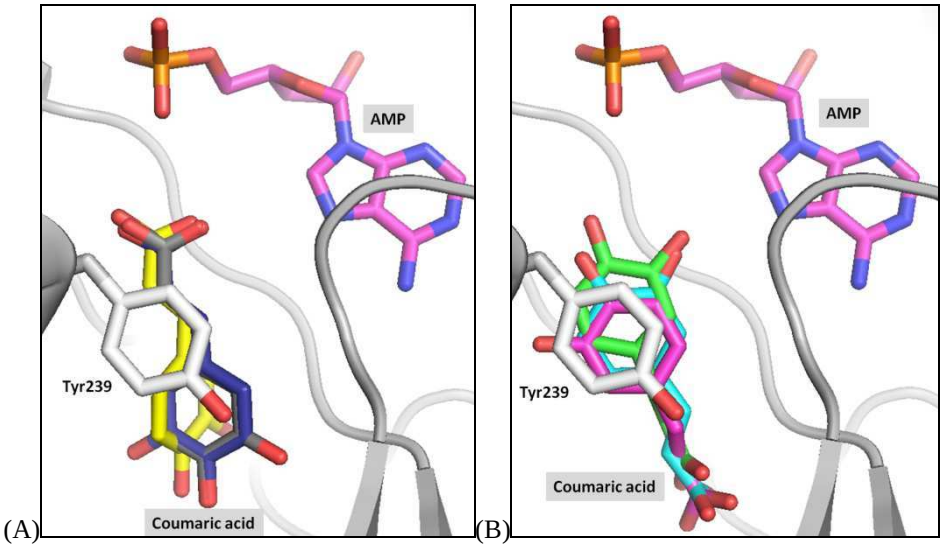
3.4.1.6. Docking of coumaric acid

Despite the fact that many attempts were made to introduce coumaric acid into crystals of apo *Nt*4CL2 and crystals of the *Nt*4CL2-CoA-AMP ternary complex, no density for this molecule was observed in the crystal structures. Docking experiments were therefore carried out using the *Nt*4CL2-CoA-AMP ternary complex crystal structure as a target for HCA binding.

The Pocket-Finder algorithm (Hendlich *et al.*, 1997) and a superimposition with PheA crystal structure (Conti *et al.*, 1997) were used to locate the phenolic substrate binding pocket (SBP) in chain A of the *Nt*4CL2-CoA-AMP crystal structure. HCAs were predicted to bind in a hydrophobic pocket formed by residues of the N-terminal domain: Gln213, Tyr239, Ser243, Met306, Ser307, Gly308, Gln331, Gly332, Gly334, Pro340, Val341, Met344 and Phe348. Docking calculations resulted in two clusters of solutions. One cluster has the carboxylate group of HCAs pointing towards the domain interface and CoA and AMP (Figure 72A), in the other cluster the HCA carboxylate group points away from these moieties (Figure 72B). For catalysis to occur the carboxylic group should be positioned close to the CoA thiol and AMP phosphate groups. Docking solutions were therefore selected according to the proximity of the carboxylic group with these latter moieties.

Figure 72: The two clusters of docking solutions for caffeic acid in the SBP predicted in *Nt*4CL2-CoA-AMP

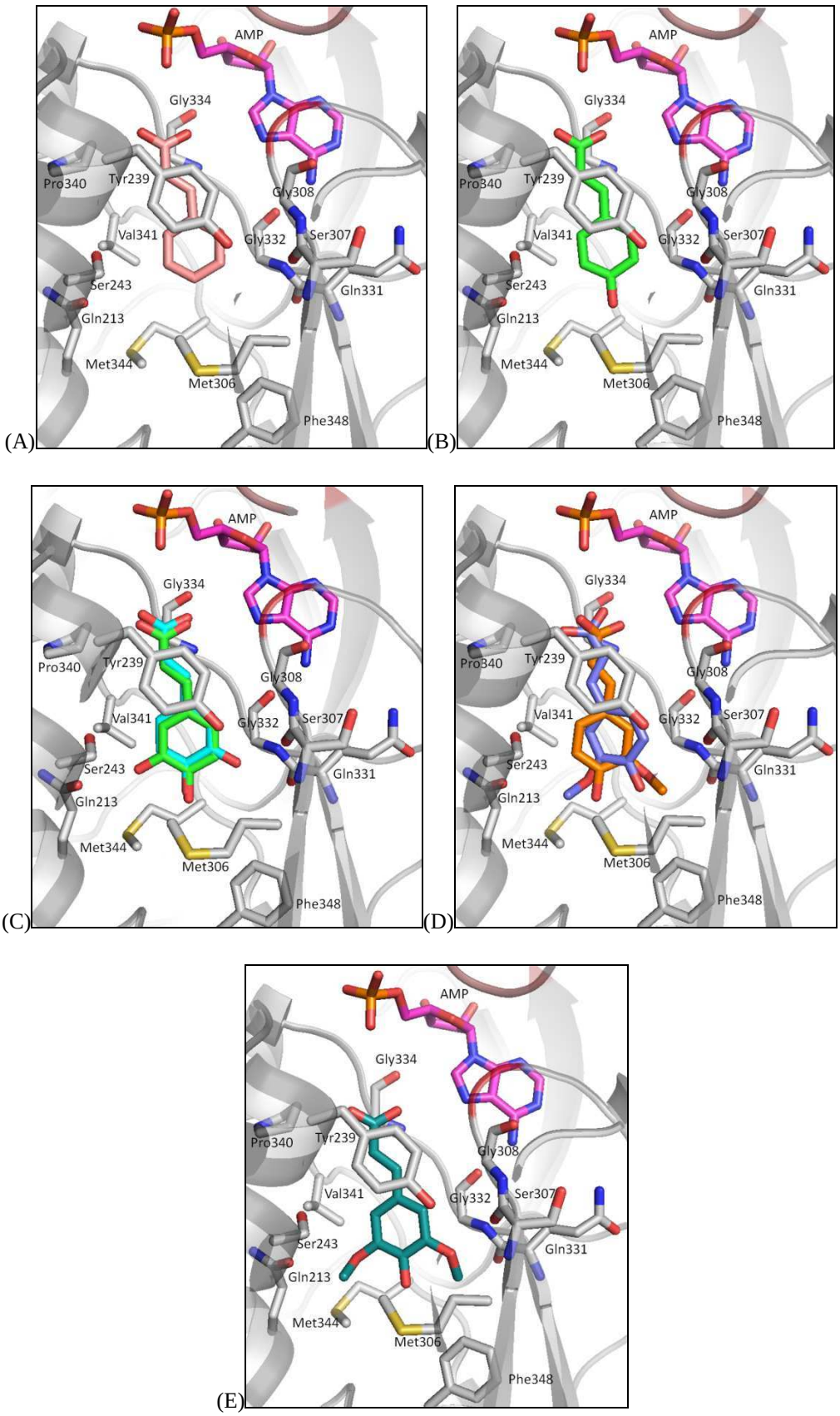
(A) Cluster 1. (B) Cluster 2.



Phenolic compounds are characterised by their hydrophobic benzenoid rings and the H-bonding potential of the hydroxyl groups. In the docking solutions chosen, the side chain of Tyr239 is ideally oriented to form a π-stacking interaction with the aromatic ring of the HCAs (Figure 73). Ideally, amino acid residues surrounding the C-3, C-4 and C-5 positions of the phenolic ring should have the capacity to interact with the HCA substrates. However, in the docking results, the C-4 hydroxyl group of coumaric, caffeic and ferulic acids are unable to be stabilised by any direct H-bonding interactions with the protein unless there is some rearrangement upon HCA binding. Moreover, C-3 and C-5 aromatic positions cannot be determined unambiguously, as shown by the two equivalent solutions for caffeic acid in Figure 73C. The Gln213 and Ser243 side-chains on one side of the SBP, or the carbonyl oxygen of Ser307 and amide hydrogen of Gly332 on the other, could potentially form H-bonds with the hydroxyl group at C-3 of caffeic acid. Docking calculations with ferulic and sinapic acids resulted in slightly different orientations of the aromatic ring, most likely because of the steric hindrance induced by the methoxyl groups at C-3 and C-5 (Figure 73D and E).

Figure 73: Cinnamic acid substrates docked in the *Nt*4CL2-CoA-AMP receptor using AutoDock Vina

*Nt*4CL2 is coloured in grey, AMP in magenta. The best solutions are represented for cinnamic (A), coumaric (B), caffeic (C), ferulic (D) and sinapic (E) acids.



3.4.2. Coffee HCT crystal structures

3.4.2.1. Crystallisation of native and K-mutant HCTs

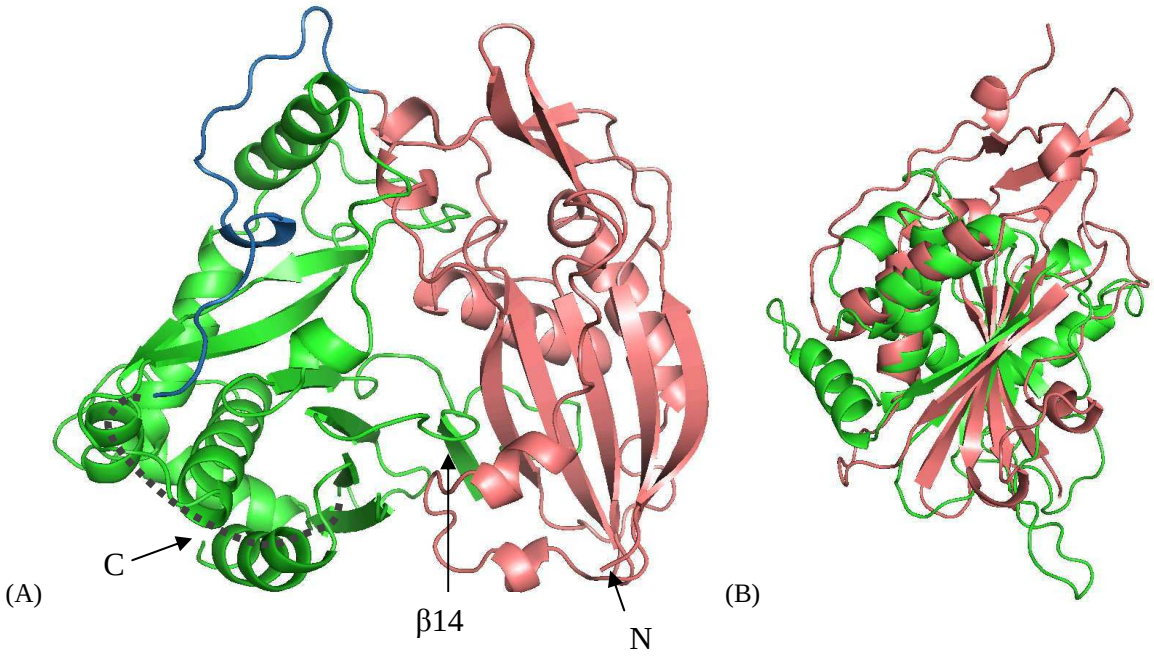
Native *Cc*HCT crystallised as thin triangular plates after several weeks at 20 °C. X-ray diffraction data to a maximum resolution of 3.0 Å were obtained. The crystals belong to a tetragonal space group ($P4_12_12$) and contain 2 molecules per AU (Table 9). The crystal structure of native *Cc*HCT was solved by MR using the crystal structure of VS (PDB: 2BGH) as a template. In order to obtain more reproducible crystals, two lysine residues (K210 and K217) located in the cross-over loop were mutated to alanine. The resulting mutant *Cc*HCT (K-mutant) crystallised after only a few days. Two crystal forms were observed: form 1 ($P2_12_12_1$) was preponderant and form 2 ($C222_1$) was obtained only occasionally in drops set up using the Cartesian robot of the EMBL-Grenoble HTX laboratory. Crystal form 1 contains two molecules per AU, while crystal form 2 contains only one monomer (Table 9). The K-mutant *Cc*HCT crystal structures were solved by MR using chain A of the native crystal as a template. Soaking and co-crystallisation experiments with CoA, 5-CQA and shikimic acid substrates did not result in any protein-ligand complexes.

3.4.2.2. Overall structure

The tertiary structure of *Cc*HCT is composed of two nearly equal-sized domains containing 16 β-strands and 19 α-helices (Figure 74A). The N-terminal domain is composed of two β-sheets, while the C-terminal domain contains one β-sheet. A strand formed by residues close to the C-terminal end of the protein (β14) participates in the formation of the main β-sheet from the N-terminal domain (Figure 74A). In both domains, α-helices pack mainly against one side of the principal β-sheet. The N- and C-terminal domains are structurally related to each other and can be aligned with a RMSD of 3.4 Å for 83 equivalent Cα atoms (Figure 74B). This domain similarity is a common structural feature of the BAHD superfamily, which has been proposed to evolve from two CAT monomers (1.6.2.71.6.2.7). In native *Cc*HCT, the domain interface buries an area of 3400 $Å^2$ stabilised with 45 H-bonds and 7 salt bridges, as calculated by PISA (Krissinel *et al.*, 2007). The domains are connected by a large cross-over loop between Pro206 to Thr224 that spans nearly 25 Å. However, there was no clear electron density for this loop from in any of the HCT crystal structures reported here. An inspection of the protein molecular surface (see below) reveals a clear solvent channel formed at the domain interface that is about 18 Å in length and 6-7 Å in width on both sides.

Figure 74: The overall structure of HCT is composed of two structurally related domains

(A) The N-terminal domain (residues 1-189) is coloured in salmon and the C-terminal domain (224-434) in green. The partially built cross-over loop is coloured in blue (190-210) and the missing residues in grey dots. The amino (N) and carboxyl (C) ends are indicated with arrows. (B) Superimposition of the N- and C-terminal domains of HCT.



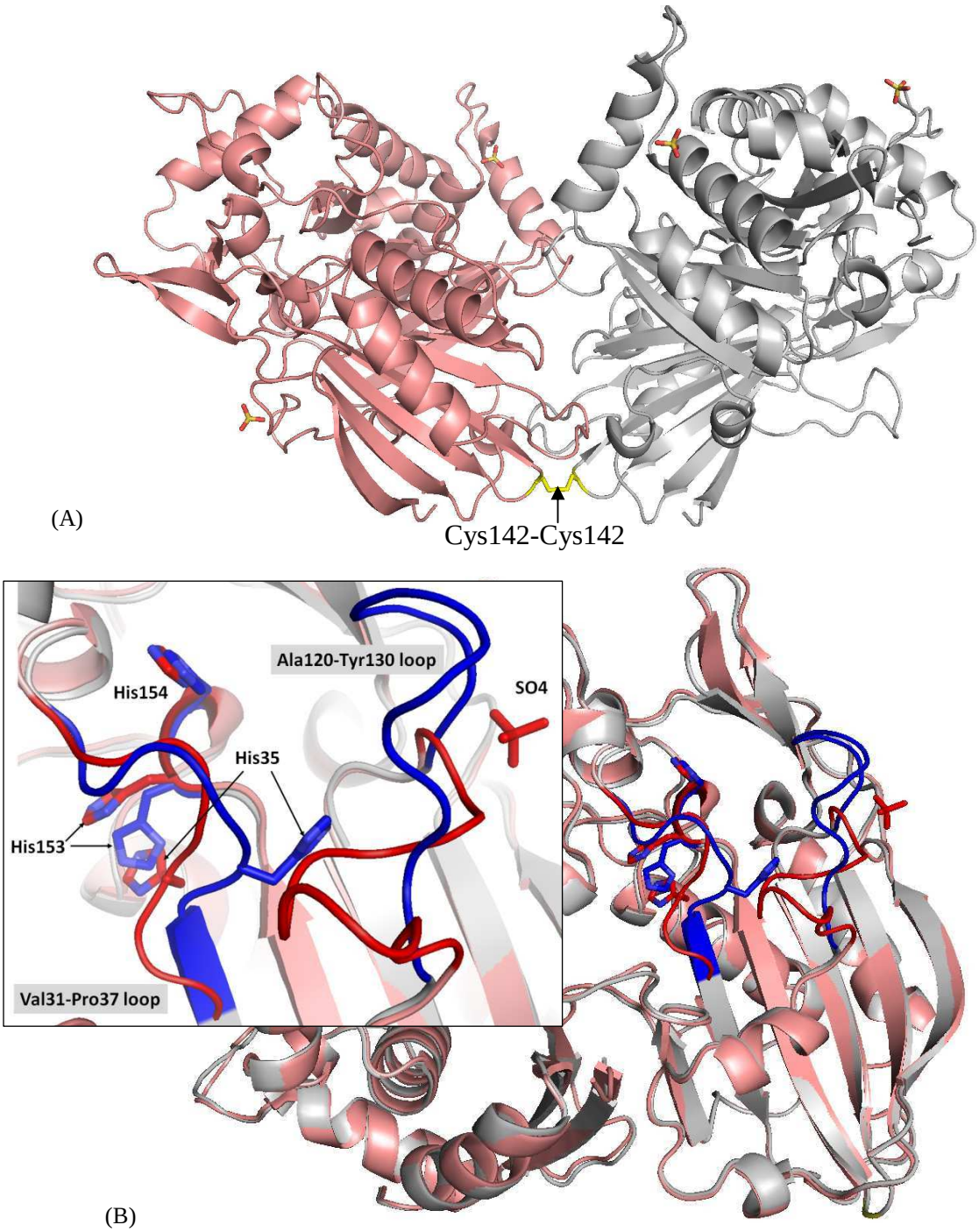
3.4.2.3. The crystal structure of native CcHCT

The main protein-protein interface in the crystal is formed between two B molecules, with a surface area of 1225 $Å^2$ strengthened by 8 H-bonds. The chain A/ B interface buries a surface area of 690 $Å^2$ per molecule and is stabilised by a covalent disulfide bond involving Cys142 of both chains (Figure 75A), 7 H-bonds and two salt bridges (Arg43/Asp103 and Asp410/Arg386). A sulfate ion is bound between two symmetry-related A molecules and interacts with the solvent-exposed Gln199 and Arg333 residues of both molecules. Two other sulfate molecules bind on the protein surface, one near Arg264, Lys242, Asp433 and Glu245, and the other near Lys240 and Ser254. Only a few structural differences were observed between the two monomers in the AU. These principally concern the Val31-Pro37 and Ala120-Tyr130 loops, and the side chain of His153 (Figure 75B). In chain B, a sulfate molecule originating from the magnesium sulfate contained in the crystallisation solution is bound on the protein surface and interacts with the side-chains of Ser129, Arg73, Glu89 and Met10. The sulfate binding most likely contributes to the different conformation of the Ala120-Tyr130 loop in chain B when compared to that in chain A where no sulfate is bound. This difference results in a conformational change of the solvent-exposed Val31-Pro37 loop and the subsequent re-orientation of the side-chain of the catalytic His153 (Figure 75B). In chain B, His153 interacts with His35 from the moving loop by π-stacking interactions

between the imidazole rings. Interestingly, Pro37 contribute to β3 strand formation in chain A, but not chain B (Figure 75B).

Figure 75: Interaction and comparison between the two molecules in the AU in native *Cc*HCT

(A) Chain A (grey) and chain B (salmon) are connected by a disulfide bond between the two Cys142. (B) Superimposition of chains A and B with the differences (chain A, blue; chain B, red) highlighted.



(A)

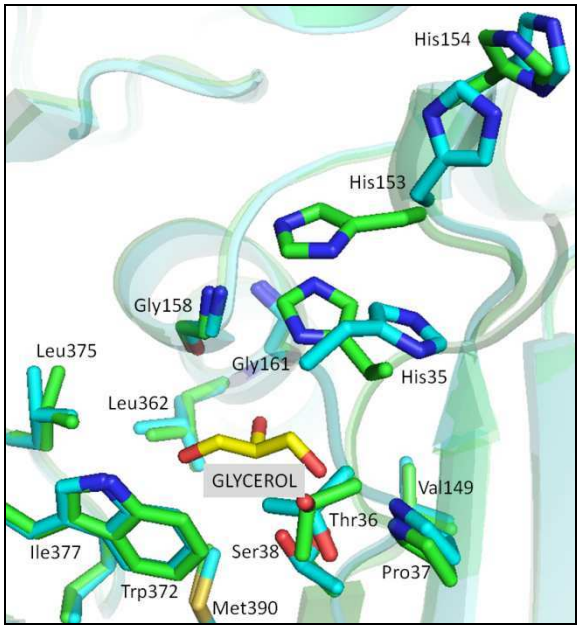
(B)

3.4.2.4. Crystal structures of K-mutant CcHCT

The two molecules in the AU of the K-mutant HCT crystal form 1 can be superposed with a RMSD of 0.42 Å for 417 equivalent Cα atoms. A difference Fourier map revealed density corresponding to a glycerol molecule, originating from the cryo-protectant mother liquor, bound in the active site channel of both chains in the AU of form 1 near residues Thr36, Pro37, Ser38, Gly158, Trp372 and Leu375 (Figure 76). The form 1 and form 2 crystal structures differ in the conformation of the Val31-Pro37 loop and the catalytic His153 side-chain (Figure 76).

Figure 76: Superimposition of K-mutant *Cc*HCT crystal forms 1 and 2 structures showing the active site region

Crystal form 1 is coloured in cyan, crystal form 2 in green. The glycerol bound to form 1 is shown in yellow.



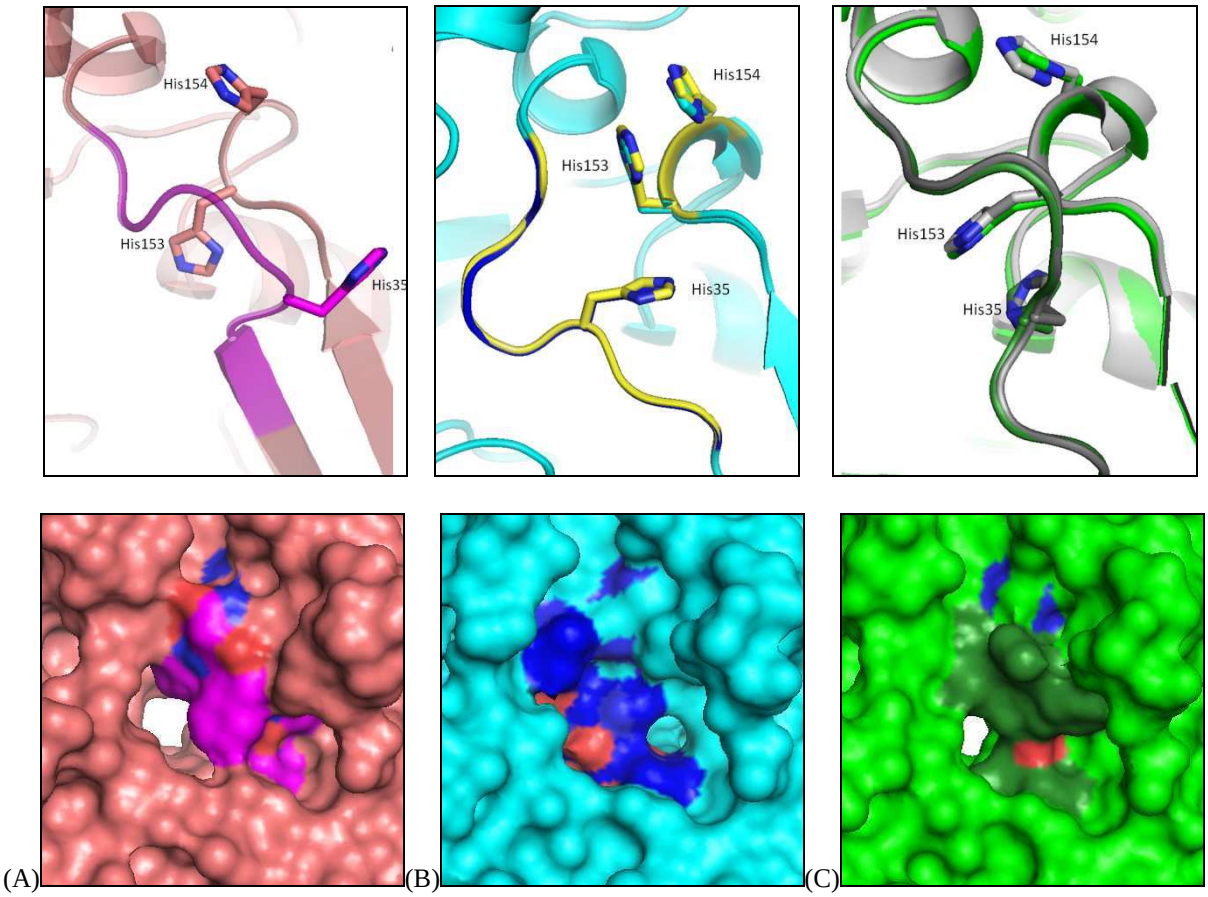
3.4.2.5. Comparison between the native and K-mutant CcHCT crystal structures

A superimposition of all HCT molecules observed in the crystal structures described here resulted in RMSD values between 0.42 Å (for 417 Cα atoms between the two chains in the AU of K-mutant form 1) and 0.72 Å (for 412 Cα atoms between native HCT chain A and mutant form 1 chain A/B). As also noted above, major differences are observed in the conformation of the Val31-Pro37 loop located in the solvent channel region (Figure 77). In the native HCT chain A structure, the three His residues are apart and the solvent channel is accessible (Figure 77A). In the K-mutant crystal form 1, in both chains in the AU, the Val31-Thr36 loop fully blocks the entrance to the active site (Figure 77B). This rearrangement also

results in the catalytic His153 forming a stacking interaction with His154. The Val31-Thr36 loop conformation in crystal form 2 and native HCT chain B may still allow access to the active site, but the catalytic His153 is now forming a stacking interaction with His35 (Figure 77C).

Figure 77: Comparison of the native and K-mutant *Cc*HCT crystal structures in the active site region

(A) Native *Cc*HCT chain A. (B) K-mutant crystal form 1, superimposition of chain A (cyan) and B (yellow). (C) K-mutant form 2 (green) superposed to native *Cc*HCT chain B (grey). All the structures are shown in the same orientation. Top: Conformation of the Val31-Pro37 loop. Bottom: Molecular surfaces viewed from the back face.



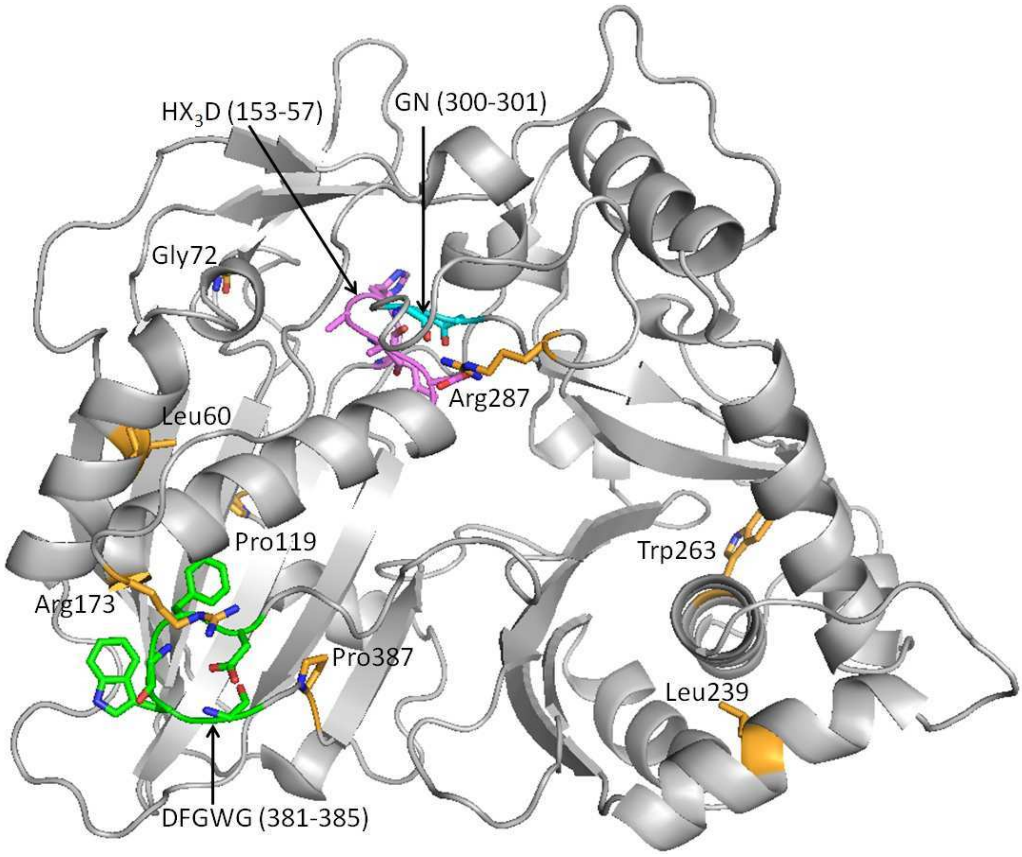
3.4.2.6. Structural comparison with other BAHD members

3.4.2.6.1. Conserved motifs and residues

The Dali server (Holm *et al.*, 2010) was used to compare the structure of native *Cc*HCT crystal structure against all the structures deposited in the PDB database (Berman *et al.*, 2002). Unsurprisingly, this comparison reveals that *Cc*HCT is structurally most similar to VS, MaT and TRI101, with RMSD values between 2.5 and 2.7 Å (Z-scores: 36.3, 35.8 and

34.4 respectively). A sequence alignment was generated using ClustalW2 (Larkin *et al.*, 2007) and ESPript 2.2 (Gouet *et al.*, 1999; Gouet *et al.*, 2003) (see Discussion section, Figure 96). This alignment was manually optimised using the secondary structures of *Cc*HCT (native chain A), VS (PDB: 2BGH), MaT (PDB: 2E1T) and TRI101 (PDB: 3B30). Only 14 amino acid residues are strictly conserved among BAHD members (Figure 78).

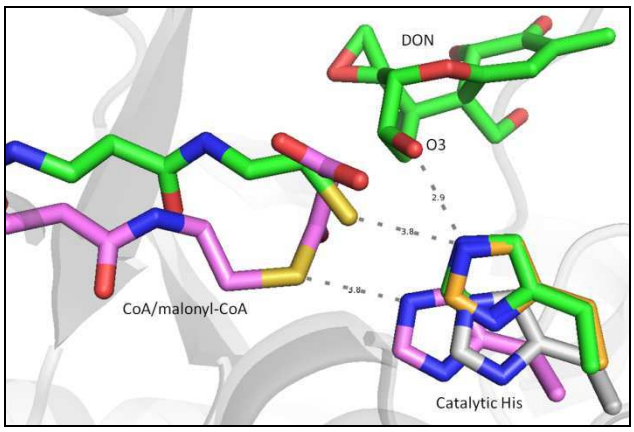
Figure 78: Native *Cc*HCT structure showing the conserved motifs and invariant residues among the BAHD superfamily



In BAHD structures, the HX_3D motif (Box I) is accessible via the solvent channel and contains the catalytic histidine residue. In both chains of the K-mutant crystal form 2, the catalytic His153 is flipped and stacks on His154. In chain B of native *Cc*HCT and in K-mutant form 2 crystal structures, the side-chain of His153 are oriented towards the solvent channel but stacks on His35 (Figure 77). The position of the imidazole ring of H153 in chain A of the native *Cc*HCT crystal structure is most similar to that seen in VS, MaT and TRI101 and this residue is appropriately placed to abstract a proton from potential substrates (Figure 79). The structure of *Cc*HCT as seen in chain A of the native crystal structure was therefore used in subsequent docking, modelling and structure comparisons.

Figure 79: Superimposition of *Cc*HCT, VS, MaT and TRI101 crystal structures in the active site region

Native *Cc*HCT chain A is coloured in grey, VS in orange, MaT-malonyl-CoA in violet and *Fg*TRI101-CoA-DON in green.



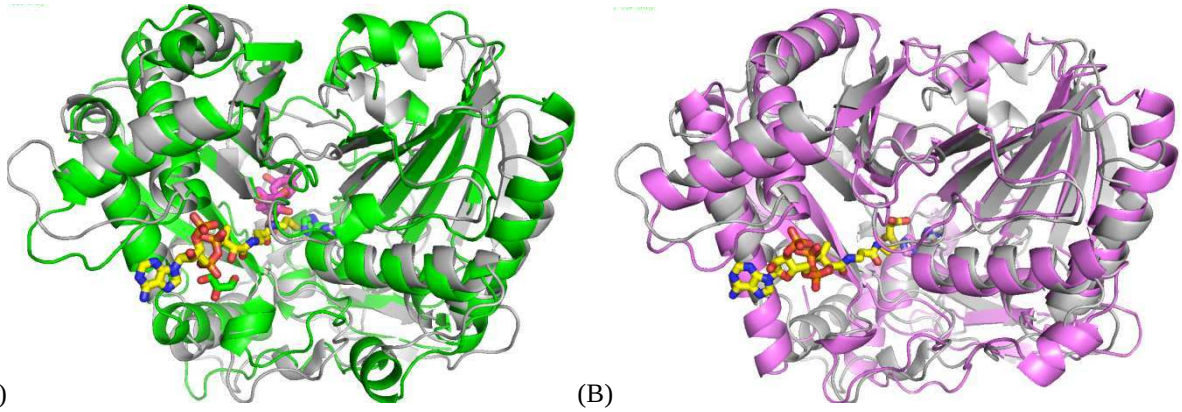
The HX_3D motif (Box I) also contains the conserved D157. The side-chain of this residue points away from the active site in *Cc*HCT and other BAHD crystal structures, where it forms a salt bridge a conserved residue of the DFGWG[R/K]P motif (Arg287 in *Cc*HCT). This interaction is probably involved in maintaining the two-domain fold of the protein. The Asn residue in the conserved Gly300-Asn301 motif is H-bonded with the main-chain oxygen of His153 only in chain B of the native HCT and the K-mutant HCT crystal structures. The DFGWG (381-385) motif is also strictly conserved among all plant BAHD enzymes. This motif is not located near the active site and most likely plays a structural rather than catalytic role (Figure 78). The role of the other conserved residues (Leu60, Gly72, Pro119, Arg173, Leu239, Trp263 and Pro387) is unclear.

3.4.2.6.2. The CoA binding site

Among all BAHD members, only the crystal structures of malonyltransferase (MaT) and trichothecene 3-*O*-acetyltransferase (TRI101) have been solved in complex with a CoA moiety. In these complex structures, CoA binds on the front face of the protein in a tunnel, of which architecture is well conserved (Unno *et al.*, 2007; Garvey *et al.*, 2008). As *Cc*HCT was never crystallised bound to CoA, its structure was superposed onto those of MaT-malonyl-CoA and *Fs*TRI1101-T2 mycotoxin (Figure 80). This enabled the identification of the residues lining a putative CoA binding site in *Cc*HCT. The adenosine moiety binds at the protein-solvent interface, allowing the pantetheine extremity of CoA to extend through the solvent channel towards the active site.

Figure 80: Superimposition of HCT with the TRI101-CoA-T2 (A) and MaT-malonyl-CoA (B) crystal structures

Native HCT chain A is coloured in grey, TRI101-CoA-T2 (PDB: 2ZBA) in green and MaT-malonyl-CoA (PDB: 2E1T) in violet. CoA/ malonyl-CoA are shown as yellow sticks and T2 as magenta sticks.



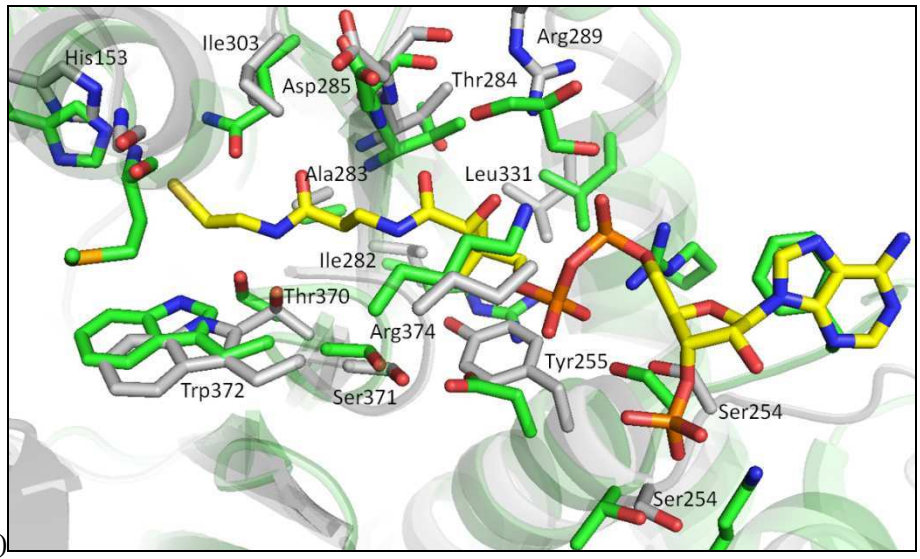
(A) (B)

The binding site for CoA is mostly associated with the C-terminal domain. In *Cc*HCT the putative CoA binding site is surrounded by the following residues: Lys240, Ser253, Ser254, Tyr255, Ala283, Thr284, Asp285, Arg289, Ile303, Leu331, Thr370, Ser371, Trp372 and Arg374. There is a relatively high conservation of the residues involved in CoA binding in MaT, TRI101 and *Cc*HCT (Figure 81A and B). However, the π-stacking interaction between the CoA adenosine moiety and a conserved aromatic residue (Phe258 in TRI101, Tyr272 in MaT) is lacking in *Cc*HCT (Figure 81C). From the sequence alignment (Figure 96), Tyr252 is the most likely residue to form such an interaction in *Cc*HCT. However, this would require a conformational rearrangement of the loop between η8 and α5 that contains Tyr252. Interestingly, Phe258 in TRI101 undergoes a similar conformational change from the apo (light green in Figure 81C) to the complex forms (green in Figure 81C). In contrast, this does not occur in MaT. The phosphate groups of CoA in *Cc*HCT probably interact via direct or water mediated H-bonding with polar residues (Ser254) or via salt bridge with charged residues (Arg289, Arg374, Lys240) and the dimethyl group with the side-chain of Tyr255.

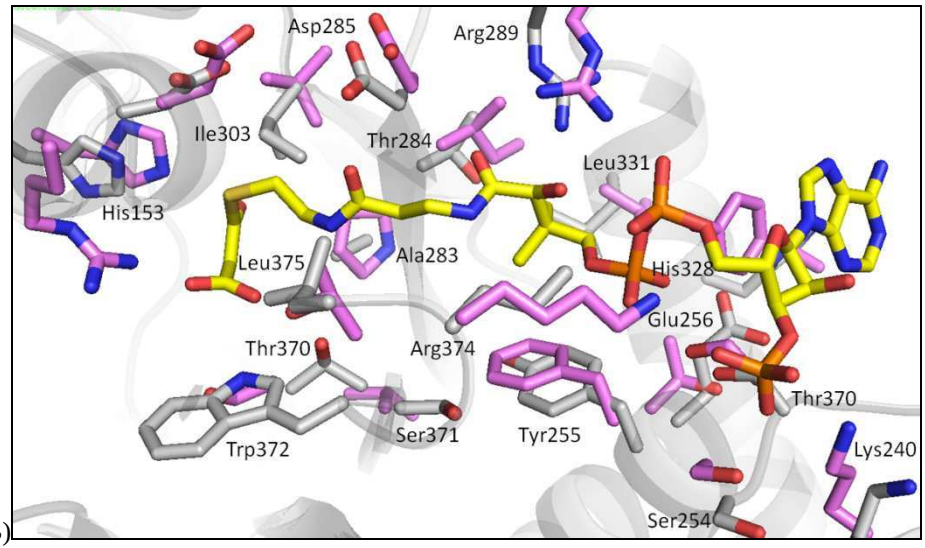
Figure 81: Superimposition of *Cc*HCT, *Fs*TRI101-CoA-T2 and MaT-malonyl-CoA crystal structures showing the CoA binding site region

(A) *Cc*HCT and *Fs*TRI101 (PDB: 2ZBA). (B) *Cc*HCT and MaT (PDB: 2E1T). (C) *Cc*HCT, MaT and TRI101 in the adenosine binding region. *Cc*HCT is coloured in grey, TRI101 in green, MaT in violet and CoA in yellow.

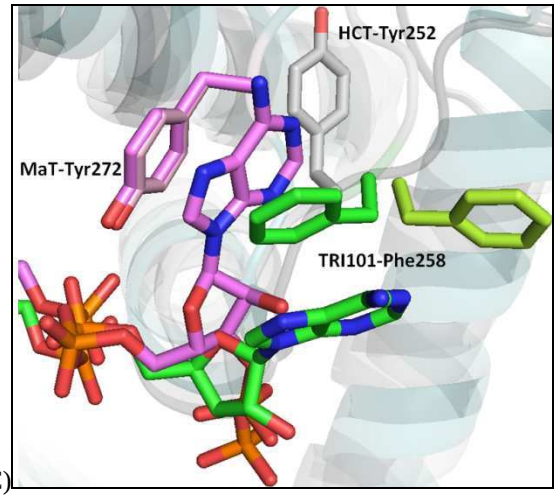
(A)



(B)



(C)

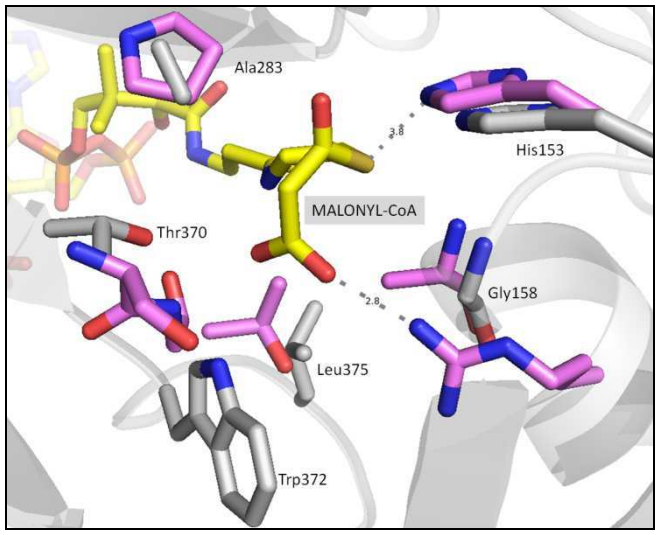


3.4.2.6.3. The acyl binding site

MaT was crystallised with malonyl-CoA bound (PDB: 2E1T). When the structures of *Cc*HCT and the MaT-malonyl-CoA complex are superposed (Figure 82), the aliphatic acyl moiety is pointing towards a pocket in *Cc*HCT which may accommodate hydroxycinnamoyl moieties in HCT/ HQT. The pocket is composed of: Thr36-Ser38, Val149, Met151, His153, Gly158, Gly161, Leu162, Trp372 and Met390. In MaT, a major interaction of the malonyl carboxylate involves Arg178 (Figure 82). This residue is not conserved in other BAHD members and is replaced by a Gly or an Ala in HCTs (Gly161 in *Cc*HCT) and a Ser in HQTs. The hydroxycinnamoyl moieties could potentially be sandwiched between the side-chain of Trp372, forming a π-stacking interaction, and Gly158 in *Cc*HCT (Figure 82). Trp372 is conserved in all BAHD members, except MaT, where it is replaced by a Thr. Gly158 is conserved in all HCTs and HQTs (HX_3DG motif), but not in the other BAHD subfamilies.

Figure 82: Superimposition of HCT and MaT-malonyl-CoA in the malonyl binding site region

MaT His170 and Arg178 interact with the malonyl-CoA. HCT native chain A is coloured in grey, MaT in violet and malonyl-CoA in yellow. For clarity, only the residues of HCT are mentioned.



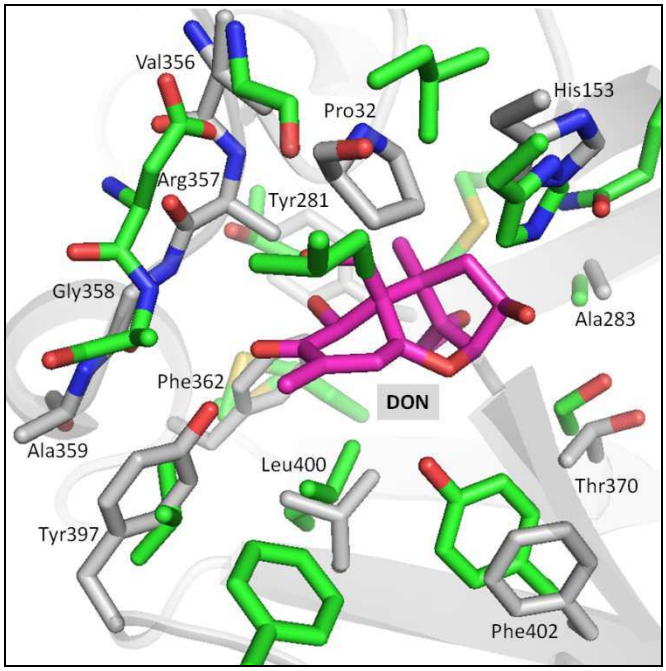
3.4.2.6.4. The acyl acceptor binding site

The only members of the BAHD superfamily that were crystallised in the presence of an acyl acceptor substrate are the *Fusarium sporotrichioides* and *graminearum* TRI101 acetyltransferases in complex with T2 mycotoxin and DON (Garvey *et al.*, 2008). The acceptor binding site is accessible from the back face of the protein, opposite of the CoA binding site. The residues lining a putative corresponding SBP for quinic/ shikimic acid in

*Cc*HCT are: Pro32, Phe34, His153, Tyr281, Ala283, Val356, Arg357, Gly358, Ala359,Phe362,Gly368, Thr370, Tyr397, Leu400 and Phe402. *Fg*TRI101-CoA-DON (PDB: 3B2S) is the only TRI101 crystal structure, where the loop containing Asp221-Ala222 at the entrance of the solvent channel is ordered. The Ala221 amide hydrogen interacts with the hydroxyl O5 of DON via H-bonding. In *Cc*HCT, the Val356-Gly358 loop occupies a similar position and it is likely that some of the residues in this loop participate in binding the acceptor substrate in *Cc*HCT (Figure 83).

Figure 83: Superimposition of the native HCT chain A structure with *Fg*TRI101-DON

The native HCT chain A is coloured in grey, *Fg*TRI101 in green and DON in magenta. For clarity, only the residues of HCT are mentioned.

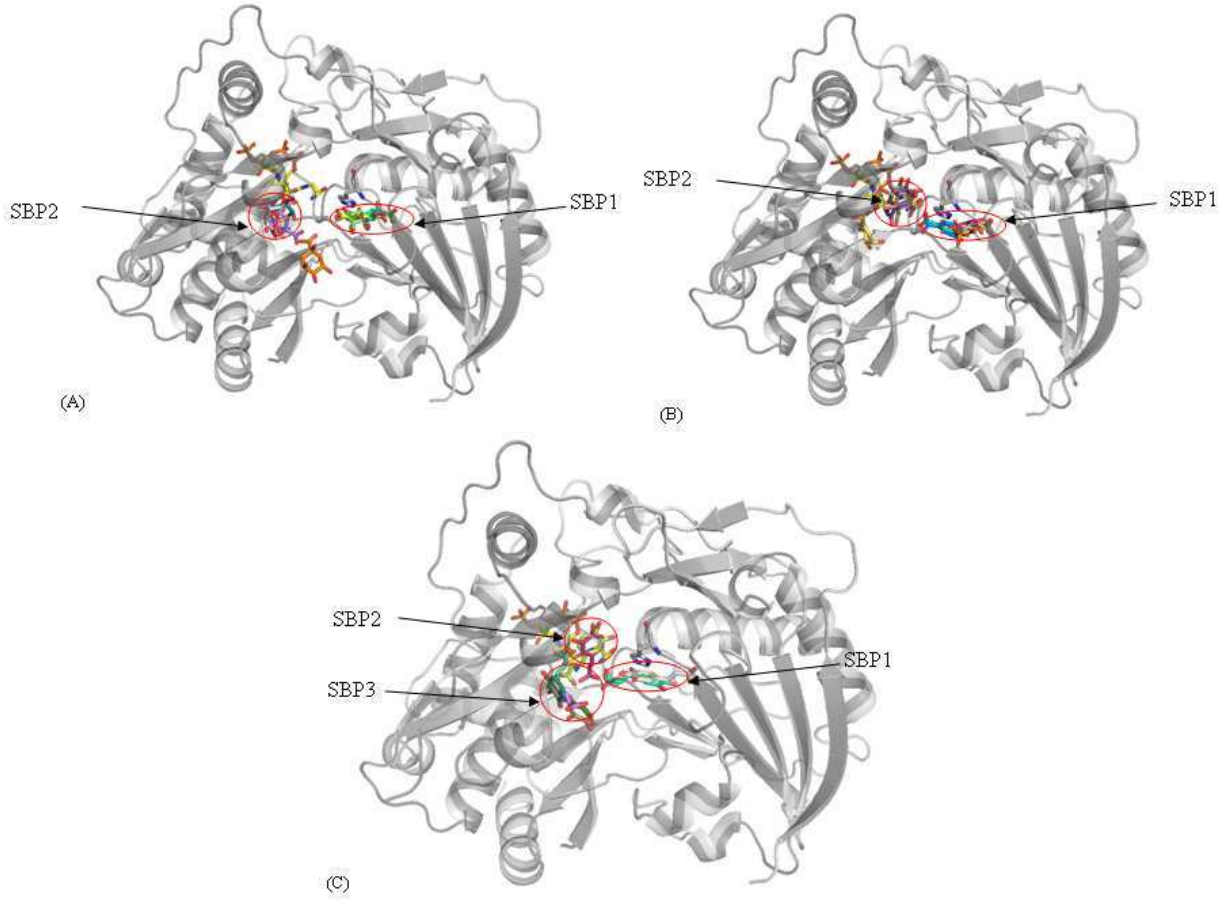


3.4.2.7. Docking calculations for CcHCT

Attempts to dock CoA into the apo-enzyme crystal structure were unsuccessful. The ligand never adopted an extended conformation; instead it randomly occupied the central part of the solvent channel. CoA was therefore initially positioned based on a superimposition of the native *Cc*HCT chain A and TRI101-CoA-T2 crystal structures. Quinic, shikimic and caffeic acids were subsequently docked in this *Cc*HCT-CoA receptor.

Figure 84: Quinic, shikimic and caffeic acid ligands docked in the *Cc*HCT-CoA receptor using AutoDock Vina

(A) Quinic acid. (B) Shikimic acid. (C) Caffeic acid.

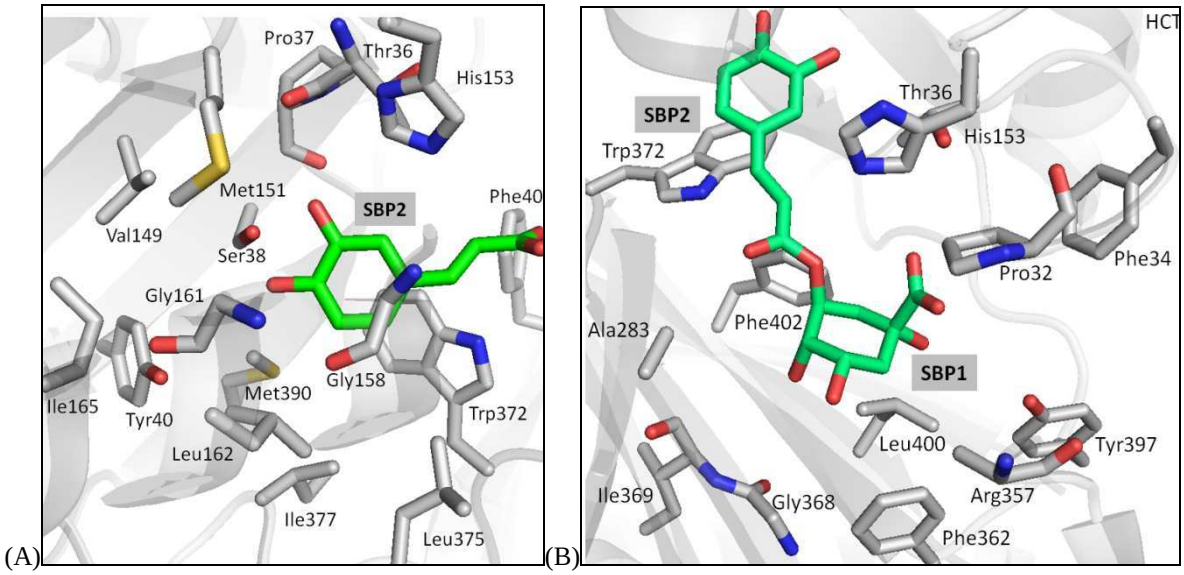


Two major clusters were found for quinic and shikimic acids, comprised of two pockets (Figure 84A and B). Meanwhile, caffeic acid occupied three major binding sites, the same two, plus another one (Figure 84C). As it was difficult to discriminate between the binding sites for the acceptor and acyl moieties, 5-CQA, 5-FQA and 5-CSA were also docked in the *Cc*HCT-CoA receptor. Figure 85A shows the best result for 5-CQA. The quinate moiety is bound preferentially in a pocket named SBP1 formed by the following residues: Pro32, Phe34, Thr36, Ile282, Ala283, Thr305, Arg357, Phe362, Gly368, Ile369, Thr370, Trp372, Tyr397, Leu400 and Phe402 (Figure 85B). The hydroxycinnamoyl moiety is preferentially bound in a pocket named SBP2 composed of the following residues: Thr36, Pro37, Ser38, Met151, His153, Gly158, Leu162, Trp372, Leu375 and Met390. As proposed from the comparison with MaT-malonyl-CoA, the phenolic ring is sandwiched between Gly158 and Trp372 and forms a π-stacking interaction with the aromatic residue (Figure 85A). In these docked ligands, the His153 side-chain is within H-bonding distance of the C-5 hydroxyl group. In addition, the Arg357 side-chain, which is not ordered in the *Cc*HCT

crystal structures, but strictly conserved among HCTs and HQTs, could be modelled to form a potential salt bridge with the carboxylate group of quinic and shikimic acids.

Figure 85: Best docking results for 5-CQA in the *Cc*HCT-CoA receptor by AutoDock Vina

Residues lining the hydroxycinnamoyl (A) and acceptor (B) moieties are shown as sticks.



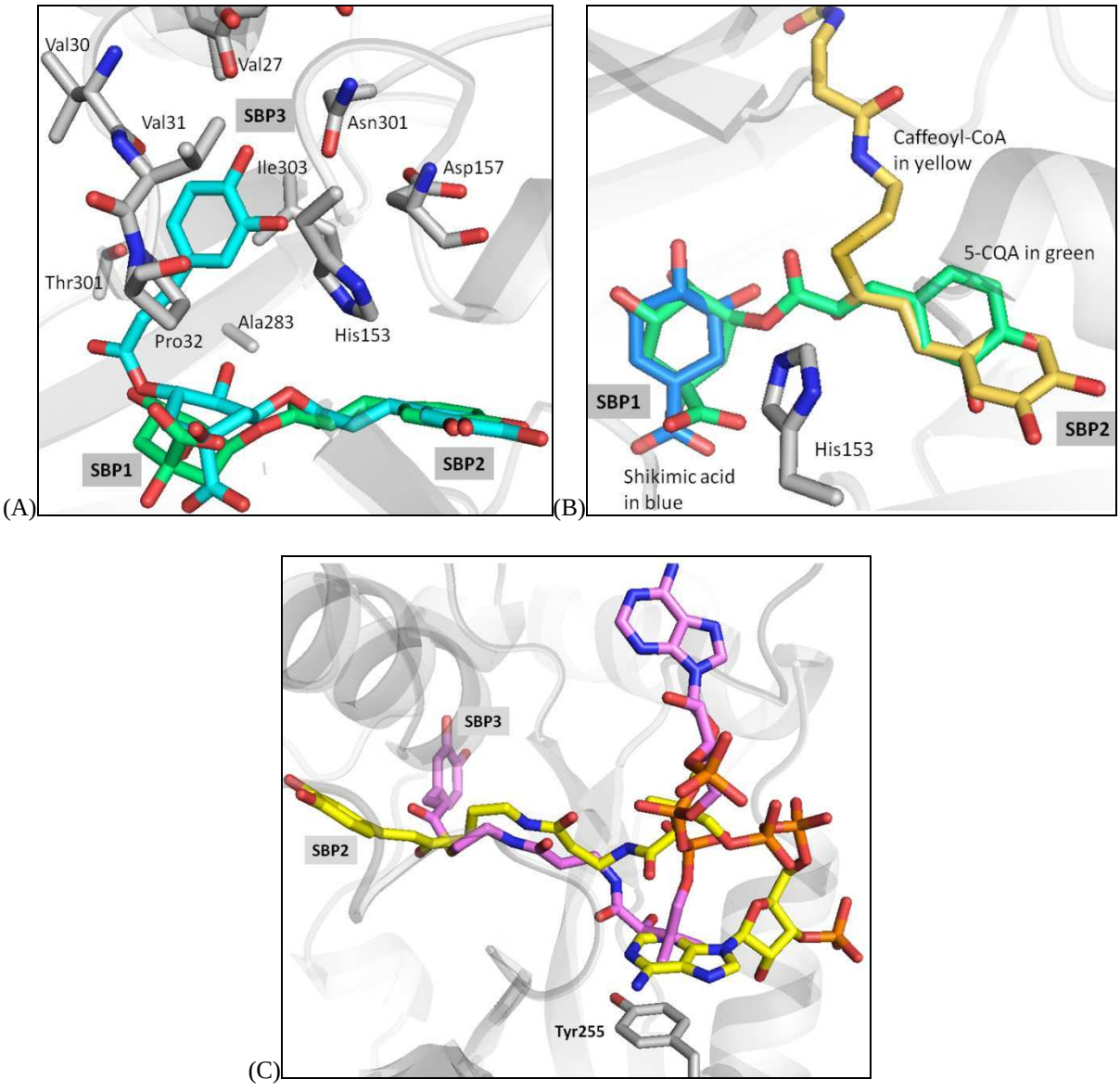
3,5-diCQA was also docked in the *Cc*HCT-CoA receptor. The best docking solutions were found with the hydroxycinnamoyl moiety at C-3 of quinic acid in SBP2, bringing the C-3 position of quinic acid close to His153 (Figure 86A). The hydroxycinnamoyl moiety at C-5 was bound in a pocket named SBP3 and composed of: Val27-Pro32, His153, Asp157, Asn301 and Ile303. The phenolic ring esterified at C-5 is sandwiched between Ile303 and Pro32. H-bonding of the hydroxyl group at C-4 is mediated with the carbonyl oxygen atoms of Val27 and Val30, and at C-5 with the Asn301 side-chain. It is therefore possible that these residues affect the substrate specificity for different cinnamic acid derivatives. The Val27-Pro32 and Asn301-Thr305 loops surround the binding pocket accommodating a second phenolic moiety. These two sequence regions may therefore influence the ability of hydroxycinnamoyl-transferases to form diesters.

To refine the CoA moiety position in *Cc*HCT, caffeoyl-CoA was docked in the HCT-shikimic acid receptor as presented in Figure 86B and C. The acyl moiety preferentially bound in two pockets, SBP2 or SBP3 (Figure 86C), as found previously for the acyl moieties of 3,5-diCQA docked in the HCT-CoA receptor. The pantetheine moiety of CoA adopted a similar conformation in most of the docking solutions similar to MaT-malonyl-CoA and

TRI101-CoA binding modes, as the dimethyl group stacked on the side-chain of Tyr255. The CoA adenosine moiety was observed in several conformations, mostly exposed to the solvent. In only one solution was the adenosine part stacked with Tyr255 side-chain. However, the molecule adopted an unusual turn (Figure 86C). Therefore, as proposed previously, the rearrangement of the loop containing the Tyr252 upon CoA binding is most likely to occur in *Cc*HCT.

Figure 86: Best docking results for 3,5-diCQA in the HCT-CoA receptor (A) and malonyl-CoA in the HCT-shikimic acid receptor (B, C) by AutoDock Vina

(A) 3,5-diCQA (cyan) superposed to the previous docking solution for 5-CQA (green). Residues lining SBP3 are shown as sticks. (B) Caffeoyl-CoA (yellow) docked in the HCT-shikimic acid (blue) receptor superposed to the previous docking solution for 5-CQA (green).



3.4.3. Coffee HQT homology-model

Despite exhaustive crystallisation screening, no crystals of native *Cc*HQT were obtained. Substitution of the lysine residues in the flexible cross-over loop to Ala, which was successful in obtaining reproducible *Cc*HCT crystals, was also carried out for *Cc*HQT. Unfortunately, the K210A/K219A mutant HQT did not crystallise. Therefore a homology-model was built using the crystal structure of native *Cc*HCT chain A as a template. *Cc*HQT-substrate models were then built using the ligands docked in *Cc*HCT (see above).

From sequence comparisons, the HX_3D motif is conserved among most HCTs (HHAAD) and most HQTs (HTLSD, except HNLSD in *Cc*HQT). In HCTs, HQTs and other BAHD acyltransferases, the residues of this motif, except for the catalytic His, are not likely to be involved in substrate binding as they do not directly line the solvent channel. From the docking carried out for *Cc*HCT, the putative binding pocket for quinic/ shikimic acid in *Cc*HCT (respectively in *Cc*HQT) involve: Pro32 (Ala32), Thr36 (Ile36), His153 (conserved), Tyr281 (conserved), Ala283 (conserved), Ile303 (Val299), Arg357 (conserved), Phe362 (conserved), Thr370 (Asn366), Leu400 (Thr396) and Phe402 (Tyr398). The putative binding site for hydroxycinnamoyl moieties in *Cc*HCT (respectively in *Cc*HQT) involves Thr36 (Ile36), Pro37 (Leu37), Ser38 (Thr38), Met151 (Val151), His153 (conserved), Gly158 (conserved), Leu162 (conserved), Trp372 (conserved), Leu375 (conserved) and Met390 (Ile386). A number of residues that surround the active site and differ between *Cc*HCT and *Cc*HQT are potentially involved in substrate discrimination. Some of these residues are specific to HCTs and HQTs and some are conserved (Figure 96).

Chapter 4. Discussion

4.1. Overview

Major advances have been made in understanding the phenylpropanoid pathway *via* the isolation and characterisation of the biosynthetic enzymes involved (Dixon *et al.*, 1996; Ferrer *et al.*, 2008). In this study, two enzyme families responsible for an essential part of the phenylpropanoid branch leading to CGAs and lignins have been investigated. The first is 4-coumarate CoA ligase (4CL), which synthesises hydroxycinnamoyl-CoA thioesters *via* an acyl-adenylate intermediate. These thioesters then serve as building blocks for CGAs and a number of other phenolics. *Cc*4CL2 was over-expressed and purified to homogeneity. An enzymatic assay however showed that the recombinant protein was not active, potentially because it appeared to form aggregates. As an alternative, *Nt*4CL2 was produced and successfully used to synthesise hydroxycinnamoyl-CoA thioester substrates that are not commercially available. Following CoA-activation, *Cc*HCT and *Cc*HQT catalyse the transfer of the hydroxycinnamoyl moieties to quinic or shikimic acid leading to CGAs. HPLC was used to characterise the reactions catalysed by these enzymes. 4CL2, HCT and HQT have been isolated from several plant species and extensively studied (Becker-Andre *et al.*, 1991; Voo *et al.*, 1995; Stuible *et al.*, 2001; Lindermayr *et al.*, 2002; Hoffmann *et al.*, 2003; Hamberger *et al.*, 2004; Comino *et al.*, 2007; Comino *et al.*, 2009). However, to our knowledge, their crystal structures have not been reported. A particular interest is to better understand the molecular basis of the reactions they catalyse and how small variations in their sequence can lead to profound alterations in substrate specificity. The crystal structures of apo-*Nt*4CL2 and *Nt*4CL2-CoA-AMP ternary complex were therefore solved and refined as were those of native and K210A/K217A mutant HCTs. The crystal structure of native *Cc*HCT was then used as a template to produce a homology model of *Cc*HQT.

4.2. *Nicotiana tabacum* 4CL2

*Nt*4CL2 was produced from a synthetic gene with optimised codons for expression in *E. coli*. The recombinant protein was purified in multiple chromatographic steps with an uncleavable C-terminal hexahistidine-tag. As shown in a previous study (Beuerle *et al.*, 2002), *Nt*4CL2 was able to convert cinnamic, coumaric, caffeic, ferulic and, very inefficiently, sinapic acids, into their corresponding CoA thioesters in the presence of Mg-

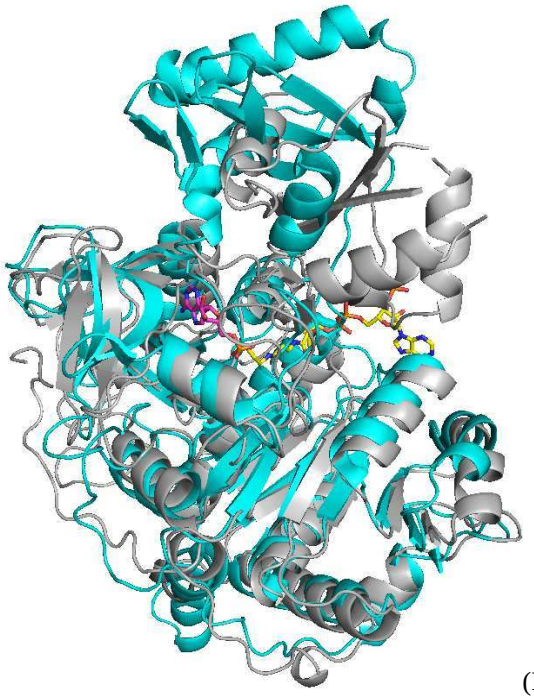
ATP. Substantial amounts of coumaroyl-CoA, caffeoyl-CoA and feruloyl-CoA were isolated using an optimised purification procedure including HPLC control analysis.

Despite rather low sequence identities (10-32 %) and versatile substrate specificities, *Nt*4CL2 shares a highly conserved tertiary structure with other adenylate-forming enzymes. *Nt*4CL2 crystallised in two conformational states that differ in the orientation of the C-terminal domain relative to the core N-terminal domain. The apo-enzyme adopts an "open" conformation (see Results section, Figure 69A) similar to that reported for firefly luciferase (PDB: 1LCI) (Conti *et al.*, 1996; Nakatsu *et al.*, 2006). The *Nt*4CL2-CoA-AMP ternary complex adopts a "closed" conformation, in which CoA and AMP bind at the domain interface (see Results section, Figure 69B). The closed configuration is similar to that adopted by human medium-chain acyl-CoA synthetase (ACSM2A) in complex with ibuprofen or AMP and butyryl-CoA (PDB: 2WD9, 3EQ6) (Figure 87B) and the structure seen for *Alcaligenes* 4-chlorobenzoate CoA ligase (CBL) in a ternary complex with AMP and 4-chlorophenacyl-CoA (PDB: 3CW9). This closed conformation is proposed to be representative of the enzyme-substrate configuration during the thioesterification step.

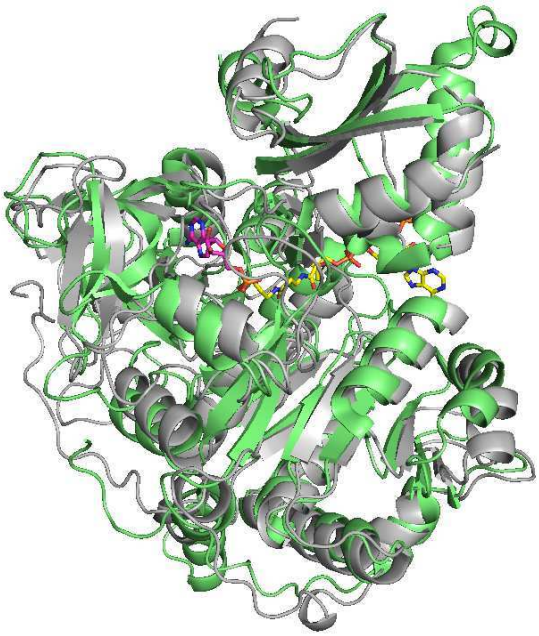
The crystal structure of the *Nt*4CL2-CoA-AMP ternary complex reveals several residues potentially involved in catalysis and substrate binding in this step (Figure 87C). These include Ser189, Thr336, Lys441, Gly444 and Gln446, which are all located at the domain interface. Lys441 and Gln446 are from the C-terminal domain and, in the open conformation, these residues are solvent-exposed and their side-chains partially disordered. Lys441 is conserved in all 4CLs but not in all adenylate-forming enzymes. In *Nt*4CL2, Lys441 could play a catalytic role by increasing the electrophilicity of the carboxyl-adenylate intermediate for nucleophilic attack of the thiol of CoA (Figure 88B). In the closed conformation, Gly444 becomes part of CoA binding pocket where it helps to correctly orient the CoA pantetheine tail and, thus, the CoA thiol group so that it appropriately positioned for thioesterification to occur (Figure 89C). Gly444 is strictly conserved in the adenylate-forming superfamily and the corresponding residue in ACSM2A is proposed to play the same role (Kochan *et al.*, 2009). In *Salmonella enterica* acetyl-CoA synthetase, mutation of this glycine to an alanine results in a protein that can catalyse the adenylation reaction but not the thioesterification step (Reger *et al.*, 2007). This supports the crucial role of this residue in CoA binding.

Figure 87: Comparison between *Nt*4CL2 and ACSM2A crystal structures

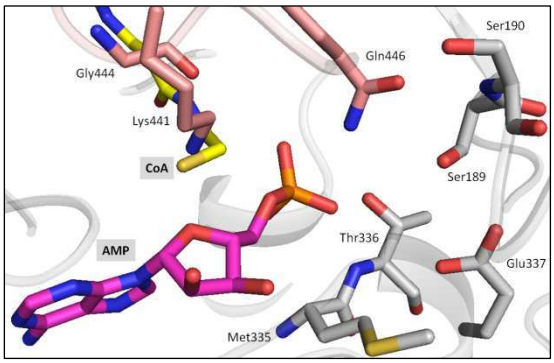
(A) Superimposition of the crystal structures of *Nt*4CL2-CoA-AMP (grey) with ACSM2A-ibuprofen (green) (PDB: 2WD9). (B) Superimposition of *Nt*4CL2-CoA-AMP (grey) with ACSM2A-ATP (cyan) (PDB: 3C5E). Comparison of (A) and (B) indicates that the structure of *Nt*4CL2-CoA-AMP observed is in the thioester-forming conformation. (C) Active site region of *Nt*4CL2-CoA-AMP, with ligands and potential residues involved in catalysis and substrate binding shown as sticks. (D) Modelling of the C-terminal domain of *Nt*4CL2 (orange) and ATP based on ACSM2A-ATP crystal structure (PDB: 3C5E) in the adenylate-forming conformation. (E) Active site region of *Nt*4CL2-ATP model.



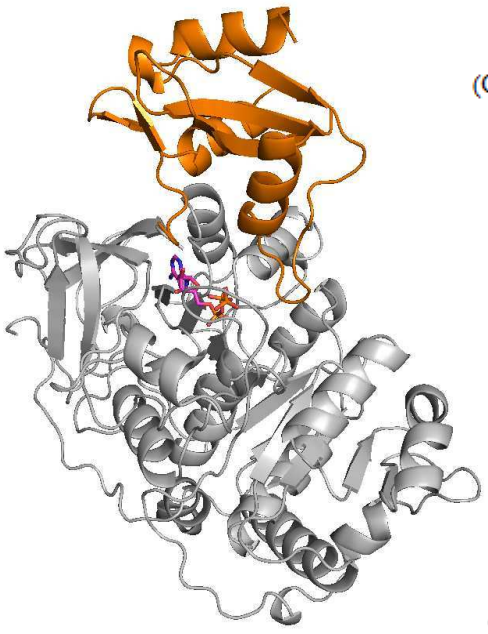
(A)



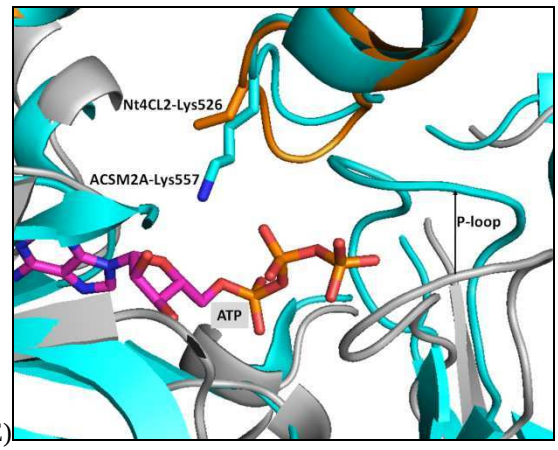
(B)



(C)



(D)

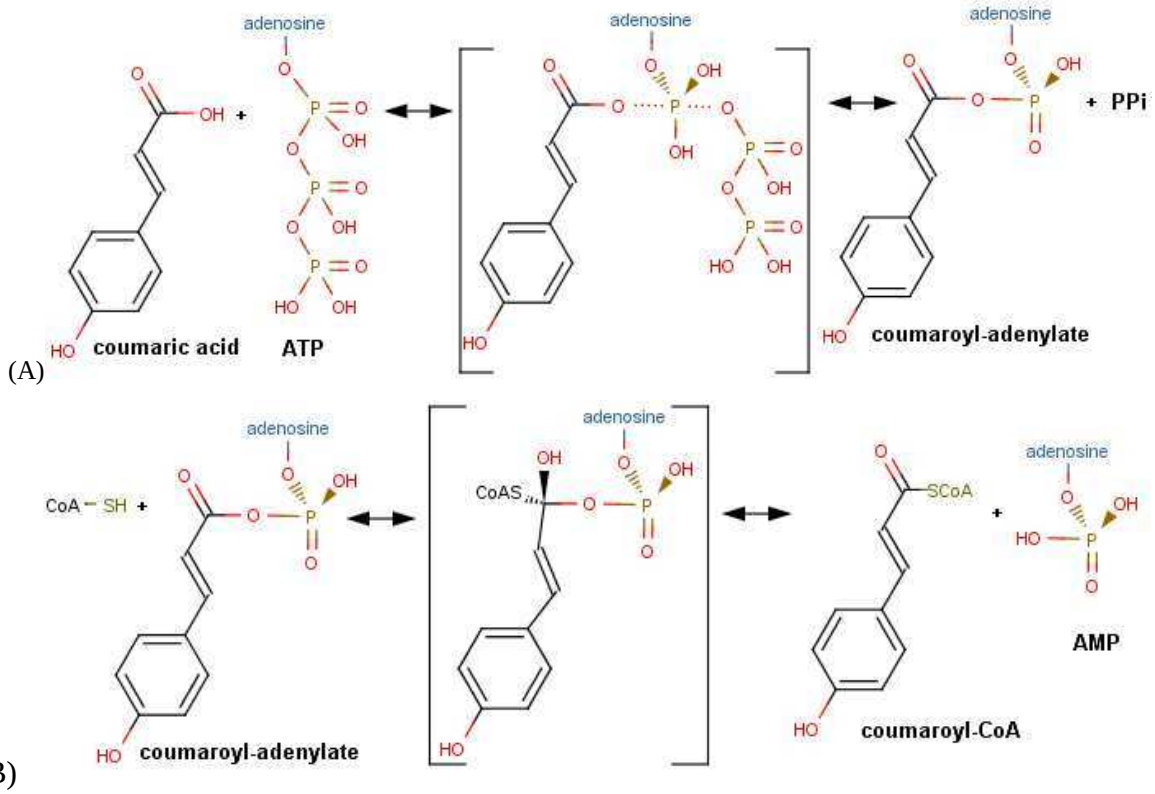


(E)

A second functional state, the adenylate-forming conformation, has been reported for ACSM2A in complex with ATP, AMP-CPP or CoA (PDB: 3C5E, 3DAY, 3GPC), *Brevibacillus brevis* PheA adenylation domain of the bacterial gramicidin S-synthetase in complex with AMP and Phe (PDB: 1AMU), and *Alcaligenes* CBL in complex with 4-chlorobenzoic acid or the adenylate intermediate (PDB: 1T5H, 1T5D, 3CW8). Thus far, CBL and ACSM2A are the only adenylate-forming enzymes to be characterised in the two functional conformations and in both cases the C-terminal domain of the protein undergoes a rotation of ~140° when compared to its position in the thioester-forming conformation. A model of the *Nt*4CL2-ATP complex in its proposed adenylate-forming conformation was thus constructed by superposing the C-terminal domain of *Nt*4CL2 with that of ACSM2A-ATP (PDB: 3C5E) and docking this, onto the structure of the N-terminal domain as seen in the *Nt*4CL2-CoA-AMP ternary complex crystal structure (Figure 87D).

Figure 88: Representation of the two-step CoA ligation catalysed by 4CL2

(A) Adenylation step. (B) Thioesterification step.



Mutational studies for several family members provide evidence that an invariant lysine is involved in catalysing the adenylation step by stabilising the transition state intermediate (Figure 88A). The enzymes catalysing this type of reaction usually employ electropositive residues (Arg, Lys, His) to orient the carboxylate and phosphoryl groups and balance the unfavourable negative charge interaction (Schmelz *et al.*, 2009). In ACSM2A, Lys557 forms a bifurcated salt bridge with the β-phosphoryl oxygen atoms of ATP for nucleophilic attack by the carboxylate substrate (Kochan et al., 2009). An additional role in binding the carboxylate substrate was proposed (Branchini *et al.*, 2000; Horswill *et al.*, 2002). The *Nt*4CL2-ATP model reveals that Lys526 is appropriately positioned to play such a catalytic role in the adenylation step (Figure 87E). In both the apo and thioester-forming conformations, the Lys526 side-chain in *Nt*4CL2 is solvent-exposed and partially disordered. This is also observed for the corresponding Lys557, which becomes ordered upon ATP binding in apo-ACSM2A. The conserved glycine-rich P-loop in ACSM2A, Thr221-Lys229, undergoes a conformational change and binds the β- and γ-phosphate of ATP (Figure 87E). The corresponding P-loop region in *Nt*4CL2, Ser189-Lys197, most likely undergoes a similar movement upon ATP binding (Figure 87E).

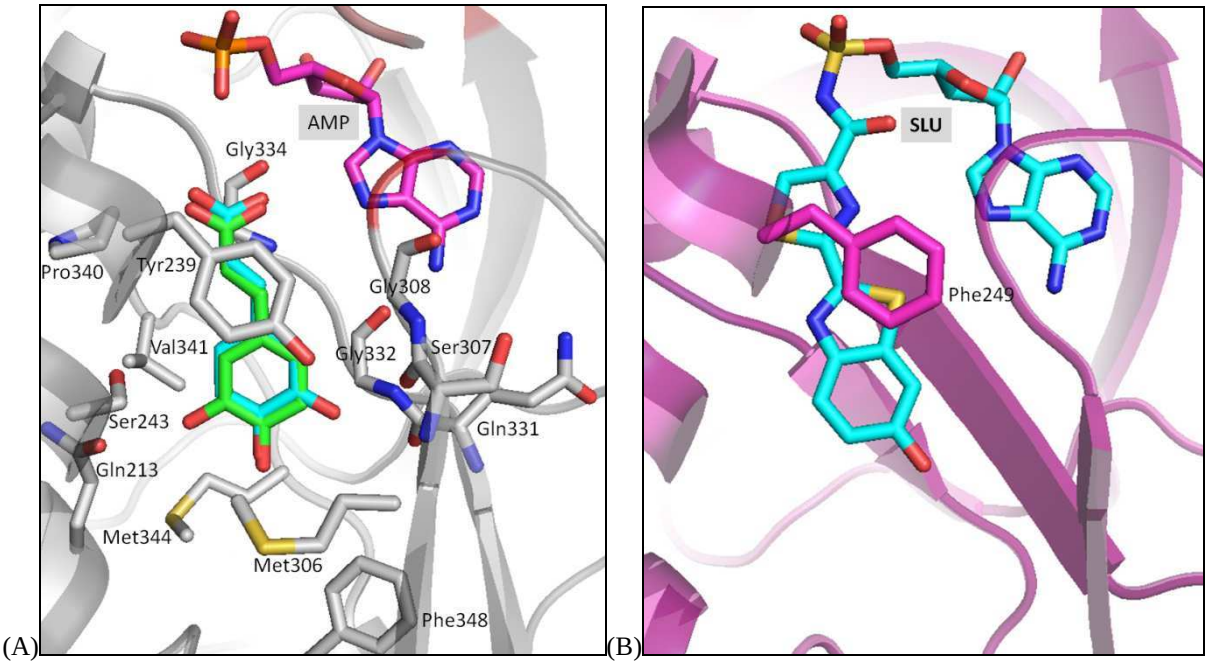
In ACSM2A, a salt bridge between Glu365 and Arg472 has been proposed to be important for precluding the ATP binding in the thioester-forming conformation. The Arg472/Glu365 salt bridge constricts the dimension of the nucleotide subsite, so that it can accommodate AMP but not ATP. Such an interaction is not occurring in *Nt*4CL2-CoA-AMP, as the corresponding residues Glu337 and Gln446 are too far apart (Figure 87C). However, the Gln446 side chain does form an extensive H-bonding network with AMP, Ser189 and Ser190, the latter also participates in H-bonding with Glu337. A similar mechanism for precluding ATP binding is therefore also likely to occur in *Nt*4CL2 (Figure 87C).

The structure of ACSM2A bound to Mg-ATP (PDB: 3C5E) showed that the magnesium ion is coordinated to the ATP β- and γ-phosphate groups and binds via a water molecule to a glutamate residue. This residue is strictly conserved among the adenylate-forming enzyme superfamily and the equivalent residue in *Nt*4CL2 (Glu337) is likely to be involved in the coordination of the Mg^{2+} ion. Another residue, His207 in CLB and Trp265 in ACSM2A, was observed to switch its position between the adenylate- and thioester-forming conformations. This residue was proposed to be involved in correctly orienting the carboxylate substrate (Kochan *et al.*, 2009; Wu *et al.*, 2009). In *Nt*4CL2, a similar conformational change of the corresponding His237 side-chain may also take place.

Despite many attempts, no crystal structure of *Nt*4CL2 in complex with coumaric acid bound could be obtained. This may be due to the limited solubility of this compound in aqueous solution. The Pocket-Finder prediction algorithm and a structure comparison with PheA bound to phenylalanine (PDB: 1AMU) enabled the location of the SBP in the *Nt*4CL2-CoA-AMP ternary complex structure. Cinnamic, coumaric, caffeic, ferulic and sinapic acids were docked into this SBP (Figure 89A), revealing a major π-stacking interaction between the phenolic moieties and the Tyr239 side-chain. This residue is conserved or replaced by Phe in other 4CLs, underlining the importance of such an aromatic residue in binding cinnamic acid substrates. Similar interactions are also seen for other adenylate-forming enzymes that bind aromatic moieties, such as Japanese firefly luciferase (PDB: 2D1R, 2D1S) (Figure 89B). However, this type of interaction is not universal as Ala and Val occupy a similar position in PheA and CBL respectively. It is therefore evident that such a stacking interaction is not a requirement for the binding of aromatic substrates by members of the adenylate-forming enzyme superfamily. Interestingly, Tyr239 undergoes a major rotamer transition when comparing the structures of apo-*Nt*4CL2 and ternary complex with CoA and AMP. This is most probably driven by nucleotide binding (see Results section, Figure 69E). Indeed, the position of the AMP phosphate group seen in the *Nt*4CL2-CoA-AMP structure would cause a steric conflict with the Tyr239 side-chain as seen in the apo structure. Tyr239 residue could also have an additional role in guiding the cinnamic acid derivatives into the SBP, as proposed for Trp265 in ACSM2A (Kochan *et al.*, 2009).

Figure 89: Representation of the substrate binding pocket for *Nt*4CL2 and luciferase

(A) Coumaric acid docked in *Nt*4CL2-CoA-AMP in π-stacking with Tyr239 side-chain. (B) Phe249 Japanese luciferase stacking with a high-energy intermediate analogue (SLU) (PDB: 2D1S).



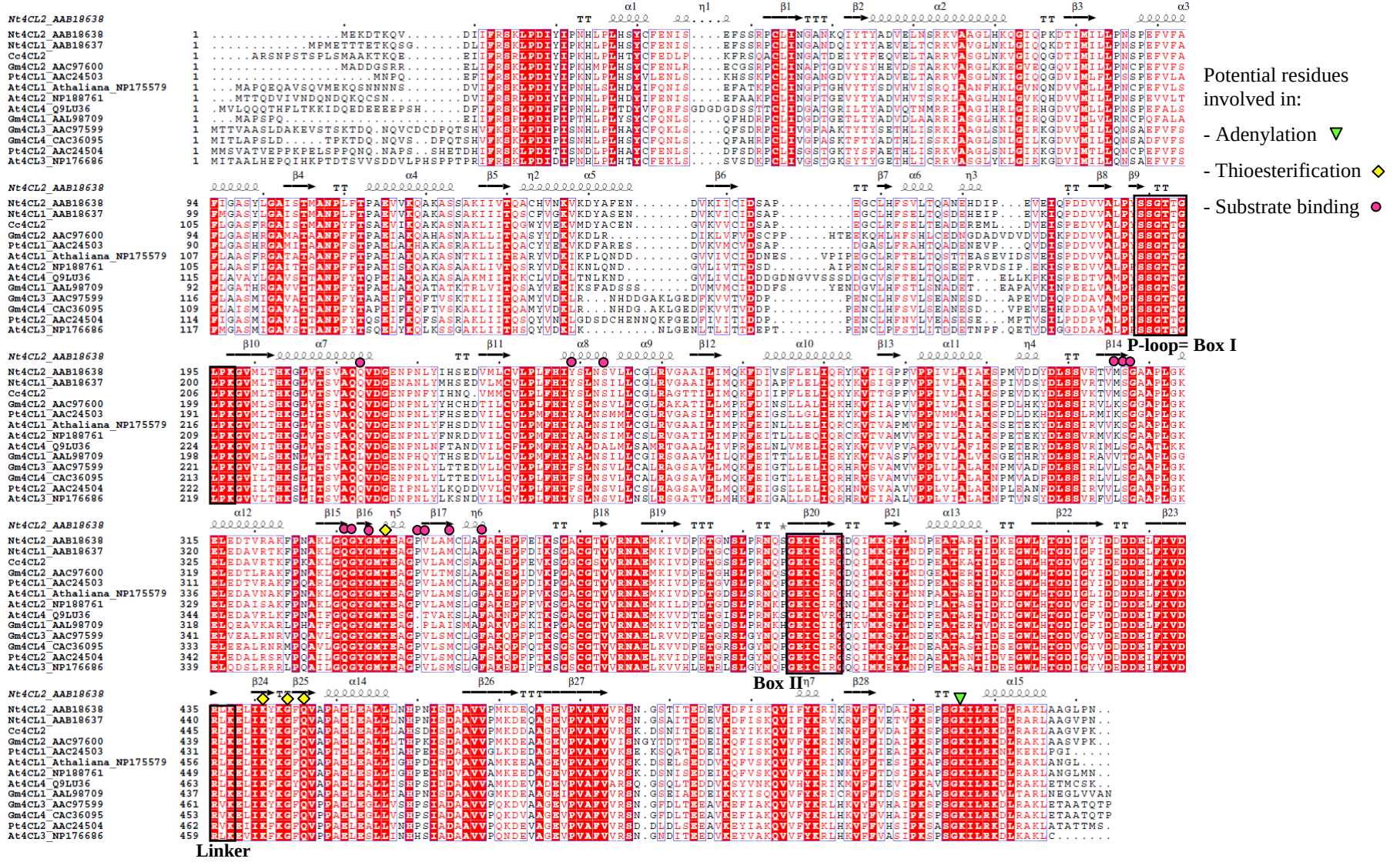
The ligand orientation was highly consistent in the case of cinnamic, coumaric and caffeic acids docked in *Nt*4CL2-CoA-AMP crystal structure. However, the hydroxyl group at position C-3 of caffeic acid could not be unambiguously positioned (Figure 89A). Two potential H-bonding interactions were observed with the Gln213 and Ser243 side-chains on one side, or the Ser307 and Gly332 main-chains on the other. Ser243 and Gly332 are strictly conserved among the 4CL family while Gln213 and Ser307 are generally conserved in most 4CL enzymes. The exceptions are *At*4CL4, where Ser243 and Gln213 are replaced by an alanine and a lysine respectively, and *Gm*4CL1, where Gln213 and Ser307 are replaced by Leu and Thr respectively. Curiously *At*4CL4 and *Gm*4CL1 are two of the rare 4CL enzymes that can efficiently activate sinapic acid (Lindermayr *et al.*, 2002; Hamberger *et al.*, 2004). A distortion of the phenolic plane orientation in the docked sinapic acid solutions in *Nt*4CL2-CoA-AMP structure is noticeable (see Results section, Figure 73E). The poor reactivity of *Nt*4CL2 towards sinapic acid was also reflected by the substantially lower binding affinities values when compared to the other hydroxycinnamic acids. Because sinapic acid carries two methoxyl groups, at the C-3 and C-5 position, size-exclusion is most likely responsible for this. As *Gm*4CL1 can activate sinapic acid, it is likely that mutation of Gln213 and Ser307 in *Nt*4CL2 to a Leu and Thr respectively, as for *Gm*4CL1, may alter 4CL2 substrate specificity.

A homology-model has been proposed for *At*4CL2 based on the PheA structure (Schneider, 2003). However, in this study, the authors failed to identify some of the key residues lining the putative SBP, in particular the conserved tyrosine that forms a π-stacking interaction with the phenyl moiety. In their model, Asn256 is proposed to play a similar role by mediating H-bonding interaction with a hydroxyl group at C-4. The present docking results indicate that such an H-bonding interaction between the hydroxyl at C-4 and the corresponding Asn242 in *Nt*4CL2 is unlikely. In fact, Asn242 is involved in an H-bonding network to Ser209 and Gln213 in *Nt*4CL2. Mutagenesis experiments targeting Met293 and Lys320 residues in *At*4CL2, corresponding to Pro279 and Met306 in *Nt*4CL2, showed that these amino acid residues are involved discriminating between ferulic and sinapic acids, most probably by steric hindrance. Met306 is part of the proposed SBP in *Nt*4CL2 and Pro279 forms a hydrophobic stacking interaction with this methionine. This suggests that these residues could potentially influence the accommodation of C-3/C-5 substituted substrates like ferulic and sinapic acids. In soybean (*Glycine max*), the deletion of a valine residue from *Gm*4CL2 and *Gm*4CL3 (Val341 in *Nt*4CL2), which is absent in *Gm*4CL1, allowed the mutant enzymes to convert sinapic acid (Lindermayr *et al.,* 2002; Lindermayr *et al.,* 2003). Val341 is indeed proximal to the C-3/C-5 position of sinapic acid docked in *Nt*4CL2. Therefore, deletion of this residue would also provide sufficient space for sinapic acid to bind (Figure 89A).

The *Nt*4CL2-CoA-AMP ternary complex crystal structure and comparison with other crystal structures (see above) enabled the identification of several putative residues involved in catalysis and substrate binding. These include: Tyr239 for substrate binding, Thr336, Lys441, Gly444 and Gln446 in the thioester-forming step, and Lys526 in the adenylate-forming step. The docking results and previous mutational studies on 4CL enzymes support the involvement of residues lining the SBP in enzyme specificity towards differently substituted cinnamic acids. These residues and the conserved motifs are highlighted in the sequence alignment presented in Figure 90. The results provided here from the structural basis for a better understanding of the *Nt*4CL2 substrate specificity.

Figure 90: Sequence and structure alignment between *Nt*4CL2 and other adenylate-forming enzymes

Conserved regions characteristic to adenylate-forming enzymes are highlighted.



4.3. *C. canephora* HCT and HQT

In the phenylpropanoid pathway, hydroxylation at C-4 of the phenylalanine aromatic ring of *t*-cinnamic acid catalysed by C4H produces 4-coumaric acid. Coumaric acid is then activated by CoA ligation (4CL), and the resulting CoA thioester is converted by HCT and HQT to coumaroylquinic or -shikimic acid. Subsequent hydroxylation at C-5 of the aromatic ring by C3H results in the formation of the most abundant CGA molecule, 5-CQA, and the less common caffeoylshikimic acid (CSA). Further hydroxylation and methoxylation leading to ferulic, 5-hydroxyferulic or sinapic conjugates comes further along the pathway (see Introduction, Figure 9). As major intermediates in the biosynthesis of guaiacyl and syringyl lignins, CGAs are widespread but usually found in relatively small amounts in higher plants. The exceptionally high levels of diverse CGA compounds found in the green *C. canephora* coffee grain is still not well understood. The importance of acyltransferases in CQA biosynthesis has prompted us to focus our analysis on the biochemical and structural properties of these enzymes.

The over-expression and purification of the recombinant *Cc*HCT was optimised using a standard protocol with an average yield of 4.5 mg/L culture. *Cc*HQT could not be over-expressed using a conventional protein tag (GST, His_6). Therefore, an N-terminal His_6-SUMO3 fusion was eventually tried and resulted in a yield of 1 mg/ L culture. During storage, purified *Cc*HCT and *Cc*HQT were partially degraded into two polypeptide fragments of ~24 kDa, corresponding to the approximate size of the N- and C-terminal domains of these proteins. This degradation could be reproduced in limited-proteolysis experiments. A similar proteolytic pattern has been observed for *D. caryophyllus* HCBT and *C. roseus* DAT (Yang *et al.*, 1997; St-Pierre *et al.*, 1998). This may be a common feature of BAHD superfamily members, which are characterised by a large inter-domain cross-over loop (Ma *et al.*, 2005; Unno *et al.*, 2007; Garvey *et al.*, 2008). From the native *Cc*HCT crystal structure, the domain interface buries 3420 $Å^3$ and is strengthened by 42 H-bonds and 5 salt bridges. This may be a reason why the polypeptide fragments of *Cc*HCT remain physically attached through a size-exclusion chromatography step. An intermolecular disulfide bond involving Cys142 was probably the key to the crystallisation of native *Cc*HCT as it changes the global surface properties of the protein. The resulting dimer is, however, not physiologically relevant as *Cc*HCT and all other BAHD acyltransferases characterised to date exist exclusively in a monomeric form (45-55 kDa). In order to facilitate the formation of such a covalent bond and thus crystal formation, oxidising and cross-linking agents were added to the protein solution

before setting up crystallisation drops. Unfortunately, the native *Cc*HCT crystals could never be reproduced with subsequent protein batches.

In an effort to optimise the reproducibility and the diffraction quality of the initial crystals, a trypsin-proteolysis resistant mutant of *Cc*HCT was produced and purified using a similar protocol as the native protein. Crystals of this K210A/K217A (K-mutant) *Cc*HCT diffracted to 1.7 Å resolution. Substitution of the lysine residues, which are characterised by a high conformational mobility, to alanine seems to enhance the crystallisability of *Cc*HCT by a reduction of the surface entropy in the cross-over loop region. Side-chain entropy is indeed negatively correlated with the involvement of some charged amino acids, such as Arg and Lys, in crystal contacts (Derewenda *et al.*, 2006; Cieslik *et al.*, 2009). However, part of the cross-over loop (residues 206-224) was not ordered in any of the HCT crystal structures.

The crystallisation conditions were very different for the native and K-mutant *Cc*HCTs, which crystallised at pH 6.0-6.5 and 8.5-9.0 respectively. The crystal structures differ mainly in the Val31-Pro37 loop and His153 side-chain conformation in the active site region. The presence of a sulfate ion bound on the protein surface near the Ala120-Tyr130 loop in the native *Cc*HCT chain A crystal structure induced a conformational change of this loop and the nearby Val31-Pro37 loop compared to chain B. The His153 side-chain in native chain A points towards the solvent channel, whereas in chain B, it forms a π-stacking interaction with the imidazole ring of His35. The structure of the K-mutant crystal form 2 is similar to that of the native chain B. In both chains A and B of crystal form 1 of the K-mutant His153 is flipped to the opposite of the solvent channel, this time forming a π-stacking interaction with His154, and the solvent channel is blocked by the Val31-Pro37 loop. A glycerol molecule interacting with the residues Thr36, Pro37, Ser38, Gly158, Trp372 and Leu375 was observed in the K-mutant form 1. The presence of this ligand bound in the active site may have influenced the Val31-Pro37 loop and His153 conformations. The differences observed in the various crystal forms suggest that the active site architecture of *Cc*HCT is flexible and may be modulated by the binding of various molecules.

In protein structures, histidine-histidine contact pairs are not uncommon and are notably involved in stabilising secondary structures (Meurisse *et al.*, 2003). The protonation state of histidine (pK_a~6) depends on the pH and local environment of the side-chain. At physiological conditions, the His-His pair has both or one of the imidazole groups deprotonated. The pair is therefore either neutral or singly positively charged and the stacking of the aromatic rings is not weakened by Coulomb repulsion (Heyda *et al.*, 2010). The proximal His153-His154 pair in *Cc*HCT corresponds to a more restricted side-chain

conformation than the distal His153-His35 pair. In both cases, π-stacking sterically constrains the free rotation of the His imidazole ring. The His153 side-chain of the native *Cc*HCT chain A is less constrained and more readily accessible. The native HCT chain A was therefore considered as the most likely functional conformation. From sequence and structure comparisons, His153 from the conserved HX_3D motif corresponds to the catalytic residue in *Cc*HCT. As expected, H153A mutant *Cc*HCT was inactive in the biochemical assays. This basic residue can extract a proton either from the acyl acceptor molecule (the hydroxyl group at C-5 of quinic acid in HCT/ HQT) or from the thiol group of CoA, depending on the direction of the reaction. The activated hydroxyl or thiol can then attack the carbonyl carbon of the acyl-CoA or CGA molecules and trans-esterification can occur.

This reaction was studied in the forward direction towards CGA biosynthesis. For this purpose, caffeoyl-, coumaroyl- and feruloyl-CoA thioesters were synthesised using *Nt*4CL2 and purified. The enzymatic activities of *Cc*HCT and *Cc*HQT were assayed with these CoA thioesters in combination with quinic and/or shikimic acid. As demonstrated in other studies on HCTs and HQTs from tomato and tobacco plants, these enzymes are able to synthesise the major CGA compound, 5-CQA, from caffeoyl-CoA and quinic acid (Hoffmann *et al*., 2003; Niggeweg *et al*., 2004). A coumaroylquinic acid and 5-FQA were also produced using *Cc*HCT and *Cc*HQT in the presence of quinic acid and the corresponding CoA thioester. Alternatively, a coumaroylshikimic acid, a caffeoylshikimic acid and a feruloylshikimic acid were synthesised in the presence of shikimic acid. One major isomer was observed in all cases, with the exception that *Cc*HCT forms three putative shikimate isomers with coumaroyl-CoA and caffeoyl-CoA. Competition assays, where both acceptor molecules are supplied, showed that *Cc*HQT transfers the acyl moiety more efficiently to quinic than to shikimic acid. Therefore, it was confirmed that the protein names have been correctly assigned (Lepelley *et al.*, 2007). Other potential acceptor molecules were tested but no activity was detected, confirming that *Cc*HCT and *Cc*HCT are specific to the alicyclic quinic and shikimic acids, as previously described for other HCT and HQT enzymes (Hoffmann *et al.*, 2003; Niggeweg *et al.*, 2004; Comino *et al.*, 2007; Comino *et al.*, 2009; Sonnante *et al.*, 2010). Esters of caffeic and coumaric acid with tryptophan are found in Robusta coffee extracts, so it would be interesting to test this potential acyl acceptor in future experiments. An additional activity was reported recently for artichoke HCT towards exogenous 3-anthranilic acid, forming a new amide conjugate with interesting biological properties, *N*-coumaroyl-3-hydroxyanthranilic acid (Moglia *et al.*, 2010). It would also be interesting to test this compound with *Cc*HCT and *Cc*HQT as well.

In the reverse direction, *Cc*HQT efficiently converts 5-CQA and 5-FQA in the presence of CoA to their corresponding CoA thioesters. HCT was less efficient towards 5-FQA. No activity towards the 3- and 4-CQA isomers was observed in the presence of *Cc*HCT or *Cc*HQT. All three isomers have been found in various plant families, despite the 5-acyl isomer being predominant in the CQA and FQA groups. During maturation and processing of the coffee beans, the 5-monoesters (CQA and FQA) clearly decrease when the 3- and 4-isomers increase (Lepelley *et al.*, 2007). It was noticed that basic pH or heating cause CGA isomerisation in solution, transforming 5-CQA into the 3- and 4-CQAs most probably by acyl migration. These and other studies (Griffiths *et al.*, 1997; (Dawidowicz *et al.*, 2010) suggest that chemical isomerisation is the major process leading to their formation *in vivo* or during processing.

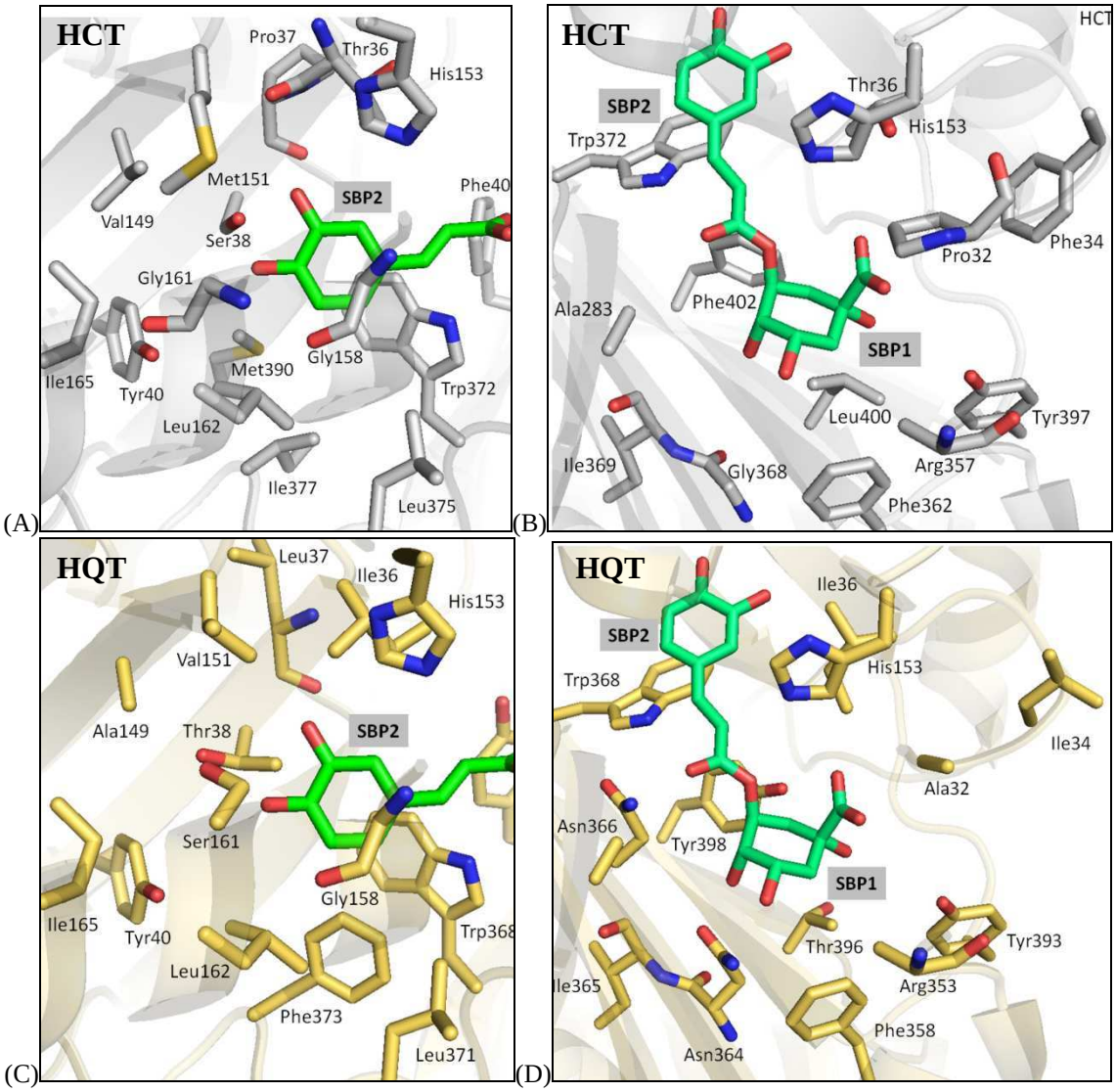
Unfortunately, no crystal structure of *Cc*HCT in complex with substrates could be obtained. As attempts to dock CoA into the crystal structure of *Cc*HCT were unsuccessful, the position of the CoA moiety was modelled based on the crystal structure of TRI101-CoA-T2. The cofactor extends from the front face of *Cc*HCT to the active site via a solvent channel. Residues involved in CoA binding seem well conserved between *Cc*HCT, MaT and TRI101. However, in the crystal structure of CcHCT no interaction stabilising the binding of the CoA adenosine moiety is seen. From a sequence alignment, Tyr252, which is conserved or replaced by a phenylalanine in all BAHD enzymes, is the residue most likely involved in binding the CoA adenosine in. However, a conformational rearrangement of the loop between η8 and α5 would be required for this residue to form a π-stacking interaction with the CoA adenosine moiety. A similar rearrangement is observed in TRI101, in which Phe258 changes its conformation upon CoA binding. This requirement for a rearrangement, in *Cc*HCT of the η8α5 loop upon CoA binding may explain why no specific binding site for the CoA adenosine could be seen in the docking solutions. Docking of caffeoyl-CoA revealed a major interaction between the CoA dimethyl group and Tyr255, similar to that seen in MaT. The NTISY (248-252) loop in *Cc*HCT is shorter in *Cc*HQT, as in potato, tobacco and tomato, but not artichoke HQTs. As the region corresponds to a loop located on the protein surface, it is most likely that it does not interfere with the architecture of the active site at the domain interface. However, this may influence the CoA binding mode between these enzymes. Mutation of Tyr252 in *Cc*HCT to an Ala resulted in a significantly reduced activity with formation of caffeic acid, which may be due to a destabilisation of CoA binding. Arg374, which is partially disordered in the HCT crystal structures, most probably interacts with CoA diphosphate group. It is conserved or replaced by a lysine in all BAHD enzymes. The importance of this residue is

shown from enzymatic assay, as the R374E mutant is inactive in the reverse reaction involving 5-CQA. In the forward reaction involving caffeoyl-CoA, it exclusively forms caffeic acid. The D157A mutant formed some caffeic acid from caffeoyl-CoA. In *Cc*HCT and other BAHD crystals structures, D157A is involved in salt-bridge interaction with Arg287. Conservation of this aspartate and arginine or lysine residues is most likely important for maintaining the protein fold. This may indicate that the enzyme architecture or stability had been affected by this amino acid substitution.

Docking of the most potent CGA molecules was carried out to identify the potential binding subsites in *Cc*HCT and the structural determinants for its substrate specificity. Docking analyses resulted in the identification of a preferential binding site for the quinic/ shikimic moiety (SBP1). The main difference between quinic and shikimic acid is the presence of a hydroxyl group at C-1 in quinic acid. In shikimic acid a double bond is present between C-1 and C-2, resulting in a different geometry of the polyol ring. Leu400 is in the vicinity of the C-1 position of the acceptor molecule in *Cc*HCT (Figure 91B). This residue is conserved among all HCTs (Figure 96). Its hydrophobic nature may therefore favour the binding of shikimic acid by van der Waals interaction in the region of the double bond. In all HQTs, it is replaced by a threonine, which may form an H-bond with the hydroxyl at C-1 of quinic acid and therefore favour the binding this acceptor molecule. In the TRI101 ternary complex crystal structures, the corresponding residue is also a leucine, which directly interacts with the T2 mycotoxin/ DON backbone ring. These observations support the prediction that this residue may be important in discriminating between alternative acceptor molecules.

Figure 91: Representation of the 5-CQA substrate binding subsites in the native *Cc*HCT crystal structure (A, B) and *Cc*HQT model (C, D)

Docked 5-CQA with residues interacting with the hydroxycinnamoyl moiety in SBP2 (A, C) and quinate moiety in SBP1 (B, D) shown as sticks. *Cc*HCT crystal structure is coloured in grey and *Cc*HQT model in gold.



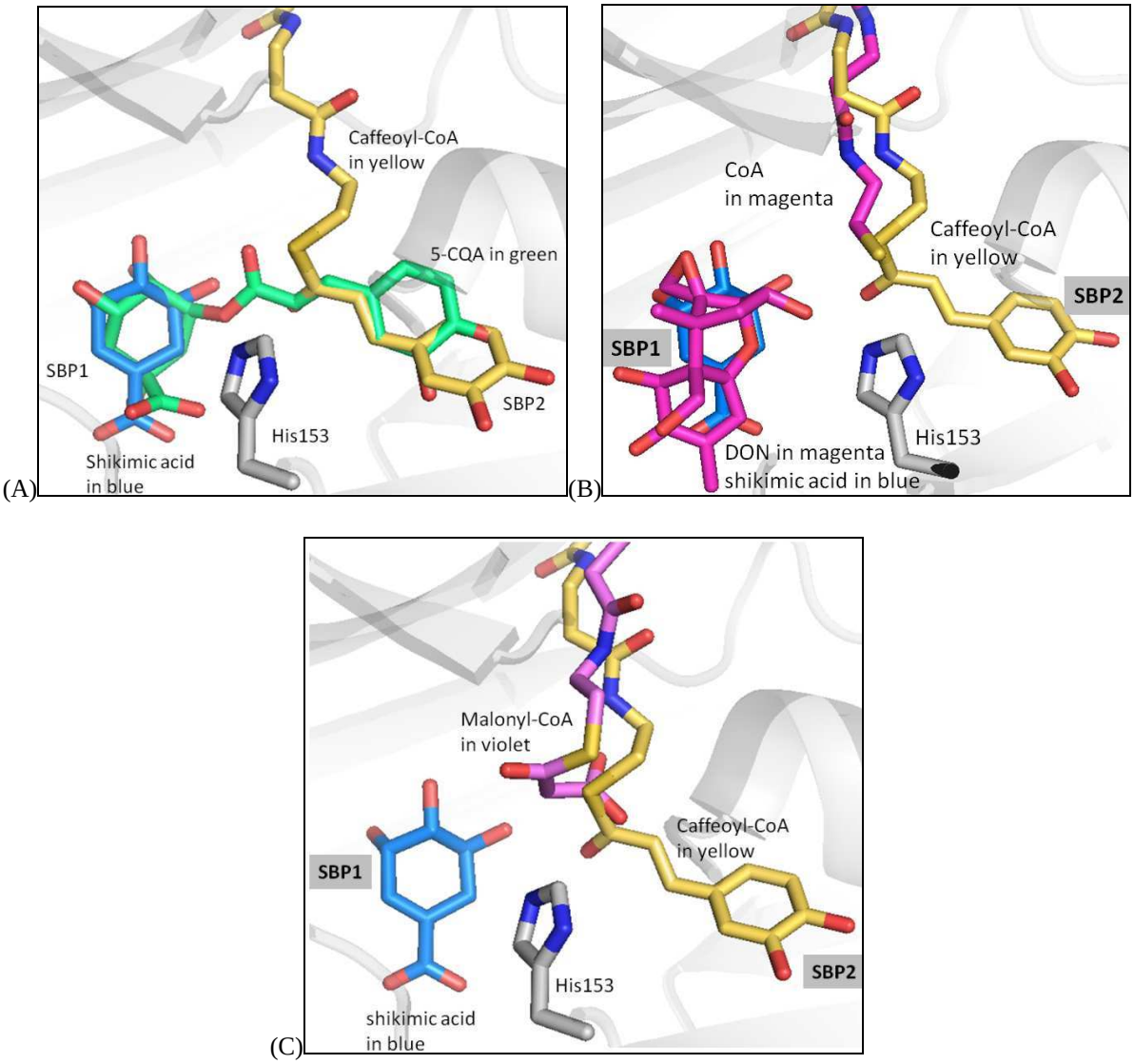
The structural determinants for the specificity found in the biochemical data concerning the phenolic preference are not obvious from the docking and structure results. *Cc*HCT transfers much less efficiently feruloyl moieties than caffeoyl and coumaroyl moieties. This substrate specificity was also reported for tobacco HCT (Hoffmann *et al.*, 2003). In contrast, *Cc*HQT accepted equivalently all three substituted hydroxycinnamoyl moieties. The docking results suggest that the Ser38 side-chain could form an H-bond with a hydroxyl in position C-4 of the aromatic ring (Figure 91A). This serine is replaced by an equivalent threonine in *Cc*HQT (Figure 91C). The Thr36 main-chain carbonyl oxygen may also form an

H-bond with OH in position C-5 of the aromatic ring in caffeoyl moieties (Figure 91A). Thr36 is replaced by an isoleucine in *Cc*HQT (Figure 91C). Met151 is also close to the C-5 hydroxyl of the docked caffeoyl moieties. Met151 is conserved in all HCTs, whereas it is replaced by a valine in most HQTs. Such a mutation may leave enough space for substitution on the C-5 by a methoxyl group and increase the binding of feruloyl moieties. This may explain why *Cc*HQT efficiently converts ferulate substrates. In artichoke HQT and HQT2, this residue is replaced by a leucine. However these enzymes have not been assayed with a ferulate substrate (Comino *et al.*, 2009; Sonnante *et al.*, 2010). The substitution of Pro37 in *Cc*HCT to a leucine in *Cc*HQT may induce a different loop conformation and increase the size of the binding pocket of *Cc*HCT to accommodate methoxylated cinnamic acids. These residues may therefore affect the substrate specificity towards the methoxylated cinnamic acids like ferulic acid. They may be structural determinants for the ability for *Cc*HCT to discriminate between hydroxyl and methoxyl substitution at C-5 of the aromatic ring. These assumptions should be verified by mutagenesis.

A structure model of artichoke HCT and HQTs was published recently (Sonnante *et al.*, 2010). From docking analysis of 5-coumaroylquinic acid and 5-CQA, the authors identified two distinct binding pockets for the caffeoyl and coumaroyl moieties. They also proposed a putative binding site for quinic/ shikimic acid that is formed by some residues corresponding to SBP2 in *Cc*HCT. However, our docking results are supported by the structure comparison with TRI101 and MaT, where the quinic/ shikimic acid moiety binds in a pocket corresponding to SBP1 (Figure 92B and C). The malonyl moiety in the MaT-malonyl-CoA complex points towards a pocket corresponding to our SBP2. A putative major π-stacking interaction was found between the phenolic moiety and the side-chain of the conserved Trp372 residue. Leu400 in the native *Cc*HCT crystal structure, corresponding to Thr396 in the *Cc*HQT model, was proposed to be an important structural determinant for the substrate preference for quinic or shikimic acid. In addition, the loop at the entrance of the solvent channel towards the quinic/ shikimic acid binding site contains an arginine residue (Arg357), conserved among all HCTs and HQTs, suggesting it may play an important role by interacting with the carboxylate group of quinic/ shikimic acid and therefore orienting the acceptor for acylation at C-5. Equivalent interactions were found in the TRI101-CoA-DON complex structure.

Figure 92: Active site of *Cc*HCT with docked ligands superposed to TRI101-CoA-DON and MaT-malonyl-CoA

(A) Docked 5-CQA (green) and caffeoyl-CoA (yellow) in *Cc*HCT-shikimic acid (blue) receptor. (B) Caffeoyl-CoA and shikimic acid solutions in *Cc*HCT superposed to DON and CoA (magenta) from TRI101. (C) Caffeoyl-CoA and shikimic acid solutions in *Cc*HCT superposed to malonyl-CoA (violet) from MaT-malonyl-CoA.

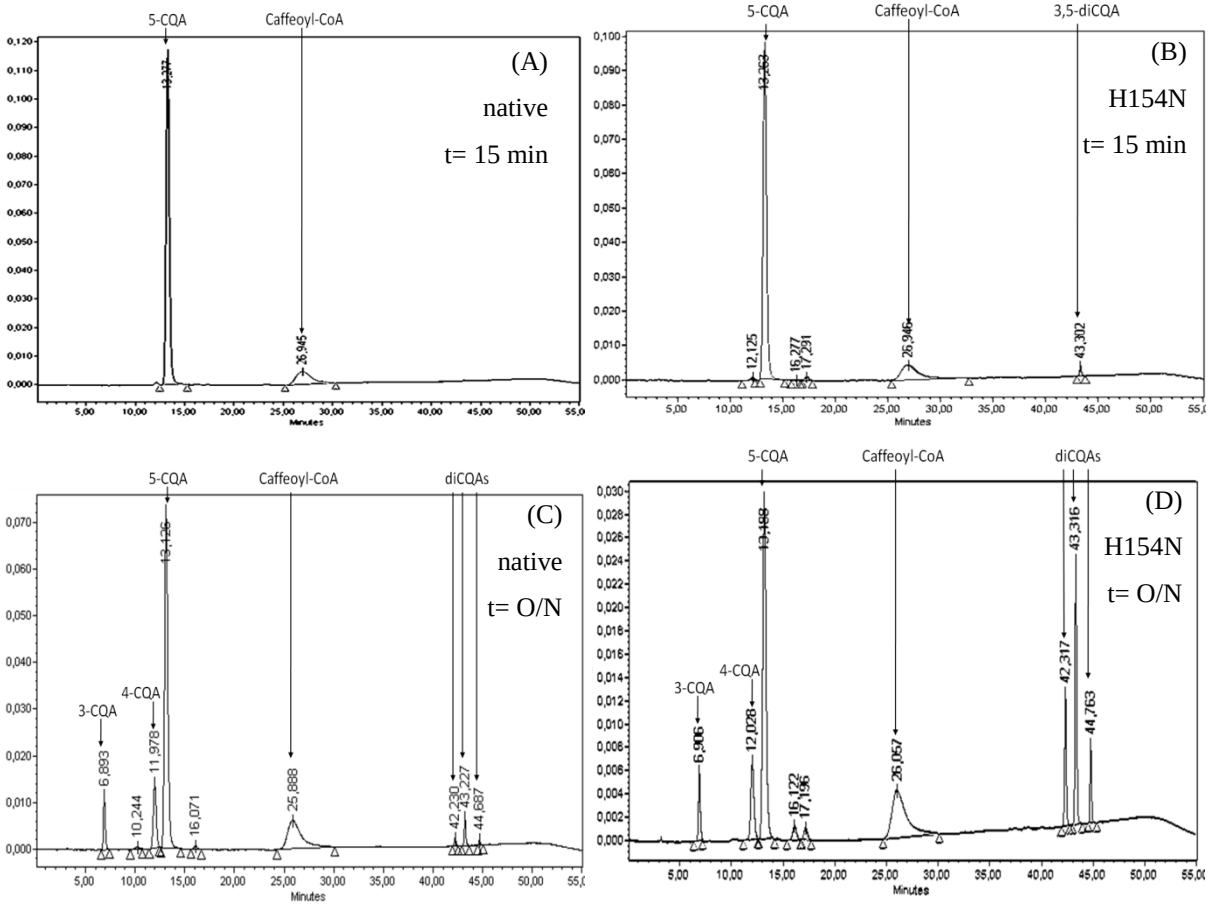


To our knowledge, the *in vitro* enzymatic synthesis of CGA diesters has never been reported to date. Initially, some diCQA peaks were observed in reactions catalysed by *Cc*HCT after overnight incubation, but only in very low amounts. The biochemical study of two mutant HCTs, H154N and H154N/A155L/A156S, revealed that these mutants could produce substantial amounts of diCQAs (Figure 93D). These mutants both target the second histidine residue of the HHAAD motif that is conserved in all HCTs (Figure 96). They were initially designed to determine if the second histidine could be involved in catalysis and to change the *Cc*HCT motif to the one found in *Cc*HQT. Interestingly, mutating the conserved HHAAD (153-157) motif in *Cc*HCT into HNLSD, as in *Cc*HQT, does not affect the preference for

shikimic over quinic acid. Remarkably though, the H154N and H154N/A155L/A156S mutant *Cc*HCTs favour the formation of diCQAs by a factor of 4 to 5 when compared to the native *Cc*HCT (Figure 93). A major peak corresponding to 3,5-diCQA was formed from 5-CQA and CoA and it always appeared first, suggesting that it was synthesised enzymatically (Figure 93). Peaks corresponding to the diCQAs and diFQAs were also observed in *Cc*HQT-catalysed reactions, and confirmed by MS analysis. However, *Cc*HCT is more efficient and therefore the native and mutant *Cc*HCTs were principally studied.

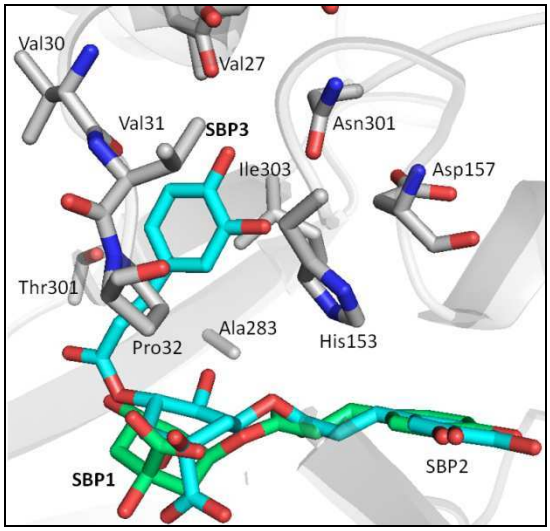
Figure 93: HPLC profiles showing the formation of diCQAs catalysed by the native and H154N mutant HCTs

Reactions containing 0.2 mM caffeoyl-CoA and 1 mM 5-CQA in 0.1 M sodium phosphate pH 6.5 were started by adding 0.5 µM native/ H154N HCT. Samples taken at t= 15 min and after overnight (O/N) incubation were analysed using HPLC-PDA with the acetonitrile method. (A) Native HCT, t= 15 min (sample 1169). (B) H154N mutant HCT, t=15 min (sample 1174). (C) Native HCT, t= O/N (sample 1184). (E) H154N mutant HCT, t= O/N (sample 1189).



As *Cc*HCT was observed to produce 3,5-diCQA, this ligand was also docked in *Cc*HCT. It resulted in a different pocket (SBP3) for the acyl moiety at C-5 of quinic acid (Figure 94). The quinic moiety was located in SBP1 and the acyl moiety at C-3 in SBP2, while bringing the ester oxygen close to the His153 side-chain. Docking of caffeoyl-CoA thioester with a shikimic acid in SBP1 results in an acyl moiety either in SBP2 or SBP3, similarly to the 3,5-diCQA. SBP3 is composed of Val27-Pro32, His153, Asp157, Asn301 and Ile303. The His154 side-chain is within 6 Å of Asn24, Asp28-Asn33, Leu74, Ile82 and Arg152, with the residues Val30, Val31 and Pro32 part of SBP3. The formation of diesters that is favoured by the H154N and H154N/A155L/A156S mutant HCTs may therefore be due to the effect of the substitution on the Val30-Asn33 loop conformation, allowing more space for the 5-CQA acyl moiety to bind in SBP3. The 5-CQA could therefore bind with the quinic acid moiety flipped compared to the initial docking results, allowing this time esterification at C-3 by a second acyl moiety from caffeoyl-CoA. It might be easier for a 5-*O*-caffeoylshikimic acid (5-CSA) to flip because the polyol ring is more planar, therefore explaining why more shikimate esters (potentially 3-CSA) are observed in HCT-catalysed reactions from shikimic acid and caffeoyl-CoA. These assumptions should be confirmed by the structural study of the mutants targeting the His154 residue, preferably in complex with the corresponding substrates. The H154N/A155L/A156S mutant HCT produced some additional caffeic acid in the forward and reverse reactions. The presence of the free acid may be explained either by the distortion of the protein architecture, which may induce a thioesterase activity when caffeoyl-CoA binds in the active site, or because a mistransfer occurs. The control reaction (no enzyme added) showed no caffeic acid formation.

Figure 94: Docked 3,5-diCQA (cyan) and 5-CQA (green) in *Cc*HCT (grey)



As the 3,5-diCQA isomer appears first and is the major diester formed in the native and H154N mutant HCT-catalysed reactions (Figure 93), the presence of the 3,4- and 4,5-diCQA isomers is probably due to the chemical isomerisation of 3,5-diCQA. CGAs have been shown to be unstable at basic pH values and upon heating, thereby resulting in a mixture of the 3-, 4- and 5-CQA or FQA isomers. Similarly, the diCQAs isomerise to a mixture of the 3,4-, 3,5- and 4,5-diesters. Previous studies investigating diCQA synthesis in artichoke failed to identify any activity towards the biosynthesis or turnover of diCQA substrates (Comino *et al.*, 2009; Menin *et al.*, 2010). In our study, the formation of diCQAs occurred in the presence of 5-CQA and CoA, or caffeoyl-CoA and 5-CQA, so the formation of caffeoyl-CoA seems to be a necessary step in diCQA formation. HCT can convert back 3,5-diCQA into 5-CQA and caffeoyl-CoA, but HQT, from our observations, cannot. A possible mechanism is presented in Figure 95.

The exciting observation that diCQAs are formed in the enzymatic assays confirms previous suggestions that 3,5-diCQA derives from 5-CQA and caffeoyl-CoA. The native *Cc*HCT crystal structure was exploited for docking analyses and homology modelling of *Cc*HQT. These results have enabled the prediction of two potential binding sites for the hydroxycinnamoyl moieties and propose a structural basis for the acceptor preference between shikimic and quinic acid. His154 proved to be important an important residue in diCQA formation, most probably by influencing the architecture of the second acyl binding pocket. These results suggest the direct involvement of *Cc*HCT and *Cc*HQT in diCQA biosynthesis in the coffee grain.

Figure 95: Possible reaction mechanism for mono- and diCQA biosynthesis

: Chemical isomerisation

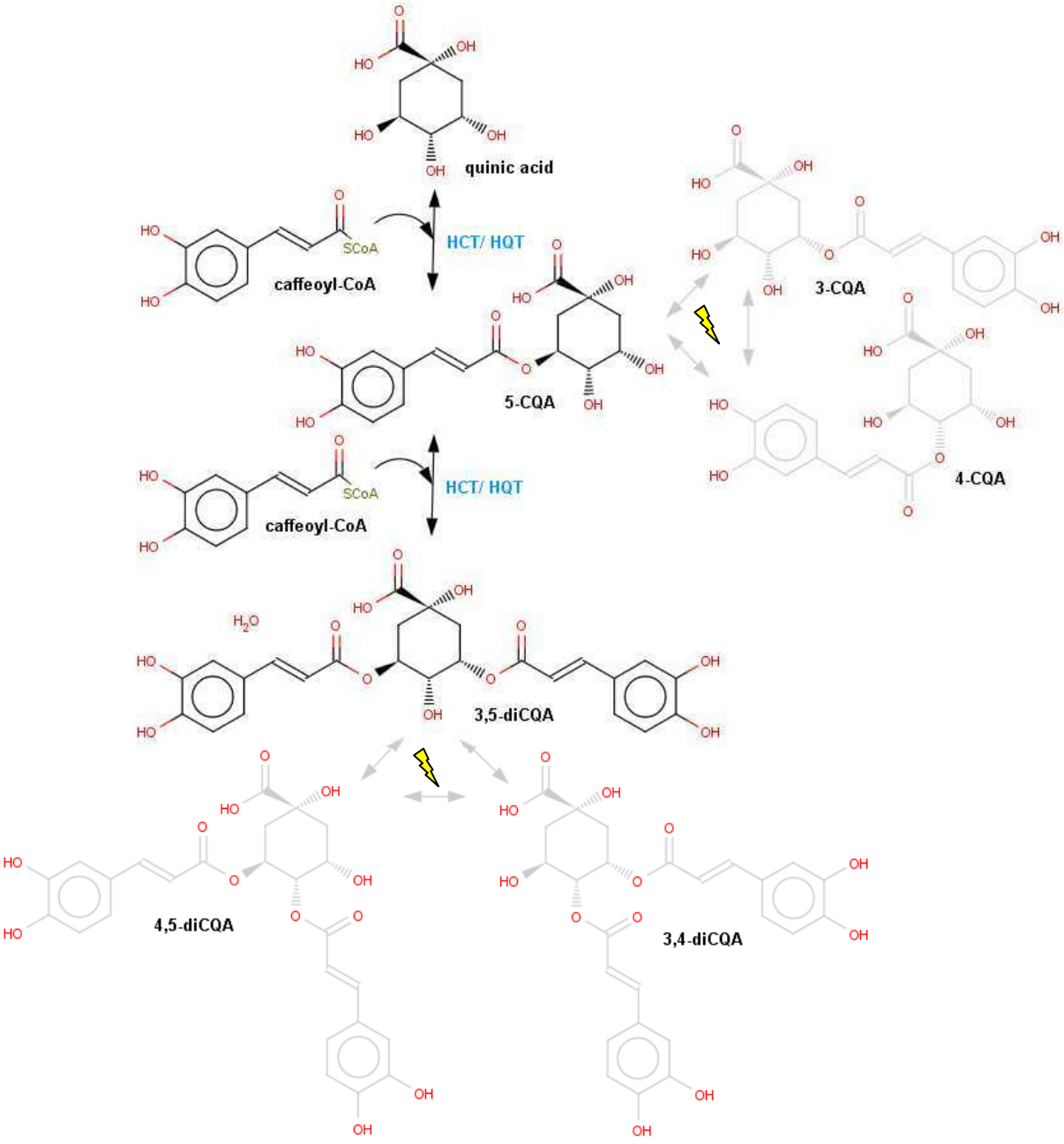
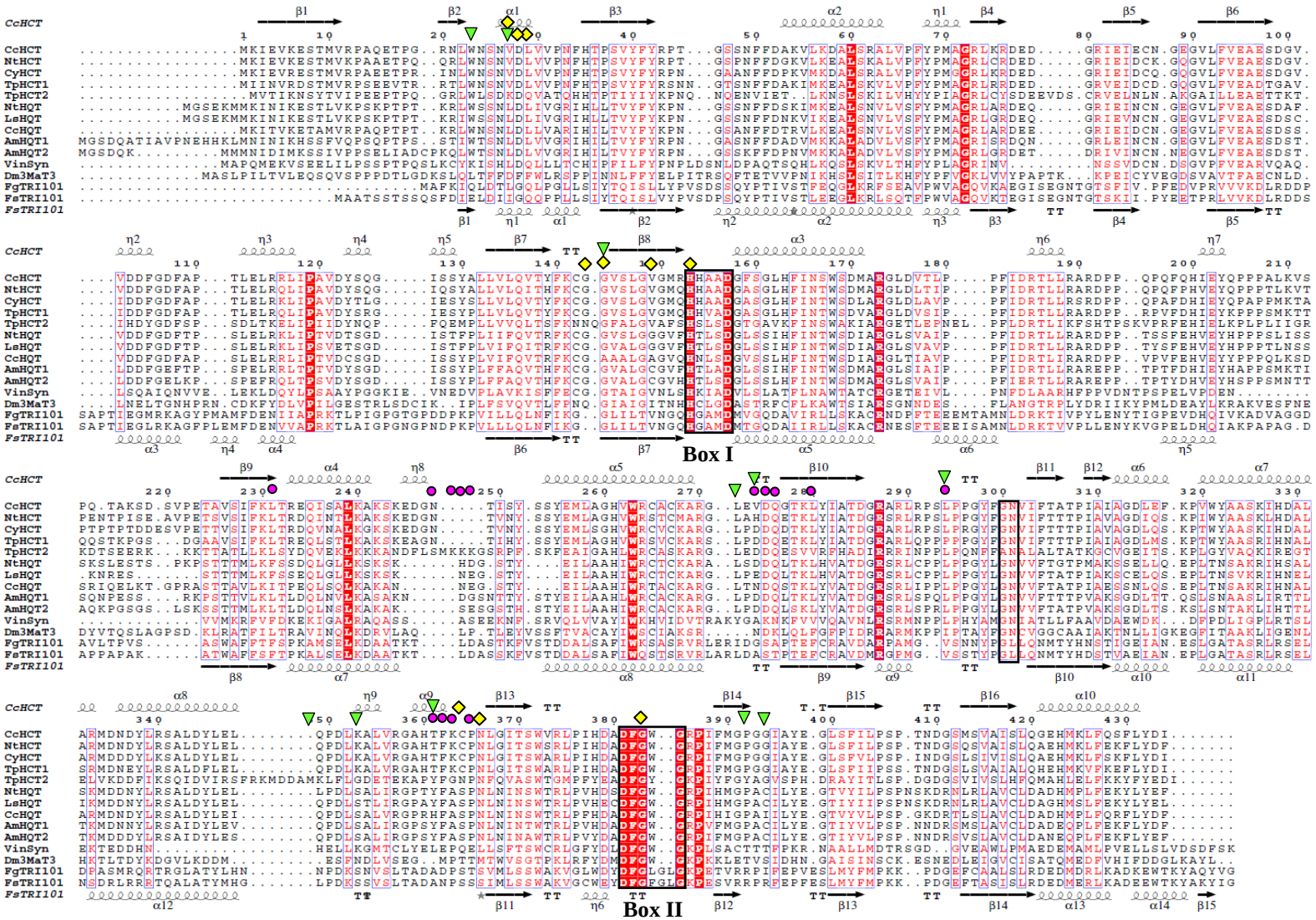


Figure 96: Sequence and structure alignment of *Cc*HCT with other hydroxycinnamoyltransferases and representative members of the BAHD superfamily

Secondary structures of the native HCT and *Fs*TRI101 (PDB: 2ZBA) are shown above and below the alignment respectively. Conserved motifs are highlighted by a box.



Putative residues involved in substrate binding

- acceptor
- acyl
- CoA

Chapter 5. Conclusion

This work has enabled the structural and biochemical characterisation of *Nt*4CL2, *Cc*HCT and *Cc*HQT, which are key enzymes in the phenylpropanoid biosynthetic pathway. The crystal structures of apo-*Nt*4CL2 and *Nt*4CL2-CoA-AMP ternary complex coupled with the comparison with the structures of other members of the adenylate-forming enzyme superfamily has allowed the identification of potential residues involved in catalysis and substrate specificity towards different substituted cinnamic acids. The crystal structures of native *Cc*HCT and a homology model of *Cc*HQT with substrate docking enabled to identify some residues, which may involved in substrate recognition and in the observed preference for quinic or shikimic acid.

In coffee plants, *Cc*HCT and *Cc*HQT appear to be responsible for the production of 5-CQA, 5-FQA and 3,5-diCQA, which can isomerise to the 3- and 4-positions, as well as mixed diesters. This study has therefore allowed an expansion of the structural diversity of the molecules that *Cc*HCT and *Cc*HQT can synthesise. The identification of particular conditions that allow the *in vitro* synthesis of CGA diesters is an exciting break-through because it has not been reported so far.

As such, a better understanding of the structure-function relationships among these enzyme families is of fundamental importance and could be exploited to produce these metabolites in plants or microbial systems for novel applications. CGAs differ with respect to their biological effects in plants or humans, and further studies are needed to compare their effectiveness as antioxidants or antimicrobial compounds. The results presented here therefore offer a means of producing diverse CGA derivatives with a high nutraceutical and pharmaceutical value.

Résumé général

Les acides chlorogéniques forment une famille d'esters formés entre des dérivés de l'acide *trans*-cinnamique avec l'acide quinique. Ces composés sont caractérisés par une grande diversité structurale et peuvent être classifiés selon l'identité, la position et le nombre des groupements acyles. Les acides hydroxycinnamiques les plus communs sont les acides coumarique (4-hydroxy), caféique (3,4-dihydroxy), férulique (3-méthoxy, 4-hydroxy) et sinapique (3,5-diméthoxy, 4-hydroxy). Les grains verts de *Coffea canephora* de la famille des Rubiacées sont connus pour avoir une teneur très élevée en acides chlorogéniques, qui sont également des composés phénoliques solubles majoritaires chez la tomate (Solanacées).

La teneur et la composition en acides chlorogéniques influencent les propriétés des aliments. Ils sont en effet partiellement dégradés par la chaleur et contribuent à la formation de produits de dégradation qui influencent les propriétés organoleptiques du café torréfié. Parmi tous les polyphénols, les acides chlorogéniques ont des propriétés antioxydantes reconnues qui aident à protéger les cellules contre les radicaux libres et le stress oxydant. Des études récentes de biodisponibilité montrent qu'ils sont facilement absorbés et métabolisés par l'organisme. Les aspects bénéfiques de ces composés comprennent, parmi d'autres, un risque diminué pour le développement de maladies cardiovasculaires et de certains cancers. Au sein de la plante, les acides chlorogéniques sont des produits du métabolisme secondaire impliqués à la fois dans son développement et sa résistance face aux contraintes de l'environnement. La biosynthèse et l'accumulation des acides chlorogéniques peut être induite à certains stades du développement des tissus et en réponse à certains stress biotiques et abiotiques (ex. stress oxydant, rayonnements du soleil). Au cours du vieillissement des feuilles, les dérivés d'acides phénoliques sont fortement accumulés dans les vaisseaux conducteurs par lignification. Les acides chlorogéniques sont en effet des intermédiaires dans la voie de biosynthèse des monomères de lignine, biopolymère caractérisé par une grande complexité chimique qui confère imperméabilité et résistance mécanique aux parois des plantes vasculaires.

La phénylalanine est produite par la voie métabolique de l'acide shikimique et est le précurseur des acides hydroxycinnamiques dans la voie générale des phénylpropanoïdes. Si certaines enzymes impliquées dans la voie des phénylpropanoïdes ont fait l'objet d'études biochimiques et structurales approfondies, des informations structurales sont nécessaires pour les acyl-CoA synthétases et acyltransférases catalysant respectivement la synthèse des thioesters hydroxycinnamoyl-CoA et des acides chlorogéniques. Trois gènes ont été isolés

chez *Coffea canephora* codant pour la 4-coumarate CoA ligase, 4CL, et les hydroxycinnamoyl-CoA quinate/shikimate hydroxycinnamoyltransférases, HQT et HCT. 4CL appartient à la superfamille des enzymes formant des adénylates. HCT et HQT diffèrent principalement par leur préférence pour l'acide quinique ou shikimique. Les acyltransférases coenzyme A-dépendantes ont été regroupées dans la superfamille nommée BAHD, dont le sigle correspond à la première lettre des quatre premiers gènes isolés. Elles utilisent pour substrats des esters de coenzyme A comme donneurs de groupements acyles. D'avantage de données structurales sont nécessaires, en particulier dans le clade auquel appartiennent HCT et HQT qui utilisent des substrats aromatiques.

Ce projet de thèse porte sur l'étude des relations structure-activité des enzymes 4CL, HCT et HQT impliquées dans la voie de biosynthèse des acides chlorogéniques et de leurs précurseurs, les thioesters de CoA. Les gènes *Cc*4cl2, *Cc*hct et *Cc*hqt ont été identifiés dans les données génétiques disponibles pour l'espèce *Coffea canephora* et clonés par Nestlé R&D à Tours (Lepelley *et al*., 2007). La sur-expression dans *E. coli* et la purification des protéines correspondantes ont été réalisé pour permettre leur caractérisation. Les méthodes utilisées comprennent des approches de biophysique pour étudier leur thermostabilité par fluorescence, mais aussi des essais enzymatiques par chromatographie liquide haute performance. Pour les études structurales, les techniques de cristallisation, de diffraction aux rayons X et de modélisation moléculaire ont été utilisées.

Les protéines recombinantes *Nt*4CL2, *Cc*HCT et *Cc*HQT ont été produites et purifiées dans une quantité et qualité nécessaires. Les thioesters de CoA des acides caféique, férulique, cinnamique et coumarique ont été synthétisés grâce à l'isoforme 2 de la 4-coumarate CoA ligase de *Nicotiana tabacum*. La structure cristalline de *Nt*4CL2 a été résolue sous sa forme native et en complexe avec CoA et AMP. Les résultats des études biochimiques pour *Cc*HCT et *Cc*HQT montrent que les protéines recombinantes sont actives et capables de produire des esters des acides quinique et shikimique. Les résultats montrent pour la première fois que l'enzyme *Cc*HCT est capable de synthétiser des acides dicaféoylquiniques. Des mutations de la protéine native ont conduit à des modifications de son activité enzymatique, notamment concernant le résidu His154 qui aurait une influence sur le mécanisme de di-acylation. Les résultats cristallographiques ont aboutis à la résolution de la première structure d'un représentant du clade V des acyltransférases CoA-dépendantes caractérisée par l'accommodation de substrats aromatiques. Des résultats de modélisation des substrats à l'intérieur des sites actifs ont été obtenus et permettent de proposer des déterminant structuraux de la spécificité de substrat de ces enzymes.

List of figures

List of tables

References

Abdullah, Y., B. Schneider and M. Petersen (2008). "Occurrence of rosmarinic acid, chlorogenic acid and rutin in Marantaceae species." Phytochemistry Letters **1**(4): 199-203.

Abdulrazzak, N., B. Pollet, J. Ehlting, K. Larsen, *et al.* (2006). "A coumaroyl-ester-3-hydroxylase insertion mutant reveals the existence of nonredundant meta-hydroxylation pathways and essential roles for phenolic precursors in cell expansion and plant growth." Plant Physiology **140**(1): 30-48.

Aerts, R. J. and T. W. Baumann (1994). "Distribution and utilization of chlorogenic acid in Coffea seedlings." Journal of Experimental Botany **45**(273): 497-503.

Arnold, K., L. Bordoli, J. Kopp and T. Schwede (2006). "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." Bioinformatics **22**(2): 195-201.

Ashihara, H., H. Sano and A. Crozier (2008). "Caffeine and related purine alkaloids: Biosynthesis, catabolism, function and genetic engineering." Phytochemistry **69**: 841–856.

Baldi, L., D. L. Hacker, M. Adam and F. M. Wurm (2007). "Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives." Biotechnology Letters **29**(5): 677-684.

Baneyx, F. (1999). "Recombinant protein expression in Escherichia coli." Current Opinion in Biotechnology **10**(5): 411-421.

Baucher, M., C. Halpin, M. Petit-Conil and W. Boerjan (2003). "Lignin: genetic engineering and impact on pulping." Critical Reviews in Biochemistry and Molecular Biology **38**(4): 305-350.

Becker-Andre, M., P. Schulze-Lefert and K. Hahlbrock (1991). "Structural comparison, modes of expression and putative cis-acting elements of the two 4-coumarate CoA ligase genes in potato." Journal of Biological Chemistry **266**(13): 8551-8559.

Berger, A., J. Meinhard and M. Petersen (2006). "Rosmarinic acid synthase is a new member of the superfamily of BAHD acyltransferases." Planta **224**(6): 1503-1510.

Berman, H. M., T. Battistuz, T. N. Bhat, W. F. Bluhm, *et al.* (2002). "The Protein Data Bank." Acta Crystallographica Section D-Biological Crystallography **58**: 899-907.

Bertrand, B., B. Guyot, F. Anthony and P. Lashermes (2003). "Impact of the coffea canephora gene introgression on beverage quality of C-arabica." Theoretical and Applied Genetics **107**(3): 387-394.

Besseau, S., L. Hoffmann, P. Geoffroy, C. Lapierre, *et al.* (2007). "Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth." Plant Cell **19**(1): 148-162.

Beuerle, T. and E. Pichersky (2002). "Enzymatic synthesis and purification of aromatic coenzyme a esters." Analytical Biochemistry **302**(2): 305-312.

Boerjan, W., J. Ralph and M. Baucher (2003). "Lignin biosynthesis." Annual Review of Plant Biology **54**: 519-546.

Branchini, B. R., M. H. Murtiashaw, R. A. Magyar and S. M. Anderson (2000). "The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase." Biochemistry **39**(18): 5433-5440.

Buglino, J., K. C. Onwueme, J. A. Ferreras, L. E. Quadri, *et al.* (2004). "Crystal structure of PapA5, a phthiocerol dimycocerosyl transferase from Mycobacterium tuberculosis." Journal of Biological Chemistry **279**(29): 30634-30642.

Burchard, P., W. Bilger and G. Weissenböck (2000). "Contribution of hydroxycinnamates and flavonoids to epidermal shielding of UV-A and UV-B radiation in developing rye primary leaves as assessed by ultraviolet-induced chlorophyll fluorescence measurements." Plant Cell & Environment **23**(12): 1373-1380.

Calabrese, J. C., D. B. Jordan, A. Boodhoo, S. Sariaslani, *et al.* (2004). "Crystal structure of phenylalanine ammonia lyase: Multiple helix dipoles implicated in catalysis." Biochemistry **43**(36): 11403-11416.

Caldwell, M. M., R. Robberecht and S. D. Flint (1983). "Internal filters - Prospects for UV-acclimation in higher plants." Physiologia Plantarum **58**(3): 445-450.

Campbell, M. M. and R. R. Sederoff (1996). "Variation in Lignin Content and Composition (Mechanisms of Control and Implications for the Genetic Improvement of Plants)." Plant Physiology **110**(1): 3-13.

Carter, C. W. and S. B. Xiang (1997). Phase improvement using conditional probability methods: Maximum entropy solvent flattening and phase permutation. Macromolecular Crystallography, Pt B. **277:** 79-109.

Cerovic, Z. G., A. Ounis, A. Cartelat, G. Latouche, *et al.* (2002). "The use of chlorophyll fluorescence excitation spectra for the non-destructive in situ assessment of UV-absorbing compounds in leaves." Plant Cell and Environment **25**(12): 1663-1676.

Chapple, C. C., T. Vogt, B. E. Ellis and C. R. Somerville (1992). "An Arabidopsis mutant defective in the general phenylpropanoid pathway." Plant Cell **4**(11): 1413-1424.

Chen, F. and R. A. Dixon (2007). "Lignin modification improves fermentable sugar yields for biofuel production." Nature Biotechnology **25**(7): 759-761.

Chen, V. B., W. B. Arendall, J. J. Headd, D. A. Keedy, *et al.* (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallographica Section D-Biological Crystallography **66**: 12-21.

Cheng (2007). "Plant Terpenoids: Biosynthesis and Ecological Functions." Journal of Integrative Plant Biology **49**(2): 179−186.

Cieslik, M. and Z. S. Derewenda (2009). "The role of entropy and polarity in intermolecular contacts in protein crystals." Acta Crystallographica Section D-Biological Crystallography **65**: 500-509.

Cle, C., L. M. Hill, R. Niggeweg, C. R. Martin, *et al.* (2008). "Modulation of chlorogenic acid biosynthesis in Solanum lycopersicum; consequences for phenolic accumulation and UV-tolerance." Phytochemistry **69**(11): 2149-2156.

Clifford, M. N. (1979). "Chlorogenic acids - Their complex nature and routine determination in coffee beans." Food Chemistry **4**(1): 63-71.

Clifford, M. N. (1985). "Chlorogenic acids. In: R. J. Clarke & R. Macrae (Eds) Coffee, Elsevier Applied Science, London." Chemistry **1**: 153-202.

Clifford, M. N. (1999). "Chlorogenic acids and other cinnamates - nature, occurrence and dietary burden." Journal of the Science of Food and Agriculture **79**(3): 362-372.

Clifford, M. N. (2000). "Chlorogenic acids and other cinnamates - nature, occurrence, dietary burden, absorption and metabolism." Journal of the Science of Food and Agriculture **80**(7): 1033-1043.

Clifford, M. N. (2004). "Diet-derived Phenols in plasma and tissues and their implications for health." Planta Medica **70**(12): 1103-1114.

Clifford, M. N., S. Marks, S. Knight and N. Kuhnert (2006). "Characterization by LC-MS(n) of four new classes of p-coumaric acid-containing diacyl chlorogenic acids in green coffee beans." Journal of the Science of Food and Agriculture **54**(12): 4095-4101.

Comino, C., A. Hehn, A. Moglia, B. Menin, *et al.* (2009). "The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway." BMC Plant Biology **9**(1): 30.

Comino, C., S. Lanteri, E. Portis, A. Acquadro, *et al.* (2007). "Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in Cynara cardunculus L." BMC Plant Biology **7**: 14.

Conti, E., N. P. Franks and P. Brick (1996). "Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes." Structure **4**(3): 287-298.

Conti, E., T. Stachelhaus, M. A. Marahiel and P. Brick (1997). "Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S." EMBO Journal **16**(14): 4174-4183.

Costa, M. A., D. L. Bedgar, S. G. A. Moinuddin, K.-W. Kim, *et al.* (2005). "Characterization in vitro and in vivo of the putative multigene 4-coumarate:CoA ligase network in Arabidopsis: syringyl lignin and sinapate/sinapyl alcohol derivative formation." Phytochemistry **66**(17): 2072-2091.

Cukovic, D., J. Ehlting, J. A. VanZiffle and C. J. Douglas (2001). "Structure and evolution of 4-coumarate : coenzyme A ligase (4CL) gene families." Biological Chemistry **382**(4): 645-654.

D'Auria, J. C. (2006). "Acyltransferases in plants: a good time to be BAHD." Current Opinion in Plant Biology **9**(3): 331-340.

Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, *et al.* (2007). "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids." Nucleic Acids Research **35**(Web Server issue): W375-383.

Dawidowicz, A. and R. Typek (2010). "Thermal stability of 5-O-caffeoylquinic acid in aqueous solutions at different heating conditions." Journal of the Science of Food and Agriculture **58**(24): 12578-12584.

De Castro, R. D. and P. Marraccini (2006). "Cytology, biochemistry and molecular changes during coffee fruit development." Brazilian Journal of Plant Physiology **18**: 175-199.

de Paulis, T., P. Commers, A. Farah, J. L. Zhao, *et al.* (2004). "4-caffeoyl-1,5-quinide in roasted coffee inhibits H-3 naloxone binding and reverses anti-nociceptive effects of morphine in mice." Psychopharmacology **176**(2): 146-153.

Derewenda, Z. S. and P. G. Vekilov (2006). "Entropy and surface engineering in protein crystallization." Acta Crystallographica Section D-Biological Crystallography **62**: 116-124.

Dixon, R. A. (2005). "Engineering of plant natural product pathways." Current Opinion in Plant Biology **8**(3): 329-336.

Dixon, R. A., C. J. Lamb, S. Masoud, V. J. Sewalt, *et al.* (1996). "Metabolic engineering: prospects for crop improvement through the genetic manipulation of phenylpropanoid biosynthesis and defense responses--a review." Gene **179**(1): 61-71.

Dixon, R. A. and N. L. Paiva (1995). "Stress-induced phenylpropanoid metabolism." Plant Cell **7**(7): 1085-1097.

Douglas, C. J. (1996). "Phenylpropanoid metabolism and lignin biosynthesis: From weeds to trees." Trends in Plant Science **1**(6): 171-178.

Drewnowski, A. (1997). "Taste preferences and food intake." Annual Review of Nutrition **17**: 237-253.

Dudareva, N., E. Pichersky and J. Gershenzon (2004). "Biochemistry of plant volatiles." Plant Physiology **135**(4): 1893-1902.

Ehlting, J., D. Buttner, Q. Wang, C. J. Douglas, *et al.* (1999). "Three 4-coumarate : coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms." Plant Journal **19**(1): 9-20.

Ehlting, J., N. J. Provart and D. Werck-Reichhart (2006). "Functional annotation of the Arabidopsis P450 superfamily based on large-scale co-expression analysis." Biochemical Society Transactions **34**: 1192-1198.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallographica Section D-Biological Crystallography **60**(Pt 12 Pt 1): 2126-2132.

Ericsson, U. B., B. M. Hallberg, G. T. Detitta, N. Dekker, *et al.* (2006). "Thermofluor-based high-throughput stability optimization of proteins for structural studies." Analytical Biochemistry **357**(2): 289-298.

Esposito, D. and D. Chatterjee (2006). "Enhancement of soluble protein expression through the use of fusion tags." Current Opinion in Biotechnology **17**: 353-358.

Evans, P. and A. McCoy (2008). "An introduction to molecular replacement." Acta Crystallographica Section D-Biological Crystallography **64**: 1-10.

Farah, A., T. de Paulis, L. C. Trugo and P. R. Martin (2005). "Effect of roasting on the formation of chlorogenic acid lactones in coffee." Journal of the Science of Food and Agriculture **53**(5): 1505-1513.

Farah, A., M. C. Monteiro, V. Calado, A. S. Franca, *et al.* (2006). "Correlation between cup quality and chemical attributes of Brazilian coffee." Food Chemistry **98**(2): 373-380.

Ferrer, J. L., M. B. Austin, C. Stewart, Jr. and J. P. Noel (2008). "Structure and function of enzymes involved in the biosynthesis of phenylpropanoids." Plant Physiology and Biochemistry **46**(3): 356-370.

Ferrer, J. L., J. M. Jez, M. E. Bowman, R. A. Dixon, *et al.* (1999). "Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis." Nature Structural Biology **6**(8): 775-784.

Finley, J. W. (2005). "Proposed criteria for assessing the efficacy of cancer reduction by plant foods enriched in carotenoids, glucosinolates, polyphenols and selenocompounds." Annals of Botany **95**(7): 1075-1096.

Flot, D., T. Mairs, T. Giraud, M. Guijarro, *et al.* (2010). "The ID23-2 structural biology microfocus beamline at the ESRF." Journal of Synchrotron Radiation **17**(1): 107-118.

Foley, S., S. Navaratnam, D. J. McGarvey, E. J. Land, *et al.* (1999). "Singlet oxygen quenching and the redox properties of hydroxycinnamic acids." Free Radical Biology and Medicine **26**(9-10): 1202-1208.

Frank, O., G. Zehentbauer and T. Hofmann (2006). "Bioresponse-guided decomposition of roast coffee beverage and identification of key bitter taste compounds." European Food Research and Technology **222**(5-6): 492-508.

Franke, R., J. M. Humphreys, M. R. Hemm, J. W. Denault, *et al.* (2002). "The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism." Plant Journal **30**(1): 33-45.

Freyer, M. W. and E. A. Lewis (2008). Isothermal titration calorimetry: Experimental design, data analysis, and probing Macromolecule/Ligand binding and kinetic interactions. Biophysical Tools for Biologists: Vol 1 in Vitro Techniques. **84:** 79-113.

Fulda, M., E. Heinz and F. P. Wolter (1994). "The fadD gene of Escherichia coli K12 is located close to rnd at 39.6 min of the chromosomal map and is a new member of the AMP-binding protein family." Molecular and General Genetics **242**(3): 241-249.

Garman, E. F. and S. M. McSweeney (2007). "Progress in research into radiation damage in cryo-cooled macromolecular crystals." Journal of Synchrotron Radiation **14**: 1-3.

Garvey, G. S., S. P. McCormick, N. J. Alexander and I. Rayment (2009). "Structural and functional characterization of TRI3 trichothecene 15-O-acetyltransferase from Fusarium sporotrichioides." Protein Science **18**(4): 747-761.

Garvey, G. S., S. P. McCormick and I. Rayment (2008). "Structural and functional characterization of the TRI101 trichothecene 3-O-acetyltransferase from Fusarium sporotrichioides and Fusarium graminearum: kinetic insights to combating Fusarium head blight." Journal of Biological Chemistry **283**(3): 1660-1669.

Gershenzon, J. and N. Dudareva (2007). "The function of terpene natural products in the natural world." Nature Chemical Biology **3**(7): 408-414.

Goodsell, D. S., G. M. Morris and A. J. Olson (1996). "Automated docking of flexible ligands: applications of AutoDock." Journal of Molecular Recognition **9**(1): 1-5.

Gouet, P., E. Courcelle, D. I. Stuart and F. Metoz (1999). "ESPript: analysis of multiple sequence alignments in PostScript." Bioinformatics **15**(4): 305-308.

Gouet, P., X. Robert and E. Courcelle (2003). "ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins." Nucleic Acids Research **31**(13): 3320-3323.

Goupy, P. M., P. J. A. Varoquaux, J. J. Nicolas and J. J. Macheix (1990). "Identification and localization of hydroxycinnamoyl and flavonol derivatives from endive (Cichorium endivia L. cv geante maraichere) leaves." Journal of Agricultural and Food Chemistry **38**(12): 2116-2121.

Grienenberger, E., S. Besseau, P. Geoffroy, D. Debayle, *et al.* (2009). "A BAHD acyltransferase is expressed in the tapetum of Arabidopsis anthers and is involved in the synthesis of hydroxycinnamoyl spermidines." Plant Journal **58**(2): 246-259.

Griffiths, D. W. and H. Bain (1997). "Photo-induced changes in the concentrations of individual chlorogenic acid isomers in potato (Solanum tuberosum) tubers and their complexation with ferric ions." Potato Research **40**(3): 307-315.

Griffiths, D. W., H. Bain and M. F. B. Dale (1992). "Development of a rapid colorimetric method for the determination of chlorogenic acid in freeze-dried potato tubers." Journal of the Science of Food and Agriculture **58**(1): 41-48.

Gulick, A. M. (2009). "Conformational Dynamics in the Acyl-CoA Synthetases, Adenylation Domains of Non-ribosomal Peptide Synthetases, and Firefly Luciferase." Acs Chemical Biology **4**(10): 811-827.

Hahlbrock, K. and D. Scheel (1989). "Physiology and molecular biology of phenylpropanoid metabolism." Annual Review of Plant Physiology and Plant Molecular Biology **40**: 347-369.

Hamberger, B. and K. Hahlbrock (2004). "The 4-coumarate:CoA ligase gene family in Arabidopsis thaliana comprises one rare, sinapate-activating and three commonly occurring isoenzymes." Proceedings of the National Academy of Sciences of the United States of America **101**(7): 2209-2214.

Harborne, J. B., C. A. Williams, J. Greenham and P. Moyna (1974). "Distribution of charged flavones and caffeoylshikimic acid in Palmae." Phytochemistry **13**(8): 1557-1559.

Hartmann, T. (2007). "From waste products to ecochemicals: Fifty years research of plant secondary metabolism." Phytochemistry **68**(22-24): 2831-2846.

Hendlich, M., F. Rippmann and G. Barnickel (1997). "LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins." Journal of Molecular Graphics & Modelling **15**(6): 359-363.

Hendrickson, W. A., A. Pahler, J. L. Smith, Y. Satow, *et al.* (1989). "Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation." Proceedings of the National Academy of Sciences of the United States of America **86**(7): 2190-2194.

Herrmann, K. M. (1995). "The shikimate pathway - Early steps in the biosynthesis of aromatic compounds." Plant Cell **7**(7): 907-919.

Herrmann, K. M. and L. M. Weaver (1999). "The shikimate pathway." Annual Review of Plant Physiology and Plant Molecular Biology **50**: 473–503.

Heyda, J., P. Mason and P. Jungwirth (2010). "Attractive interactions between side chains of histidine-histidine and histidine-arginine-based cationic dipeptides in water." Journal of Physical Chemistry B **114**(26): 8744-8749.

Hoffmann, L., S. Besseau, P. Geoffroy, C. Ritzenthaler, *et al.* (2004). "Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis." Plant Cell **16**(6): 1446-1465.

Hoffmann, L., S. Besseau, P. Geoffroy, C. Ritzenthaler, *et al.* (2005). "Acyltransferase-catalysed p-coumarate ester formation is a committed step of lignin biosynthesis." Plant Biosystems **139**(1): 50-53.

Hoffmann, L., S. Maury, F. Martz, P. Geoffroy, *et al.* (2003). "Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism." Journal of Biological Chemistry **278**(1): 95-103.

Holm, L. and P. Rosenstrom (2010). "Dali server: conservation mapping in 3D." Nucleic Acids Research **38**: W545-W549.

Horswill, A. R. and J. C. Escalante-Semerena (2002). "Characterization of the propionyl-CoA synthetase (PrpE) enzyme of Salmonella enterica: Residue Lys592 is required for propionyl-AMP synthesis." Biochemistry **41**(7): 2379-2387.

Hu, W. J., A. Kawaoka, C. J. Tsai, J. H. Lung, *et al.* (1998). "Compartmentalized expression of two structurally and functionally distinct 4-coumarate : CoA ligase genes in aspen (Populus tremuloides)." Proceedings of the National Academy of Sciences of the United States of America **95**(9): 5407-5412.

Hunt, I. (2005). "From gene to protein: a review of new and enabling technologies for multiparallel protein expression." Protein Expression and Purification **40**: 1-22.

Ilari, A. and C. Savino (2008). "Protein structure determination by X-ray crystallography." Bioinformatics **452**(3).

IUPAC (1976). "Nomenclature of cyclitols." Biochemical Journal **153**: 23-31.

Iwashina, T. (2003). "Flavonoid function and activity to plants and other organisms." Biological Sciences in Space **17**(1): 24-44.

Jaiswal, R. (2010). "Profile and Characterization of the Chlorogenic Acids in Green Robusta Coffee Beans by LC-MSn: Identification of Seven New Classes of Compounds." Journal of Agricultural and Food Chemistry **58**: 8722–8737.

Jiang, D. S. and D. G. Peterson (2010). "Role of hydroxycinnamic acids in food flavor: a brief overview." Phytochemistry Reviews **9**(1): 187-193.

Jogl, G. and L. Tong (2003). "Crystal structure of carnitine acetyltransferase and implications for the catalytic mechanism and fatty acid transport." Cell **112**(1): 113-122.

Johnston, K. L., M. N. Clifford and L. M. Morgan (2003). "Coffee acutely modifies gastrointestinal hormone secretion and glucose tolerance in humans: glycemic effects of chlorogenic acid and caffeine." American Journal of Clinical Nutrition **78**(4): 728-733.

Kabsch, W. (2010). "XDS." Acta Crystallographica Section D-Biological Crystallography. **66**(2): 125-132.

Kahnt, G. (1967). "Trans-cis-equilibrium of hydroxycinnamic acids during irradiation of aqueous solutions at different pH." Phytochemistry **6**(5): 755-&.

Kane, J. F. (1995). "Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli." Current Opinion in Biotechnology **6**(5): 494-500.

Kleywegt, G. J. (1996). "Use of non-crystallographic symmetry in protein structure refinement." Acta Crystallographica Section D-Biological Crystallography **52**: 842-857.

Kochan, G., E. S. Pilka, F. von Delft, U. Oppermann, *et al.* (2009). "Structural Snapshots for the Conformation-dependent Catalysis by Human Medium-chain Acyl-coenzyme A Synthetase ACSM2A." Journal of Molecular Biology **388**(5): 997-1008.

Koshiro, Y., M. C. Jackson, R. Katahira, M. L. Wang, *et al.* (2007). "Biosynthesis of chlorogenic acids in growing and ripening fruits of Coffea arabica and Coffea

canephora plants." Zeitschrift Fur Naturforschung Section C-a Journal of Biosciences **62**(9-10): 731-742.

Kost, T. A. and J. P. Condreay (1999). "Recombinant baculoviruses as expression vectors for insect and mammalian cells." Current Opinion in Biotechnology **10**(5): 428-433.

Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." Journal of Molecular Biology **372**(3): 774-797.

Ku, H. M., T. Vision, J. P. Liu and S. D. Tanksley (2000). "Comparing sequenced segments of the tomato and Arabidopsis genomes: Large-scale duplication followed by selective gene loss creates a network of synteny." Proceedings of the National Academy of Sciences of the United States of America **97**(16): 9121-9126.

Ky, C. L., J. Louarn, S. Dussert, B. Guyot, *et al.* (2001). "Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild Coffea arabica L. and C-canephora P. accessions." Food Chemistry **75**(2): 223-230.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, *et al.* (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Lattanzio, V., P. A. Kroon, S. Quideau and D. Treutter (2008). Plant Phenolics - Secondary Metabolites with Diverse Functions. Recent Advances in Polyphenol Research, Vol 1. F. Daayf and V. Lattanzio. Oxford, Blackwell Publishing. **1:** 1-35.

Lee, D. and C. J. Douglas (1996). "Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family - cDNA structure, gene inheritance and expression, and properties of recombinant proteins." Plant Physiology **112**(1): 193-205.

Leloup, V., A. Louvrier and R. Liardon (1995). "Degradation mechanisms of chlorogenic acids during roasting." Proc. 16th Int. Sci. Coll. Coffee (Kyoto) ASIC, Paris: 192-198.

Lepelley, M., G. Cheminade, J. G. McCarthy, V. Pétiard, *et al.* (2007). Polynucleotides encoding phenylpropanoid and flavonoid biosynthetic pathway enzymes in coffee.

Lepelley, M., G. Cheminade, N. Tremillon, A. Simkin, *et al.* (2007). "Chlorogenic acid synthesis in coffee: An analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in C. canephora." Plant Science **172**(5): 978-996.

Leslie, J. R. (1988). "Structure of Chloramphenicol Acetyltransferase at 1.75- angstrom Resolution." Proceedings of the National Academy of Sciences of the United States of America **85**(12): 4133–4137.

Li, M., Z. Su and J. Janson (2004). "In vitro protein refolding by chromatography procedures." Protein Expression and Purification **33**: 1-10.

Liang, Y. (2008). "Applications of isothermal titration calorimetry in protein science." Acta Biochimica Et Biophysica Sinica **40**(7): 565-576.

Lin, C., L. A. Mueller, J. Mc Carthy, D. Crouzillat, *et al.* (2005). "Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts." Theorical and Applied Genetics **112**(1): 114-130.

Lindermayr, C., J. Fliegmann and J. Ebel (2003). "Deletion of a single amino acid residue from different 4-coumarate : CoA Ligases from soybean results in the generation of new substrate specificities." Journal of Biological Chemistry **278**(5): 2781-2786.

Lindermayr, C., B. Mollers, J. Fliegmann, A. Uhlmann, *et al.* (2002). "Divergent members of a soybean (Glycine max L.) 4-coumarate : coenzyme A ligase gene family - Primary structures, catalytic properties, and differential expression." European Journal of Biochemistry **269**(4): 1304-1315.

Long, F., A. A. Vagin, P. Young and G. N. Murshudov (2008). "BALBES: a molecular-replacement pipeline." Acta Crystallographica Section D-Biological Crystallography **64**: 125-132.

Lowe, D. (2009). "The great acetonitrile shortage." Corante.com.
Luo, J., C. Fuell, A. Parr, L. Hill, *et al.* (2009). "A Novel Polyamine Acyltransferase Responsible for the Accumulation of Spermidine Conjugates in Arabidopsis Seed." Plant Cell **21**(1): 318-333.
Ma, X., J. Koepke, A. Bayer, V. Linhard, *et al.* (2004). "Vinorine synthase from Rauvolfia: the first example of crystallization and preliminary X-ray diffraction analysis of an enzyme of the BAHD superfamily." Biochimica et Biophysica Acta **1701**(1-2): 129-132.
Ma, X., J. Koepke, S. Panjikar, G. Fritzsch, *et al.* (2005). "Crystal structure of vinorine synthase, the first representative of the BAHD superfamily." Journal of Biological Chemistry **280**(14): 13576-13583.
Mahesh, V., R. Million-Rousseau, P. Ullmann, N. Chabrillange, *et al.* (2007). "Functional characterization of two p-coumaroyl ester 3'-hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis." Plant Molecular Biology **64**(1-2): 145-159.
Maier, V. P., D. M. Metzler and A. F. Huber (1964). "3-3-Caffeoylshikimic acid (dactylifric acid) and its isomers a new class of enzymic browning substrates." Biochemical and Biophysical Research Communications **14**(2): 124-128.
Malakhov, M. P., M. R. Mattern, O. A. Malakhova, M. Drinker, *et al.* (2004). "SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins." Journal of Structural and Functional Genomics **5**(1-2): 75-86.
Malissard, M., S. Zeng and E. G. Berger (1999). "The yeast expression system for recombinant glycosyltransferases." Glycoconjugate Journal **16**(2): 125-139.
Matsuda, F., K. Morino, R. Ano, M. Kuzawa, *et al.* (2005). "Metabolic flux analysis of the phenylpropanoid pathway in elicitor-treated potato tuber tissue." Plant and Cell Physiology **46**(3): 454-466.
McCarthy, A. A., S. Brockhauser, D. Nurizzo, P. Theveneau, *et al.* (2009). "A decade of user operation on the macromolecular crystallography MAD beamline ID14-4 at the ESRF." Journal of Synchrotron Radiation **16**: 803-812.
McCarty, M. F. (2005). "A chlorogenic acid-induced increase in GLP-1 production may mediate the impact of heavy coffee consumption on diabetes risk." Medical Hypotheses **64**(4): 848-853.
McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, *et al.* (2007). "Phaser crystallographic software." Journal of Applied Crystallography **40**: 658-674.
McDougall, B., P. J. King, B. W. Wu, Z. Hostomsky, *et al.* (1998). "Dicaffeoylquinic and dicaffeoyltartaric acids are selective inhibitors of human immunodeficiency virus type 1 integrase." Antimicrobial Agents and Chemotherapy **42**(1): 140-146.
McPherson, A. (1982)." Preparation and analysis of protein crystals." Wiley.
Menin, B., C. Comino, A. Moglia, Y. Dolzhenko, *et al.* (2010). "Identification and mapping of genes related to caffeoylquinic acid synthesis in *Cynara cardunculus* L." Plant Science **179**: 338-347.
Meurisse, R., R. Brasseur and A. Thomas (2003). "Aromatic side-chain interactions in proteins. Near- and far-sequence His-X pairs." Biochimica et Biophysica Acta **1649**: 85-96.
Moglia, A., C. Comino, L. S., R. de Vos, *et al.* (2010). "Production of novel antioxidative phenolic amides through heterologous expression of the plant's chlorogenic acid biosynthesis genes in yeast. ." Metabolic Engineering **12**(3): 223-232.
Mondolot, L., P. La Fisca, B. Buatois, E. Talansier, *et al.* (2006). "Evolution in caffeoylquinic acid content and histolocalization during Coffea canephora leaf development." Annals of Botany **98**(1): 33-40.

Moriguchi, T., R. J. A. Villegas, T. Kondo and M. Kojima (1988). "Isolation of 1-O-trans-para-coumaroyl-beta-deuterium-glucopyranose from sweet-potato roots and examination of its role in chlorogenic acids biosynthesis." Plant and Cell Physiology **29**(7): 1221-1226.

Murata, M., H. Okada and S. Homma (1995). "Hydroxycinnamic acid derivatives and p-coumaroyl-(L)-tryptophan, a novel hydroxycinnamic acid derivative, from coffee beans." Bioscience Biotechnology and Biochemistry **59**(10): 1887-1890.

Murata, M., M. Sugiura, Y. Sonokawa, T. Shimamura, *et al.* (2002). "Properties of chlorogenic acid quinone: Relationship between browning and the formation of hydrogen peroxide from a quinone solution." Bioscience Biotechnology and Biochemistry **66**(12): 2525-2530.

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallography Section D-Biological Crystallography **53**: 240-255.

Nair, R. B., Q. Xia, C. J. Kartha, E. Kurylo, *et al.* (2002). "Arabidopsis CYP98A3 mediating aromatic 3-hydroxylation. Developmental regulation of the gene, and expression in yeast." Plant Physiology **130**(1): 210-220.

Nakatsu, T., S. Ichiyama, J. Hiratake, A. Saldanha, *et al.* (2006). "Structural basis for the spectral difference in luciferase bioluminescence." Nature **440**(7082): 372-376.

Nardini, M., E. Cirillo, F. Natella and C. Scaccini (2002). "Absorption of phenolic acids in humans after coffee consumption." Journal of Agricultural and Food Chemistry **50**(20): 5735-5741.

Natella, F., M. Nardini, F. Belelli and C. Scaccini (2007). "Coffee drinking induces incorporation of phenolic acids into LDL and increases the resistance of LDL to ex vivo oxidation in humans." American Journal of Clinical Nutrition **86**(3): 604-609.

Niggeweg, R., A. J. Michael and C. Martin (2004). "Engineering plants with increased levels of the antioxidant chlorogenic acid." Nature Biotechnology **22**(6): 746-754.

Noirot, M., C. Campa, et al. (2004)." Discontinuous distribution of caffeine content among wild Coffea species." 20th International Conference on Coffee Science (ASIC), Bangalore, India.

Ohiokpehai, O., G. Brumen and M. N. Clifford (1983). "The chlorogenic acids content of some peculiar green coffee beans and the implications for beverage quality." 10th International Colloquium on the Chemistry of Coffee (ASIC): 177-186.

Pantoliano, M. W., E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, *et al.* (2001). "High-density miniaturized thermal shift assays as a general strategy for drug discovery." J Biomol Screen **6**(6): 429-440.

Petersen, M., Y. Abdullah, J. Benner, D. Eberle, *et al.* (2009). "Evolution of rosmarinic acid biosynthesis." Phytochemistry **70**(15-16): 1663-1679.

Petersen, M. and M. S. Simmonds (2003). "Rosmarinic acid." Phytochemistry **62**(2): 121-125.

Pichersky, E. and J. Gershenzon (2002). "The formation and function of plant volatiles: perfumes for pollinator attraction and defense." Current Opinion in Plant Biology **5**(3): 237-243.

Raes, J., A. Rohde, J. H. Christensen, Y. Van de Peer, *et al.* (2003). "Genome-wide characterization of the lignification toolbox in Arabidopsis." Plant Physiology **133**(3): 1051-1071.

Ranjeva, R., G. Alibert and A. M. Boudet (1977). "Metabolism of phenolic compounds in petunia. 6. Studies on biosynthesis of chlorogenic acidand naringenin by isolated chloroplasts." Plant Science Letters **10**(3): 235-242.

Ravelli, R. B. G. and E. F. Garman (2006). "Radiation damage in macromolecular cryocrystallography." Current Opinion in Structural Biology **16**(5): 624-629.

Reger, A. S., J. M. Carney and A. M. Gulick (2007). "Biochemical and crystallographic analysis of substrate binding and conformational changes in Acetyl-CoA synthetase." Biochemistry **46**(22): 6536-6546.

Reger, A. S., R. Wu, D. Dunaway-Mariano and A. M. Gulick (2008). "Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-Chlorobenzoate : CoA ligase." Biochemistry **47**(31): 8016-8025.

Reverter, D. and C. D. Lima (2006). "Structural basis for SENP2 protease interactions with SUMO precursors and conjugated substrates." Nature Structural & Molecular Biology **13**(12): 1060-1068.

Rice-Evans, C., N. Miller and G. Paganga (1997). "Antioxidant properties of phenolic compounds." Trends in Plant Science **2**: 152-159.

Rice-Evans, C. A., N. J. Miller and G. Paganga (1996). "Structure-antioxidant activity relationships of flavonoids and phenolic acids." Free Radical Biology and Medicine **20**(7): 933-956.

Ritter, H. and G. E. Schulz (2004). "Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase." Plant Cell **16**(12): 3426-3436.

Ritter, H. and G. E. Schulz (2004). "Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase." Plant Cell **16**: 3426-3436.

Rogers, W. J., S. Michaux, M. Bastin and P. Bucheli (1999). "Changes to the content of sugars, sugar alcohols, myo-inositol, carboxylic acids and inorganic anions in developing grains from different varieties of Robusta (Coffea canephora) and Arabica (C. arabica) coffees." Plant Science **149**(2): 115-123.

Rohde, A., K. Morreel, J. Ralph, G. Goeminne, *et al.* (2004). "Molecular phenotyping of the pal1 and pal2 mutants of Arabidopsis thaliana reveals far-reaching consequences on and carbohydrate metabolism." Plant Cell **16**(10): 2749-2771.

Ruelas, C., M. Tiznado-Hernandez, A. Sanchez-Estrada, M. Robles-Burgueno, *et al.* (2006). "Changes in phenolic acid content during Alternaria alternata in tomato fruit." J Phytopathology **154**(4): 236-244.

Saraste, M., P. R. Sibbald and A. Wittinghofer (1990). "The P-loop - a common motif in ATP-binding and GTP-binding proteins." Trends in Biochemical Sciences **15**(11): 430-434.

Scalbert, A., C. Morand, C. Manach and C. Remesy (2002). "Absorption and metabolism of polyphenols in the gut and impact on health." Biomed Pharmacother **56**(6): 276-282.

Schmelz, S. and J. H. Naismith (2009). "Adenylate-forming enzymes." Current Opinion in Structural Biology **19**(6): 666-671.

Schneider, K., L. Kienow, E. Schmelzer, T. Colby, *et al.* (2005). "A new type of peroxisomal acyl-coenzyme A synthetase from Arabidopsis thaliana has the catalytic capacity to activate biosynthetic precursors of jasmonic acid." Journal of Biological Chemistry **280**(14): 13962-13972.

Schoch, G., S. Goepfert, M. Morant, A. Hehn, *et al.* (2001). "CYP98A3 from Arabidopsis thaliana is a 3 '-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway." Journal of Biological Chemistry **276**(39): 36566-36574.

Schuttelkopf, A. W. and D. M. F. van Aalten (2004). "PRODRG: a tool for high-throughput crystallography of protein-ligand complexes." Acta Crystallographica Section D **60**(8): 1355-1363.

Shahidi, F. and A. Chandrasekara (2010). "Hydroxycinnamates and their in vitro and in vivo antioxidant activities." Phytochemistry Reviews **9**(1): 147-170.

Shomura, Y., I. Torayama, D. Y. Suh, T. Xiang, *et al.* (2005). "Crystal structure of stilbene synthase from Arachis hypogaea." Proteins-Structure Function and Bioinformatics **60**(4): 803-806.

Sondheimer, E. (1964). "Chlorogenic acids and related depsides." Botanical Review **30**(4): 667-712.

Sonnante, G., R. D'Amore, E. Blanco, C. L. Pierri, *et al.* (2010). "Novel Hydroxycinnamoyl-Coenzyme A Quinate Transferase Genes from Artichoke Are Involved in the Synthesis of Chlorogenic Acid." Plant Physiology **153**(3): 1224-1238.

Souza, C. D., B. Barbazuk, S. G. Ralph, J. Bohlmann, *et al.* (2008). "Genome-wide analysis of a land plant-specific acyl : coenzymeA synthetase (ACS) gene family in Arabidopsis, poplar, rice and Physcomitrella." New Phytologist **179**(4): 987-1003.

St-Pierre, B. and V. De Luca (2000). Evolution of acyltransferase genes: Origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. Evolution of Metabolic Pathways. J. T. Romeo, R. Ibrahim, L. Varin and V. DeLuca. Kidlington, Pergamon-Elsevier Science Ltd. **34:** 285-315.

St-Pierre, B., P. Laflamme, A. M. Alarco and V. De Luca (1998). "The terminal O-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A-dependent acyl transfer." Plant Journal **14**(6): 703-713.

Stalmach, A., W. Mullen, D. Barron, K. Uchida, *et al.* (2009). "Metabolite profiling of hydroxycinnamate derivatives in plasma and urine after the ingestion of coffee by humans: identification of biomarkers of coffee consumption." Drug Metabolism and Disposition **37**(8): 1749-1758.

Stalmach, A., H. Steiling, G. Williamson and A. Crozier (2010). "Bioavailability of chlorogenic acids following acute ingestion of coffee by humans with an ileostomy." Archives of Biochemistry and Biophysics **501**(1): 98-105.

Strack, D. (1997). "Phenolic metabolism." Plant Biochemistry eds Dey, P. M., Harborne, J. B., Academic Press, 387-416.

Stuible, H. P., D. Buttner, J. Ehlting, K. Hahlbrock, *et al.* (2000). "Mutational analysis of 4-coumarate : CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes." Febs Letters **467**(1): 117-122.

Stuible, H. P. and E. Kombrink (2001). "Identification of the substrate specificity-conferring amino acid residues of 4-coumarate : coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties." Journal of Biological Chemistry **276**(29): 26893-26897.

Sullivan, M. L. (2009). "A Novel Red Clover Hydroxycinnamoyl Transferase Has Enzymatic Activities Consistent with a Role in Phaselic Acid Biosynthesis." Plant Physiology **150**(4): 1866-1879.

Sullivan, M. L. and R. Zarnowski (2010). "Red clover coumarate 3'-hydroxylase (CYP98A44) is capable of hydroxylating p-coumaroyl-shikimate but not p-coumaroyl-malate: implications for the biosynthesis of phaselic acid." Planta **231**(2): 319-328.

Terwilliger, T. C. (1999). "Reciprocal-space solvent flattening." Acta Crystallographica Section D-Biological Crystallography **55**: 1863-1871.

Terwilliger, T. C. (2000). "Maximum-likelihood density modification." Acta Crystallographica Section D-Biological Crystallography **56**: 965-972.

Terwilliger, T. C. (2003). SOLVE and RESOLVE: Automated structure solution and density modification. Macromolecular Crystallography, Pt D. **374:** 22-37.

Tholl, D. (2006). "Terpene synthases and the regulation, diversity and biological roles of terpene metabolism." Current Opinion in Plant Biology **9**(3): 297-304.

Trott, O. and A. J. Olson (2010). "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading." Journal of Computational Chemistry **31**(2): 455-461.

Ulbrich, B. and M. H. Zenk (1979). "Partial purification and properties of hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferase from higher plants." Phytochemistry **18**(6): 929-933.

Unno, H., F. Ichimaida, H. Suzuki, S. Takahashi, *et al.* (2007). "Structural and mutational studies of anthocyanin malonyltransferases establish the features of BAHD enzyme catalysis." Journal of Biological Chemistry **282**(21): 15812-15822.

Ververidis, F., E. Trantas, C. Douglas, G. Vollmer, *et al.* (2007). "Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part II: Reconstruction of multienzyme pathways in plants and microbes." Biotechnology Journal **2**(10): 1235-1249.

Villegas, R. J. A. and M. Kojima (1986). "Purification and characterization of hydroxycinnamoyl D-glucose quinate hydroxycinnamoyl transferase in the root of sweet potato, Ipomoea batatas Lam." Journal of Biological Chemistry **261**(19): 8729-8733.

Villegas, R. J. A., T. Shimokawa, H. Okuyama and M. Kojima (1987). "Purification and characterization of chlorogenic acid - chlorogenate caffeoyl transferase in sweet potato roots." Phytochemistry **26**(6): 1577-1581.

Voo, K. S., R. W. Whetten, D. M. Omalley and R. R. Sederoff (1995). "4-Coumarate-coenzyme A ligase from loblolly-pine xylem - Isolation, characterization and complementary DNA cloning." Plant Physiology **108**(1): 85-97.

Walter, T. S., C. Meier, R. Assenberg, K. F. Au, *et al.* (2006). "Lysine methylation as a routine rescue strategy for protein crystallization." Structure **14**(11): 1617-1622.

Wang, Y. and C. T. Ho (2009). "Polyphenolic Chemistry of Tea and Coffee: A Century of Progress." Journal of Agricultural and Food Chemistry **57**(18): 8109-8114.

Waters, E. R. (2003). "Molecular adaptation and the origin of land plants." Molecular Phylogenetics and Evolution **29**(3): 456-463.

Wink, M. (2003). "Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective." Phytochemistry **64**(1): 3-19.

Winkel-Shirley, B. (1999). "Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways." Physiologia Plantarum **107**(1): 142-149.

Winkel-Shirley, B. (2001). "Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology." Plant Physiology **126**(2): 485-493.

Wu, R., A. S. Reger, X. F. Lu, A. M. Gulick, *et al.* (2009). "The Mechanism of Domain Alternation in the Acyl-Adenylate Forming Ligase Superfamily Member 4-Chlorobenzoate: Coenzyme A Ligase." Biochemistry **48**(19): 4115-4125.

Yang, Q., K. Reinhard, E. Schiltz and U. Matern (1997). "Characterization and heterologous expression of hydroxycinnamoyl/ benzoyl-CoA:anthranilate N-hydroxycinnamoyl/ benzoyltransferase from elicited cell cultures of carnation, Dianthus caryophyllus L." Plant Mol. Biol. **35**(6): 777-789.

Ziouti, A., C. ElModafar, A. ElMandili, E. ElBoustani, *et al.* (1996). "Identification of the caffeoylshikimic acids in the roots of the date palm, principle fungitoxic compounds vis-a-vis Fusarium-oxysporum f sp albedinis." Journal of Phytopathology-Phytopathologische Zeitschrift **144**(4): 197-202.