The HOTHEAD Protein: Assessing Enzymatic Activity using Computational and Recombinant Protein Expression Approaches

By

Eric Le Dreff-Kerwin

A thesis
Presented to the University of Waterloo
In fulfillment of the
Thesis requirement for the degree of
Master of Science
In
Biology

Waterloo, Ontario, Canada, 2019

Author's Declaration

I hereby declare that I am the sole author of this thesis.

This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In *Arabidopsis thaliana*, a number of genes regulating cuticle synthesis have been identified by virtue of organ fusion phenotype. One such gene, *HOTHEAD (HTH)* was among those originally identified by this phenotype but its exact role in cuticle formation has proven challenging to determine. Previous bioinformatic work has identified that the HTH protein is a member of the GMC oxidoreductase family, and shares peptide sequence identity with a mandelonitrile lyase and an alcohol dehydrogenase that are within the same protein family. This thesis work investigated the potential enzymatic function of HTH by comparing a structural model to two structural analogs. The structure model of HTH, as determined in this study, shows that HTH shares certain conserved features of GMC proteins. The aim of this research also included isolating a recombinantly expressed HTH protein from *Escherichia coli* and initial work in *Pichia pastoris*. Protein isolation attempts in *Escherichia coli* failed to yield active HTH protein, potentially due to the lack of post-translational modifications. Continuing the work using the eukaryote, *Pichia pastoris*, should solve the issue. Finally, crude extracts from *Arabidopsis thaliana* floral tissue were obtained to test for the presence of HTH enzymatic activity but showed no significant mandelonitrile lyase enzymatic activity, potentially due to the inability to successfully extract the HTH protein from the floral tissue.

Acknowledgements

I would like to thank my supervisors, Dr. Susan Lolle and Dr. Todd Holyoak, for all of their help, continuous support and being great examples of great scientists; ones that I always tried to live up to throughout my degree. Even during the most challenging of times, they were always ready to get me back on my feet.

Thank you to my committee members, Dr. Barbara Moffatt and Dr. David Rose, for their valuable comments, and for being continuously willing to help advance this project. To the current and past members of the Lolle and Holyoak labs; Dr. Pearl Chang, Therese Francom, Serena Thamilselvan, Saima Hossain, Saeer Adeel, Jackline Ponomariov and Christina Doherty from the Lolle lab, and Matt McLeod, Iain Wallace, Ben Morrow, Norman Tran. Sarah Barwell, Julia Solonenka and Dr. Bidyadhar Das from the Holyoak lab; as well as Nardo Nava-Rodriguez, Nicole Fraser and the other Rose lab members. Thank you all for your support, even if it was one of the many late nights working in the lab.

Thank you to Quinn Abram and Dr. Tania Rodriguez-Ramos of the Dixon lab for constantly putting up with my random visits to their lab in search for answers and tips on how to improve my technique.

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations and Acronyms

5'UTR: 5' Untranslated region
ACS6: 1-Aminocyclopropane-1-carboxyllate synthase 6
ADE2: Adenine deaminase 2
ADH: Alcohol dehydrogenase
$AmSO_4$: Ammonium sulfate
AOX1: Alcohol oxidase 1
bp: Base pairs
cDNA: Complementary DNA
CYP: Cytochrome P450
Ds: Dissociation element
DCA: Dicarboxylic acid
DNA: Deoxyribonucleic acid
ECL: Enhanced chemiluminescent
eda-17: Embryo development arrest 17
EDTA: Ethylenediaminetetraacetic acid
EMS: Ethyl methanesulfonate
En/Spm: Enhancer/suppressor mutator
ER: Endoplasmic reticulum
EV: Empty vector
FA: Fatty acid
FAD: flavin adenine dinucleotide
FAD-HNL: Flavin adenine dinucleotide dependent hydroxynitrile lyase
FG1: Female gametophyte development stage 1
FG2: Female gametophyte development stage 2
FG3: Female gametophyte development stage 3
FG8: Female gametophyte development stage 8
g: Grams
GFP: Green fluorescent protein
GPAT: Glycerol-3 phosphate acyltransferase
HCN: Hydrogen cyanide
HTH: HOTHEAD
HTH-FP: Fluorescent protein-tagged HTH construct
HTH_SUMO: HOTHEAD_pE_SUMOstar
IDEN: Sequence identity
IgG: Immunoglobulin G
IPTG: Isopropyl β-D-1-thiogalactopyranoside
kd: Kilobase pair
kDa: Kilo Dalton
LACS: Long-chain acyl-coenzyme A synthase
LB: Luria broth
MDL: Mandelonitrile lyase
MEB1: Membrane of ER Body 1
MEB2: Membrane of ER Body 2

mg: Milligram
min: Minutes
ml: Millilitre
mRNA: Messenger ribonucleic acid
ng: Nanogram
nm: Nanometer
OD: Optical density
P-6 DG: Polyacrylamide desalting gel with 6,000 molecular weight exclusion limit
PCR: Polymerase chain reaction
PAGE: Polyacrylamide gel electrophoresis
RMSD: Root-mean-square deviation
SDS: Sodium dodecyl sulfate
TBS-T: Tris-buffer saline with Tween 20
TEV: Tobacco etch virus
TM-Score: Template modeling score
UV: Ultraviolet
VLCFA: Very-long-chain fatty acid
WT: Wild type
YFP: Yellow fluorescent protein
ΔA_{275}: Change of absorbance of light measure at wavelength 275 nm
µg: Microgram
µL: Microlitre
µM: Micromolar
µm: Micrometer
α: Alpha
β: Beta
ω: Omega

DNA Code:
A: Adenine
C: Cytosine
G: Guanine
T: Thymine

Protein Code:
Ala (A): Alanine
Arg (R): Arginine
Asn (N): Asparagine
Asp (D): Aspartic acid
Cys (C): Cysteine
Gln (Q): Glutamine
Glu (E): Glutamic acid
Gly (G): Glycine
His (H): Histidine
Ile (I): Isoleucine
Leu (L): Leucine

Lys (K): Lysine
Met (M): Methionine
Phe (F): Phenylalanine
Pro (P): Proline
Ser (S): Serine
Thr (T): Threonine
Trp (W): Tryptophan
Tyr (Y): Tyrosine
Val (V): Valine

Chapter 1

General Introduction

1.1 The Plant Cuticle

1.1.1 Cuticle Function

Terrestrial plants have acquired a variety of adaptations to aid in their survival on dry land. One such adaptation is the plant cuticle. The cuticle is a hydrophobic membrane that acts as an extra layer separating the aerial tissue of the plant from its dynamic environment. The cuticle protects against ultraviolet radiation, minimizes surface dirt buildup, regulates the flow of gases, nutrients and molecules, and protects against most pathogens (Fich, Segerson, & Rose, 2016; Yeats & Rose, 2013). An interesting exception is a pathogenic attack from *Botrytis cinereal*. *Arabidopsis thaliana* exhibiting a loss in cuticle integrity will be more effective at protecting itself from *B. cinereal* when compared to *A. thaliana* with an intact cuticle (Fich et al., 2016).

The cuticle is also important for proper organ partitioning; in mutants with perturbed cuticle function (or during normal carpel development of some plant species), epidermal cells have been shown to fuse together when they come into direct contact with each other. A water-soluble signaling molecule identified in *Catharanthus roseus* travels between carpel cells separated by a permeable barrier causing cell dedifferentiation (Verbeke & Walker, 1986). In these carpels, altered cuticle integrity appears to allow this signalling molecule to travel between the cells it separates and is a natural process.

1.1.2 Cuticle Composition

The cuticle is a thin multilayered membrane typically varying from 0.1-10 μm in thickness, although it has been observed to be thinner in *A. thaliana*, ranging from 0.02-0.1 μm (Riederer & Schreiber, 2001; Shumborski, Samuels, & Bird, 2016). The cuticle is divided into the cuticular layer and the cuticle proper. The cuticle proper is further divided into the intracuticular wax, the epicuticular wax film and the epicuticular wax crystals. Each layer is

made up of a mix of cutin monomers embedded in wax. Higher levels of cutin can be found in the cuticular layer, while higher levels of wax are in the cuticle proper (Yeats & Rose, 2013). The composition varies between species, making the exact structure of the cuticle difficult to determine. For example, the *A. thaliana* cuticle is known to be dicarboxylic fatty acid (DCA) rich with higher levels of glycerol compared to other species (Kurdyukov et al., 2006). This leads to the belief that the composition of this particular cuticle is different from most other plants (Yang et al, 2016).

1.1.3 Biosynthesis Pathway

Cuticle synthesis occurs in the underlying epidermal cells. The formation of cutin and wax monomers begin in the plastids with *de novo* biosynthesis of C16 and C18 fatty acids (FA). These are converted to CoA thioesters by long-chain acyl-coenzyme A synthase (LACS) and transported into the endoplasmic reticulum (ER) lumen via an unknown mechanism (Yeats & Rose, 2013). Once in the ER, the cuticle biosynthesis pathway divides into the cutin and the wax biosynthesis pathways.

Typical wax monomers include very-long-chain fatty acids (VLCFA) ranging in size from C20 to C34 which include alkanes, alcohols, ketones and wax esters (Yeats & Rose, 2013). To form the wax monomer, the FAs are elongated by fatty acid elongase then reduced to fatty aldehydes and primary alcohols. They can also be reduced and decarbonylated to alkanes, then converted into ketones and secondary alcohols (Yeats & Rose, 2013). The wax monomers are then transported out of the ER, across the cell wall, to the plant tissue surface.

Typical cutin monomers include hydroxy FAs, glycerol, dicarboxylic acids and phenylpropanoids. Terminal and mid-chain oxidation will occur once the C16 and C18 FAs are converted to CoA thioesters by LACS (Yeats & Rose, 2013). Terminal oxidation is believed to

be performed by CYP68A enzymes, coded for by the cytochrome P450 (*CYP*) gene, while mid-chain oxidation is believed to be performed by CYP77A enzymes. Finally, the CoA is replaced with a glycerol-3-phosphate by glycerol-3-phosphate acyltransferase (GPAT) (Fich et al., 2016; Yeats & Rose, 2013).

Various putative mechanisms for cutin and wax monomer export and assembly have been proposed (Pollard, Beisson, Li, & Ohlrogge, 2008). The most recent research suggests that the cutin monomers are transported through plasma membrane embedded ABC transporters, while wax monomers exit the cell at sites of close contact between the ER and the plasma membrane (Fich et al., 2016).

1.2 HOTHEAD

1.2.1 The *HOTHEAD* Gene

A list of known genes involved in various aspects of cuticle formation has been compiled through molecular analyses of *A. thaliana* mutants (Yeats & Rose, 2013). Most genes noted in this list have a known function in cuticle formation, while others have yet to be determined. One of those genes with an unknown function, identified in *A. thaliana*, is *HOTHEAD* (*HTH*) (locus: At1g72970) (Figure 1). This gene maps to chromosome 1 of *A. thaliana* and contains 6 exons and 5 introns. In the Columbia ecotype, the *HTH* gene is 4,843 base pairs (bp) in length. This accounts for the 2834 bp transcribed region and its 2009 bp promoter and 5'-untranslated region (5'UTR) (Krolikowski et al., 2003).

1.2.2 *HOTHEAD* Mutant Alleles

HTH was initially identified through forward genetic screens (Lolle, Hsu, & Pruitt, 1998). *A. thaliana* seeds treated with ethyl methanesulfonate (EMS) were visually screened at various life stages for the presence of organ fusion. This phenotype suggests a loss of cuticle integrity.

This study identified eleven mutant *hth* alleles, *hth-1* through *hth-11*, with varying degrees of floral bud fusion. The relative position of the mutation in each *hth* allele is shown in Figure 1 (Krolikowski et al., 2003). Two of these mutant alleles, *hth-8* and *hth-10* also showed evidence of ovule abnormalities (Lolle et al., 1998). These 11 alleles are a result of a substitution mutation due to the EMS treatment. Most of these mutations alter amino acids in the peptide sequence. Two of these alleles, *hth-1* and *hth-9*, translate into truncated proteins due to the introduction of a stop codon and the alteration of a splice junction sequence, respectively (Lolle et al., 1998).

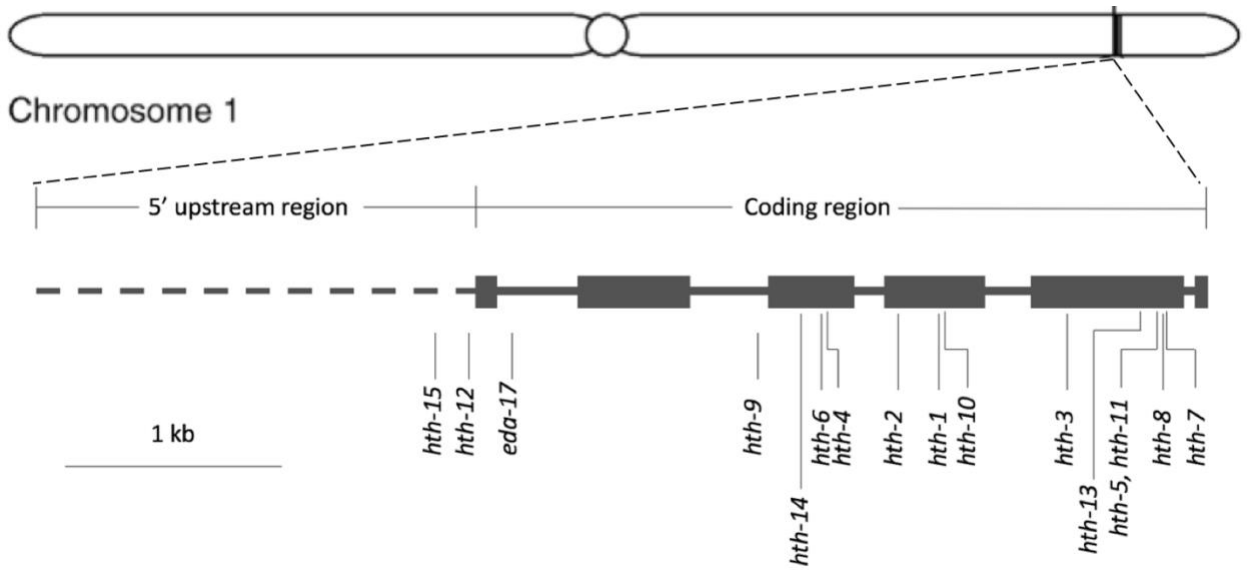


Figure 1: The HTH gene model. The *HTH* gene location on chromosome 1 of *A. thaliana* and the relative position of the various *hth* mutant alleles. The large boxes represent the six exons and the solid lines represent the introns. The position of the single point mutations (*hth-1* to *hth-11*), transposon insertion sites (*hth-12* and *eda-17*) and T-DNA insertion sites (*hth-13* to *hth-15*) are shown (Krolikowski et al., 2003; Kurdyukov et al., 2006). *Based on Figure 1.8 from Chang, 2016.*

In another study performed by Kurdyokov et al., mutant *A. thaliana* seeds were generated by enhancer/suppressor mutator (En/Spm) transposon mutagenesis (Kurdyukov et al., 2006). The *hth-12* mutant was identified via gel blot and hybridization (Figure 1). *A. thaliana* plants harboring the *hth-12* mutant allele showed floral bud organ fusion believed to be due to a loss of

cuticle integrity. Transmission electron microscopy images showed that the cuticle of the *hth-12* mutant was not continuous, as seen in wild type (WT) *A. thaliana* cuticle (Kurdyukov et al., 2006). The composition of the *hth-12* mutant cuticle was compared to the WT Arabidopsis cuticle. The mutant plant cuticle showed an increase in ω-hydroxy FAs and a decrease in ω-oxo FAs. Both of these molecules are precursors to α-,ω-dicarboxylic acids, a known cutin monomer in the plant cuticle. The cuticular wax composition did not appear to be affected by the *hth-12* mutation (Kurdyukov et al., 2006). This change in cutin composition due to the *hth-12* mutant allele links *HTH* to the cutin biosynthesis pathway.

Pagnussat et al. transformed *A. thaliana* plants with a dissociation element (Ds) transposon (Pagnussat et al., 2005) and identified a novel *hth* mutant allele, *embryo sac development arrest-17* (*eda-17*) containing a Ds transposon with the first intro of the *HTH* gene (Figure 1). This is the only *hth* allele with a mutation located within this region. This screen identified numerous mutants with female gametophyte defects. Plant lines containing the Ds transposon were tested for segregation ratio distortion in kanamycin resistance since the Ds element included a kanamycin resistance gene (Pagnussat et al., 2005). Female gametophyte development can be divided into two phases: megasporogenesis and megagametogensis (Drews & Koltunow, 2011). One megaspore is present at the end of megasporogensis and enters megagametogenesis. This process is divided into eight stages, starting with female gametophyte stage 1 (FG1) and finishing with an embryo sac at the end of female gametophyte stage 8 (FG8) (Christensen, King, Jordan, & Drews, 1997). An *A. thaliana* harboring the *eda-17* allele will arrest at female gametophyte stage 2 (FG2) of megagametogenesis, a stage where the nucleus undergoes mitosis without cell division (Christensen et al., 1997). This suggests that the *HTH* gene is important for embryo sac development.

Most *hth* mutants share a similar phenotype, postgenital organ fusion, which is due to a loss in cuticle integrity. This fusion is most prominent in the floral buds, once the plant has reached sexual maturity. It was this phenotype that lead to the suggestion that *HTH* is involved in cuticle synthesis, most probably in the cutin biosynthesis pathway. But other phenotypes suggest that HTH is likely involved in more than just cuticle synthesis. For example, the *eda-17* mutant causes female gametophyte development arrest at the two-nuclear stage (Lolle et al., 1998; Pagnussat et al., 2005).

1.2.3 The HOTHEAD Protein

The HTH protein contains 594 amino acid residues with a predicted molecular mass of 65.3 kDa and an inferred isoelectric point of 9.75 (locus: At1g72970) (Figure 2) (Krolikowski et al., 2003). The protein is predicted to be globular and contain a cleavable N-terminal signal peptide spanning residues 1 to 19 (Chang, 2016).

HTH is considered to be a part of the glucose-methanol-choline (GMC) oxidoreductase protein family, due to the presence of conserved N-terminal and C-terminal domains typical of the family. This family contains a variety of proteins with oxidoreductase enzymatic activity (Cavener, 1992). The GMC oxidoreductase N-terminal domain contains a Rossmann fold, indicating the presence of a flavin adenine dinucleotide (FAD) binding site (Bannwarth et al., 2004). Based on previous bioinformatic work, the HTH protein has a putative FAD binding site between residues 64 to 91 which is found within the predicted GMC N-terminal domain (Chang, 2016). The GMC oxidoreductase C-terminal domain is known as the substrate binding domain. All GMC proteins are predicted to have a conserved histidine active site residue (His466 based on choline oxidase from *Arthrobacter globiformis* (PDB ID: 3nne; EC 1.1.3.17)), while most will have a second active site residue, asparagine (Asn510 in cholesterol oxidase from *A.*

globiformis) (Dreveny et al., 2001). Both of these residues are found within this conserved C-terminal domain (Li et al., 1993). Based on previous bioinformatic work, the HTH protein has three putative active site residues, His529, Ile527 and Asn567, all of which are found within this conserved C-terminal domain (Chang, 2016).

The HTH protein is also predicted to contain certain post-translational modifications: 4 putative phosphorylation sites and 5 putative N-linked glycosylation sites (Chang, 2016). None of the known *hth* mutants are found at the locations of these putative post-translational modifications. HTH has been experimentally shown to be a glycoprotein based on size shifts following endoglycosidase treatment of a HTH::GFP protein extracted from transgenic plant lines (Chang, 2016). Whether these modifications are important for protein structure and/or function has yet to be determined.

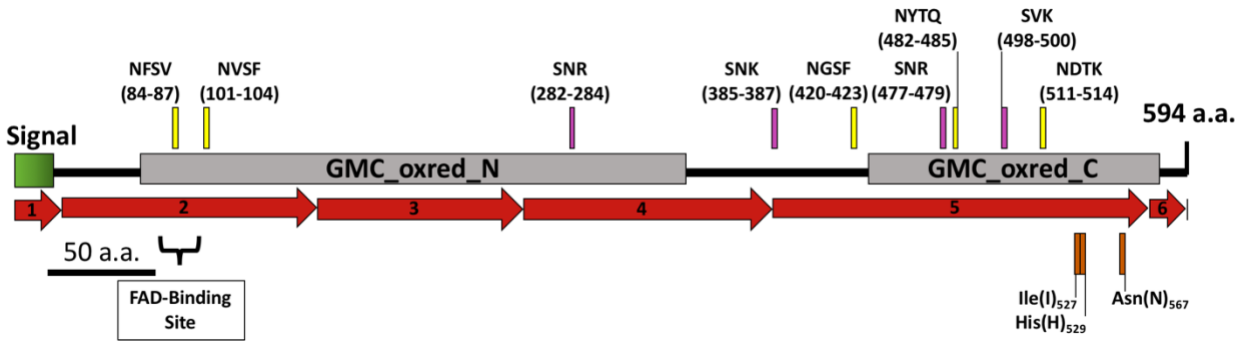


Figure 2: Schematic representation of the HTH protein. The red arrows indicate the six exons of the *HTH* gene. The first 19 residues are predicted to be a part of the signal peptide (green rectangular box). The two conserved GMC oxidoreductase domains are represented by grey rectangular boxes. The conserved N-terminal domain contains the FAD-binding site while the C-terminal domain contains the three putative active site residues (red markers). The relative position of the N-linked glycosylation sites (yellow markers) and phosphorylation sites (purple markers) are indicated. *Based on Figure 4.1 from Chang, 2016*.

1.2.4 HOTHEAD Protein Localization

Chang analyzed HTH localization in *A. thaliana* transgenic plants transformed with a fluorescent protein tagged HTH (Chang, 2016). HTH is first observed in *A. thaliana* ovules at

female gametophyte stage 3 (FG3) of megagametogenesis, prior to fertilization (Chang, 2016). At this stage, two nuclei are present and travel to opposite ends of the embryo sac while a vacuole forms in the middle (Christensen et al., 1997). It is at this stage that HTH is present at the chalazal pole and remains there until anthesis, the point where floral buds open, after which HTH becomes more diffuse (Chang, 2016). Three days after anthesis, HTH localizes in the integument and persists until seed maturity (Chang, 2016).

On a gross level, the HTH mRNA appears to be ubiquitously distributed in *A. thaliana* epidermal and subepidermal cells (Krolikowski et al., 2003). It has been observed in leaves, stems, inflorescences, siliques and roots (Krolikowski et al., 2003). This does not necessarily mean that the HTH protein also localizes in all these tissues since it may move to nearby cells. HTH protein has been detected in seedlings, flowers and siliques (Chang, 2016).

At the cellular level, HTH localizes in the endoplasmic reticulum (ER) within epidermal cells, even though it does not have a canonical KDEL retention signal (Chang, 2016). Instead, it has three KDEL-like sequences in areas predicted to be found on the protein surface: KDEK at position 270, KKEL at position 310 and KNEL at position 387 (Chang, 2016). None of the known *hth* mutants are found within these KDEL-like sequences.

HTH has also been observed in ER-derived structures known as ER bodies (Chang, 2016). These predominantly exist in young *A. thaliana* plants within cotyledons, hypocotyls and roots. The number of ER bodies decreases during senescence from the base to the tip of the cotyledons. Hypocotyl and root ER body levels do not appear to decrease (Matsushima et al., 2003). Rosette leaves of mature plants typically do not have any ER bodies unless the tissue is wounded where upon a large accumulation of ER bodies can be found localized around the damaged tissue.

Besides HTH, ER bodies contain various proteins: PYK10, NAI2, membrane of ER body 1 (MEB1) and membrane of ER body 2 (MEB2) (Yamada et al., 2013). PYK10 is a β-glucosidase which belongs to a group of enzymes known to produce compounds that deter pests, activate glycosylated hormones and break down the cell wall (Matsushima et al., 2003). The exact role of the PYK10 protein in ER bodies is undetermined but it appears to be important for plant defence (Vinther et al., 2008). The MEB1 and MEB2 proteins embed themselves in the ER body membrane and are believed to be involved in metal ion homeostasis (Yamada et al., 2013). This could explain why ER bodies are constitutively expressed in roots, a tissue constantly exposed to high levels of metal ions. Finally, NAI2 is responsible for the accumulation of PYK10, MEB1 and MEB2 in ER bodies (Yamada et al., 2008). A *nai2* mutant will have reduced and elongated ER bodies. PYK10, MEB1 and MEB2 levels will not be affected in this mutant but they will not localize to ER bodies (Yamada et al., 2008).

Based on localization and composition, ER bodies are believed to be important for plant defence (Matsushima et al., 2003). PYK10 most likely aids in deterring pests and breaking down cell walls while MEB1 and MEB2 ensure metal ion levels do not reach toxic levels. The role of HTH in these ER-derived structures is currently unknown.

1.3 Alcohol Dehydrogenase

The potential enzymatic function of the HTH protein has not been confirmed. As previously mentioned, Kurdyukov et al. demonstrated that the cuticle of a *hth-12* mutant *A. thaliana* has an increase in ω-hydroxy FAs and a decrease in ω-oxo FA, a precursor to the α-,ω-dicarboxylic acid cutin monomer (Kurdyukov et al., 2006). This appears to be due to the loss of alcohol dehydrogenase (ADH) enzymatic activity (Figure 3A). Knowing that most *hth* mutants cause a loss in cuticle integrity, it is logical that the HTH protein acts as an ADH in the cutin

biosynthesis pathway. Based on the overall peptide sequence, HTH does share 25% peptide sequence identity to long chain ω-fatty alcohol dehydrogenase from *Candida* (Kurdyukov et al., 2006). *Candida* metabolizes carbon dioxide via a membrane-bound ω-oxidation pathway that involves a flavin-dependent aldehyde dehydrogenase (Vanhanen et al., 2000). The substrate specificity of this enzyme varies. For example, in *C. tropicalis*, the enzyme is most active on long-chain fatty alcohols but can also accept other substrates such as ω-hydroxy FA, secondary alcohols and dodecan-2-ol (Vanhanen et al., 2000).

Most GMC proteins catalyze a redox reaction which involves FAD. The N5 atoms in the flavin interacts with the substrate to donate or accept a hydride anion (Dreveny, Andryushkova, Glieder, Gruber, & Kratky, 2009). The exact conformation of the active site is not strictly conserved due to the variability in substrate size, ranging from cholesterol to small lactates (Fraaije & Mattevi, 2000). GMC proteins have two conserved residues within the active site: His466 and Asn510 (numbering based on choline oxidase from *A. globiformis*) (Yue, Kass, Sampson, & Vrielink, 1999). These residues form a hydrogen bond network, to ensure that the substrate is properly oriented and in proximity with the N5 atom of FAD (Dreveny et al., 2009).

1.4 Mandelonitrile Lyase

On the other hand, the overall peptide sequence of HTH shares 41% peptide sequence identity to mandelonitrile lyase (MDL) from *Prunus* (Kurdyukov et al., 2006). Wu & Poulton identified the localization of MDL in *Prunus* seeds (Wu & Poulton, 1991). MDL was primarily localized to the cell wall with lower amounts present in protein bodies (Wu & Poulton, 1991). In plants, protein bodies are important for protein storage in seeds (Saberianfar et al., 2016; Schmidt, 2013). Thayer & Conn also looked at MDL localization in *Sorghum* leaves and observed it in the cytoplasm of mesophyll cells (Thayer & Conn, 1981).

MDL is an FAD-dependent hydroxynitrile lyase (FAD-HNL) that is a part of the same protein family as HTH, the GMC oxidoreductase protein family (Dreveny et al., 2009). Most proteins in this family perform some sort of oxidation or reduction reaction, aided by the FAD co-factor, specifically the N5 atom in the flavin (Dreveny et al., 2009). FAD-HNLs are an exception since they perform cyanogenesis and their FAD co-factor does not appear to play an enzymatic role. Rather, the FAD seems to be important for the protein's structural integrity as an apo-HNL is found to be inactive, unstable and degrade within minutes (Dreveny et al., 2001). As its name suggests, MDL uses mandelonitrile as its substrate to produce benzaldehyde and hydrogen cyanide (HCN) (Figure 3B). FAD-HNLs are believed to have evolved from a common ancestor shared with aryl-alcohol oxidases, since both proteins are a part of the GMC oxidoreductase protein family and share a similar product, benzaldehyde (Dreveny et al., 2009).

As previously stated, the FAD co-factor in most GMC proteins is used for an oxidoreduction reaction resulting in the hydroxyl group of the substrate being in proximity of the N5 atom of FAD. FAD-HNLs are typically (R)-selective, meaning they exclusively use substrates which are (R)-stereoisomers (Andexer et al., 2012). Due to this substrate specificity in MDL, the hydroxyl group of mandelonitrile will be oriented away from the flavin while the cyano group will be oriented towards the flavin. Therefore, the hydroxyl group is positioned at a distance that is too far to allow hydride transfer with the N5 of the FAD cofactor.

In contrast to the alcohol dehydrogenase members of the GMC family, FAD-HNLs have two conserved histidine residues to facilitate chemistry: His459 and His497 (numbering based on the FAD-HNL peptide from *Prunus dulcis* (PDB ID: 1ju2; EC: 4.1.2.10)) (Dreveny et al., 2009). These residues are found within the active site that is connected to the surface of the protein via a hydrophobic tunnel. The active site is observed to possess a negative electrostatic potential that

is hypothesized to assist in bringing the substrate in proximity to the active site residues (Dreveny et al., 2001). His497 is considered a general base that hydrogen bonds with the hydroxyl group of mandelonitrile to help stabilize the substrate (Dreveny et al., 2001). His459 on the other hand mediates the protonation of the cyanide ion, allowing for the production of hydrogen cyanide (Dreveny et al., 2009).

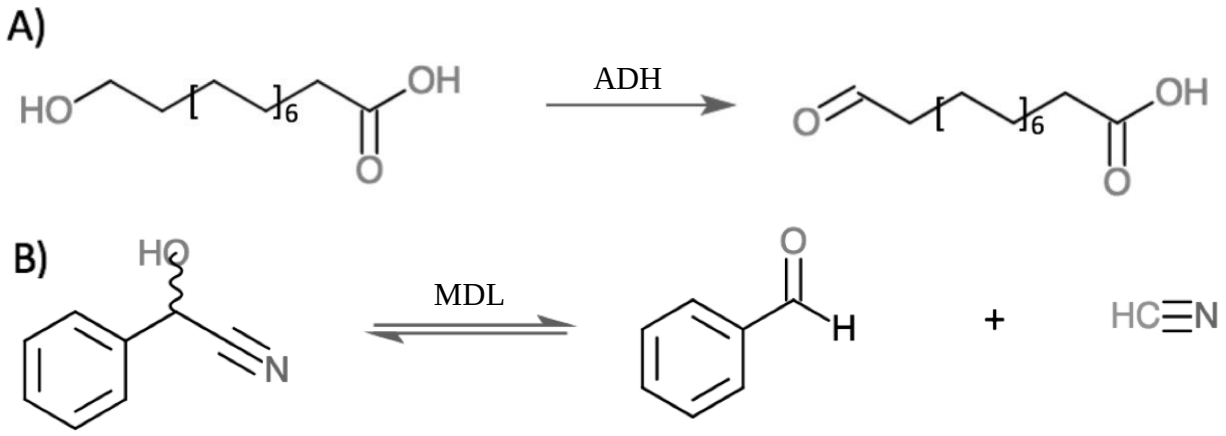


Figure 3: Schematic of the potential enzymatic activity of HTH. (A) Dehydrogenation of ω-dicarboxylic FA to ω-oxo FA by a ω-hydroxy fatty acyl dehydrogenase (ADH). (B) Cyanogenesis of mandelonitrile to benzaldehyde and hydrogen cyanide by a mandelonitrile lyase (MDL).

1.4.1 The Role of Hydrogen Cyanide in Plants

If HTH has a similar enzymatic activity to FAD-HNLs and produces HCN, it is important to consider the role of HCN in plants. This molecule is highly regulated and is crucial for various metabolic, physiological and developmental processes, be they direct or indirect (Siegień & Bogatek, 2006). HCN is a small, simple gaseous molecule that can easily diffuse to non-cyanide producing tissues. Plants can produce HCN multiple ways: catabolism of cyanogenic glycosides, as a side product to the ethylene biosynthesis pathway or as a product of glyoxylates in photorespiration and hydroxylamine (Siegień & Bogatek, 2006). A hydroxynitrile lyase will catabolise an aliphatic or an aromatic containing hydroxynitrile to produce hydrogen cyanide and

aldehyde or ketone. Mandelonitrile lyase will use mandelonitrile as its substrate to produce benzaldehyde and HCN (Figure 3B).

1.4.2 HCN & Plant Stress Response

Cyanide can be very toxic and is typically produced in response to an environmental stress, along with ethylene. HCN levels must be tightly regulated in plants to prevent self-harm. Even at very low concentrations, it is known to stunt growth, reduce plant size and decrease chlorophyll content (J. M. Smith & Arteca, 2000). When produced, the tissue is quick to metabolise the cyanide. Cyanoalanine synthase uses HCN and L-cysteine to produce 3-cyano-L-alanine, which is then converted into aspartic acid or asparagine by *A. thaliana* Nitrilase 4 (Piotrowski & Volmer, 2006). Certain areas of the plant may still have non-lethal levels of HCN which is believed to trigger the plant's response to stress (J. M. Smith & Arteca, 2000). This is seen in tobacco plants, whereby a non-lethal concentration of HCN enhances the resistance to the tobacco mosaic virus (Chivasa & Carr, 1998). It is believed that cyanide is also important for stress relief, as non-lethal concentrations decrease the growth rate of the infected cells to conserve energy and decrease the risk of heritable damage (May, Vernoux, Leaver, Van Montagu, & Inzé, 1998).

ER bodies are also produced in response to environmental stress. Matsushima et al. demonstrated that ER bodies were induced in rosette leaves treated with methyl jasmonate, Rosette leaves do not appear to contain ER bodies when the plant is not under stress (Matsushima et al., 2003). The induced ER bodies are limited to the peripheral regions exposed to the stressor showing a localization of ER bodies in that area (Chang, 2016; Matsushima et al., 2003). ER bodies have been linked to plant defence based on their composition and localization (Matsushima et al., 2003). The HTH protein has been observed in these structures, suggesting

that HTH may be involved in stress response (Chang, 2016). If HTH were to have MDL enzymatic activity, this may explain its presence in ER bodies. MDL enzymes have not been previously shown to localize to ER bodies.

A comparison between HTH and MDL has shown sequence similarity overall, although differences in amino acid conservation has been noted between residues 552 and 577 (numbering based on the HTH peptide sequence from *A. thaliana*), a region within the conserved C-terminal substrate binding domain (Krolikowski et al., 2003). One of the HTH putative active site residues, Asn567 (numbering based on the HTH peptide sequence from *A. thaliana*) is found within this region. This may indicate that these proteins have distinct enzymatic function. A subsequent comparison of *Arabidopsis* and rice HTH to black cherry and almond MDL was also performed and an unrooted phylogenetic tree showed that all the HTH sequences cluster in a distinct group from MDL (Krolikowski et al., 2003). This further suggests that although the enzymatic function of these two proteins may be related to each other, they are most likely distinct.

1.4.3 HCN and Seed Dormancy

HCN has also shown concentration dependent effects on dormant seeds. HCN is produced during the pre-germination period, even in seeds that do not contain cyanogenic glycosides. It is believe that the source of HCN in non-cyanogenic seeds originates from cyanogenic storage substances (Esashi et al., 1991). Seed germination is stimulated when low amounts of HCN are present, but the effects are only observed once HCN is removed. If HCN remains present, germination is blocked (Maruyama et al., 1996). The mode of cyanide action in pre-germinated seeds is still unclear. Many putative modes of cyanide action have been

proposed, forming a complex web of potential interactions, but its importance during germination could be further indicative of HTH's hydroxynitrile lyase activity.

A transgenic plant expressing GFP or yellow fluorescent protein (YFP) tagged HTH construct (HTH-FP) has been observed as early as the FG3 stage of megagametogenesis and remains present through seed maturation (Chang, 2016). Cutin monomers and certain genes previously identified to be important in cuticle synthesis have been identified in the seed coat (Molina, Bonaventure, Ohlrogge, & Pollard, 2006; Molina, Ohlrogge, & Pollard, 2008). Many *hth* mutants have bigger seeds than WT due to seed coat defects (Chang, 2016), but this does not explain HTH's role during megagametogenesis. The HTH protein has also been linked to female gametophyte development. The *eda-17* mutant will arrest at the FG2 stage of megagametogenesis, causing ovule defects and female sterility (Pagnussat et al., 2005). Interestingly, *eda17* is the only *hth* allele with a mutation within the first 1kb of the 5' end of the gene (Lolle, unpublished results). This likely indicates that this region of the HTH gene may be essential for female gametophyte development, although that role is still unclear. This link between HCN production and seed dormancy may explain HTH's presence during megagametogenesis.

1.4.4 HCN & Nitrate Assimilation

In addition to pathogen defence and germination, HCN is proposed to be involved in nitrate assimilation (Solomonson & Barber, 1990) (Figure 4). Nitrogen reductase is a key enzyme for nitrate assimilation. When oxidized, nitrate reductase can convert nitrate to nitrite then complex with HCN and become inactivated (Siegień & Bogatek, 2006). Nitrite is used to form hydroxyloamine which is an intermediate used to form amino acids (Solomonson & Spehar, 1977). HCN can be dissociated from nitrogen reductase by exposing the complex to light

or oxygen, linking this process to photorespiration (Solomonson & Barber, 1990). Crude extracts from *Chlorella* have indicated that HCN is formed from glyoxylate and hydroxyloamine, a product of photorespiration and an intermediate in nitrate assimilation, respectively. More glyoxylate is produced under high intracellular oxygen levels, ultimately producing more HCN. When intracellular carbon dioxide levels are low, HCN can inactivate nitrogen reductase and prevent the accumulation of toxic molecules (Solomonson & Spehar, 1977).

HCN's role in plants seems widespread and diverse. It appears to be able to act in a concentration dependent manner (stress response and germination) or act directly with enzymes (nitrate assimilation). Since it has direct and indirect effects on various pathways, it is unclear how involved HCN is throughout the entire plant. Further research is needed to gain a better understanding of its role in plants. If HTH has hydroxynitrile lyase enzymatic activity, the current knowledge on HCN could be an explanation as to why it is found in various tissues.

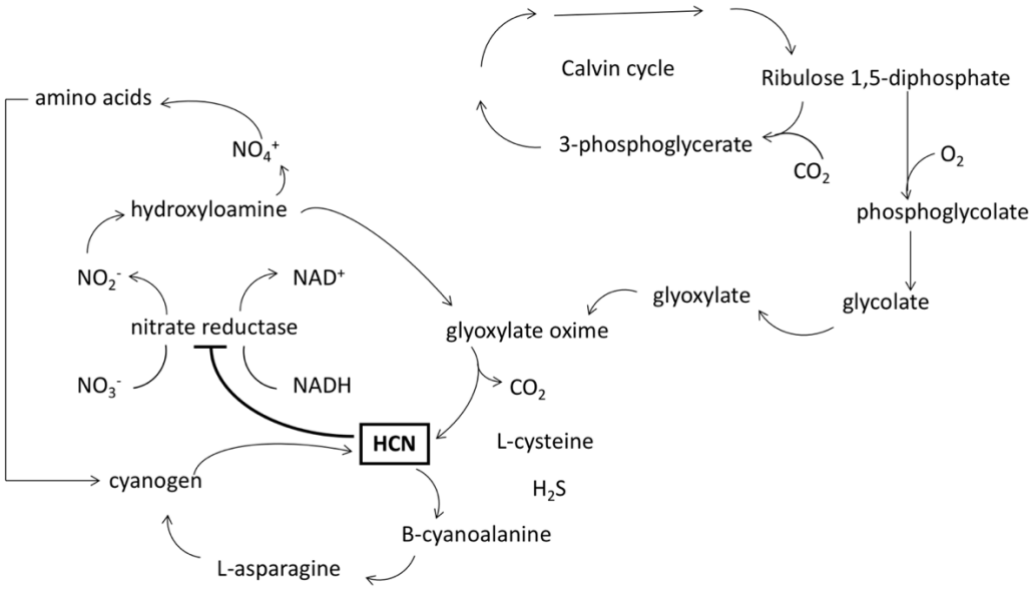


Figure 4: Cyanide-mediated assimilation of nitrate. Hydrogen cyanide (HCN) will bind and inactivate nitrate reductase. HCN levels will increase as hydroxyloamine and oxygen levels increase. *Based on Figure 4 from Siegień & Bogatek, 2006*.

1.4.5 Moonlighting Proteins

Based on the current knowledge, it is possible that HTH may be multifunctional and could therefore be a moonlighting protein. Moonlighting is a term used to describe proteins that perform more than one activity. These types of proteins are found in many species including plants (Jeffery, 2017). For example, arogenate dehydratase 2 from *A. thaliana* has arogenate dehydratase activity, but it has also been suggested that this protein is also important for chloroplast division, playing a non-enzymatic role (Rad et al., 2018). Another example is plastidial NAD^+-dependent malate dehydrogenase from *A. thaliana*. This protein is a part of a large protease complex important for protein processing during plastid biogenesis. It has also been reported that this protein also converts malate to oxaloacetate (Schreier et al., 2018; S. M. Smith, 2018). Very few plant moonlighting proteins have been characterized opening up the possibility that many have yet to be discovered. Gaining a better understanding of the HTH protein could help elucidate its potential enzymatic activity and give a better insight into its potential moonlighting function.

The most direct way to determine if a protein has a particular enzymatic function is to perform enzymatic assays. Although this may be easy to perform, the challenge lies in obtaining an active isolate of the protein of interest. A previous attempt has been made to isolate and test a bacterially expressed recombinant HTH protein for MDL and ADH enzymatic activity (Chang, 2016). These tests showed no enzymatic activity from the isolated HTH protein. This could indicate that HTH does not perform either enzymatic function, but it is more likely that the isolated protein was not active. In this previous attempt, the HTH gene was cloned into the pMal-c4x vector and expressed in *E. coli*, a common recombinant protein expression system (Chang,

2016). Once HTH protein production was induced and cells were lysed, very little HTH protein was obtained in the soluble fraction.

1.5 Objectives of the Thesis

The objective of the research outlined in this thesis was to expand the current knowledge of the HTH protein function using two approaches. The first approach focused on computational analysis. The second approach directly tested HTH protein for MDL activity using recombinant HTH protein or by isolating crude protein from plant lines expressing HTH translationally fused to GFP. A model of the HTH protein was compared to two structural analogs to gain a better understanding of the HTH protein structure and potential enzymatic function, with a focus on conserved active site residues. N-linked glycosylation sites conserved in hydroxynitrile lyases were compared to the putative sites found in HTH using a predicted structural model. Ligands predicted to bind to the HTH model were also analyzed. A recombinant HTH protein with a C-terminal His tag was expressed in *E. coli* then isolated using nickel chromatography. Adjustments to the isolation protocol were made to optimize the isolation of the recombinant protein, with current work focusing on using *Pichia pastoris* to express HTH. Finally, crude extracts from *A. thaliana* floral tissue were collected to test for mandelonitrile lyase enzymatic activity.

Chapter 2

Structural Analysis of the HOTHEAD Protein Model

2.1 Introduction

The cuticle is an important adaptation that helps terrestrial plants survive on land. A loss of cuticle integrity has proven to cause negative effects on the plant (Fich et al., 2016; Yeats & Rose, 2013). To date, many genes have been linked to cuticle formation but the exact role of some of these genes has yet to be determined, one of them being the *HTH* gene (Yeats & Rose, 2013).

The potential enzymatic function of the HTH protein has not been determined. It does contain three KDEL-like sequences (Chang, 2016) and localizes to the ER, the organelle where the majority of the cuticle biosynthesis pathway occurs (Chang, 2016; Yeats & Rose, 2013). A role for the HTH protein in cuticle synthesis has been proposed: an alcohol dehydrogenase that leads to the production of a cutin monomer, α-,ω-dicarboxylic acid (Kurdyukov et al., 2006). HTH does share sequence identity to an ADH from *Candida* but it has higher sequence identity to an MDL from *Prunus* (Kurdyukov et al., 2006). MDL is an FAD-HNL that produces HCN, a small gaseous molecule known for being toxic and linked to plant defence (Siegień & Bogatek, 2006). The HTH protein appears to localize to ER bodies, structures believed to be involved in plant defence (Chang, 2016). If HTH is a protein involved in cyanogenesis, its presence in ER bodies could be further justified.

The structure of the HTH protein is currently unknown due to the inability to isolate it in the active form making it quite challenging to determine its role in *A. thaliana*. Until a protein isolation protocol is developed, other approaches must be considered to elucidate on HTH's enzymatic function. Thanks to technological advancements, protein models can be easily developed and have proven to be useful tools when the actual protein structure is unknown. Homology modeling, also known as comparative modeling, allows for the production of a

protein structural model using known protein structures (Yang & Zhang, 2015). I-TASSER is one of those software programs which uses multiple-threading alignments (Yang & Zhang, 2015). This means that multiple proteins templates are aligned to a single protein sequence to gain an overall more accurate prediction of the unknown protein structure (Peng & Xu, 2011).

In this chapter, a model of the HTH protein was developed using the I-TASSER software (Roy, Kucukural, & Zhang, 2010; Yang et al., 2015; Yang & Zhang, 2015; Zhang, 2008). Using Chimera (Pettersen et al., 2004), an interactive molecule visualization program, the HTH protein model was compared to a choline oxidase from *A. globiformis* (PDB: 3nne) as well as an FAD-HNL from *P. dulcis* (PDB: 1ju2). The location of known key active site residues in both of these proteins were compared to the residues in the HTH protein model to gain a better understanding on HTH's potential enzymatic function. The binding site of potential ligands were also analyzed using COACH (Yang, Roy, & Zhang, 2013b, 2013a) and SwissDock (Grosdidier, Zoete, & Michielin, 2011; Grosdidier, 2011).

2.2 Methods

2.2.1 HOTHEAD Protein Structural Model Analysis

The complete HTH peptide sequence was input into the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy, Kucukural, & Zhang, 2010; Yang et al., 2015; Yang & Zhang, 2015; Zhang, 2008). The HTH structural model was compared to two of the structural analogs, hydroxynitrile lyase from *P. dulcis* (PDB ID: 1ju2) and choline oxidase from *A. globiformis* (PDB: 3nne), using Chimera (Pettersen et al., 2004). The two structures were superimposed based on peptide sequence and secondary structures. Various ligands were predicted using the COACH server and analyzed using Chimera (Yang, Roy, & Zhang, 2013b, 2013a). Potential ligand docking sites were further investigated using the SwissDock server (Grosdidier, Zoete, & Michielin, 2011; Grosdidier, 2011) and analyzed using Chimera.

2.3 Results & Discussion

2.3.1 Hypothesized Enzymatic Function of HTH

As it stands, the HTH protein is hypothesized to either have alcohol dehydrogenase activity or hydoxynitrile lyase activity. The HTH peptide sequence has 25% peptide sequence identity to an alcohol dehydrogenase from *Candida* (Kurdyukov et al., 2006). Work with the *hth-12* mutant has shown increased levels of ω-hydroxy FAs and decreased levels of ω-oxo FAs in *A. thaliana* cuticle. ω-oxo FA is a precursor to α-,ω-dicarboxilic FA, one of the major cutin monomers present in the cuticle (Kurdyukov et al., 2006). This gives a clear link to HTH's involvement in cuticle synthesis and potentially seed coat development, but it does not answer why HTH localizes in ER bodies or appears to be involved in embryogenesis.

In comparison to the *Candida* alcohol dehydrogenase, HTH has higher peptide sequence identity to hydroxynitrile lyase from *Prunus*, at 41% (Kurdyukov et al., 2006). This enzyme is a part of the GMC oxidoreductase protein family, like HTH (Krolikowski et al., 2003). Phylogenetic analysis of the *HTH* gene and hydroxynitrile lyases from *Prunus* showed that the HTH sequences cluster in a distinct group from the HNL sequences, suggesting that these enzymes have similar but distinct enzymatic functions (Krolikowski et al., 2003). Whether HTH shares the same enzymatic function as HNLs is yet to be determined.

2.3.2 Structural Analogs to the HTH Protein Model

Knowing the structure of a protein can be useful to better understand its function. Since the HTH protein has yet to be successfully isolated, the structure has not been elucidated. However, due to advancements in computational structure prediction algorithms, this research was able to generate a model of the HTH protein using the I-TASSER software that offers some possible functional insights (Figure 5). This protein model contains the complete HTH peptide

sequence. Structural analogs to the HTH model have been compiled, all of which are members of the GMC oxidoreductase protein family (Table 1).

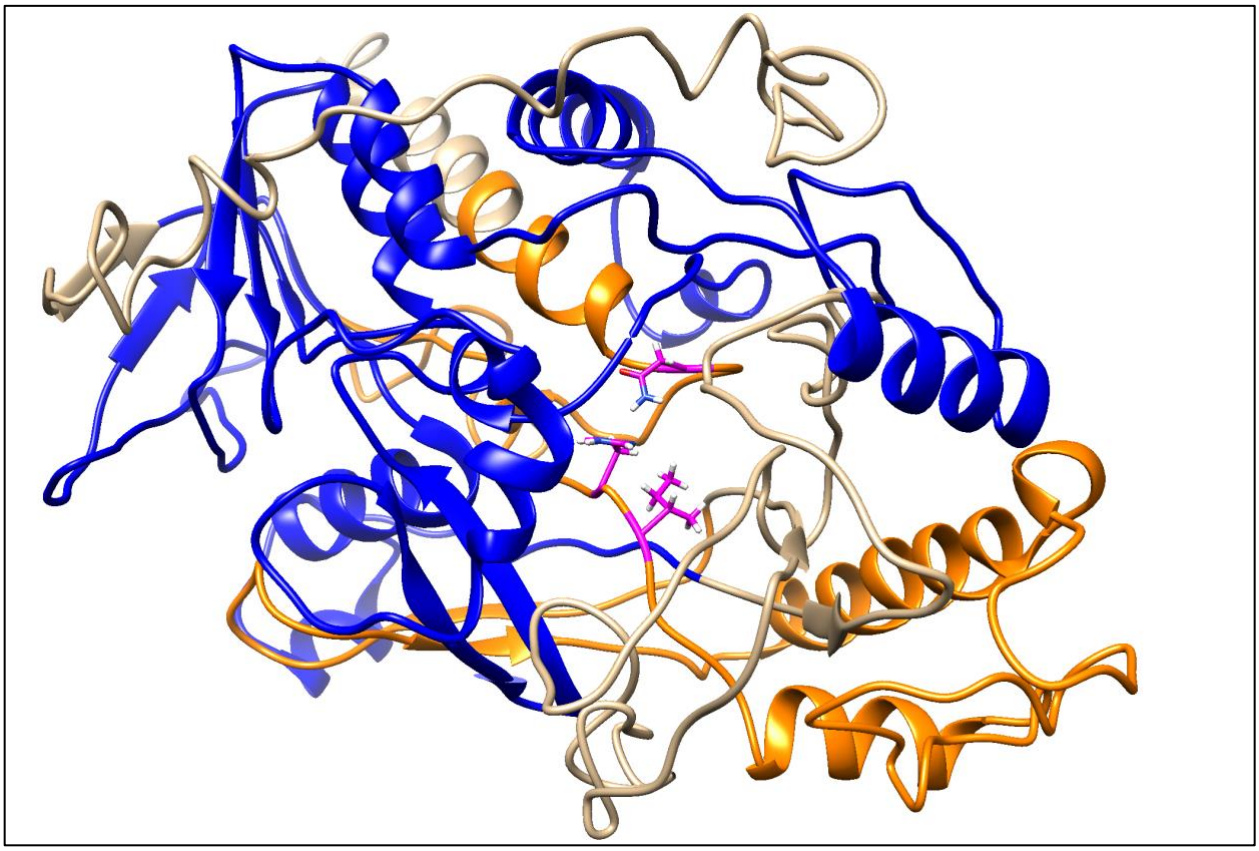


Figure 5: Ribbon diagram of the predicted HTH model. The HTH model was generated with I-TASSER and visualized in Chimera. The portion of the HTH model that is a part of the GMC N-terminal domain (residues 63-335) is represented in blue. The portion of the HTH model that is a part of the HMC C-terminal domain (residues 431-577) is represented in orange. The three putative active site residues (Ile527, His529 and Asn567) are represented in magenta.

These analogs are ranked based on their template modeling score (TM-Score), root-mean-square deviation (RMSD), percent identity (IDEN), and coverage. The RMSD is a quantitative measure of the similarity between two proteins. This value relies on a distance measurement between equivalent residues in each structure to a refence point. Proteins with a lower RMSD are more similar than protein with a higher RMSD (Kufareva & Abagyan, 2012). The TM-Score is another quantitative measure of the similarity between two proteins. Where the

RMSD measurement weighs the distance between each residue evenly, the TM-Score will weigh shorter distances between equivalent residues more strongly than equivalent residues with longer distances, making it a more accurate measurement (Zhang & Skolnick, 2004). The TM-Score is a value from 0-1 with 1 indicating two completely identical protein structures (Kufareva & Abagyan, 2012). The IDEN is the percentage of identical residues in the aligned protein sequences. The coverage is the percentage of residues structurally aligned.

Table 1: Structural analogs to the HTH protein model. List of the top 10 identified structural analogs in PDB (protein data bank; https://www.rcsb.org/) compared to the HTH protein model predicted by I-TASSER. Proteins are ranked from the protein with the highest structural similarity to the lowest structural similarity, based on various parameters: template modeling score (TM-score) root-mean-square deviation (RMSD), sequence identity (IDEN) and sequence coverage (Coverage). The TM-score and RMSD indicate the similarity between two protein structures but the TM-score is independent of the peptide length (Kufareva & Abagyan, 2012; Zhang & Skolnick, 2004). The IDEN measures identical residues between two peptide sequences. The coverage measures similar residues between two peptide sequences. All structural analogs are GMC oxidoreductases. The top three structural analogs are FAD-dependent hydroxynitrile lyases.

Structural Analog	TM-score	RMSD	IDEN	Coverage
1ju2 (hydroxynitrile lyase)	0.846	1.56	0.370	0.869
3red (hydroxynitrile lyase)	0.845	1.62	0.364	0.869
5eb4 (hydroxynitrile lyase)	0.833	1.67	0.363	0.859
3nne (choline oxidase)	0.747	2.68	0.238	0.805
4h7u (pyranose dehydrogenase)	0.745	3.27	0.215	0.828
3q9t (formate oxidase)	0.736	2.79	0.211	0.798
5I68 (alcohol oxidase)	0.731	3.02	0.195	0.801
3fim (aryl-alcohol-oxidase)	0.729	2.78	0.215	0.791
1gpe (glucose oxidase)	0.728	3.34	0.187	0.810
1gal (glucose oxidase)	0.726	3.04	0.213	0.798

The proteins listed in Table 1 are members of the GMC oxidoreductase protein family. This means they have an N-terminal FAD-binding domain and the C-terminal substrate binding domain, which contains the active site residues (Li et al., 1993). Most of the predicted HTH protein analogs use FAD to perform a redox enzymatic reaction, except for the top three, 1ju2, 3red, and 5eb4. These three are all FAD-HNLs from almond (Dreveny et al., 2009) sharing

common features to other GMC oxidoreductase proteins, but do not appear to use their FAD cofactor for enzymatic purposes, at least not for a redox reaction (Dreveny et al., 2001).

2.3.3 Conserved Active Site Residues in FAD-Dependent Hydroxynitrle Lyase

It is predicted that all FAD-HLNs have two conserved active site residues: His459 and His497 (numbering based on the *P. dulcis* HNL peptide sequence; Figure 6B). HTH has three putative active site residues: Ile527, His529 and Asn567 (numbering based on the A. *thalian* HTH peptide sequence) (Chang, 2016). By using Chimera, the HTH protein model was superimposed over 1ju2, the FAD-HNL from *P. dulcis* (Figure 6D). HTH His529 overlaps with 1ju2 His459. The His459 in 1ju2 causes the protonation of the cyanide ion, allowing for the production of hydrogen cyanide (Dreveny et al., 2009). Since HTH has the same residues in this position, it could have the same function. HTH Asn567 overlaps with 1ju2 His497. The His497 in 1ju2 is considered a general base that hydrogen bonds with the hydroxyl group of mandelonitrile to help stabilize the substrate (Dreveny et al., 2001).

1ju2 also has a tyrosine at position 457 which could hydrogen bond with the substrate (Dreveny et al., 2009) but based on pKa values of the residue R-group, His497 is more likely to play that role (Dreveny et al., 2001). HTH has an asparagine in the same position as 1ju2 His497 which is not ionizable under physiological conditions. This means that it cannot be as strong of a general base like histidine or tyrosine. Histidine may be conserved in FAD-HNLs to ensure a strong general base is present to properly orient the substrate within the active site. Since HTH does not contain this key conserved residue, it is reasonable to suggest that HTH will not possess MDL enzymatic activity.

Neither of the conserved active site residues in 1ju2 (His459 and His497) overlap with HTH Ile527. It is in the same position as Tyr457 in 1ju2. Isoleucine is an aliphatic residue and it

likely not an active site residue. It could potentially hydrogen bond with the substrate to help hold it in place or create a favourable environment within the active site. Once further research on potential HTH substrates is conducted, the involvement of Ile527 in the binding process should be clarified.

2.3.4 Conserved Active Site Residues in GMC Proteins

It is predicted that GMC oxidoreductases have two conserved active site residues: His466 and Asn510 (numbering based on the Choline Oxidase peptide from *A. globiformis*) (Figure 6C) (Dreveny et al., 2009). These residues form a well-defined hydrogen bond network that ensures the substrate is in the correct position and reacts with the N5 molecule of the FAD co-factor (Fraaije & Mattevi, 2000; Yue et al., 1999). As described previously, HTH has three putative active site residues: Ile527, His529 and Asn567 (Chang, 2016). Chimera was used to superimpose 3nne over the HTH protein model (Figure 6E). HTH His529 overlaps with 3nne His466 and HTH Asn567 overlaps with 3nne Asn510. Neither of the conserved active site residues in 3nne overlap with HTH Ile527. Since HTH contains an asparagine and histidine in the same location as the conserved residues of 3nne, they could be important for proper enzymatic function by forming hydrogen bonds with the substrate. Since HTH does contain both conserved redox required active site residues, the structural model is more consistent with the proposal that HTH has ADH enzymatic activity.

If HTH were to have ADH enzymatic activity, this could explain why it localizes in the ER of epidermal cells (Chang, 2016). As it has been previously proposed, HTH appears to be involved in cuticle synthesis, a process that primarily takes places in the ER of epidermal cells (Yeats & Rose, 2013). But this protein has been found in areas of the plant where cuticle synthesis does not occur, such as ER bodies in *A. thaliana* hypocotyl cells (Chang, 2016). These

are structures that have been linked to plant defence (Matsushima et al., 2003). One of the known *hth* mutant alleles, *eda-17*, causes embryo sac development arrest, linking HTH to early events in female gametophyte development. These results raise many questions about HTH's role in *A. thaliana*, one of which being: Is HTH a moonlighting protein? Moonlighting proteins are proteins that have more than one function and typically arise from some sort of evolutionary split and resource allocation (Jeffery, 2017). Very few moonlighting proteins have been identified in plants (Rad et al., 2018) meaning that many have yet to be discovered. HTH may be one of them.

2.3.5 Predicted Ligand Binding Site on the HTH Protein Model

Potential ligands to the HTH protein were predicted using COACH software (Yang, Roy, & Zhang, 2013b, 2013a). This generated a list of ligands, their predicted binding site and a coinciding C-score. The C-score is a confidence score of the prediction based on the threading template alignments and convergence parameters of the structure assembly simulations (Roy, Kucukural, & Zhang, 2010; Yang et al., 2015; Yang & Zhang, 2015; Zhang, 2008). This score ranges from -5 to 2 with a higher value signifying a higher confidence.

The cofactor with the highest C-score was FAD at 0.78 (Figure 7A). This further supports the belief that HTH is an FAD-dependent protein and a member of the GMC oxidoreductase protein family. Whether HTH uses FAD for enzymatic purposes or simply for structural integrity us unclear. Taking a closer look at the predicted binding site of FAD, it is possible to see that the flavin portion of the molecule comes in close proximity to the three putative active site residues in HTH (Figure 7A). No portion of the FAD molecule comes into contact with the predicted FAD-binding site (residues 64-91 in HTH). This is likely due to the inaccuracies in this simple docking experiment.

Benzaldehyde and mandelonitrile were the two other ligands predicted to complex with the HTH protein model. Although benzaldehyde and mandelonitrile are fairly similar and close in overall size, their predicted binding sites are in different areas (Figure 7B-C). Benzaldehyde was predicted to bind in proximity to the putative active site residues, 4.3 Å from His529 in the HTH protein model, while mandelonitrile is predicted to bind to the surface of the HTH protein, 12 Å from Ile527 in the HTH protein model, the closest putative active site residue. Unfortunately, benzaldehyde and mandelonitrile have very low C-scores, at 0.04 and 0.01, respectively. Due to the low C-score, it is not possible to say with confidence that either ligands bind to the HTH protein. No other ligands were predicted to bind to the HTH protein model.

An issue with the ligand docking predictions is that the FAD co-factor is not included when the model is input into the software. This generates a large empty space in the predicted location of FAD within the HTH model (data not shown). This was further exemplified with the SwissDock software (Grosdidier, Zoete, & Michielin, 2011; Grosdidier, 2011). When mandelonitrile was input as a potential ligand, it was predicted to bind to the surface of the HTH model as well as within the FAD binding site. A similar issue could be happening within the active site. Since no ligands are included when the protein model is being created, the residues may be oriented differently to accommodate that empty space. This could be preventing the software algorithms to predict certain ligands bind within this site. To account for this, the residues around the empty spaces could be manually adjusted in Chimera to see if ligands can fill the space, although this may not be an accurate representation of the model.

These simple docking experiments were largely a failure. The results should not be strongly interpreted due to the reasons discussed above. A more conplex docking study could be undertaken to provide stronger evidence for or against the binding of these ligands.

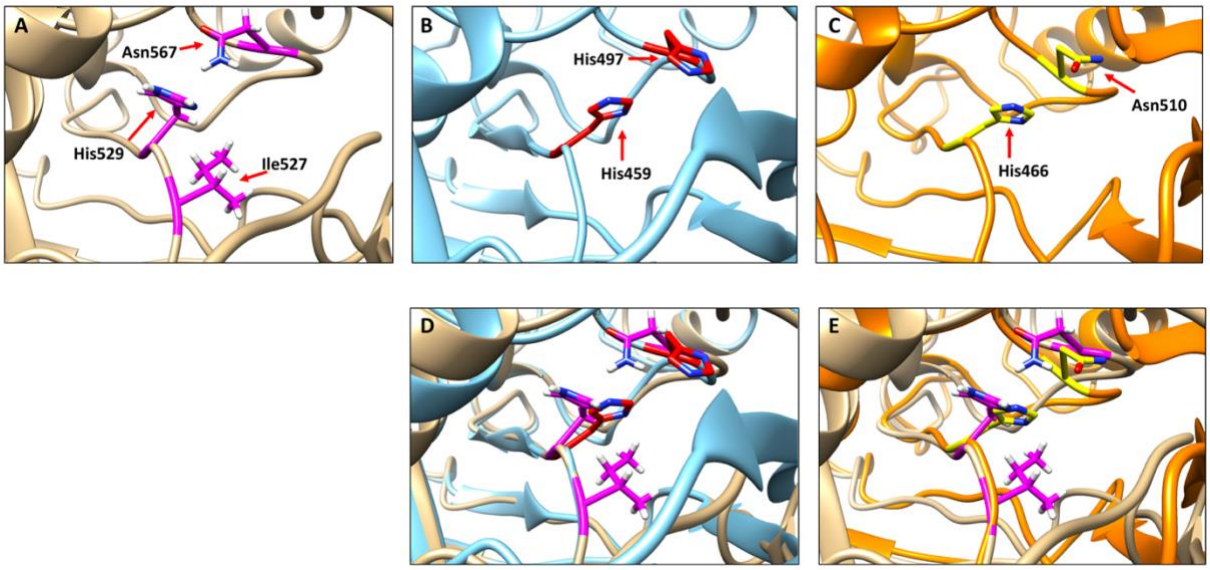


Figure 6: Ribbon diagrams comparing the putative active site residues of the HTH protein to conserved active site residues of an FAD-HNL (1ju2) & a GMC protein (3nne). The HTH model was generated using I-TASSER and visualized in Chimera. (A, B, C) The active site residues of HTH, 1ju2 and 3nne. (A) The HTH protein model is represented in beige. The three putative active site residues (Ile527, His529, Asn567) are represented in magenta. (B) The 1ju2 protein is represented in blue. The two conserved active site residues (His459 and His497) are represented in red. (C) The 3nne protein is represented in orange. The two conserved active site residues (His466 and Asn510) are represented in yellow. (D, E) The HTH protein model superimposed over the 1ju2 and 3nne based on secondary structure using Chimera. (D)The HTH protein model (beige) superimposed over 1ju2 (blue). HTH His529 overlaps with 1ju2 His459 while HTH Asn567 overlaps with 1ju2 His497. (E) The HTH protein model (beige) superimposed over 3nne (orange). HTH His529 overlaps with 3nne His466 while HTH Asn567 overlaps with 3nne Asn510.

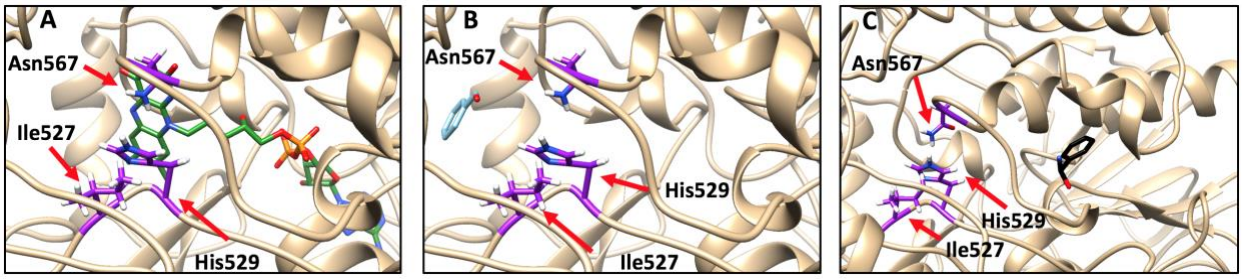


Figure 7: Ribbon diagrams of the HTH model-ligand complexes. The HTH model complexes were generated with COACH and visualized in Chimera. The HTH protein model is represented in beige. The putative active site residues are purple. (A) The most probable binding site of FAD (green) within the HTH protein model. The ligand looks to bind nicely within the FAD binding domain with the N5 molecule located in proximity of the active site. (B) The most probable binding site of benzaldehyde (blue) within the HTH protein model. The ligand is found within the active site. Its hydroxyl group is shown to be ~4.2Å from the closest putative active site residue, His529. (C) The most probable binding site of mandelonitrile (black) within the HTH protein model. The ligand is found binding to the surface of the protein structure.

2.3.6 Comparing Hydroxynitrile Lyase N-Linked Glycosylation to the HOTHEAD Model

N-linked protein glycosylation is a highly conserved process in eukaryotes (Burda & Aebi, 1999). It begins in the ER with the formation of a glycan. This oligosaccharide is found within the ER lumen and bound to the membrane. It is transferred to the protein by oligosaccharyl transferase (Burda & Aebi, 1999). The glycan will bond with the protein's asparagine with the proper recognition sequence (Burda & Aebi, 1999). At this point, chaperone proteins will recognize the glucose residues on the glycan and aid in protein folding (Helenius et al., 1997). Glucosidase will cleave off the glucose molecules, allowing the folded protein to dissociate from the chaperone proteins and travel to the Golgi apparatus (Helenius et al., 1997). Once in the Golgi apparatus, the oligosaccharide will be subjected to further modifications (Burda & Aebi, 1999; Schwarz & Aebi, 2011).

All known FAD-HNLs are predicted to have three conserved N-linked glycosylation sites: Asn118, Asn135 and Ans352 (numbering based on the *P. dulcis* HNL peptide sequence)

(Dreveny et al., 2001). These sites all have the same recognition sequences of N-X-S/T (i.e., Asp, any amino acid, Ser or Thr) (Dreveny et al., 2001). Based on previous bioinformatic work, the HTH protein is predicted to have 5 putative N-linked glycosylation sites, all with the same recognition sequences as FAD-HNLs: Asn84, Asn101, Asn420, Asn482 and Asn511 (Chang, 2016). These glycosylation sites are all found on the surface of the protein (Figure 8). It is interesting to note however that when the HTH protein model is superimposed with the top structural analog, based on secondary structures, none of the putative glycosylation sites align with the three conserved FAD-HNL N-linked glycosylation sites. The alignment of these glycosylation sites should be further investigated once the structure of the HTH protein is confirmed.

HTH has been experimentally shown to be a glycoprotein based on size shifts following endoglycosidase treatment of a HTH::GFP protein extracted from transgenic plant lines (Chang, 2016). FAD-HNLs have varying amounts of glycosylation which may not be important to the enzyme's function as de-glycosylation of the purified enzyme from *P. dulcis* did not affect enzymatic activity (Dreveny et al., 2001). This de-glycosylation occurred after the protein had been folded so it is likely that the N-linked glycosylation was important for proper protein folding. It is unknown whether glycosylation is important for HTH protein function.

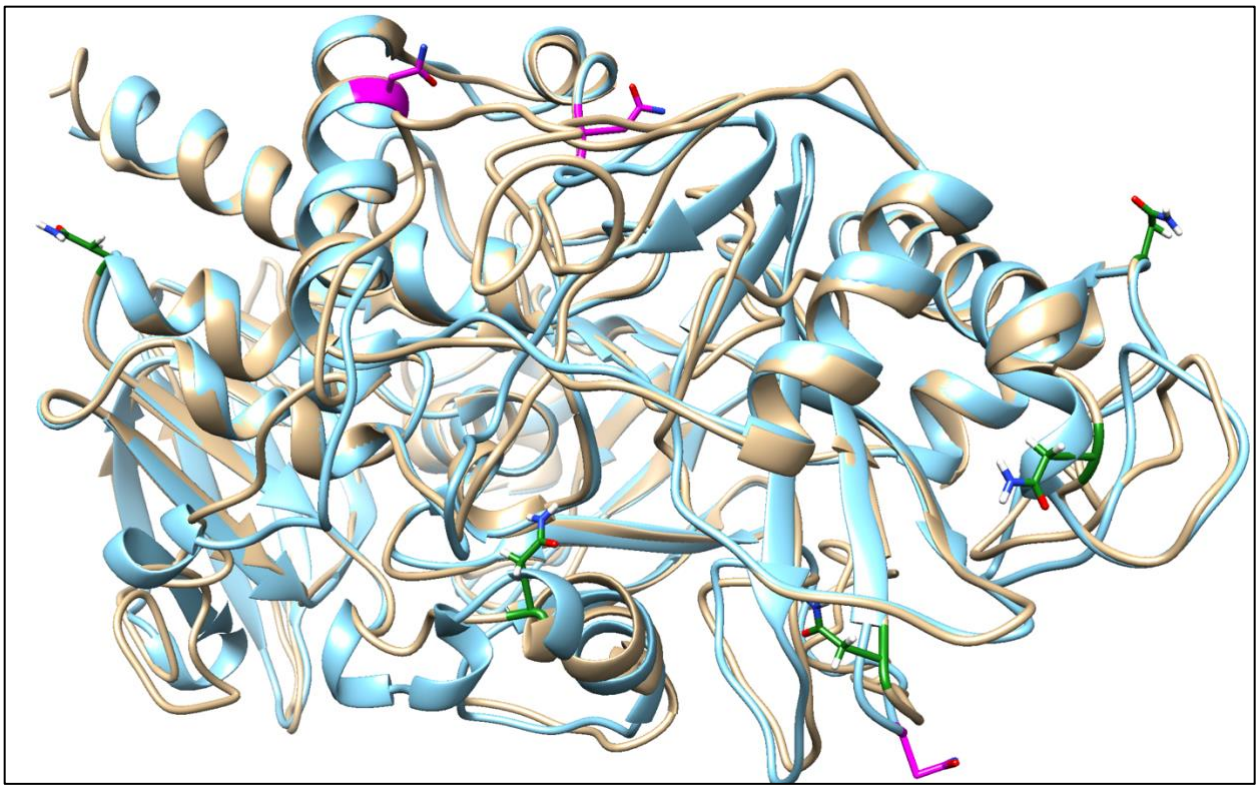


Figure 8: Ribbon diagrams of the HTH model & FAD-HNL (1ju2) N-linked Glycosylation sites. The HTH model complexes were generated with COACH and visualized in Chimera. The HTH protein model is represented in beige. The 5 putative N-linked glycosylation sites (Asn84, Asn101, Asn420, Asn482 and Asn511) are represented in green. The 1ju2 protein is represented in blue. The 3 conserved N-linked glycosylation sites (Asn118, Asn135 and Ans352) are represented in purple. All putative N-linked glycosylation sites are found on the surface of the HTH protein model. None of the glycosylation sites in the HTH model overlap with the glycosylation sites in 1ju2.

2.4 Conclusions

The protein model predicted by I-TASSER has been useful to learn more about the HTH protein. This model was built by aligning the HTH peptide sequence to proteins with known structures. Developing a model of a protein has proven to be an effective method to determine the overall structure of a protein, especially proteins with sequence identities ≥ 25% (Giorgetti, Raimondo, Miele, & Tramontano, 2005). Although not as precise as the actual HTH structure, since the model is based on the structure of other proteins, it is a useful tool for structural and computational analysis. The fact that FAD was omitted from the model did create inaccuracies that could have effects on the overall structure. Once a protein isolation protocol is developed for the HTH protein, its structure can be determined and used to validate the findings presented in this chapter.

The software used to predict potential ligand docking sites omitted FAD, causing inaccurate predictions. Residues likely changed orientation to fill in the empty space typically occupied by FAD. Due to these slight changes in the protein, FAD was unable to bind to its predicted binding site (residues 64-91 in HTH). Even though MDL and benzaldehyde, two molecules that bind to FAD-HNLs, were predicted to bind to the HTH model, their confidence scores were too low to state that they truly bind to HTH. These results could not be used to further validate the conclusions drawn from the conserved active site residue analysis. A more robust docking prediction software should be used in the future to determine potential HTH protein ligands.

The HTH protein shares sequence identity to a GMC protein with redox activity and an FAD-HNL (Kurdyukov et al., 2006). Based on the results presented in Sections 2.3.3 and 2.3.4, the HTH protein model does not contain the conserved active site residues of FAD-HNLs and it

does contain the conserved active site residues of GMC oxidoreductase proteins (Figure 6). These residues are important for forming a hydrogen bond network with the substrate to ensure that it is properly positioned in proximity to the N5 atom of the FAD (Yue et al., 1999). This could mean that HTH has ADH enzymatic activity.

N-linked glycosylation occurs within the ER and the Golgi apparatus. This post-translational modification is important for proper protein folding (Schwarz & Aebi, 2011). If a glycoprotein is lacking its required N-linked glycosylation, it most likely will not fold properly. It was shown that FAD-HNLs have three conserved N-linked glycosylation sites (Dreveny et al., 2001) that do not overlap with the putative N-linked glycosylation sites in HTH. If an FAD-HNL is de-glycosylated, it can still remain active (Dreveny et al., 2001). The importance of N-linked glycosylation in the HTH protein remains to be determined.

HTH has been implicated in several different processes: cuticle, embryogenesis and plant defence. It may have more than one function which would classify it as a moonlighting protein, such as arogenate dehydratase 2 (Rad et al., 2018) and plastidial NAD^+-dependent malate dehydrogenase (Schreier et al., 2018; S. M. Smith, 2018). To date, very few moonlighting proteins have been identified in plants. Hopefully, once the actual HTH protein structure is determined, further research can be done to gain a better understanding of its potential enzymatic activity and its role in *A. thaliana*.

Chapter 3

Recombinant HOTHEAD Protein Isolation & Enzymatic Assays

3.1 Introduction

A significant bottleneck in assigning a definitive role for HTH in the cuticle synthetic pathway has been the inability to isolate and characterize the HTH protein. Being able to develop an isolation protocol would be beneficial and bring new insight into the HTH's role in *A. thaliana* growth and development.

A recombinant protein isolation protocol can be very challenging to develop since many variables must be considered to optimize the protocol and ensure that enough active protein is isolated (Lee, 2017). Once optimized, this opens the door for further research opportunities. A protein's structure can be determined via x-ray crystallography and its potential enzymatic function can be assessed with various enzymatic assays. An attempt has been previously made to isolate the HTH protein (Chang, 2016). In this case, HTH was expressed in *E. coli* using the pMal-c4x vector but the strategy yielded little to no protein.

This chapter documents various attempts to isolate a recombinant HTH protein, expressed in *E. coli*. The construct design, induction trials, initial isolation attempts, and all troubleshooting steps are discussed. Further optimization is still required, with current work focusing on inducing HTH in a yeast expression system. In an alternate strategy to obtain active HTH, crude protein extracts were obtained from transgenic *A. thaliana* plant lines. These extracts were tested for potential MDL enzymatic activity.

3.2 Materials

The HTH gene was synthesized and cloned into the SUMOstar_Kan vector by GenScript (GenScript Biotech Corporation; Piscataway, New Jersey, United States) and inserted into BL21(DE3) competent cells creating pE_SUMOstar_Kan. The SUMOstar_Kan vector was obtained from LifeSensors (LifeSensors; Devault, Pennsylvania, United States). The GeneJET Plasmid Miniprep Kit was purchased from Thermo Fisher Scientific (Thermo Fisher Scientific; Waltham, Massachusetts, Unites States). All restriction enzymes were purchased from New England Biolabs (New England Biolabs; Ipswish, Massachusetts, United States). The 1 kilobase pair (kd) DNA size ladder was purchased from GeneDireX (GeneDireX Inc; Taiwan). The Protease Arrest and EDTA were purchased from G Biosciences (G Biosciences, St. Louis, Missouri, United States). The Ultracel 10 kDa Ultrafiltration Disc was purchased from MilliporeSigma (Millipore Sigma; Burlington, Massachusetts, Unites States). The Ni-NTA Agarose was purchased from Qiagen (Qiagen; Germany). The Bio-gel P-6 DG (polyacrylamide desalting gel with 6,000 molecular weight exclusion limit) media was purchased from Bio-Rad (Bio-Rad; Hercules, California, United States). The UNOsphere Q and the Bio-Scale Mini UNOsphere S 5 mL cartridge were purchased from Bio-Rad (Bio-Rad; Hercules, California, United States). The Q5 Site-Directed mutagenesis kit was purchased from New England Biolabs (New England Biolabs; Ipswish, Massachusetts, United States). All primers were synthesized by Sigma (Sigma-Aldrich Life Science; St. Louis, Missouri, United States), unless otherwise specified. The LC1 Grower Mix was purchased from SunGro Sunshine (SunGro Sunshine; Alberta, Ontario). The mandelonitrile lyase from Almonds was purchased from Sigma (Sigma-Aldrich Life Science; St. Louis, Missouri, United States). The mandelonitrile was purchased from Sigma (Sigma-Aldrich; St. Louis, Missouri, United States). The anti-GFP antibody was

purchased from Abcam (Abcam Biotechnology Company; Cambridge, United Kingdom). The anti-rabbit IgG antibody conjugated to horseradish peroxidase was purchased from Sigma (Sigma-Aldrich; St. Louis, Missouri, United States). All other general lab chemicals were of the highest grade available and purchased from various companies.

3.3 Methods

3.3.1 Synthetic HTH cDNA Sequence Design

The HTH peptide sequence was input into the SignalP 4.1 server to determine if HTH has a signal peptide (http://www.cbs.dtu.dk/services/SignalP-4.1/). The output graph indicated the C-score, S-score and Y-score of the first 70 residues in the peptide sequence (Petersen, Brunak, Heijne, & Nielsen, 2011). The first 19 residues are predicted to comprise the signal peptide. The predicted peptide sequence of the synthetic cDNA was aligned with the WT HTH peptide sequence to ensure proper design. The nucleotide bases that code for this region of the HTH protein were therefore omitted in the synthetic HTH cDNA sequence, to optimize for protein expression in bacteria and facilitate isolation. The HTH protein without this signal peptide has a predicted mass of 63 kDa. The synthetic cDNA was synthesized and cloned by GenScript (GenScript Biotech Corporation; Piscataway, New Jersey, United States) into the pE_SUMOstar_Kan vector (HTH_SUMO) (Appendix A) and the pET24a+ vector (HTH_pET24a+) (Appendix B).

3.3.2 Heat Shock Bacteria Transformation

A 100 µL aliquot of BL21(DE3) competent cells was gently mixed with 5 ng of HTH_SUMO vector DNA and incubated on ice for 30 minutes. Cells were heat shocked for 30 seconds at 42°C and added to 400 µL Luria broth (LB) (1 mg/mL tryptone, 1 mg/mL NaCl and 0.5 mg/mL yeast extract) then incubated at 37°C for 1 hour. Cells were spread on LB agar plates

supplemented with 50 μg/mL kanamycin, 40 μg/mL X-gal and 50 μg/mL isopropyl β-D-1-thiogalactopyranoside (IPTG) then incubated overnight at 37°C. White colonies were picked and added to separate flasks containing 50 mL LB supplemented with 50 mg/mL kanamycin then incubated overnight at 37°C. A 25% glycerol stock was prepared by mixing 250 μL of the cell resuspension with 250 μL 50% glycerol. The glycerol stock was stored at -80°C. The same heat shock protocol was repeated to transform BL21(DE3) with HTH_pET24a+.

3.3.3 Bacterial Plasmid Minipreparation

Bacterial plasmid DNA was isolated from transformed BL21(DE3) cells using the GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific; Waltham, Massachusetts, United States). 5-10 mL of BL21(DE3) transformed with HTH_SUMO was resuspended via vortex in 250 μL resuspension solution supplemented with 2.5 μL RNaseA. 250 μL lysis solution was added and mixed by inversion and 350 μL neutralizing solution was immediately added to the lysate. The cells were mixed by inversion and incubated on ice for 5 minutes before centrifugation for 5 minutes at 13,000 × g. Up to 600 μL of the supernatant was applied to a Thermo Scientific GeneJET Spin column. 185 μL wash solution supplemented with 315 μL 100% EtOH was added to the spin column and centrifuged for 1 minutes at 13,000 × g. The wash step was repeated a second time. 50 μL elution buffer was added directly to the spin column membrane and incubated at room temperature for 2 minutes before centrifugation for 2 minutes at 13,000 × g. The final column flow through, containing the plasmid, was stored at -20°C.

3.3.4 HTH_SUMO Restriction Endonuclease Digestion

HTH_SUMO plasmid isolated from transformed BL21(DE3) was digested with restriction endonucleases to ensure that the HTH_SUMO vector taken up by the competent cells was unaltered. One sample of plasmid was incubated with *Nco*I (New England Biolabs; Ipswich,

Massachusetts, United States), another sample was incubated with *Nru*I (New England Biolabs), and a final sample was incubated with both restriction endonucleases. All samples as well as an undigested HTH_SUMO plasmid were run on a 0.8% agarose gel casted with 0.1 μg/mL ethidium bromide (Thermo Fisher Scientific). DNA bands were visualized under UV light. A 1 Kb DNA size ladder (GeneDireX Inc.; Taiwan) run concurrently on the same gel was used to determine the apparent band sizes. Colonies with plasmids yielding digested bands of the appropriate size were kept and store at -80°C.

The same diagnostic digest protocol was used on HTH_pET24A+ plasmid isolated from BL21(DE3) using endonucleases *Xho*I (New England Biolabs) and *Nde*I (New England Biolabs). This protocol was also used on *hth* mutant alleles generated using the Q5 Site-Direted Mutagenesis kit (New England Biolabs) cloned into the pE_SUMOstar_Kan vector using endonucleases *Cla*I (New England Biolabs) and *Xho*I (New England Biolabs).

3.3.5 HTH Protein Autoinduction Set up

50 mL of autoinduction media (1 mg/mL tryptone, 0.5 mg/mL yeast extract, 1M $MgSo_4$, 0.1% (v/v) 1000 × TRACE elements (50 mM $FeCl_3$, 20 mM$CaCl_2$, 10 mM MnCl2, 10 mM $ZnSO_4$, 2 mM $CoCl_2$, 2mM $CuCl_2$, 2mM $NiCl_2$, 2mM Na_2MoO_4, 2mM Na_2SeO_3, 2mM H_3BO_3), 5% (v/v) 20 × NPS (66 mg/mL $(NH_4)_2SO_4$, 136 mg/mL KH_2PO_4, 142 mg/mL Na_2・HPO_4), 1% (v/v) 50% glycerol, 0.2% (v/v) 25% glucose, 2% (v/v) 10% lactose) was supplemented with 50 mg/mL kanamycin and 1 mL of a 50 mL BL21(DE3) starter culture resuspension successfully transformed with HTH_SUMO (Fox & Blommel, 2017). Induction media inoculated with transformed cells was incubated at 20°C for 24 hours. Media was centrifuged for 10 minutes at 5,100 × g. Cells were resuspended and sonicated using the QSonica Q500 sonicator fitted with the standard ½ inch diameter probe (QSonica LLC; Newtown, Connecticut, Unites States). The

cells were sonicated for a total of 2 minutes at 75% amplitude. This 2-minute length of time was divided into 5 second pulses with a 55 second interval between each one. Therefore, the total sonication procedure took 24 minutes. The cell lysate was subsequently centrifuged for 45 minutes at 12,000 × g and a sample of the supernatant and the pellet were run on a 12% SDS-PAGE gel to determine the protein induction profile.

3.3.6 Nickel Chromatography of HTH_SUMO Protein

The induction protocol outlined in Section 3.3.5 was scaled up by increasing the autoinduction media volume to 4 L and the HTH_SUMO transformed BL21(DE3) starter culture volume to 100 mL. Cell pellets were stored at -80°C prior to lysis and purification, when necessary, and thawed on ice before use. Cells were resuspended in 100 mL Buffer 1 (25 mM HEPES, 10% glycerol, 200 mM NaCl, 10 mM imidazole, pH 7.5) supplemented with 1 × BioScience Protease Arrest and 5 mM ethylenediaminetetraacetic acid (EDTA) (G Biosciences, St. Louis, Missouri, United States). Resuspended cells were sonicated using the same protocol outlined in 3.2.8 and subsequently centrifuged for 45 minutes at 20,000 × g. 360 mg/mL ammonium sulfate ($(NH_4)_2SO_4$) was added to the supernatant and mixed for 30 minutes before centrifugation for 30 minutes at 20,000 × g. The pellet was resuspended in 50 mL Buffer 1.

20 mL Ni-NTA Agarose was added to a column (5 cm × 20 cm) (Qiagen; Germany). Before use, the nickel resin was stripped and recharge. The column was washed with 200 mL 0.2 M acetic acid and 6 M Guanidine hydrochloride and rinsed with 500 mL dH_2O. The column was then washed with 300 mL 2% SDS solution and rinsed with 400 mL dH_2O. The column was then washed with 500 mL 70% EtOH and rinsed with 500 mL dH_2O. The column was then washed with 500 mL 100mM EDTA and rinsed with 500 mL dH_2O. The column was washed with 500

mL $NiSO_4$ and rinsed with 500 mL dH_2O. Finally, 300 mL of Buffer 1 was passed through the column.

The pellet resuspension was mixed with the nickel resin for 30 minutes and applied to a column (5 cm × 20 cm). Buffer 1 was washed through the nickel column until the eluate reached an OD_{280} below 0.1. 2 mL eluate fractions were collected as Buffer 2 (25 mM HEPES, 300 mM imidazole, pH 7.5) was washed through the nickel column. All fractions with an OD_{280} above 0.1 were mixed together and concentrated to 10 mL using an Amicon stir cell fitted with an Ultracel 10 kDa Ultrafiltration Disc (Millipore Sigma; Burlington, Massachusetts, Unites States), pressurized under compressed nitrogen gas. The concentrated protein sample was added to a column (1.5 cm × 30 cm) filled with Bio-Gel P-6 DG Media (Bio-Rad). 2 mL eluate fractions were collected as Buffer 3 (25 mM HEPES, pH 7.5) was washed through the nickel column. All fractions with an OD_{280} above 0.1 were pooled. 1 mL of isolated SUMO protease was mixed overnight at 4°C with the pooled fractions to cleave off the 6 × His tag. The nickel column was washed with 300 mL Buffer 3. The protein solution was mixed with the nickel resin on column for 30 minutes. 2 mL eluate fractions were collected as Buffer 3 was washed through the nickel column. All fractions with an OD_{280} above 0.1 were mixed together and concentrated down to 10 mL using a same Amicon stir cell fitted with an Ultracel 10 kDa Ultrafiltration Disc (Millipore Sigma), pressurized under compressed nitrogen gas. The sample was concentrated down to 1 mL using a 30 kDa Amicon Ultra-15 Centrifugal Filter Unit centrifuged for 10 minutes at 2,800 × g. Samples were flash frozen in 20 µL aliquots using liquid nitrogen and stored at -80°C. All steps were performed at 4°C or on ice, unless otherwise specified.

3.3.7 Anion Exchange Chromatography

Protein fractions isolated via nickel chromatography were thawed on ice. The P-6 DG column was washed with 300 mL Buffer A (25 mM HEPES (pH 7.5)). The thawed protein fractions were added to the column. 2 mL eluate fractions were collected as Buffer A was washed through the column. All fractions with an OD_{280} above 0.1 were mixed together. Up to 1 mL of pooled eluate was injected into the BioLogic Duo Flow Chromatography System (FPLC) set up with the UNOsphere Q cartridge (Bio-Rad; Hercules, California, United States). 35 mL of Buffer A was pumped through the FPLC at a rate of 1.5 mL/min. The protein was injected onto the UNOsphere cartridge at a rate of 1.5 mL/min. 15 mL of Buffer A was pumped through the FPLC at a rate of 0.25 mL/min. A gradient change from Buffer A to Buffer B (25 mM HEPES, 0.5 M NaCl, pH 7.5) was set at a rate of 1.5 mL/min for 100 mL until only Buffer B was pumping through the FPLC for another 35 mL. 2 mL fractions were collected throughout the protocol. A UV sensor was zeroed and used to measure the presence of protein as the solution was pumped out of the machine. All aforementioned cation exchange chromatography steps were performed at 4°C or on ice, unless otherwise specified. The fractions containing protein were run on a 12% SDS-PAGE gel.

3.3.8 Cation Exchange Chromatography

Protein fractions isolated via nickel chromatography were thawed on ice. The P-6 DG column was washed with 300 mL Buffer A (25 mM HEPES (pH 7.5)). The thawed protein fractions were added to the column. 2 mL eluate fractions were collected as Buffer A was washed through the column. All fractions with an OD_{280} above 0.1 were mixed together. Up to 1 mL of pooled eluate was injected into the BioLogic Duo Flow Chromatography System (FPLC) set up with the Bio-Scale Mini UNOsphere S 5 mL cartridge (Bio-Rad). 35 mL of Buffer A was

pumped through the FPLC at a rate of 0.5 mL/min. The protein was injected onto the UNOsphere cartridge at a rate of 0.25 mL/min. 15 mL of Buffer A was pumped through the FPLC at a rate of 0.25 mL/min. A gradient change from Buffer A to Buffer B (25 mM HEPES, 0.5 M NaCl, pH 7.5) was set at a rate of 0.25 mL/min for 100 mL until only Buffer B was pumping through the FPLC for another 35 mL. 2 mL fractions were collected throughout the protocol. A UV sensor was zeroed and used to measure the presence of protein as the solution was pumped out of the machine. All aforementioned cation exchange chromatography steps were performed at 4°C or on ice, unless otherwise specified. The fractions containing protein were run on a 12% SDS-PAGE gel.

The protein fractions were also passed through the same cation exchange column at pH 6. The same protocol was used with the exception of the buffer. A 25 mM MES buffer (pH 6) was used for buffer A and a 25 mM MES buffer (pH 6) with 0.5 M NaCl was used for Buffer B.

3.3.9 Protein Solubilization via Arginine Treatment

BL21(DE3) cells transformed with HTH_SUMO were induced using the protocol outlined in Section 3.3.2. Six aliquots of 1 mL induced started culture were used. Once the cells were lysed and the supernatant was removed, 500 µL PBS (pH 7.5) was added to the pellets. Arginine was added to each aliquot, at increasing concentrations: 0M, 0.25 M, 0.5 M, 1M, 1.5M and 2M. Solutions were mixed via pipette and incubated at 4°C overnight. The solution was centrifuged for 30 minutes at 12,000 × g and a sample of the supernatant and the pellet were run on a 12% SDS-PAGE gel to determine the protein profile.

3.3.10 Mass Spectrometry Analysis

Mass spectrometry analysis was performed on samples of the final fraction obtained from nickel chromatography to determine the protein profile. The protein sample was run on a 4%

SDS-PAGE gel. Five bands around the 70 kDa range were cut out of the acrylamide gel and sent to the Sick Kids' SPARC BioCentre Molecular Analysis facility (Toronto Hospital for Sick Children; Ontario, Canada). The protein profile of each band was visualized on the Scaffold Viewer software (Proteome software; Portland, Oregon, United States).

3.3.11 Nickel Chromatography using HTH_pET24a+

The induction protocol outlined in Section 3.3.5 was scaled up by increasing the autoinduction media volume to 4 L and the HTH_pET24a+ transformed BL21(DE3) starter culture volume to 100 mL. Cell pellets were stored at -80°C prior to lysis and purification, when necessary, and thawed on ice before use. Cells were resuspended in 100 mL Buffer 1 (25 mM HEPES, 10% glycerol, 200 mM NaCl, 10 mM imidazole, pH 7.5) supplemented with 1 × BioScience Protease Arrest and 5 mM ethylenediaminetetraacetic acid (EDTA) (G Biosciences). Resuspended cells were sonicated using the same protocol outlined in 3.1.8. The cell lysate was subsequently centrifuged for 45 minutes at 20,000 × g. 360 mg/mL. $(NH_4)_2SO_4$ was added to the supernatant and mixed for 30 minutes before centrifugation for 30 minutes at 20,000 × g. The pellet was resuspended in 50 mL Buffer 1. The nickel column was prepared as previously described in Section 3.3.6.

The resuspended pellet was mixed with the nickel resin for 30 minutes and applied to a column (5 cm × 20 cm). Buffer 1 was washed through the nickel column until the eluate reached an OD_{280} below 0.1. Two mL eluate fractions were collected as Buffer 2 (25 mM HEPES, 300 mM imidazole, pH 7.5) was washed through the nickel column. All fractions with an OD_{280} above 0.1 were mixed together and concentrated to 10 mL using an Amicon stir cell fitted with an Ultracel 10 kDa Ultrafiltration Disc, pressurized compressed nitrogen gas. The concentrated protein sample was added to a P-6DG column. 2 mL eluate fractions were collected as Buffer 3

(25 mM HEPES, pH 7.5) was washed through the nickel column. All fractions with an OD_{280} above 0.1 were mixed together and concentrated down to 10 mL using a same Amicon stir cell fitted with an Ultracel 10 kDa Ultrafiltration Disc, pressurized under gaseous nitrogen from a compressed nitrogen tank. The sample was concentrated down to 1 mL using a 30 kDa Amicon Ultra-15 Centrifugal Filter Unit centrifuged for 10 minutes at 2,800 × g. Samples were flash frozen in 20 μL aliquots using liquid nitrogen and stored at -80°C. All steps were performed at 4°C or on ice, unless otherwise specified.

3.3.12 Chemical Competent *Saccharomyces cerevisiae* Transformation

HTH cDNA without the nucleotides coding for the signal peptide, obtained from *A.thalian* mRNA reverse transcription (Chang, 2016), was PCR amplified using the forward 5'-ATTCGCAGCATCCTCCGCATTAGCTGGGGGATCCACTAGTTCCACTGCCTCTAAAGGTAAAG-3' and reverse 5'- GGTGCTCGAGGCCCTGAAAATACAGGTTTTCTGCGGCCGCACTAAACACCAGCTTTGTTTC-3' primers (Sigma-Aldrich Life Science; St. Louis, Missouri, United States). 15 μL PCR product was mixed with 5 μL fish sperm DNA, 5 μL linearized and dephosphorylated pPinkNN vector (Nava & Amidi, unpublished results) and 200 μL chemical competent *S. cerevisiae* (Amidi & Heinisch, unpublished results). The cell mixture was gently mixed and incubated at 37°C for 10 minutes. 1 mL Solution B (40% PEG 1500, 100 mM Bicine pH 8.4) was added to the suspension, gently mixed and incubated at 30°C for 1 hour. The suspension was centrifuged, and excess supernatant was removed. The pellet was resuspended in the remaining 200 μL supernatant then plated on $Ura3^-$ agar plates and incubated at 30°C for 3 days.

3.3.13 HTH cDNA Q5 Site-Directed Mutagenesis

The Q5 Site-Directed Mutagenesis kit (New England Biolabs) was used to introduce mutations in the synthetic HTH DNA vector (HTH_SUMO) described in Section 3.3.1. Primer sets were designed using the NEBaseChanger software (https://nebasechanger.neb.com/). Mutations were introduced following the mutagenesis kit protocol and the mutated HTH_SUMO vectors transformed via heat shock into DH 5-α competent cells. Cell mixture was plated on LB agar plates supplemented with 50 μg/mL kanamycin, 40 μg/mL X-gal and 50 μg/mL IPTG and incubated at 37°C overnight. White colonies were picked and added to separate flasks containing 50 mL LB supplemented with 50 mg/mL kanamycin then incubated overnight at 37°C. A 25% glycerol stock was prepared by mixing 250 μL of the cell resuspension with 250 μL 50% filtered sterilized glycerol. The glycerol stock was stored at -80°C.

3.3.14 DNA Sequence Analysis

Six primers were designed specifically for sequencing HTH_SUMO vectors (Sigma-Aldrich). 5'-AAGAAGGCTGAT GGAAGCGT-3', 5'- CTGACCCTGTGGCAAAAAGC-3', and 5'-TTTTGGTCAGAGCCCGG AAA-3' anneal to the DNA in the forward direction. 5-TTTCACCAGCTTCGGATCCC-3', 5'-AATACCGCTCAGCATCAGCA-3' and 5'-CCCATTCGCCAATCCGGATA-3' anneal to the DNA in the reverse direction (Appendix I). Primers were spaced no more than 1000 bp apart to obtain a full coverage of the cDNA sequence. Sequencing was performed by The Centre for Applied Genomics (Toronto Hospital for Sick Children; Ontario, Canada). Sequencing results were analysed to ensure correct mutations were introduced into the HTH cDNA sequence.

3.3.15 Plant material and Growth Conditions

Previously, the *HTH$_{pro}$:HTH-G3GFP* construct (Chang, 2016) was cloned into the pGWB650 vector (Akamura et al., 2010) and the pCAMBIA vector (Sinclair et al., 2009) then transformed into *hth-9* mutant plants to create the P650 and P103 plant lines, respectively (Figure 19). *hth-9 A. thaliana* seeds, in the Wassilewskija ecotype, were previously described by Lolle et al. (1998). The pGWB650 vector contains the 2,834 bp *HTH* gene and its 2,009 bp putative promoter with a C-terminal green fluorescent protein tag, a modified GFP protein (Chang, 2016; Kawakami & Watanabe, 1997). The pCAMBIA vector contained the *HTH* gene under the control of the CaMV 35S promoter with a C-terminal G3 green fluorescent protein tag (Chang, 2016). Plant lines were also transformed with the empty vectors and denoted as P650EV and P103EV respectively.

Seeds from these four plant lines, as well as untransformed *hth-9* seeds, were sown onto moistened Sun Gro Sunshine LC1 Grower Mix (Sun Gro Sunshine; Alberta, Canada) and stratified at 4°C for 4 days. Plants were maintained in a Conviron PGW36 growth chamber at 20-23°C and illuminated with a mixture of incandescent and fluorescent lights (140 – 170 μmol m^{-2} sec^{-1} at pot level) with a 24-hour photoperiod. Plants were watered as needed, typically three times a week.

3.3.16 Floral Tissue Crude Protein Extraction

Plants were collected from the growth chamber once floral buds were formed and ready to open, which occurred approximately 4 weeks after being sown. Fresh floral bud tissue was homogenized in a small glass homogenizer using a 2:1 Extraction buffer (100 mM Tris pH7.5, 150 mM NaCl, 10% glycerol, 0.1% Triton-X, 1X BioScience ProteaseArrest) to floral tissue ratio (2 mL extraction buffer to 1 mg tissue) to help extract the protein from the tissue. The tissue

homogenate was centrifuged twice for 30 minutes at 12,000 × g to remove insoluble plant material. 180 mg/mL $(NH_4)_2SO_4$ was mixed with the supernatant and incubated for 30 minutes at 4°C and the supernatant collected after centrifugation 30 minutes at 12,000 × g. 420 mg/mL $(NH_4)_2SO_4$ was mixed with the supernatant, incubated for 30 minutes at 4°C and the pellet collected by centrifugation (30 minutes at 12,000 × g). The final pellet fraction was resuspended in 200 μL extraction buffer.

3.3.17 *In vitro* Mandelonitrile Lyase Enzymatic Assay

To determine if HTH had hydroxynitrile lyase enzymatic activity, the final pellet fraction from the crude floral protein extract was tested for activity using an *in vitro* enzymatic assay (Jorns, 1979). Crude extracts from the P650, P650EV, P103, P103EV and untransformed *hth-9* plant lines were used for the assay. A positive control using a commercially available MDL extracted from Almond (pH 5.4; Sigma-Aldrich Life Sciences) was also included in the assay. Recombinant HTH protein induced in *E. coli* and isolated via nickel chromatography were also tested for MDL enzymatic activity.

Enzymatic activity was determined by monitoring cleavage of mandelonitrile into benzaldehyde and HCN at 25°C by following the protocol developed by Jorns (1979). The assay contained 95 mM sodium acetate (pH 5.4), 2.8 mM mandelonitrile, 1.6% (v/v) EtOH and 0.2 unit MDL. The absorbance of the assay solution was monitored at 275 nm in a plastic cuvette with a 1 cm light path until readings stabilized, roughly 3 minutes. The reaction was initiated by addition of mandelonitrile (Sigma-Aldrich Life Sciences). The OD_{275} of the reaction was continuously monitored for 15 minutes using the Cary Win UV Kinetics software (Agilent Technologies; Mississauga, Ontario, Canada) and a Cary 100 UV-Vis spectrometer (Agilent Technologies). Enzymatic activity was calculated using the slope of ΔOD_{275nm}. The change in

absorbance from the moment mandelonitrile was added to the assay until 15 minutes after was measured by using the "rule of least squared" cursor option provided in the UV Kinetics software (Agilent Technologies).

3.3.18 Western Blot Analysis of Floral Bud Crude Protein Extracts

The homogenate and final pellet fraction of the P650 crude protein extracts and the homogenate of the *hth9* crude protein extracts were used for protein immune-detection via Western blotting. The total protein concentration was determined using the Quick Start Bradford 1 × Dye Reagent (BioRad) according to the protocol (Bio-Rad, 2012). 30 μg of protein was run on a 10% SDS-PAGE gel then transferred to a polyvinylidene fluoride transfer membrane soaked in transfer buffer (25 mM Tris, 192 mM Glycine, 10% methanol) supplemented with 0.1% SDS. The transfer was performed on a Trans-Blot SD Semi-Dry Transfer cell (Bio-Rad). The membrane was stained with Ponceau-S (0.1% (w/v) in 5% acetic acid) to visualize protein bands and then washed with milliQ water before soaking in blocking buffer (TBS-T (20 mM Tris pH 7.5, 300 mM NaCl, 0.1% (v/v) Tween-20) plus 5% skimmed milk). The membrane was sealed in an airtight bag and incubated overnight at 4°C in 2 mL blocking buffer supplemented with 1:3,000 anti-GFP antibody (Biotechnology Company; Cambridge, United Kingdom; ab6556). The membrane was washed with TBS-T then incubated in blocking buffer supplemented with 1:30,000 anti-rabbit IgG antibody conjugated to horseradish peroxidase (Sigma-Aldrich Life Sciences; A0545) for 1 hour at room temperature. The membrane was washed with TBS-T. Clarity Western ECL Substrates (BioRad) were mixed and added to the membrane and chemiluminescence visualized on the ChemiDoc MP Imaging System (BioRad) running the Image Lab 5.2.1 software.

3.4 Results & Discussion

This research attempted to improve recombinant HTH protein yield compared to previous attempts which used a maltose-binding protein tag (Chang, 2016), through transformation of a synthetic HTH cDNA sequence designed for optimal bacterial expression using the pE_SUMOstar vector. While the attempts to isolate the HTH protein were unsuccessful, important research implications, including the potential factors inhibiting the isolation of HTH, and corresponding implications for HTH functionality, were revealed. Overall, the findings from this research indicate that the most likely reason for unsuccessful HTH isolation stems from the use of *E. coli* as an expression system. *E. coli* is an organism commonly used for recombinant protein production, isolation and activity assays due to it having well characterized genetics, as well as being inexpensive, easy to work with and efficient at protein production in comparison to other systems (Baneyx, 1999). This research aimed to improve protein yields induced in this bacterial expression system. This research has created a foundation for work in a eukaryotic system and has identified concrete next steps for future research.

3.4.1 HTH Protein Construct Design

A significant bottleneck in assigning a definitive role for HTH in the cuticle synthetic pathway is the inability to isolate and characterize the HTH protein. Previous attempts to express HTH in *E. coli* using the pMal-c4x vector yielded little to no protein (Chang, 2016).

To try and improve yields, a synthetic HTH cDNA sequence was designed for optimal expression in bacteria, omitting the nucleotides coding for the predicted signal peptide, as determined by the SignalP 4.1 data (Figure 9) (Petersen et al., 2011). SignalP uses an artificial neural network and a hidden Marvol model prediction to determine the signal peptide of a peptide sequence (Nielsen, 2017) SignalP 4.1 generated a C-score, S-score and Y-score for the

HTH cDNA sequence to predict the location of the signal peptide and its cleavage site. The C-score indicates the location of the signal peptide cleavage site. It is expected to be high at the position immediately after the cleavage site. The S-score determines if an amino acid is a part of the signal peptide. An amino acid with an S-score close to 1 is predicted to be a part of the signal peptide. A sharp drop in the S-score is expected after the cleavage site. The Y-score uses the C-score and S-score values to give a better prediction of the signal peptide cleavage site. It considers the high peaks from the C-score and finds the one located in the steep slope generated by the S-score to determine the cleavage site (Petersen et al., 2011). Based on the SignalP results presented in Figure 9, it is highly likely that the HTH protein has a signal peptide spanning residues 1-19.

Synonymous nucleotide changes were introduced to optimize codons for bacterial expression while preserving the amino acid sequence of the HTH protein. These changes take many factors into account to ensure biased codons are used and key sequences that cause translational pausing are removed (Kudla, Murray, David, & Plotkin, 2009; Li, Oh, & Weissman, 2012). A sequence alignment between the HTH peptide sequence and the synthetic peptide shows perfect amino acid sequence alignment with the exception of the first 19 amino acids at the N-terminus of the WT sequence, the predicted signal peptide (Figure 10). The removal of these amino acids should have no effect on the protein structure or expression in *E. coli* since, under native condition, it is predicted to be cleaved off in plant cells once it is localized to the ER. It should be noted that the HTH protein with the signal peptide removed has a predicted mass of 63 kDa, instead of 65 kDa.

The pE_SUMOstar vector from GenScript was chosen in hopes that it would help with protein induction and yield active soluble protein (Zuo et al., 2005). This vector is fully

compatible with the pET expression system and has also been shown to aid with protein expression and solubility (Malakhov et al., 2004; Marblestone et al., 2006; Zuo et al., 2005). This system uses a vector that has been engineered to work with mutated *E. coli* strains denoted as BL21(DE3) and ensures that the protein of interest is only expressed when induced with lactose or a lactose mimic such as IPTG, allowing for a highly regulated protein production system (Sørensen & Mortensen, 2005). The vector is designed to insert a 12 kDa N-terminal SUMO fusion protein (Figure 13A) that acts as a folding chaperone containing a 6 × His tag to facilitate protein purification. The SUMO fusion can be cleaved off during the purification with a SUMO protease creating a final protein product with no additional N-terminal amino acids. The SUMO fusion protein and SUMO protease are removed during a secondary pass through a nickel resin. The protein of interest will elute because it no longer has an affinity for the nickel resin.

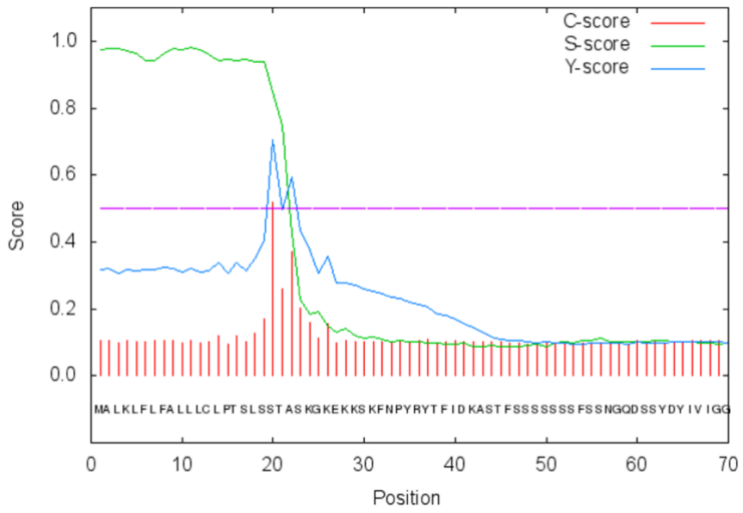


Figure 9: HTH signal peptide prediction. The HTH peptide sequence was inputted into the SignalP 4.1 Server to predict the presence of a signal peptide sequence. The C-score, S-score and Y-score of the first 70 residues are graphed. C-score is highest at residue 20 and indicates a cleavage site between residues 19 and 20. The S-score indicates the presence of a signal peptide spanning the first 19 residues, with a steep slope downwards starting around residue 20. The Y-score which is at its highest at residue 20, further validate the results obtained from the C-score and S-score. It appears as though the HTH protein has a cleavable signal peptide spanning the first 19 residues.

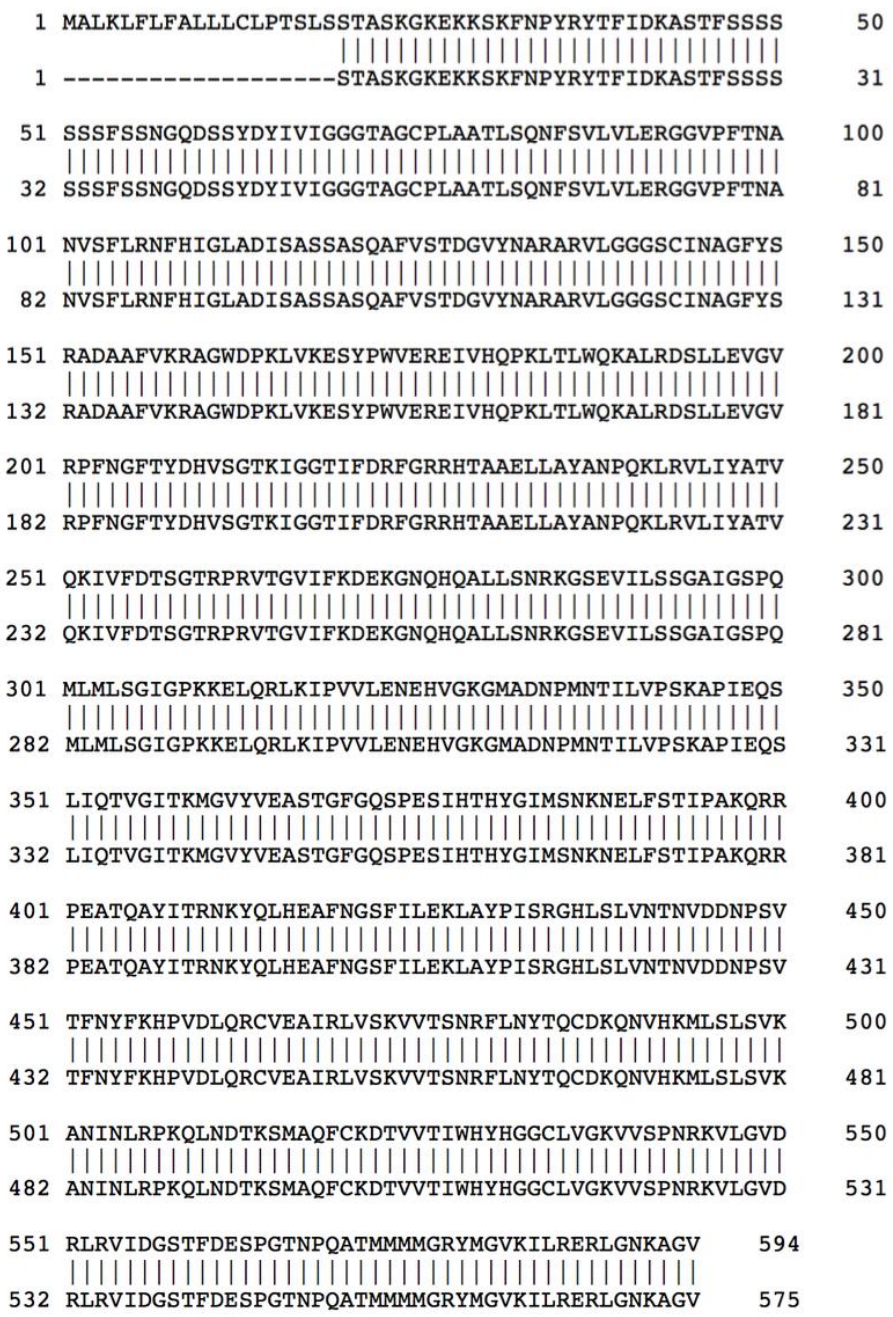


Figure 10: HTH peptide sequence alignment. The peptide sequence translated from the synthetic HTH cDNA sequence (bottom sequence) was compared to the WT HTH plant peptide sequence (top sequence). The two sequences have perfect alignment except for the residues predicted to code for the signal sequence (lower dashed line).

3.4.2 Protein Isolation in *E. coli*

HTH_SUMO plasmids from BL21(DE3) cells were analyzed via diagnostic restriction enzyme digest to ensure the cells were transformed with the plasmid insert (Figure 11). Positive transformants were subsequently used for HTH protein production. Initial induction attempts showed that the pE_SUMOstar system is capable of inducing large quantities of the HTH protein although only a small portion was soluble (Figure 12). By using the cleavable N-terminal 6 × His tag, nickel chromatography was used to isolate the HTH protein from the soluble fraction (Figure 13A). A few non-specific bands still remained after the final elution off the nickel column (Figure 13B). The final concentrated fraction was used to test for MDL activity (Jorns, 1979) and showed no enzymatic activity.

Ion exchange chromatography was used to try isolating HTH from the non-specific bands. Anion exchange was performed at pH 7.5 using the UNOsphere Q cartridge (Bio-Rad). Cation exchange was performed at pH 7.5 and 6 using the UNOsphere S cartridge (Bio-Rad) None of the ion exchange trials appeared to successfully remove contaminating bands, as exemplified with the cation exchange chromatography performed in MES (Figure 13C). An attempt was made to buffer exchange the recombinant HTH into a pH 5 buffer, but the protein appeared to pellet out of solution during the buffer exchange in the Bio-Gel P-6 DG resin.

3.4.3 Protein Solubilization via Arginine Treatment

Based on the initial induction pattern shown in Figure 12, a significant amount of protein was present in the pellet however, very little seemed to be soluble. To try and bring more protein into solution, pellets of induced BL21(DE3) cells transformed with HTH_SUMO were treated with arginine. It has been previously shown that arginine can help with protein refolding, allowing the protien to become soluble (Tischer, Lilie, Rudolph, & Lange, 2010). Once

BL21(DE3) cells transformed with HTH_SUMO were lysed, the pellet was mixed with varying amounts of arginine to attempt solubilization of the protein stuck in the pellet (Appendix G). The arginine was not removed from the pellet before running the sample on the SDS-PAGE, creating inaccurate patterns in the lanes containing those samples. A bit of protein may have come into solution when the pellet was treated with 0.25 M and 1 M arginine, but this was not further investigated. It was decided that other avenues would be explored instead of further pursuing arginine treatment. This could be something to further investigate in the future.

3.4.3 Mass Spectrometry Analysis

Protein bands from the final nickel chromatography fraction were resolved via SDS-PAGE gel (Figure 14A) and sent for mass spectrometry analysis at the Sick Kids' SPARC BioCentre Molecular Analysis facility (Toronto Hospital for Sick Children; Ontario, Canada). The normalized spectral counts for each band are shown in Figure 14B. A normal spectral count is synonymous with the quantitative value of each protein in the individual samples (Lundgren & Hwang, 2010). HTH is present in all samples but is most prominent in band 5 (Table 2). The other proteins are native to *E. coli*, including high levels of chaperone proteins. Considering the initial induction pattern (Figure 12), it is evident that the majority of the induced HTH protein appears to remain in inclusion bodies and does not go into solution, i.e., is soluble. Inclusion bodies form when misfolded protein aggregates together, making it insoluble (Baneyx & Mujacic, 2004). The low levels of HTH protein found in the soluble fraction likely have many chaperone proteins bound to them to aid in their solubilization. Unfortunately, the presence of these chaperones indicates that HTH was largely misfolded even with the inclusion of the SUMO fusion protein, meaning that it was inactive and unusable for structure analysis or for performing

enzymatic activity assays. The isolated protein fractions were also tested for MDL enzymatic activity and showed no evidence of activity.

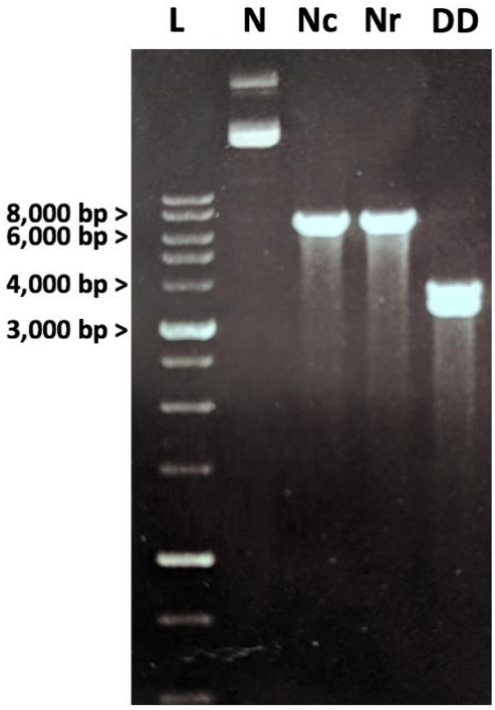


Figure 11: Restriction enzyme digestion of the HTH_SUMO vector. HTH cloned into pE_SUMOstar was isolated from BL21(DE3) competent cells. The banding pattern of the DNA run on a 0.8% agarose gel cast with ethidium bromide was used to determine if transformation was successful. Bands were observed under UV light. The undigested vector (N) will not be linearized. *Nco*I (Nc) and *Nru*I (Nr) both linearize the vector to create a single 7,356 band. Mixing both restriction enzymes with the vector (DD) will create a 3,829 bp band and a 3,527 bp band. All band sizes were determined by comparing to the 1 Kb DNA size ladder (L).

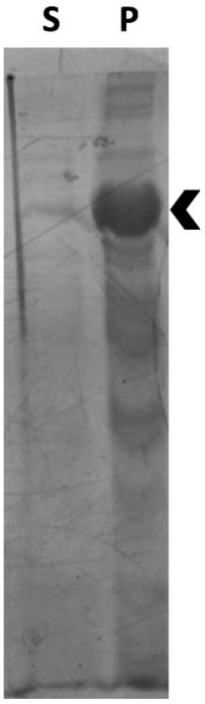


Figure 12: Autoinduction protein profile. BL21(DE3) competent cells transformed with the synthetic HTH cDNA cloned into pE_SUMOstar were induced via autoinduction for 24 hours at 20°C. Cells were lysed. A sample of the soluble fraction (S) and the pellet (P) were run on a 12% SDS-PAGE gel. The HTH protein fused to the SUMO fusion protein has an expected size of 75 kDa (arrowhead). A prominent 75 kDa band can be seen in the pellet fraction. A less prominent band of equal size can be seen in the soluble fraction. Based on this gel, it appears as though the SUMO expression system and autoinduction protocol successfully induces the production of HTH protein, but most is not soluble.

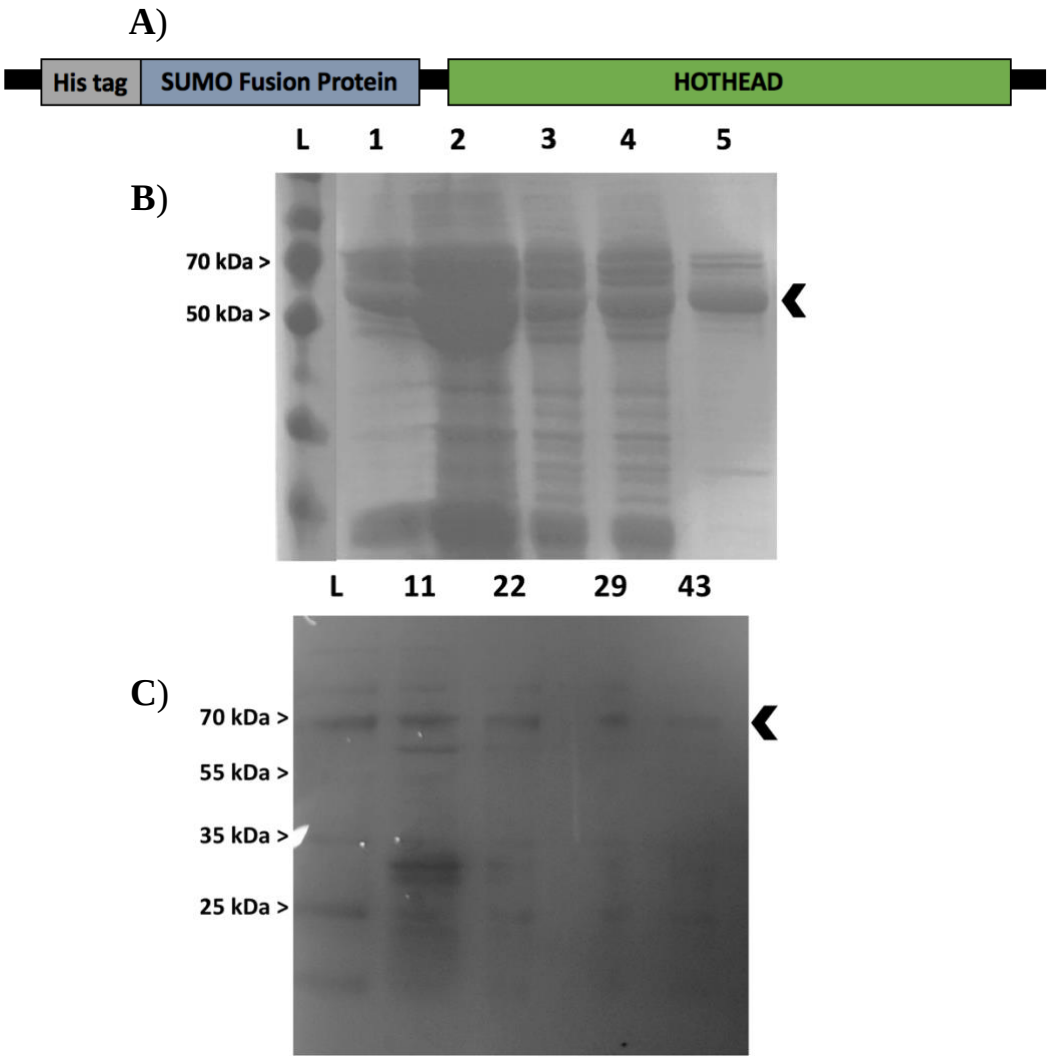


Figure 13: HTH_SUMO protein isolations. (A) Schematic of the synthetic HTH cDNA sequence cloned into the pE_SUMOstar vector. The 6 × His tag and SUMO fusion protein are located at the N-terminus of the HTH protein. (B) Protein fraction samples collected during nickel chromatography run on a 12% SDS-PAGE gel. The HTH protein has an expected size of 75 kDa when fused with the SUMO fusion protein and 63 kDa once the SUMO fusion protein is cleaved off (arrowhead). Bands around the 70 kDa size can be seen eluting off the nickel column (1). These bands remain present when the sample is concentrated down to ~10 mL (2), exchanged into a 25 mM HEPES buffer (3) and after cleavage of the SUMO fusion protein (4). Protein is eluted off the final nickle column with a majority of the protein being around 63 kDa (5). It appears as though the HTH protein was successfully isolated although contaminating bands are also present in the final fraction. All band sizes were determined by comparing to the ladder (L). (C) Protein fraction samples collected during cation exchange in MES pH 6 run on a 12% SDS-PAGE gel. Four fractions (11, 22, 29 and 43) were selected because the UV sensor indicated that they contained higher amounts of protein. The 63 kDa band (arrowhead) can be seen in all fractions and appears to be continuously eluting off the column, never strongly binding to the resin.

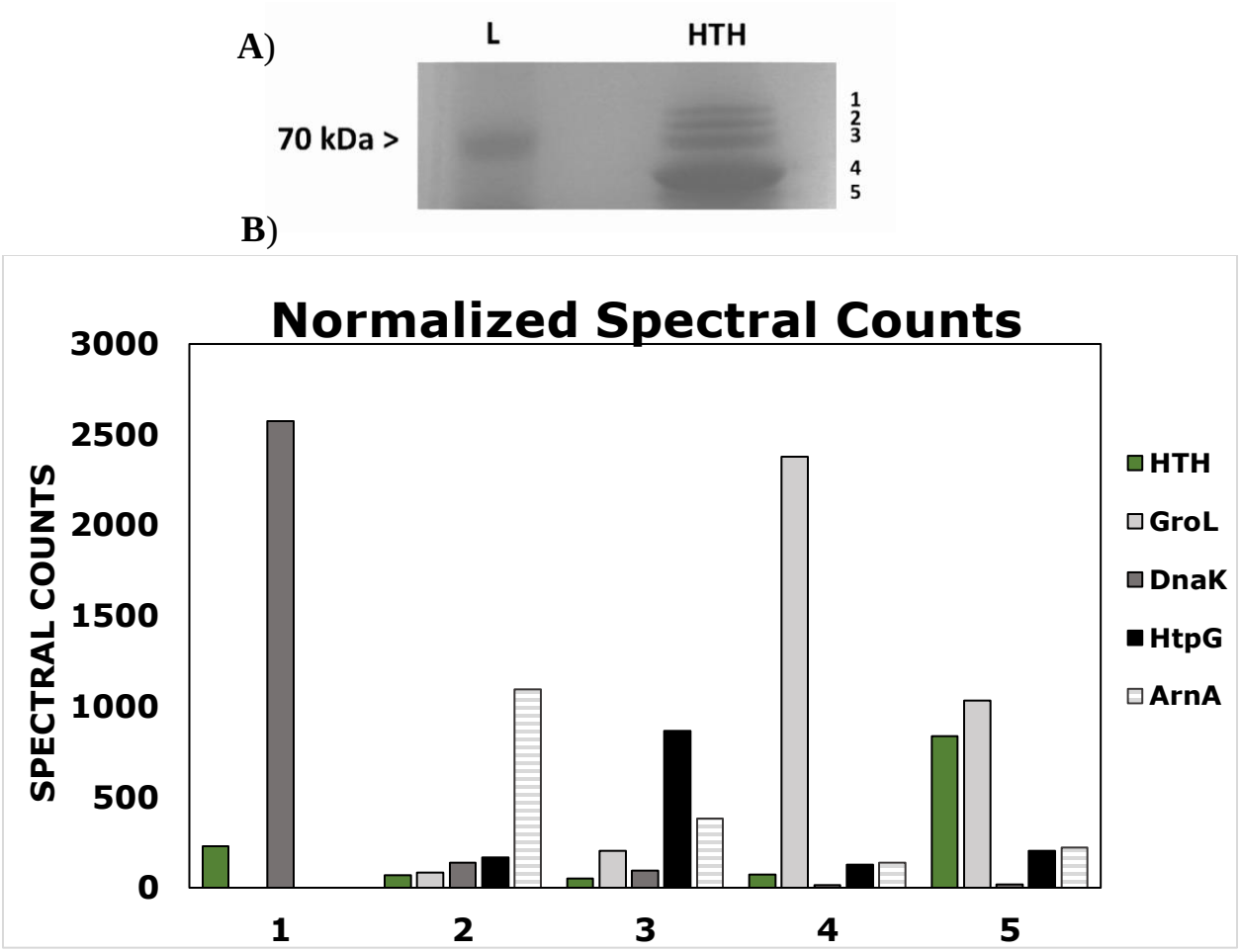


Figure 14: Mass spectrometry analysis of the protein bands expressed in the pE_SUMOstar vector. (A) The final protein fraction after nickel chromatography resolved on a 4% SDS-PAGE gel (HTH). Bands around the 70 kDa range were isolated and sent for mass spectrometry analysis at the Sick Kids SPARC Molecular Analysis facility. Band sizes were determined by comparing to the protein size ladder (L). A total of five bands (1-5) were sent for analysis. (B) Normalized spectral counts of the four most prominent proteins in each band. HTH (green) appears to be present in all bands and most prominent in the fifth band. GroL (light grey), DnaK (dark grey) and HtpG (black) are ArnA (stripped lines) are all proteins native to *Escherichia coli* and appear to be in each sample at varying amounts.

Table 2: Most prominent protein in each band sent for mass spectrometry analysis.

Bands	Most Prominent Protein	Weight
1	Chaperone protein (DnaK)	70 kDa
2	Bifunctioanl polymyxin resistance protein (ArnA)	75 kDa
3	Chaperone protein (HtpG)	71 kDa
4	60 kDa chaperone (GroL)	57 kDa
5	HOTHEAD (HTH)	65 kDa

3.4.4 Potential Factors Inhibiting Protein Isolation from *E. coli*

Although the methods used in this research were purposefully chosen with the goal of maximizing protein expression, one or more factors caused the protein to not isolate in its active form. Therefore, it is necessary to explore other factors in the protein isolation protocol, as many elements may influence the likelihood of successful protein isolation. One such factor that may influence the ability of a particular construct to yield soluble protein is temperature. Proteins can be induced at various temperatures and ideally temperature is optimized at each step of a protein isolation protocol. Warmer temperatures will accelerate the induction step, while a colder temperatures will decelerate the process (Baneyx, 1999). If induction occurs too quickly, it is possible that large amounts of protein may stick together and go into inclusion bodies, thus preventing the protein from going into the soluble fraction. Although induction of the HTH protein was performed at 20°C, it is possible that this temperature was still too high. It is clear that substantial protein was being induced but very little was going into the soluble fraction. Performing induction at a temperature as low as 10°C could be considered in future research. This low-temperature induction method has been successful for aggregation-prone proteins (Vasina & Baneyx, 1997). The high levels of chaperone proteins, however, suggest that HTH was not correctly folded. This indicates that there likely are other factors that inhibit the production of soluble properly folded and active HTH protein.

The components of the induction media could also be a factor affecting protein isolation. The HTH protein is predicted to be a member of the GMC oxidoreductase protein family; these proteins all have a conserved N-terminal domain containing an FAD binding site, suggesting that HTH is an FAD-dependent protein. The lack of FAD in the induction media most likely means that the HTH protein will not fold properly and be insoluble. While FAD may be added to the induction media in attempt to mitigate this issue, the approach used in this research excluded it, as isolation of other FAD dependent proteins with well-established isolation protocols do not include FAD in their induction media (Doubayashi et al., 2011; Maeda et al., 2009, 2010). It is believed that the native levels of FAD in the expression system are sufficient enough to ensure the protein folds properly, remains stable and is enzymatically active. Thus, the induction media seems less likely to be the cause preventing HTH from folding properly.

The peptide sequence of HTH contains a stretch of consecutive serine residues from residues 47 to 56, located close to the N-terminus of the HTH protein. It is possible that this represents an intrinsically disordered domain. Often shown to play important biological roles for proteins, such an unstructured region can bind to other proteins, leading to the formation of inclusion bodies. These protein aggregates are insoluble and prevent the protein from going into the soluble fraction (Baneyx & Mujacic, 2004). The first 56 residues could be removed from the current HTH construct to investigate if this unstructured region is a factor causing low soluble protein yields. It is worth restating that this recombinant protein does not contain the predicted HTH signal peptide. It may be of interest to express the complete HTH peptide sequence in *E. coli* using the pE_SUMO vector expression system to see if results vary.

Another possible explanation for the lack of soluble protein may be the presence of a histidine tag on the N-terminus of the HTH protein. Previous work where *hth-9* mutant *A.*

thaliana were transformed with a N-terminus tagged gene construct failed to rescue the mutant phenotype. Once a C-terminal tagged gateway cassette was used, the rescue was successful, restoring mutant plants to wild type phenotypes (Chang, 2016). High levels of chaperone proteins co-purifying with the recombinant HTH protein (Figure 14; Table 2) and the high levels of protein in the pellet during induction (Figure 12) suggests that HTH is not folded properly. The low levels of HTH protein found in the soluble fraction likely have many chaperone proteins bound to them to aid in their solubilization. This suggests that HTH is highly unfolded. The lack of MDL enzymatic activity also suggests that HTH is not active. Thus, it was hypothesized that the pE_SUMOstar vector may be an inappropriate expression vector for isolating HTH since it adds a 6 × His Tag and the SUMO fusion protein to the N-terminus of the HTH protein. As discussed below, another vector was therefore chosen that only adds tags to the C-terminal end of the protein.

The pET24a+ vector uses the pET expression system (Studier & August, 1991). This is the same induction system used when HTH was cloned into the pE_SUMOstar vector. This means that the same induction protocol can be used. This vector adds a 6 × His Tag on the C-terminal end of the protein of interest but lacks the SUMO fusion protein (Figure 16A; Appendix B), removing the capability to cleave off the 6 × His Tag. The HTH protein without the signal peptide and the addition of the 6 × His tag should have a mass of 64 kDa. Plasmids from BL21(DE3) competent cells transformed with HTH cloned into the pET24a+ were analyzed via diagnostic restriction enzyme digest to ensure the plasmids isolated from transformed cells contained the desired recombinant HTH construct (Figure 15). HTH protein production was induced in HTH_pET24a+ transformed cells and HTH was isolated via nickel chromatography. Once the final fraction was run on an SDS-PAGE gel (Figure16B), the banding pattern was

compared to the final fraction obtained from the HTH_SUMO protein isolation (Figure 13B). It is apparent that more protein bands are present in the HTH_pET24a+ protein isolate. This is due to the fact that this vector does not contain the cleavable SUMO fusion protein, the 6 × His tag cannot be cleaved off and HTH cannot be negatively selected for with the nickel column. Taking a closer look around the expected size range of HTH in the SDS-PAGE (Figure 16B), it appears as though the few co-purifying proteins are still present, solely based on the protein band patterns. The presence of these co-purifying proteins suggests that the chaperone proteins are most likely still present and that HTH protein is not active. The final HTH_pET24a+ protein fraction was not tested for MDL enzymatic activity. In order to have an accurate comparison between the banding pattern of the final HTH_SUMO fraction and HTH_pET24a+, both should be run on the same gel and compared directly. Protein fractions should also be tested for MDL enzymatic activity.

Although *E. coli* is commonly used for protein expression, it does have limitations, one of which being that bacterial cells cannot perform many types of post-translational modifications. Bioinformatic work on HTH has shown that it potentially has 4 phosphorylation sites and 5 N-linked glycosylation sites (Chang, 2016). It is possible that these post-translational modifications are essential for the proper folding of the HTH protein. Therefore, expressing the HTH protein in a prokaryotic system may prevent it from folding properly and being isolated in its active form.

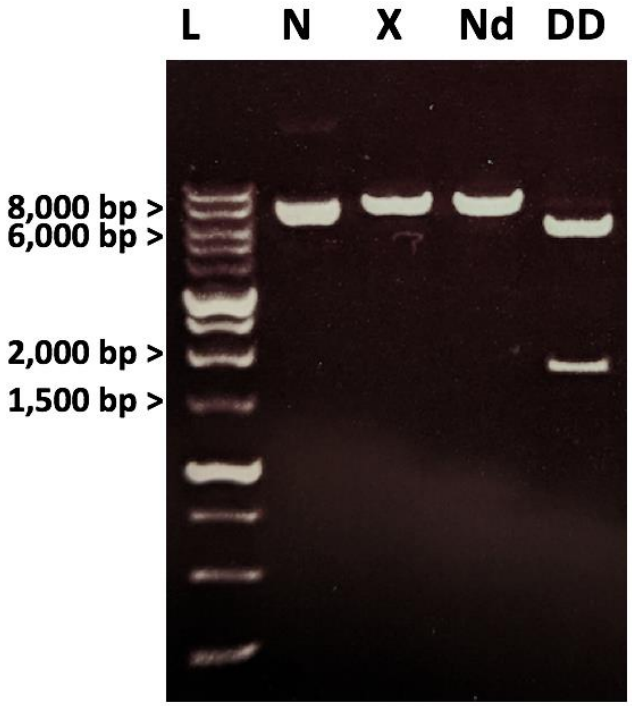


Figure 15: Restriction enzyme digestion of the HTH_pET24a vector. HTH cloned into pET24a+ was isolated from BL21(DE3) competent cells. The banding pattern of the DNA run on a 0.8% agarose gel cast with ethidium bromide was used to determine if transformation was successful. Bands were observed under UV light. The undigested vector (N) will not be linearized. *Xho*I (X) and *Nde*I (Nd) both linearize the vector to create a single 8,962 bp band. Mixing both restriction enzymes with the vector (DD) will create a 5,230 bp band and a 1,732 bp band. All band sizes were determined by comparing to the 1 Kb DNA size ladder (L).

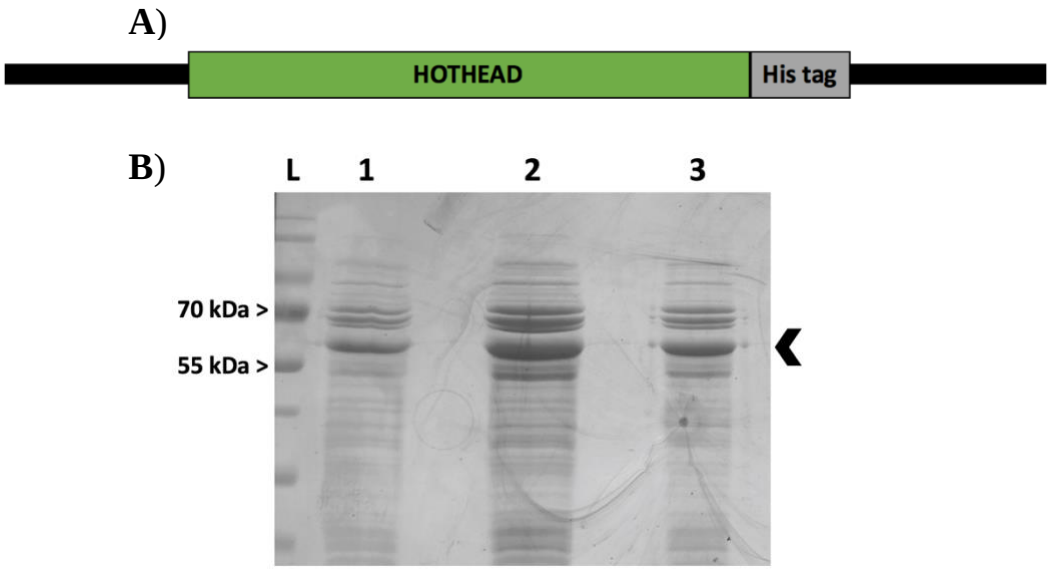


Figure 16: HTH_pET24a protein isolation. (A) Schematic representation of the synthetic HTH cDNA sequence cloned into the pET24a+ vector. The 6 × His tag is located at the C-terminus of the HTH protein. (B) Protein fraction samples collected during nickel chromatography run on a 12% SDS-PAGE gel. The HTH protein has an expected size of 64 kDa (arrowhead). A significant 64 kDa band can be seen eluting off the nickel column (1). A 64 kDa band remains present when the sample is concentrated down to ~10 mL (2) and exchanged into a 25 mM HEPES buffer containing to imidazole (3). Based on the presence of a 64 kDa band, it appears as though the HTH protein was successfully isolated although contaminating bands are also present in the final fraction. All band sizes were determined by comparing to the protein size ladder (L).

3.4.5 Eukaryotic Expression System

Multiple expression systems exist for protein purification. They each have their advantages and disadvantages which should be considered before deciding on one to use for protein isolation. One of the expression systems is *Pichia pastoris*, a yeast species. This eukaryote is commonly used for protein expression due to simplicity of use and cost effectiveness. It also allows for a variety of post-translational modifications, the most significant for this research being N-linked glycosylation (Maeda et al., 2009; Wegner, 1990). HTH is predicted to have N-linked glycosylation sites (Chang, 2016). This post-translational modification is known for being important for protein folding (Helenius et al., 1997). Although

FAD-HNLs also have N-linked glycosylation sites, they are not essential for protein integrity once the protein is folded. It is unclear if HTH will remain active if unglycosylated.

Initial steps have been taken to work with the pPinkNN expression vector within *P. pastoris* (Nava & Amidi, unpublished results). This vector has been designed to add a 6 × His tag to the C-terminus of the protein of interest, which is cleavable by the tobacco etch virus (TEV) site. The α-mating factor secretory signal on the N-terminus of the protein of interest facilitates isolation (Figure 17A) (Cereghino, Cereghino, Ilgen, & Cregg, 2002) by targeting the protein for secretion into the media. *S. cerevisiae* is used to clone the cDNA sequence of the protein of interest into the pPinkNN vector via homologous recombination. The HTH cDNA sequence from *A. thalinana* has been PCR amplified with the appropriate homologous recombination recognition sequences flanking HTH cDNA sequences (Figure 17B). Once HTH has been cloned into the pPinkNN vector, the plasmid can be transformed into competent *P. pastoris* cells.

P. pastoris is methylotropic meaning it is capable of using methanol as a sole carbon source through the oxidation of methanol to formaldehyde (Faber, Harder, Ab, & Veenhuis, 1995). Alcohol oxidase performs this reaction using molecular oxygen. Large amounts of alcohol oxidase are typically translated since it does not have high affinity to oxygen. The alcohol oxidase (*AOX1*) gene is the major producer of this enzyme and its expression is regulated and induced by methanol (Ellis et al., 1985). The PichiaPink expression system takes advantage of this by putting the expression of the protein of interest under the control of the *AOX1* promoter so that the protein is only produced when methanol is added to the media.

A mutant strain of *P. pastoris* denoted as PichiaPink is used for this expression system. These cells have a mutated adenine deaminase 2 (*ADE2*) gene that encodes a carboxylase involved in the synthesis of purine. PichiaPink cells will appear pink due to an accumulation of

purine precursors and will not be able to grow on media lacking adenine (Gedvilaite & Sasnauskas, 1994; Zonneveld & Van Der Zanden, 1995). Cells that successfully maintain the pPinkNN vector will obtain a functional *ADE2* gene, allowing them to grow on media lacking adenine. Initial steps have been taken to work with this system (Figure 17). Primers have been designed to introduce the homologous recombination sequence on either end of the HTH cDNA sequence. As of yet, *HTH* has not been successfully cloned into the pPinkNN vector. Future research will continue to investigate if successful isolation of the active HTH protein is possible using this expression system.

Plants can also be used as a recombinant expression system. They will also allow for post-translational modifications. It takes longer to obtain protein in plants when compared to *E. coli* and yeast. It was considered to collaborate with PlantForm Corporation (Guelph, Ontario) to try and induce HTH in plants. Ultimately, the decision was made to attempt expressing recombinant HTH in yeast instead of plants.

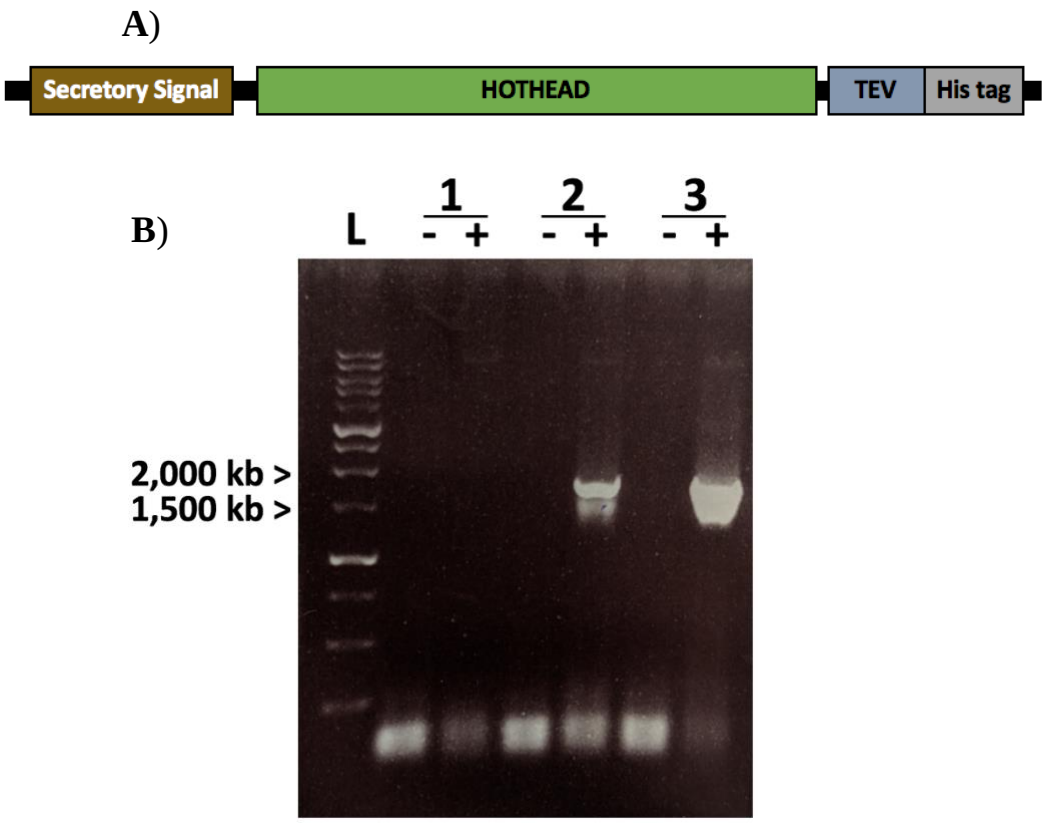


Figure 17: PichiaPink expression system. (A) Schematic representation of the HTH cDNA sequence cloned into the pPinkNN vector. The 6 × His tag and TEV site are located at the C-terminus of the HTH protein. The secretory signal is located at the N-terminus of the HTH protein. (B) PCR amplification product of the HTH cDNA sequence. The cDNA sequence was amplified using primers that added base pairs on either end to allow for homologous recombination with the pPinkNN vector once transformed into *S. cerevisiae*. PCR products were run on a 0.8% agarose gel caster with ethidium bromide. Bands were visualized under UV light. Three different PCR conditions were tested (1-3). Each reaction included a negative control (-) and the PCR product (+). PCR condition 3 yielded the most product. 10% DMSO was added to the PCR mix and used at an annealing temperature set to 52°C.

3.4.6 HTH Mutant Constructs

The *HTH* gene has eleven known mutant alleles created through substitution mutations. Most mutants encode for a different amino acid residue in the peptide sequence with the exception of *hth-1* and *hth-9*. These mutant alleles are predicted to translate into truncated proteins due to the introduction of a stop codon and the alteration of a splice junction sequence,

respectively. These eleven mutation alleles have substitutions located in exons 3, 4 and 5, with the exception of *hth-9* which is located in the second intron (Figure 1).

A prominent phenotype of mutations in the *hth* gene is the fusion between floral organs. This is believed to be due to a loss of cuticle integrity. This phenotype is evident in all *hth* mutants, although to varying degrees, providing genetic evidence for the functional importance of each of these target residues as well as the importance of the HTH protein in cuticle formation. Once a protein isolation protocol has been developed, working with HTH mutants will be an efficient method to compare differences to the WT protein in terms of structure and enzymatic function, and will allow more to be learned about key residues.

Versions of HTH_SUMO were created containing the different *hth* mutations (*hth-1* to *hth-11*), excluding *hth-9*, using the Q5 Site Directed Mutagenesis Kit (New England Biolabs). Once made and transformed into competent DH 5-α competent *E. coli* cells, these mutations were confirmed via restriction endonuclease diagnostic digest (Figure 18), and further validated through DNA sequencing with the aid of custom primers (Appendix H-J). Successfully generated mutant constructs were store at -80°C in glycerol stocks.

Unfortunately, results from protein isolation attempts in *E. coli* indicate that this system is not a suitable expression system, and thus the synthesized HTH cDNA sequence cannot be used in a eukaryotic system because it is optimized for bacterial protein expression. Most of these mutant proteins would not be expected to have fully wild type levels of enzymatic activity. The amount of HTH activity can be correlated to the severity in the floral bud fusion phenotype. For example, the *hth-9* mutant has a predicted truncated protein (Krolikowski et al., 2003) which means that it is likely to be inactive. An *A. thaliana* plant harboring this allele has a very severe floral organ fusion (Krolikowski et al., 2003). On the other hand, an *A. thaliana* plant harboring

the *hth-5* allele has a less severe floral organ fusion (Krolikowski et al., 2003), meaning that the protein still has some activity, although not as much as the WT. The *hth-5* mutant causes Pro564 to be replaced by a serine (Krolikowski et al., 2003). This is only 3 resides away from Asn567, one of the putative active site residues (Chang, 2016). It is interesting that the fusion phenotype for this mutant is weak.

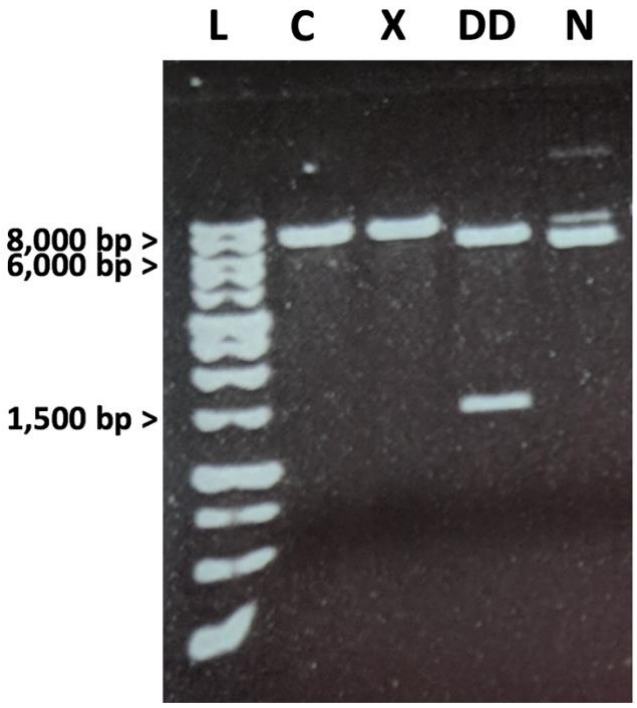


Figure 18: Restriction enzyme digestion of the HTH_SUMO vector. HTH mutant sequences cloned into the pE_SUMOstar vector were isolated from XL1-Blue competent cells. The banding pattern of the DNA run on a 0.8% agarose gel cast with ethidium bromide was used to determine successful mutagenesis. Bands were observed under UV light. *Cla*I (C) and *Xho*I (X) both linearize the vector to create a single 7,356 bp band. Mixing both restriction enzymes with the vector (DD) will create a 5,944 bp band and a 1,412 bp band. The undigested vector (N) will not be linearized. All band sizes were determined by comparing to the 1 Kb DNA size ladder (L).

3.4.7 MDL Enzymatic Assay

P650, P650EV, P103, P103EV and an untransformed *hth-9 A. thaliana* plant line (Chang, 2016) were used to perform floral bud crude protein extractions. Mandelonitrile lyase enzymatic activity assays were performed on these extracts to test the potential enzymatic function of HTH. P650 plant lines appear WT and should have native levels of HTH protein since the *HTH* gene is under the control of its native promoter. HTH mRNA expression levels from this construct have not been measured and it appears to indicate that there may be a slightly higher level of protein (Chang, unpublished results). Chang also verified HTH protein expression by immunoblotting (Chang, 2016) and detected immunoreactive protein species were detected in seedlings, flower buds and siliques (Chang, 2016). P103 plant lines appear WT and should have higher levels of the HTH protein since the *HTH* gene is under the control of the 35S promoter. Sunikumar et al. looked at tissue-specific expression of CaMV35S promoter linked to GFP in cotton (Sunilkumar et al., 2002). This study demonstrated GFP expression in all tissue types, although levels were much lower in floral tissue (Sunilkumar et al., 2002). P650EV, P103EV and untransformed *hth-9* plant lines retain floral fusion phenotype presumably due to a loss of cuticle integrity because these lines lack a fully functional HTH protein.

The *in vitro* mandelonitrile lyase enzymatic assays were used to test extracts obtained from these plant lines. All samples showed a change in absorbance when the substrate and the enzyme sample were mixed together (Appendix K). The specific activity of each sample was determined based on the absorbance reads (Table 3) (Jorns, 1979). It was expected that the P103 plant line would have greater enzymatic activity due to the *HTH* gene being under the control of the 35S promoter, however 35S expression appears to be fairly low in floral tissue (Sunilkumar et al., 2002). The specific activity for all the assays were very low and comparable to one

another. The fact that no differences were observed between transgenic plant lines indicate that HTH is unlikely to be driving cleavage. This assay was performed at pH 5.4 while the crude extraction buffer was at pH 7.5. A previous attempt was made to exchange HTH into a pH 5 solution during the ion exchange chromatography trials. This appeared to cause HTH to precipitate out of solution. This change in pH when HTH is added to the MDL assay solution could have a negative effect on protein activity and create an environment where HTH may not be active.

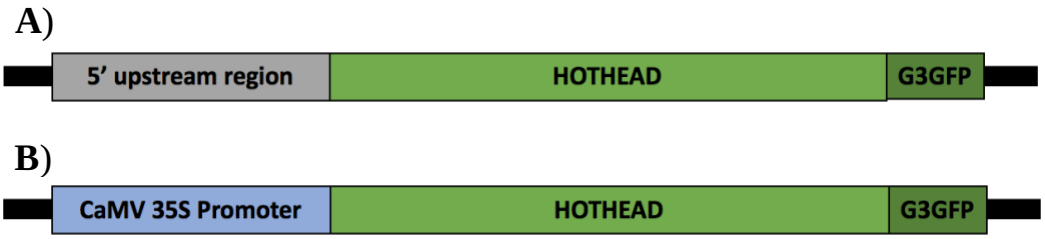


Figure 19: HTHpro:HTH-FP constructs. (A) Schematic representation of the *HTH* gene cloned into the pGWB650 vector. The *HTH* gene was cloned into the vector along with its 5' upstream region, which contains the gene's promoter. A G3 GFP tag is located downstream of the *HTH* gene to place it at the C-terminus of the HTH protein. (B) Schematic representation of the *HTH* gene cloned into the pCAMBIA vector. The *HTH* gene was cloned into the vector along with the CaMV 35S promoter. A G3 GFP tag is located downstream of the *HTH* gene to place it at the C-terminus of the HTH protein.

Since the assay was performed on crude protein extracts, it is difficult to claim that the change in absorbance was due to the presence of HTH. *A. thaliana* does have a hydroxynitrile lyase that is not FAD-dependent (locus: At5g10300) (Andexer et al., 2012). The hydroxynitrile lyase can be found at low levels in various tissues including the sepals of mature flowers (Winter et al., 2007), a tissue present in the plant homogenate used for the crude protein extraction. When referring to the chirality center of mandelonitrile, the non-FAD-dependent enzyme present in floral tissue is (S)-selective, meaning that it exclusively uses substrates which are (S)-

stereoisomer. On the other hand, FAD-dependent hydroxynitrile lyases are typically (R)-selective, meaning it exclusively uses substrates which are (R)-stereoisomer (Andexer et al., 2012). It is possible that this non-FAD dependent hydroxynitrile lyase is present in solution. Since a racemic mixture of mandelonitrile was used as the substrate for the enzymatic assay, it is not possible to say what drove the change in absorbance during the MDL assay.

Table 3: Spectrophotometric mandelonitrile lyase activity assay. Final crude protein fractions were directly mixed with mandelonitrile and continuously monitored for the production of benzaldehyde to determine enzymatic activity at an absorbance 275 nm. The Δ275/min was determined by absorbance detected over a period of 15 minutes. Values are the mean. Experiments were repeated three times.

	MDL	**P650**	**P650EV**	**P103**	**P103EV**	**hth9**
ΔA_{275}/min	0.0104	0.0083	0.0085	0.0041	0.0079	0.0100
Specific Activity (U/min/mg)	6930 ±634.8773	0.0061 ±0.0003	0.0026 ±0.0002	0.0037 ±0.0007	0.0025 ±0.0006	0.0076 ±0.0005

To detect the presence of HTH, the initial homogenate and final pellet fractions from the P650 plant line crude extracts were tested for the presence of the green fluorescent protein (GFP) tag via Western blot (Figure 20). These fractions were chosen to see if HTH is present when the plant tissue is homogenized and at the end of the extraction process. The Western blot showed no detectable recombinant HTH::GFP in any of the extracts. It is therefore possible that insufficient HTH protein may have been isolated since it could not be detected in the extracts. This is likely due to the less harsh extraction method used to ensure HTH was not denatured. Triton-X, a mild detergent, was used to ensure that the proteins were not denatured. A stronger detergent could be considered but it must not cause denaturation. Since HTH has been shown to localize in the ER membrane, it proves to be challenging to isolate the protein from a homogenized crude extract.

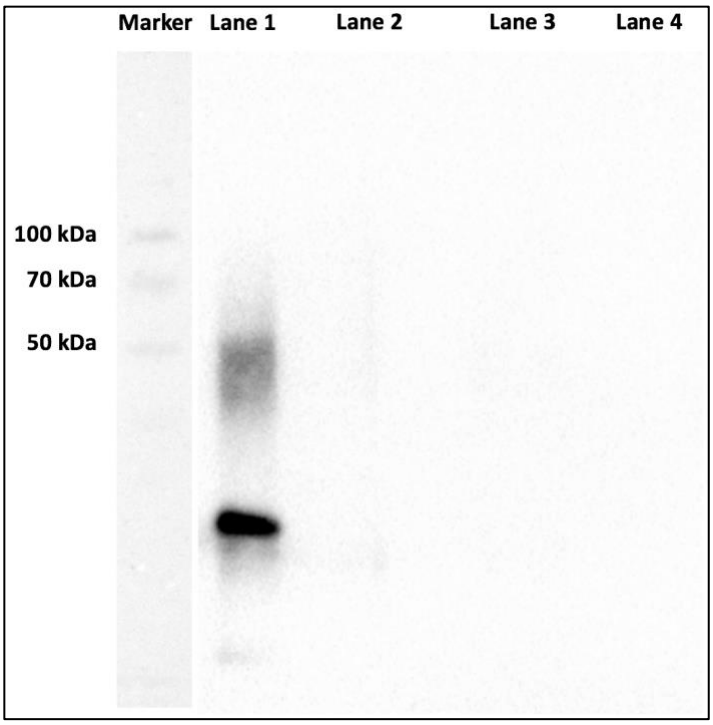


Figure 20: Western blot analysis of HTH protein using crude protein extract from HTHpro:HTH-G3GFP cloned into the pGWB650 vector. Probed with anti-GFP antibody, isolated GFP (lane 1) was detected at ~30 kDa. No band corresponding to HTH::GFP was detected in the P650 crude extract homogenate (lane 2) and the P650 final fraction (lane 3). A negative control lacking the HTH construct (lane 4) also did not detect HTH::GFP.

3.5 Concluding remarks

Attempting to isolate the HTH protein has proven to be challenging. Even though protein induction was successful in *E. coli,* isolation attempts indicated that HTH was not active even when stabilized by the SUMO fusion protein. Mass spectrometry analysis did confirm that HTH was present in the final fraction after nickel chromatography, although other chaperone proteins appear to be co-purifying. This provides insight into why the isolated protein was inactive, namely that recombinant HTH, while soluble, was likely misfolded. The current hypothesis is that HTH recombinantly expressed in *E. coli* may lack post-translational modifications that are essential to HTH folding and enzymatic activity. Initial steps have been taken to test this hypothesis, through work in a yeast expression system that may address this final hurdle and allow for successful isolation of recombinant, active HTH. If this hypothesis proves true, available *hth* mutant constructs will not be useful as they were generated in the synthetic cDNA sequence optimized for bacterial expression.

A change in absorbance was observed in the crude protein extracts from *A. thaliana* floral tissue when tested for mandelonitrile lyase enzymatic activity, although it was very low. Considering that *A. thaliana* has an endogenous hydroxynitrile lyase expressed in mature floral tissue, it is not possible to say which protein caused the change in absorbance, if any. Based on the Western blot analysis, no detectable amount of HTH::GFP was observed indicating that HTH was not successfully extracted from the floral tissue, or it was extracted at a very low amount. It is unclear what enzyme drove MDL activity but it may not be due to HTH::GFP. A new extraction protocol may need to be considered to successfully isolate HTH and test for enzymatic activity, such as the removal or chaperone proteins recently discussed by Morales et al. (2019).

Overall, this research will contribute to the development of a new HTH protein isolation protocol. Using the limited existing research of the HTH protein, this work developed and tested hypotheses to investigate the possible factors that contribute to the unsuccessful isolation. While the protocol is still being developed, this research has reduced the list of potential factors that may hinder protein isolation and has informed possible directions for future work.

Appendices

Appendix A: HTH_pESUMO vector sequence. Nucleotide sequence coding for the synthesized HTH cDNA cloned into pE_SUMOstar_Kan provided by GenScript (GenScript Biotech Corporation; Piscataway, New Jersey, United States). The nucleotides that code for the pE_SUMOstar_Kan vector are in black. The nucleotides that code for the HTH peptide sequence are in blue.

AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTA
ACTTTAAGAAGGAGATATACCATGGGTCATCACCATCATCATCACGGGTCCCTGCAGGACTCAGAAGTCAATCAAGAAGCTAAGC
CAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCAAGATCAA
AAAGACCACTCCTTTAAGAAGGCTGATGGAAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAACGTTCTTGTACGAC
GGTATTGAAATTCAAGCTGATCAGACCCCTGAAGATTTGGACATGGAGGATAACGATATTATTGAGGCTCACCGCGAACAGATTG
GAGGTTGAGACCACTAGTGGTACCGGTCTCAGCACCGCGAGCAAGGGCAAAGAAAAGAAAAGCAAGTTCAACCCGTATCGTTAC
ACCTTCATCGACAAAGCGAGCACCTTCAGCAGCAGCAGCAGCAGCAGCTTCAGCAGCAACGGTCAGGACAGCAGCTACGATTAT
ATCGTGATTGGTGGCGGTACCGCGGGTTGCCCGCTGGCGGCGACCCTGAGCCAAAACTTCAGCGTGCTGGTTCTGGAGCGTGGCG
GTGTTCCGTTTACCAACGCGAACGTGAGCTTCCTGCGTAACTTTCACATCGGTCTGGCGGACATTAGCGCGAGCAGCGCGAGCCA
GGCGTTTGTGAGCACCGATGGTGTTTACAACGCGCGTGCGCGTGTTCTGGGCGGTGGCAGCTGCATTAACGCGGGTTTCTATAGCC
GTGCGGACGCGGCGTTTGTTAAGCGTGCGGGCTGGGATCCGAAGCTGGTGAAAGAGAGCTACCCGTGGGTGGAGCGTGAAATCG
TTCACCAGCCGAAGCTGACCCTGTGGCAAAAAGCGCTGCGTGACAGCCTGCTGGAAGTGGGTGTTCGTCCGTTCAACGGCTTTAC
CTATGACCACGTGAGCGGTACCAAAATCGGTGGCACCATCTTCGATCGTTTTGGTCGTCGTCATACCGCGGCGGAACTGCTGGCGT
ATGCGAACCCGCAGAAGCTGCGTGTTCTGATTTATGCGACCGTGCAAAAAATCGTTTTCGACACCAGCGGTACCCGTCCGCGTGT
GACCGGCGTTATTTTTAAGGATGAGAAAGGTAACCAGCACCAAGCGCTGCTGAGCAACCGTAAGGGCAGCGAAGTGATTCTGAG
CAGCGGTGCGATTGGCAGCCCGCAGATGCTGATGCTGAGCGGTATTGGCCCGAAGAAAGAGCTGCAACGTCTGAAGATCCCGGT
GGTTCTGGAGAACGAACACGTTGGTAAAGGCATGGCGGATAACCCGATGAACACCATCCTGGTGCCGAGCAAGGCGCCGATCGA
ACAGAGCCTGATTCAAACCGTTGGTATCACCAAAATGGGCGTGTACGTTGAAGCGAGCACCGGTTTTGGTCAGAGCCCGGAAAGC
ATTCACACCCACTATGGTATCATGAGCAACAAGAACGAGCTGTTTAGCACCATTCCGGCGAAACAGCGTCGTCCGGAAGCGACCC
AAGCGTACATCACCCGTAACAAGTATCAACTGCACGAGGCGTTCAACGGTAGCTTTATTCTGGAAAAACTGGCGTACCCGATCAG
CCGTGGCCACCTGAGCCTGGTGAACACCAACGTTGACGATAACCCGAGCGTTACCTTCAACTATTTTAAGCACCCGGTGGACCTG
CAGCGTTGCGTTGAGGCGATCCGTCTGGTGAGCAAAGTGGTTACCAGCAACCGTTTCCTGAACTACACCCAGTGCGATAAGCAAA
ACGTTCACAAAATGCTGAGCCTGAGCGTGAAGGCGAACATTAACCTGCGTCCGAAGCAGCTGAACGACACCAAAAGCATGGCGC
AATTTTGCAAAGATACCGTGGTTACCATCTGGCACTACCACGGTGGCTGCCTGGTGGGCAAGGTGGTTAGCCCGAACCGTAAAGT
GCTGGGCGTTGACCGTCTGCGTGTTATCGACGGTAGCACCTTCGATGAGAGCCCGGGTACCAACCCGCAGGCGACCATGATGATG
ATGGGTCGTTATATGGGCGTGAAAATCCTGCGTGAGCGTCTGGGCAACAAGGCGGGTGTGTAACTCGAGCACCACCACCACCACC
ACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGG
GGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCG
CATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTC
CCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGG
CACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTT
GGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGG
GATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTA
CAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCA
TGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAA
AAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACT
CGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAAT
CCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCA
TCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAA
TCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCT
GTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATT
CCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCA
TCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATC
CATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAG
CAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGG
ATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGA
TCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGT
TAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACAC
AGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGA
GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTAT
CTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAA
CGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTG
GATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAA
GCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATC
TGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACA

CCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCA
GAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCT
GCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGT
TTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGC
TCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGA
CCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCG
ATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATG
TTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGG
CAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCT
GCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACA
GGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCAT
GATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAA
GGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAA
ACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTT
CACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCC
AGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGA
TATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGT
GGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCT
GAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCG
CGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGG
TGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA
TAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCAT
CGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTG
GAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCG
CCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGC
ATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGC
GAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAG
GTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGC
CACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCG
CCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCG

Appendix B: HTH_pET24a+ vector sequence. Nucleotide sequence coding for the synthesized HTH cDNA cloned into pET24a+ provided by GenScript (GenScript Biotech Corporation; Piscataway, New Jersey, United States). The nucleotides that code for the pET24a+ vector are in black. The nucleotides that code for the HTH peptide sequence are in blue.

ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTATTGCTCAGCG
GTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGCACACCCGC
CTTGTTGCCCAGACGCTCACGCAGGATTTTCACGCCCATATAACGACCCATCATCATCATGGTCGCCTGCGGGTTGGTACCCGGGC
TCTCATCGAAGGTGCTACCGTCGATAACACGCAGACGGTCAACGCCCAGCACTTTACGGTTCGGGCTAACCACCTTGCCCACCAG
GCAGCCACCGTGGTAGTGCCAGATGGTAACCACGGTATCTTTGCAAAATTGCGCCATGCTTTTGGTGTCGTTCAGCTGCTTCGGAC
GCAGGTTAATGTTCGCCTTCACGCTCAGGCTCAGCATTTTGTGAACGTTTTGCTTATCGCACTGGGTGTAGTTCAGGAAACGGTTG
CTGGTAACCACTTTGCTCACCAGACGGATCGCCTCAACGCAACGCTGCAGGTCCACCGGGTGCTTAAAATAGTTGAAGGTAACGC
TCGGGTTATCGTCAACGTTGGTGTTCACCAGGCTCAGGTGGCCACGGCTGATCGGGTACGCCAGTTTTTCCAGAATAAAGCTACCG
TTGAACGCCTCGTGCAGTTGATACTTGTTACGGGTGATGTACGCTTGGGTCGCTTCCGGACGACGCTGTTTCGCCGGAATGGTGCT
AAACAGCTCGTTCTTGTTGCTCATGATACCATAGTGGGTGTGAATGCTTTCCGGGCTCTGACCAAAACCGGTGCTCGCTTCAACGT
ACACGCCCATTTTGGTGATACCAACGGTTTGAATCAGGCTCTGTTCGATCGGCGCCTTGCTCGGCACCAGGATGGTGTTCATCGGG
TTATCCGCCATGCCTTTACCAACGTGTTCGTTCTCCAGAACCACCGGGATCTTCAGACGTTGCAGCTCTTTCTTCGGGCCAATACCG
CTCAGCATCAGCATCTGCGGGCTGCCAATCGCACCGCTGCTCAGAATCACTTCGCTGCCCTTACGGTTGCTCAGCAGCGCTTGGTG
CTGGTTACCTTTCTCATCCTTAAAAATAACGCCGGTCACACGCGGACGGGTACCGCTGGTGTCGAAAACGATTTTTTGCACGGTCG
CATAAATCAGAACACGCAGCTTCTGCGGGTTCGCATACGCCAGCAGTTCCGCCGCGGTATGACGACGACCAAAACGATCGAAGAT
GGTGCCACCGATTTTGGTACCGCTCACGTGGTCATAGGTAAAGCCGTTGAACGGACGAACACCCACTTCCAGCAGGCTGTCACGC
AGCGCTTTTTGCCACAGGGTCAGCTTCGGCTGGTGAACGATTTCACGCTCCACCCACGGGTAGCTCTCTTTCACCAGCTTCGGATC
CCAGCCCGCACGCTTAACAAACGCCGCGTCCGCACGGCTATAGAAACCCGCGTTAATGCAGCTGCCACCGCCCAGAACACGCGCA
CGCGCGTTGTAAACACCATCGGTGCTCACAAACGCCTGGCTCGCGCTGCTCGCGCTAATGTCCGCCAGACCGATGTGAAAGTTAC
GCAGGAAGCTCACGTTCGCGTTGGTAAACGGAACACCGCCACGCTCCAGAACCAGCACGCTGAAGTTTTGGCTCAGGGTCGCCGC
CAGCGGGCAACCCGCGGTACCGCCACCAATCACGATATAATCGTAGCTGCTGTCCTGACCGTTGCTGCTGAAGCTGCTGCTGCTG
CTGCTGCTGAAGGTGCTCGCTTTGTCGATGAAGGTGTAACGATACGGGTTGAACTTGCTTTTCTTTTCTTTGCCCTTGCTCGCGGTG
CTCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTTATCCGCTCACAATTCCCCTATAGTGAGTC
GTATTAATTTCGCGGGATCGAGATCTCGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCT
GGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGG
TGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAA
CCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGAGATCCCGGACACCATCGAATGGCGCAAAACCTT
TCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGA
GTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAA
GCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCA
CCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGT
GTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATT
AACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCA
GACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAA
ATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAAT
TCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTT
CCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGG
ATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTG
CTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGG
TGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACA
GGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACCGGGATCTCGACCGAT
GCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTA
TCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGG
CCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGC
AGGCCATTATCGCCGGCATGGCGGCCCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTG
CCTTACTGGTTAGCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACA
ACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCAT
CGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTCTGGT
CCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGC
ATCCTCTCTCGTTTCATCGGTATCATTACCCCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAA
CCGCCCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGATGAACAGGC
AGACATCTGTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCT
GACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGG
GTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAG
CAGATTGTACTGAGAGTGCACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTC
CGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTAT
CCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTG
CTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC
TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCC
TTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGC
TGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTT
ATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCT

AACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGA
AGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAACAATAAAACTG
TCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTCTAGGCCGCGATTAAATTCCA
ACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAA
GCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGG
CTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGG
GAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTG
CATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTG
GTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTC
ACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTG
GACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGG
CTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAAGAATTAATT
CATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAA
ATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAA
ATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGG
ACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTC
GAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAG
AAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCG
CCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCA

Appendix C: Peptide sequence alignment between HOTHEAD and alcohol dehydrogenase from Candida albicans. The sequence alignment was performed using Clustal Omega. The Asterix (*) indicates position with a conserved residue. The colon (:) indicates the position of two residues with strongly similar properties. The period (.) indicates the position of two residues with weakly similar properties.

```
CLUSTAL O(1.2.4) multiple sequence alignment


sp|Q9S746|HTH_ARATH       MALKLFLFALLLCLPTSLSSTASKGKEKKSKFNPYRYTFIDKASTFSSSSSSSFSSNGQD  60
sp|P43067|ADH1_CANAX      -----------------MSEQI-----------------------PKTQKAVVFDTNGGQ  20
                                          :*.                        .:..:  *.:** :

sp|Q9S746|HTH_ARATH       SSYDYIVIGGGTAGCPLAATLSQNFSVLVLERGGVPFTNANVSFLRNFHIGLA--DISAS  118
sp|P43067|ADH1_CANAX      LVYKD---------------------------YPVPTPKPNELLIHVKYSGVCHTDLHAR  53
                            *.                            **  : *  :::  : *:.  *: *

sp|Q9S746|HTH_ARATH       SASQAFV---STDGVYNA-------------------RARVLGGGSCINAGFYSRADAAF  156
sp|P43067|ADH1_CANAX      KGDWPLATKLPLVGGHEGAGVVVGMGENVKGWKIGDFAGIKWLNGSCMSCEFCQQGAEPN  113
                          ...  :.      * ::.                    .    .***:.. * .:.

sp|Q9S746|HTH_ARATH       VKR---AGWDPKLVKESYPWVEREIVHQPKLTLWQ-KALRDSLLEVGVRPFNGFTYD---  209
sp|P43067|ADH1_CANAX      CGEADLSGYTHDGSFEQYAT--ADAVQAAKIPAGTDLANVAPILCAGVTVYKALKTADLA  171
                            .   :*:  .   *.*     : *:  *:      *    :* .**  ::.:.

sp|Q9S746|HTH_ARATH       -----HVSGTKIGGTIFDRFGRRHTAAELLAYANPQKLRVLIYATVQ-KIVF-DTSGTR-  261
sp|P43067|ADH1_CANAX      AGQWVAISGA--GGGL---------GSLAVQYARAMGLRVVAIDGGDEKGEFVKSLGAEA  220
                                :**:  ** :         .:  : **.   ***:     : *  * .: *:.

sp|Q9S746|HTH_ARATH       -------PRVTGVIFKDEKGNQHQALLSNRKGSEVILSSGAIGSPQMLMLSGIGPKKELQ  314
sp|P43067|ADH1_CANAX      YVDFTKDKDIVEAVKKATDGGPHGAINVSVSEKAIDQSVEYVRPLGKVVLVGLPAHAKV-  279
                                   :. .: *  .*. * *:  . . . :  *   :     ::* *:  : ::

sp|Q9S746|HTH_ARATH       RLKIPVVLENEHVGKGMADNPMNTILVPSKAPIEQSLIQTVGITKMGVYVEASTGFGQSP  374
sp|P43067|ADH1_CANAX      ----------------------------TAPVFDAVVKSI--EIKGSYVGNRKD-----  303
                                                .**: :::::::     * **   ..

sp|Q9S746|HTH_ARATH       ESIHTHYGIMSNKNELFSTIPAKQRRPEATQAYITRNKYQLHEAFNGSFILEKLAYPISR  434
sp|P43067|ADH1_CANAX      ------------------------------------------------------------  303


sp|Q9S746|HTH_ARATH       GHLSLVNTNVDDNPSVTFNYFKHPVDLQRCVEAIRLVSKVVTSNRFLNYTQCDKQNVHKM  494
sp|P43067|ADH1_CANAX      -------------TAEAIDFFSR--GLIKCPIKIVGLSDLP--------------EVFK-  333
                                     : ::::*.:  .* :*   *  :*.:               :*.*

sp|Q9S746|HTH_ARATH       LSLSVKANINLRPKQLNDTKSMAQFCKDTVVTIWHYHGGCLVGKVVSPNRKVLGVDRLRV  554
sp|P43067|ADH1_CANAX      --------------L--------------------MEEGKILGRYVLDTSK---------  350
                                                  . * ::*: *  . *

sp|Q9S746|HTH_ARATH       IDGSTFDESPGTNPQATMMMMGRYMGVKILRERLGNKAGV      594
sp|P43067|ADH1_CANAX      ---------------------------------------      350
```

Appendix D: Peptide sequence alignment between HOTHEAD and (R)-hydroxynitrile lyase from Prunus mume. The sequence alignment was performed using Clustal Omega. The Asterix (*) indicates position with a conserved residue. The colon (:) indicates the position of two residues with strongly similar properties. The period (.) indicates the position of two residues with weakly similar properties.

```
CLUSTAL O(1.2.4) multiple sequence alignment

sp|Q9S746|HTH_ARATH         ----------MALKLFLFALLLCLPTSLSSTASKGKEKKSKFNPYRYTFIDKASTFSSSS        50
tr|W8VKF9|W8VKF9_PRUMU      MEKSTMSAILLVLHLFVLHLQYSEVHSLAT------TSDHDFSYLSFAY-----------        43
                                      :.*:**:: *  .   **::       .. .*.   :::

sp|Q9S746|HTH_ARATH         SSSFSSNGQDSSYDYIVIGGGTAGCPLAATLSQNFSVLVLERGGVPFTNANVSFLRNFHI        110
tr|W8VKF9|W8VKF9_PRUMU      --DATDLELEGSYDYVIVGGGTSGCPLAATLSEKYKVLVLERGSLPTAYPNVLTADGFVY        101
                              . :.   :.****:::****:*********:::.*******.:* :  **    .*

sp|Q9S746|HTH_ARATH         GLADI-SASSASQAFVSTDGVYNARARVLGGGSCINAGFYSRADAAFVK--RAGWDPKLV        167
tr|W8VKF9|W8VKF9_PRUMU      NLQQEDDGKTPVERFVSEDGIDNVRGRVLGGTSMINAGVYARANTSIYSASGVDWDMDLV        161
                            .* :  ...:  : *** **: *.*.***** * ****.*:**:::: .   ..** .**

sp|Q9S746|HTH_ARATH         KESYPWVEREIVHQPKLTLWQKALRDSLLEVGVRPFNGFTYDHVSGTKIGGTIFDRFGRR        227
tr|W8VKF9|W8VKF9_PRUMU      NQTYEWVEDTIVYKPNSQSWQSVTKTAFLEAGVHPNHGFSLDHEAGTRITGSTFDNKGTR        221
                            :::* ***  **::*:   **.. : ::**.**:* :**: ** :**:* *: **. * *

sp|Q9S746|HTH_ARATH         HTAAELLAYANPQKLRVLIYATVQKIVFDTSGTRPRVTGVIFKDEKGNQHQALLSNRKGS        287
tr|W8VKF9|W8VKF9_PRUMU      HAADELLNKGNSNNLRVGVHASVEKIIFSNAA-GLTATGVIYRDSNGTPHQAFVR--SKG        278
                            *:* ***  .* ::*** ::*:*:**:*..:.    .****::*.:*. ***::   . .

sp|Q9S746|HTH_ARATH         EVILSSGAIGSPQMLMLSGIGPKKELQRLKIPVVLENEHVGKGMADNPMNTILVPSKAPI        347
tr|W8VKF9|W8VKF9_PRUMU      EVIVSAGTIGTPQLLLLSGVGPRVYLSSLNIPVVLSHPYVGQFLHDNPRNFINILPPNPI        338
                            ***:*:*:**:**:*:***:**:  *. *:*****.: :**: : *** * * :    **

sp|Q9S746|HTH_ARATH         EQSLIQTVGITKMGVYVEASTGFGQSPESIHTHYGIMSNKNELFSTIPAKQRRPEATQAY        407
tr|W8VKF9|W8VKF9_PRUMU      EPTIVTVLGISNDFYQCS----FSS-----------LPFTTPPF--------------GF        369
                            * ::: .:**::     .    *..           :  ..  *              .:

sp|Q9S746|HTH_ARATH         ITRNKYQLHEAFNGSFILEKLAYPISRGHLSLVNTNVDDNPSVTFNYFKHPVDLQRCVEA        467
tr|W8VKF9|W8VKF9_PRUMU      FPSTSYPLPNSTF-AHFANKVAGPLSYGSLTLTSSNVRVSPNVKFNYYSNPTDLSHCVSG        428
                            :  ..* * ::   :.: :*:* *:* * *:*..:**  .*.*.***:.:*.**.:**..

sp|Q9S746|HTH_ARATH         IRLVSKVVTSNRFLNYTQCDKQNVHKMLSLSVKANINLRPKQLNDTKSMAQFCKDTVVTI        527
tr|W8VKF9|W8VKF9_PRUMU      MKKIGELLSTDALKPYKVEDLPGIEGFNILGI-----PLPKDQTDDAAFETFCRESVASY        483
                            :: :.:::::: :  *.  *  .:. :  *.:       **: .*  ::  **:::*.:

sp|Q9S746|HTH_ARATH         WHYHGGCLVGKVVSPNRKVLGVDRLRVIDGSTFDESPGTNPQATMMMMGRYMGVKILRER        587
tr|W8VKF9|W8VKF9_PRUMU      WHYHGGCLVGKVLDGDFRVTGINALRVVDGSTFPYTPASHPQGFYLMLGRYVGIKILQER        543
                            ************:. : :* *:: ***:*****  :*.::**.  :*:***:*:***:**

sp|Q9S746|HTH_ARATH         LGNKAGV------------ 594
tr|W8VKF9|W8VKF9_PRUMU      SASDLKILDSLKSAASFVF 562
                             ...  :
```

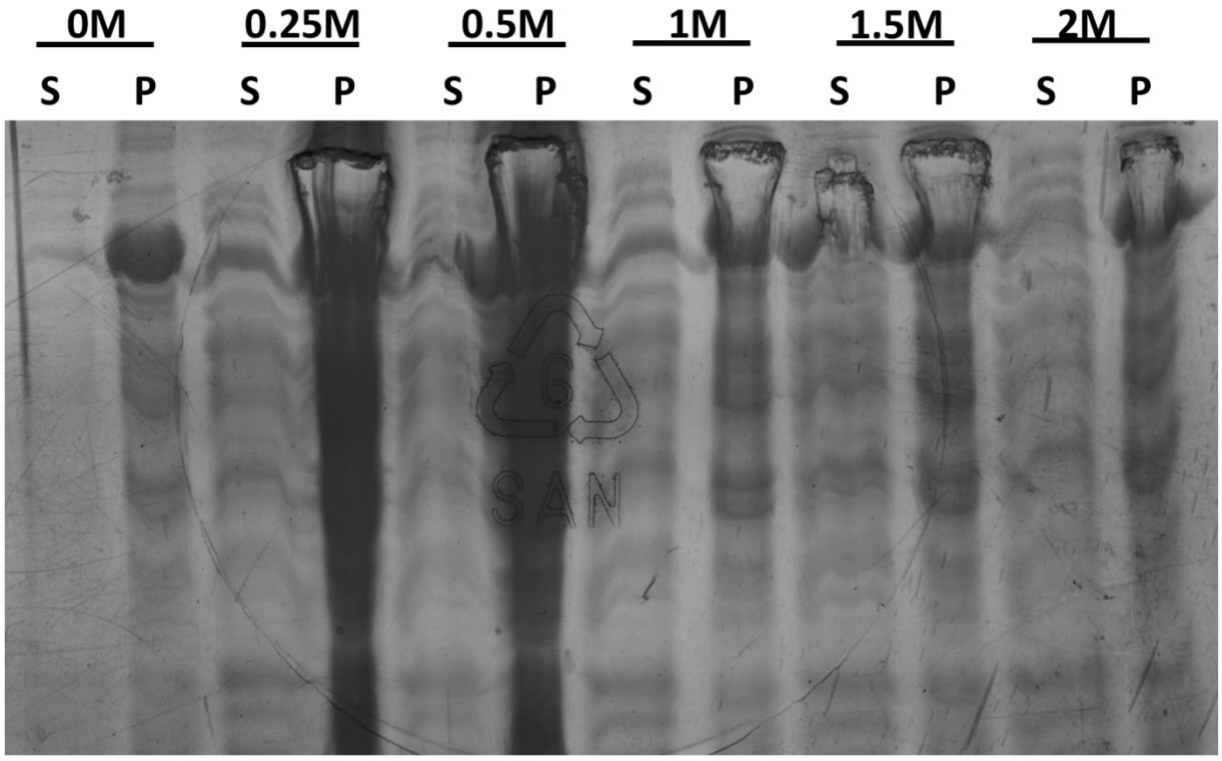
Appendix E: Peptide sequence alignment between HOTHEAD and choline oxidase from A. globiformis. The sequence alignment was performed using Clustal Omega. The Asterix (*) indicates position with a conserved residue. The colon (:) indicates the position of two residues with strongly similar properties. The period (.) indicates the position of two residues with weakly similar properties.

```
CLUSTAL O(1.2.4) multiple sequence alignment


sp|Q9S746|HTH_ARATH       MALKLFLFALLLCLPTSLSSTASKGKEKKSKFNPYRYTFIDKASTFSSSSSSSFSSNGQD  60
sp|Q7X2H8|CHOX_ARTGO      -------------------------------------------------MHIDNIENLSD  11
                                                                     .  .* .*

sp|Q9S746|HTH_ARATH       SSYDYIVIGGGTAGCPLAATLSQNFSVLVL--ERG----GVPFTNANVSFLRNFHIGLA-  113
sp|Q7X2H8|CHOX_ARTGO      REFDYIVVGGGSAGAAVAARLSEDPAVSVALVEAGPDDRGVPEVLQLDRWMELLESGYDW  71
                           .:****:***:**. :** **:: :* *   * *    *** .     ::. :. *

sp|Q9S746|HTH_ARATH       --DISASSASQAFVSTDGVYNARARVLGGGSCINAG--FYSRADA-----AFVKRAGWDP  164
sp|Q7X2H8|CHOX_ARTGO      DYPIEPQENGN-----SFMRHARAKVMGGCSSHNSCIAFWAPREDLDEWEAKYGATGWNA  126
                             *. .. .:     . : :***:*:** *. *:   *::  :      *    :**:

sp|Q9S746|HTH_ARATH       KLVKESYPWVEREIVHQPKLTLWQKALRDSLLEVGVRPFNGFTY-DH------------V  211
sp|Q7X2H8|CHOX_ARTGO      EAAWPLYKRLETNEDAGPDAPHHGDSGPVHLMNVPPKDPTGVALLDACEQAGIPRAKFNT  186
                          : .   *  :* :    *.     .:    *::*  :  .*.:  *             .

sp|Q9S746|HTH_ARATH       SGTKIGGTIFDRFGRRHTA---AELLAYANP----QKLRVLIYATVQKIVFDTSGTRPRV  264
sp|Q7X2H8|CHOX_ARTGO      GTTVVNGANFFQINRRADGTRSSSSVSYIHPIVEQENFTLLTGLRARQLVFDADR---RC  243
                          . * :.*: * ::.**  .   :. ::* :*    ::: :*    .:::***:.    *

sp|Q9S746|HTH_ARATH       TGVIFKDE-KGNQHQALLSNRKGSEVILSSGAIGSPQMLMLSGIGPKKELQRLKIPVVLE  323
sp|Q7X2H8|CHOX_ARTGO      TGVDIVDSAFGHTH----RLTARNEVVLSTGAIDTPKLLMLSGIGPAAHLAEHGIEVLVD  299
                          *** : *.  *: *         .**:**:***.:*::********  .* .  * *:::

sp|Q9S746|HTH_ARATH       NEHVGKGMADNPMNTILVPSKAPIEQSLIQTVGITKMGVYVEASTGFGQSPESIHTHYGI  383
sp|Q7X2H8|CHOX_ARTGO      SPGVGEHLQDHPEGVVQFEAKQPMVAESTQ---WWEIGIFTPTEDGLDR--PDLMMHYGS  354
                          .  **: : *:* ..: . :* *:  .  *     ::*::. :. *:.:   .:  ***

sp|Q9S746|HTH_ARATH       MSNKNELFSTIPAKQRRPEATQAYITRNKYQLHEAFNGSFILEKLAYPISRGHLSLVNTN  443
sp|Q7X2H8|CHOX_ARTGO      VPFDMNTL-----RHGYP---------------TTENGFSLTPNVTHARSRGTVRLRSRD  394
                          :  . : :     ::  *               : **  :  ::::  *** : * . :

sp|Q9S746|HTH_ARATH       VDDNPSVTFNYFKHPVDLQRC--VEAIRLVSKVVTSNRFLNYTQCDKQNVHKMLSLSVKA  501
sp|Q7X2H8|CHOX_ARTGO      FRDKPMVDPRYFTDPEGHDMRVMVAGIRKAREIAAQPAMAEWTGR-------ELSPGVE-  446
                          . *:* *  .**..* . :    * .** . ::.:.  : ::*          ** .*:

sp|Q9S746|HTH_ARATH       NINLRPKQLNDTKSMAQFCKDTVVTIWHYHGGCLVGKV------VSPNRKVLGVDRLRVI  555
sp|Q7X2H8|CHOX_ARTGO      --------AQTDEELQDYIRKTHNTVYHPVGTVRMGAVEDEMSPLDPELRVKGVTGLRVA  498
                                 :  :.: :: :.*  *::*  *   :* *      :.*: :* **  ***

sp|Q9S746|HTH_ARATH       DGSTFDESPGTNPQATMMMMGRYMGVKILRERLGNKAGV---------      594
sp|Q7X2H8|CHOX_ARTGO      DASVMPEHVTVNPNITVMMIGERCADLIRSARAGETTTADAELSAALA      546
                          *.*.: *   .**: *:**:*.  .  *   * *:.: .
```

Appendix F: Peptide sequence alignment between HOTHEAD and mandelonitrile lyase from P. dulcis. The sequence alignment was performed using Clustal Omega. The Asterix (*) indicates position with a conserved residue. The colon (:) indicates the position of two residues with strongly similar properties. The period (.) indicates the position of two residues with weakly similar properties.

```
CLUSTAL O(1.2.4) multiple sequence alignment


sp|Q9S746|HTH_ARATH       --------MALKLFLFALLLCLPTSLSSTASKGKEKKSKFNPYRYTFIDKASTFSSSSSS  52
sp|Q945K2|MDL2_PRUDU      MEKSTMSAILLVLYIFVLHLQYSEV----HSLATTSDHDFSYLSFAY-------------  43
                                  : * *::*.* *          * .. .. .*.   :::

sp|Q9S746|HTH_ARATH       SFSSNGQDSSYDYIVIGGGTAGCPLAATLSQNFSVLVLERGGVPFTNANVSFLRNFHIGL  112
sp|Q945K2|MDL2_PRUDU      DATDLELEGSYDYVIVGGGTSGCPLAATLSEKYKVLVLERGSLPTAYPNVLTADGFVYNL  103
                          . :.   :.****:::****:*********:::.*******.:* :  **    .*  .*

sp|Q9S746|HTH_ARATH       ADI-SASSASQAFVSTDGVYNARARVLGGGSCINAGFYSRADAAFVK--RAGWDPKLVKE  169
sp|Q945K2|MDL2_PRUDU      QQEDDGKTPVERFVSEDGIDNVRGRVLGGTSIINAGVYARANTSIYSASGVDWDMDLVNQ  163
                           :  ...:  : *** **: *.*.***** * ****.*:**:::: .   ..** .**::

sp|Q9S746|HTH_ARATH       SYPWVEREIVHQPKLTLWQKALRDSLLEVGVRPFNGFTYDHVSGTKIGGTIFDRFGRRHT  229
sp|Q945K2|MDL2_PRUDU      TYEWVEDTIVYKPNSQSWQSVTKTAFLEAGVHPNHGFSLDHEEGTRITGSTFDNKGTRHA  223
                          :* ***  **::*:   **.. : ::**.**:* :**: ** .**:* *: **. * **:

sp|Q9S746|HTH_ARATH       AAELLAYANPQKLRVLIYATVQKIVFDTSGTRPRVTGVIFKDEKGNQHQALLSNRKGSEV  289
sp|Q945K2|MDL2_PRUDU      ADELLNKGNSNNLRVGVHASVEKIIFSNAP-GLTATGVIYRDSNGTPHQAFVR--SKGEV  280
                          * ***  .* ::*** ::*:*:**:*..:     .****::*.:*. ***::   . .**

sp|Q9S746|HTH_ARATH       ILSSGAIGSPQMLMLSGIGPKKELQRLKIPVVLENEHVGKGMADNPMNTILVPSKAPIEQ  349
sp|Q945K2|MDL2_PRUDU      IVSAGTIGTPQLLLLSGVGPESYLSSLNIPVVLSHPYVGQFLHDNPRNFINILPPNPIEP  340
                          *:*:*:**:**:*:***:**:. *. *:*****.: :**: : *** * * :    ***

sp|Q9S746|HTH_ARATH       SLIQTVGITKMGVYVEASTGFGQSPESIHTHYGIMSNKNELFSTIPAKQRRPEATQAYIT  409
sp|Q945K2|MDL2_PRUDU      TIVTVLGISNDFYQCS----FSSLP-----------FTTPPFG--------------FFP  371
                          ::: .:**::     .    *.. *            ..  *.              ::

sp|Q9S746|HTH_ARATH       RNKYQLHEAFNGSFILEKLAYPISRGHLSLV-NTNVDDNPSVTFNYFKHPVDLQRCVEAI  468
sp|Q945K2|MDL2_PRUDU      SASYPLPNSTFA-HFASKVAGPLSYGSLTLKSSSNVRVSPNVKFNYYSNLTDLSHCVSGM  430
                            .* * ::  . .: .*:* *:* * *:*  .:**  .*.*.***:.: .**.:**..:

sp|Q9S746|HTH_ARATH       RLVSKVVTSNRFLNYTQCDKQNVHKMLSLSVKANINLRPKQLNDTKSMAQFCKDTVVTIW  528
sp|Q945K2|MDL2_PRUDU      KKIGELLSTDALKPYKVEDLPGVEGFNILGI-----PLPKDQTDDAAFETFCRESVASYW  485
                          : :.::::::: :  *.  *  .*. :  *.:       **: .*  ::  **:::*.: *

sp|Q9S746|HTH_ARATH       HYHGGCLVGKVVSPNRKVLGVDRLRVIDGSTFDESPGTNPQATMMMMGRYMGVKILRERL  588
sp|Q945K2|MDL2_PRUDU      HYHGGCLVGKVLDGDFRVTGINALRVVDGSTFPYTPASHPQGFYLMLGRYVGIKILQERS  545
                          ***********:. : :* *:: ***:*****  :*.::**.  :*:***:*:***:**

sp|Q9S746|HTH_ARATH       GNKAGV------------    594
sp|Q945K2|MDL2_PRUDU      ASDLKILDSLKSAASLVL    563
                          ...  :
```

Appendix G: Arginine treatment. Various concentrations of arginine were added to the pellet of lysed Bl21(DE3) transformed with HTH_SUMO and induced via autoinduction. The soluble (S) and pellet (P) fractions were run on a 12% SDS-PAGE for each concentration of Arginine: 0M, 0.25M, 0.5M, 1M, 1.5M and 2M.



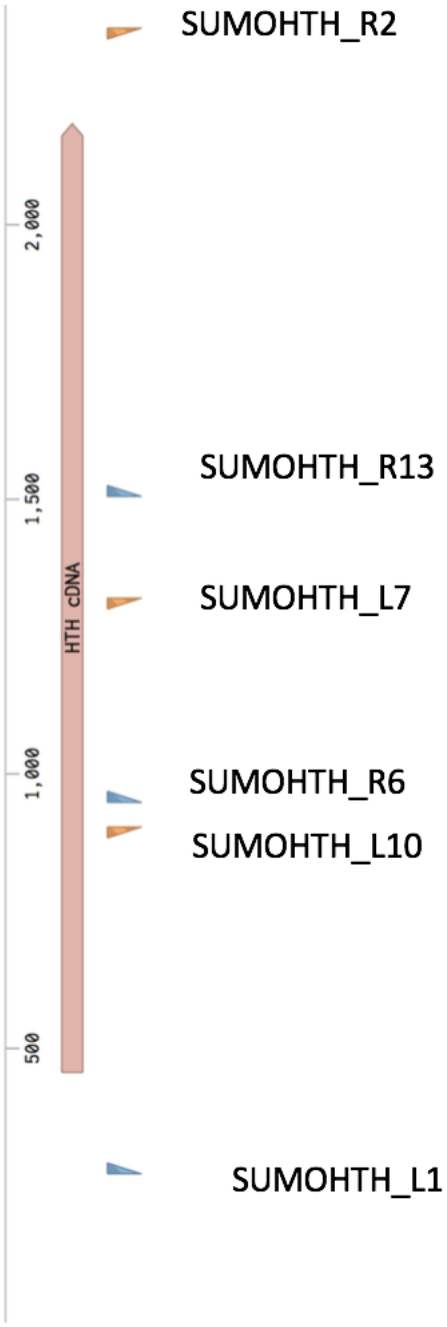
Appendix H: List of primer sets used for site-directed mutagenesis. List of mutants and the primers designed for site-directed mutagenesis. Primers were designed using NEBaseChanger.

Mutant	Direction	Sequence
hth-1	Forward	GAGCCTGATTTAAACCGTTGGTATC
	Reverse	TGTTCGATCGGCGCCTTG
htt-2	Forward	CTGAGCAGCGATGCGATTGGC
	Reverse	AATCACTTCGCTGCCCTTAC
hth-3	Forward	GATCAGCCGTAGCCACCTGAG
	Reverse	GGGTACGCCAGTTTTTCC
hth-4	Forward	TTTTGGTCGTTGTCATACCGC
	Reverse	CGATCGAAGATGGTGCCA
hth-5/11	Forward	ATTAACCTGCGTCCGAAGCA
	Reverse	AGACGCTCACGCAGGATTTT
hth-6	Forward	CAAAATCGGTAGCACCATCTTCGATC
	Reverse	GTACCGCTCACGTGGTCA
hth-7	Forward	ATTAACCTGCGTCCGAAGCA
	Reverse	AGACGCTCACGCAGGATTTT
hth-8	Forward	TGAGAGCCCGAGTACCAACCC
	Reverse	AGACGCTCACGCAGGATTTT
hth-10	Forward	CAAACCGTTGATATCACCAAAATG
	Reverse	AATCAGGCTCTGTTCGATC
Exon5 Deletion	Forward	CCAAGCGTACATCACCCGTA
	Reverse	TTCGGGCTAACCACCTTGC

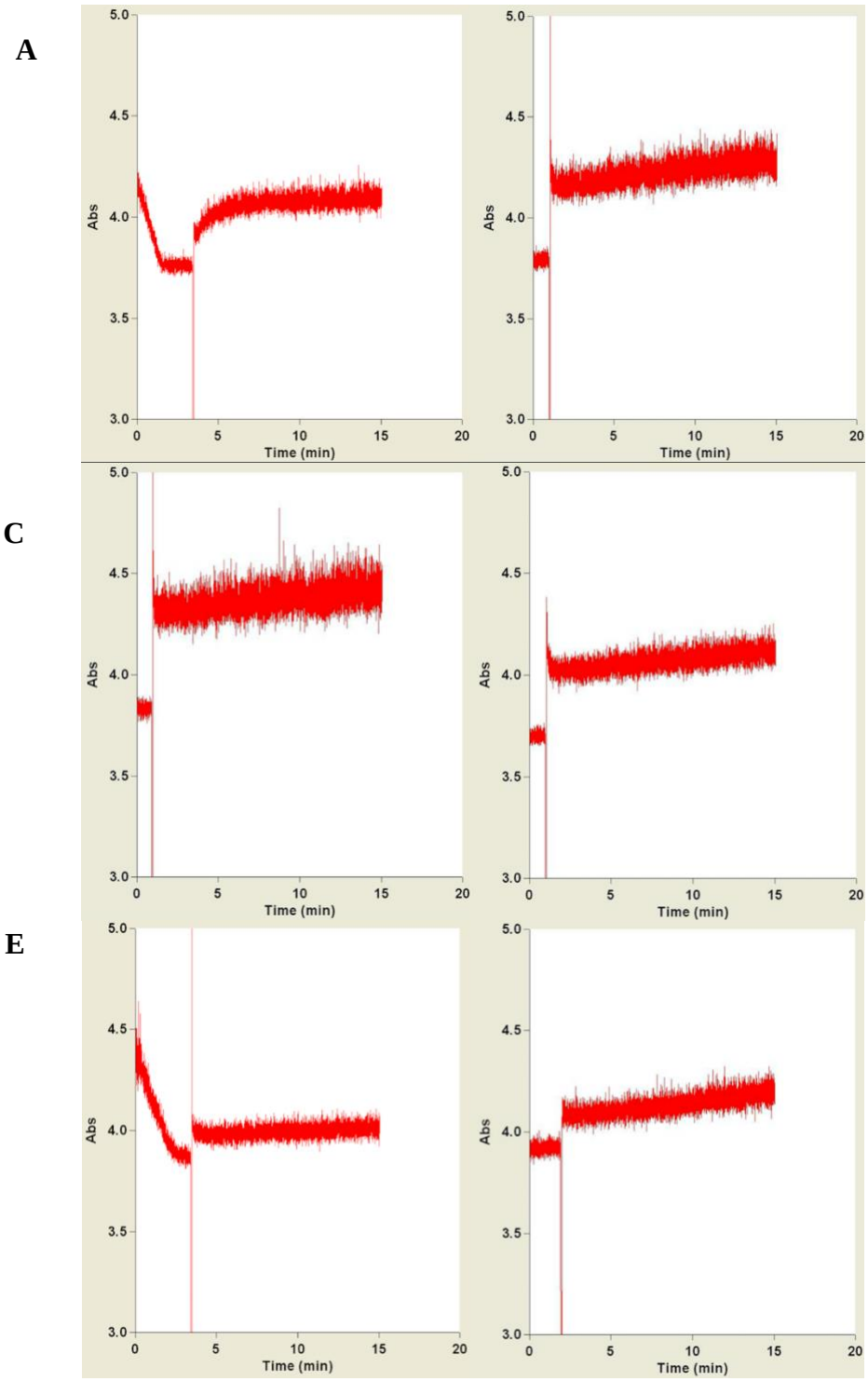
Appendix I: List of sequencing primers. Primers were designed to sequence the synthetic cDNA sequence cloned into pE_SUMOstar. The primers denoted with an asterisk (*) were used to sequence HTH mutants at the Centre for Applied Genomics (Toronto Hospital for Sick Children; Ontario, Canada).

Primer	Sequence
SUMOHTH_L1 **	AAGAAGGCTGATGGAAGCGT
SUMOHTH_L2	TTATTGAGGCTCACCGCGAA
SUMOHTH_L3	GCGAGCAAGGGCAAAGAAAA
SUMOHTH_L4	GTCAGGACAGCAGCTACGAT
SUMOHTH_L5	TGCTGGTTCTGGAGCGTG
SUMOHTH_L6	CATCGGTCTGGCGGACATTA
SUMOHTH_L7 **	CTGACCCTGTGGCAAAAAGC
SUMOHTH_L8	TCGATCGTTTTGGTCGTCGT
SUMOHTH_L9	CCGAAGAAAGAGCTGCAACG
SUMOHTH_L10 **	TTTTGGTCAGAGCCCGGAAA
SUMOHTH_L11	CCAAGCGTACATCACCCGTA
SUMOHTH_L12	TCCGTCTGGTGAGCAAAGTG
SUMOHTH_L13	AACGACACCAAAAGCATGGC
SUMOHTH_R1	CACGCTCCAGAACCAGCA
SUMOHTH_R2 **	TTTCACCAGCTTCGGATCCC
SUMOHTH_R3	CTTTTTGCCACAGGGTCAGC
SUMOHTH_R4	ACGACGACCAAAACGATCGA
SUMOHTH_R5	CGCTGGTGTCGAAAACGATT
SUMOHTH_R6 **	AATACCGCTCAGCATCAGCA
SUMOHTH_R7	CGTACACGCCCATTTTGGTG
SUMOHTH_R8	AACGCCTCGTGCAGTTGATA
SUMOHTH_R9	GACGCAGGTTAATGTTCGCC
SUMOHTH_R10	GCCATGCTTTTGGTGTCGTT
SUMOHTH_R11	CACGCCCATATAACGACCCA
SUMOHTH_R12	CAAGACCCGTTTAGAGGCCC
SUMOHTH_R13 **	CCCATTCGCCAATCCGGATA

Appendix J: Sequencing primer locations on the synthetic HTH cDNA sequence cloned into pE_SUMOstar. Six primers were used to completely sequence the synthetically synthesized HTH cDNA cloned in pE_SUMOstar. Three primers sequenced in the forward direction (HTHSUMO_L1, HTHSUMO_L7 and HTHSUMO_L10). Three primers sequenced in the reverse direction (HTHSUMO_R2, HTHSUMO_R6, HTHSUMO_R11). Primers were spaced no more than 1000 bp apart to ensure complete coverage of the sequence.
Appendix C: MDL Enzymatic Assay Traces. All assays were performed on the Cary Win UV Kinetics software and a Cary 100 UV-Vis spectrometer (Agilent Technologies). Six fractions were tested: (A) MDL, (B) *hth-9*, (C) P560, (D) P650EV, (E) P103 and (F) P103EV.



Appendix K: Crude protein extraction absorbance trace from MDN enzymatic assays. The change in absorbance was determined using Cary Win UV Kinetics software and a Cary 100 UV-Vis spectrometer (Agilent Technologies). Six samples were tested: (A) *Prunus* MDL (Sigma-Aldrich Life Sciences), (B) untransformed *hth-9 A. thaliana*, (C) P650, (D) P650EV, (E) P103 and (F) P103EV.



References

Akamura, S. N., Ano, S. M., Anaka, Y. T., Hnishi, M. O., Akamori, C. N., Raki, M. A., … Shiguro, S. I. (2010). Gateway Binary Vectors with the Bialaphos Resistance Gene , bar , as a Selection Marker for Plant Transformation. *Bioscience, Biotechnology, and Biochemistry*, *74*(6), 1315–1319. https://doi.org/10.1271/bbb.100184

Andexer, J. N., Staunig, N., Eggert, T., Kratky, C., Pohl, M., & Gruber, K. (2012). Hydroxynitrile Lyases with α/β-Hydrolase Fold: Two Enzymes with Almost Identical 3D Structures but Opposite Enantioselectivities and Different Reaction Mechanisms. *ChemBioChem*, *13*(13), 1932–1939. https://doi.org/10.1002/cbic.201200239

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*, *10*(5), 411–421. https://doi.org/10.1016/S0958-1669(99)00003-8

Baneyx, F., & Mujacic, M. (2004). Recombinant protein folding and misfolding in Escherichia coli. *Nature Biotechnology*, *22*(11), 1399–1408. https://doi.org/10.1038/nbt1029

Bannwarth, M., Bastian, S., Heckmann-pohl, D., Giffhorn, F., & Schulz, G. E. (2004). Crystal Structure of Pyranose 2-Oxidase from the White-Rot Fungus. *Biochemistry*, *43*(37), 11683–11690. https://doi.org/10.1021/bi048609q

Bio-Rad. (2012). Quick Start™ Bradform Protein Assay Instruction Manual. https://doi.org/10.1080/00173139309428968

Burda, P., & Aebi, M. (1999). The dolichol pathway of N-linked glycosylation. *Biochemica et Biophysica Acta*, *1426*(2), 239–257.

Cavener, D. R. (1992). GMC Oxidoreductases A Newly Defined Family of Homologous

Catalytic Activities. *Journal of Molecular Biology*, *223*(2), 811–814.

Cereghino, G. P. L., Cereghino, J. L., Ilgen, C., & Cregg, J. M. (2002). Production of recombinant proteins in fermenter cultures of the yeast Pichia pastoris. *Current Opinion in Biotechnology*, *13*(4), 329–332. https://doi.org/10.1016/S0958166902003300

Chang, P. P. (2016). *Localization and possible functions of Arabidopsis HOTHEAD protein*. University of Waterloo.

Chivasa, S., & Carr, J. P. (1998). Cyanide Restores N Gene-Mediated Resistance to Tobacco Mosaic Virus in Transgenic Tobacco Xxpressing Salicylic Acid Hydroxylase. *The Plant Cell*, *10*(9), 1489–1498. https://doi.org/10.1105/tpc.10.9.1489

Christensen, C. A., King, E. J., Jordan, J. R., & Drews, G. N. (1997). Megagametogenesis in Arabidopsis wild type and the Gf mutant. *Sex Plant Reproduction*, *10*(1), 49–64.

Doubayashi, D., Ootake, T., Maeda, Y., Oki, M., Tokunaga, Y., Sakurai, A., … Uchida, H. (2011). Formate Oxidase, an Enzyme of the Glucose-Methanol-Choline Oxidoreductase Family, Has a His-Arg Pair and 8-Formyl-FAD at the Catalytic Site. *Bioscience, Biotechnology, and Biochemistry*, *75*(9), 1662–1667. https://doi.org/10.1271/bbb.110153

Dreveny, I., Andryushkova, A. S., Glieder, A., Gruber, K., & Kratky, C. (2009). Substrate Binding in the FAD-Dependent Hydroxynitrile Lyase from Almond Provides Insight into the Mechanism of Cyanohydrin Formation and Explains the Absence of Dehydrogenation Activity. *Biochemistry*, *48*(15), 3370–3377. https://doi.org/10.1021/bi802162s

Dreveny, I., Gruber, K., Glieder, A., Thompson, A., & Kratky, C. (2001). The Hydroxynitrile Lyase from Almond: A Lyase that Looks Like an Oxidoreductase. *Structure*, *9*(9), 803–815. https://doi.org/10.1016/S0969-2126(01)00639-6

Drews, A. G. N., & Koltunow, A. M. G. (2011). The Female Gametophyte. *The Arabidopsis*

Book, *2011*(9). https://doi.org/10.1199/tab.0155

Ellis, S. B., Brust, P. F., Koutz, P. J., Waters, A. N. N. F., Harpold, M. M., Gingeras, T. R., & Al, E. E. T. (1985). Isolation of Alcohol Oxidase and Two Other Methanol Regulatable Genes from the Yeast Pichia pastoris. *Molecular and Cellular Biology*, *5*(5), 1111–1121.

Esashi, Y., Isuzugawa, K., Matsuyama, S., Ashino, H., & Hasegawa, R. (1991). Endogenous evolution of HCN during pre-germination periods in many seed species. *Physiologia Plantarum*, *83*(1), 27–33.

Faber, K. N., Harder, W. I. M., Ab, G., & Veenhuis, M. (1995). Review : Methylotrophic Yeasts as Factories for the Production of Foreign Proteins. *Yeast*, *11*(4), 1331–1344.

Fich, E. A., Segerson, N. A., & Rose, J. K. C. (2016). The Plant Polyester Cutin: Biosynthesis, Structure, and Biological Roles. *Annual Review of Plant Biology*, *67*, 207–233. https://doi.org/10.1146/annurev-arplant-043015-111929

Fox, B. G., & Blommel, P. G. (2017). Autoinduction of Protein Expression. *Current Protocols in Protein Science*, *Chapter 5*(Unit 5), 23. https://doi.org/10.1002/0471140864.ps0523s56.Autoinduction

Fraaije, M. W., & Mattevi, A. (2000). Flavoenzymes : diverse catalysts with recurrent features. *Trends in Biochemical Science*, *25*(3), 126–132.

Gedvilaite, A., & Sasnauskas, K. (1994). Control of the expression of the ADE2 gene of the yeast Saccharomyces cerevisiae. *Current Genetics*, *25*(6), 475–479.

Giorgetti, A., Raimondo, D., Miele, A. E., & Tramontano, A. (2005). Evaluating the usefulness of protein structure models for molecular replacement. *Bioinformatics*, *21*(2), 72–76. https://doi.org/10.1093/bioinformatics/bti1112

Grosdidier, A., Zoete, V., & Michielin, O. (2011). SwissDock, a protein-small molecule docking

web service based on EADock DSS. *Nucleic Acids Research*, *39*, 270–277. https://doi.org/10.1093/nar/gkr366

Helenius, A., Trombetta, E. S., Hebert, D. N., & Simons, J. F. (1997). Calnexin, calreticulin and the folding of glycoproteins. *Trends in Cell Biology*, *7*(5), 193–200.

Jeffery, C. J. (2017). Protein moonlighting: what is it , and why is it important? *Phylosophical Transactions of the Royal Society of London*, *19*(373), 1738.

Jorns, M. S. (1979). Mechanism of Catalysis by the Flavoenzyme Oxynitrilase. *The Journal of Biological Chemistry*, *254*(23), 12145–12152.

Kawakami, S., & Watanabe, Y. (1997). Use of Green Fluorescent Protein as a Molecular Tag of Protein Movement in vivo. *Plant Biology*, *14*(2), 127–130.

Krolikowski, K. A., Victor, J. L., Wagler, T. N., Lolle, S. J., & Pruitt, R. E. (2003). Isolation and characterization of the Arabidopsis organ fusion gene HOTHEAD. *The Plant Journal*, *35*(4), 501–511. https://doi.org/10.1046/j.1365-313X.2003.01824.x

Kudla, G., Murray, A. W., David, T., & Plotkin, J. B. (2009). Coding-Sequence Determinants of Gene Expression in Escherichia coli. *Science*, *324*(5924), 255–258. https://doi.org/10.1126/science.1170160

Kufareva, I., & Abagyan, R. (2012). Methods of protein structure comparison. *Methods of Molecular Biology*, *857*, 231–257. https://doi.org/10.1007/978-1-61779-588-6

Kurdyukov, S., Faust, A., Trenkamp, S., Bär, S., Franke, R., Efremova, N., … Yephremov, A. (2006). Genetic and biochemical evidence for involvement of HOTHEAD in the biosynthesis of long-chain α-,ω-dicarboxylic fatty acids and formation of extracellular matrix. *Planta*, *224*(2), 315–329. https://doi.org/10.1007/s00425-005-0215-7

Lee, C. (2017). A Simple Outline of Methods for Protein Isolation and Purification.

Endocrinology and Metabolism, *32*(1), 18–22.

Li, G. W., Oh, E., & Weissman, J. S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, *484*(7395), 538–541. https://doi.org/10.1038/nature10965

Li, J., Vrielink, A., Brick, P., & Blow, D. M. (1993). Crystal Structure of Cholesterol Oxidase Complexed with a Steroid Substrate: Implications for Flavin Adenine Dinucleotide Dependent Alcohol Oxidases. *Biochemistry*, *32*(43), 11507–11515. https://doi.org/10.1021/bi00094a006

Lolle, S. J., Hsu, W., & Pruitt, R. E. (1998). Genetic Analysis of Organ Fusion in Arabidopsis thaliana. *Genetics*, *149*(2), 607–619.

Lundgren, D. H., & Hwang, S. (2010). Role of spectral counting in quantitative proteomics. *Expert Review of Proteomics*, *7*(1), 39–53.

Maeda, Y., Doubayashi, D., Oki, M., Nose, H., Sakurai, A., Isa, K., … Uchida, H. (2009). Expression in Escherichia coli of an unnamed protein gene from Aspergillus oryzae RIB40 and cofactor analyses of the gene product as formate oxidase. *Bioscience, Biotechnology, and Biochemistry*, *73*(12), 2645–2649. https://doi.org/10.1271/bbb.90497

Maeda, Y., Doubayashi, D., Ootake, T., Oki, M., Mikami, B., & Uchida, H. (2010). Crystallization and preliminary X-ray analysis of formate oxidase, an enzyme of the glucose-methanol-choline oxidoreductase family. *Structural Biology and Crystallization Communications*, *66*(9), 1064–1066. https://doi.org/10.1107/S1744309110028605

Malakhov, M. P., Mattern, M. R., Malakhova, O. A., Drinker, M., Weeks, S. D., & Butt, T. R. (2004). SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *Journal of Structural and Functional Genomics*, *5*(1–2), 75–86.

https://doi.org/10.1023/B:JSFG.0000029237.70316.52

Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., & Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Science*, *15*(1), 182–189. https://doi.org/10.1110/ps.051812706.for

Maruyama, A., Yoshiyama, M., Adachi, Y., Tani, A., Hasegawa, R., & Esashi, Y. (1996). Promotion of Cocklebur Seed Germination by Allyl, Sulfur and Cyanogenic Compounds. *Plant and Cell Physiology*, *37*(8), 1054–1058. https://doi.org/10.1093/oxfordjournals.pcp.a029053

Matsushima, R., Hayashi, Y., Yamada, K., Shimada, T., Nishimura, M., & Hara-Nishimura, I. (2003). The ER Body, a Novel Endoplasmic Reticulum-Derived Structure in Arabidopsis. *Plant and Cell Physiology*, *44*(7), 661–666. https://doi.org/10.1093/pcp/pcg089

May, M. J., Vernoux, T., Leaver, C., Van Montagu, M., & Inzé, D. (1998). Glutathione homeostasis in plants: implications for environmental sensing and plant development. *Journal of Experimental Botany*, *49*(321), 649–667. Retrieved from http://jxb.oxfordjournals.org/cgi/content/abstract/49/321/649

Molina, I., Bonaventure, G., Ohlrogge, J., & Pollard, M. (2006). The lipid polyester composition of Arabidopsis thaliana and Brassica napus seeds. *Phytochemistry*, *67*(23), 2597–2610. https://doi.org/10.1016/j.phytochem.2006.09.011

Molina, I., Ohlrogge, J. B., & Pollard, M. (2008). Deposition and localization of lipid polyester in developing seeds of Brassica napus and Arabidopsis thaliana. *The Plant Journal*, *53*(3), 437–449. https://doi.org/10.1111/j.1365-313X.2007.03348.x

Morales, E. S., Parcerisa, I. L., & Ceccarelli, E. A. (2019). A novel method for removing

contaminant Hsp70 molecular chaperones from recombinant proteins. *Protein Science*, *28*(4), 800–807. https://doi.org/10.1002/pro.3574

Nielsen, H. (2017). Predicting secretory proteins with SignalP. *Methods of Molecular Biology*, *1611*, 59–73. https://doi.org/10.1007/978-1-4939-7015-5

Pagnussat, G. C., Yu, H.-J., Ngo, Q. A., Rajani, S., Mayalagu, S., Johnson, C. S., … Sundaresan, V. (2005). Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. *Development*, *132*(3), 603–614. https://doi.org/10.1242/dev.01595

Peng, J., & Xu, J. (2011). A multiple-template approach to protein threading. *Proteins*, *79*(6), 1930–1939. https://doi.org/10.1002/prot.23016.A

Petersen, T. N., Brunak, S., Heijne, G. Von, & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Piotrowski, M., & Volmer, J. J. (2006). Cyanide metabolism in higher plants : cyanoalanine hydratase is a NIT4 homolog. *Plant Molecular Biology*, *61*(1–2), 111–122. https://doi.org/10.1007/s11103-005-6217-9

Pollard, M., Beisson, F., Li, Y., & Ohlrogge, J. B. (2008). Building lipid barriers : biosynthesis of cutin and suberin. *Trends in Plant Science*, *13*(5), 236–246. https://doi.org/10.1016/j.tplants.2008.03.003

Rad, S. A., Clayton, E. J., Cornelius, E. J., Howes, T. R., & Kohalmi, S. E. (2018). Moonlighting proteins: putting the spotlight on enzymes. *Plant Signaling & Behavior*, *13*(10), 1–6. https://doi.org/10.1080/15592324.2018.1517075

Riederer, M., & Schreiber, L. (2001). Protecting against water loss: analysis of the barrier properties of plant cuticles. *Journal of Experimental Botany*, *52*(363), 2023–2032.

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*(4), 725–738. https://doi.org/10.1038/nprot.2010.5

Saberianfar, R., Sattarzadeh, A., Joensuu, J. J., & Kohalmi, S. E. (2016). Protein Bodies in Leaves Exchange Contents through the Endoplasmic Reticulum. *Frontiers in Plant Science*, *7*, 1–16. https://doi.org/10.3389/fpls.2016.00693

Schmidt, S. R. (2013). Protein Bodies in Nature and Biotechnology. *Molecular Biology*, *54*(2), 257–268. https://doi.org/10.1007/s12033-012-9563-2

Schreier, T. B., Cléry, A., Schläfli, M., Galbier, F., Stadler, M., Demarsy, E., … Zeeman, S. C. (2018). Plastidial NAD-Dependent Malate Dehydrogenase : A Moonlighting Protein Involved in Early Chloroplast Development through Its Interaction with an FtsH12-FtsHi Protease Complex. *Plant Cell*, *30*(8), 1745–1769. https://doi.org/10.1105/tpc.18.00121

Schwarz, F., & Aebi, M. (2011). Mechanisms and principles of N-linked protein glycosylation. *Current Opinion in Structural Biology*, *21*(5), 576–582. https://doi.org/10.1016/j.sbi.2011.08.005

Shumborski, S. J., Samuels, A. L., & Bird, D. A. (2016). Fine structure of the Arabidopsis stem cuticle: effects of fixation and changes over development. *Planta*, *244*(4), 843–851. https://doi.org/10.1007/s00425-016-2549-8

Siegień, I., & Bogatek, R. (2006). Cyanide action in plants - From toxic to regulatory. *Acta Physiologiae Plantarum*, *28*(5), 483–497. https://doi.org/10.1007/BF02706632

Sinclair, A. M., Trobacher, C. P., Mathur, N., Greenwood, J. S., & Mathur, J. (2009). Peroxule extension over ER-defined paths constitutes a rapid subcellular response to hydroxyl stress. *The Plant Journal*, *59*(2), 231–242. https://doi.org/10.1111/j.1365-313X.2009.03863.x

Smith, J. M., & Arteca, R. N. (2000). Molecular control of ethylene production by cyanide in Arabidopsis thaliana. *Physiologia Plantarum*, *109*(2), 180–187. https://doi.org/10.1034/j.1399-3054.2000.100210.x

Smith, S. M. (2018). Moonlighting NAD + Malate Dehydrogenase Is Essential for Chloroplast Biogenesis. *Plant Cell*, *30*, 1663–1664. https://doi.org/10.1105/tpc.18.00474

Solomonson, L. P., & Barber, M. J. (1990). Assimilatory Nitrate Reductase: Functional Properties and Regulation. *Annual Review of Plant Biology*, *41*(1), 225–253.

Solomonson, L. P., & Spehar, A. M. (1977). Model for the regulation of nitrate assimilation. *Nature*, *265*(5592), 373–375. https://doi.org/10.1038/266309a0

Sørensen, H. P., & Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of Biotechnology*, *115*(2), 113–128. https://doi.org/10.1016/j.jbiotec.2004.08.004

Studier, F. W., & August, R. (1991). Use of Bacteriophage T7 Lysozyme to Improve an Inducible T7 Expression System. *Journal of Molecular Biology*, *219*(1), 37–44.

Sunilkumar, G., Mohr, L., Lopata-finch, E., Emani, C., & Keerti, S. (2002). Developmental and tissue-specific expression of CaMV 35S promoter in cotton as revealed by GFP. *Plant Molecular Biology*, *50*(3), 463–474.

Thayer, S. S., & Conn, E. E. (1981). Subcellular Localization of Dhurrin fl-Glucosidase and

Hydroxynitrile Lyase in the Mesophyll Cells of Sorghum Leaf Blades '. *Plant Physiology*, *67*(4), 617–622.

Tischer, A., Lilie, H., Rudolph, R., & Lange, C. (2010). L-Arginine hydrochloride increases the solubility of folded and unfolded recombinant plasminogen activator rPA. *Protein Science*, *19*(9), 1783–1795. https://doi.org/10.1002/pro.465

Vanhanen, S., West, M., Kroon, J. T. M., Lindner, N., Casey, J., Cheng, Q., … Slabas, A. R. (2000). A consensus Sequence for Long-chain Fatty-acid Alcohol Oxidases from Candida Identifies a Family of Genes Involved in Lipid ω-Oxidation in Yeast with Homologues in Plants and Bacteria. *The Journal of Biological Chemistry*, *275*(6), 4445–4452. https://doi.org/10.1074/jbc.275.6.4445

Vasina, J. A., & Baneyx, F. (1997). Expression of Aggregation-Prone Recombinant Proteins at Low Temperatures: A Comparative Study of the Escherichia coli cspA and tac Promoter Systems. *Protein Expression and Purification*, *9*(2), 211–218. https://doi.org/10.1006/prep.1996.0678

Verbeke, J. A., & Walker, D. B. (1986). Morphogenetic factors controlling differentiation and dedifferentiation of epidermal cells in the gynoecium of Catharanthus roseus. *Planta*, *168*(1), 43–49. Retrieved from papers2://publication/uuid/5E4EEDCC-9AAA-4C5C-BF16-3038C7A34563

Vinther, A., Jørgensen, K., Jørgensen, C., Michelle, S., Sánchez-pérez, R., Lindberg, B., & Bak, S. (2008). β-Glucosidases as detonators of plant chemical defense. *Phytochemistry*, *69*(9), 1795–1813. https://doi.org/10.1016/j.phytochem.2008.03.006

Wegner, G. H. (1990). Emerging applications of the methylotrophic yeasts. *FEMS Microbiology Reviews*, *87*(3–4), 279–284. https://doi.org/10.1016/0378-1097(90)90467-5

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., & Provart, N. J. (2007). An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLoS ONE*, *2*(8), 718. https://doi.org/10.1371/journal.pone.0000718

Wu, H., & Poulton, J. E. (1991). Immunocytochemical Localization of Mandelonitrile Lyase in Mature Black Cherry (Prunus serotina Ehrh.) Seeds. *Plant Physiology*, *96*(4), 1329–1337.

Yamada, K., Nagano, A. J., Nishina, M., Hara-nishimura, I., & Nishimura, M. (2008). NAI2 Is an Endoplasmic Reticulum Body Component That Enables ER Body Formation in Arabidopsis thaliana. *Plant Cell*, *20*(9), 2529–2540. https://doi.org/10.1105/tpc.108.059345

Yamada, K., Nagano, A. J., Nishina, M., Hara-nishimura, I., & Nishimura, M. (2013). Identification of Two Novel Endoplasmic Reticulum Body-Speci fi c Integral Membrane Proteins. *Plant Physiology*, *161*(January), 108–120. https://doi.org/10.1104/pp.112.207654

Yang, J., Roy, A., & Zhang, Y. (2013a). BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Research*, *41*(D1), D1096–D1103. https://doi.org/10.1093/nar/gks966

Yang, J., Roy, A., & Zhang, Y. (2013b). Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, *29*(20), 2588–2595. https://doi.org/10.1093/bioinformatics/btt447

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, *12*(1), 7–8. https://doi.org/10.1038/nmeth.3213

Yang, J., & Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research*, *43*(W1), W174–W181. https://doi.org/10.1093/nar/gkv342

Yang, W., Pollard, M., Li-Beisson, Y., & Ohlrogge, J. (2016). Quantitative analysis of glycerol in dicarboxylic acid-rich cutins provides insights into Arabidopsis cutin structure. *Phytochemistry*, *130*, 159–169. https://doi.org/10.1016/j.phytochem.2016.03.017

Yeats, T. H., & Rose, J. K. C. (2013). The Formation and Function of Plant Cuticles. *Plant Physiology*, *163*(1), 5–20. https://doi.org/10.1104/pp.113.222737

Yue, Q. K., Kass, I. J., Sampson, N. S., & Vrielink, A. (1999). Crystal Structure Determination of Cholesterol Oxidase from Streptomyces and Structural Characterization of Key Active Site Mutants. *Biochemistry*, *38*(14), 4277–4286. https://doi.org/10.1021/bi982497j

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*(40), 1–8. https://doi.org/10.1186/1471-2105-9-40

Zhang, Y., & Skolnick, J. (2004). Scoring Function for Automated Assessment of Protein Structure Template Quality. *Proteins: Structure, Function, and Bioinformatics*, *57*(4), 702–710. https://doi.org/10.1002/prot.20264

Zonneveld, B. J. M., & Van Der Zanden, A. L. (1995). The Red ade Mutants of Kluyveromyces lactis and their Classification by Complementation with Cloned A DE7 or ADE2 Genes from Saccharomyces cerevisiae. *Yeast*, *11*(9), 823–827.

Zuo, X., Li, S., Hall, J., Mattern, M. R., Tran, H., Shoo, J., … Butt, T. R. (2005). Enhanced expression and purification of membrane proteins by SUMO fusion in escherichia coli. *Journal of Structural and Functional Genomics*, *6*(2–3), 103–111. https://doi.org/10.1007/s10969-005-2664-4