Structural Characterization of the Flavonoid Enzyme Complex

Christopher David Dana

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Biology

B.S.J. Winkel, Chair
D.R. Bevan
J.M. McDowell
J.C. Sible
R.A. Walker

July 8, 2004
Blacksburg, VA

Keywords: Flavonoids, Arabidopsis, chalcone synthase, chalcone isomerase

Copyright 2004, Christopher Dana
Abstract

Flavonoid biosynthesis is an important secondary metabolic pathway in higher plants with a range of vital functions in plants and animals. This pathway has been developed as a model system for the study of multi-enzyme complexes. The goal of the work presented here was to structurally characterize a series of loss-of-function chalcone synthase (CHS) alleles and to define the molecular basis of the interaction between CHS and the second enzyme of flavonoid biosynthesis, chalcone isomerase (CHI).

CHS proteins encoded by five previously characterized alleles were characterized by homology modeling in an effort to explain the alterations in function, stability, and dimerization exhibited by these variants. Four of the encoded proteins have a single amino acid substitution and the fifth is a truncated protein resulting from a frameshift. Models for each of these proteins were generated in silico and analyzed after molecular dynamics simulations. This analysis suggested reasons for changes in catalytic ability and stability for three of the five CHS variants.

To characterize the molecular basis of the CHS-CHI interaction, a model was developed using X-ray crystallography, small-angle neutron scattering (SANS), in silico docking, molecular dynamics simulations, and yeast 2-hybrid analyses. These enzymes appear to be interacting in a manner that could facilitate the flow of intermediates from one active site to another. These experiments also identified a series of amino acids that appear to be involved in the interaction, which are currently undergoing alteration and
analysis using a yeast 2-hybrid assay to verify the authenticity of the model. The data presented herein could be used in future engineering experiments to alter pathway flux to control the levels or types of flavonoid endproducts, resulting in more nutritious plants or flowers with novel pigments.

These experiments advance the study of the structure of multi-enzyme complexes, an area that currently contains little information. As well, this is the first known use of SANS for the investigation of the architecture of metabolons. The techniques described herein could easily be applied to other systems in an effort to better understand the organization of multi-enzyme complexes and the implications of these assemblies on metabolic regulation.
Dedication

This work is dedicated to my family, especially my parents and brother, David, Donna, and William Dana, who have been my biggest cheerleaders throughout this endeavor. They have always urged me to pursue my dreams, and without them and the rest of my family, I surely would not be who or where I am today.
Acknowledgements

There are, indeed, many people to thank for their help and support during my time at Virginia Tech. First and foremost, I would like to thank my advisor, Dr. Brenda Winkel, for allowing me to add my two cents to this amazing story. She patiently listened to my sometimes hair-brained schemes on how to approach an experiment, and offered invaluable support and guidance when something didn’t go quite as expected. I also would like to express my gratitude to Dr. David Bevan, Dr. John McDowell, Dr. Jill Sible, and Dr. Rich Walker for serving on my advisory committee and offering nothing but support throughout my time here. Finally, I would like to thank all of the laboratories in both Derring Hall and Fralin Biotechnology Center that have provided advice, assistance, and facilities over the years.

This work could have not been done without the support of my collaborators. Dr. David Bevan and Jory Zmuda provided invaluable assistance and advice on the molecular dynamics simulations. Dr. Joe Noel, Dr. Joe Jez, Dr. Steph Richards, Mike Austin, and Marianne Bowman at the Salk Institute introduced me to protein crystallography. Dr. Susan Krueger at the NIST center for neutron research, introduced me to small angle neutron scattering, and with her unwavering help and support, this project has come to fruition.

I would also like to thank the members of the Winkel lab, both past and present for their support and general joviality, which kept the environment of the lab good. I thank Dr. David Saslowsky for guiding me along and putting up with my incessant questions when I first got here. Many thanks go to Dr. Mike Santos, who handled my teasing quite well, and was a wonderful mentor for the molecular biology lab. My
gratitude also goes to Daniel “The Notorious HPLC” Owens for the many conversations about science, life, and music over the years, for they have brightened my day on more than one occasion. I would also like to thank Dr. Anne Walton, Dr. Ujwala Warek, Melissa Ramirez, Ana Frazzon, and Michelle Barthet for their help and advice over the years. In the cast of Winkel lab members, the group of undergrads cannot be overlooked. Over the years, there have been many wonderful undergrads that have come through the lab. My thanks go out to all of them and everything they contributed to the project and to me, but most especially, Ashley Duda, Sandra Jabre, Anna Leung, Eric McFadden, Anna Watson, Lori Hill, and Liz Tucker.

I have been extremely fortunate to have been surrounded by an amazing group of people in my time here. First and foremost, I would like to thank James Glazebrook, director of the New River Valley Symphony at Virginia Tech for allowing me the honor of playing with his orchestra during my time here, it provided many hours of enjoyment, learning, and stress relief. I would also like to thank Dr. Robert Goebel at James Madison University for many hours of laughter from the puns, the reading suggestions and helping me to look at things beyond science. Special thanks go to Becky Abler, Tom Bunt, Mike Crandall, and Laura Moffett for their amazing friendship and support, which has meant more to the completion of this dissertation than any of you will know. Many thanks also go out to Tracey and Douglas Slotta for all of the support and kindness they have shown me over the years. Friends in “The Other Group” have provided hours of laughter and fun, without which I would have gone mad. For those friends who weren’t mentioned here, your support and friendship has not gone unnoticed and has been a source of strength for me.
Table of Contents

Chapter 1: Literature Review.................................................................1

Introduction..................................................................................2

Metabolons..................................................................................3

Static vs. Dynamic Metabolons..................................................4

Metabolic Complexes and Whole Cell Visualization..................4

Phenylpropanoid and Flavonoid Biosynthesis.........................5

Regulation of Flavonoid Biosynthesis.........................................8

Evidence for a Flavonoid Metabolon........................................9

*Arabidopsis* Flavonoid Biosynthesis as a Model System for Studying Metabolic Pathways...........................................11

Structural Biology of Flavonoid Metabolism..........................12

Chalcone Synthase.................................................................13

Chalcone Isomerase.................................................................14

Methods to Determine Interactions Between Proteins...............16

Small-Angle Neutron Scattering..............................................16

Surface Plasmon Resonance..................................................17

Project Goals/Rational.................................................................20

References Cited.........................................................................22
Chapter 2: Interpretation of Functional Effects of Mutations in *Arabidopsis* Chalcone Synthase Based on Molecular Modeling

Abstract

Introduction

Methods

Results

Discussion

Acknowledgements

References Cited

Chapter 3: Structural Characterization of the Chalcone Synthase-Chalcone Isomerase Interaction

Abstract

Introduction

Materials and Methods

Site Directed Mutagenesis of pET32a(+)

Creation of Expression Constructs

Expression and Purification of pCD1-CHS and CHI
Appendix B: SPR Analysis of Flavonoid Enzymes…………………………………………139
List of Tables and Figures

Chapter 1

Figure 1.1: Schematic of the flavonoid pathway in *Arabidopsis thaliana* ................................................................. 36

Figure 1.2: Reaction catalyzed by CHS ................................................. 37

Figure 1.3: Structure of *Medicago sativa* CHS2 .............................. 38

Figure 1.4: Close up of CHS2 active site ......................................... 39

Figure 1.5: The reaction catalyzed by CHI ....................................... 40

Figure 1.6: The structure of CHI from *Medicago sativa* ................. 41

Figure 1.7: Schematic representation of SANS ............................... 42

Figure 1.8: Schematic representation of the principle of SPR ............ 43

Chapter 2

Table 2.1: Biochemical and genetic characteristics of *Arabidopsis* CHS alleles ................................................................. 63

Table 2.2: Dimensions of the CHS proteins ................................... 64

Table 2.3: Phi-psi torsion angles within the CHS active site .......... 65

Table 2.4: Distances between CoA binding residues ...................... 66

Figure 2.1: RMSD of the backbone atoms for wild-type CHS ........ 67

Figure 2.2: Multiple sequence alignment of selected regions of a representative group of plant type III polyketide synthases .......... 69

Figure 2.3: Structure of the *Arabidopsis* wild-type CHS enzyme .... 71
Figure 2.4: Models of the *Arabidopsis* wild-type and *TT4* CHS enzymes…………………………………………………………………..73

Chapter 3

Table 3.1: Crystallographic data………………………………………………110
Table 3.2: Primers used for PCR and site-directed mutagenesis for yeast 2-hybrid constructions……………………………………………………111
Table 3.3: Yeast 2-hybrid analysis of amino acid substitutions…………112
Table 3.4: Neutron Scattering Parameters for CHS and CHI in solution…………………………………………………………………………113
Table 3.5: RMSD and distance measurements of secondary structure and residues proposed to be involved in the CHS-CHI interaction……114

Figure 3.1: Schematic of the flavonoid pathway in *Arabidopsis thaliana*……………………………………………………………………115
Figure 3.2: CHS crystal structure………………………………………………116
Figure 3.3: Residues on CHS tested for effects on interactions with other flavonoid enzymes by yeast 2-hybrid……………………………………117
Figure 3.4: Normalized SANS data…………………………………………118
Figure 3.5: Data fit of the docked CHS-CHI complex……………………120
Figure 3.6: Proposed structure of the CHS-CHI enzyme complex……121
Figure 3.7: Movements of CHI beta sheets……………………………………122
Figure 3.8: Residues proposed to participate in the CHS-CHI interaction………………………………………………………………………123
Figure 3.9: Close up of the interaction interface between CHS and CHI……………………………………………………………………...124

Figure 3.10: Multiple sequence alignments of *Arabidopsis* CHS and CHI protein sequences……………………………………………………….125

Appendix A

Figure A.1: Homology models of *Arabidopsis* Flavonoid Enzymes…..137

Appendix B

Figure B.1: Representative SPR output………………………………...141

*Curriculum vitae*……………………………………………………………………… ..143
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANR</td>
<td>anthocyanidin reductase</td>
</tr>
<tr>
<td>ANS</td>
<td>anthocyanidin synthase</td>
</tr>
<tr>
<td>BD</td>
<td>binding domain</td>
</tr>
<tr>
<td>C4H</td>
<td>cinnamate 4-hydroxylase</td>
</tr>
<tr>
<td>CHI</td>
<td>chalcone isomerase</td>
</tr>
<tr>
<td>CHS</td>
<td>chalcone synthase</td>
</tr>
<tr>
<td>DFR</td>
<td>dihydroflavonol 4-reductase</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>F3H</td>
<td>flavanone 3-hydroxylase</td>
</tr>
<tr>
<td>F3’H</td>
<td>flavonoid 3’-hydroxylase</td>
</tr>
<tr>
<td>FLS</td>
<td>flavonol synthase</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>I(0)</td>
<td>forward scattering intensity</td>
</tr>
<tr>
<td>IGF</td>
<td>insulin growth factor</td>
</tr>
<tr>
<td>OAS-TL</td>
<td>O-acetylserine (thiol) lyase</td>
</tr>
<tr>
<td>PAL</td>
<td>phenylalanine ammonia lyase</td>
</tr>
<tr>
<td>PKS</td>
<td>polyketide synthase</td>
</tr>
<tr>
<td>PS</td>
<td>pyrone synthase</td>
</tr>
<tr>
<td>Rg</td>
<td>radius of gyration</td>
</tr>
<tr>
<td>rER</td>
<td>rough endoplasmic reticulum</td>
</tr>
</tbody>
</table>
**RMSD** – root mean square deviation

**SANS** – small-angle neutron scattering

**SAT** – serine acetyltransferase

**SPR** – surface plasmon resonance

**STS** – stilbene synthase

**TA** – transactivation domain

**TCA** – tricarboxylic acid

*tt – transparent testa*

**TRX** – thioredoxin
Chapter 1

Literature Review
**Introduction**

A driving question in the study of biochemistry is how metabolic pathways are organized within the cell. Evidence for organized metabolism dates back to the 1950s when David Green first isolated the enzymes of the tricarboxylic acid (TCA) cycle as a complex (Green, 1958). Subsequent ultracentrifugation experiments with *Neurospora* (Zalokar, 1960) and *Euglena* (Kempner and Miller, 1968) indicated that the aqueous cytoplasm contains little or no free floating proteins or enzymes and that the interior of the cell is highly organized. These seminal experiments revolutionized the concept of cellular organization and laid the groundwork for future experiments aimed at defining metabolic organization within the cell. The advent of interactome analyses (Ito et al., 2001; Li et al., 2004; Sanchez et al., 1999) provided new insights into the extent to which cellular proteins are able to interact. As more examples of multienzyme complexes, or metabolons, were identified, the notion that cellular metabolism is arranged as a highly organized network of metabolic complexes gained widespread acceptance.

In our laboratory, we are investigating multi-enzyme complex formation in flavonoid biosynthesis. Flavonoids are secondary metabolites that are important both for the plant and for organisms that use plants for nutrition. Helen Stafford first suggested that flavonoid metabolism is organized as one or more enzyme complexes (Stafford, 1974a; Stafford, 1974b), and later experiments provided evidence for this (for example, Burbulis and Winkel-Shirley, 1999; Fritsch and Grisebach, 1975; Hrazdina and Jensen, 1992). One of our goals is to determine the molecular basis of the interaction between the first two enzymes in this pathway, chalcone synthase (CHS) and chalcone isomerase (CHI), in the hope that this information will lead to a better understanding of the architecture of enzyme complexes in this and other systems.
Metabolons

An enzyme complex is a group of two or more enzymes that interact to catalyze related metabolic reactions, allowing for higher catalytic efficiency than if the enzymes were not associated. Metabolic pathways can obtain such an efficiency by controlling the flux of substrate and product thru channels between proteins (Ovádi and Srere, 2000). Metabolic channeling also protects intermediates that are unstable or rare and maintains concentration gradients within the complex (Mathews, 1993). The mechanics behind metabolic channeling include the covalent binding of intermediates to the active site, a physical barrier acting as a tunnel, direct site-to-site transfer, transfer in an unstirred layer, and electrostatic channels (reviewed in Ovádi and Srere, 2000).

The TCA cycle represents one of the best-characterized multienzyme complexes. Since David Green isolated the TCA cycle complex, the functional and structural properties of the TCA cycle complex have been well characterized in the laboratory of the late Paul Srere (Robinson et al., 1987; Srere, 1985; Srere, 1987; Sumegi et al., 1993; Vélot et al., 1997; Vélot and Srere, 2000). Srere’s work eventually led to the first three dimensional model of the citrate synthase-malate dehydrogenase enzyme complex though in silico docking of the crystal structures (Ovádi and Srere, 2000). There are now many more examples of this level of organization, and it has been suggested that all metabolic pathways organize in this way (Beeckmans et al., 1994; Beeckmans et al., 1989; Gontero et al., 1988; Robinson et al., 1987).
Static vs. Dynamic Metabolons

Enzyme complexes have been classified as being either static or dynamic (Welch, 1977). Generally, static complexes are very stable over time and the interactions between the enzymes are so strong that upon purification, the enzymes remain associated because little energy is needed to maintain the integrity of the complex (Ovádi and Srere, 2000). In contrast, a dynamic enzyme complex is held together by relatively weak forces and is dependent upon constant energy dissipation (Welch, 1977) and is composed of multiple branch points where many enzymes could compete for one substrate (Ovádi and Srere, 2000). It is generally thought that static enzyme systems possess no branches, whereas dynamic enzyme systems may contain many branches (Ovádi and Srere, 2000). Examples of static enzyme complexes are the $\alpha$-keto acid dehydrogenase complex, the proteasome, and proteins involved in DNA, RNA, and protein synthesis (reviewed in Ovádi and Srere, 2000; Winkel, 2004). Two examples of dynamic enzyme systems are the TCA cycle complex (Robinson et al., 1987) and enzymes involved in carbamoyl phosphate metabolism (Massant and Glansdorff, 2004).

Metabolic Complexes and Whole Cell Visualization

There is a plethora of in vitro and in vivo evidence for protein-protein interactions and macromolecular crowding within the cell. It has only recently been possible to visualize this directly; however, there have been few examples of whole cell visualization showing enzyme complexes and, ultimately, macromolecular crowding. An initial representation of
macromolecular crowding came from David Goodsell’s drawings of *E.coli* ribosomes, DNA, and RNA to shape and scale and illustrated that the interior of the cell is crowded (Goodsell, 1991).

Electron tomography, an established technique, has recently gained new popularity for visualization of macromolecular organization within cells. By using this technique, Hoppe et al. (1974) were able to describe a complex of enzymes involved in fatty acid biosynthesis. As computational power increased, electron tomography could be used to visualize organelles and later whole cells. Marsh, et. al (2001) recently used this technology to show that the environment around the Golgi apparatus in pancreatic β cells is extremely congested. Elegant work performed in Wolfgang Baumeister’s lab using electron tomographic imaging of whole *Dictyostelium* cells showed that the cell is crowded with actin networks, membranes, and macromolecular assemblies (Medalia et al., 2002).

**Phenylpropanoid and Flavonoid Biosynthesis**

The phenylpropanoid pathway converts phenylalanine into a variety of secondary metabolites including lignins, sinapate esters, and stilbenes (Figure 1.1) that have a multitude of *in planta* functions, such as reproduction, defense against fungal infections, and signaling in pathogenesis and symbiosis (reviewed in Winkel-Shirley, 2001). Phenylpropanoid biosynthesis also feeds into the flavonoid pathway, which converts coumaroyl-CoA and malonyl-CoA into flavonoids, isoflavonoids, flavonol glycosides, anthocyanidins, and condensed tannins (Figure 1.1). In plants, these compounds have been postulated to function in reproduction, defense against the biotic and abiotic environment, and signaling in pathogenesis and symbiosis. These metabolites, along with sinapate esters derived from sinapic acid (Landry et al., 1995; Li et al.,
1993), protect the plant by absorbing light in the damaging UV-B spectrum (between 280 and 320 nm), acting as natural sunscreens for the plant. Flavonoids have been shown to accumulate at the epidermal layer of leaf tissue, pollen where UV-B exposure is the highest as well as in the epidermis of fruit (Chapple et al., 1992; Day, 1993; Jordan et al., 1998; Landry et al., 1995; Solovchenko and Schmitz-Eiberger, 2003; Strack et al., 1985).

Other *in planta* activities for flavonoids include defense against pathogen attack, signaling molecules, and regulators of auxin transport. Defense-related compounds, called phytoalexins, are produced when the plant comes under attack from pathogenic fungi or bacteria and, in the case of bacteria, are potent anti-microbial agents (Liu and Dixon, 2001). Certain flavonoids such as naringenin, quercetin, kaempferol, and rutin, are constitutively produced and thought to have some anti-microbial properties (Paiva, 2000). However, when a plant comes under attack from fungal pathogens, some species produce glyceollin and medicarpin (both isoflavonoid pterocarbons, and are also potent anti-microbial agents) in the foliage surrounding the infection site (Paiva, 2000). In legumes, a specific class of flavonoids, the isoflavonoids, act as signaling compounds that initiate the symbiosis of the nitrogen-fixing rhizobia (Hirsch et al., 2001). Flavonoids have also been implicated in the regulation of auxin transport (Winkel-Shirley, 2002). Recent studies showed that in *Arabidopsis* plants that are unable to produce naringenin, auxin distribution resembles that of a plant overproducing auxin (Brown et al., 2001; Buer and Muday, 2004; Murphy et al., 2000). Normal auxin distribution is restored upon the application of exogenous naringenin.

Flavonoids are also responsible for pigmentation in flower, fruit, and vegetative tissue (especially blue, red, and purple coloration), functioning as attractants for pollinators and seed dispersers. These pigments have been the phenotype observed in some important scientific
breakthroughs such as Mendel’s elucidation of genetics through the study of phenotypes in pea (*Pisum sativa*), one being flower coloration. Altering the pigmentation of flowers is of intense interest to the horticultural industry (Forkmann and Martens, 2001). The first example of metabolic engineering of flower pigmentation came involved introduction of the maize gene encoding dihydroflavonol 4-reductase (DFR) into *Petunia*, yielding flowers with a novel orange color (Meyer et al., 1987). Recently, the color of *Phalenopsis* flowers were altered by a transient transfection method using a putative flavonoid-3’,5’-hydroxylase from the same organism (Su and Hsu, 2003).

Flavonoids exhibit a host of nutritional and medicinal properties including anti-oxidant activities, which are useful in reducing the risk of cardiovascular disease and some cancers (Dixon and Steele, 1999). Other beneficial properties of flavonoids include prevention of bacterial infections, enzyme inhibition, and free radical scavenging (reviewed in Havsteen, 2002). Daidzin, an isoflavonoid, is a potent anti-dipsotropic agent that has been used as a Chinese herbal remedy for alcohol dependence for over a millennium (Keung, 2003). In Asian societies, the low incidence of cardiovascular disease and cancer has been, in part, linked to the isoflavones produced in soybean, which is regularly consumed (Sarkar and Li, 2003). Dark chocolate, interestingly, has the highest concentration of anti-oxidant flavonoids among commonly-consumed foods (Lotito and Fraga, 2000; Serafini et al., 2003).

Creating plants with elevated levels of antioxidant flavonoids is a focus of recent metabolic engineering projects. Work done at Unilever Research in the United Kingdom and Plant Research International in the Netherlands demonstrated that expression of petunia CHI in tomato plants induces a 78-fold increase of fruit peel flavonols without any change in visible phenotype (Muir et al., 2001). Interestingly, when the fruit was processed, a majority of the
flavonols were retained. Yu et al. (2003) were able to increase the levels of isoflavones in soybean seed by suppressing flavanone 3-hydroxylase (F3H), the entry point to flavonol, condensed tannin, and anthocyanin biosynthesis, and activating expression of two transcription factors required for isoflavonoid biosynthesis. These metabolic engineering efforts could create edible plants that supplement diets with anti-oxidants to help protect against cancer and other diseases.

**Regulation of Flavonoid Biosynthesis**

Flavonoid biosynthesis is tightly controlled by a number of internal and external factors. It can be modulated by development, exposure to UV-B radiation, high intensity and blue light, and both low and high temperatures (reviewed in Winkel-Shirley, 2002). The molecular basis of this regulation has been intensely studied in parsley, petunia, snapdragon, maize, and Arabidopsis and numerous transcriptional regulators have been identified and characterized (reviewed in Marles et al., 2003; Sainz et al., 1997; Weisshaar and Jenkins, 1998).

In maize, two sets of transcription factors, the R family of regulatory proteins and the C1 myb-domain-containing transcription factors, appear to control the transcription of the entire flavonoid pathway (Taylor and Briggs, 1990). However, other plants that contain homologues of the maize transcription factors do not exhibit the same breadth of control over the pathway. For example, an1, 2, 4, and 11 from petunia and delila from snapdragon all appear to be homologues of the R family of genes in maize, but control the transcription of a more defined subset of genes (Martin and Gerats, 1993; Spelt et al., 2000).
Much is also known about the regulation of flavonoid biosynthesis in *Arabidopsis*. Some of the transcription factors that control flavonoid biosynthesis in *Arabidopsis* have been identified by characterization of the transparent testa (tt) mutants. For example, *TTG1*, *TTG2*, *TT2*, *TT8*, and *TT16* encode transcription factors that control DFR and anthocyanidin reductase (ANR) expression in the seed coat, and have been proposed to cooperate in controlling the accumulation of proanthocyanidins in developing seed coats (Johnson et al., 2002; Nesi et al., 2000; Nesi et al., 2002; Nesi et al., 2001; Walker et al., 1999). However, there are transcription factors that are ubiquitous in their ability to regulate the expression of genes. ICX1 is such an example which has been implicated in the negative regulation of both light and non-light responses (Wade et al., 2003).

**Evidence for a Flavonoid Metabolon**

Helen Stafford’s original hypothesis that phenylpropanoid metabolism is organized as one or more enzyme complexes (Stafford, 1974a; Stafford, 1974b) has been supported by experiments performed in many laboratories, including our own (reviewed in Winkel, 2004; Winkel-Shirley, 1999). Through microsomal fractionation experiments, Czichi and Kindl (1977) demonstrated a close association of the first two enzymes of general phenypropanoid metabolism [phenylalanine ammonia lyase (PAL) and cinnamate 4-hydroxylase (C4H)] with the endoplasmic reticulum. Later evidence for channeling between these enzymes was obtained in experiments where $^3$H-L-phenylalanine was fed to tobacco cell cultures. Little $^3$H-trans-cinnamic acid (the product of PAL) was identified in the cells, suggesting that this product was rapidly channeled
through the general phenylpropanoid pathway (Rasmussen and Dixon, 1999). Evidence for a multi-enzyme complex and channeling has also been obtained for the first two enzymes of isoflavonoid biosynthesis through *in planta* localization and feeding experiments (Liu and Dixon, 2001).

The first experimental evidence for protein-protein interactions within the flavonoid pathway came in cell fractionation experiments that demonstrated that the enzymes are located on microsomal or endoplasmic reticulum (ER) membranes (Fritsch and Grisebach, 1975). Sucrose density gradients of *Hippeastrum* petals demonstrated activity of PAL, C4H, CHS, and UDP-glucose flavonoid glucosyltransferase in fractions corresponding to ER (Wagner and Hrazdina, 1984). Fractionation experiments in buckwheat showed that CHS activity corresponded with the rough ER (rER) and C4H, a P450 hydroxylase and purported membrane anchor for the pathway (Hrazdina and Wagner, 1985). Through immunolocalization, Hrazdina’s laboratory was able to show that CHS associated with the cytoplasmic face of the rER (Wagner and Hrazdina, 1984). Building on these results, affinity chromatography, co-immunoprecipitation, and yeast two-hybrid experiments provided the first direct evidence for interactions within flavonoid metabolism (Burbulis and Winkel-Shirley, 1999). Further evidence was obtained from immunolocalization assays in fixed cells (Saslowsky and Winkel-Shirley, 2001). These assays showed that CHS and CHI co-localize and were consistent with a previously-proposed role for flavonoid 3’-hydroxylase (F3’H) as a membrane anchor for the complex (Saslowsky and Winkel-Shirley, 2001; Stafford, 1990). It has also been demonstrated that by expressing a phage-derived antibody to CHI in transgenic plants, the activity of CHI is altered, as is the overall flavonoid composition of the plant (Santos et al., 2004). This antibody could prevent CHI from assembling with other flavonoid enzymes. These recent findings have
changed the overall concept of the flavonoid metabolon from one of a linear arrangement of enzymes (Hrazdina, 1992; Stafford, 1974a; Stafford, 1974b) to a multi-enzyme complex with contacts between multiple proteins (Burbulis and Winkel-Shirley, 1999). The next challenge is to determine the molecular basis of these interactions. A greater understanding of how enzyme complexes form and how this assembly is regulated may facilitate the engineering of plants with enhanced agronomic, nutritional or medicinal characteristics.

**Arabidopsis Flavonoid Biosynthesis as a Model System for Studying Metabolic Pathways**

*Arabidopsis* has been developed as a model system for the study of plant genetics, biochemistry and molecular biology because of its ease of growth, short generation time, utility as a genetic system, and the availability of an efficient and simple transformation system (Page and Grossniklaus, 2002). In our laboratory, *Arabidopsis* is being used to provide a better insight into the flavonoid biosynthetic machinery and enzyme complexes as a whole. The *Arabidopsis* flavonoid biosynthetic pathway is very well characterized and mutations in flavonoid genes can be easily identified based on the seed coat color (Shirley et al., 1995). The use of the *tt* mutants has provided important new knowledge on the function of flavonoids for the plant. For example, the *tt4(2YY6)* mutant (which carries a null allele for CHS) was used to show that, unlike the situation in maize and petunia, flavonoids do not play a role in male fertility in *Arabidopsis* (Burbulis et al., 1996b). Another interesting result is the determination that sinapate esters and flavonoids play a critical function as sunscreens (Landry et al., 1995; Li et al., 1993). These mutants have also played an important part in characterizing the organization of the pathway as a
multi-enzyme complex (Burbulis et al., 1996a; Burbulis and Winkel-Shirley, 1999; Saslowsky and Winkel-Shirley, 2001).

Genes encoding eight flavonoid enzymes have been isolated in *Arabidopsis*. These enzymes are CHS, CHI, F3H, flavonol synthase (FLS1-6), F3’H, DFR, ANS (also known as leucoanthocyanidin dioxygenase (LDOX)), and ANR (also known as BANYULS (BAN)) (Burbulis et al., 1996a; Pelletier et al., 1999; Pelletier et al., 1997; Pelletier and Shirley, 1996; Saslowsky and Winkel-Shirley, 2001; Shirley and Goodman, 1993; Xie de et al., 2004). All of these enzymes are encoded by single copy genes with the exception of FLS, which is present in six copies, at most three of which are thought to be catalytically active (Owens, Walton, and Winkel, unpublished data). Five of these enzymes, CHS, CHI, F3H, FLS, ANS, and ANR, have been overexpressed in *E. coli* and used to raise polyclonal antibodies (Cain et al., 1997; Pelletier et al., 1997; Xie de et al., 2004).

**Structural Biology of Flavonoid Metabolism**

Understanding the structure and function of a protein on an atomic level can lead to a better understanding the catalytic properties of a protein. Structures have been solved for three flavonoid enzymes over the past several years, CHS2 and CHI from *Medicago sativa*, and ANS from *Arabidopsis* (Ferrer et al., 1999; Jez et al., 2000b; Turnbull et al., 2001; Wilmouth et al., 2002). As detailed below, the structures of these enzymes have allowed researchers to better understand how these enzymes function catalytically and also investigate ways to alter their substrate specificity and understand their evolution.
Chalcone Synthase: CHS catalyzes the first committed reaction in flavonoid biosynthesis in which one molecule of 4-coumaroyl-CoA and 3 molecules of malonyl-CoA are used to create naringenin chalcone, the starter molecule for the ensuing steps in flavonoid metabolism (Figure 1.2). CHS is a member of the plant type III polyketide synthase family, which encompasses all CHS enzymes as well as a variety of ‘CHS-like’ enzymes that catalyze similar polyketide reactions in plants (Austin and Noel, 2003). The structure (Figure 1.3) and the reaction mechanism of CHS2 from *Medicago sativa* was solved in the laboratory of Joe Noel at the Salk Institute (Ferrer et al., 1999; Jez et al., 2000a; Jez et al., 2001; Jez et al., 2000d; Jez and Noel, 2000). The CHS structure is made up of an aßaßa core motif that is conserved among all type III polyketide synthases as well as enzymes containing a thiolase fold, such as ketoacyl thiolase (Austin and Noel, 2003). CHS bears a striking resemblance, both in sequence and structure, to thiolase enzymes such as ketoacyl synthases and thiolases, which provides evidence for the hypothesis that CHS evolved from an enzyme of fatty acid biosynthesis as first suggested by Verwoert, et al. (1992).

Catalytically-active CHS is a bi-lobed homodimer containing two distinct active sites. The CHS homodimer is held together by identical six-residue loops in each opposing monomer that contribute a methionine to the wall of the active site of the neighboring subunit. The CHS active site cavity is buried within the structure of the protein and is accessible only through a 16 Å long channel. However, the CHS crystal structure has shown that this channel is actually too small for the product to leave CHS, suggesting that it is a dynamic structure that can change dimensions to accommodate the entry or exit of different substrates and products (Ferrer et al., 1999).
The CHS reaction starts by the loading of coumaroyl-CoA onto the active site cysteine (Fig. 4), which is conserved among all thiolase-fold enzymes (Austin and Noel, 2003). This allows the CoA moiety to be released through a decarboxylation reaction. Malonyl-CoA is then freed of its CoA moiety by a second decarboxylation reaction and attached to the coumarate bound in the active site. Two more malonyl-CoA units are added until a linear tetraketide chain is created, which is circularized to form the chalcone product through an aromatase-like mechanism that is not yet completely understood. The overall reaction is carried out by a ‘catalytic triad’ of conserved residues that play critical roles in the decarboxylation and condensation reactions required for catalysis (Austin and Noel, 2003). The first of these residues, the active site cysteine, was initially identified by site directed mutagenesis studies performed in Gerhard Schröder’s group (Lanz et al., 1991). This cysteine is conserved across all thiolase-fold enzymes and situated in a loop between two buried amphipathic α-helices within the central core of the enzyme (Austin and Noel, 2003).

Chalcone Isomerase: CHI catalyzes the second step in flavonoid biosynthesis, closing the center ring (Figure 1.5) and creating the basic flavonoid backbone. The CHI reaction can occur spontaneously, however there is a $10^7$ increase in reaction rate when the reaction is performed by CHI (Bednar and Hadcock, 1988). There is also some evidence that CHI is post-translationally modified, possibly through a thiol linkage, which could anchor CHI to the ER or aid in the association of CHI with other enzymes in the complex (Burbulis and Winkel-Shirley, 1999).

The structure (Figure 1.6) and catalytic mechanism of CHI from Medicago sativa were also solved in Joe Noel’s laboratory (Jez et al., 2000c; Jez et al., 2002). The structure resembles an upside down flower bouquet with an open β sandwich fold (Jez et al., 2000c). The overall
CHI fold is maintained in this enzyme across higher plant species. The fold is also predicted to be present in some bacterial (gamma proteobacteria) and fungal proteins (Gensheimer and Mushegian, 2004), dispelling the initial notion that this fold is unique to CHI. The functions of these ‘CHI-like’ proteins are unknown, however.

The cyclization reaction catalyzed by CHI is facilitated by an initial binding of the tetrahydroxychalcone through a hydrogen bond between the hydroxyl group at the 7 position of the substrate to the hydroxyl on threonine 190, which has also been suggested to function in determining substrate preference (Jez et al., 2002). A further hydrogen bonding network is required in order to catalyze the reaction. One water molecule, in particular, is hydrogen bonded to tyrosine 106, which is thought to activate the water molecule and is hydrogen bonded to the substrate. Upon activation, the water molecule attacks the carbon-carbon double bond, thus reducing it and allowing for the formation of a bond to close the center ring on the substrate (Jez et al., 2002). CHI appears to have evolved to be stereospecific in the reaction it catalyzes. *Medicago* CHI prefers the S-isomer of tetrahydroxychalcone over the R-isomer by an apparent 100,000:1 ratio (Bednar and Hadcock, 1988). In order for CHI to efficiently catalyze the conversion of the R-isomer of tetrahydroxychalcone, there would have to be significant changes in the active site structure in order to accommodate the additional stereochemistry (Jez et al., 2000c). This stereospecificity raises the questions, does CHS always produce the S-isomer and, if not, is there a use for the R-isomer in plant secondary metabolism? Unfortunately, the answers are not yet known.
Methods to Determine Interactions Between Proteins

When investigating interactions between proteins, it is helpful to have multiple techniques available. Many of the original studies that demonstrated protein-protein interactions relied on techniques such as precipitation of protein aggregates, chromatography, and ultracentrifugation (Ovádi and Srere, 2000). More recent techniques such as fluorescence resonance energy transfer (Day et al., 2001), affinity gel electrophoresis (Beeckmans et al., 1989), and yeast two-hybrid analysis (Ito et al., 2001) allow for the identification of two or more interacting proteins. Techniques such as surface plasmon resonance refractometry (Rich and Myszka, 2000), analytical ultracentrifugation (Lebowitz et al., 2002), and isothermal titration calorimetry (Weber and Salemme, 2003) are capable of giving more detailed information such as kinetic binding constants, reaction kinetics, and the thermodynamics of interactions. However, these techniques cannot approach the atomic level of resolution that is needed for determining the structure of a complex. The best ways to gain an appreciation of the structure of an enzyme-complex on an atomic level are through co-crystallography of the proteins, small angle neutron scattering (SANS) (Koch et al., 2003), and in silico protein-protein docking algorithms (Halperin et al., 2002). For the purpose of this introduction, two of the above techniques, SANS and surface plasmon resonance refractometry, will be discussed.

Small-Angle Neutron Scattering: SANS is a technique that uses neutrons to probe three-dimensional structure of biological macromolecules in a non-destructive manner. Whereas X-ray diffraction is typically used to study atomic-level structures (under 10 Å), SANS is generally used to investigate larger structures (10 to 1000 Å) (Koch et al., 2003). Briefly, a sample is
exposed to a monochromatic, cold neutron beam generated by either a nuclear reactor or a spallation source (details are shown in Fig. 7). The neutrons are scattered upon contact with the sample due to interactions with the individual atoms. Neutron scattering intensities are recorded by a two-dimensional detector, which can be moved in the both the y- and z- axis with respect to the sample. Typically, neutrons are scattered through small angles (less than 1 degree) when compared to traditional diffraction angles (more than 10 degrees). The scattering pattern is then analyzed computationally to determine the overall size and shape of the molecules in solution.

The data obtained from these measurements do not provide a high resolution structure like traditional crystallography, but rather a low resolution solution structure into which high resolution crystal structures can be fit in order to develop models of macromolecular assemblies. SANS can also be used to determine regions of proteins that are highly mobile in crystal structures. SANS has been used in a number of biological applications, contributing to determination of the structure of the 30S ribosomal subunit (Capel et al., 1987), elucidating better ways to crystallize a protein by determining the amount of protein that is aggregated (Velev et al., 1998), and characterizing the structure of the GroEL/ES chaperonin complex (Krueger et al., 2003). Beyond biological applications, SANS has been used in the study of polymers, complex fluids, materials science and condensed matter (Koch et al., 2003).

**Surface Plasmon Resonance:** Another powerful technique that can be used to determine molecular interactions between proteins is SPR refractometry (or spectroscopy). This approach uses monochromatic light to detect interactions between different molecules via a change in refractive index that occurs during complex formation or dissociation (Rich and Myszka, 2000). Briefly, protein is first immobilized onto a gold-coated glass slide that has been coupled with oil
to a sapphire crystal through which monochromatic light is passed (Fig. 8). The slide (or chip) has a defined surface chemistry through a thiol linkage onto the gold. Two of the more common surface chemistries currently being used are carboxy-methyl dextran self assembling monolayers and metal affinity chemistries (Rich and Myszka, 2000). The former involves covalently linking the protein (typically, any surface amine groups are utilized) to the surface chemistry whereas the latter generally involves non-covalent coupling via a hexa-histadine tag. Once protein is bound onto the slide, unbound protein is washed off, and solution containing a second protein is passed over the slide. Interaction between the two is detected by a change in refractive angle which is detected by a photo diode array and reported on the monitor.

SPR can be used not only to detect protein interactions, but also to measure kinetics, affinity constants, the thermodynamics of an interaction, and, in some cases, to determine the stoichiometry and mechanism of interaction (Morton and Myszka, 1998; Myszka et al., 1998). Recent advances have also coupled SPR with mass spectrometry to identify interacting proteins as well as determining interaction domains (Nedelkov and Nelson, 2003a). Various software packages exist to aid in analyzing the results of SPR experiments, such as CLAMP99 (Myszka and Morton, 1998), which allows for the determination of association and dissociation values for an interaction.

One of many examples of current research using SPR is with the human immunodeficiency virus (HIV). Areas of the HIV life cycle that have been intensely studied with SPR are binding, mapping of epitopes for antibody development, and effects of inhibitory drugs on viral machinery. This work has given a unique insight into how HIV functions (reviewed in Rich and Myszka, 2003). These studies focused on antibody-antigen interactions
with the hopes of finding a magic bullet that will affect the ability of HIV to infect and replicate in the human host.

An example of biosynthetic enzymes being studied by SPR is the cysteine synthase complex. Previous evidence from yeast two-hybrid analysis indicated that serine acetyltransferase (SAT) and $O$-acetylserine (thiol) lyase (OAS-TL) interact to form a complex that is involved in cysteine biosynthesis (Bogdanova and Hell, 1997). These initial studies also defined some of the interaction domains between SAT and OAS-TL by means of truncation series. SPR was subsequently used to determine the binding kinetics of these two enzymes when they are activated in the presence of sulfur (Berkowitz et al., 2002).

A recent innovation has been the addition of mass spectroscopy at the conclusion of the SPR assay (Nedelkov and Nelson, 2001). The technique of SPR is carried out in the same manner as in conventional applications, except that the immobilized protein has to be covalently attached to the slide to inhibit leaching. Once the SPR experiment is complete, interacting proteins are eluted, leaving the initial protein bound to the slide, and these proteins are subjected to matrix-assisted laser desorption-ionization time-of-flight (MALDI-TOF) MS (Nedelkov and Nelson, 2003b). This technique allows for the effective identification of previously-unknown interacting proteins. An example of the application of SPR-MS is the identification of the protein complex makeup of insulin-like growth factors (IGF-1 and -2) from human plasma (Nedelkov et al., 2003). In separate experiments, antibodies against IGF-1 or -2 were used to immobilize IGF-1 and -2 along with any other proteins interacting with them from human plasma, which were later identified by MS. It was determined that both growth factors are present in human plasma, suggesting that IGF-1, -2 and a previously unknown truncated form of IGF-2 make up a protein complex \textit{in vivo} (Nedelkov et al., 2003).
**Project Goals/Rationale**

There has been a variety of experimental evidence for interactions between enzymes involved in flavonoid metabolism, both *in vitro* and *in vivo*. These data provide a broad understanding of interactions within the metabolon, but no details regarding the molecular basis of these interactions. It would seem that the next logical step would be to investigate this problem on an atomic scale, i.e., determining how the enzymes physically interact with each other and identifying the specific residues and atoms involved. Data from such experiments would provide more information, not only on the organization of the flavonoid metabolon, but also for metabolic organization in general. These insights could be used to create stronger (or weaker) interactions to modulate pathway flux and engineer plants with higher levels of flavonoids important for flower pigments, (e.g., create a blue rose), or as antioxidants.

This project was divided into two major objectives. First, the structural implications of amino acid substitutions were investigated by molecular dynamics simulations. The second part of this project studied the functional and structural components of flavonoid biosynthesis, specifically the interactions between CHS, CHI, and F3H. X-ray crystallography and molecular modeling were utilized to determine or deduce the structure of proteins involved in flavonoid biosynthesis. Amino acid substitutions were introduced on the surface of CHS to test for effects on the interaction of CHS with CHI in a yeast two-hybrid assay. SANS, which had not previously been used for the study of metabolic complexes, was used to generate a structural model of the interacting proteins. Next, computational docking studies were performed to develop a model that fit the SANS data. The results from this project provide new insights into
the self assembly of enzymes involved in flavonoid biosynthesis, with implications for the organization of enzyme complexes in general.
References Cited:


Figure 1.1: Schematic of the flavonoid pathway in Arabidopsis thaliana. The pathway creates a diverse set of compounds used in UV protection, signaling, defense, and pigmentation. CHS represents that first committed step in the pathway, where 4-coumaroyl-CoA (from phenylpropanoid biosynthesis) and 3 units of malonyl-CoA (from fatty acid metabolism) are used as substrates. Structures in the figure represent the three enzymes in the pathway that crystal structures have been solved for, CHS (pdb id:1BI5) and CHI (pdb id:1EYQ) (both from Medicago sativa), and anthocyanidin synthase (ANS) (pdb id:1GP6) (from Arabidopsis thaliana). Structures in orange are homology models of dihydroflavonol 4-reductase (DFR), flavanone 3-hydroxylase (F3H), flavonoid 3’ hydroxylase (F3’H), and flavonol synthase (FLS). Other enzyme names are abbreviated as follows: anthocyanidin reductase (ANR), cinnamate 4-hydroxylase (C4H), p-coumarate:CoA ligase (4CL), leucoanthocyanidin reductase (LAR), phenylalanine ammonia-lyase (PAL). Arrows labeled in grey indicate branches of the pathway not present in Arabidopsis.
Figure 1.2. Reaction catalyzed by CHS. CHS utilizes 4-coumaroyl-CoA and 3 units of malonyl-CoA in a Claisen condensation reaction to create tetrahydroxychalcone. The atoms on tetrahydroxychalcone which are labeled in black come from 4-coumaroyl-CoA and the atoms in red come from malonyl-CoA.
Figure 1.3. Structure of *Medicago sativa* CHS2 (Ferrer et al., 1999). The CHS protein is composed of two identical monomers (illustrated in purple and yellow). The red atoms indicate the methionine that is utilized as part of the active site wall by the partner monomer. The arrow indicates the substrate/product channel leading to the buried active site. The structure was generated using SwissPDB Viewer (Guex and Peitsch, 1997) and rendered using POV-Ray 3.1 (www.povray.org).
Figure 1.4. Close up of CHS2 active site (Ferrer et al., 1999). F215 and F265 are the ‘gatekeeper’ residues that are proposed to control the entry of substrate. H303 and N336 are involved with the binding of substrate and removal of the CoA moieties while C164 holds the substrate during the reaction.
Figure 1.5. The reaction catalyzed by CHI. The C ring of tetrahydroxychalcone is closed in a cyclization reaction to create (2S)-naringenin, which serves as the basic backbone for the downstream flavonoid metabolic reactions.
Figure 1.6. The structure of CHI from *Medicago sativa* crystallized with its product (Jez et al., 2000c). The structure maintains an overall open β-sandwich fold. The arrow indicates the location of the product, (2S)-Naringenin.
Figure 1.7. Schematic representation of SANS. Cold neutrons are emitted from the neutron reactor or spallation source and are then selected for the appropriate wavelength by the velocity selector. A series of lenses then focus the monochromatic neutron beam before passage through the sample. Upon hitting the sample, scattered neutrons are recorded by the 2-D detector and turned into electrical impulses, which are then transmitted to a data collection computer. Non-scattered neutrons are blocked by a beamstop made of neutron-dense material. (adapted from Koch et al., 2003).
Figure 1.8: Schematic representation of the principle of SPR. Polarized light at 760 nm is directed through the sapphire prism onto a glass slide with evaporated Ti and Au and a surface chemistry has been coupled. The change in observed resonance angle (?) is recorded by a photodetector when protein binds to the surface chemistry (adapted from Lahiri et al., 1999).
Chapter 2

Interpretation of functional effects of mutations in *Arabidopsis* chalcone synthase based on molecular modeling

Christopher D. Dana¹, David R. Bevan² and Brenda S.J. Winkel¹

¹Department of Biology and Fralin Biotechnology Center and ²Department of Biochemistry, Virginia Tech

Manuscript to be submitted to Journal of Molecular Modeling

Key Words: molecular dynamics, allelic series, chalcone synthase (CHS), transparent testa (*tt*)
Abstract

Chalcone synthase (CHS) catalyzes the first committed step in flavonoid biosynthesis, a major pathway of plant secondary metabolism. An allelic series for the Arabidopsis CHS locus, tt4, was previously characterized at the gene, protein, and end product levels (Saslowsky et al., 2000). The substitutions identified in each of these alleles were used to generate models for five of the tt4 proteins based on the Arabidopsis CHS crystal structure (Chapter 3) in order to determine how mutations may affect protein structure and function. tt4(85) contains a substitution that apparently disrupts catalytic activity due a change in the CoA binding face geometry. The tt4(UV113) allele disrupts enzymatic activity while maintaining a stable protein, which could be caused by a stiffening of the loop containing the critical active site residue, C169. In contrast, the tt4(38G1R) mutation results in a truncated protein that is both catalytically inactive, does not dimerize, and accumulates at reduced levels, presumably due to the loss of three beta sheets. However, no structural basis could be discerned for the temperature-dependent loss-of-function of the enzymes encoded by tt4(UV01) and tt4(UV25).
Introduction

The type III polyketide synthases (PKSs) comprise a structurally-related class of enzymes that are key to the production of an array of biologically-active natural products in plants and some bacteria (Austin and Noel, 2003; Moore et al., 2002). These enzymes, also known as homodimeric iterative PKSs, include the extensively-studied chalcone synthase (CHS), an enzyme that is ubiquitous among plants and that catalyzes the first committed step in the synthesis of flavonoids. In the CHS reaction, one molecule of \( p \)-coumaroyl CoA and three molecules of malonyl CoA are used to generate a tetraketide that is then cyclized into 4,2',4',6'-tetrahydroxychalcone. This molecule serves as a backbone for the production of a wide array of flavonoid compounds that function in such diverse roles as flower pigmentation, UV protection, signaling, male fertility, and defense against microbial pathogens (Winkel-Shirley, 2001).

Alfalfa CHS was the first type III PKS, and the first flavonoid enzyme, for which a crystal structure was solved (Ferrer et al., 1999). Together with the subsequent determination of structures for mutagenized forms of this enzyme and for 2-pyrone synthase (PS) from daisy (Jez et al., 2002; Jez et al., 2000b), significant insights have been gained into the molecular basis of the reaction series catalyzed by the type III PKSs. For example, it was known that, although each of the two active sites in CHS and the closely related enzyme, stilbene synthase (STS), function independently, dimerization is required for activity (Tropf et al., 1995). It is now clear that this is because a methionine in each monomer helps shape the active site cavity of the adjoining monomer. In addition, it has been shown that each active site is buried within the enzyme and that substrates enter via a long (16 Å in alfalfa CHS) CoA binding tunnel [Ferrer, 1999 #79]. The three key catalytic residues are located within an initiation/elongation cavity, one lobe of which forms a coumaroyl binding pocket, while the other accommodates and
determines the length of the growing polyketide chain. Although CHS and PS use different substrates, produce different products, and exhibit 74% amino acid sequence identity, the overall architecture of the two enzymes is nearly identical. Remarkably, it has been possible to engineer CHS for PS activity by altering just three residues that affect the overall dimensions of the initiation/elongation cavity (Jez et al., 2002).

Several recent studies have utilized molecular dynamics simulations in efforts to elucidate the functional properties of a protein containing amino acid substitutions. For example, Ceruso et al. (2004) were able to suggest a nucleotide exchange mechanism in the Ga subunit of transducin by modeling a series of substitutions in residues that had been previously characterized biochemically. Another study, on the molecular switch present in the glucocorticoid receptor, used molecular dynamics simulations to show that distinct substitutions that result in a gain-of-function are caused by two different structural mechanisms (Stockner et al., 2003).

We previously characterized an allelic series of mutations for Arabidopsis CHS, which is defined by the TT4 locus (Saslowsky et al., 2000). The seven tt4 alleles included in this analysis were found to have diverse effects on the accumulation of CHS mRNA, the dimerization and stability of the enzyme, and the accumulation of flavonoid end products. Two of the alleles also exhibited temperature sensitivity. A homology model was initially generated for the wild-type Arabidopsis enzyme based on the Medicago CHS2 crystal structure, and the locations of the five tt4 mutations resulting in single amino acid changes or a small deletion were mapped on this model (Ferrer et al., 1999). This preliminary analysis suggested that several of the mutations could have significant effects on the structure of the CHS enzyme. To further explore this possibility, individual homology models have been generated for these five altered proteins
based on the recently-solved crystal structure for *Arabidopsis* CHS (Chapter 3). These models provide insights into the phenotypic and biochemical consequences that arise from either point substitutions or truncation in the enzyme in three of these alleles.

**Methods**

A multiple sequence alignment of 15 representative members of the plant type III PKS superfamily was generated using Clustal W (Thompson et al., 1994) using the Megalign program in DNAStar (Madison, WI). Included in this analysis were CHS from *Arabidopsis* (accession number P13114), alfalfa (P30074), maize (SYZMCC), parsley (S42523), raspberry (AAK15174), snapdragon (P06515), and tea (P48387); acridone synthase 2 from rue (Q9FSC0); bibenzyl synthase from orchid (S71619); coumaroyl triacetic acid synthase from hydrangea (BAA32733); dihydropinosylvin-forming stilbene synthase from pine (Q02323); PS from daisy (P48391); trihydroxystilbene synthase sequences from grape (P51070) and peanut (S00334); and valerophenone synthase (O80400) from hops.

Homology models were generated for five of the *Arabidopsis* tt4 alleles [tt4(85), tt4(UV113), tt4(38G1R), tt4(UV01), and tt4(UV25)] based on the crystal structure for *Arabidopsis* CHS (Chapter 3). The deduced sequences of the mutant proteins (Saslowsky et al., 2000; Shirley et al., 1995) were first aligned independently with alfalfa CHS2 using Biology Workbench (http://workbench.sdsc.edu). Five models were produced for each allele using Modeller6 of Sali and Blundell (1993) and an average model was generated for each protein by coordinate averaging. The averaged structures were minimized using the Sander module of Amber7 for 200 steps using steepest descent (Pearlman et al., 1995). Models were then solvated with water, and sodium ions were added to give the system a net charge of zero. Minimization
of the solvent and counterions was performed for 200 steps as described above, followed by a minimization of the entire system. Molecular dynamics was then performed using the Sander module of Amber7 for a total of 500 ps with a time step of 2.0 fs and a structure collected at every 1 ps. All simulations were performed using 4-8 processors on Virginia Tech’s Laboratory for Advanced Scientific Computing and Applications Linux cluster. The dynamics simulation was divided into the following components: the first 80 ps was a heating phase to raise the system temperature from 0 K to either 285 K, 293 K, or 300 K, the next 120 ps was a constant volume equilibration phase, and the final 300 ps was a constant pressure phase. Final models were generated by coordinate averaging of the last 100 ps of the dynamics simulation. The solvent and sodium ions were manually stripped from the average structures and the systems were minimized again as described above. Assessments of the models were performed using Procheck on the Whatif server (http://swift.embl-heidelberg.de/servers/). Root mean squared deviation (RMSD) calculations for the last 100 ps of the dynamics simulations were performed using the Carnal module of Amber7 (a representative RMSD graph is shown in Figure 2.1). Distances between alpha carbons, phi-psi torsion angles for residues, and hydrogen bond networks were determined using the Carnal module of Amber7 using the last 100 ps of the dynamics simulations. To determine the structural differences between the TT4 proteins and the wild-type structure, the structures were individually aligned by the least squares fit function within Swiss-Pdb Viewer, version 3.7 (Guex and Peitsch, 1997) and RMSD was calculated using the same program. All images of protein structures were generated using Swiss-Pdb Viewer (Guex and Peitsch, 1997) and rendered using POV-Ray version 3.1a (www.povray.org).
Results

Seven alleles of the *Arabidopsis* CHS gene, induced by either EMS or gamma radiation, were previously characterized at the gene, protein, and end product levels (Table 2.1) (Burbulis et al., 1996; Saslowsky et al., 2000; Shirley et al., 1995). One of these alleles [*tt4(UV118a)*] contains a complex rearrangement, while the other six contain point mutations that either disrupt splicing to generate a severely-truncated coding region [*tt4(2YY6)*], introduce a stop codon to truncate the coding region by 15 codons [*tt4(38G1R)*], or alter a single amino acid in the gene product [*tt4(85), tt4(UV01), tt4(UV25), and tt4(UV113)*]. Most of the mutations completely abolish the accumulation of flavonoids in affected plants, although one appears to be leaky and two others display a temperature-sensitive phenotype (unpublished observations and Saslowsky et al., 2000) (Table 2.1).

To further investigate the effects of small changes in CHS protein sequence on the properties of the enzyme, the proteins containing single amino acid changes or the C-terminal truncation were first examined in the context of current information on structure-function relationships in CHS. A multiple sequence alignment generated using 14 representative members of the plant type III PKS superfamily, including *Arabidopsis* CHS, illustrates that each of the *tt4* mutations affect residues that are very highly, if not absolutely, conserved in these enzymes (Figure 2.2) and that most are also located in highly-conserved stretches of sequence. Although all five mutations significantly affect flavonoid accumulation, none affect residues with known functional importance. Mapping of the affected residues on a homology model of the *Arabidopsis* wild-type enzyme previously showed that many are located in or near important functional domains, including the dimerization interface and the catalytic center, suggesting that
these mutations may have other, indirect effects on various aspects of enzyme function
(Saslowsky et al., 2000) (Figure 2.3). To examine the structural effects of these mutations in
further detail, the crystal structure for the wild-type Arabidopsis enzyme (Chapter 3) was first
subjected to energy minimizations in the presence of solvent and counterions, followed by
molecular dynamics simulations, as described in the Methods. Models were then developed and
similarly refined for the products of the five mutant alleles, tt4(85), tt4(UV01), tt4(UV25),
tt4(UV113), and tt4(38G1R). The effects of the altered residues and the C-terminal truncation
on the overall structure of the enzyme and on the dimensions and architecture of both the active
site and CoA binding site were examined by calculating the RMSD values, measuring the
distances between alpha carbons, and the phi and psi angles.

*tt4*(85)

The *tt4*(85) allele was the first *Arabidopsis* CHS mutation to be reported (Koornneef, 1990). Although plants carrying this allele accumulate wild-type levels of CHS protein that
appears to dimerize normally, flavonoid levels in these plants are reduced to almost undetectable
levels (data not shown, Saslowsky et al., 2000). The *tt4*(85) allele carries a point mutation that
results in replacement of a glycine at position 268 with serine (G268S) (Shirley et al., 1995).
This glycine residue appears to be absolutely conserved in all type III PKSs and is located in a
beta sheet (β11) in linear and spatial proximity to amino acids involved in CoA binding (L273
and K275) as well as to determinants of substrate specificity (I260 and F271) and product length
(G262) (Jez et al., 2000a) (Figures 2.2 and 2.4), and near an arginine (R265) that has been shown
to be essential for activity of raspberry CHS1 (Zheng et al., 2001). A way of measuring the
difference between two proteins is to fit one onto another structure and calculate the RMSD
between the two. The RMSD for the tt4(85) when compared to the wild-type structure is 2.48 Å, also indicative of significant structural differences between the two proteins. To investigate the effect of the G268S substitution on the geometry of the active site, the distances between the alpha carbons of the four key active site residues were measured (Table 2.2). The results indicate that the active site histidine (H309) appears closer to the critical active site cysteine (C169), but is further from the active site asparagine (N342) and phenylalanine (F220) in tt4(85) (Table 2.2). In the wild-type enzyme, C169 functions as a nucleophile, removing the CoA moiety from either coumaroyl-CoA or malonyl-CoA and binding the reaction intermediates during polyketide formation (Ferrer et al., 1999). Both the phi and psi angles for this residue appear to be distorted in tt4(85), indicating that the side chain may be in an unfavorable orientation for catalysis. The phi angle for F271, which is one of two ‘gatekeeper’ residues just outside the active site, is also much larger than wild-type (Table 2.3), which could affect entry of substrate into the active site. Moreover, almost all of the residues involved in CoA and substrate binding appear to be further apart than in the wild-type protein (Table 2.4). The increase in these distances, some being almost 2 Å larger than that of wild-type, indicates that this protein might be incapable of binding or correctly positioning the CoA moiety within the active site. The homology model of the tt4(85) protein also indicates that helix a3, which contains three of the CoA binding residues is distorted relative to the wild-type, with a kink in the middle of this helix (Figure 2.4). Although, the tt4(85) protein does retain some activity, as indicated by the leaky phenotype of mutant plants, it provides a good example of how a single amino acid change can have substantial structural and mechanistic effects on enzyme function.
The \texttt{tt4(UV113)} mutation results in the substitution of proline for threonine at position 174 (T174P). This highly-conserved amino acid is located in helix \( \alpha 7 \) (Figure 2.2; Figure 2.4C) and is five residues C-terminal to the active site cysteine. As in the case of \texttt{tt4(85)}, plants carrying this mutation produce wild-type levels of CHS protein that appears to dimerize normally (Saslowsky et al., 2000). However, no detectable flavonoids are produced. T174 appears to be absolutely conserved in CHS enzymes and most other type III PKSs, with the exception of valerophenone synthase from Hops, where there is a lysine at this position (Figure 2.2). Homology modeling indicates that, although the \texttt{tt4(UV113)} protein maintains its overall tertiary structure, the spatial relationship of the active site cysteine (C169) to the other active site residues is altered (Table 2.2). There are also changes in the phi angle for the C169 backbone relative to the wild-type protein (Table 2.3), which could indicate that the side chain is being forced into an orientation that is inaccessible to the substrate. Measurement of the distances between the alpha carbons of the CoA binding residues indicates that, for the most part, the geometry of these residues is unchanged. However, insertion of a proline residue into the \( \alpha 7 \) helix is likely to reduce the flexibility of this region, which is indicated by an decrease in the RMSD for the \( \alpha 7 \) helix (1.09 ± 0.071 Å for \texttt{tt4(UV113)} and 1.49 ± 0.091 Å for wild-type). When the same measurement is performed for residues 166-172, the loop containing the active site cysteine, the RMSD is 0.614 ± 0.106 Å for UV113 and 0.882 ± 0.093 Å for wild-type. Thus, although the T174P substitution appears to have little effect on the overall structure of the active site pocket, it could affect the dynamic movement of the active site cysteine during catalysis. The \texttt{tt4(UV113)} substitution has an even more dramatic effect on enzyme function than the one in \texttt{tt4(85)}, presumably due to a loss in flexibility of the loop containing the active site cysteine.
The point mutation in *tt4*(38G1R) shifts the reading frame starting at codon 381, introducing 3 altered residues and a premature stop codon so that 15 residues are lost from the C terminus. The protein accumulates at much lower levels than the wild-type CHS, is catalytically inactive, and does not form dimers (Table 2.1) (Saslowsky et al., 2000). The first two residues altered by this mutation are part of the conserved GFGPG loop that provides a scaffold for the cyclization reaction in type III PKSs (Austin and Noel, 2003; Suh et al., 2000), and are located on the opposite side of the protein from the dimerization interface (Figure 2.4D).

The model for *tt4*(38G1R) bears the most striking changes in overall tertiary structure of all the altered proteins (Figure 2.4). Due to the truncation, beta sheet 15, which appears to be required for the formation and stabilization of a series of three beta sheets, is lost. The RMSD between this structure and that of the wild-type enzyme is 2.92 Å, indicating significant differences in the secondary and tertiary structures of the two proteins. The homology model also indicates that the residues comprising beta sheets 14 and 7 in the wild-type no longer retain any beta sheet structure in *tt4*(38G1R). Interestingly, the loss of this series of beta sheets also affects one of the active site residues, H309, which functions with N342 as an electron sink to stabilize the transition state during the transfer of substrate intermediates to C169 (Austin and Noel, 2003). This residue is located at one end of beta sheet 14 in the wild-type protein and is moved in *tt4*(38G1R) (Table 2.2). Moreover, the phi angles for both the gatekeeper amino acids, F220 and F271, and the psi angles for F271 are significantly different than in the wild-type protein (Table 2.3), which suggests that these residues could be obstructing the active site entrance. Also, the CoA binding residues are substantially further apart than in the wild-type protein (Table 2.4), indicating that these residues may not be able to bind substrate. However,
from the analyses performed in this study, no reason for the loss of dimerization was determined. These results suggest the apparent lack of catalytic activity may result from the loss of three beta sheets required for protein stability and catalytic function.

*tt4*(UV01)

*tt4*(UV01) is a temperature-sensitive allele that results in the replacement of a threonine with an isoleucine at position 49 (T49I). As is the case for G268 in the *tt4*(85) allele, this residue appears to be invariant in all type III PKSs (Figure 2.2). Position 49 is near the CoA binding residues in both the primary sequence and the tertiary structure (Figure 2.3). This amino acid change does not appear to affect the ability of the enzyme to dimerize, consistent with the fact that it is located on the opposite side of the enzyme from the dimerization interface, but does appear to decrease protein stability at all temperatures (Saslowsky et al., 2000) (Table 2.1). These plants exhibit a temperature-dependent reduction in the accumulation of flavonoid end products, with almost no flavonoids present at 27°C, very low levels at 20°C (the optimum growth temperature for *Arabidopsis*), and moderate levels at 12°C (Saslowsky et al., 2000).

After subjecting the *tt4*(UV01) homology model to molecular dynamics simulations at 27°, 20°, and 12°C, and superimposing the resulting structures on that of the wild-type protein, no substantial changes in overall secondary or tertiary structure were apparent. The phi and psi angles do not deviate substantially from the wild-type for the majority of the residues with the exception of F271 (Table 2.3), which is involved in substrate specificity. Interestingly, the phi angle for F271 is significantly different from wild-type at all three temperatures investigated. The dimensions of the CoA binding tunnel did vary with temperature, however the greatest changes occurred at lower temperatures, where the enzyme exhibits some catalytic activity.
Overall, the homology model did not provide any clues regarding the temperature-sensitive phenotype of the tt4(UV01) enzyme.

**tt4(UV25)**

The *tt4*(UV25) allele is characterized by the replacement of arginine at position 334 with a cysteine (R334C) (Saslowsky et al., 2000). This mutation is located eight residues N-terminal to the active site asparagine (N342) and is highly conserved among type III PKSs. The known exceptions are two confirmed and two suspected isoforms of acridone synthase from *rue* (Junghanns et al., 1995) (accession numbers S60241, Q9FSC0, Q9FSC1, and Q9FSC2, respectively), which have conservative substitutions of lysine at this position (Figure 2.2). Similar to tt4(UV01), the tt4(UV25) enzyme is present at reduced levels in mutant plants relative to wild-type and exhibits a temperature-dependent reduction in flavonoid accumulation, however, it also appears to have lost the ability to form homodimers (Table 2.1). Molecular dynamics simulations were performed at 12º, 20º, and 27º C. Analysis of the dihedral angles of the active site residues as well as the distances between the active site residues, CoA binding residues, and substrate specificity residues, provided no obvious explanation for the phenotypes of the tt4(UV25) enzyme. However, the RMSD values are indicative of a structure that is not very stable at 12º C. Also, the RMSD between the wild-type and the UV25 structures at 12º C is 2.50 Å, indicating that the two structures are different from each other. Both of these results are opposite to what was expected, as the protein should be more stable at lower temperatures. Interestingly, this substitution occurs in the region of the protein that is proposed to be important for interaction with CHI (Figure 2.3) (Chapter 3). Perhaps the alteration of this residue affects this predicted interaction domain in a temperature-dependent manner, thus impacting not only
CHS protein accumulation, but its ability to function as part of a proposed flavonoid enzyme complex.

Discussion

This study indicates that homology modeling can be a useful tool for the analysis of structure-function relationships in enzymes such as CHS. An allelic series of mutations in the CHS gene had previously been characterized on the molecular and genetic level. For most of the alleles, molecular dynamics simulations of homology models were able to suggest a structural basis for the loss of function, usually due to distortion of the active site pocket, an overall change in the dimensions of the CoA binding tunnel dimensions, the loss of flexibility in loop, or a loss of substantial tertiary structure. The molecular dynamics simulations indicated that the substitutions caused changes in the structures of tt4(85), tt4(UV113), and tt4(38G1R) that may explain the changes in function and accumulation. However, in the case of the temperature-sensitive alleles, the molecular dynamics were unable to provide an explanation for the temperature-dependent loss of function and stability. Also, for the tt4(38G1R) and tt4(UV01) alleles, this study was unable to provide any predictions for the loss of dimerization of the protein. It is important to note that the dynamics performed in this study were done on the monomeric form of CHS, thus simulations on the dimeric form of the protein could provide additional useful information. Future studies could also include in silico docking of substrate into the CHS active site to determine if the proteins are able to bind the CoA moiety, which is required for proper positioning of the substrate.
Modeling and molecular dynamics simulations can also be applied to other systems where biochemical data exist for amino acid substitutions in order to gain information about effects on secondary and tertiary structure. These methods could be used as an inexpensive approach to rationally engineering a gain or loss of function within proteins. It might also be possible to engineer novel catalytic activities by first modeling amino acid substitutions and then, using quantum mechanical calculations, simulate the reaction that would be performed.

Moreover, the techniques described in this study could be used to elucidate the architecture of multi-enzyme complexes. First, through *in silico* docking techniques and molecular dynamics simulations, the identity and location of residues that contribute to protein-protein interactions could be identified and then verified by experimental techniques such as yeast two-hybrid and surface plasmon resonance refractometry. Once the interaction has been identified, it might then be possible to engineer stronger or weaker interactions through amino acid substitutions in order to modify metabolic flux creating new activities or causing the accumulation of specific end products.
Acknowledgments

The authors would like to thank Jory Zmuda Ruscio for her assistance with the molecular dynamics simulations. This work was supported by the National Science Foundation (grants MCB-9808117 and MCB-0131010) and by grants from the Virginia Tech Graduate Research Development Project to C.D.D.
References


Table 2.1 Biochemical and genetic characteristics of *Arabidopsis* CHS alleles.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Mutation$^a$</th>
<th>CHS protein levels$^b$</th>
<th>Allele type</th>
<th>Flavonoid glycoside levels$^c$</th>
<th>Homodimer crosslinking$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild-type</td>
<td>N.A.$^c$</td>
<td>++++</td>
<td>Wild-type</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>tt4(85)$^e$</td>
<td>G268S</td>
<td>++++</td>
<td>Leaky</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>tt4(UV113)</td>
<td>T174P</td>
<td>++++</td>
<td>Null</td>
<td>-</td>
<td>+++</td>
</tr>
<tr>
<td>tt4(38G1R)</td>
<td>P381S, G382H, L383C, T384stop</td>
<td>+</td>
<td>Null</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>tt4(UV01)</td>
<td>T49I</td>
<td>++</td>
<td>Temperature-sensitive</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>tt4(UV25)</td>
<td>R334C</td>
<td>++</td>
<td>Temperature-sensitive</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

$^a$ Amino acid(s) altered in the predicted protein.
$^b$ Determined by immunoblot analysis (Saslowsky et al., 2000).
$^c$ Determined by HPLC analysis (data not shown, Saslowsky et al., 2000).
$^d$ Determined using disuccinimidyl glutarate and *N*-succinimidyl-[4-vinylsulfonyl] benzoate (Saslowsky et al., 2000).
$^e$ N.D., not determined.
Table 2.2. Dimensions\(^a\) of the CHS proteins

<table>
<thead>
<tr>
<th></th>
<th>27°C</th>
<th>20°C</th>
<th>12°C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Arabidopsis</td>
<td>tt4 (85)</td>
<td>tt4 (UV113)</td>
</tr>
<tr>
<td>RMSD(^b) vs. wt</td>
<td>N/A 2.48</td>
<td>1.79</td>
<td>2.92</td>
</tr>
<tr>
<td>RMSD(^c) over 100 ps</td>
<td>1.81 (0.06)</td>
<td>2.07 (0.06)</td>
<td>1.87 (0.06)</td>
</tr>
<tr>
<td>F220 to N342</td>
<td>4.75 (0.22)</td>
<td>4.78 (0.19)</td>
<td>4.74 (0.18)</td>
</tr>
<tr>
<td>C169 to H309</td>
<td>7.74 (0.38)</td>
<td>6.76* (0.22)</td>
<td>7.17 (0.29)</td>
</tr>
<tr>
<td>C169 to F220</td>
<td>10.78 (0.42)</td>
<td>11.86* (0.33)</td>
<td>10.49 (0.31)</td>
</tr>
<tr>
<td>H309 to N342</td>
<td>8.18 (0.22)</td>
<td>9.14* (0.30)</td>
<td>8.48 (0.24)</td>
</tr>
<tr>
<td>C169 to N342</td>
<td>9.19 (0.32)</td>
<td>9.34 (0.27)</td>
<td>8.61* (0.27)</td>
</tr>
<tr>
<td>F220 to H309</td>
<td>12.21 (0.33)</td>
<td>13.44* (0.33)</td>
<td>12.64 (0.26)</td>
</tr>
</tbody>
</table>

\(^a\)Values are in angstroms and represent the average of distances between alpha carbons of the indicated residues in the final 100 structures generated by dynamic simulation. Numbers in parentheses are standard deviations.

\(^b\)RMSD values, which indicate overall similarity of structures relative to Arabidopsis wild-type CHS, were calculated using Swiss-PdbViewer v3.7.

\(^c\)RMSD values, which indicate the overall stability of the Ca backbone, was calculated using Amber7 (Pearlman et al., 1995). An asterisk (*) indicates a 1 Å or greater deviation from the wild-type.
Table 2.3. Phi-psi torsion angles within the CHS active site\textsuperscript{a}.

<table>
<thead>
<tr>
<th></th>
<th>27°C</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>20°C</th>
<th></th>
<th></th>
<th></th>
<th>12°C</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wild-type</td>
<td>tt4 (UV113)</td>
<td>tt4 (38G1R)</td>
<td>tt4 (UV01)</td>
<td>tt4 (UV25)</td>
<td>Wild-type</td>
<td>tt4 (UV01)</td>
<td>tt4 (UV25)</td>
<td>Wild-type</td>
<td>tt4 (UV01)</td>
<td>tt4 (UV25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C169 $\phi$</td>
<td>-58.63 (10.61)</td>
<td>41.46* (10.3)</td>
<td>-44.50 (12.1)</td>
<td>-62.49 (10.0)</td>
<td>-49.95 (10.3)</td>
<td>-58.62 (10.2)</td>
<td>-54.47 (8.23)</td>
<td>-60.72 (9.2)</td>
<td>-57.40 (10.46)</td>
<td>-56.92 (7.9)</td>
<td>-54.29 (10.0)</td>
<td>-48.26 (12.1)</td>
<td></td>
</tr>
<tr>
<td>C169 $\psi$</td>
<td>-19.71 (14.3)</td>
<td>-100.98* (11.8)</td>
<td>-37.84 (12.0)</td>
<td>-7.04 (10.5)</td>
<td>-13.00 (10.3)</td>
<td>-28.12 (8.5)</td>
<td>-44.50* (12.1)</td>
<td>-77.83 (13.7)</td>
<td>-63.77 (9.1)</td>
<td>-75.98 (9.8)</td>
<td>-83.21 (10.1)</td>
<td>-80.89 (9.7)</td>
<td></td>
</tr>
<tr>
<td>F220 $\phi$</td>
<td>-65.59 (11.0)</td>
<td>-56.17 (8.9)</td>
<td>-67.35 (9.0)</td>
<td>-76.58* (9.6)</td>
<td>-68.8 (8.4)</td>
<td>-61.47 (10.7)</td>
<td>-72.94 (11.6)</td>
<td>-87.83 (13.7)</td>
<td>-60.72 (9.2)</td>
<td>-57.40 (10.46)</td>
<td>-56.92 (7.9)</td>
<td>-54.29 (10.0)</td>
<td></td>
</tr>
<tr>
<td>F220 $\psi$</td>
<td>142.6 (10.2)</td>
<td>133.1 (8.7)</td>
<td>137.1 (7.4)</td>
<td>150.2 (8.1)</td>
<td>136.4 (9.6)</td>
<td>143.5 (10.3)</td>
<td>147.5 (8.1)</td>
<td>153.3 (7.2)</td>
<td>140.7 (8.7)</td>
<td>156.59 (7.0)</td>
<td>152.9 (6.4)</td>
<td>152.17 (7.0)</td>
<td></td>
</tr>
<tr>
<td>H309 $\phi$</td>
<td>-68.6 (10.3)</td>
<td>-66.6 (9.5)</td>
<td>-72.5 (11.0)</td>
<td>-68.5 (11.1)</td>
<td>-74.0 (11.6)</td>
<td>-73.42 (12.4)</td>
<td>-72.5 (11.0)</td>
<td>-71.7 (13.2)</td>
<td>-71.3 (10.0)</td>
<td>-74.0 (12.0)</td>
<td>-66.7 (10.7)</td>
<td>-66.0 (11.6)</td>
<td></td>
</tr>
<tr>
<td>H309 $\psi$</td>
<td>120.0 (7.5)</td>
<td>108.4 (8.6)</td>
<td>118.5 (7.2)</td>
<td>115.1 (6.7)</td>
<td>118.6 (7.0)</td>
<td>115.2 (7.1)</td>
<td>113.2 (6.8)</td>
<td>113.7 (5.9)</td>
<td>115.1 (7.1)</td>
<td>114.8 (6.5)</td>
<td>115.4 (6.4)</td>
<td>115.5 (5.8)</td>
<td></td>
</tr>
<tr>
<td>N342 $\phi$</td>
<td>-87.02 (11.7)</td>
<td>-73.5 (8.5)</td>
<td>-89.6 (18.9)</td>
<td>-83.1 (10.9)</td>
<td>-128.7* (23.7)</td>
<td>-81.4 (10.4)</td>
<td>-79.8 (10.2)</td>
<td>-79.9 (12.2)</td>
<td>-81.6 (12.6)</td>
<td>-81.42 (10.0)</td>
<td>-78.3 (11.4)</td>
<td>-79.8 (10.6)</td>
<td></td>
</tr>
<tr>
<td>N342 $\psi$</td>
<td>96.0 (10.7)</td>
<td>97.6 (10.4)</td>
<td>93.8 (14.0)</td>
<td>99.7 (8.0)</td>
<td>135.21* (12.0)</td>
<td>96.5 (9.7)</td>
<td>104.1 (12.2)</td>
<td>112.4 (12.9)</td>
<td>94.9 (10.1)</td>
<td>96.5 (10.7)</td>
<td>112.6* (16.3)</td>
<td>103.3 (10.2)</td>
<td></td>
</tr>
<tr>
<td>F271 $\phi$</td>
<td>-78.5 (15.3)</td>
<td>-148.0* (9.0)</td>
<td>-126.0* (34.6)</td>
<td>-56.5* (8.7)</td>
<td>-146.1* (10.2)</td>
<td>-155.5* (8.9)</td>
<td>-149.9 (34.2)</td>
<td>-77.3* (19.4)</td>
<td>-88.3* (18.4)</td>
<td>-89.6 (22.2)</td>
<td>-150.8* (8.4)</td>
<td>-146.7* (8.4)</td>
<td></td>
</tr>
<tr>
<td>F271 $\psi$</td>
<td>120.7 (8.7)</td>
<td>138.7* (11.4)</td>
<td>144.3* (8.0)</td>
<td>134.2* (9.8)</td>
<td>136.5* (13.2)</td>
<td>143.0* (11.3)</td>
<td>129.2 (9.2)</td>
<td>149.5* (7.9)</td>
<td>121.2 (12.2)</td>
<td>127.7 (36.7)</td>
<td>133.9* (103.5)</td>
<td>125.5 (9.1)</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}Values are in degrees. The phi angle ($\phi$) indicates rotation around the N-C\(_\alpha\) bond, while the psi angle ($\psi$), indicates rotation around the C\(_\alpha\)-C bond. Positive and negative values indicate right-handed and left-handed rotation, respectively. Standard deviations are given in parentheses. An asterisk (*) indicates rotation 15º greater than wild-type.
Table 2.4. Distances between CoA binding residues.

<table>
<thead>
<tr>
<th></th>
<th>Arabidopsis CHS</th>
<th>tt4 (85)</th>
<th>tt4 (UV113)</th>
<th>tt4 (38G1R)</th>
<th>tt4 (UV01)</th>
<th>tt4 (UV25)</th>
<th>Arabidopsis CHS</th>
<th>tt4 (UV01)</th>
<th>tt4 (UV25)</th>
<th>Arabidopsis CHS</th>
<th>tt4 (UV01)</th>
<th>tt4 (UV25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K67 to K275</td>
<td>17.7 (0.5)</td>
<td>19.2* (0.8)</td>
<td>16.6* (0.4)</td>
<td>18.8 (0.7)</td>
<td>18.3 (0.5)</td>
<td>17.1 (0.7)</td>
<td>17.6 (0.6)</td>
<td>17.8 (0.6)</td>
<td>17.4 (0.5)</td>
<td>16.8 (0.6)</td>
<td>17.1 (0.5)</td>
<td>17.0 (0.4)</td>
</tr>
<tr>
<td>K67 to L273</td>
<td>15.7 (0.4)</td>
<td>16.4 (0.8)</td>
<td>14.8* (0.4)</td>
<td>16.9* (0.8)</td>
<td>16.0 (0.5)</td>
<td>15.1 (0.6)</td>
<td>15.3 (0.6)</td>
<td>15.8 (0.6)</td>
<td>15.4 (0.5)</td>
<td>15.3 (0.8)</td>
<td>13.8* (0.5)</td>
<td>15.3</td>
</tr>
<tr>
<td>K60 to K275</td>
<td>18.0 (0.4)</td>
<td>20.8* (0.6)</td>
<td>17.4 (0.6)</td>
<td>19.9* (0.5)</td>
<td>18.9* (0.5)</td>
<td>17.8 (0.7)</td>
<td>18.8 (0.5)</td>
<td>19.9 (0.6)</td>
<td>18.0 (0.4)</td>
<td>17.8 (0.5)</td>
<td>18.7 (0.5)</td>
<td>18.0</td>
</tr>
<tr>
<td>K60 to L273</td>
<td>15.8 (0.3)</td>
<td>17.9* (0.4)</td>
<td>15.5 (0.4)</td>
<td>17.2* (0.5)</td>
<td>15.8 (0.5)</td>
<td>15.3 (0.5)</td>
<td>15.8 (0.4)</td>
<td>16.9* (0.7)</td>
<td>16.1 (0.4)</td>
<td>15.8 (0.5)</td>
<td>15.0 (0.5)</td>
<td>15.7</td>
</tr>
<tr>
<td>R63 to K275</td>
<td>18.8 (0.4)</td>
<td>21.0* (0.6)</td>
<td>18.2 (0.5)</td>
<td>20.6* (0.5)</td>
<td>19.5 (0.6)</td>
<td>18.4 (0.7)</td>
<td>19.1 (0.5)</td>
<td>17.8* (0.6)</td>
<td>19.0 (0.4)</td>
<td>18.2 (0.7)</td>
<td>18.9 (0.5)</td>
<td>18.5</td>
</tr>
<tr>
<td>R63 to L273</td>
<td>16.7 (0.4)</td>
<td>18.0* (0.5)</td>
<td>16.3 (0.3)</td>
<td>18.2* (0.7)</td>
<td>16.7 (0.6)</td>
<td>16.0 (0.6)</td>
<td>16.3 (0.4)</td>
<td>15.8* (0.6)</td>
<td>16.9 (0.3)</td>
<td>16.4 (0.8)</td>
<td>15.3* (0.5)</td>
<td>16.4</td>
</tr>
<tr>
<td>K67 to A314</td>
<td>8.9 (0.3)</td>
<td>11.0* (0.6)</td>
<td>9.2 (0.4)</td>
<td>9.4 (0.5)</td>
<td>9.6 (0.3)</td>
<td>9.4 (0.5)</td>
<td>9.2 (0.4)</td>
<td>9.6 (0.4)</td>
<td>9.7 (0.4)</td>
<td>9.0 (0.4)</td>
<td>9.5 (0.4)</td>
<td>9.0</td>
</tr>
<tr>
<td>F220 to F271</td>
<td>10.5 (0.3)</td>
<td>10.6 (0.3)</td>
<td>10.6 (0.4)</td>
<td>10.1 (0.3)</td>
<td>10.5 (0.4)</td>
<td>10.3 (0.3)</td>
<td>10.7 (0.3)</td>
<td>9.6 (0.3)</td>
<td>9.1 (0.4)</td>
<td>10.9 (0.5)</td>
<td>10.6 (0.3)</td>
<td>11.1</td>
</tr>
<tr>
<td>K60 to I260</td>
<td>20.6 (0.4)</td>
<td>22.4* (0.5)</td>
<td>20.9 (0.4)</td>
<td>21.9* (0.4)</td>
<td>20.5 (0.4)</td>
<td>20.4 (0.6)</td>
<td>20.1 (0.4)</td>
<td>20.9 (0.7)</td>
<td>20.8 (0.4)</td>
<td>20.9 (0.7)</td>
<td>20.4 (0.5)</td>
<td>20.2</td>
</tr>
<tr>
<td>K60 to G262</td>
<td>22.8 (0.4)</td>
<td>22.8 (0.4)</td>
<td>22.2 (0.4)</td>
<td>23.3 (0.5)</td>
<td>21.8 (0.4)</td>
<td>21.9 (0.5)</td>
<td>22.4 (0.4)</td>
<td>22.6 (0.7)</td>
<td>23.1 (0.5)</td>
<td>22.0 (0.4)</td>
<td>22.8 (0.5)</td>
<td>22.2</td>
</tr>
<tr>
<td>R63 to I260</td>
<td>21.3 (0.5)</td>
<td>22.4* (0.5)</td>
<td>21.3 (0.3)</td>
<td>22.3* (0.5)</td>
<td>20.8 (0.5)</td>
<td>20.7 (0.7)</td>
<td>20.3 (0.4)</td>
<td>20.6 (0.7)</td>
<td>21.4 (0.4)</td>
<td>21.7 (0.9)</td>
<td>20.4* (0.5)</td>
<td>20.6</td>
</tr>
<tr>
<td>R63 to G262</td>
<td>23.9 (0.4)</td>
<td>23.8 (0.4)</td>
<td>23.4 (0.4)</td>
<td>24.5* (0.5)</td>
<td>23.2 (0.5)</td>
<td>23.0 (0.5)</td>
<td>23.5 (0.4)</td>
<td>22.9 (0.7)</td>
<td>24.1 (0.4)</td>
<td>22.9 (0.5)</td>
<td>23.3 (0.5)</td>
<td>23.8</td>
</tr>
<tr>
<td>K67 to I260</td>
<td>19.7 (0.6)</td>
<td>20.3 (0.6)</td>
<td>18.9 (0.4)</td>
<td>20.2 (0.7)</td>
<td>19.1 (0.5)</td>
<td>18.8 (0.9)</td>
<td>18.6 (0.6)</td>
<td>18.3 (0.6)</td>
<td>19.2 (0.6)</td>
<td>20.3 (0.9)</td>
<td>18.1* (0.5)</td>
<td>18.8*</td>
</tr>
<tr>
<td>K67 to G262</td>
<td>22.9 (0.4)</td>
<td>22.7 (0.6)</td>
<td>22.2 (0.7)</td>
<td>23.5 (0.6)</td>
<td>22.9 (0.6)</td>
<td>22.2 (0.6)</td>
<td>22.9 (0.5)</td>
<td>21.4* (0.5)</td>
<td>22.6 (0.5)</td>
<td>21.7 (0.5)</td>
<td>21.8 (0.5)</td>
<td>23.2*</td>
</tr>
</tbody>
</table>
Dana, et al., Figure 2.1
Figure 2.1. The RMSD of the backbone atoms for wild-type CHS from the starting confirmation throughout the simulation.
Dana, et al., Figure 2.2
Figure 2.2. Multiple sequence alignment of selected regions of a representative group of plant
type III polyketide synthases. CHS, acridone synthase (ACS), bibenzyl synthase (BBS),
coumaroyl triacetic acid synthase (CTAS), dihydropinosylvin-forming stilbene synthase (DPS),
2-pyrone synthase (PS), trihydroxystilbene synthase (THS), and valerophenone synthase (VPS)
sequences from a variety of plant species were aligned using CLUSTAL W (Thompson et al.,
1994). Highlighting indicates catalytic residues (red), CoA binding residues (purple), and active
site residues that are structural (yellow), control polyketide length (green), or determine substrate
specificity (blue) (Jez et al., 2000b). Residues that are altered in the four alleles are indicated
with asterisks; numbering is relative to the Arabidopsis CHS protein sequence. Percentages to
the right of the sequences refer to amino acid identity relative to Arabidopsis CHS. Pink dots
below the sequences indicate residues that are absolutely conserved among type II PKSs; green
dots indicate those conserved in CHS but not other type II PKSs. Elements of secondary
structure in the crystal structure of Arabidopsis CHS are identified below the sequences.
Dana, et al., Figure 2.3
**Figure 2.3.** Structure of the *Arabidopsis* wild-type CHS enzyme. The structure represented here is the 100 ps average from the molecular dynamics simulations at 27º C. Labeled on this structure are the active site (red), determinants of substrate specificity (orange), product length determinant (green), amino acids responsible for the binding of the CoA moiety (blue), the locations of the tt4 substitutions or truncation (purple), and the residues proposed to be involved in the interaction between CHS and CHI (yellow). A, front view; B, side view.
Dana, et al., Figure 2.4
Figure 2.4. Models of the *Arabidopsis* wild-type and tt4 CHS enzymes. The structures were generated based on the crystal structure CHS from *Arabidopsis* as described in the Methods. Active site residues (wild-type) and substituted residues (alleles) are labeled in red. The secondary elements labeled in green indicate structural changes relative to the wild-type. A, 27º C wild-type CHS structure; B, tt4(85) with alpha helix a3, showing disorder, and beta sheets ß10 and ß11 labeled to indicate the location of the substitution; C, tt4(UV113) with alpha helix a7 and the loop containing the active site cysteine labeled; D, tt4(38G1R) with the areas corresponding to the location of beta sheets ß7, ß12, ß13, and ß14 in the wild-type labeled to identify the loss of secondary structure resulting from the truncation.
Chapter 3

Structural Characterization of the Chalcone Synthase-Chalcone Isomerase Interaction

Christopher D. Dana¹, Stéphane B. Richard², Susan Krueger³, Joseph P. Noel², and Brenda S. J. Winkel¹

¹Department of Biology and Fralin Biotechnology Center, Virginia Tech, ²Structural Biology Laboratory, The Salk Institute for Biological Sciences, and ³NIST Center for Neutron Research
Abstract

Biosynthetic pathways are often organized within cells as groups of interacting enzymes known as enzyme complexes or metabolons. Although this organization appears to be widespread in biological systems, it is still poorly understood. The flavonoid pathway of Arabidopsis thaliana has been developed in our laboratory as a model for studying enzyme complexes (Burbulis and Winkel-Shirley, 1999). Several of the enzymes of this pathway have been shown to interact in vitro, and have also been found to be co-localized in roots (Burbulis and Winkel-Shirley, 1999; Saslowsky and Winkel-Shirley, 2001). Our studies are currently focusing on the first enzymes of the flavonoid biosynthetic pathway, chalcone synthase (CHS) and chalcone isomerase (CHI), which catalyze the condensation and cyclization of coumaroyl-CoA and 3 malonyl-CoA into the central flavonoid backbone. We propose a structure of the Arabidopsis CHS-CHI bienzyme complex as deduced by X-ray crystallography, small-angle neutron scattering (SANS), in silico docking, and molecular dynamics simulations. Surface-exposed residues thought to be involved in this interaction have been altered and tested by yeast 2-hybrid analysis. These experiments provide the first structural insights into the organization of flavonoid biosynthesis as well as establishing SANS as a technique that can be used in the study of metabolic organization.
Introduction

The study of multi-enzyme complexes, or metabolons, dates to the late 1950s with the isolation of a complex of enzymes involved in the citric acid cycle (Green, 1958). As time progressed, many more examples of multi-enzyme complexes were discovered, introducing the idea that biosynthetic pathways are organized as metabolons and leading to a broader understanding of how the cytoplasm of the cell is organized. Metabolic complexes can be characterized as either static, which are very stable and require little energy for association (Ovadi and Srere, 2000), or dynamic, requiring constant energy input for complex integrity (Welch, 1977). This organization also has the potential to offer numerous advantages to the cell such as coordinated flow of substrate from one active site to another, high local substrate concentration, and sequestering of substrates as protection for the both the cell and the pathway (Ovadi and Srere, 2000). Metabolic complexes also provide a mechanism by which the types of end products being produced can be quickly and effectively altered. It has even been suggested that all biosynthetic pathways are organized as multi-enzyme complexes (Beeckmans et al., 1994; Beeckmans et al., 1989; Gontero et al., 1988; Robinson et al., 1987).

Flavonoids constitute an important class of secondary metabolites produced in higher plants. Chalcone synthase (CHS), catalyzes the first enzymatic reaction in flavonoid biosynthesis, the condensation of three malonyl-CoA (from acetate metabolism) molecules with coumaroyl-CoA (from the general phenylpropanoid pathway) to create tetrahydroxychalcone (Figure 3.1). Chalcone isomerase (CHI), catalyzes the creation of naringenin, which serves as the main flavonoid backbone for
ensuing enzymatic reactions that create flavonols, anthocyanins, condensed tannins, and iso-flavonoids (Figure 3.1). The products of general flavonoid biosynthesis have diverse roles in plants, including serving as pigments in seeds, flowers and fruits and as UV sunscreens (reviewed in Winkel-Shirley, 2002). These plant products also have anti-oxidant, anti-dipsotrophic, and anti-cancer properties in mammals (reviewed in Havsteen, 2002). Recent metabolic engineering efforts have been directed at altering the amounts and types of these compounds for nutritional enhancement of plants as well as engineering plants with novel flower pigmentation (Forkmann and Martens, 2001; Muir et al., 2001; Yu et al., 2003).

Some 30 years ago, Helen Stafford speculated that the flavonoid pathway assembles as a multi-enzyme complex (Stafford, 1974a; Stafford, 1974b). The first indirect evidence for the existence of a flavonoid metabolon was published by Fritsch and Griseback (1975) who showed that operationally-soluble enzymes involved in flavonoid biosynthesis localize to the endoplasmic reticulum (ER). Later sucrose density and immunolocalization experiments verified that enzymes of both the general phenylpropanoid and flavonoid pathways are localized to the cytoplasmic face of the ER (Hrazdina and Wagner, 1985; Hrazdina et al., 1987; Wagner and Hrazdina, 1984). More recent work demonstrated the first direct evidence of interactions between flavonoid enzymes via affinity chromatography, yeast 2-hybrid analysis, and co-immunoprecipitation experiments (Burbulis and Winkel-Shirley, 1999). Immunofluorescence and immunoelectron microscopy experiments have shown that CHS and CHI co-localize at the ER and that the intracellular distribution of these enzymes is altered in an flavonoid 3’-hydroxylase (F3’H) mutant, consistent with a previously-
proposed role for F3’H as a membrane anchor for the complex (Stafford, 1990). This work also showed an asymmetric distribution in root cortex cells that points to a possible physiological significance of the organization in the control of auxin transport by flavonoids. Additional evidence for a flavonoid metabolon comes from the expression of anti-CHI antibodies in transgenic plants. These antibodies did not appear to change the catalytic activity of CHI, but did alter the production of flavonoids in seedlings, possibly by interfering with assembly of CHI into the flavonoid metabolon (Santos et al., 2004). Deducing the architecture of the flavonoid complex and the mechanisms governing its assembly and localization could yield critical new insights for effective engineering of this and other metabolic systems.

Small angle neutron scattering (SANS) is a technique that probes three-dimensional structure in a non-destructive manner (Koch et al., 2003). Data from SANS experiments provides a low resolution structure for molecules in an aqueous environment, independent of the packing inherent in crystal structures. This technique has many useful applications in polymer science, materials science, and in the life sciences. Examples of the latter include aiding in the determination of the structure of the 30S ribosomal subunit (Capel et al., 1987), elucidating better ways to crystallize a protein by determining the amount of protein that is aggregated (Velev et al., 1998), structural characterization of the GroEL/ES complex (Krueger et al., 2003), and structural changes in SecA resulting from nucleotide binding (Bu et al., 2003).

SANS, together with in silico docking algorithms and molecular dynamics simulations were used to develop a model of the CHS-CHI bienzyme complex. To further define the interactions between CHS and other enzymes in flavonoid biosynthesis,
four surface-exposed phenylalanine and tryptophan residues on CHS were converted to alanine and assayed by yeast 2-hybrid analysis. The molecular dynamics simulations helped identify a series of residues that may be involved in an electrostatic interaction between CHS and CHI that are being assessed in parallel 2-hybrid experiments. By understanding the interaction between CHS and CHI, it is hoped that a better perspective on metabolon architecture can be achieved, which in turn can be used in experiments to better understand protein complex assembly and regulation as a whole.

Materials and Methods

Site Directed Mutagenesis of pET32a (+)

The pET32a (+) vector was modified to allow cleavage of both the thioredoxin (TRX) and hexa-histidine tags from the protein of interest, leaving just one spurious amino acid (serine) on the N-terminus. Mutagenesis of pET32a(+) was performed using the Quik Change site-directed mutagenesis kit (Stratagene La Jolla, CA). Primers were designed to incorporate an NcoI site next to the thrombin proteolytic cleavage site. The primers were 5’-GGTGCCACGCGGTTCCATG

GTATGAAAGAA ACC-3’, with a second primer containing the complementary sequence. The underlined portion of the primer denotes the sequence that was altered. Mutagenesis was performed with Pfu Turbo (Stratagene) under the following conditions: 95°C for 30 s as an initial denaturing step followed by 18 cycles of 95°C for 30 s, 55°C for 1 min, and 68°C for 12 min. The PCR product was treated with DpnI (Promega) for 1 h at 37°C to digest the methylated template DNA, then used to transform E. coli XL1-Blue MRF’ (Novagen) cells using a heat shock method (Ausubel et al., 1989). The transformed cells were grown overnight at
37°C on LB\textsubscript{Amp100} plates. Confirmation of successful mutagenesis was via PCR using the primer, 5’-GGTGCCACGCGGTTCCAT-3’, and the T7 terminator primer and then sequenced using the BigDye Terminator kit (Stratagene) in order to verify sequence fidelity. This modified pET32a plasmid was given the designation pCD1.

**Creation of Expression Constructs**

The CHS and CHI coding regions were subcloned into the \textit{NcoI} and \textit{NotI} or \textit{SalI} sites, respectively, of the pCD1 expression vector to create pCD1-CHS and pCD1-CHI. PCR was performed using \textit{Pfu} turbo (Stratagene) and the following primers: 5’-CATGC\texttt{CATG}GTGATGGCTGGTGC-3’ and 5’-AGCGGCCGCT\texttt{TAGAGAGGAGGAACGC} TGTG-3’ for CHS and 5’-CATG\texttt{CCATGGGAA}ATG\texttt{TCTTTCATCCAACGC}G-3’ and 5’-\texttt{CGGTGACTGATTCTTCTTGTGCTAG}-3’ for CHI, with pCHS.cr2 or pCHI.cr as the templates (Pelletier and Shirley, 1996). Reactions were carried out under the following conditions: 95°C for 2 min as an initial denaturing step followed by 50 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 2 min and then a final elongation step of 72°C for 10 min. The PCR products were then digested with \textit{NcoI} and \textit{NotI} (CHS) or \textit{SalI} (CHI). T4 DNA Ligase (Promega, Madison, WI) was used to insert these fragments into the pCD1 expression vector. The ligation reaction was then used to transform \textit{E. coli} DH5\textsubscript{a} cells by electroporation (Dower et al., 1988). Transformed colonies were recovered on LB\textsubscript{Amp100} medium and screened by isolating plasmids using an alkaline lysis miniprep (Ausubel et al., 1989; Birnboim and Doly, 1979) which were then digested with with \textit{NcoI} and \textit{NotI} (CHS) or \textit{SalI} (CHI). Positive clones were sequenced using the BigDye
Terminator kit (Stratagene, La Jolla, CA) in order to verify sequence fidelity. Plasmids were then used to transform *E.coli* BL21(DE3) pLys-S (Novagen) using a heat shock method (Ausubel et al., 1989) for expression and purification of recombinant protein.

**Expression and Purification of pCD1-CHS and -CHI**

A single colony containing either pCD1-CHS or pCD1-CHI was used to inoculate LB<sub>Amp100-Cm34</sub>, which was grown overnight, then diluted 1:100 into the same medium. The cultures were grown to mid-log phase (OD<sub>595</sub> = 0.5-0.7) at 37ºC (approximately 4h), then isopropyl-beta-D-thiogalactopyranoside (Alexis Biochemicals) was added to a final concentration of 0.5 mM to induce protein expression. The cultures were grown at room temperature with shaking (250 RPM) for a further 4 h. The cells were harvested by centrifugation at 2500 x g for 10 min at 4ºC and frozen at -80º C until protein purification.

Frozen cells were resuspended in lysis buffer (50 mM Tris, pH 8.0; 150 mM NaCl; 10% glycerol; 1% Tween-20 and 10 mM β-mercaptoethanol, approximately 33 ml of lysis buffer per one liter of culture) and resuspended by stirring at 4ºC for 30 min. The mixture was then sonicated six times for 30 sec each time on ice, with 2 min between sonication. This lysate was centrifuged at 21,000 x g for 1 h at 4ºC. The supernatant (crude extract) was passed over a Ni-NTA (Qiagen) column, which was packed and equilibrated with 10 bed volumes of ddH<sub>2</sub>O and then 10 bed volumes of lysis buffer. The column was washed with 10 bed volumes of lysis buffer and then with the same volume of wash buffer (50 mM Tris, pH 8.0; 500 mM NaCl; 10 mM imidazole, pH 8.0; 10%
glycerol; and 10 mM β-mercaptoethanol). Elution buffer (lysis buffer containing 250 mM imidazole, pH 8.0, but no Tween-20) was then passed over the column and four fractions were collected (0.5, 2, 1 and 1 bed volumes). The fractions were analyzed by SDS-PAGE in order to determine purity. Next, the TRX and hexa-histadine fusion partner was removed by the addition of the protease, thrombin (Sigma) (1:1000 dilution of a 1U/µl solution), into the protein preparation during dialysis into lysis buffer, performed using SpectraPor 12-14,000 MWCO dialysis tubing (Spectrum), overnight at 4ºC, with stirring.

The dialyzed sample was analyzed by SDS-PAGE to confirm complete cleavage of the TRX and hexa-histadine fusion partner from the protein of interest. The protein preparation was then passed over a benzamidine-sepharose 4B (Amersham Biosciences) to remove the thrombin. The TRX and hexa-histidine partner was removed by passing the protein solution back over the Ni-NTA column and collecting the flow-through.

After sample purity was assessed by SDS-PAGE, the protein was dialyzed into 25 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM dithiothreitol (DTT), and 30-50% glycerol (in order to concentrate the protein). The final purification step was carried out using a Superdex 200 (Amersham Biosciences) gel filtration column on an Äkta FPLC system at a flow rate of 0.5 ml/min with a pressure limit of 0.5 MPa. The column was prepared by first passing over 10 bed volumes of ddH₂O and then the same volume of FPLC buffer (25 mM HEPES, pH 7.5, 100 mM NaCl, and 2-5 mM DTT). Fractions were collected, assessed by SDS-PAGE, and those containing proteins of the correct size were pooled. Protein concentration was determined either by the Bio-Rad microassay or based on an A₂₈₀ extinction coefficient calculated for each of the proteins. The protein preparation
was then concentrated to 20-30 mg/ml using Centriprep and Centricon (Millipore) centrifugal filter devices.

**Determination of the CHS Crystal Structure**

CHS crystals were grown by vapor diffusion in hanging drops containing a 1:1 mixture of protein and crystallization buffer (16% PEG 8000 and 0.1 M TAPS pH 7.7) at 4°C. The crystals were stabilized in 18% PEG 8000, 0.1M TAPS, pH 7.7, and 15% glycerol. CHS crystals grew in the space group P2₁ with 4 monomers per asymmetric unit cell, with unit cell dimensions of \(a= 55.06\), \(b=139.27\) and \(c=109.60\). Diffraction data were collected from a single crystal mounted in a cryoloop and flash frozen in a nitrogen stream at 105K. Data were collected at beamline 7-1 at the Stanford Synchrotron Radiation Laboratory (SSRL 7-1) on a 30 cm MAR imaging plate detector. All images were indexed and integrated using DENZO and the reflections merged with SCALEPACK (Otwinowski and Minor, 1997). Data reduction was completed using programs in CCP4 (Collaborative Computational Project, 1994) (Table 3.1). Structures were determined using AMoRE (Navaza, 1994) with an *Arabidopsis* CHS homology model (Saslowsky et al., 2000) as the starting structure. Refinement of the protein model was performed with CNS (Brunger et al., 1998). Model building was carried out with the program O (Jones et al., 1991) using SIGMAA-weighted \(|2F_o-F_c|, |F_o-F_c|\) electron density maps (Collaborative Computational Project, 1994). Refinement consisted of iterative cycles of simulated annealing, positional refinement, and B-value refinement. Model quality was assessed with PROCHECK (Laskowski et al., 1993). All structure images
were manipulated with SwissPDB Viewer (Guex and Peitsch, 1997) and rendered with POVRay (www.povray.org).

**Multiple Sequence Alignment of CHS Proteins**

A multiple sequence alignment of 15 representative members of the plant type III PKS superfamily was generated using Clustal W (Thompson et al., 1994) using the Megalign program in DNAStar (Madison, WI). Included in this analysis were CHS from *Arabidopsis* (accession number P13114), alfalfa (P30074), maize (SYZMCC), parsley (S42523), raspberry (AAK15174), snapdragon (P06515), and tea (P48387); acridone synthase 2 from rue (Q9FSC0); bibenzyl synthase from orchid (S71619); coumaroyl triacetic acid synthase from hydrangea (BAA32733); dihydroxynosedylvin-forming stilbene synthase from pine (Q02323); PS from daisy (P48391); trihydroxystilbene synthase sequences from grape (P51070) and peanut (S00334); and valerophenone synthase (O80400) from hops.

**Site-Directed Mutagenesis**

Mutagenesis of CHS was carried out using the pCD1-CHS construct and the QuikChange® system. The sequences of the primers used for F293A, F305A, W306A, and W373A are given in Table 3.2. Mutagenesis was performed with *Pfu* Turbo (Stratagene, La Jolla, CA) under the following conditions: 95°C for 30 s as an initial denaturing step followed by 16 cycles of 95°C for 30 s, 55°C for 1 min, and 68°C for 7
min. Recovery of modified CHS plasmid constructs was as described as above for construction of pCD1. Confirmation of the mutated CHS coding region was by PCR using the appropriate confirmation primers (Table 3.2) and by sequencing.

**Yeast Two-Hybrid Analysis**

The mutated CHS coding regions were amplified using the primers listed in Table 3.2, digested with \textit{XhoI} and \textit{NotI}, and then cloned into the yeast 2-hybrid vectors, pBI-880 (containing the Gal4\textsubscript{1-147} DNA-binding domain) and pBI-881 (containing the Gal4\textsubscript{768-881} transactivation domain) (Kohalmi et al., 1998). The ligation reactions were used to transform \textit{E.coli} DH10b cells by electroporation (Dower et al., 1988). Clones were screened by PCR using the same primers used for amplification. The recombinant plasmids were isolated using an alkaline lysis miniprep method (Ausubel et al., 1989; Birnboim and Doly, 1979) and then used to transform \textit{Saccharomyces cerevisiae} HF7c cells (Feilotter et al., 1994) in pairwise combinations using the method of Gietz and Schiestl (1995). Double transformants were selected on leu-, trp- synthetic dextrose minimal medium (Ausubel et al., 1989). Interactions between fusion proteins were assayed after 4 d at 30º C on leu-, trp-, his- minimal medium containing 5 mM 3-aminotriazole (Sigma).
Small-Angle Neutron Scattering

Protein samples for SANS were purified as described above, then diluted to the desired concentration (0.75, 1, 2, or 4 mg/ml) in FPLC buffer containing 5% glycerol. For $^2$H$_2$O studies, the samples were diluted to the appropriate concentration in FPLC buffer prepared in $^2$H$_2$O (Cambridge Isotopes) from stock solutions, which were made in H$_2$O, (resulting in approximately 9.5% H$_2$O in the final solution), and dialyzed overnight using Slide-A-Lyzer 10,000 MWCO cassettes (Pierce) at 4ºC against $^2$H$_2$O buffer with one buffer change at approximately 3 h.

SANS experiments were performed on the CHRNS NG3 30-m SANS (Glinka et al., 1998) at the NIST Center for Neutron Research (NCNR). The neutron wavelength, $\lambda$, was 5.50 Å with a wavelength spread, $\Delta\lambda/\lambda$, of 0.150. Sample-to-detector distances of 5.0 and 1.5 m were used to obtain a wavevector transfer range of $0.009 \text{ Å}^{-1} = Q = 0.3335 \text{ Å}^{-1}$, where $Q = 4\pi\sin(\theta)/\lambda$ and $2\theta$ is the scattering angle. Data were collected at 25ºC with total protein concentrations ranging from 0.75 to 4 mg/ml.

Neutron intensities were corrected for scattering from the buffers and from background neutrons. Data were placed on an absolute scale by normalizing the scattered intensity to the incident beam flux. Finally, the data were radially averaged to produce scattered intensity, $I(Q)$, vs $Q$ curves. Distance distribution functions, $P(r)$, were calculated using the GNOM program (Semenyuk and Svergun, 1991), along with the radius of gyration, $R_g$, forward scattered intensity, $I(0)$, and the maximum dimension, $D_{\text{max}}$. $R_g$ and $I(0)$ values calculated from Guinier plots (Guinier and Fournet, 1955) were used as a guide to determine the best $P(r)$ function.
**In Silico Docking**

The CHS and CHI structures were docked using the geometric-hashing methodology within the program Protein-Protein Dock (PPD) (Norel et al., 1994; Norel et al., 1995; Norel et al., 1999; Norel et al., 2001). PPD performs rigid body docking of one molecule into another via three steps: generation of a molecule’s critical points by a molecular surface calculation, docking of the molecules using the critical points through a matching algorithm, and evaluation of the solution. Intensities for each of the docked complexes were scored against the SANS 1 mg/ml data set using the program CRYSON (Svergun et al., 1995).

**Molecular Dynamics Simulations**

The CHS-CHI complex that best fit the SANS data was refined by molecular dynamics simulations. The structure was first solvated with water, and sodium ions were added to give the entire system a net charge of zero. Minimization of the solvent and counterions was performed for 200 steps using Amber8 using steepest descent (Pearlman et al., 1995). Molecular dynamics was then performed using the Sander module of Amber8 for a total of 5 ns with a time step of 2.0 fs and a structure being collected every 1 ps. The dynamics simulations were performed using 16 processors on Virginia Tech’s Laboratory for Advanced Scientific Computing and Applications Linux cluster. The resulting structure was again compared against the SANS data as described above. Interatomic distances, hydrogen bonding patterns, and root mean square deviation
(RMSD) values were determined using the Carnal program within Amber7 (Pearlman et al., 1995). An average model was generated by coordinate averaging of the last 100 ps of the dynamics simulation. The solvent and sodium ions were manually stripped from the average structure and the system was minimized again as described above. The final, minimized model was again compared against the SANS H$_2$O and $^2$H$_2$O data.

**Multiple Sequence Alignments of CHS and CHI**

Multiple sequence alignments of CHS and CHI from different *Arabidopsis* ecotypes, members of the *Brassicaceae* family, and diverse plant species were generated using ClustalW (Thompson et al., 1994) using the Megalign program in DNASTar (Madison, WI). The ecotypes included in the CHS alignment are as follows, with abbreviations and accession numbers given in parentheses: *Arabidopsis thaliana* Landsberg erecta (Ler P13114), Maerk (Mrk-0 CAF04429), Kashmir (Kas-1 CAF04430), Llagostera (Li-0 CAF04432), Columbia (Col-0 CAF04433), and Lipowiec (Lip-0 CAF04434). The CHI alignment included the following ecotypes: Landsberg erecta (Ler AAA32766), Canary (Can-0 CAB94969), Champex (Cha-0 CAB94970), Condhara (CAB94972), Granollers (Gran1 CAB94973), Graponne (Grap1 CAB94974), Ibel Tazekka (Ita-0 CAB94975), Kashmir (Kas-1 CAB94976), Mechtshausen (Me-0 CAB94977), Muehlen (Mh-0 CAB94978), Monte (Mr-0 CAB94979), Wilp (Wlp2 CAB94982), Cape Verdi (Cvi-0 CAB94983), Eden (Ed1 CAB94984), Graz (Gr-5 CAB94985), Lulep (Lu1 CAB94986), Perm (Per-1 CAB94987), Richmond (Ri-0 CAB94988), Turk Lake (Tul-0 CAB94989), Yosemite (Yo-0 CAB94990), Rubezhnoe CAB94991) and other species.
(Rub-1 CAB94991), Ruds Vedby (RV8 CAD42211), and Tvärminne (TV5 CAD42220).

The sequence alignment for CHS from the Brassicaceae were A. thaliana (P13114), Aethionema grandiflora (AAF23557), A. lyrata (CAF04461), A. griffithiana (AAF23568), A. halleri (CAF04425), Arabis alpina (Q9SEP4), Cardamine amara (Q9SEP2), Aubrieta deltoidea (AAF23584), Barbarea vulgaris (AAF23583), Arabis turrita (AAF23582), Capsella rubella (AAF23581), Arabis procurrens (AAF23580), Arabis pauciflora (AAF23577), Arabis parishii (AAF23576), Arabis lyallii (AAF23574), Arabis lignifera (AAF23573), Arabis jacquinii (AAF23572), Arabis hirsuta (AAF23571), Halimolobos perplexa (AAF23569), Arabis glabra (AAF23566), Arabis fendleri (AAF23565), Arabis drummondii (AAF23564), and Arabis blepharophylla (AAF23562). Sequences from Brassicaceae CHI family members were A. thaliana (accession number P13114), china aster (CA91930), orange CHS1 (BAA81664), clove pink (CA91923), morning glory CHS1 (BAA75310) and CHS2 (BAA36224), kudzu vine (BAA01075), maize (P24825), alfalfa CHS2 (P30074), petunia (AAF60297), radish (AAB87072), rice (CA61955), and grape (BAA31259). The analysis using CHI proteins included sequences from the following species: A. thaliana (AAA32766), china aster (CA91921), orange (BAA36552), clove pink (CA06202), morning glory (AAB86474), kudzu vine (BAA09795), maize (S41570), alfalfa CHI1 (P28012), petunia CHI1 (ISPJCB) and CHI2 (ISPJA1), radish (AAB87071), rice (AAO65886), and grape (CAA53577).
Results

CHS Structure

In order to solve the crystal structure of *Arabidopsis* CHS, recombinant protein was expressed, purified, assayed for activity, and crystallized. The structure was solved at 1.71 Å resolution with molecular replacement using a homology model for *Arabidopsis* CHS (Saslowsky et al., 2000). The *Arabidopsis* CHS crystal structure (Figure 3.2A) bears substantial similarity to that of *Medicago sativa* CHS2 (pdb id: 1BI5) (Ferrer et al., 1999). When overlaid, the *Arabidopsis* and *Medicago* CHS structures are nearly identical, with the exception of a four amino acid insertion at the N-terminus of the *Arabidopsis* protein (Figure 3.2B). The RMSD between the two proteins is 0.66 Å, indicating very few structural differences. *Arabidopsis* CHS maintains the same aβαβα core motif as other enzymes in the plant type III polyketide synthase superfamily (Ferrer et al., 1999). The amino acids involved in catalysis, substrate specificity, and CoA binding, are all conserved and in a similar location as in the *Medicago* CHS structure.

Experimental Evidence for a Role of Specific Residues in CHS in Assembly of the Flavonoid Enzyme Complex

Ma et al. (2003) were able to identify phenylalanine and tryptophan residues that are grouped together on the surface of a protein as important for interactions through analysis of binding sites using the structure comparison algorithm MUSTA (Leibowitz et
al., 2001a; Leibowitz et al., 2001b). CHS residues F293, F305, W306, and W373 were selected for analysis as these residues are surface-exposed on the same area of CHS. Interestingly, each of these residues is highly conserved among plant type III polyketide synthases. Both F293 and F305 are strictly conserved across 15 representative members of this enzyme super-family, while W306 is conserved in all but soybean, and W373 has substitutions in CHS and two stilbene synthases from pine.

The effects of these amino acid substitutions on the ability of CHS to interact with other flavonoid enzymes were assayed using a yeast 2-hybrid system after alanine substitution using site-directed mutagenesis. This system had previously been used by Burbulis and Winkel-Shirley (1999) to identify interactions among wild-type CHS, CHI, and dihydroflavonol 4-reductase (DFR). In this experiment, each of the CHS variants was fused to either the binding domain (BD) or the transcriptional activation domain (TA) of the yeast GAL4 protein and tested in all pairwise combinations with Arabidopsis CHI and DFR. The wild-type constructs were tested to reproduce the original experiments, showing interactions between the Gal4-TA CHS and Gal4-BD CHI constructs as well as between the Gal4-TA DFR and the Gal4-BD CHS constructs (Table 3.3). Remarkably, each of the four introduced changes appeared to disrupt both of these interactions. These results suggest that F293, F305, W306, and W373 define a surface domain that functions in the interactions of CHS with several other flavonoid enzymes. It is also possible that these substitutions have indirect effects on other surface domains involved in protein-protein interactions. However, there are no indications of significant structural changes from homology modeling and molecular dynamics simulations (data not shown).
Small-Angle Neutron Scattering

SANS was used to examine the structure of the CHS and CHI bi-enzyme system in solution. SANS data were collected for CHS and CHI at 0.75 and 1 mg/ml in either a H$_2$O or $^2$H$_2$O buffer. The resulting scattered intensity curves are shown in Figure 3.4. The data were normalized to a scattered intensity of $I(Q = 0) = 1.0$ so that the shapes of the curves can be more easily compared. In this case, data were collected to a $Q$ value of 0.3335 Å$^{-1}$, which translates into a resolution of approximately 33 Å using the relation $D = 2p/Q_{\text{max}}$ (Jacrot et al., 1976).

Neutron scattering parameters calculated from the scattered intensity curves using both Guinier and distance distribution function, $P(r)$, analyses are shown in Table 3.4. Although there were slight differences, the data from the $P(r)$ analyses generally concurred with the data from the Guinier approximation. Data from the $P(r)$ analysis are considered to be more reliable in comparisons to model structures because $P(r)$ analysis uses the entire $Q$ range to calculate the radius of gyration and forward scattering intensity [$I(0)$] values.

Interestingly, there seemed to be a difference in the makeup of the CHS-CHI complex in H$_2$O and $^2$H$_2$O. From I(0), it is possible to calculate a rough molecular weight estimate by the following equation:

$$Molecular\ Weight = \frac{I(0)N_A}{c(\Delta \rho \nu)^2}$$  \hspace{1cm} \text{Equation 1}$$

where $c$ is the concentration in g/cm$^3$, $\Delta \rho$ is the contrast [scattering length density of the protein minus that of the solvent (for protein in $^2$H$_2$O: -3.40*10$^{10}$ cm$^{-2}$ and for protein in
H$_2$O: 2.36*10$^{10}$ cm$^{-2}$], $\nu$ bar is the partial specific volume (0.73 cm$^3$/g) and I(0) is the intensity at Q=0 (cm$^{-1}$) from the normalized scattering data. When the molecular weight was calculated for the 1 mg/ml data sets, there was a significant difference in the predicted molecular weight of the complex in H$_2$O and $^2$H$_2$O (Table 3.4). It appears that two CHI monomers interact with each CHS dimer in H$_2$O, whereas only one CHI monomer interacts with the dimer in $^2$H$_2$O. $^2$H$_2$O easily exchanges from H on solvent-exposed carbons and nitrogens with $^2$H. This result suggests that the presence of $^2$H on the protein interferes with the interaction, possibly by disrupting either hydrogen bonds or electrostatic bonds between CHS and CHI.

**In Silico Docking and Molecular Dynamics Analysis of the CHS-CHI Complex**

The interaction of CHS and CHI was modeled using the crystal structures of *Arabidopsis* CHS and CHI (J.P.N. unpublished results) with the program Protein-Protein Dock (Norel et al., 1994; Norel et al., 1995; Norel et al., 1999; Norel et al., 2001), which performs a rigid body docking of the two molecules based on the surface shape and charge complementarily. This process produced 55 potential docking solutions that were compared to the SANS intensity data using CRYSON (Svergun et al., 1995). The proposed complex model that best fit the SANS data (Figure 3.5) was then subjected to molecular dynamics simulations to allow for any structural changes resulting from the interaction. The interaction of CHI with CHS in this model did not obscure either the substrate channel or the surface exposed-residues on CHS involved in the binding of
substrate. Interestingly, the location of the CHS-CHI interface did not include the residues that were analyzed in the yeast 2-hybrid experiments.

The molecular dynamics simulation did not identify any significant structural changes at the interaction interface, as indicated by a RMSD value of 2.33 Å for the entire system. In particular, two alpha helices on CHI (a3 and a7) and one on CHI (a11) remained relatively static (Table 3.5). The dynamics simulations identified three salt bridges that are present between charged residues in the two enzymes: between E327 and E328 on CHS and K18 on CHI, between R71 on CHS and E219 on CHI, and K361 of CHS and E87 CHI (Figures 3.9 and 3.10). The distances between the alpha carbon atoms remained fairly constant over the last 100 ps of the simulation, with standard deviations no greater than 0.31 Å, and, in most cases, the residues moved closer to each other during the simulation (Table 3.5). The model also predicted a hydrogen bond between T69 of CHS and Q222 on CHI (Figures 3.9 and 3.10). To further investigate the function of these residues in the CHS-CHI interaction, each is currently being altered to alanine for analysis in the yeast 2-hybrid system.

CHS and CHI each exhibited one relatively mobile region throughout the dynamics simulations. In CHS, this region corresponded to the first 13 residues, which is solvent exposed in the model. The most mobile regions in the entire complex were sheets β4 and β5 of CHI (Figure 3.6). Over the last 1 ns of simulation, the alpha carbon of residue G53, a residue located in the turn between β4 and β5, moved 5.61 Å (Figure 3.7) with a RMSD of 2.43 ± 0.682 Å. Interestingly, β4 and β5 help form the naringenin binding cleft and include residues involved in the active site hydrogen bonding network.
(Jez et al., 2000). It is possible that the presence of bound substrate stabilizes these two sheets.

**Conservation of Residues Involved in the CHS-CHI Interaction**

In order to investigate conservation of residues predicted by these experiments to be involved in the CHS-CHI interaction, multiple sequence alignments were generated using representative CHS and CHI sequences from different ecotypes of *Arabidopsis thaliana*, members of the *Brassicaceae* family, and more distantly-related plant species for which sequences of both enzymes are known (Figure 3.10). For both CHS and CHI there is absolute conservation of these residues as well as the surrounding sequences among proteins from the various ecotypes of *Arabidopsis* (Figure 3.10). However, only two of the residues in CHS, -R71 and -E328, are absolutely conserved among members of the *Brassicaceae* family. In other *Brassicaceae* species, CHS-E327 is an alanine. Interestingly, CHS-K361 is conserved in 14 of the species surveyed, and the other species contain a glutamic acid at this position, which has the opposite charge (Figure 3.10). Unfortunately, the corresponding CHI sequence is not available for these species to determine whether compensatory changes have occurred in the other enzyme.

Investigation of the sequences from distantly-related plant species indicated many differences between CHS and CHI at the proposed interaction interface. Across the CHS species, two conserved residues were observed, R71 and E328, with few non-conserved substitutions (Figure 3.10). Among the CHI proteins, the only proposed interaction
residue that appears to be conserved is E87. China aster contains the only significant
difference, with an alanine at this position; in all other species surveyed, position 87
retains a negative charge (Figure 3.10C).

Discussion

This study presents an experimental model for the interaction of CHS and CHI,
which catalyze the first two enzymatic reactions in flavonoid biosynthesis. Previous work
had implicated specific surface-exposed phenylalanine and tryptophan residues as
‘hot spots’ that are critical for protein-protein interactions (Ma et al., 2003). However,
from the data presented here, it appears that the surface-exposed phenylalanines and
tryptophans on CHS that were tested are not directly involved in the interaction between
CHS and CHI, but may play an indirect role by maintaining the structure of adjacent
domains that function in the interactions.

The model presented here encompasses only the first two enzymes of the
flavonoid pathway. However, many other questions relating to the flavonoid metabolon
remain to be answered, including the organization of the entire complex both \textit{in vitro} and
\textit{in vivo} as well as the effects of substrates and end products on its stability and
composition. The next step in the characterization of the flavonoid metabolon is to
determine how other flavonoid enzymes interact with the CHS-CHI bienzyme complex.
Previous studies from our laboratory indicate that DFR and F3H also interact with CHS
(Burbulis and Winkel-Shirley, 1999) as part of the flavonoid metabolon. Further
structural and SANS studies could help elucidate how these proteins interact with each
other. Surface plasmon resonance refractometry is a relatively new technology that could be used to determine the kinetics of protein association and dissociation. Amide hydrogen/deuterium exchange-MS experiments could also give information on both the rates of association and dissociation as well as the solvent-accessible area present on a protein (Mandell et al., 2001). Once more is known about the complex as a whole, it might then be possible to alter the flux of the pathway to lead to the production of novel flower coloration or more nutritious plants.

SANS is strictly an *in vitro* technique, and *in vivo* verification of potential models is required. Validation of a SANS-developed model can be performed by site-directed mutagenesis and yeast two-hybrid analysis as described here. Another way to validate a model of such a complex could be through tandem affinity purification (TAP) tagging. Recently, TAP tagging, coupled with mass spectrometry, has been used to identify protein complexes present in yeast (Gavin et al., 2002; Graumann et al., 2004), mammals (Brajenovic et al., 2004; Knuesel et al., 2003), and plants (Rohila et al., 2004) and could easily be applied to the flavonoid metabolon.

The SANS experiments presented herein investigated two proteins in an environment that is not crowded, unlike the environment present in the cell. Future SANS experiments should be performed in the presence of a crowding agent, such as polyethylene glycol (PEG), to provide a crowded environment, similar to the cell. PEG has been shown previously to serve as a suitable crowding agent in studying metabolic flux (Rohwer et al., 1998) and interactions (Kozer and Schreiber, 2004), however, there are no examples of PEG being used as a crowding agent for specific SANS experiments.
Currently, little is known about the structure of metabolic enzyme complexes. One of the few structural models of such organization is of enzymes involved in the citric acid cycle (Vélot et al., 1997). The work presented here represents the first effort to use SANS in the determination of the structure of a metabolic enzyme complex. SANS coupled with \textit{in silico} docking algorithms provided a powerful approach for predicting the structure of the CHS-CHI bienzyme complex, which is now being tested experimentally. These techniques are likely to find applications in the study of other metabolic complexes where at least some structural information is available.
Acknowledgements

The authors would like to thank Mike Austin and Marianne Bowman of the Noel lab for assistance with CHS purification and crystallization, the staff at the NIST Center for Neutron Research for support with the SANS experiments, and Dr. David Bevan and Jory Zmuda Ruscio for assistance with the molecular dynamics simulations. This work was supported by the National Science Foundation (grants MCB-9808117 and MCB-0131010) and by a grant from the Virginia Tech Graduate Research Development Project to CDD. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program. We also acknowledge the National Institute of Standards and Technology, U.S. Department of Commerce, in providing facilities used in this work which are supported by the National Science Foundation under Agreement No. DMR-9986442.
**References Cited:**


isomerase in tomato results in fruit containing increased levels of flavonols.


Arabidopsis seedlings. Coordinate regulation with chalcone synthase and

characterization of the Krebs tricarboxylic acid cycle metabolon. J Biol Chem
262, 1786-1790.

purification tag and methods for isolation of protein heterocomplexes from plants.
Plant J 38, 172-181.

Implications of macromolecular crowding for signal transduction and metabolite
channeling. Proc Natl Acad Sci USA 95, 10547-10552.

metabolism in Arabidopsis using a phage-derived antibody. Molecular Breeding,
in press.


25, 459-486.


Table 3.1: Crystallographic data

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>P2₁</td>
</tr>
<tr>
<td>Unit Cell dimensions (Å)</td>
<td>$a = 55.06$, $b = 139.27$, $c = 109.60$</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.08</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>29.387 - 1.710</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>152185</td>
</tr>
</tbody>
</table>
Table 3.2: Primers used for PCR and site directed mutagenesis for yeast 2-hybrid constructions

A: Mutagenesis Primers (sections underlined indicate the location of the mutation):

- **F293A (sense):** 5'-GAAGAGTCTAGACGAAGCGGGCAAAACCTTTGGGATAGTGAC-3'
- **F293A (antisense):** 5'-GTCACTTATCCCCAAAGGTGTTCGCCCTGCTTAGACTCTTC-3'
- **F305A (sense):** 5'-TGACTGGAACTCCCTCGCAGCTGGATAGCCCACCCTG-3'
- **F305A (antisense):** 5'-CAGGGTGGGCTATCCATGCGAGGGAGTTCCAGTCA-3'
- **W306A (sense):** 5'-CTGGAACTCCCTTTGCAATAGCCCACCCTGGAG-3'
- **W306A (antisense):** 5'-CAGGAGAAGGGTTGGAGGCAAGTGTCTTGGTGGTTTCG-3'
- **W373A (sense):** 5'-CAGGAGAAGGGTTGGAGGCAAGTGTCTTGGTGGTTTCG-3'
- **W373A (antisense):** 5'-CGAAACCAAACAAAGACACCTGCCCTCAAACCCTTCTCCTG-3'

B: CHS amplification primers for yeast 2-hybrid constructs:

- **CHS forward primer:** 5'-CCGCTCGAGCATGGTGATGGCTGGTGC-3'
- **CHS reverse primer:** 5'-AGCGGCCGCTTAGAGAGGAACGCTGTG-3'

C: Confirmation primers:

- **F293A confirmation:** 5'-CCCCAAAGGTGGTTAGC-3'
- **F305A confirmation:** 5'-GGGTGGGCTATCCATGC-3'
- **W306A confirmation:** 5'-CGAAACCAAACAAAGACACCTGCCCTCAAACCCTTCTCCTG-3'
- **W373A confirmation:** 5'-CGAAACCAAACAAAGACACCTGCCCTCAAACCCTTCTCCTG-3'
Table 3.3: Yeast 2-hybrid analysis of amino acid substitutions.

<table>
<thead>
<tr>
<th>Gal4-DB construct</th>
<th>Gal4-TA construct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CHS</td>
</tr>
<tr>
<td>CHS</td>
<td>+/-</td>
</tr>
<tr>
<td>CHI</td>
<td>+++</td>
</tr>
<tr>
<td>DFR</td>
<td>+/-</td>
</tr>
<tr>
<td>CHSF293A</td>
<td>N/T</td>
</tr>
<tr>
<td>CHSF305A</td>
<td>N/T</td>
</tr>
<tr>
<td>CHSW306A</td>
<td>N/T</td>
</tr>
<tr>
<td>CHSW373A</td>
<td>N/T</td>
</tr>
</tbody>
</table>

+++ : Growth of yeast cells
+/-, - : No growth of yeast cells
N/T: Not tested
Table 3.4
Neutron scattering parameters for CHS and CHI in solution

<table>
<thead>
<tr>
<th></th>
<th>Gunier analysis</th>
<th>P(r) Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R_g$ (Å)</td>
<td>I(0)</td>
</tr>
<tr>
<td>CHS/CHI, H$_2$O (1 mg/ml)</td>
<td>28.57 ± 0.827</td>
<td>0.0838</td>
</tr>
<tr>
<td>CHS/CHI, H$_2$O (0.75 mg/ml)</td>
<td>30.99 ± 1.188</td>
<td>0.0380</td>
</tr>
<tr>
<td>CHS/CHI, D$_2$O (1 mg/ml)</td>
<td>28.97 ± 0.242</td>
<td>0.120</td>
</tr>
<tr>
<td>CHS/CHI, D$_2$O (0.75 mg/ml)</td>
<td>28.47 ± 0.330</td>
<td>0.0615</td>
</tr>
</tbody>
</table>

Calculated molecular weigh of the CHS/CHI complex$^a$

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CHS/CHI, H$_2$O (1 mg/ml)</td>
<td>174 kDa</td>
</tr>
<tr>
<td>CHS/CHI, D$_2$O (1 mg/ml)</td>
<td>117 kDa</td>
</tr>
</tbody>
</table>

Predicted weight for complex composed of 1 CHS homodimer and 2 CHI monomers 140 kDa

Predicted weight for complex composed of 1 CHS homodimer and 1 CHI monomer 113 kDa

$^a$Molecular weights from SANS data were calculated using equation 1.
Table 3.5: RMSD and distance measurements of secondary structure and residues proposed to be involved in the CHS-CHI interaction.


<table>
<thead>
<tr>
<th></th>
<th>CHS a1</th>
<th>CHS β1</th>
<th>CHS β3</th>
<th>CHI a7</th>
<th>CHI a3</th>
<th>CHI β1</th>
<th>CHS a1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average of last</td>
<td>1.72</td>
<td>0.911</td>
<td>1.22</td>
<td>1.04</td>
<td>1.75</td>
<td>0.927</td>
<td>3.32</td>
</tr>
<tr>
<td>100 ps</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.04)</td>
<td>(0.09)</td>
<td>(0.07)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>Average of last</td>
<td>1.56</td>
<td>0.852</td>
<td>1.34</td>
<td>0.991</td>
<td>1.56</td>
<td>0.934</td>
<td>3.43</td>
</tr>
<tr>
<td>1 ns</td>
<td>(0.13)</td>
<td>(0.10)</td>
<td>(0.14)</td>
<td>(0.06)</td>
<td>(0.18)</td>
<td>(0.11)</td>
<td>(0.23)</td>
</tr>
</tbody>
</table>

(B): Distance averages of proposed residues involved in the CHS-CHI interaction.

<table>
<thead>
<tr>
<th>CHS residue #</th>
<th>E327 CD</th>
<th>E327 CA</th>
<th>E328 CD</th>
<th>E328 CA</th>
<th>T69 OG1</th>
<th>T69 CA</th>
<th>R71 NH1</th>
<th>R71 CA</th>
<th>K361 NZ</th>
<th>K361 CA</th>
<th>K361 NZ</th>
<th>K361 CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>and atom type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHI residue #</td>
<td>K18 NZ</td>
<td>K18 CA</td>
<td>K18 NZ</td>
<td>K18 CA</td>
<td>Q222 CD</td>
<td>Q222 CA</td>
<td>E218 CA</td>
<td>E218 CA</td>
<td>T86 OG1</td>
<td>T86 CA</td>
<td>E87 OE2</td>
<td>E87 CA</td>
</tr>
<tr>
<td>and atom type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average over last</td>
<td>3.63</td>
<td>9.00</td>
<td>3.31</td>
<td>9.77</td>
<td>5.12</td>
<td>8.19</td>
<td>3.98</td>
<td>10.6</td>
<td>2.88</td>
<td>9.90</td>
<td>2.78</td>
<td>11.3</td>
</tr>
<tr>
<td>100 ps of simulation</td>
<td>(0.31)</td>
<td>(0.29)</td>
<td>(0.13)</td>
<td>(0.29)</td>
<td>(0.42)</td>
<td>(0.26)</td>
<td>(0.37)</td>
<td>(0.31)</td>
<td>(0.11)</td>
<td>(0.20)</td>
<td>(0.12)</td>
<td>(0.19)</td>
</tr>
</tbody>
</table>
Figure 3.1: Schematic of the flavonoid pathway in *Arabidopsis thaliana*. The pathway creates a diverse set of compounds used in UV protection, signaling, defense, and pigmentation. CHS represents that first committed step in the pathway, where 4-coumaroyl-CoA (from phenylpropanoid biosynthesis) and 3 units of malonyl-CoA (from fatty acid metabolism) are used as substrates. Structures in the figure represent the three enzymes in the pathway that crystal structures have been solved for, CHS (pdb id:1BI5) and CHI (pdb id:1EYQ) (both from *Medicago sativa*), and anthocyanidin synthase (ANS) (pdb id:1GP6) (from *Arabidopsis thaliana*). Structures in orange are homology models of dihydroflavonol 4-reductase (DFR), flavanone 3-hydroxylase (F3H), flavonoid 3′ hydroxylase (F3′H), and flavonol synthase (FLS). Other enzyme names are abbreviated as follows: anthocyanidin reductase (ANR), cinnamate 4-hydroxylase (C4H), p-coumarate:CoA ligase (4CL), leucoanthocyanidin reductase (LAR), phenylalanine ammonia-lyase (PAL). Arrows labeled in grey indicate branches of the pathway not present in *Arabidopsis*.
Figure 3.2: CHS crystal structure. A. Structure of the CHS dimer. Each monomer is labeled in a different color. B. Overlay of *Medicago sativa* CHS2 (gray) with *Arabidopsis thaliana* CHS (purple).
Figure 3.3: Residues on CHS tested for effects on interactions with other flavonoid enzymes by yeast 2-hybrid analysis. The front (a) and side (b) views. The surface exposed phenylalanines and tryptophans selected for alanine substitutions are labeled in blue and the active site residues are labeled in red.
Dana, et al. Figure 3.4
Figure 3.4. Normalized SANS data. The CHS/CHI data in either H$_2$O (A) or $^2$H$_2$O (B) solvent. Normalized distance distribution function, $P(r)$, for the CHS-CHI complex in H$_2$O (C) or $^2$H$_2$O (D) solvent. The CHS-CHI data at 0.75 mg/ml is labeled as x and the CHS-CHI data at 1.0 mg/ml is labeled as ?.
Figure 3.5: Data fit of the Docked CHS-CHI Complex. The optimized structure was fit to the 1 mg/ml data set from the SANS analysis. Data fit was performed using CRYSON (Svergun, et al., 1995).
Figure 3.6: Proposed structure of the CHS-CHI enzyme complex. The CHS homodimer is represented in purple, CHI is in yellow, active site residues are labeled in red and the residues involved in CoA binding are labeled in blue. (A) front view, (B) view from 45° on the left, (C) side view, and (D) top view.
Figure 3.7: Movement of beta sheets β3a and β3b of CHI at 100 ps intervals over the last 500 ps of dynamics simulation. Color labels are as follows, red: 4500 ps, orange: 4600 ps, yellow: 4700 ps, green: 4800 ps, blue: 4900 ps, and violet: 5000 ps. Over the course of the dynamics simulations the residue on the loop connecting the two beta sheets, glutamine 53 moves 5.61 Å.
Figure 3.8: Residues proposed to participate in the CHS-CHI interaction. (a) front view and (b) top view. Residues contributed to the interaction by CHS are labeled in blue and the residues contributed by CHI are labeled in orange. In all of the structures, the active site residues are labeled in red.
Figure 3.9: Close up of the interaction interface between CHS and CHI from the (a) front and (b) top. CHS is in purple, CHI is in yellow. The residues proposed to be important in this interaction are labeled with: basic residues in blue, acidic residues in red, and polar residues are in green.
Dana, et al. Figure 3.10A
<table>
<thead>
<tr>
<th>Species</th>
<th>CHS Sequence</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td>CDKSTIRKRHM..................LGLKEEKMRA..................RKSAKDGV...</td>
<td>100%</td>
</tr>
<tr>
<td>Aethionema grandiflora</td>
<td>V..............................................K...RE...AV...</td>
<td>89.0%</td>
</tr>
<tr>
<td>Arabidopsis lyrata</td>
<td>-..............................M..............................</td>
<td>98.0%</td>
</tr>
<tr>
<td>Arabidopsis griffithiana</td>
<td>M..............................M..............................</td>
<td>99.0%</td>
</tr>
<tr>
<td>Arabidopsis halleri</td>
<td>-..............................M..............................</td>
<td>98.5%</td>
</tr>
<tr>
<td>Arabis alpine</td>
<td>M..............................................K...........................................</td>
<td>96.7%</td>
</tr>
<tr>
<td>Cardamine amara</td>
<td>....................................................K...........................................</td>
<td>95.7%</td>
</tr>
<tr>
<td>Aubrieta deltoidea</td>
<td>M..............................A..K...K...EE...A...</td>
<td>93.2%</td>
</tr>
<tr>
<td>Barbarea vulgaris</td>
<td>M..............................A....K...................................................</td>
<td>94.9%</td>
</tr>
<tr>
<td>Arabis turrita</td>
<td>M..............................................K...........................................</td>
<td>93.4%</td>
</tr>
<tr>
<td>Capsella rubella</td>
<td>M..............................................K...........................................</td>
<td>98.5%</td>
</tr>
<tr>
<td>Arabis procurrens</td>
<td>M..............................A....K...VE...A...</td>
<td>93.7%</td>
</tr>
<tr>
<td>Arabis pauciflora</td>
<td>M..............................................A...E..A...</td>
<td>89.4%</td>
</tr>
<tr>
<td>Arabis parishii</td>
<td>M..............................................A...........................................</td>
<td>98.0%</td>
</tr>
<tr>
<td>Arabis lyallii</td>
<td>M..............................................A...........................................</td>
<td>98.0%</td>
</tr>
<tr>
<td>Arabis lignifera</td>
<td>M..............................................A...........................................</td>
<td>97.7%</td>
</tr>
<tr>
<td>Arabis jacquinii</td>
<td>M..............................................K...VE...A...</td>
<td>93.4%</td>
</tr>
<tr>
<td>Arabis hirsute</td>
<td>M..............................................K...VE...A...</td>
<td>93.7%</td>
</tr>
<tr>
<td>Halimolobos perplexa</td>
<td>M..............................................K...........................................</td>
<td>98.7%</td>
</tr>
<tr>
<td>Arabis glabra</td>
<td>M..............................................K...........................................</td>
<td>98.2%</td>
</tr>
<tr>
<td>Arabis fendleri</td>
<td>M..............................................A...........................................</td>
<td>98.0%</td>
</tr>
<tr>
<td>Arabis drummondii</td>
<td>M..............................................A...........................................</td>
<td>97.5%</td>
</tr>
<tr>
<td>Arabis blepharophylla</td>
<td>M..............................................K...VE...A...</td>
<td>93.2%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>CHI Sequence</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td>PAVTKLHVD......................RGKTTHELTE...........................................</td>
<td>100%</td>
</tr>
<tr>
<td>Arabidopsis lyrata</td>
<td>S..............................................S...........................................</td>
<td>89.4%</td>
</tr>
</tbody>
</table>

Dana, et al. Figure 3.10B
Figure 3.10: Multiple sequence alignments of *Arabidopsis* CHS and CHI protein sequences against other ecotypes of *Arabidopsis thaliana* (A), members of the *Brassicaceae* family (B), and different plant species. A dot represents the same amino acids as the Landsberg erecta ecotype of *Arabidopsis thaliana*. Residues proposed to be important for the interaction are shaded and dashed lines indicate the interacting pairs of residues. Boxes indicate conservative changes. Percentages to the right of the sequences refer to total amino acid identity relative to the *Arabidopsis* proteins.
Chapter 4

Conclusions
The organization of metabolism has become a focus of study over the past 40 years. In our laboratory, the *Arabidopsis* flavonoid biosynthetic pathway has been developed as a model system for the study of metabolic organization. More than 30 years ago, Helen Stafford suggested that enzymes in flavonoid biosynthesis are organized as a metabolon (Stafford, 1974a; Stafford, 1974b), and subsequent experiments investigating both the subcellular localization, channeling and co-fractionation (Hrazdina, 1992; Winkel-Shirley, 1999) and the direct interactions of the enzymes (Burbulis and Winkel-Shirley, 1999) have supported her hypothesis. This work laid the foundation for the structural studies of the flavonoid complex described herein, the goals of which were two-fold: 1) to structurally characterize an allelic series of the gene encoding chalcone synthase (CHS), the first enzyme of the flavonoid pathway, using homology modeling and molecular dynamic simulations, and 2) to determine the molecular basis of the complex formed by interaction between CHS and the second flavonoid enzyme, chalcone isomerase (CHI).

Computational analyses contributed significantly to this project. *In silico* techniques were used to develop structural models for a series of CHS variants that had been characterized on the gene, protein and endproduct levels (Saslowsky et al., 2000). The single amino acid substitutions or truncation characterizing these variants result in proteins with altered activity, stability, and dimerization ability. These computational techniques provide an efficient way to predict the effect of these changes on local and global protein structure without solving the crystal structure. However, this methodology yielded few insights into the molecular basis of the temperature-sensitive variants.
Coupling these *in silico* techniques with quantum mechanical algorithms should aid in the determination of catalytic mechanisms leading to the engineering of proteins with altered activities. Previous molecular dynamics and quantum mechanics studies into the reaction mechanism of *Medicago* CHI demonstrated that the substrate, tetrahydroxchalcone, undergoes significant conformational changes during catalysis to naringenin (Hur et al., 2004). These same types of computational approaches could be applied in preliminary engineering efforts to determine the catalytic or structural roles of specific residues.

For the first time, a three-dimensional model for the interactions between enzymes involved in flavonoid metabolism has been developed. Using a series of both experimental and computational techniques, we were able to determine the structure of CHS, suggest a model for the interaction between CHS and CHI, and test the residues predicted to be important for the interaction using site-directed mutagenesis and yeast two-hybrid analysis. Surface plasmon resonance refractometry (SPR) showed promise as a way to determine the binding affinities of the CHS-CHI complex (C. Dana and B. Winkel, unpublished data) and could be expanded to other interacting enzymes in the pathway.

The next important step will be to expand these analyses to other enzymes in the pathway in order to obtain a complete picture of the flavonoid metabolon, both on a biochemical and structural level. Structures for most of the other flavonoid enzymes have been determined either through X-ray crystallography [CHS, CHI, and anthocyanidin synthase (Wilmouth et al., 2002)] or by homology modeling [flavanone 3-hydroxylase, flavonol synthase, and dihydroflavonol 4-reductase (C. Dana, D. Owens, and B. Winkel, unpublished data)] and could be used in SANS studies to piece together
the overall structure of the complex. However, flavanone 3’-hydroxylase (F3’H), the proposed membrane anchor, would be more difficult to include in this type of analysis due to its N-terminal transmembrane domain. An alternative approach that may be useful in this situation is neutron reflectivity, a technique similar to SANS that uses neutrons to probe membrane-associated molecules. As with SANS, a structural picture could be gained by docking the protein structures using \textit{in silico} algorithms followed by molecular dynamics simulations. Briefly, F3’H would be expressed in mammalian cells and purified with the transmembrane domain incorporated into a membrane bilayer. The purified protein could then be subjected to neutron reflectivity to confirm initial structure of F3’H to that of a previously generated homology model (C. Dana and B. Winkel, unpublished data). Adding the other flavonoid enzymes sequentially to the immobilized F3’H could then be used to help build a general model of the metabolon. There are, as yet, no examples of neutron reflectivity being used in this manner. However, a new instrument at the NIST center for neutron research, the advanced neutron diffractometer/reflectometer, has been designed to be used for investigations into the structure of membrane-bound proteins as well as biological membrane structure. This type of analysis could lead to the definition of the structure of the complex as well as the channels available for substrate movement between active sites and determining the molecular basis of the interaction between proteins in the metabolon.

The experiments presented herein represent a significant leap forward in the understanding of the structure of the flavonoid metabolon. The structure of the CHS-CHI provides, for the first time, details about the interfaces required for the interaction between these two cooperating proteins. Once the residues required for the interaction
are verified by *in vitro* assays, it might then be possible to start metabolic engineering to lead to novel flavonoid production in *Arabidopsis*, such as the production of specific pigments. To rationally design a change in pathway flux or endproduct levels, it is imperative that the structure and biochemistry of the remaining parts of the pathway be determined. This information could lead to eventual engineering efforts of the flavonoid pathway in other plants to produce more nutritious plants or flowers with novel pigmentation.

The CHS-CHI bienzyme model helped to establish SANS as a technology through which the structure of a metabolic complex can be determined. To our knowledge, this investigation was the first application of SANS in the study of metabolic complex structure and assembly, for which little is known. Our findings suggest that SANS offers a powerful approach in characterizing multi-enzyme systems and should have applications in other systems where little is known about the organization of the complex, adding to the paltry knowledge of metabolon organization. The data gained from these experiments can also be used to gain more information on the structure of metabolic complexes and then later, in metabolic engineering projects to alter flux within a pathway in such a way to alter substrate flux or create novel endproducts beneficial to animals.
References Cited:

*Arabidopsis* flavonoid biosynthetic pathway. *Proc Natl Acad Sci U S A* 96,
12929-12934.

Metabolism in Plants, H. A. Stafford, ed.

acid catalysis, water expulsion, and enzyme reorganization in Medicago savita


Stafford, H. A. (1974b). Possible multi-enzyme complexes regulating the formation of
C₆-C₃ phenolic compounds and lignins in higher plants. *Rec Adv Phytochem* 8,
53-79.

Wilmouth, R. C., Turnbull, J. J., Welford, R. W., Clifton, I. J., Prescott, A. G., and
Schofield, C. J. (2002). Structure and mechanism of anthocyanidin synthase from
*Arabidopsis thaliana*. *Structure (Camb)* 10, 93-103.

Winkel-Shirley, B. (1999). Evidence for enzyme complexes in the phenylpropanoid and
flavonoid pathways. *Physiol Plant* 107, 142-149.
Appendix A

Homology Modeling of *Arabidopsis* flavonoid enzymes.
Homology models were generated for flavonone 3-hydroxylase, flavonoid 3’-hydroxylase, and dihydroflavonol 4-reductase based on the crystal structures for *Arabidopsis* anthocyanidin reductase (Turnbull et al., 2004; Wilmouth et al., 2002), human cytochrome P450 2C9 (Williams et al., 2003) and *E. coli* UDP-galactose-4-epimerase (Liu et al., 1997), respectively. Alignment of the amino acid sequences was performed using Modeller6 of Sali and Blundell (1993). Five models were produced for each protein and an average model was generated for each protein by coordinate averaging. The averaged structures were minimized using the Sander module of Amber7 for 200 steps using steepest descent (Pearlman et al., 1995). All structure images were manipulated with SwissPDB Viewer (Guex and Peitsch, 1997) and rendered with POV-Ray (www.povray.org).
Figure A.1: Homology models of flavanone 3-hydroxylase (a), flavonoid 3’ hydroxylase (b), and dihydroflavonol 4-reductase (c).
References Cited:


Appendix B

SPR Analysis of Flavonoid Enzymes
Real-time surface plasmon resonance refractometry was performed on a Leica AR600 automatic refractometer equipped with a 10% planar carboxymethyl PEG/PEG mixed self assembled monolayer (CM PEG/PEG m SAM) gold chip (Reichert, Inc). Onto the CM PEG/PEG mSAM chip a NTA moiety was attached by NHS/EDC coupling and was then charged with NiCl₂ (Lahiri et al., 1999) All experiments were carried out at 25°C with a constant flow rate of 0.3 ml/min in PBS-T buffer. The chip was allowed to equilibrate in PBS-T for 30 minutes before the start of the experiment. Upon the start of the experiment, each following solution was passed over the flow cell for 10 minutes each starting first with PBS-T, then 10µM BSA, a PBS-T wash step, next the his-tagged protein to be immobilized is passed over the chip (anywhere from 1-3 µM), another PBS-T wash is performed to remove any unbound protein, the unbound protein to be studied is next passed over the chip (again, anywhere from 1-3 µM), followed by a final wash with PBS-T.

The his-tagged protein to be immobilized onto the slide consisted of thioredoxin-his fusions of chalcone synthase (CHS) or chalcone isomerase (CHI) that were expressed and purified as in Chapter 3 except that the fusion partner was not cleaved. The unbound protein consisted of CHS or CHI which was expressed and purified as in Chapter 3 with cleavage of the fusion partner.
Figure B.1. Representative SPR output. In these experiments either his::TRX::CHI (a) or his::TRX::CHS (b) was immobilized and CHS or CHI, respectively, was passed over the immobilized protein to detect for interactions between the two.
References Cited:

Christopher David Dana
http://main.biol.vt.edu/department/faculty/winkel/chris.htm
709 Earl of Chesterfield Lane 107 Fralin Biotechnology Center
Virginia Beach, VA 23454 Virginia Tech
(540)449-9770 Blacksburg, VA 24061
e-mail: danacd@danaenterprises.org (540)231-9375

Education
Virginia Polytechnic Institute and State University, Blacksburg, VA
Ph.D. Student (expected graduation date: July, 2004)
Major: Biology
QCA: 3.4

James Madison University, Harrisonburg, VA
Majors: Integrated Science and Technology; Biotechnology & Engineering/Manufacturing Concentrations
German
Minor: Music
Graduation Date: May 1999
Cumulative GPA: 3.4

Work Experience and Professional Development
Virginia Tech
Dates: 8/99-12/03
Graduate Teaching Assistant
• Teach upper level molecular biology laboratory class (Fall 2000-present)
• Assisted with development of current molecular biology laboratory manual
• Taught freshman level general and principles of biology laboratory class (Fall 1999-Spring 2000)

PPD Pharmaco (Contract Research Organization)
Dates: 5/98-8/98
2244 Dabney Road
Richmond, VA 23230
• Received GLP training
• Developed database-tracking system for audits
• Developed database for all archived records on site

James Madison University
Dates: 8/96-5/99
• Laboratory technician
• Assisted with teaching of biotechnology labs
• Assisted with development of new laboratory experiments for the junior level biotechnology class
Ondal France E.U.R.L. (Subsidiary of Wella AG)
Dates: 5/97-7/97
2. Rue Dennis Papin - B.P. 70305
57203 Sarreguemines Cedex, France
- Laboratory technician analyzing products for conformance to specifications
- Assisted with ISO 9002 certification efforts

Wella Manufacturing of Virginia, Inc.
Dates: 5/96-7/96
4650 Oakley’s Lane
Richmond, VA 23231
- Assisted with ISO 9002 certification efforts

Extracurricular Activities and Affiliations
- Serving on the Structural Biology Search Committee
- Principal 2nd Violinist, New River Valley Symphony (9/99-present)
- Summer Musical Enterprises, Violinist in performance of “King and I” (5/03-8/03)
- Treasurer, Biology Graduate Student Association (1/00-present)
- Member: Virginia Academy of Science
- Member: Sigma Xi Scientific Honor Society
- Alpha Phi Omega: National Co-ed Community Service Fraternity; Spring, 1999: Chapter parliamentarian
- Golden Key National Honor Society

Presentations
- Structural Analysis of Flavonoid Enzyme Interactions in *Arabidopsis*. Presentation to Dean DellaPenna’s Lab at Michigan State University. October 2003.
- Characterizing the Functional and Structural Domains in Flavonoid Enzymes. NIST Neutron Division. February 2003. (Received small Honorarium for this talk)
• Characterizing the Functional and Structural Domains in Arabidopsis Flavonoid Enzymes. Botany Seminar. April 2001

Posters Presented

• A.S. Duda, C.D. Dana, and B. Winkel. Substrate Specificity of Dihydroflavonol 4-Reductase. ACS undergraduate research symposium. Poster was awarded “Best Poster Presentation.”
• C.D. Dana, D.R. Bevan, and B. Winkel. Homology Modeling of the tt4 Mutants in Arabidopsis thaliana. 17th Annual Research Symposium of Virginia Tech, March 2001. Received 2nd place ($150.00 award)

Publications

Meetings Organized

- Gordon Research Conference “Stage Invasion.” A brief symposium before the start of the GRC for graduate students and post docs to present their research. Queen’s College, Oxford, UK. August 2002. (This is not affiliated with the GRC).

Grants Applied for and Received

- GSA Travel Fund award, $200, with matching funds from department, for travel to the Gordon Research Conference on Macromolecular Organization and Cell Function (August 2002).
- NIST Center for Neutron Research, ~$850, to attend the NIST summer school on Neutron Scattering (June 2002).
- Joint Institute for Neutron Sciences, $1000 ($500 matching from department), to attend workshop on “Using Neutrons to Probe Structure and Dynamics in Biological Systems” at Oak Ridge National Laboratory (April 2002).
- GSA Travel Fund award, $100, Fall 2001, for travel to Phytochemical Society of North America annual meeting in Oklahoma City, OK (August 2001).
- Phytochemical Society Travel Award, $150, for travel to Phytochemical Society of North America annual meeting in Oklahoma City, OK (August 2001).
- Graduate Research Development Project, $500, with matching funds from biology department (Spring 2001).
- GSA Travel Fund award, $300, with matching funds from biology department, Fall 2000, for travel to Gordon Research Conference on Macromolecular Organization and Cell Function (August 2000).
- Graduate Research Development Project, $300, with matching funds from biology department (Spring 2000).

Other

- Supervised the training and work of 4 undergraduate researchers: Ashley Duda (5/03-current), Eric McFadden (5/01-5/03), Elizabeth Tucker (5/00 to 8/00), and Anna Watson (5/01-8/01).
- Redesigned and maintain Winkel Lab web page.

Collaborators

- Dr. Joseph Noel, Salk Institute. Crystallization of Flavonoid Enzymes.
- Dr. Susan Krueger, NCNR, NIST. Solution Structure of the Flavonoid Multienzyme Complex.
• Dr. Craig Nessler, Virginia Tech, PPWS. Homology Modeling of Alkaloid O-Methyltransferases.