Duplication of Alcohol Dehydrogenases Unlocks the Chemical Diversity of the Medicinal Plant *Catharanthus roseus*

Anna Katerina Stavrinides

This thesis is submitted in fulfilment of the requirements of the degree of Doctor of Philosophy at the University of East Anglia

Department of Biological Chemistry

John Innes Centre

Norwich

September 2016

Abstract

This thesis details the discovery and characterisation of biosynthetic enzymes implicated in the monoterpene indole alkaloid (MIA) pathway of the medicinal plant *Catharanthus roseus*. The MIA pathway is characterised by a plethora of different carbon skeletons, which are derived from the central pathway intermediate strictosidine aglycon. Given the biological importance of these compounds, there is great interest in identifying the enzymes that catalyse the formation of these different carbon skeletons, as well as understanding the mechanistic basis for how the pathway is controlled at this critical step in the pathway.

The discovery of the first enzyme capable of reducing the reactive aglycon intermediate to a heteroyohimbine-type MIA (tetrahydroalstonine) opened the door to discovery of many other heteroyohimbine synthases (HYS, Chapter 2). These enzymes share a degree of sequence identity and are all members of the Medium-Chain Dehydrogenase/Reductases. Interestingly, these HYSs catalysed the formation of different ratios of several heteroyohimbine stereoisomers. A detailed mutation screen, together with protein crystallography, deuterium labelling, and *in silico* docking enabled us to propose a catalytic mechanism for these enzymes and how the ratio of products is controlled (Chapter 3). An investigation into a different family of reductases, the Short-Chain Dehydrogenase/Reductases, revealed another enzyme capable of reducing the reactive strictosidine aglycon intermediate (Chapter 4). After extensive NMR characterisation, the enzymatic product was found to possess an unusual carbon skeleton different to that of the heteroyohimbines.

Characterisation of the phylogeny of these enzymes revealed that they have undergone numerous duplication events (Chapter 5). The HYSs appear to have undergone multiple duplications and neofunctionalisation that has given rise to at least one other biosynthetic enzyme which acts in one of the downstream MIA pathway branches. A study of a large reductase duplication locus provides evidence that pathway clustering in plants arises through translocation of biosynthetic genes.

The discovery of these reductases has provided us with an unprecedented opportunity to study the dynamics of the branch-point of the MIA pathway. These discoveries constitute an important step towards the elucidation of the MIA pathway in *C. roseus* and in the many related MIA producing plants.

Acknowledgements

This PhD has been one of the hardest things I have ever done. Although I sometimes asked myself why I put myself through this ordeal, in the end I hope it has been worth it. The last four years have been a journey of discovery, and I am not the same person as I was when I began, professionally and personally.

I have met some great people, made some amazing friends, and learned a lot about science and life. The Biological Chemistry department, and the whole John Innes Centre, has been an inspiring place to work in. There are so many brilliant people who are doing great science and who have fascinating stories to tell. I wish I had more time to talk with and learn from everyone.

I owe this success to the excellent supervision that Sarah O'Connor offered over the years. She has been a great mentor, and she allowed me to follow through with my ideas and let me work independently. It would not have been possible without the support and help from Lorenzo Caputi and Evangelos Tatsis. Lorenzo has been a great friend over the years and has always been there to offer his help and expertise. He has also been the level-headed advisor in times of need and he was a real blessing to me and the team as a whole. Evangelos was indispensable for my education in the chemistry aspects of this PhD. He is a great lab partner and his work ethic never ceases to impress me. I am very grateful to have worked with him.

My time in the group would not have been the same without Dorota Jakubczyk. She is a great friend and could always lend me a shoulder to cry on when necessary. Thuy Dang is a very kind and inspiring person and I'm happy I had the chance to get to know her better over the years. Omar Kamileen will always hold a special place in my heart, his easy-going and cheerful personality brightens up the lab and always made my days a bit more manageable.

I would also like to thank the past group member Weslee Glenn for being welcoming and for passing on his enzymes at the end of his PhD. Weslee taught me not to take things too seriously, and at the hardest times I made an effort to remember this lesson. Hajo Kries was also a great group member who encouraged me with my crazy projects and shared my enthusiasm when many others did not. Thank you to current members Don Nguyen, Benjy Lichman, and Scott Farrow with whom I had many interesting conversations.

Throughout my PhD one person has really stood out and supported me more than anyone else, even though he was miles away. I owe my sanity to my loving partner, Alex, who has never ceased to encourage me and was always there to hear my rants. He brought a breath of fresh air to my time as a PhD student and really helped me put things back into perspective when I lost sight of the path. I can't express how grateful I am to have him in my life.

I owe a lot to my mother, Aileen, who has always been a role model to me. I admire her for her outgoing and happy personality and general optimism. She can always sense when I am down and is always there for me with kind words. She has been my crutch since my teenage years, and I would not have got this far without her love and support.

My father, Stavros, encouraged me to follow my passions and trusted my decisions. I know he would have been very proud of me and I wish he were here to share this moment with me. This thesis is dedicated to his memory.

Contents

List of Tables

List of Figures

Chapter 2

Thesis Abbreviations Used

10HGO	10- hydroxygeraniol-oxidoreductase
19-EA	19-epiajmalicine
7-DLGT	7-deoxyloganetic acid-O-glucosyl transferase
7-DLH	7-deoxyloganic acid hydroxylase
ACN	Acetonitrile
ADH	Alcohol Dehydrogenase
AJM	Ajmalicine
AKR	Aldo-keto reductase
AS	Anthranilate synthase
BIA	Benzylisoquinoline alkaloid
BPF	box P-binding factor
CAD	Cinnamyl alcohol dehydrogenase
COSY	Homonuclear correlation spectroscopy
CPR	Cytochrome P450 reductase
D4H	Desacetoxyvindoline 4-hydroxylase
DAT	Deacetylvindoline acetyltransferase
DMAPP	Dimethylallyl pyrophosphate
DXR	1-deoxy-D-xylulose-5-phosphate reductoisomerase
DXS	1-deoxy-D-xylulose-5-phosphate synthase
EDTA	Ethylenediaminetetraacetic acid
EtOAc	Ethyl acetate
FA	Formic acid
G10H	Geraniol 10-hydroxylase
GBF	G-box binding factor
H2BC	Heteronuclear multiple-bond correlation over two bonds spectroscopy
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HSQC	Heteronuclear single quantum correlation
HYS	Heteroyohimbine Synthase
IO	Iridoid oxidase
IPP	Isopentenyl pyrophosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside

ISY	Iridoid synthase
LAMT	Loganic acid methyltrasferase
MAT	Minovincine 19-hydroxy-O-acetyltransferase
MDR	Medium chain Dehydrogenase/Reductase
MeJa	Methyl Jasmonate
MeOH	Methanol
MEP	2-C-Methyl-D-erythritol 4-phosphate
MIA	Monoterpene Indole Alkaloid
MYC2	MYC2 transcription factor
NMR	Nuclear magnetic resonance
NMT	N-methyltransferase
ORCA3	octadecanoid responsive Catharanthus AP2-domain protein
P450	Cytochrome P450
PAGE	Polyacrylamide Gel Electrophoresis
Prx1	peroxidase
ROESY	Rotatin frame nuclear Overhauser effect spectroscopy
SAD	Sinapyl alcohol dehydrogenase
SDR	Short-chain dehydrogenase/reductase
SDS	Sodium Dodecyl Sulfate
SGD	Strictosidine beta-glucosidase
SLS	Secologanin synthase
STR	Strictosidine synthase
T16H	Tabersonine 16-hydroxylase
T16OMT	16-hydroxytabersonine-O-methyltransferase
T19H	Tabersonine 19-hydroxylase
T3O	Tabersonine-3-oxygenase
T3R	Tabersonine-3-reductase
TCEP	Tris(2-carboxyethyl)phosphine
TDC	Tryptophan decarboxylase
TEA	Triethylamine
TFA	Trifluoroacetic acid
THA	Tetrahydroalstonine
THAS	Tetrahydroalstonine Synthase

TIA	Terpene indole alkaloid
TLC	Thin-layer chromatography
WGC	Whole genome contig
WRKY1	WRKY transcription factor 1
ZCT	Transcription factor IIIA-type zinc finger family

Chapter 1

Introduction

1.1 Chemical diversity of medicinal plants

Plants, as sedentary beings preyed upon by many herbivores, have had to evolve methods of defence. Often this defence can come in the form of the plant's secondary metabolism. A well-studied example is found in the brassicas, which produce glucosinolates that are unpalatable to many animals (Giamoustaris and Mithen, 1995, Hopkins et al., 2009). Another large group of plant secondary metabolites are the alkaloids. These compounds are characterised by a bitter taste and therefore can serve as a deterrent or antifeedant. Many alkaloids have been linked to animal toxicity. Many also have neurological effects, acting as antagonists to neurotransmitters; most alkaloids are derived from amino acids and are structurally similar to neurotransmitters, which are also made from amino acids (Wink and Schimmer, 2010, Strauss et al., 2002). Plants have been extensively studied for centuries to isolate, identify and understand these bioactive components that are the products of secondary metabolism.

Contrary to primary metabolism, which is shared among large groups of living organisms, secondary metabolism can be delimited to only one or a small group of related organisms (Dewick, 2011). Studying the secondary metabolome of plants has led to the discovery of different compounds with novel structural components, some of which have been used in the clinic (Saklani and Kutty, 2008, Cushnie et al., 2014). Many fundamental studies of plant extracts have been published and show that the field is very active and prolific (for example Google Scholar lists 30 000+ publications between 2006 and 2016 under the search term "medicinal plant antimicrobial screening"). Despite all these efforts, most plant species on earth have never been analysed for their chemical constituents and even fewer have been tested for biological activity (Mahesh and Satish, 2008). The potential of discovery of new compounds with medical benefits to humans from plants is certain.

However, discovery of bioactive compounds alone is not enough to translate to efficient use in the clinic. These natural products are often present in only trace amounts in the host tissues and chemical synthesis can be prohibitively long and expensive. A deeper understanding of the biosynthesis of those compounds will provide the scientific community with the knowledge necessary to engineer production of those beneficial compounds in large amounts. This could come about through many different ways: 1) improved breeding of plants using marker assisted selection for varieties expressing the biosynthetic genes at higher levels; 2) genetically modified plants, for example expression of the pathway in a tissue that is faster growing or more easily harvested; 3) genetic engineering of microorganisms such as baker's yeast (*Saccharomyces cerevisiae*) which can attain higher mass faster and is compatible with current fermenter technologies. Furthermore, a better understanding of the enzyme mechanisms and the enzymatic specificity can facilitate enzyme engineering for production of unnatural natural products.

Some plant families present extraordinary richness in their secondary metabolism. Of notable mention is the Ranunculales clade which comprises many species with medicinally important secondary metabolites. Perhaps the best known example in this clade is the opium poppy (*Papaver somniferum*), which has been used by humans throughout history for its medicinal properties. The order Gentianales (of the Asterid clade) also comprises many species rich in secondary metabolites used by humans throughout history. The Gentianales family Apocynaceae has been well studied in large part due to two economically-important medicinal plants: *Rauvolfia serpentina* and *Catharanthus roseus*. The bioactive components of these plants are alkaloids, a class of natural products that contain at least one atom of nitrogen in a heterocycle. This thesis will focus on these tryptophan-derived indole alkaloids and how their chemical diversity is generated in *C. roseus*.

1.2 Indole alkaloids

The Indole alkaloids are a very large and diverse family of more than 2000 natural products (O'Connor and Maresh, 2006). The structural diversity of the indole alkaloids is mirrored in their wide clinical use from anti-tumour to antimalarial agents (fig. 1).

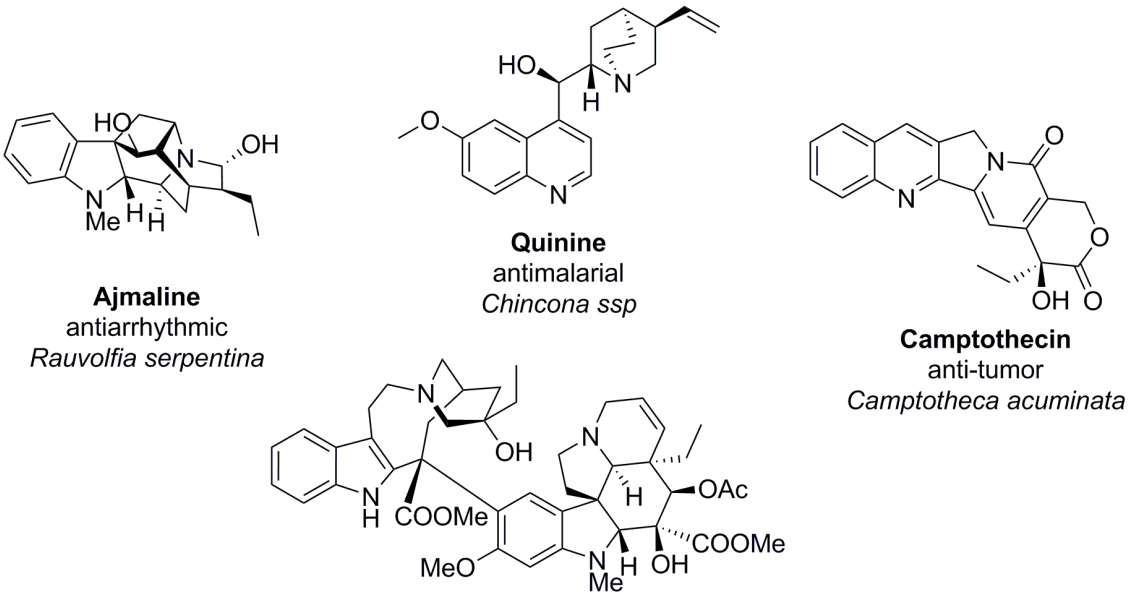


Figure 1: Selected indole alkaloids with medicinal uses. Indole alkaloids currently used as treatments in the clinic, including the plant from which they were first identified.

A major subclass of indole alkaloids, the monoterpene indole alkaloids (MIA), are derived from tryptamine (produced by decarboxylation of the amino acid tryptophan) and one molecule of terpene in the form of secologanin (fig. 2) (Szabó, 2008). Tryptamine and secologanin are condensed to form the central intermediate of the pathway. During chemical condensation of these two moieties, two isomers are produced with opposite stereochemistry at C-3: vincoside

(3-*R*) and strictosidine (3-*S*) (Battersby et al., 1969)(fig. 2). However when these compounds are condensed enzymatically only the strictosidine isomer is produced (fig. 3). Initial confusion about the precursor of the terpene indole alkaloids was dispelled by Stöckigt and Zenk (1977) who showed that proteins crudely purified from indole alkaloid-producing plants produce only the strictosidine diastereomer, with 3-(*S*) configuration. Feeding with [3-³H]-labelled strictosidine indicated that the 3-(*S*) diastereomer is the precursor to all monoterpene indole alkaloids, including those with 3-(*R*) configuration (Rueffer et al., 1978)(fig. 3). Therefore strictosidine is the universal precursor to thousands of monoterpene indole alkaloids found in numerous plant species.

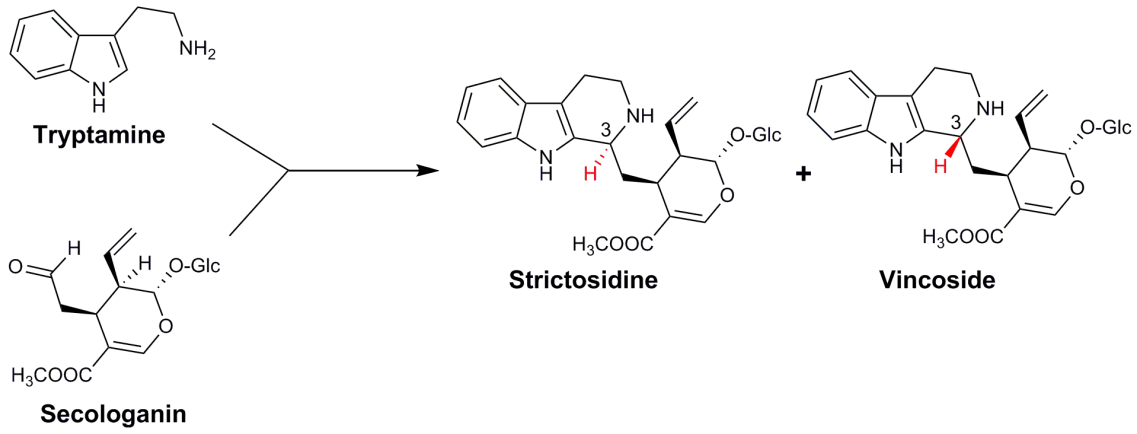


Figure 2: Chemical condensation of tryptamine and secologanin. This reaction yields vincoside and strictosidine, the two diastereomers at C-3. Only strictosidine is produced enzymatically.

1.3 Strictosidine Synthase

The enzyme responsible for this stereoselective condensation is Strictosidine Synthase (STR, EC 4.3.3.2). It was first discovered in *C. roseus* in 1977 (Stöckigt and Zenk, 1977, Treimer and Zenk, 1979) and was cloned for the first time from *Rauvolfia serpentina* in 1988 (Kutchan et al., 1988). STR produces strictosidine through a Pictet-Spengler condensation (Stöckigt et al., 2011) of secologanin and tryptamine. This enzyme catalyses the first committed step in MIA biosynthesis and has been extensively studied.

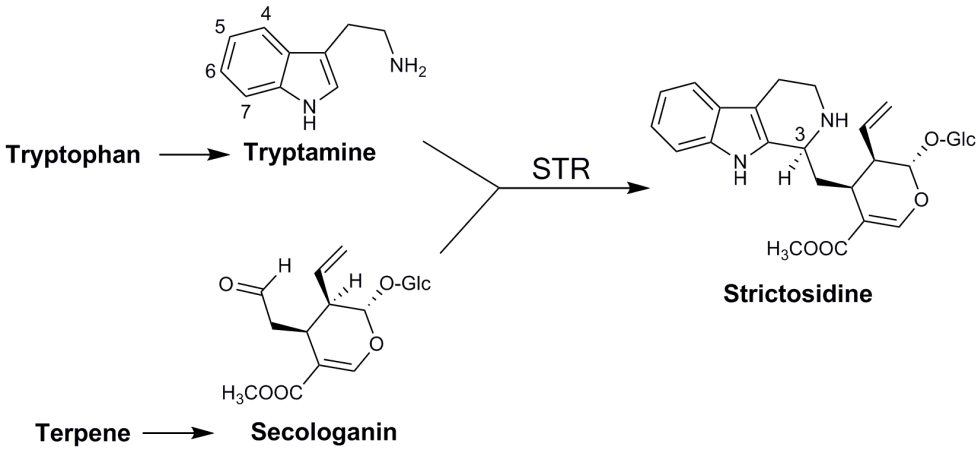


Figure 3: Strictosidine biosynthesis from tryptamine and secologanin.

STR has strict product stereospecificity; it only produces the 3-(*S*) diastereomer. It also has limited substrate specificity and it does not accept other amines such as histamine, tyramine or tryptophan and it does not tolerate substitutions to the C-5 and C-6 positions of the tryptamine indole ring (McCoy et al., 2006). Furthermore, the catalytic efficiency of the enzyme is not affected by the presence of downstream alkaloids such as ajmalicine (AJM), catharanthine, or vindoline (Mizukami et al., 1979), which signifies that these downstream alkaloids do not bind to the enzyme active site. Possibly negative feedback loops occur at other points in the pathway but are not necessary at the step of strictosidine formation.

Crystal structures of STR were used to understand the mechanism of this enzymatic transformation and to design mutations to broaden its substrate specificity (Loris et al., 2007, Chen et al., 2006). The broad-specificity engineered STR with the point mutation in the active site cavity (V214M) was transformed into *C. roseus* hairy root lines (Runguphan and O'Connor, 2009). These lines were fed commercially available tryptamine analogs and the resulting alkaloids were purified and characterised. It was found that the substituted tryptamines were accepted by the mutated STR and a variety of MIAs was found to have incorporated the analog. AJM and serpentine had incorporated all three analogs tested (chlorinated, methylated or brominated at position 10 of tryptamine) whereas tabersonine and catharanthine had not incorporated the brominated analog. This demonstrates how crystallisation of an enzyme can point to the amino acids which help control substrate specificity. Furthermore, these experiments illustrate the potential flexibility of the biosynthetic enzymes of this pathway and highlight there is room for manipulation of the substrate specificity in order to expand the chemical diversity.

1.4 Deglycosylation of strictosidine

Strictosidine β-D-glycosidase (SGD) is a key enzyme in the biosynthetic pathways of plants that produce monoterpene indole alkaloids. SGD cleaves the glucose moiety of strictosidine, thus releasing a reactive intermediate which can rearrange to form the precursor to the downstream alkaloids such as AJM, catharanthine and vindoline (fig. 5). This reactive intermediate however also has another role, that of defence. The dialdehyde form of the intermediate is able to react with free amines, such as those found on proteins, and is also able to crosslink DNA, making strictosidine aglycon a formidable defence compound (Guirimand et al., 2010). Interestingly, strictosidine is stored in the vacuole of the cell and SGD is present in the nucleus (Guirimand et al., 2010). The spatial segregation of these two signifies that strictosidine can be classified as a phytoanticipin molecule (Morant et al., 2008). This segregation has been reported before for many combinations of β-glucosidases and their substrates in plants (Morant et al., 2008). An example of this is the dhurrin cyanogenic glycoside from *Sorghum bicolor* which is used as a defence compound (Tattersall et al., 2001)(fig. 4).

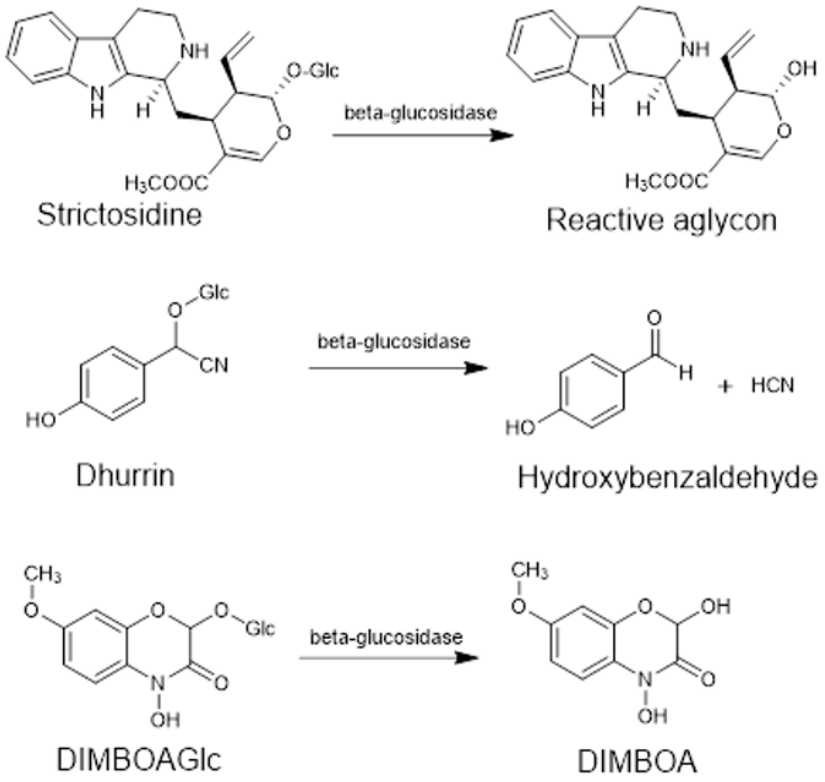


Figure 4: Aglycon formation from strictosidine, dhurrin, and DIMBOAGlc. Redrawn from Morant et al. (2008).

The biological role of strictosidine in the plant is not well understood. Efforts to elucidate the effect on pathogens and insects have been few. A study conducted recently found that compounds with tetrahydro-beta-carboline moieties (such as strictosidine) have a slight retardant effect on the growth of the generalist herbivore *Spodoptera littoralis* (Sudžuković et al., 2016), but that did not affect the lifecycle of this insect. Strictosidine had an inhibitory effect on the growth of fungal plant pathogens such as *Fusarium oxysporum* and *Cladosporium cucumerinum*. But when strictosidine was incubated in the presence of SGD, the inhibitory effect was stronger (Luijendijk et al., 1996). This suggests strictosidine could be involved in plant defence, particularly when coupled with SGD which can release the reactive intermediate. Wounded leaves would release the reactive intermediate thus preventing more fungal growth or acting as an antifeedant for insect herbivores (Guirimand et al., 2010).

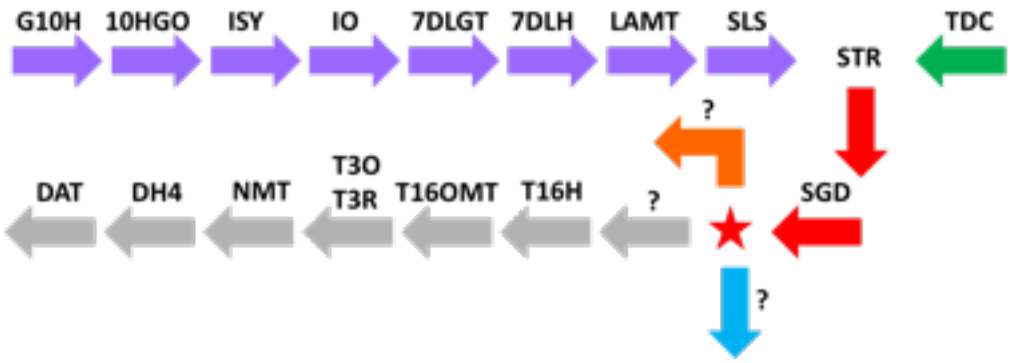


Figure 5: *C. roseus* general MIA pathway. The upstream part of the pathway that generates the monoterpene precursor is shown in lilac; tryptophan decarboxylation in green; generation of strictosidine and the central intermediate strictosidine aglycon in red; vindoline biosynthesis in gray; catharanthine in blue; AJM biosynthesis in orange. The star represents the central intermediate strictosidine aglycon, and question marks indicate pathway steps not characterised at the onset of this thesis. Enzymes: G10H: geraniol-10-hydroxylase; 10HGO: 10-hydroxygeraniol-oxidoreductase; ISY: iridoid synthase; IO: iridoid oxidase; 7-DLGT: 7-deoxyloganetic acid glucosyl transferase; 7-DLH: 7-deoxyloganic acid hydroxylase; LAMT: loganic acid O-methyltransferase; SLS: secologanin synthase; TDC: tryptophan decarboxylase; STR: strictosidine synthase; SGD: strictosidine glucosidase; T16H: tabersonine-16-hydroxylase; T16OMT: 16-hydroxytabersonine-O-methyltransferase; T3O: tabersonine-3-oxygenase; T3R: tabersonine-3-reductase; NMT: N-methyltransferase; D4H: desacetoxyvindoline-4-hydroxylase; DAT: deacetylvindoline-4-O-acetlytransferase.

Contrary to the limited substrate promiscuity of STR, SGD tolerates a wide range of substitutions to the indole moiety of strictosidine, but is not very active against strictosidine analogs with substitutions on the secologanin moiety (Yerkes et al., 2008). *C. roseus* SGD also deglycosylates the strictosidine isomer vincoside. Therefore, SGD appears to be more permissive of unnatural substrates.

1.5 Spatiotemporal localisation

As with many plant processes, secondary metabolic pathways can be induced upon elicitation. Adaptive responses to the environment or to an aggressor can help the plant regulate its use of resources. For that reason secondary metabolites can fluctuate in concentration both in response to the seasons, and in response to pathogen or herbivore attack. Targeted induction can also provide a way to produce a secondary metabolite only in the tissues where it is needed, for example the roots, the bark, or the leaves.

The plant hormone methyl jasmonate (MeJa) has an elicitation effect on the expression of some parts of the pathway, but appears to act in different ways for upstream (tryptophan decarboxylase, TDC) or downstream enzymes (desacetoxyvindoline-4-hydroxylase, D4H) (Vázquez-Flotaand and De Luca, 1998). The metabolic flux through the early part of the pathway seems to be very tightly regulated. During efforts to engineer the production of 'unnatural' natural products in *C. roseus* hairy root lines, Glenn et al. (2011) attempted to overexpress TDC to increase the flux from the indole pathway of primary metabolism into the indole alkaloid secondary metabolism. The authors attempted to overexpress TDC in a constitutive manner in hairy roots which

unfortunately was lethal, even when the hairy roots were grown on media supplemented with 500 µM of L-tryptophan. The tight regulation by MeJa of the upstream pathway is further evidenced by the successful use of transcriptomes of MeJa-elicited *C. roseus* tissues for discovery of genes in that part of the pathway (Miettinen et al., 2014, Geu-Flores et al., 2012).

Further studies were focused on discovering the transcription factors of the pathway in *C. roseus*. This is of importance as genetically engineered plants or cell cultures overexpressing the pathway biosynthetic enzymes could produce more downstream metabolites, thus reducing the cost of the valuable end-product, such as the anti-cancer agents vinblastine and vincristine. Hairy root lines were developed by Sun and Peebles (2015) to overexpress the ORCA3 transcription factor which has been shown to control the expression of multiple pathway genes (van der Fits and Memelink, 2000). They discovered that in this line, compared to the same line without induction of ORCA3 expression, many genes in the pathway (both upstream and downstream) were upregulated, but this did not produce a statistically significant increase in any strictosidine-derived alkaloid they measured (serpentine, AJM, catharanthine, hörhammericine) with the exception of lochnericine (56% increase, $p<0.05$). However, the authors noticed a statistically significant decrease in tabersonine (-97%, $p<0.05$) and also a slight decrease in SGD mRNA levels based on RT-qPCR. Interestingly, in this ORCA3 overexpressing line, they noted that many negative transcription factors (ZCT1, ZCT2, ZCT3, GBF1, GBF2, GBF3) were upregulated. To counter the SGD downregulation, Sun and Peebles (2015) developed another hairy root line which overexpressed both ORCA3 and SGD under the control of an inducible promoter. In this line they achieved a ten-fold overexpression of SGD and they observed a statistically significant increase in all the alkaloids they measured. These results indicate that the regulation of the pathway is not controlled by a single transcription factor, but is instead more complex and involves multiple positive and negative regulators. SGD is not under the control of the ORCA3 transcription factor, and is in fact downregulated when ORCA3 is upregulated. SGD appears to be a rate limiting step at this part of the pathway, at least in hairy roots, based on the increase of MIAs observed when SGD is upregulated.

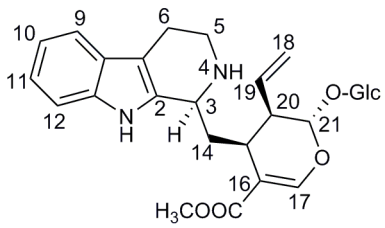
The MIA biosynthetic pathway in *C. roseus* has been the subject of many studies relating to its spatiotemporal localisation. The various MIA it produces can be found in different tissues (van der Heijden et al., 2004, McCoy and O'Connor, 2006, Laflamme et al., 2001). This clearly suggests one of two scenarios are possible; either the enzymes responsible for the different products are expressed differentially in different tissues, or the different MIA are actively transported for sequestration in different tissues. There is strong evidence that the first scenario is more likely, based on the expression of the pathway biosynthetic enzymes in different tissues and cell types (Guirimand et al., 2011b, Verma et al., 2012, Murata et al., 2008). However, these results also point to the fact that there must be transport of metabolic intermediates between cell types in order for the final products to be produced (St-Pierre et al., 1999).

Furthermore, it was discovered that even when parts of the pathway are all expressed in the same tissue (leaves for example) certain enzymes are expressed in different cell types. The pathway involves at least 4 different cell types in the leaves (Courdavault et al., 2014). The internal phloem associated parenchyma cells are the location of the upstream part of the pathway, from geraniol to loganic acid biosynthesis (fig. 9). The latter is transported, through an unknown transporter, to the leaf epidermis cells where the central part of the biosynthesis takes place, namely the synthesis of secologanin, strictosidine, and downstream to catharanthine and 16-methoxytabersonine (fig. 9 and 10). Finally, vindoline biosynthesis occurs in the laticifers and idioblasts (Courdavault et al., 2014, De Luca et al., 2014).

Adding an additional layer of complexity, many enzymes of the pathway are localised to intracellular compartments. The upstream part of the pathway occurs in the plastid until geraniol is released to the cytosol. Many enzymes downstream from geraniol are soluble in the cytosol or are anchored to the ER membrane (cytochrome P450 enzymes) (Courdavault et al., 2014, Guirimand et al., 2011a, Guirimand et al., 2011b). Interestingly though, STR, along with its product strictosidine, is localised to the vacuole and the next biosynthetic enzyme, SGD, is localised to the nucleus (Guirimand et al., 2010). This spatial separation of STR from the cytosol and of SGD from the strictosidine pool has implications on the substrate availability for each of the enzymes. It is apparent that transport of secologanin and tryptamine, the substrates of STR, must occur in order for strictosidine to be produced. Similarly, transport of strictosidine out of the vacuole must be controlled by a transporter in order for SGD to access its substrate and produce substrates for the downstream parts of the pathway. The flow of strictosidine out of the vacuole is could be one of the bottlenecks in this pathway and understanding how it is controlled will be a breakthrough in engineering higher vinblastine-producing plants.

1.6 Structural rearrangements unlock chemical diversity in MIA

When strictosidine is deglycosylated it produces a reactive and unstable intermediate (Brown and Chapple, 1974, Husson et al., 1977, Kan-Fan and Husson, 1979). Early studies using ^{14}C-labelled tryptophan feeding with *C. roseus* seedlings helped establish the sequence of reactions leading to the downstream alkaloids from strictosidine onwards (fig. 6)(Scott et al., 1971). It is clear from these early experiments that the deglycosylation of strictosidine unlocks its potential to rearrange and form new carbon skeletons. It was also noted that strictosidine can non-enzymatically lose its glucose moiety which then can lead to the rearrangement of the backbone through reaction of the N4 with a free aldehyde (Smith, 1968).



Strictosidine

Figure 6: Strictosidine aglycon numbering.

Rearrangement of strictosidine aglycon occurs due to the reactive aldehyde groups on carbon 17 and 21. As illustrated in fig. 7 either of these aldehydes can attack the nitrogen to produce a cyclised molecule. This occurs after rotation of the molecule around either the C-14-C-15 bond or the C-15-C-20 bond (blue or pink arrow). Rotation around the C-14-C-15 bond results in cyclisation through the C-17 whereas rotation around the C-15-C-20 bond results in cyclisation through the C-21. In the first case this leads to generation of the vallesiachotaman skeleton, and the second to the corynanthean skeleton. The second scenario further leads to the generation of chemical diversity which arises through equilibration and/or cyclisation of the final ring. This cyclisation can either occur through attack of the C-18 onto C-17 (which leads to the yohimbine alkaloids) or through attack of the hydroxyl of C-17 onto C-19 (which leads to cathenamine and the heteroyohimbine alkaloids). Finally, if reduction of the C-21-N bond occurs before cyclisation then a different mode cyclisation can take place. This involves the intermediate geissoschizine which can undergo rearrangements to produce preakuammicine. This undergoes further rearrangements (through enzymatic catalysis on dehydrosecodine) to yield the products catharanthine and tabersonine.

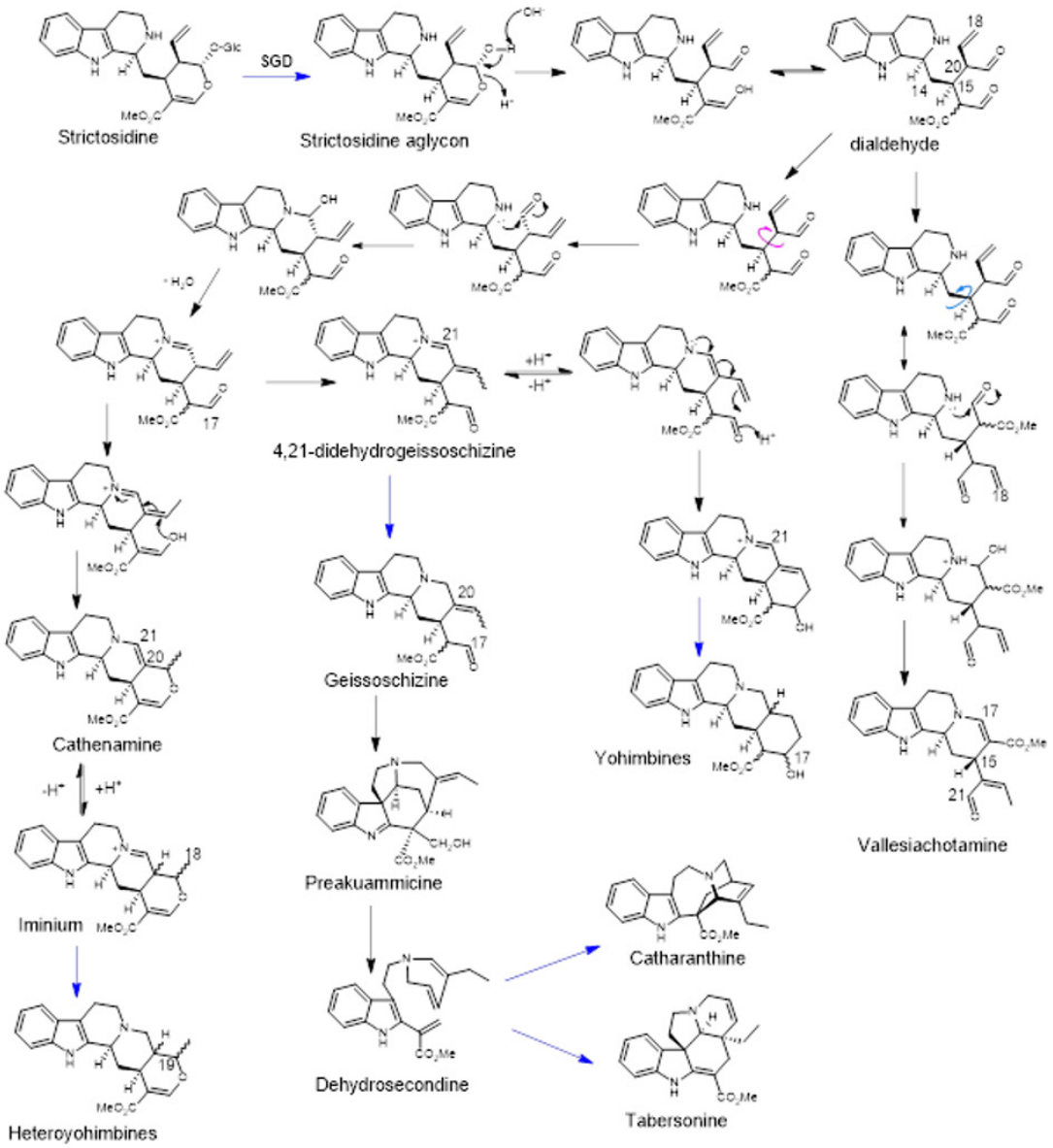
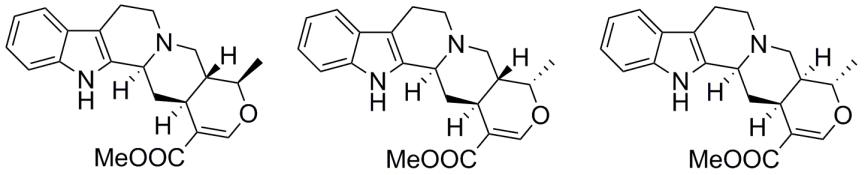


Figure 7: Strictosidine aglycon and its structural rearrangements. Steps illustrated with blue arrows must occur through enzymatic catalysis. Pink and blue arrows indicate the rotation of the molecule around the C-15 and C-20 bond and the C-14 and C-15 bond.

During experiments using cell-free crude enzyme assays Stöckigt et al. (1977) identified a compound that accumulated when tryptamine and secologanin were supplied as substrates. This compound was identified as 20,21-didehydroajmalicine which was given the trivial name cathenamine (fig. 7). The authors then went on to demonstrate that it is the direct precursor to the ajmalicine-type alkaloids, the heteroyohimbines (fig. 8). Cathenamine was also purified from leaves of *Guettarda eximia* (Rubiaceae) by Husson et al. (1977) and it was noted that the compound was not stable and was confirmed to be a precursor to the heteroyohimbine tetrahydroalstonine (THA) by reduction with $NaBH_4$. Therefore, it was clear the precursor molecule to the heteroyohimbines was an unstable intermediate.



19-epiajmalicine **Ajmalicine** **Tetrahydroalstonine**

Figure 8: Heteroyohimbine alkaloids of *C. roseus*. The alkaloids ajmalicine (AJM), 19-epiajmalicine (19-EA), and tetrahydroalstonine (THA) differ by the conformation at carbons 20 and 19.

Assays involving crude protein extracts from *C. roseus* (callus, leaves, and stems) indicated that soluble proteins were able to convert tryptamine and secologanin into primarily AJM and geissoschizine (Scott and Lee, 1975). Early studies involving reduction of cathenamine with *C. roseus* cell suspension culture crude cell protein extract indicated that some cell lines of *C. roseus* produced different ratios of the three heteroyohimbines (AJM:19-epiajmalicine (19-EA):THA) (Zenk, 1980). It was also demonstrated that the heteroyohimbine synthases are reductases which utilise NADPH to reduce the aglycon. The existence of some cell lines which could produce only one (or different ratios) of the three heteroyohimbines points to the existence, in *C. roseus*, of discreet reductases with different product stereospecificities.

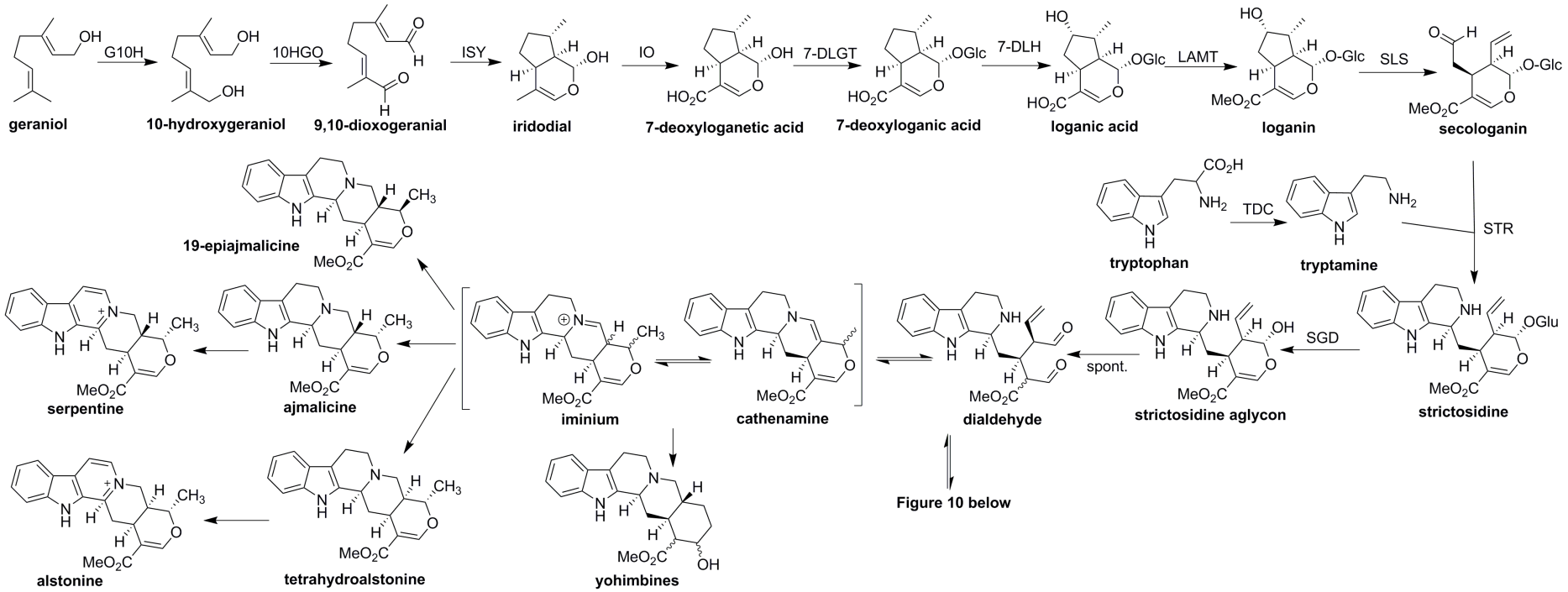


Figure 9: Pathway of *C. roseus* up to the heteroyohimbine alkaloids. The arrow from the dialdehyde leads to the figure below, where the downstream part of the pathway continues. The enzymes which were known are detailed. Enzymes: G10H: geraniol-10-hydroxylase; 10HGO: 10-hydroxygeraniol-oxidoreductase; ISY: iridoid synthase; IO: iridoid oxidase; 7-DLGT: 7-deoxyloganetic acid glucosyl transferase; 7-DLH: 7-deoxyloganic acid hydroxylase; LAMT: loganic acid O-methyltransferase; SLS: secologanin synthase; TDC: tryptophan decarboxylase; STR: strictosidine synthase; SGD: strictosidine glucosidase.

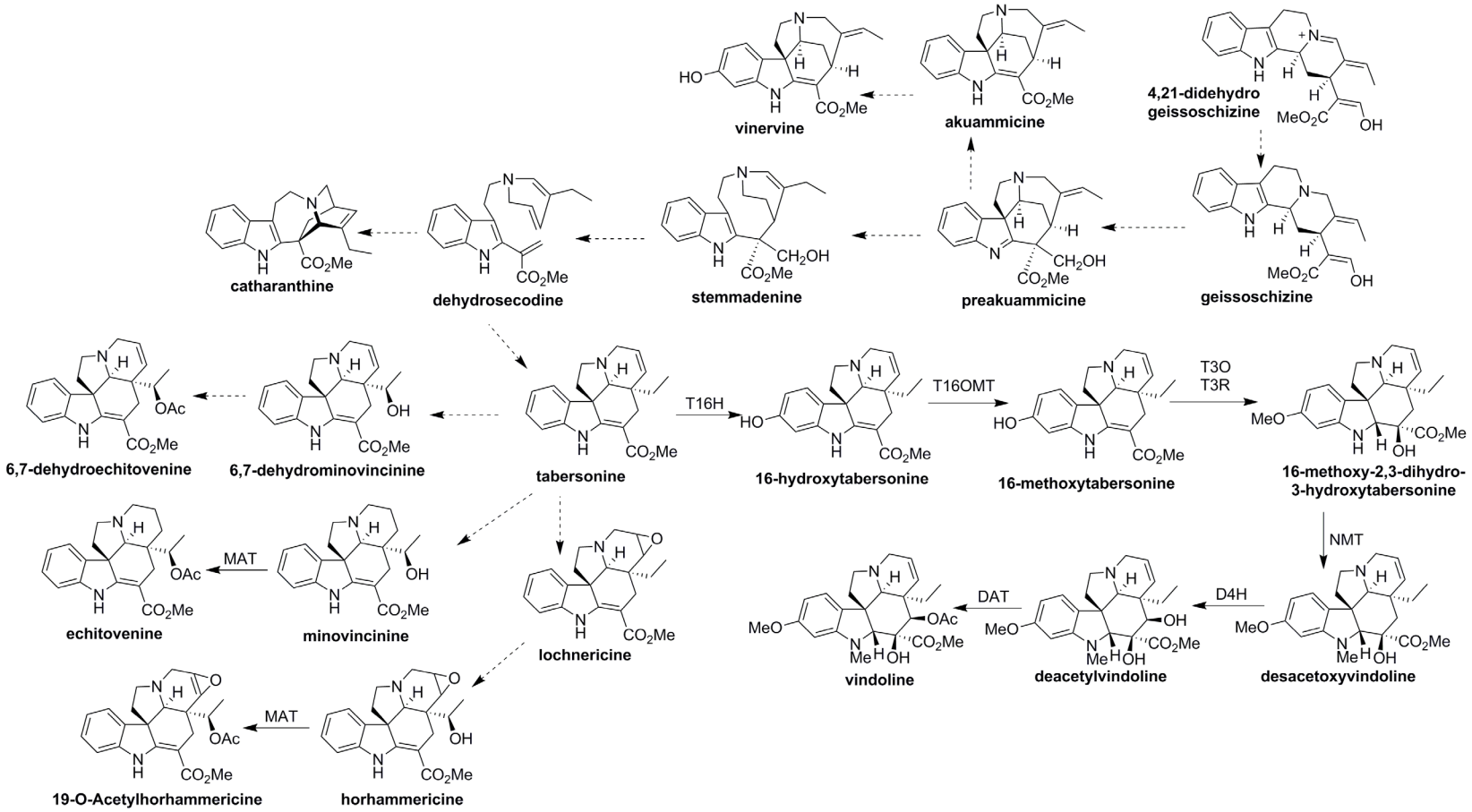


Figure 10: Proposed pathway of *C. roseus* from the strictosidine dialdehyde onwards. Figure continued from fig. 9. Dashed arrows represent steps for which biosynthetic enzymes have not been discovered. Vindoline and catharanthine are combined to form vinblastine. Enzymes: T16H: tabersonine-16-hydroxylase; T16OMT: 16-hydroxytabersonine-O-methyltransferase; T3O: tabersonine-3-oxygenase; T3R: tabersonine-3-reductase; NMT: N-methyltransferase; D4H: desacetoxyvindoline-4-hydroxylase; DAT: deacetylvindoline-4-O-acetlytransferase; MAT: minovincine-O-acetyltransferase.

1.7 Dehydrogenases

Dehydrogenases are able to remove two hydrogen atoms from one substrate and pass them onto another substrate, thus oxidizing the donor and reducing the acceptor (fig. 11)(Dewick, 2011, Persson et al., 2008). A pyridine nucleotide, nicotinamide adenine dinucleotide (NAD^+) or nicotinamide adenine dinucleotide phosphate ($NADP^+$) is often used as one substrate as cofactor. Usually NAD^+ is used in oxidation reactions and NADPH in reduction reactions. One of the hydrogens involved either originates from or is passed onto the cofactor, and the other hydrogen (a proton) is from the solvent or an acidic amino acid residue.

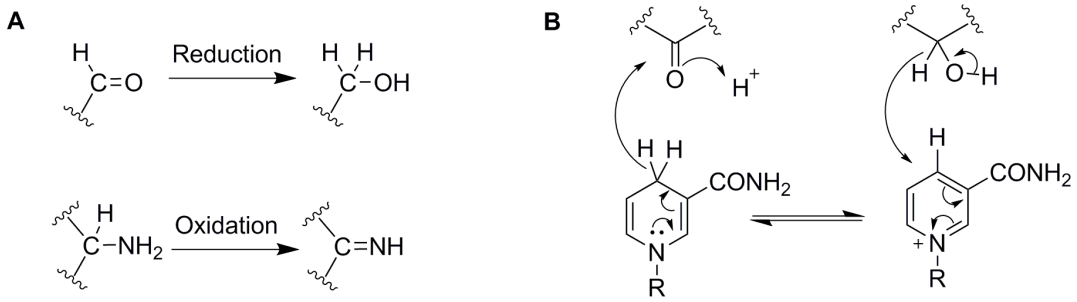


Figure 11: Oxidation and reduction reactions. A: general reaction scheme for reduction of a carbon-oxygen double bond (top row), and a general reaction scheme for oxidation of a carbon-nitrogen bond (bottom row). B: a general reaction scheme for oxidation-reduction reaction involving the cofactor NADP(H) with the R group representing the adenine dinucleotide phosphate moiety of the cofactor. Reactions redrawn from Dewick (2011).

The Rossmann protein fold is responsible for the NAD(P) cofactor binding in dehydrogenases (Rossmann et al., 1974). It is made up of a twisted parallel β-sheet core flanked on either side by α-helices. The Rossmann fold has been found to be one of the most common protein folds due to the functional versatility of reductase enzymes (Kavanagh et al., 2008, Kallberg and Persson, 2006). Enzymes with a Rossmann fold are present in every kingdom of life and carry out a large variety of oxidoreductions (Jörnvall et al., 2010). These enzymes appear to share a common ancestor which, after duplications and diversification over the course of evolution, still retain the same basic Rossmann fold but have evolved modified structures and extra domains (Jörnvall et al., 2010, Kavanagh et al., 2008, Riveros-Rosas et al., 2003). The major families within the Rossmann-fold enzyme superfamily are the Short-chain (SDR) and the Medium-chain dehydrogenase/reductases (MDR) (Kavanagh et al., 2008, Jörnvall et al., 1999).

1.8 The Medium-Chain Dehydrogenase/Reductase family

The MDR enzymes typically consist of a Rossmann fold/nucleotide binding domain, and a substrate binding domain. These enzymes usually are about 350 amino acids long which differentiate them from the SDRs which are usually about 250 amino acids long. The Alcohol Dehydrogenases (ADH) were first discovered from yeast and were the subject of many detailed kinetic analyses. The classic liver alcohol dehydrogenase can reduce or oxidise carbon-oxygen bonds of various alcohols (Dalziel and Dickinson, 1966). These enzymes usually catalyse the transfer of the pro-*R*-hydride of the NAD(P)H cofactor to the substrate (fig. 12).

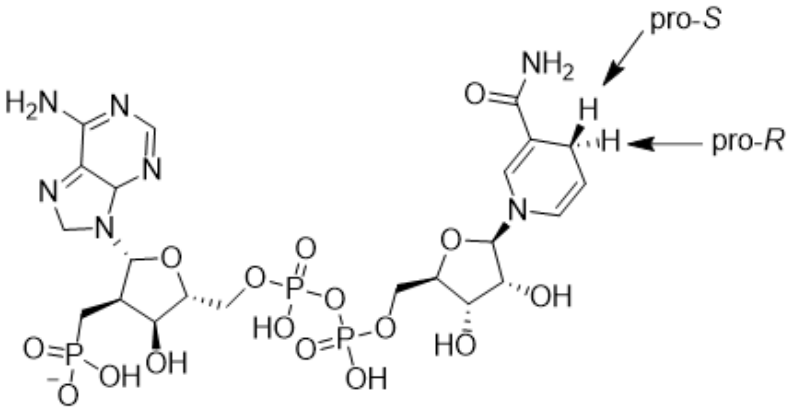


Figure 12: Reduced nicotinamide adenine dinucleotide phosphate (NADPH). The pro-*S* and pro-*R* hydrides are illustrated.

The MDRs in plants have undergone a large expansion, especially in the cinnamyl alcohol dehydrogenase (CAD) family (Riveros-Rosas et al., 2003). A plant-specific branch of the CADs has evolved in plants, the elicitor-inducible defense-related proteins (ELI3, Logemann et al. (1997)). These are specifically upregulated by pathogen elicitors and wounds (Riveros-Rosas et al., 2003).

The CAD enzyme group is one of the better studied group of plant MDRs. CAD and its closely related paralog Sinapyl alcohol dehydrogenase (SAD) have been purified and studied from many different plant species (Knight et al., 1992, O'Malley et al., 1992, Sarni et al., 1984, Wyrambik and Grisebach, 1979, Wyrambik and Grisebach, 1975, Mansell et al., 1974). These enzymes are of direct importance to the pulping and paper industry as well as forage crop agriculture as they are critical players in the lignification of plant tissues (fig. 13, (Halpin et al., 1994)).

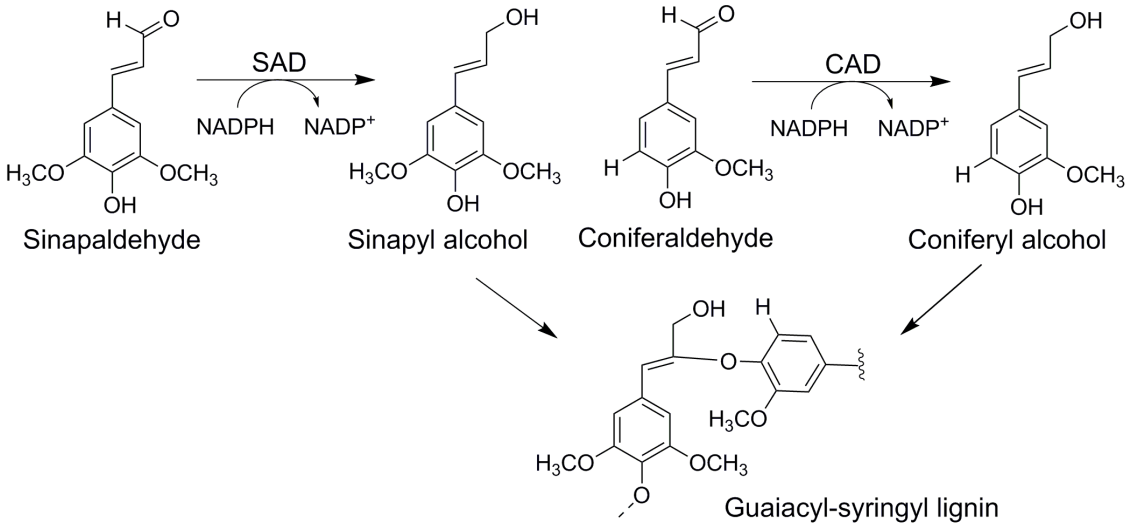


Figure 13: Reaction catalysed by Sinapyl and Cinnamyl alcohol dehydrogenases. The monolignol units coniferyl alcohol and sinapyl alcohol are polymerised to yield guaiacyl-syringyl lignin.

In *C. roseus* two MDRs have been discovered, outside of this thesis work, to participate in the MIA biosynthetic pathway. Upstream, the 10-hydroxygeraniol oxidoreductase (10HGO, (Miettinen et al., 2014)) oxidises both hydroxyl groups of 10-hydroxygeraniol to give rise to the dialdehyde 9,10-dioxogeranial (fig. 9). The second MDR member in the *C. roseus* MIA biosynthetic pathway

is the recently discovered tabersonine-3-reductase (T3R, Qu et al. (2015)) which acts in concert with a P450 enzyme (T3O) to hydroxylate a late-stage biosynthetic intermediate to produce 16-methoxy-2,3-dihydro-3-hydroxytabersonine (fig. 10).

A similar MDR was recently discovered in the related medicinal plant *Rauvolfia serpentina*, which also produces MIAs. This enzyme, vomilenine reductase (RsVR2, Geissler et al. (2015)) reduces the carbon-carbon double bond of the substrate vomilenine and produces 19,20-dihydrovomilenine (fig. 14). Based on homology to CAD and SAD enzymes, RsVR2 probably originated from evolution from a duplication of a CAD or SAD enzyme in *R. serpentina* (Geissler et al., 2015).

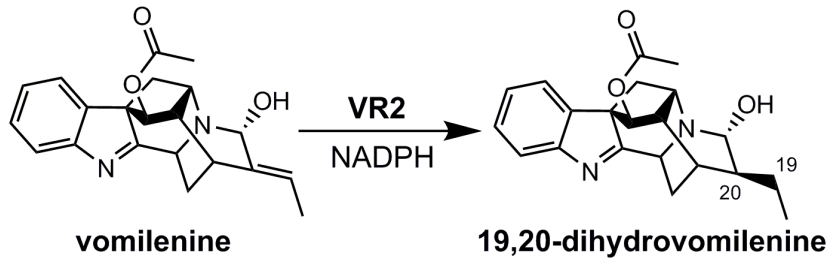


Figure 14: Reduction of vomilenine by RsVR2. Reduction occurs at the 19-20 double bond.

1.9 The Short-Chain Dehydrogenase/Reductase family

The SDRs are functionally and structurally very diverse (Jörnvall et al., 1995). One of the first SDRs to be characterised as such was the *Drosophila melanogaster* alcohol dehydrogenase (Winberg and McKinleymckee, 1994, Winberg and McKinleymckee, 1988) which was originally thought to be similar to liver alcohol dehydrogenase, an MDR. Soon though it was understood that the two enzymes differ structurally (SDRs do not contain Zinc ions as most MDRs) and they catalyse the transfer of the pro-*S* hydride of NAD(P)H to the substrate. SDRs are usually characterised by a conserved motif in the active site (YxxxK), which is part of the typical catalytic triad Ser-Tyr-Lys (Filling et al., 2002).

The SDRs have been reported to be present in very high numbers in the genomes of plants compared to animals, fungi, and prokaryotes (Moummou et al., 2012). It was found that in plants this family has undergone significant diversification, with some SDR subfamilies emerging after the separation of vascular plants from Bryophytes. Interestingly, the more diverged subfamilies are involved in the plant secondary metabolism or in developmental processes, and not so much in primary metabolism which claims more conserved SDRs (Moummou et al., 2012).

Many examples of SDRs in secondary plant metabolism exist. A well-known pathway is that of the benzylisoquinoline alkaloids (BIA), which are based on the amino acid tyrosine as opposed to tryptophan that is used for MIA production. The lengthy BIA pathway contains many branch pathways found in different plants. One SDR of great pharmaceutical importance is salutaridine reductase (SalR) from opium poppy (*Papaver somnifeum*, (Ziegler et al., 2006)) which acts upstream during the biosynthesis of morphine, a widely used analgesic (fig. 15 left). Another example of an SDR from BIA biosynthesis is sanguinarine reductase (SanR) from

the California poppy (*Eschscholzia californica*) (Weiss et al., 2006) which is implicated in the benzophenanthridine alkaloid branch of BIA (fig. 15 right).

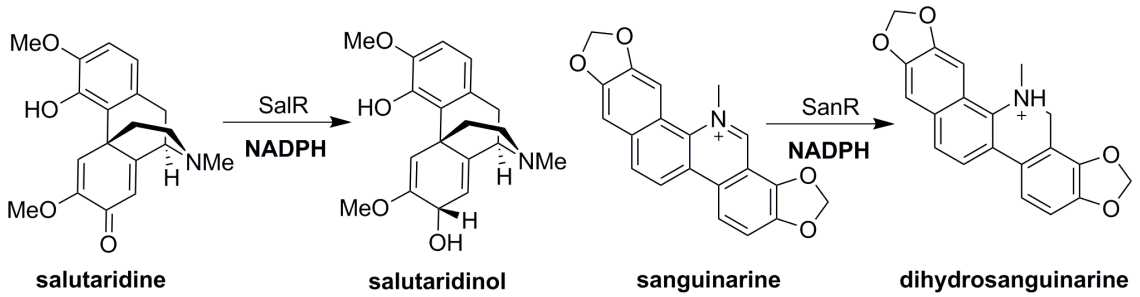


Figure 15: Reactions catalysed by SalR and SanR, SDR enzymes of the BIA pathway. The reduction carried out by SalR (left) and the reduction carried out by SanR (right).

In *C. roseus* one critical step of iridoid biosynthesis is known to be carried out by a member of the SDR family. The iridoid synthase (ISY, Geu-Flores et al. (2012)) reduces the substrate, 10-oxogeranial, which then undergoes cyclisation to yield the cyclised product nepetalactol (iridodial, fig. 9). This enzyme is critical for the formation of MIA as it produces the first cyclised intermediate in the pathway towards the monoterpene secologanin (fig. 9). It is expected that similar enzymes carrying out the same reaction exist in all plants which produce iridoids, and in fact a similar enzyme has been discovered in Olive (*Olea europaea*) (Alagna et al., 2016).

1.10 Stereoselectivity of Dehydrogenases/Reductases

Reductions carried out by dehydrogenases with NADPH are often stereoselective (Dewick, 2011). However, there exist examples of enzymes which can reduce two different isomers (substrate promiscuity) or produce two isomers (product promiscuity). An example of both can be found in the biosynthetic pathway of menthol. The eight-step biosynthetic pathway for menthol production from the well-studied and economically important plant *Mentha x piperita* contains multiple dehydrogenation steps (Croteau et al., 2005). Three of the steps are carried out by members of the SDR family, and one by a member of the MDR family (fig. 16). This last enzyme is not stereoselective and produces both (-)-menthone and (+)-isomenthone (fig. 16).

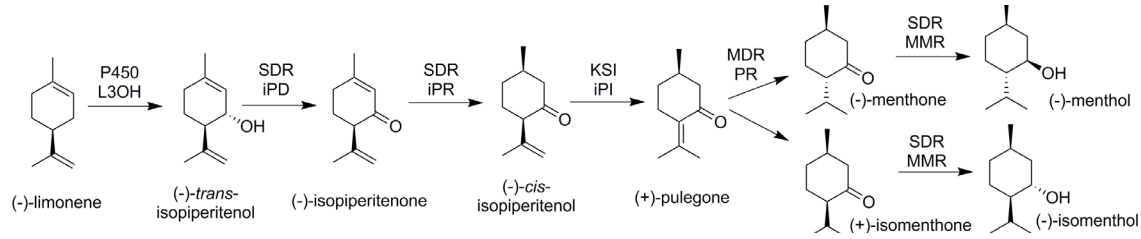


Figure 16: Menthol biosynthesis pathway from the cyclised terpene limonene. The enzymes responsible are indicated as well as the family each enzyme belongs to. In order they are: L3OH: limonene-3-hydroxylase (a P450 family member); iPD: *trans*-isopoperitenol dehydrogenase (an SDR); iPR: isopiperitenone reductase (SDR); iPI: isopiperitenone isomerase (a Ketosteroid Isomerase, KSI); PR: pulegone reductase (an MDR); MMR: menthone-to-menthol reductase (SDR).

Interestingly, the reduction carried out by iPR is stereoselective and produces only (+)-pulegone. The reduction mechanism and an explanation for this stereoselectivity were given by Lygidakis et al. (2016) after crystallisation of the enzyme in presence of the cofactor. The mechanism proposed includes an important amino acid in the active site of this enzyme which is an acidic glutamate rather than the usual tyrosine of the catalytic tetrad of SDRs. Lygidakis et al. (2016) argue that this causes the substrate to bind in a different manner which allows the alkene to be the acceptor of the cofactor hydride rather than the carbonyl (fig. 17). When the glutamate was mutated to a tyrosine iPR lost its isopiperitenone reductase ability and instead became a ketoreductase. This confirms the plasticity of the active site amino acids and demonstrates the chemical flexibility that is possible given even a single mutation to a biosynthetic enzyme.

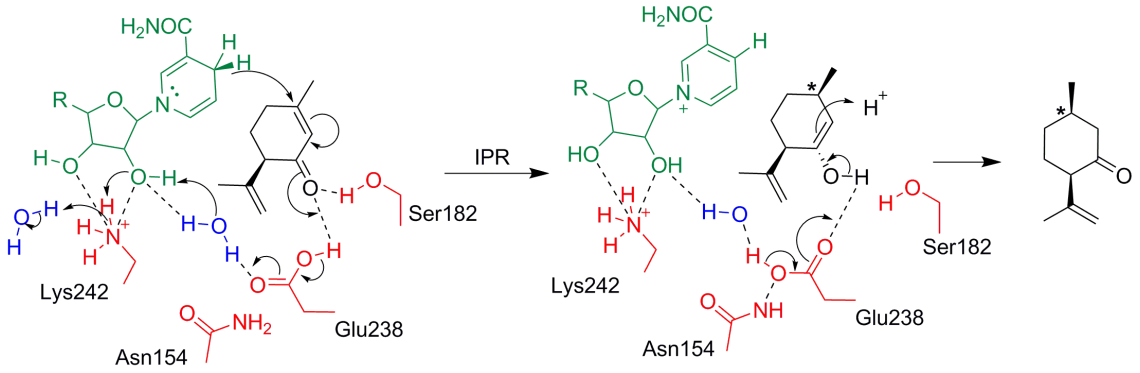


Figure 17: Stereoselective reduction of (–)-isopiperitenone to produce (+)-*cis*-isopulegone. Proposed reduction mechanism of Isopiperitenone Reductase (IPR), redrawn from (Lygidakis et al., 2016). The cofactor is illustrated in green, water molecules in blue, enzyme residues in red and the substrate and product in black. The generated stereocentre is indicated with an asterisk.

From a biocatalytic perspective, the stereoselectivity of these reductases is very interesting. Efforts have been made to produce enzymes with the opposite stereoselectivity for the product. Such optically-pure compounds can form the building blocks for further chemical synthesis. Mutations to the active site of a stereospecific MDR from *Candida parapsilosis* were attempted in order to switch the stereoselectivity (Wang et al., 2014). These mutations were able to cause the substrate (an aryl ketone) to flip orientation in the active site, resulting in a product with the opposite stereochemistry. Remarkably the product was changed from primarily (*R*) to mostly (*S*) (>99.9 enantioselectivity in both cases). This work highlighted how the variable active site residues, which might not be taking part in catalysis, can have a large effect on the binding orientation of the substrate and therefore the stereoselectivity of the enzyme.

The study of these biosynthetic enzymes can enrich our understanding of the enzymatic mechanisms governing the chemical diversity found in the natural world. This added knowledge can be applied to enzyme engineering to improve existing enzyme catalytic functions or for enzymatic synthesis of valuable compounds.

1.11 Scope of this thesis

The large chemical diversity found in the MIA biosynthetic pathways is generated from the unstable intermediate strictosidine aglycon. Despite many years of effort, and a reasonable understanding of the enzymatic steps involved in the pathway, no biosynthetic gene immediately following deglycosylation of strictosidine had been elucidated at the onset of this project. To elucidate the genes responsible for the enzymatic reactions, we relied on data generated previously by Zenk (1980) and the newly sequenced *C. roseus* transcriptome (Góngora-Castillo et al., 2012). A screen of candidate clones from *C. roseus* was carried out by functional expression in *E. coli* followed by *in vitro* assay, as described in Chapter 2. During the course of this thesis, a number of MDRs displaying activity against deglycosylated strictosidine were characterised in further detail, particularly to determine the mechanism of reduction. These enzymes have been named heteroyohimbine synthases.

As reported in Chapter 3, the discovery of the heteroyohimbine synthases with different product profiles provides a unique system to study the structure-function relationship controlling the product specificity among this family of enzymes. In addition, as described in Chapter 4, the discovery of an enzyme (ADH10) from a different family of reductases, the SDR family, with a very different product but a similar reduction mechanism has allowed us to probe the dynamics of the unstable strictosidine aglycon substrate. This comparison has given us an unprecedented first look into why the chemical diversity of the MIA is so large.

Finally, an in-depth comparison of the identified enzyme families (Chapter 5) indicated that the HYSs discovered are probably derived from primary metabolism enzymes of plants and more specifically from CAD enzymes. The origins of ADH10 however are murkier, in part because of the complexity of the SDR family in plants. Gene duplications appear to be the driving force behind the expansion of the chemical diversity as they provide the primary material for diversification and neofunctionalisation.

The discovery of the HYS and ADH10 in *C. roseus* constitutes the first discovery of enzymes acting directly downstream of SGD. This is an important step forwards in the elucidation of biosynthetic pathways in many plants which rely on strictosidine aglycon to produce the chemical diversity of their secondary metabolism. Understanding the dynamics of this complicated system at a more detailed level can provide insight to the evolution and the generation of chemical diversity, not only for plant but also for any organism with a rich secondary metabolism. The enzymes discovered here can give clues to the nature of the enzymes implicated in similar steps in other plants and can also expand the enzymatic toolbox for biocatalysis applications.

1.12 References

ALAGNA, F., GEU-FLORES, F., KRIES, H., PANARA, F., BALDONI, L., O'CONNOR, S. E. & OSBOURN, A. 2016. Identification and Characterization of the Iridoid Synthase Involved in Oleuropein Biosynthesis in Olive (Olea europaea) Fruits. *Journal of Biological Chemistry*, 291, 5542-5554.

BATTERSBY, A., BURNETT, A. & PARSONS, P. 1969. Alkaloid biosynthesis. Part XV. Partial synthesis and isolation of vincoside and isovincoside: biosynthesis of the three major classes of indole alkaloids from vincoside. Journal of the Chemical Society C: Organic, 1193-1200.

BROWN, R. T. & CHAPPLE, C. L. 1974. Biomimetic conversion of vincoside into heteroyohimbine alkaloids. *Journal of the Chemical Society, Chemical Communications*, 0, 740-742.

CHEN, S., GALAN, M. C., COLTHARP, C. & O'CONNOR, S. E. 2006. Redesign of a Central Enzyme in Alkaloid Biosynthesis. *Chemistry & Biology*, 13, 1137-1141.

COURDAVAULT, V., PAPON, N., CLASTRE, M., GIGLIOLI-GUIVARC'H, N., ST-PIERRE, B. & BURLAT, V. 2014. A look inside an alkaloid multisite plant: the Catharanthus logistics. *Current Opinion in Plant Biology*, 19, 43-50.

CROTEAU, R. B., DAVIS, E. M., RINGER, K. L. & WILDUNG, M. R. 2005. (−)-Menthol biosynthesis and molecular genetics. *Naturwissenschaften*, 92, 562-577.

CUSHNIE, T. T., CUSHNIE, B. & LAMB, A. J. 2014. Alkaloids: An overview of their antibacterial, antibiotic-enhancing and antivirulence activities. *International journal of antimicrobial agents*, 44, 377-386.

DALZIEL, K. & DICKINSON, F. M. 1966. The kinetics and mechanism of liver alcohol dehydrogenase with primary and secondary alcohols as substrates. *Biochem*. J., 100, 34-46.

DE LUCA, V., SALIM, V., THAMM, A., MASADA, S. A. & YU, F. 2014. Making iridoids/secoiridoids and monoterpenoid indole alkaloids: progress on pathway elucidation. *Current Opinion in Plant Biology*, 19, 35-42.

DEWICK, P. M. 2011. *Medicinal Natural Products: A Biosynthetic Approach*, John Wiley & Sons.

FILLING, C., BERNDT, K. D., BENACH, J., KNAPP, S., PROZOROVSKI, T., NORDLING, E., LADENSTEIN, R., JÖRNVALL, H. & OPPERMANN, U. 2002. Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *Journal of Biological Chemistry*, 277, 25677-25684.

GEISSLER, M., BURGHARD, M., VOLK, J., STANIEK, A. & WARZECHA, H. 2015. A novel cinnamyl alcohol dehydrogenase (CAD)-like reductase contributes to the structural diversity of monoterpenoid indole alkaloids in Rauvolfia. *Planta*, 1-12.

GEU-FLORES, F., SHERDEN, N. H., COURDAVAULT, V., BURLAT, V., GLENN, W. S., WU, C., NIMS, E., CUI, Y. & O'CONNOR, S. E. 2012. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature*, 492, 138-142.

GIAMOUSTARIS, A. & MITHEN, R. 1995. The effect of modifying the glucosinolate content of leaves of oilseed rape (Brassica napus ssp. oleifera) on its interaction with specialist and generalist pests. *Annals of Applied Biology*, 126, 347-363.

GLENN, W. S., NIMS, E. & O'CONNOR, S. E. 2011. Reengineering a Tryptophan Halogenase To Preferentially Chlorinate a Direct Alkaloid Precursor. *Journal of the American Chemical Society*, 133, 19346-19349.

GÓNGORA-CASTILLO, E., CHILDS, K. L., FEDEWA, G., HAMILTON, J. P., LISCOMBE, D. K., MAGALLANES-LUNDBACK, M., MANDADI, K. K., NIMS, E., RUNGUPHAN, W., VAILLANCOURT, B., VARBANOVA-HERDE, M., DELLAPENNA, D., MCKNIGHT, T. D., O'CONNOR, S. & BUELL, C. R. 2012. Development of Transcriptomic Resources for Interrogating the Biosynthesis of Monoterpene Indole Alkaloids in Medicinal Plant Species. *PLoS ONE*, 7, e52506.

GUIRIMAND, G., COURDAVAULT, V., LANOUE, A., MAHROUG, S., GUIHUR, A., BLANC, N., GIGLIOLI-GUIVARC'H, N., ST-PIERRE, B. & BURLAT, V. 2010. Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? *Bmc Plant Biology*, 10.

GUIRIMAND, G., GUIHUR, A., GINIS, O., POUTRAIN, P., HERICOURT, F., OUDIN, A., LANOUE, A., ST-PIERRE, B., BURLAT, V. & COURDAVAULT, V. 2011a. The subcellular organization of

strictosidine biosynthesis in *Catharanthus roseus* epidermis highlights several trans-tonoplast translocations of intermediate metabolites. *Febs Journal*, 278, 749-763.
GUIRIMAND, G., GUIHUR, A., POUTRAIN, P., HERICOURT, F., MAHROUG, S., ST-PIERRE, B., BURLAT, V. & COURDAVAULT, V. 2011b. Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *Journal of Plant Physiology*, 168, 549-557.
HALPIN, C., KNIGHT, M. E., FOXON, G. A., CAMPBELL, M. M., BOUDET, A. M., BOON, J. J., CHABBERT, B., TOLLIER, M.-T. & SCHUCH, W. 1994. Manipulation of lignin quality by downregulation of cinnamyl alcohol dehydrogenase. *The Plant Journal*, 6, 339-350.
HOPKINS, R. J., VAN DAM, N. M. & VAN LOON, J. J. 2009. Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annual review of entomology*, 54, 57-83.
HUSSON, H.-P., KAN-FAN, C., SÉVENET, T. & VIDAL, J.-P. 1977. Structure de la cathénamine intermédiaire clé de la biosynthèse des alcaloides indoliques. *Tetrahedron Letters*, 18, 1889-1891.
JÖRNVALL, H., HEDLUND, J., BERGMAN, T., OPPERMANN, U. & PERSSON, B. 2010. Superfamilies SDR and MDR: From early ancestry to present forms. Emergence of three lines, a Zn-metalloenzyme, and distinct variabilities. *Biochemical and Biophysical Research Communications*, 396, 125-130.
JÖRNVALL, H., HÖÖG, J.-O. & PERSSON, B. 1999. SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Letters*, 445, 261-264.
JÖRNVALL, H., PERSSON, B., KROOK, M., ATRIAN, S., GONZALEZ-DUARTE, R., JEFFERY, J. & GHOSH, D. 1995. Short-chain dehydrogenases/reductases (SDR). *Biochemistry*, 34, 6003-6013.
KALLBERG, Y. & PERSSON, B. 2006. Prediction of coenzyme specificity in dehydrogenases/reductases. *FEBS Journal*, 273, 1177-1184.
KAN-FAN, C. & HUSSON, H.-P. 1979. Isolation and biomimetic conversion of 4,21-dehydrogeissoschizine. *Journal of the Chemical Society, Chemical Communications*, 1015-1016.
KAVANAGH, K. L., JÖRNVALL, H., PERSSON, B. & OPPERMANN, U. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families: The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cellular and Molecular Life Sciences*, 65, 3895-3906.
KNIGHT, M. E., HALPIN, C. & SCHUCH, W. 1992. Identification and characterisation of cDNA clones encoding cinnamyl alcohol dehydrogenase from tobacco. *Plant Molecular Biology*, 19, 793-801.
KUTCHAN, T. M., HAMPP, N., LOTTSPEICH, F., BEYREUTHER, K. & ZENK, M. H. 1988. The CDNA clone for strictosidine synthase from Rauvolfia-serpentina - DNA-sequence determination and expression in Escherichia-coli. *Febs Letters*, 237, 40-44.
LAFLAMME, P., ST-PIERRE, B. & DE LUCA, V. 2001. Molecular and Biochemical Analysis of a Madagascar Periwinkle Root-Specific Minovincinine-19-Hydroxy-O-Acetyltransferase. *Plant Physiology*, 125, 189-198.
LOGEMANN, E., REINOLD, S., SOMSSICH, I. E. & HAHLBROCK, K. 1997. A novel type of pathogen defense-related cinnamyl alcohol dehydrogenase. *Biological chemistry*, 378, 909-914.
LORIS, E. A., PANJIKAR, S., RUPPERT, M., BARLEBEN, L., UNGER, M., SCHÜBEL, H. & STÖCKIGT, J. 2007. Structure-Based Engineering of Strictosidine Synthase: Auxiliary for Alkaloid Libraries. *Chemistry & Biology*, 14, 979-985.
LUIJENDIJK, T. J. C., VAN DER MEIJDEN, E. & VERPOORTE, R. 1996. Involvement of strictosidine as a defensive chemical in *Catharanthus roseus*. *Journal of Chemical Ecology*, 22, 1355-1366.
LYGIDAKIS, A., KARUPPIAH, V., HOEVEN, R., NÍ CHEALLAIGH, A., LEYS, D., GARDINER, J. M., TOOGOOD, H. S. & SCRUTTON, N. S. 2016. Pinpointing a Mechanistic Switch Between Ketoreduction and "Ene" Reduction in Short-Chain Dehydrogenases/Reductases. *Angewandte Chemie International Edition*, 55, 9596-9600.
MAHESH, B. & SATISH, S. 2008. Antimicrobial activity of some important medicinal plant against

plant and human pathogens. *World journal of agricultural sciences*, 4, 839-843.
MANSELL, R. L., GROSS, G. G., STÖCKIGT, J., FRANKE, H. & ZENK, M. H. 1974. Purification and properties of cinnamyl alcohol dehydrogenase from higher plants involved in lignin biosynthesis. *Phytochemistry*, 13, 2427-2435.
MCCOY, E., GALAN, M. C. & O'CONNOR, S. E. 2006. Substrate specificity of strictosidine synthase. *Bioorganic & Medicinal Chemistry Letters*, 16, 2475-2478.
MCCOY, E. & O'CONNOR, S. E. 2006. Directed biosynthesis of alkaloid analogs in the medicinal plant *Catharanthus roseus*. *Journal of the American Chemical Society*, 128, 14276-14277.
MIETTINEN, K., DONG, L., NAVROT, N., SCHNEIDER, T., BURLAT, V., POLLIER, J., WOITTIEZ, L., VAN DER KROL, S., LUGAN, R., ILC, T., VERPOORTE, R., OKSMAN-CALDENTEY, K.-M., MARTINOIA, E., BOUWMEESTER, H., GOOSSENS, A., MEMELINK, J. & WERCK-REICHHART, D. 2014. The seco-iridoid pathway from *Catharanthus roseus*. *Nature Communications*, 5, 3606.
MIZUKAMI, H., NORDLOV, H., LEE, S.-L. & SCOTT, A. I. 1979. Purification and properties of strictosidine synthetase (an enzyme condensing tryptamine and secologanin) from *Catharanthus roseus* cultured cells. *Biochemistry*, 18, 3760-3763.
MORANT, A. V., JORGENSEN, K., JORGENSEN, C., PAQUETTE, S. M., SANCHEZ-PEREZ, R., MOLLER, B. L. & BAK, S. 2008. beta-glucosidases as detonators of plant chemical defense. *Phytochemistry*, 69, 1795-1813.
MOUMMOU, H., KALLBERG, Y., TONFACK, L., PERSSON, B. & REST, B. 2012. The Plant Short-Chain Dehydrogenase (SDR) superfamily: genome-wide inventory and diversification patterns. *BMC Plant Biology*, 12, 1-17.
MURATA, J., ROEPKE, J., GORDON, H. & DE LUCA, V. 2008. The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell*, 20, 524-542.
O'CONNOR, S. E. & MARESH, J. J. 2006. Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Natural Product Reports*, 23, 532-547.
O'MALLEY, D. M., PORTER, S. & SEDEROFF, R. R. 1992. Purification, Characterization, and Cloning of Cinnamyl Alcohol Dehydrogenase in Loblolly Pine (Pinus taeda L.). *Plant Physiology*, 98, 1364-1371.
PERSSON, B., HEDLUND, J. & JÖRNVALL, H. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cellular and Molecular Life Sciences*, 65, 3879-3894.
QU, Y., EASSON, M. L. A. E., FROESE, J., SIMIONESCU, R., HUDLICKY, T. & DE LUCA, V. 2015. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proceedings of the National Academy of Sciences*, 112, 6224-6229.
RIVEROS-ROSAS, H., JULIÁN-SÁNCHEZ, A., VILLALOBOS-MOLINA, R., PARDO, J. P. & PIÑA, E. 2003. Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily. *European Journal of Biochemistry*, 270, 3309-3334.
ROSSMANN, M. G., MORAS, D. & OLSEN, K. W. 1974. Chemical and biological evolution of nucleotide-binding protein. *Nature*, 250, 194-199.
RUEFFER, M., NAGAKURA, N. & ZENK, M. H. 1978. Strictosidine, the common precursor for monoterpenoid indole alkaloids with 3 α and 3 β configuration. *Tetrahedron Letters*, 19, 1593-1596.
RUNGUPHAN, W. & O'CONNOR, S. E. 2009. Metabolic reprogramming of periwinkle plant culture. *Nat Chem Biol*, 5, 151-153.
SAKLANI, A. & KUTTY, S. K. 2008. Plant-derived compounds in clinical trials. *Drug discovery today*, 13, 161-171.
SARNI, F., GRAND, C. & BOUDET, A. M. 1984. Purification and properties of cinnamoyl-CoA reductase and cinnamyl alcohol dehydrogenase from poplar stems (Populus X euramericana). *European Journal of Biochemistry*, 139, 259-265.
SCOTT, A. I. & LEE, S.-L. 1975. Biosynthesis of the indole alkaloid. Cell-free system from *Catharanthus roseus*. *Journal of the American Chemical Society*, 97, 6906-6908.
SCOTT, A. I., REICHARDT, P. B., SLAYTOR, M. B. & SWEENY, J. G. 1971. Mechanisms of indole alkaloid

biosynthesis. Recognition of intermediacy and sequence by short-term incubation. *Bioorganic Chemistry*, 1, 157-173.

SMITH, G. N. 1968. Strictosidine: a key intermediate in the biogenesis of indole alkaloids. *Chemical Communications* (London), 912-914.

ST-PIERRE, B., VAZQUEZ-FLOTA, F. A. & DE LUCA, V. 1999. Multicellular compartmentation of *Catharanthus roseus* alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. *The Plant Cell*, 11, 887-900.

STÖCKIGT, J., ANTONCHICK, A. P., WU, F. & WALDMANN, H. 2011. The Pictet–Spengler Reaction in Nature and in Organic Chemistry. *Angewandte Chemie International Edition*, 50, 8538-8564.

STÖCKIGT, J., HUSSON, H., KAN-FAN, C. & ZENK, M. 1977. Cathenamine, a central intermediate in the cell free biosynthesis of ajmalicine and related indole alkaloids. *Journal of the Chemical Society, Chemical Communications*, 164-166.

STÖCKIGT, J. & ZENK, M. H. 1977. Isovincoside (strictosidine), the key intermediate in the enzymatic formation of indole alkaloids. *FEBS Letters*, 79, 233-237.

STRAUSS, S. Y., RUDGERS, J. A., LAU, J. A. & IRWIN, R. E. 2002. Direct and ecological costs of resistance to herbivory. *Trends in Ecology & Evolution*, 17, 278-285.

SUDŽUKOVIĆ, N., SCHINNERL, J. & BRECKER, L. 2016. Phytochemical meanings of tetrahydro-β-carboline moiety in strictosidine derivatives. *Bioorganic & Medicinal Chemistry*, 24, 588-595.

SUN, J. & PEEBLES, C. A. M. 2015. Engineering overexpression of ORCA3 and strictosidine glucosidase in *Catharanthus roseus* hairy roots increases alkaloid production. *Protoplasma*, 1-10.

SZABÓ, L. 2008. Rigorous Biogenetic Network for a Group of Indole Alkaloids Derived from Strictosidine. *Molecules*, 13, 1875.

TATTERSALL, D. B., BAK, S., JONES, P. R., OLSEN, C. E., NIELSEN, J. K., HANSEN, M. L., HØJ, P. B. & MØLLER, B. L. 2001. Resistance to an herbivore through engineered cyanogenic glucoside synthesis. *Science*, 293, 1826-1828.

TREIMER, J. F. & ZENK, M. H. 1979. Purification and Properties of Strictosidine Synthase, the Key Enzyme in Indole Alkaloid Formation. *European Journal of Biochemistry*, 101, 225-233.

VAN DER FITS, L. & MEMELINK, J. 2000. ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science*, 289, 295-297.

VAN DER HEIJDEN, R., JACOBS, D. I., SNOEIJER, W., HALLARED, D. & VERPOORTE, R. 2004. The Catharanthus alkaloids: Pharmacognosy and biotechnology. *Current Medicinal Chemistry*, 11, 607-628.

VÁZQUEZ-FLOTAAND, F. A. & DE LUCA, V. 1998. Jasmonate modulates development-and light-regulated alkaloid biosynthesis in *Catharanthus roseus*. *Phytochemistry*, 49, 395-402.

VERMA, P., MATHUR, A. K., SRIVASTAVA, A. & MATHUR, A. 2012. Emerging trends in research on spatial and temporal organization of terpenoid indole alkaloid pathway in *Catharanthus roseus*: a literature update. *Protoplasma*, 249, 255-268.

WANG, S. S., NIE, Y., XU, Y., ZHANG, R. Z., KO, T. P., HUANG, C. H., CHAN, H. C., GUO, R. T. & XIAO, R. 2014. Unconserved substrate-binding sites direct the stereoselectivity of medium-chain alcohol dehydrogenase. *Chemical Communications*, 50, 7770-7772.

WEISS, D., BAUMERT, A., VOGEL, M. & ROOS, W. 2006. Sanguinarine reductase, a key enzyme of benzophenanthridine detoxification. *Plant, Cell & Environment*, 29, 291-302.

WINBERG, J. O. & MCKINLEYMCKEE, J. S. 1988. THE ADHS ALLELOENZYME OF ALCOHOL-DEHYDROGENASE FROM DROSOPHILA-MELANOGASTER - VARIATION OF KINETIC-PARAMETERS WITH PH. *Biochemical Journal*, 255, 589-599.

WINBERG, J. O. & MCKINLEYMCKEE, J. S. 1994. DROSOPHILA-MELANOGASTER ALCOHOL-DEHYDROGENASE - PRODUCT-INHIBITION STUDIES. *Biochemical Journal*, 301, 901-909.

WINK, M. & SCHIMMER, O. 2010. Molecular modes of action of defensive secondary metabolites. *Functions and biotechnology of plant secondary metabolites*, 39, 21-161.

WYRAMBIK, D. & GRISEBACH, H. 1975. Purification and Properties of Isoenzymes of Cinnamyl-

Alcohol Dehydrogenase from Soybean-Cell-Suspension Cultures. *European Journal of Biochemistry*, 59, 9-15.

WYRAMBIK, D. & GRISEBACH, H. 1979. Enzymic Synthesis of Lignin Precursors. *European Journal of Biochemistry*, 97, 503-509.

YERKES, N., WU, J. X., MCCOY, E., GALAN, M. C., CHEN, S. & O'CONNOR, S. E. 2008. Substrate specificity and diastereoselectivity of strictosidine glucosidase, a key enzyme in monoterpene indole alkaloid biosynthesis. *Bioorganic & Medicinal Chemistry Letters*, 18, 3095-8.

ZENK, M. H. 1980. Enzymatic Synthesis of Ajmalicine and Related Indole Alkaloids. *Journal of Natural Products*, 43, 438-451.

ZIEGLER, J., VOIGTLANDER, S., SCHMIDT, J., KRAMELL, R., MIERSCH, O., AMMER, C., GESELL, A. & KUTCHAN, T. M. 2006. Comparative transcript and alkaloid profiling in Papaver species identifies a short chain dehydrogenase/reductase involved in morphine biosynthesis. *Plant Journal*, 48, 177-192.

Chapter 2

Discovery and characterisation of heteroyohimbine synthases of *Catharanthus roseus*

2.1 Introduction

2.1.1 The heteroyohimbines

The heteroyohimbine alkaloids belong to the group of corynanthean MIAs. They differ at the configuration of the hydrogens and methyl of carbons 3, 20, and 19 (fig. 18). Some members of this group display biological activity and have been used in the clinic. For example, 19-epiajmalicine (19-EA) can bind to benzodiazepine receptors in the brain of rats (Ai et al., 1997) and ajmalicine (AJM) is an α1-adrenergic receptor antagonist (Roquebert and Demichel, 1984, Li et al., 2004). Furthermore, alstonine, an oxidised form of tetrahydroalstonine (THA) displays anxiolytic and antipsychotic activity (Costa-Campos et al., 2004, Elisabetsky and Costa-Campos, 2006).

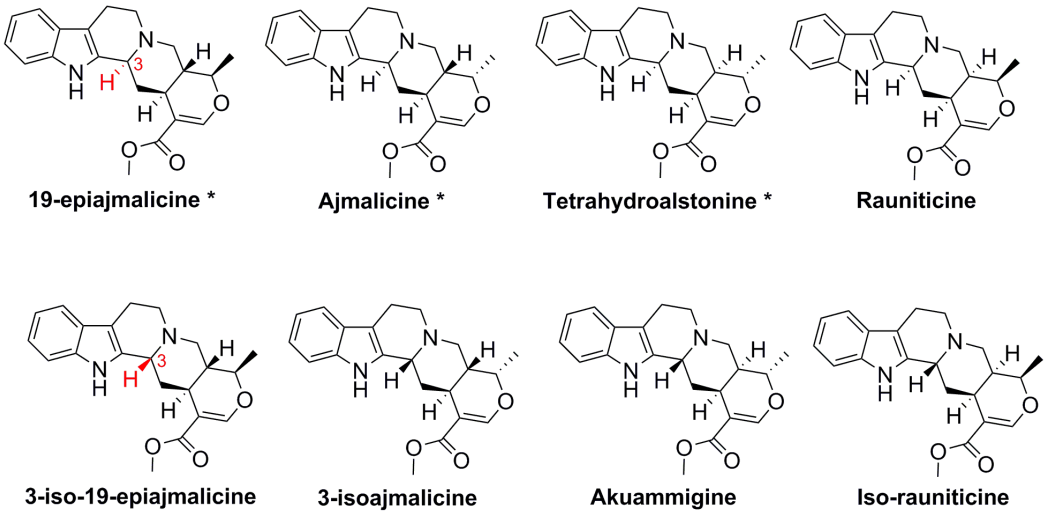


Figure 18: The eight heteroyohimbines. The top row (with (*S*) stereochemistry at C-3) are derived from strictosidine, and the bottom row (with (*R*) stereochemistry at C-3) are derived from the strictosidine isomer, vincoside. The three heteroyohimbines marked with an asterisk have been reported to be found in *C. roseus*. All molecules are stereoisomers and have the same predicted mass, 352.178.

2.1.2 Heteroyohimbine biosynthesis in *C. roseus*

C. roseus produces three heteroyohombine diastereomers,[1] AJM, 19-EA, and THA, with AJM being the most abundant. Heteroyohimbine alkaloids are derived from strictosidine after

1 The fourth strictosidine-derived heteroyohimbine, rauniticine, is mentioned to be found in *C. roseus* only by Stöckigt et al. (1976) but the authors do not provide a reference to the claim. Whole alkaloid extracts of the leaves or roots of *C. roseus* performed in our lab contain multiple peaks with *m/z* of 353, the largest of which co-elutes with an authentic standard of AJM. As no other reference can be found for the presence of this alkaloid in *C. roseus* other than the passing mention made by Stöckigt et al. (1976), and the absence of an authentic standard to compare to our plant extracts, for the remainder of this thesis I will assume it is not found in this plant.

it is deglycosylated by SGD. Studies done by Stöckigt et al. (1980) using deuterated solvent and cofactors illustrated how AJM, 19-EA, and THA can be formed directly from an isomer of strictosidine aglycon, dehydrogeissoschizine, when supplied to a *C. roseus* cell suspension culture protein extract. These workers illustrated that dehydrogeissoschizine can isomerise in solution into cathenamine and 19-epicathenamine, which can then both isomerise into the iminium tautomers (fig. 19). They proposed that these iminium forms are then acted upon by NADPH-dependent reductases which produce AJM, 19-EA, and THA (fig. 19).

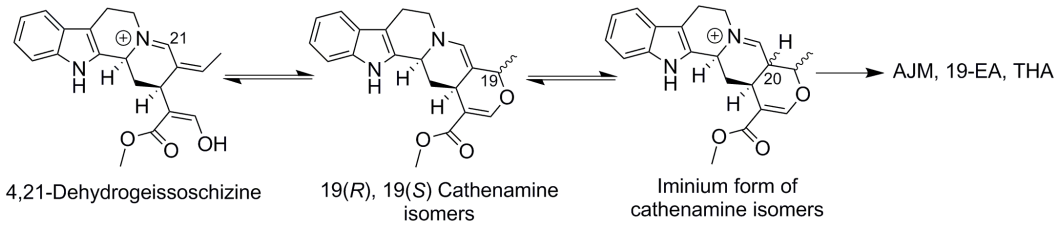


Figure 19: Proposed mechanism of biosynthesis of heteroyohimbines from dehydrogeissoschizine. Adapted from Stöckigt et al. (1980). Dehydrogeissoschizine equilibrates in solution to the two isomers 19(*R*)-cathenamine and 19(*S*)-cathenamine. This in turn can equilibrate into the iminium form by protonation at C-20.

During efforts to elucidate how the MIA pathway progresses through secologanin and tryptamine Stöckigt et al. (1976) showed that AJM, 19-EA, and THA can all be produced simultaneously *in vitro* using crude protein extracts of *C. roseus*. The authors prepared a crude protein extract from cell suspension cultures of *C. roseus* and incubated it with tryptamine, secologanin and various cofactors. The resulting alkaloid mixture was subjected to purification by TLC and it was shown that in presence of NADPH or NADH, three heteroyohimbines are formed, AJM, 19-EA, and THA. In absence of a suitable reduced cofactor another alkaloid accumulated, which if purified and incubated with the crude enzyme preparation and NADPH, was converted into AJM and 19-EA. The authors identified this compound as the intermediate in the reaction but could not identify its structure. Given the knowledge we now have about the pathway, the authors were generating strictosidine from secologanin and tryptamine with STR present in the crude protein mixture, and subsequently deglycosylating strictosidine with SGD also present in the crude enzyme preparation. When Stöckigt et al. (1976) repeated this *in vitro* experiment with [2-^{14}C]-labelled tryptamine they calculated that 26% of the labelled tryptamine was converted into these heteroyohimbines. Furthermore, they mention that although rauniticine (see footnote 1) and geissoschizine (fig. 20) are found in *C. roseus* plants, these compounds were not produced in these *in vitro* experiments.

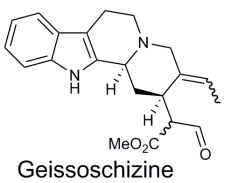


Figure 20: The MIA geissoschizine.

Later experiments attempted to purify the enzyme or enzymes responsible for these reactions. An enzyme displaying tetrahydroalstonine synthase activity was purified from cell suspension cultures of different Apocynaceae plants including *C. roseus*, *C. ovalis*, *Picralima nitida*, *Rhazya stricta* and *Vinca herbacea* (probably *Vinca minor*), all of which are known THA producers (Hemscheidt and Zenk, 1985). The *C. roseus* cell suspension culture that was used is an unpublished cell-line the authors claimed produced only THA and no other heteroyohimbine alkaloids. This fact indicates that there exist some enzymes which specifically produce one heteroyohimbine and not the others. The presence of tetrahydroalstonine synthase (THAS) activity was compared between these plants and also with enzymes purified from control tissues from *Beta vulgaris* (Chenopodiaceae), *Malus domestica* (Rosaceae), and *Solanum marginatum* (Solanaceae), plants that do not produce heteroyohimbine alkaloids. The authors found that the tissue of the cell suspension line of *C. roseus* had the most active enzyme (0.98 pkat/mg protein), and the second-best producer was *R. stricta* (0.64 pkat/mg protein). Different levels of activity were recorded for the enzymes purified from the heteroyohimbine-producing plants but no activity was recorded in the reactions using the control plants.

To characterise the enzyme purified from *C. roseus* the authors determined its molecular weight based on a gel filtration column against commercially available protein standards and found that it was 81 ± 2.4 KDa. They observed that the enzyme could only turnover the strictosidine aglycon isomer cathenamine when supplied with NADPH and not with NADH. Kinetic studies found that the enzyme had a K_m for cathenamine of 62 µM (pH 6.6, 30 ˚C). Hemscheidt and Zenk (1985) also purified similar enzymes from another cell suspension line of *C. roseus*. The purified enzymes from this line could turnover cathenamine and produce AJM and 19-EA, which led these researchers to hypothesise that there exist different enzymes in *Catharanthus* which are responsible for the production of different ratios of heteroyohimbines found in *C. roseus*.

Given this information, it is clear that multiple enzymes in *C. roseus* must be able to produce heteroyohimbines with different stereocentres. The heteroyohimbine biosynthetic enzymes appear to be soluble, NADPH-dependent reductases, and approximately 80 KDa. *C. roseus* expresses multiple NADPH-dependent reductases, both short-chain and medium-chain.

2.1.3 Known reduction steps in MIA biosynthesis

Medium-chain Dehydrogenase/Reductases (MDRs) are a type of Alcohol Dehydrogenase (ADH) enzyme. They are a very ancient and diverse protein superfamily which was already found in the last universal common ancestor of all life on earth (Riveros-Rosas et al., 2003). An in-depth study of the diversity and evolution of this superfamily (Riveros-Rosas et al., 2003) revealed that there are three macrofamilies of MDRs, each further split into separate Clusters of Orthologous Groups (COGs) (Tatusov et al., 1997).

According to macrofamily delimitation the authors suggested Cinnamyl Alcohol Dehydrogenases (CAD) and Elicitor-inducible defence-related proteins (ELI3) belong to the COG1064 along with Yeast-ADHs (Y-ADH) (Riveros-Rosas et al., 2003). CAD are found in plants, fungi, protista, and bacteria. Conversely, ELI3 are only found in plants. Y-ADHs are not found in plants, but can be found in fungi, animals, archaea, and bacteria. An unrooted tree produced by Riveros-Rosas et al. (2003) placed the CAD and ELI3 families as sister groups, and it seems apparent that the ELI3s evolved from an ancient duplication of a CAD specifically in plants.

The MIA biosynthetic pathways are predicted to contain numerous reduction steps. Recent work on the elucidation of these pathways has revealed roles for both short-chain dehydrogenases and medium-chain dehydrogenases. An MDR has been discovered to act upstream in the pathway, at the oxidation of 10-hydroxygeraniol to 10-oxogeranial. This involves a two-step oxidation of the substrate involving NAD^+ as cofactor. This reaction is catalysed by 10-hydroxygeraniol-oxidoreductase (10HGO, fig. 21, (Miettinen et al., 2014)).

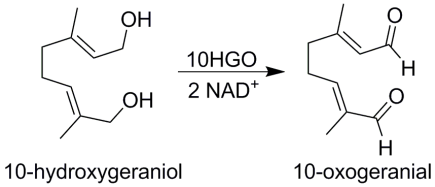


Figure 21: Reaction catalysed by 10-Hydroxygeraniol-Oxidoreductase (10HGO).

Additionally, an MDR from *C. roseus* was recently demonstrated to have activity further downstream at the tabersonine-to-vindoline branch of the vinblastine biosynthetic pathway (Qu et al., 2015). This enzyme, in concert with a cytochrome P450, is capable of reducing tabersonine into 3-hydroxy-2,3-dihydrotabersonine (fig. 22). This enzyme was named T3R and is referenced as such throughout this thesis.

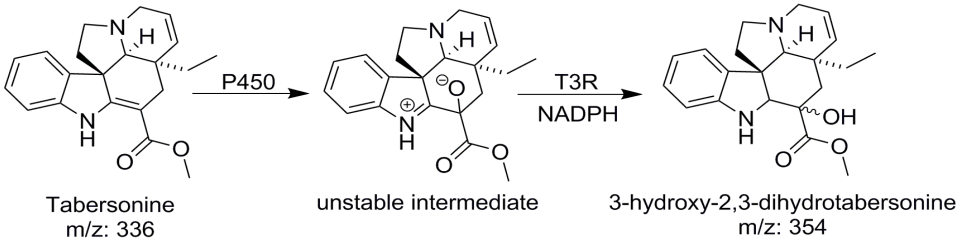


Figure 22: Conversion of the *C. roseus* MIA tabersonine into 3-hydroxy-2,3-dihydrotabersonine by the action of a cytochrome P450 followed by reduction by T3R.

2.1.4 Discovery of the heteroyohimbine synthases

Synthesis of heteroyohimbines also constitutes a reduction step in the MIA pathway, though at the start of this work, the identity of the enzymes that catalysed the formation of these structures was not known. In this chapter, I describe the discovery of the genes that encode heteroyohimbine synthases. Based on the preliminary work of Zenk and Stockigt described above, the most obvious candidate for the heteroyohimbine synthases would be an MDR. The transcriptome of *C. roseus* was indispensable for querying which MDRs could be responsible

for heteroyohimbine biosynthesis. During a screening of MDRs from *C. roseus* which are highly induced by application of MeJa at the seedling stage, an MDR was identified which could reduce cathenamine and produce primarily THA. This enzyme was named Tetrahydroalstonine Synthase 1 (THAS1). This initial success was further used to identify more candidates from the transcriptome based on sequence similarity to THAS1. This approach was successful and resulted in the identification of multiple heteroyohimbine synthases (HYSs), most of which are specific for THA production (THAS2-4), and one of which was capable of producing all 3 *C. roseus* heteroyohimbines (HYS). The enzymes discovered are described as hetetroyohimbine synthases (HYSs) to simplify discussion.

Discovery of these HYSs concludes the biosynthetic pathway from geraniol to the heteroyohimbines of *C. roseus*. This is the first report of genes that encode enzymes that act on the reactive intermediate strictosidine aglycon. These results provide insights into how the chemical diversity of the alkaloids is generated at this crucial branchpoint in MIA biosynthesis.

2.2 Results

2.2.1 Identification and sequence analysis of THAS1

Seedlings of *C. roseus* have been shown to actively synthesize the MIA whereas shoots, and in general older plant material, does not present such active biosynthesis (Scott et al., 1971). Therefore, expression profiles of biosynthetic enzymes at different plant growth stages are critical to the discovery of new enzymes taking part in this pathway. The transcriptomic dataset generated by Góngora-Castillo et al. (2012) was used as the basis for the work described in this chapter.

Hierarchical clustering analysis performed previously on this transcriptomic database (Geu-Flores et al., 2012) revealed multiple gene candidates that seemed to be co-regulated with other genes in the monoterpene indole alkaloid pathway. One of these genes (Cr024553), named THAS1 for reasons detailed later in this chapter, was of particular interest as it is physically present on the genome (WG contig v.2.A 371, Buell and Kim, in preparation) in close proximity to an SLS homolog (Cr024556). This gene is expressed in sterile seedlings and induced approximately two-fold at both five and twelve hours after the application of methyl-jasmonate (fig. 23). It is also expressed in immature and mature leaves (6.5 and 2.2 fpkm respectively) and also highly expressed in stem tissue (58.3 fpkm). Together, the expression profile and genome location pointed to a role in MIA biosynthesis.

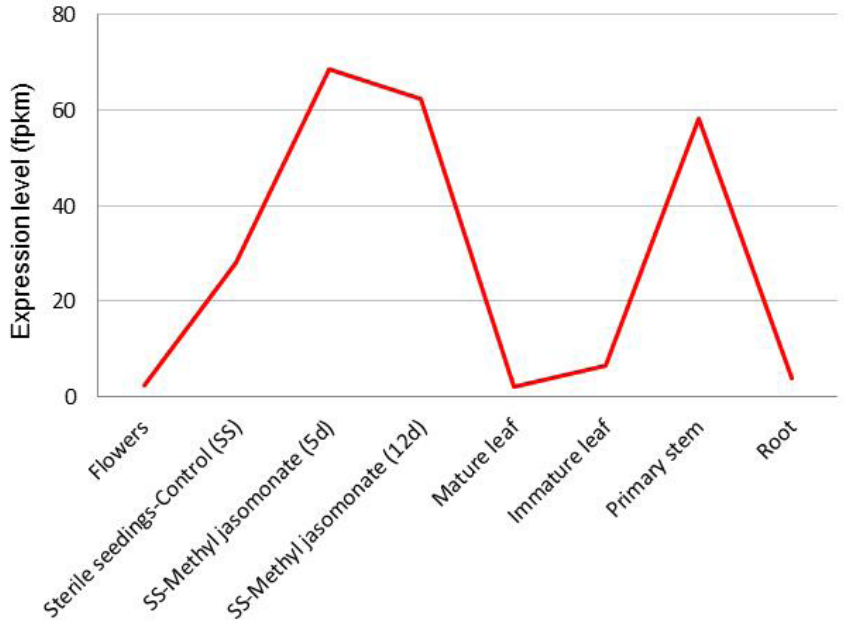


Figure 23: Expression profile of THAS1 in *C. roseus* tissues. The expression level (fpkm) is plotted against the various tissues including flowers, sterile seedlings, sterile seedlings 5 days after metyl-jasmonate application, sterile seedlings 12 days after MeJa application, mature and immature leaf, primary stem, and roots.

The gene cloned from cDNA was 1071 bp in length and coded for a member of the ADH family, specifically the Medium-chain Dehydrogenase/Reductase (MDR) subfamily (Appendix 1 for cDNA and protein sequences). When submitted to BLAST against the NCBI protein database the closest characterised homolog was a Sinapyl Alcohol Dehydrogenase (SAD (Bomati and Noel, 2005)) from the economically important tree species *Populus tremuloides* (PDB accession code: 1YQD). The two sequences have 64% identity at the amino acid level. THAS1 was also found to share 63% amino acid identity with the *C. roseus* biosynthetic enzyme Cr10HGO, which catalyses the conversion of 10-hydroxygeraniol to 10-oxogeranial early in the alkaloid pathway (fig. 9, NCBI accession code: Q6V4H0, (Miettinen et al., 2014)).

PtSAD, which is implicated in lignin biosynthesis (Li et al., 2001), catalyses the reduction of the aldehyde group on sinapaldehyde to an alcohol group, and has been crystallised to 2 Å resolution (Bomati and Noel, 2005). PtSAD has been shown to be active as a homodimer, utilises NADPH as cofactor, and binds two Zn^{2+} ions per subunit (Bomati and Noel, 2005). Initial characterisations of a similar enzyme CAD isolated from plant material indicated that zinc was necessary for catalysis (Wyrambik and Grisebach, 1979). Many MDRs contain two zinc ions, one coordinated in the enzyme active site (catalytic zinc) and one positioned near the surface of the protein (structural zinc).

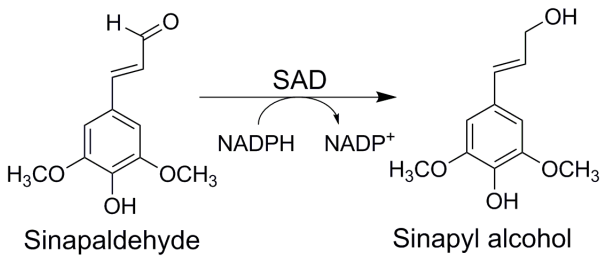


Figure 24: Reduction of sinapaldehyde to sinapyl alcohol, catalysed by Sinapyl Alcohol Dehydrogenase (SAD).

The protein product of the cloned THAS1 gene was predicted to be a dimeric zinc-binding metalloprotein, consistent with other MDRs which usually contain two zinc ions, one near the active site, termed “catalytic zinc” and one in a lobe at the surface of the enzyme, termed “structural zinc”. When analysed with zinc-prediction software (PredZinc (Shu et al., 2008), and ZincExplorer (Chen et al., 2013)) the residues C54, H76, C107 ,C110 , and C113 were predicted to form a zinc coordination sphere in both servers (table 1). Some predicted Zinc coordinating residues were not predicted in both servers.

Table 1: results of Zinc coordinating amino acids prediction of THAS1 residues using ZincExplorer online server. The score reflects the confidence of the prediction (most confident prediction gives a score of 1)

Zinc atom	Residue	Score
Zinc atom 1	Cysteine 54	0.889883
	Histidine 76	0.739041
	Cysteine 162	0.700843
Zinc atom 2	Cysteine 107	0.815173
	Cysteine 110	0.768885
	Cysteine 113	0.718252
	Cysteine 121	0.737481

These cysteines of THAS1 align to cysteines on the amino acid sequence of PtSAD. However, not all these cysteines take part in Zn^{2+} binding, as shown by the crystal structure of SAD (Bomati and Noel, 2005) in which it was shown that the catalytic Zn^{2+} binding residues are Cys50, His72, Cys166 and a water molecule which is coordinated with Ser52 (fig. 25). The crystal structure of PtSAD indicates that Ser52 is implicated in the catalysis of sinapaldehyde reduction and His55 is coordinated to the ribose of the cofactor (Bomati and Noel, 2005). Alignment of PtSAD with THAS1 indicates that a glutamate (Glu59) is equivalent to His55 of SAD and a tyrosine (Tyr56) is equivalent to Ser52. Tyrosine is not easily ionisable and glutamate has an opposite charge to histidine; this raises some interesting questions about the active site of THAS1 and the catalysis mechanism.

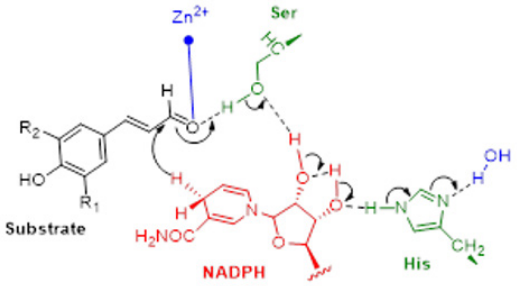


Figure 25: SAD reduction mechanism. Sinapaldehyde (substrate, in black) is reduced through hydride attack from the cofactor, NADPH (in red). The catalytic Zn^{2+} ion activates the carbon-oxygen bond. The active site serine provides the second hydrogen during the reduction and an electron transport chain is achieved through the hydrogen-binding network, via the cofactor ribose, and the coordinating histidine and finally into the bulk solvent.

Interestingly, the THAS1 protein sequence was found to contain a Class 5 nuclear localisation signal which is plant-specific (Kosugi et al., 2009). The amino acids 214-KKKR-217 are likely responsible for transport of this protein into the nucleus.

2.2.2 Initial functional characterisation of THAS1

The THAS1 gene was cloned into the *E. coli* IPTG-inducible expression vector pOPINF which contains an N-terminal His_6-tag for purification with a nickel column. Expression was better and higher protein yields were achieved when a richer medium was used, such as Terrific Broth or 2x YT media, and high amounts of protein (4 mg/L of culture) was achieved when IPTG between 1 mM and 0.01 mM were used. Concentration of the protein was possible up to approximately 20 mg/mL but beyond that it precipitated out of solution. The purified protein was aliquoted in 5 – 10 µL aliquots in sterile Eppendorf tubes. It was discovered that THAS1 lost activity after freezing and thawing and therefore was stored in small single-use aliquots. Inclusion of glycerol in the storage buffer did not improve the stability after thawing. Freezing of the aliquots in liquid nitrogen before storage at – 20 ˚C was necessary to retain activity; aliquots placed at – 20 ˚C without snap freezing were found to be inactive upon thawing. All purified ADH proteins reported in this thesis were snap frozen and stored in the same manner after discovering this fact.

Gel filtration chromatography revealed that in a typical preparation the enzyme was usually present in two isoforms, a monomer and a dimer (fig. 26, left panel). This has been observed for similar enzymes in the past (Wyrambik and Grisebach, 1975). Both isoforms were active (fig. 26, right panel), but only the fractions containing the dimer was stored for future use as the dimer forms of ADH are reported to keep their activity longer (Wyrambik and Grisebach, 1975).

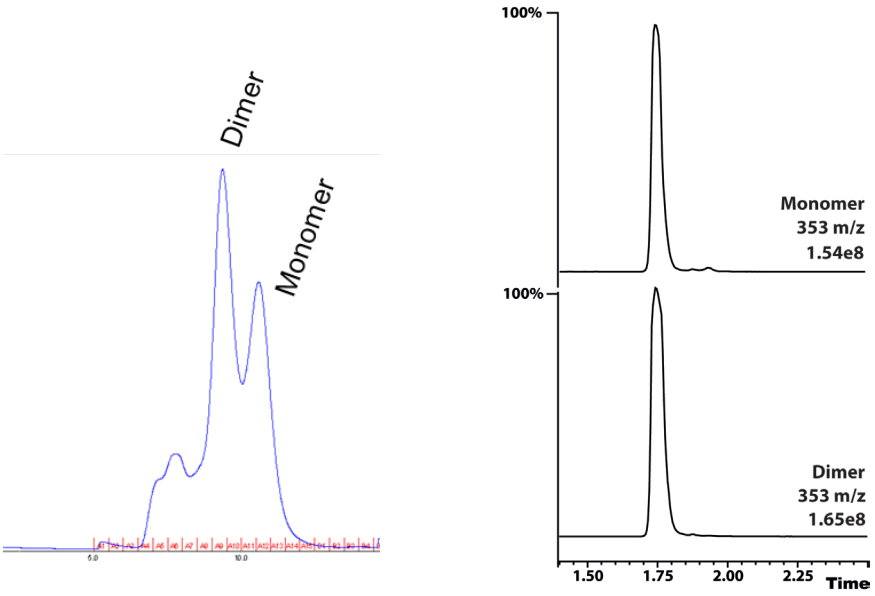


Figure 26: Gel filtration trace and LC-MS assays of THAS1 with dimer isozyme and monomer isozyme labelled. Left panel: Gel filtration trace of eluate from nickel column purification, the early bump constitutes high-molecular weight contaminants. The largest peak corresponds to the dimer form of THAS1 (approx. 65 KDa) and the last peak corresponds to the monomer form of THAS1 (approx. 42 KDa). Right panel: LC-MS chromatogram traces of the product of THAS1 assays in presence of strictosidine aglycon and NADPH.

Enzyme reactions were initially monitored by LC-MS. LC-MS measurements of the enzyme reactions were performed using targeted methods (Multiple Reaction Monitoring- MRM). Such targeted techniques allow better resolution of peaks as the detector selectively monitors the ions for the specified parent mass and also the specified fragmentation pattern. Thus each trace for the specified ion is free of all other contaminating ions. This method allows use of low concentrations of substrate and product and can also distinguish between ions which co-elute. Here, an MRM specifically developed from a heteroyohimbine authentic standard was used with reactions using purified strictosidine and also using a fresh enzymatic reaction of strictosidine aglycon.

Purified THAS1 was tested with deglycosylated strictosidine *in vitro*. The enzyme was able to reduce this substrate in the presence of the cofactor NADPH and produced a compound of mass 353 *m/z* (fig. 27 trace a). In absence of SGD, THAS1, substrate or reduced cofactor no product was formed (fig. 27, traces c to g). THAS1 was able to utilise NADH as well but produced lower amounts of product compared to the reaction utilising NADPH after 30 minutes incubation. The assays were also done in the presence or absence of the metal chelating agent ethylenediaminetetraacetic acid (EDTA, fig. 27, inset). There did not appear to be any significant reduction in activity of THAS1 with 5 mM of EDTA. While this suggests that metal is not required for THAS activity, the complete removal of the metal was not verified spectroscopically.

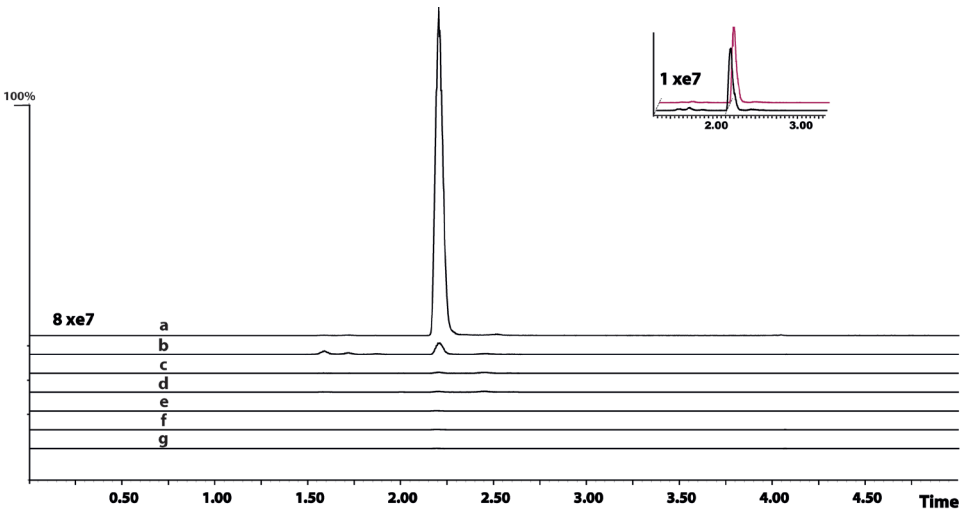


Figure 27: Chromatograms of THAS1 reaction along with controls. Enzyme reactions were performed at 25 ˚C for 30 minutes. Total ion chromatogram for heteroyohimbines *m/z* 353. Trace a: THAS1 (50 nM), SGD (6 nM), strictosidine (200 µM), NADPH (200 µM); b: NADPH replaced with NADH; c: boiled THAS1; d: no THAS1; e: no strictosidine; f: no SGD; g: no SGD, no THAS1. Inset: Assay with EDTA (black, 5 mM) and without EDTA (red).

THAS1 was also tested for activity in the reverse reaction, by supplying 300 µM THA as substrate together with $NADP^+$ as cofactor. This reaction did not yield significant amounts of cathenamine (fig. 28). The trace of 351 is of low intensity signal, and most probably corresponds to the normal variation in sensitivity of the MS detector or contamination of the authentic standard of THA. However, it is not possible to rule out the possibility that THAS1 does oxidise THA in the reverse reaction, but it is clear that the K_m is very low compared to its K_m for the normal reduction.

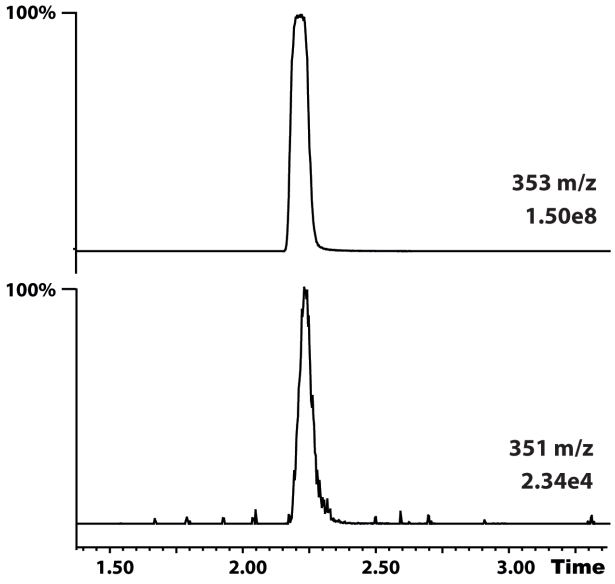


Figure 28: LC-MS trace of THAS1 reaction with THA and NADP+ as substrate. The substrate (top panel, 353 *m/z*), and the product trace (bottom panel, 351 *m/z*).

2.2.3 Characterisation of THAS1 product

There are multiple alkaloids reported to be found in *C. roseus* which have a mass of 353 *m/z* and it was not possible to determine the identity of the product based on mass alone. Authentic standards for AJM and THA were available, but unfortunately when tested using standard LC-MS conditions ("fast method" described in the Materials and Methods section below) these stereoisomers co-eluted. Even after development of LC conditions that could effectively separate these stereoisomers ("separation method"), as one of the potential heteroyohimbine diastereomers (rauniticine) is unavailable commercially, it is not possible to rule out the possibility it could co-elute with one of the others. To unambiguously identify which product(s) THAS1 produced, an NMR characterisation was necessary. A large-scale reaction using the purified enzyme and deglycosylated strictosidine was separated using silica thin layer chromatography (TLC) (fig. 29).

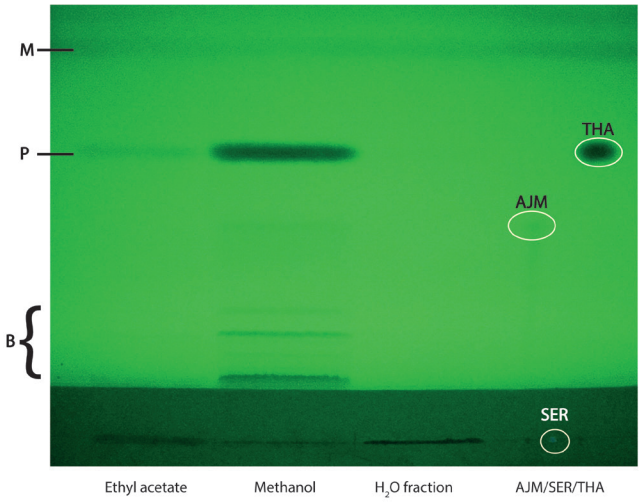


Figure 29: TLC of THAS1 large scale reaction. The loading areas of the various extracts are indicated on the figure; Ethyl acetate: the ethyl acetate extract of the water fraction after the MeOH extraction; Methanol: the methanol fraction; H_2O fraction: what was left over after both MeOH and ethyl acetate extraction. The authentic standards were loaded to the right of the plate in the order AJM, serpentine (SER), and THA. M indicates the migration front; P indicates the migration of the major product; B indicates the unknown by-products of the reaction.

Extraction of the dried reaction by MeOH appears to have extracted most constituents; only trace amounts of the major product and the lowest by-product can be seen in the lane with the ethyl acetate extract (fig. 29). The water fraction does not appear to contain any visible traces of heteroyohimbines. AJM and THA migrated and separated very well on silica and the THA standard migrated higher than the AJM and corresponded to the major product of the reaction. The serpentine authentic standard did not migrate in these conditions and appeared bluish under this UV illumination. There appears to be a slight band of product visible at the same R_f as AJM in the MeOH extract but ^1H NMR of this product was not of high enough signal to allow identification of that alkaloid. However, LC-MS measurement of the extract from that band confirmed that it is a heteroyohimbine (fig. 30). Other minor products were visible on

the TLC but unfortunately were not in high enough amounts for NMR. TLC purification of the major product yielded approximately 500 µg of material which was enough to obtain an NMR measurement in deuterated chloroform. Comparison of the proton shifts with published shifts of the heteroyohimbine alkaloids (Lounasmaa and Kan, 1980) indicated that the major purified product is THA (fig. 31).

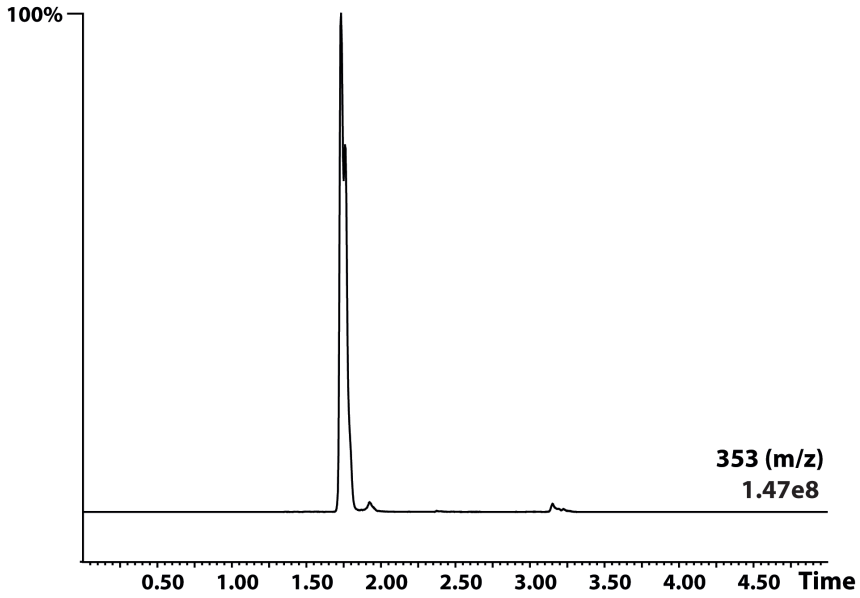


Figure 30: LC-MS trace of the TLC band corresponding to AJM. Targeted MRM trace of the heteroyohimbine isolated from the AJM band from the large-scale TLC purification of THAS1 reaction.

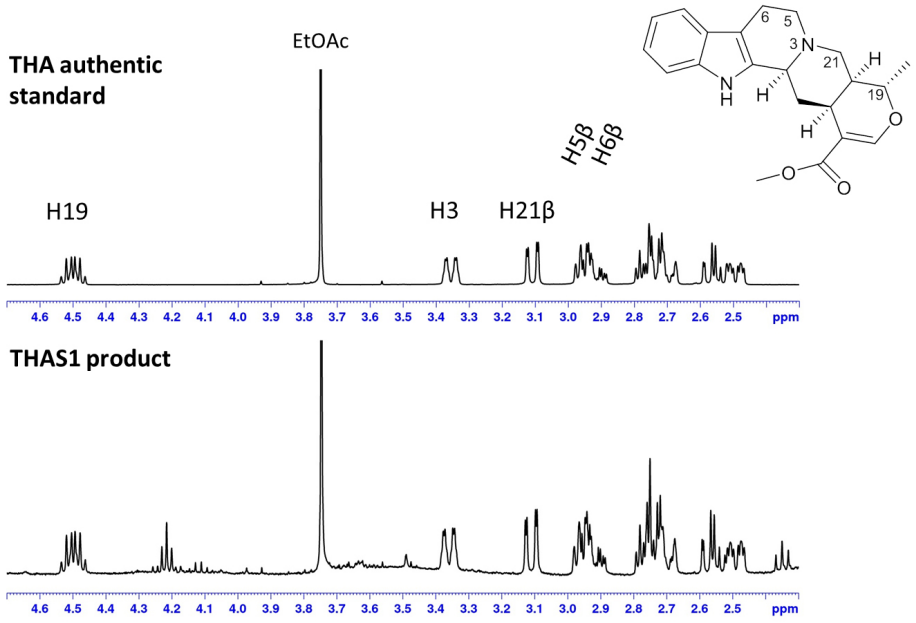


Figure 31: NMR traces of an authentic standard of THA and the purified major product of the large-scale THAS1 reaction. The structure of THA is included for reference.

2.2.4 Screening for similar enzymes with THAS activity

The discovery of an MDR from *C. roseus* which produces a heteroyohimbine alkaloid allowed a first positive hit from which to continue the search for other MDRs which could reduce deglycosylated strictosidine. The DNA sequence of THAS1 was BLASTed against the *C. roseus* transcriptome and resulted in only a few hits. However, BLASTing of the protein sequence with tblastn gave more results. Most hits had a predicted 45-55% identity with the THAS1 amino acid sequence, but there were some transcripts which showed higher identity, above 65%. The hits selected are Cr021691 (THAS2), Cr010119 (THAS3), Cr032583 (THAS4 and HYS, see section below), Cr017994, Cr021541 (T3R), and Cr027234 (amino acid sequence percent identities are displayed in table 2).

SAD from *P. tremuloides* was also BLASTed against the same database in order to provide negative controls as the hits from THAS1 BLAST search are biased towards high identity with THAS1. Various hits were retrieved when PtSAD was BLASTed and seven of them were picked due to the predicted presence of a serine or a threonine in the active site at the equivalent position to the catalytic serine of PtSAD. These hits selected were the transcripts Cr011702, Cr030442, Cr006840, Cr022770, Cr033537, Cr033062, and Cr2141 (amino acid sequence percent identities are displayed in table 2).

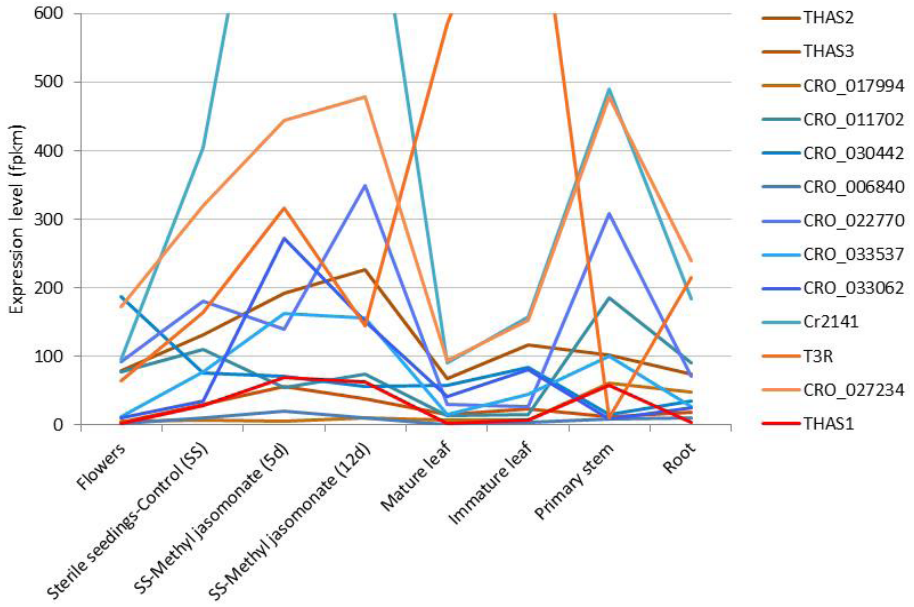


Figure 32: Expression level of the MDRs cloned in the initial HYS screen plotted against various tissues (HYS and THAS4 are not displayed, see below and Chapter 5 for discussion of this). The MDRs which were similar to CAD and SAD enzymes (i.e. with Ser or Thr in the active site) are illustrated in shades of blue. The MDRs with other residues in the predicted active site are illustrated in shades of orange and brown. THAS1 is displayed in red. For clarity the highest expression points (Cr2141 in the two MeJa application conditions, 1042 and 886 fpkm respectively, and T3R in immature leaf, 889 fpkm) are not displayed.

All the MDRs cloned display some expression in leaf tissues (fig. 32). By far the most highly expressed transcript is that of T3R, which contrary to almost all other transcripts is not expressed

in the stem. Cr2141 is the highest expressed MDR overall, including leaf tissue, where it is surpassed only by T3R. Cr2141 is also very highly induced by MeJa application at the seedling stage, which indicates it could be under the control of a MeJa-responsive promoter.

Sequencing of two Cr032583 (THAS4 and HYS) clones revealed that at least two very similar (96.2% identity) MDR genes were cloned. Interestingly, the differences were mostly present in the substrate binding domain of the enzyme, with the cofactor binding domain largely unchanged. The two clones have 14 synonymous (dS) substitutions between them and 20 non-synonymous (dN) substitutions. This results in a dN/dS ratio of 1.42 which suggests positive selection has been acting at this locus, however it is not possible to determine which gene the selection is acting on with only two sequences for comparison. These genes have likely arisen from a duplication, but which gene is experiencing the positive selection cannot be investigated without more information about the duplication, or sequencing of other examples of this duplication. The evolutionary origin of the MDR duplications will be discussed in detail in Chapter 5.

Alignment of the candidate MDR sequences revealed that most of each protein is similar in sequence (fig. 33). This is in agreement with the pairwise identities (table 2). However there appears to be a relatively large gap in the alignment that does not have identity in any of the proteins. The region between position 140 and 170 appears to have low identity in all these candidates. Most importantly there appear to be numerous indels in this section of the sequences. THAS2 and THAS3 have more amino acids in this section of the protein than all the others (24 and 21 more amino acids for THAS3 and THAS2 respectively compared to the THAS1 sequence). The sequence of candidate Cr017994 does not align well up to position 93 and this suggests the cloning of this candidate might not have been correct. This candidate also did not express when heterologously expressed in *E. coli*. The enzymes most similar to each other are THAS4 and HYS which share almost 94% identity at the amino acid level (fig. 33). The lowest identity (38.1%) is shared between Cr011702 and THAS3. THAS1 is most similar to HYS and THAS4 (76.3 and 75.7% respectively).

Table 2: Percentage of amino acid sequence identity between MDR candidates cloned. Heatplot illustrates the high identitieswith red and low identities with yellow.

	THAS1	THAS3	THAS2	THAS4	HYS	Cr017994	Cr011702	Cr030442	Cr006840	Cr022770	Cr033537	Cr033062	Cr2141	T3R	Cr027234
THAS1		52.6	54.3	75.7	76.3	52.7	44.3	46.3	54.3	58.6	53.7	54.0	58.3	67.4	54.9
THAS3	52.6		63.9	53.1	52.5	48.3	38.1	42.4	48.9	52.5	51.1	49.7	51.7	50.4	47.9
THAS2	54.3	63.9		56.3	56.6	48.8	39.5	44.1	49.1	53.6	51.3	50.8	53.1	51.8	50.5
THAS4	75.7	53.1	56.3		93.9	52.9	43.8	45.3	51.7	55.4	51.9	53.0	57.3	65.6	53.6
HYS	76.3	52.5	56.6	93.9		53.4	43.8	45.3	53.3	55.4	51.6	53.8	57.6	65.4	53.0
Cr017994	52.7	48.3	48.8	52.9	53.4		46.0	48.4	60.3	56.3	51.5	49.6	59.3	50.1	48.2
Cr011702	44.3	38.1	39.5	43.8	43.8	46.0		43.4	50.4	46.3	43.1	45.0	47.6	43.7	39.7
Cr030442	46.3	42.4	44.1	45.3	45.3	48.4	43.4		55.6	45.3	44.7	43.3	53.8	45.9	44.1
Cr006840	54.3	48.9	49.1	51.7	53.3	60.3	50.4	55.6		58.3	56.0	53.5	62.5	48.6	50.4
Cr022770	58.6	52.5	53.6	55.4	55.4	56.3	46.3	45.3	58.3		65.8	63.0	63.2	54.6	53.2
Cr033537	53.7	51.1	51.3	51.9	51.6	51.5	43.1	44.7	56.0	65.8		75.9	60.8	48.9	47.3
Cr033062	54.0	49.7	50.8	53.0	53.8	49.6	45.0	43.3	53.5	63.0	75.9		58.3	50.0	47.5
Cr2141	58.3	51.7	53.1	57.3	57.6	59.3	47.6	53.8	62.5	63.2	60.8	58.3		49.6	52.2
T3R	67.4	50.4	51.8	65.6	65.4	50.1	43.7	45.9	48.6	54.6	48.9	50.0	49.6		47.4
Cr027234	54.9	47.9	50.5	53.6	53.0	48.2	39.7	44.1	50.4	53.2	47.3	47.5	52.2	47.4	

39

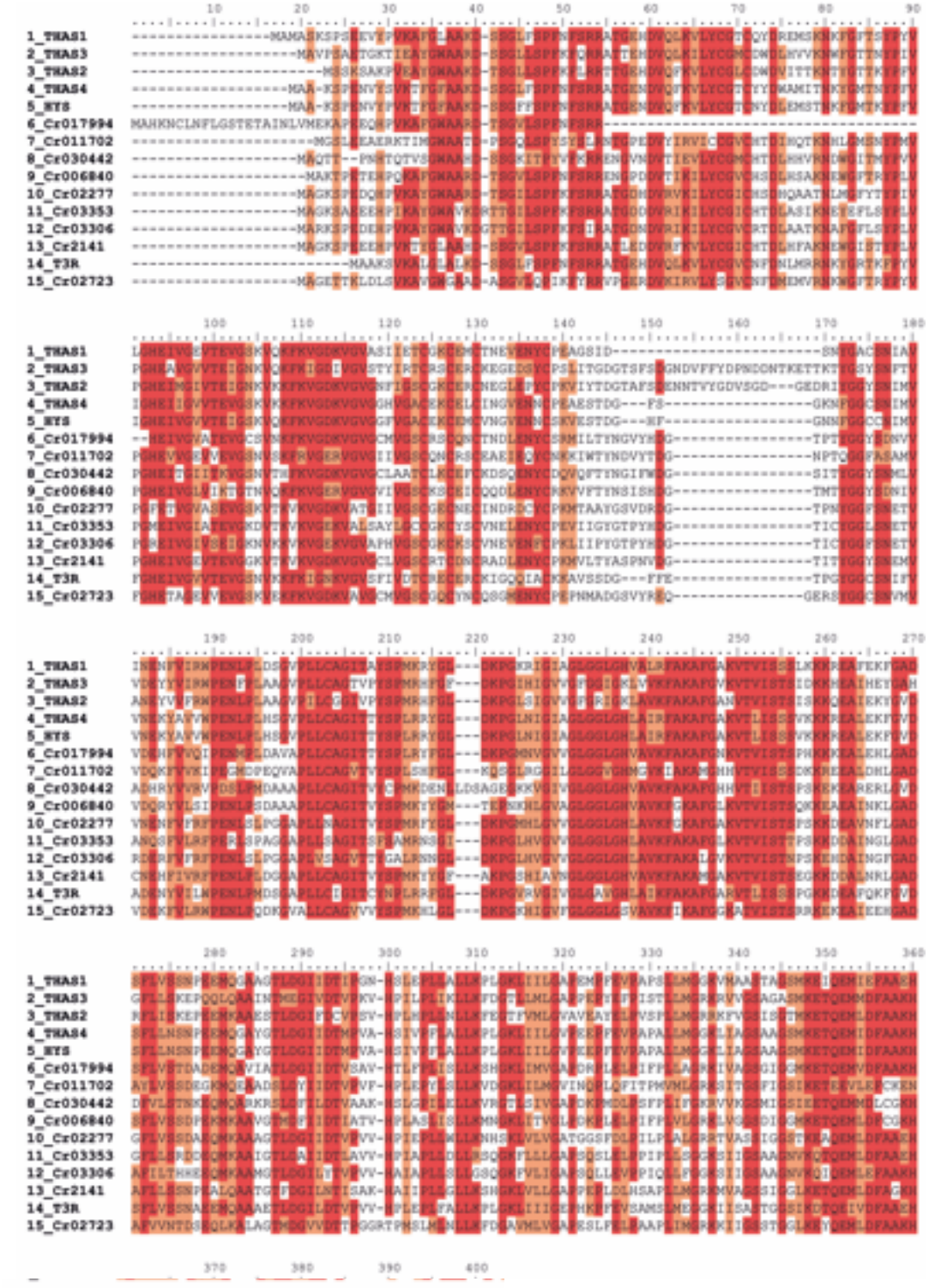


Figure 33: Protein alignment of candidate MDRs. Identical sites are coloured red and sites with more than 65% identity are coloured orange. The predicted active site equivalent amino acid to the SAD/CAD serine or threonine is at position 73 and is not conserved among these enzymes. The Rossman fold GXGXXG motif is conserved among all the candidates and starts at position 231.

After heterologous expression in *E. coli*, all candidates were tested in triplicate for production of heteroyohimbines. Cr017994 could not be expressed in *E. coli* and was not tested. Quantification of products was done using the "separation method" by LC-MS (fig. 34). THAS2, 3, and 4 all produce primarily THA, with THAS2 being the most promiscuous of all the THAS. HYS produced the three heteroyohimbines and is named Heteroyohimbine Synthase to reflect this promiscuity.

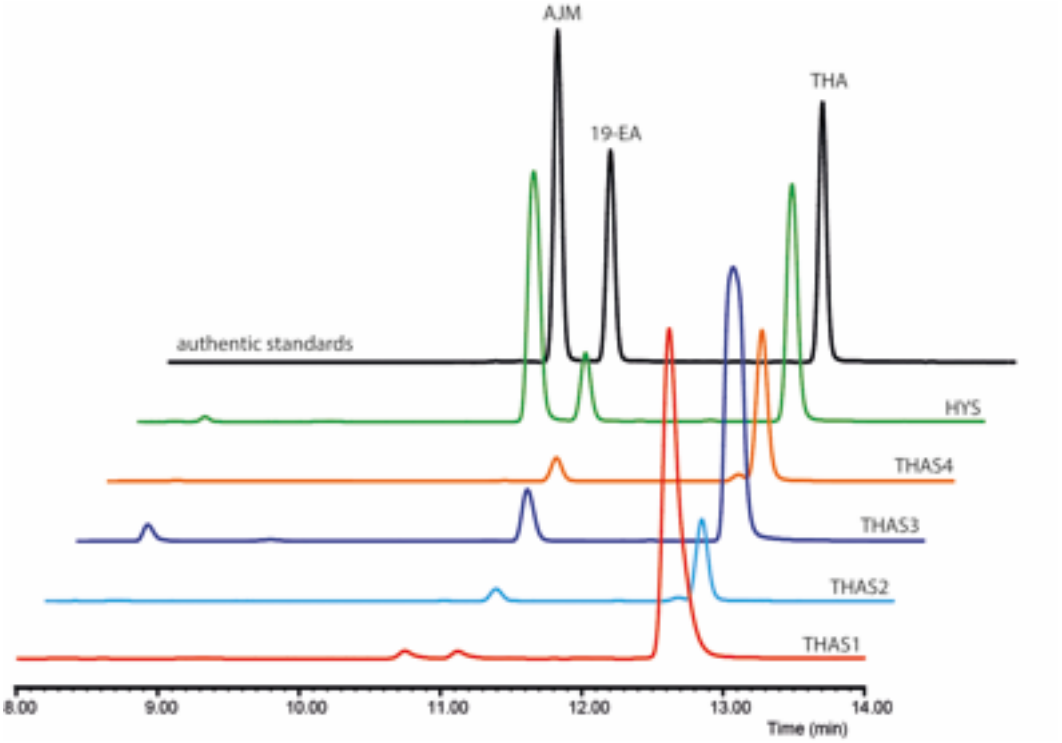


Figure 34: LC-MS traces of product formed after reaction of strictosidine aglycon with the HYSs. The trace in black is a mixture of the authentic standards for AJM, 19-EA, and THA. THAS1 is illustrated in red.

All other MDRs tested converted less than 1% of the initial strictosidine substrate into a heteroyohimbine, with Cr033062 being the best producer among them with 0.45% conversion. This enzyme is not very specific but the extremely low activity strongly suggests it is not a HYS. T3R and Cr033537 were initially chosen because they lack the serine or threonine active site amino acid that the classic SAD/CAD enzymes have (position 73 in alignment fig. 33), but did not show any significant HYS activity (table 3). A detailed mutational study of the active site will be described in Chapter 3.

Table 3: Results of enzyme assay analysis by LC-MS of MDRs that were screened for heteroyohimbine synthase activity. The total conversion refers to the yield compared to the starting concentration of strictosidine aglycon. The measured concentrations of the three identified heteroroyohimbines (AJM, 19-EA, THA) define the product ratio; values are rounded up for clarity. The conversion of all enzymes was compared to conversion of THAS1.

		Product ratio			Conversion
MDRs	Total conversion	Ajmalicine	19-epiajmalicine	Tetrahydroalstonine	compared to THAS1
THAS1	19%	2%	2%	97%	100%
THAS3	23%	0.5%	12%	88%	123%
THAS2	19%	2%	12%	87%	29%
THAS4	11%	1%	10%	89%	57%
HYS	22%	35%	15%	50%	115%
Cr011072	0%	0%	0%	100%	0%
Cr030442	0%	0%	0%	100%	0%
Cr006840	0%	0%	0%	100%	0%
Cr022770	0%	0%	0%	100%	0%
Cr033537	0%	0%	0%	100%	0%
Cr033062	0.5%	17%	5%	78%	2%
Cr2141	0%	0%	0%	100%	0%
T3R	0%	0%	0%	100%	0%
Cr027234	0%	0%	0%	100%	0%

2.2.5 pH and temperature optima

Analysis of the pH optimum of THAS1 was done by calculating the rate of the reaction by taking three timepoints for a single pH into account. All regressions had a high correlation, the minimum was R^2=0.97. The rate of each replicate was averaged and the standard deviation calculated. Results are displayed in figure 35.

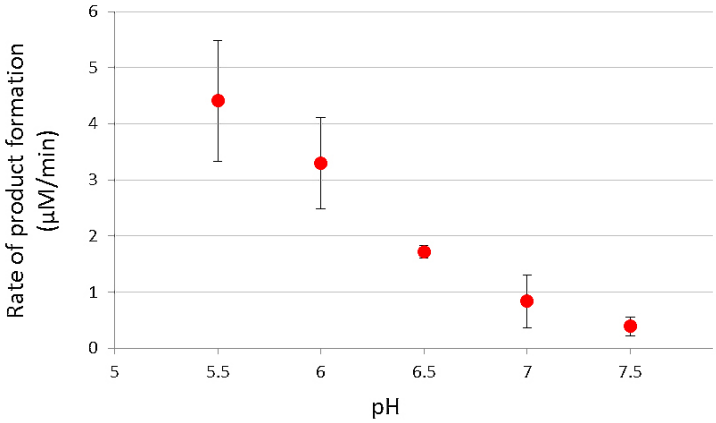


Figure 35: pH optimum of THAS1. The rate of product formation of THA (red) is plotted against the pH of the reaction. Error bars represent the standard deviation of the triplicate.

It appears the rate of product formation increases with decreasing pH, and THAS1 performed better at the lowest pH assayed, pH 5.5. The average of pH 5.5 is higher than that at pH 6.0

but the difference is not statistically significant (Student's one-tailed t-test t-statistic = 1.35). However there was a significant difference between the rate at pH 5.5 and at 6.5 (Student's one-tailed t-test t-statistic = 3.52).

The temperature optimum was assayed by varying the temperature of the assay by increments of 5 ˚C. The rate of reaction was calculated and plotted for the two replicates; results are displayed in figure 36. The data suggest that higher temperatures increase the rate of THA formation for THAS1. After these first results were obtained (fig. 36 left panel) another assay was done with a temperature range between 40 and 65 ˚C but unfortunately the variation among the replicates did not allow for any meaningful analysis of the data. Nevertheless, the temperatures above 55 ˚C showed low product formation at 3 minutes (fig. 36, right panel). Product formation seems to not be possible above 55 ˚C, which indicates the enzyme is not stable above that temperature.

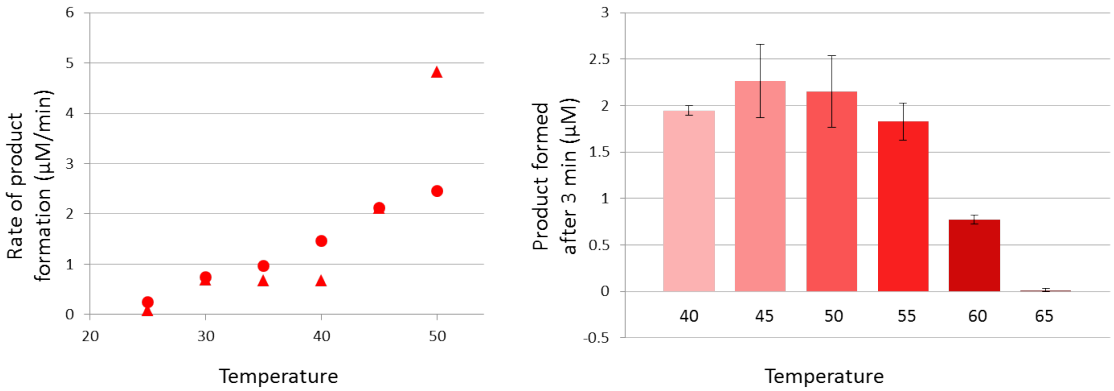


Figure 36: Temperature optimum of THAS1. Left panel: Rate of product formation is plotted according to temperature, triangles and dots represent two replicates. Right panel: Mean product formed in assay conditions after 3 minutes incubation at each temperature (mean of 3 replicates). Error bars represent the standard deviation.

The product formation with HYS at different pH values was assessed after running the samples on the LC-MS using the "fast method". The reaction in HEPES pH 7.5 produced a large amount of product at timepoint 1 minute. HYS appears to be very active and therefore comparison between the 1 and the 2-minute timepoints was not meaningful. This was exaggerated by the appearance of double peaks, which are perhaps the stereoisomers presenting some separation in some of the runs. Integration of the response area was done for both peaks where they were separated and the values were combined (table 4). It appears HEPES buffer at pH 7.5 resulted in the largest amount of product formed. It is not clear why the response area for the 2 minute timepoints is lower than that at 1 minute. Comparisons between 1 minute and 2 minutes are therefore not meaningful, especially because at 1 minute the products in the high-activity assays were close to saturation of the detector (above 1.0 e^8).

Table 4: pH assay of HYS product formation.

	Peak response area	
Buffer used	1 min	2 min
MES pH 5.5	635443	850858
Phosphate pH 6.0	2860224	2399075
Citrate pH 6.0	2183272	1097576
MES pH 6.5	1713416	900544
HEPES pH 7.0	3973651	1970816
Phosphate pH 7.0	18810523	15516128
HEPES pH 7.5	22848140	21184542
Phosphate pH 7.5	18346649	14730331

2.2.6 Kinetics of THAS1 with strictosidine aglycon

To determine the steady state kinetic constants of THAS1 for strictosidine aglycon and NADPH, kinetics were attempted by monitoring the rate of product formation at different initial concentrations of strictosidine aglycon by LC-MS. Monitoring the consumption of the substrate is not possible because strictosidine aglycon is present in solution in multiple forms and can interconvert between the forms. It is not known which isomers of strictosidine aglycon are available to THAS1 for reduction.

After optimisation of THAS1 concentration and the assay timepoints the results were analysed using the software SigmaPlot. The data were plotted in a Michaelis-Menten plot (fig. 37 and 38) and the K_m was calculated both for strictosidine aglycon and for the cofactor, NADPH.

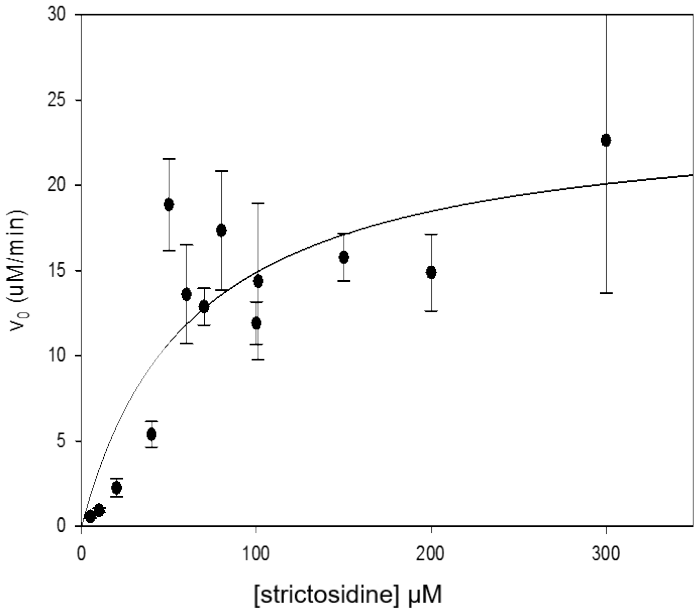


Figure 37: Michaelis-Menten plot of THAS1 with varying concentrations of strictosidine aglycon. Product formation rate (µmol/min) is plotted against the initial concentration of strictosidine aglycon (µM). The concentration of NADPH was kept constant at 500 µM.

Assays with varying amounts of strictosidine aglycon were not straightforward; the product precipitates in the assay conditions above approximately 50 µM and this is reflected in the errors associated with the different measurements by LC-MS (fig. 37). For reasons that are not clear the kinetics done with strictosidine aglycon seem to follow a sigmoidal pattern at low concentrations. This would suggest a threshold of substrate concentration needs to be passed before THAS1 can achieve a certain rate of product formation. However, given that the substrate can equilibrate into multiple isoforms this phenomenon could also be produced if at low concentrations the equilibration into the accepted substrate isoform is smaller than at higher concentrations. Because of the doubt surrounding the actual THAS1 substrate concentration the sigmoidal character of the measured product formation rate is not taken into account. Despite these caveats the THAS1 K_m for strictosidine aglycon was calculated to be 58.29 ± 20.7 µM.

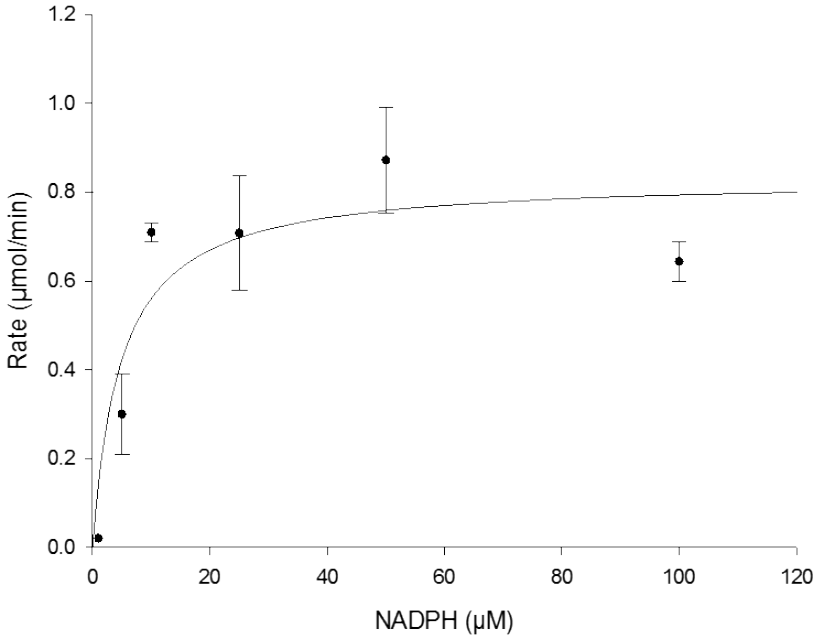


Figure 38: Michaelis-Menten plot of THAS1 with varying concentrations of NADPH. Product formation rate (µmol/min) is plotted against the initial concentration of NADPH (µM). The concentration of strictosidine aglycon was kept constant at 300 µM.

Kinetics done with varying NADPH concentration also displayed large error. The K_m for NADPH is calculated to be 4.84 ± 2.5 µM.

2.2.7 ITC

Isothermic titration calorimetry of THAS1 with NADPH and deglycosylated strictosidine was carried out at constant temperature to determine the order of substrate binding and dissociation constants (k_d). It was discovered that the NADPH cofactor binds first with a k_d of 1.5 ± 0.1 µM (ΔH (cal/mol) 2310 ± 123.2; ΔS (cal/mol/deg) 34.2 ± 0.3, fig. 39) which is consistent with reported values for other MDRs (Lee et al., 2013). Titration assays with deglycosylated strictosidine and THAS1 in the absence of any cofactor did not give any signal, indicating that deglycosylated

strictosidine does not bind to the active site in the absence of NADPH. Titration with deglycosylated strictosidine was also attempted on THAS1-NADPH complex in order to measure the k_d for the second substrate. However, when the substrate concentration reached approximately 50 µM precipitation was observed and this caused aberrant signals which could not be processed.

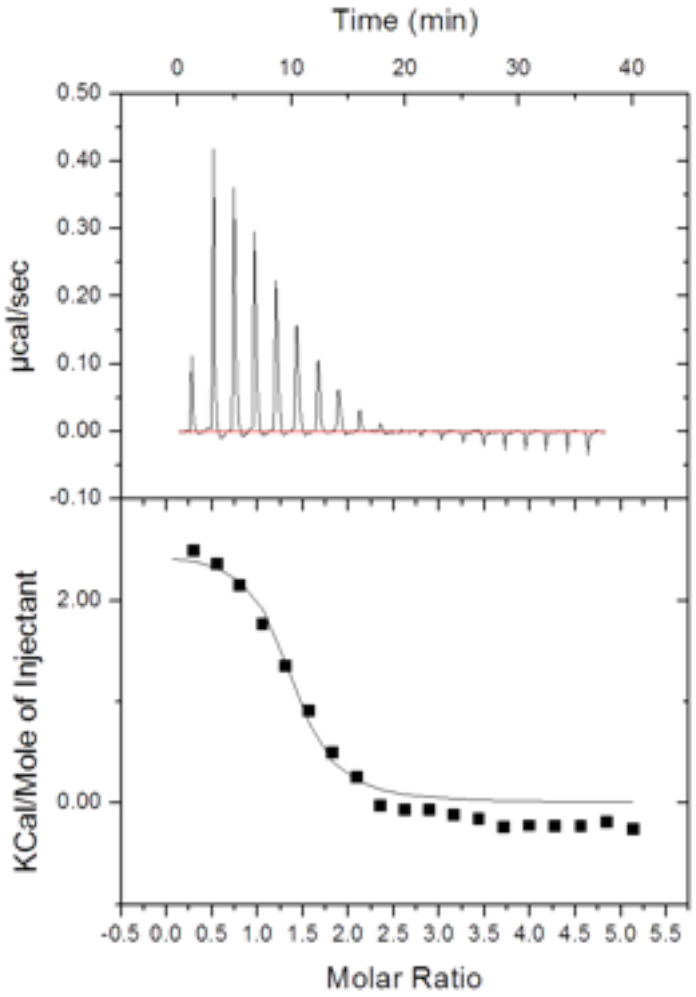


Figure 39: Isothermic titration calorimetry for NADPH binding to THAS1. Titration with NADPH indicates that binding of the cofactor to the enzyme is an exothermic process and THAS1 has a k_d of 1.5 ± 0.1 µM. Bottom panel illustrates the molar ratio between the injectant (NADPH) and the target (THAS1) is 2.0, i.e. approx. 2 moles of NADPH bind per dimer of THAS1.

2.2.8 Pull-down of THAS1 with SGD

Collaborators have demonstrated that THAS1 localises to the same compartment as SGD (the nucleus) suggesting there might be a biological role for this common localisation (Stavrinides et al., 2015). To test if there is a protein-protein interaction between SGD and THAS1, SGD was used as bait in a pull-down assay of THAS1. His_6-tagged SGD was loaded onto a nickel-column and an untagged preparation of THAS1 (fig. 40 A) was loaded onto the column. Aliquots of the different steps were analysed by SDS-PAGE (fig. 40). Although most of the loaded THAS1 appears to have not been retained by the column (fig 40 B lanes 8 to 12) a small amount was retained and eluted

together with SGD (fig 40 B lanes 14 to 16). These results were suggestive of an interaction between SGD and THAS1, but not conclusive. However, BiMFC experiments performed by a collaborator confirmed the interaction *in vivo*.

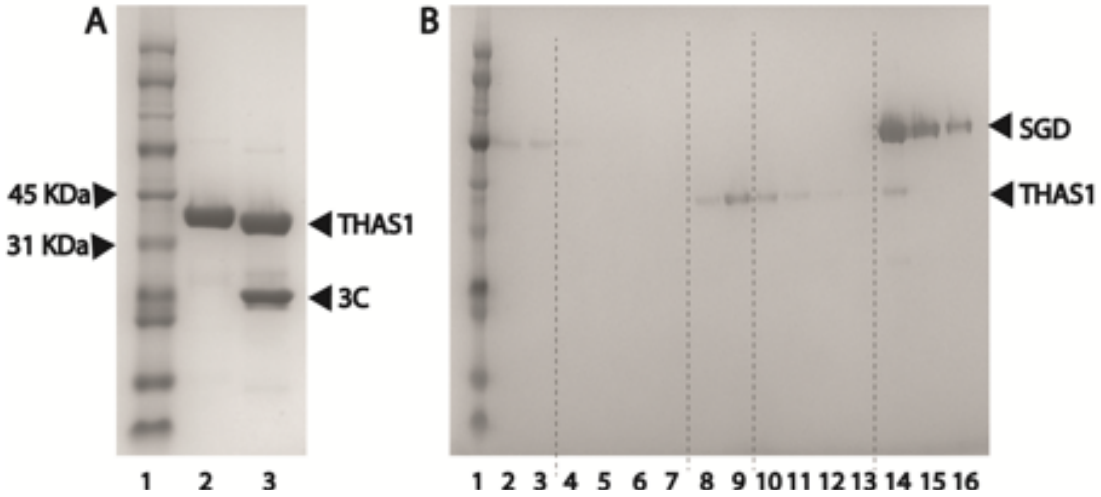


Figure 40: Cleavage of His_6-tag of THAS1 and pull down with SGD. A: SDS-PAGE analysis of His_6-tag cleavage by 3C protease, lane 1: Protein ladder; lane 2: Un-cleaved His_6-THAS1; lane 3: His_6-tag cleaved THAS1 and 3C protease. B: Pull down of THAS1 with SGD. Lane 1: Protein ladded; lane 2-3: SGD loading onto nickel column; lane 4-7: washes; lane 8-9: THAS1 loading onto nickel column; lane 10-11: washes; lane 14-16: elution of bound SGD and THAS1.

2.3 Discussion

2.3.1 Heteroyohimbines in *C. roseus*

Tetrahydroalstonine (fig. 41) is a heteroyohimbine alkaloid derived from the reduction of deglycosylated strictosidine. Deglycosylated strictosidine is a reactive compound which can undergo multiple rearrangements to generate different backbones (see fig. 7, Chapter 1). There are four possible heteroyohimbine stereoisomers that can be produced from deglycosylated strictosidine, three of which have been reported to be found in *C. roseus*. These heteroyohimbines differ at the stereochemistry at the C-20 and at C-19. When the ring D closes to form the five member ring system of deglycosylated strictosidine, the methyl group (C-18) can be either in (*R*) or (*S*) configuration to yield cathenamine and epi-cathenamine, respectively (fig. 19). Cathenamine can give rise to THA or AJM, presumably resulting from the stereochemistry of the iminium tautomer at C-20. It is hypothesized that this iminium is then reduced to form either THA or AJM. Ring closure of the strictosidine aglycon to yield 19-epicathenamine analogously gives rise to the heteroyohimbines 19-EA and rauniticine (which is not reported to be found in *C. roseus*). It is unclear why there is not an equal distribution of all four heteroyohimbines in *C. roseus*. The enzymatic redundancy described here indicates there could be more undiscovered HYSs that could account for the product distribution observed in the plant. There could also be enzymes of different eductase enzyme families acting as HYSs. The function of the heteroyohimbines *in planta* is not understood, and without a better understanding it is not possible to draw conclusions about the significance of these product ratios. The mechanism of reduction will be detailed in Chapter 3.

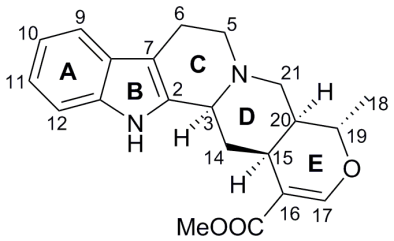


Figure 41: Tetrahydroalstonine

2.3.2 THAS1 discovery and characterisation

Previously reported hierarchical clustering of the transcriptome of *C. roseus* identified several genes annotated as ADHs which appeared to be co-regulated with previously characterised upstream pathway genes. An MDR from this list was cloned from leaf cDNA of *C. roseus* and expressed heterologously in *E. coli*. This enzyme, when incubated in the presence of NADPH and strictosidine aglycon, produces primarily THA, as verified by NMR, with small amounts of AJM and 19-EA also produced. THAS1 is also able to utilise NADH, though at rates too low to be accurately quantified.

An attempt was made to determine the K_m of THAS1 for the strictosidine aglycon/cathenamine and NADPH substrate. Efforts to accurately measure the steady-state kinetics of this enzyme were complicated by several factors. First, strictosidine aglycon reacts with nucleophiles, opening the possibility that the substrate, especially at higher concentrations, reacts with components in the reaction or with amines present on the surface of the enzyme. Furthermore, strictosidine aglycon exists as several isomers that presumably interconvert. If only one of these isomers is the correct substrate for the ADH, then estimating the concentration of the actual substrate in the reaction will be a challenge. Measuring equilibration rates *in vitro* would likely have limited relevance *in vivo*. Nevertheless, given these caveats, we obtained estimated K_m and k_{cat} values. Unfortunately the measurement error was very large for strictosidine at concentrations above 50 µM and therefore the final K_m calculation has a large error associated with it. Despite these complications the K_m for strictosidine aglycon was calculated to be 58.29 ± 20.7 µM. Interestingly, for concentrations of strictosidine below 50 µM the rate of product formation appears to follow a sigmoidal pattern. It is not clear why this phenomenon is observed, but sigmoidal curves are usually interpreted as representing an allosteric enzymatic activity. This behaviour in the case of strictosidine aglycon could be due to the substrate interconversion between different forms; at low concentrations of strictosidine aglycon the substrate of THAS1, cathenamine, might not be present at high enough concentrations. THAS1 was calculated to have a K_m for NADPH of 4.84 ± 2.5 µM. The binding constant was also calculated for the cofactor, NADPH, using ITC (k_d of 1.5 ± 0.1 µM), and was discovered that two molecules of NADPH can bind to the THAS1 dimer. These values for NADPH are comparable to values obtained for similar characterised enzymes (Lee et al., 2013).

Assays to determine the pH optimum of the enzyme revealed that THAS1 has a faster rate at lower pH values. Assays to determine the temperature optimum were not as straightforward as the pH optimum assays and the results are not as reproducible, but the enzyme appears to perform best between 45 and 55 ˚C; this is in line with results found for other similar MDRs (Mansell et al., 1974). It is not clear if this increase in catalytic efficiency is only due to facilitation of the enzymatic catalysis or whether substrate equilibration or solubility plays a large part in this effect. Similarly, it is not clear if the increase in activity at lower pH is solely due to the enzyme catalysis becoming more efficient or rather the solubility or equilibration of the substrate speeding up. For these reasons and because of limited substrate availability temperature and pH optima were not determined for all the HYSs.

2.3.3 Discovery of other HYS

The initial discovery of THAS1 prompted a more in-depth study of other *C. roseus* MDRs. By submitting the protein sequence of THAS1 to a BLAST against the *C. roseus* transcriptome it was possible to identify similar MDRs. The genes for these other candidates were cloned and expressed in *E. coli* in the same way as THAS1. Surprisingly one transcript, Cr032583, when cloned resulted in two distinct genes with identity at the cDNA level of 96%. This level of variation is too high to be explained by allelic variation and therefore the two versions of the gene were considered as two independent but similar genes. It is not clear why the transcriptome assembly was not able to differentiate between the two genes.

Four other MDRs were identified which produce primarily THA, based on LC-MS assays and co-elution with authentic standards. They yield varying amounts of product ratios with THA: AJM: 19-EA from 50:35:15 (HYS) to 96:2:2 (THAS1 table 3). These other HYSs, although of the same enzyme family as THAS1, are quite diverged. Pairwise alignment between the protein sequences reveals the highest homology is between THAS4 and HYS, which appear to be recently-diverged copies. Interestingly, these two clones were found to have strikingly different product profiles, with THAS4 being selective for THA production and HYS presenting a more promiscuous product profile. The next highest homologies are between THAS2, HYS, and THAS1. Perhaps more interesting is the fact that THAS2 and THAS3 share only approx. 53% identity with THAS1 but are still tetrahydroalstonine synthases. This reveals the large flexibility of the enzyme scaffold for production of heteroyohimbines.

All the MDRs cloned code for proteins of approximately 350 amino acids. The HYS discovered converted varying amounts of the initial strictosidine aglycon substrate, but none converted more than 23.2% (fig. 34). This is a comparable conversion to the 26% conversion described by Stöckigt et al. (1976) and it is possible one of the enzymes described here was the one purified by Hemscheidt and Zenk (1985). The enzymes are of similar size and catalyse the same reaction, but it is not possible to say with certainty which enzyme(s) Hemscheidt and Zenk (1985) had purified.

2.3.4 Interaction with upstream enzymes

A class 5 nuclear localisation signal (Kosugi et al., 2009) was discovered in the protein sequence of THAS1. Strictosidine biosynthesis is carried out in the vacuole of *C. roseus* cells (McKnight et al., 1990), and the following step, deglycosylation by SGD is done in the nucleus (Guirimand et al., 2010). The spatial organisation of the pathway is of great importance and the possibility that THAS1 is also present in the nucleus in close proximity to SGD is very intriguing. To test whether THAS1 and SGD have a protein-protein interaction between them a pull-down was done using His-tagged SGD and non-tagged THAS1.

The results of the pull-down of THAS1 with SGD suggests that these two successive enzymes have protein-protein interactions but the interaction seems to be weak given the small amount of THAS1 that co-eluted with SGD. It is not possible to comment on the specificity of the interactions from this experiment. The nuclear localisation signal is present in the closest paralogs of THAS1, THAS4 and HYS, and this suggests these enzymes might too interact with SGD in the nucleus. Indeed, fluorescently-tagged THAS1 and HYS are localised to the nucleus of transformed *C. roseus* protoplasts as demonstrated by collaborators (Stavrinides et al., 2015, Stavrinides et al., 2016). These collaborators have also been able to show that these enzymes interact with SGD in the nucleus of transformed *C. roseus* protoplasts which suggests the localisation serves a role.

2.4 Conclusion

The discovery of these heteroyohimbine synthases in *C. roseus* completes the biosynthetic pathway of the heteroyohimbines. Much of the MIA structural diversity is thought to be generated in the first few steps after the deglycosylation of strictosidine. Discovery of the HYSs is the culmination of many decades of research and constitutes an important step of the branching of the pathway after strictosidine deglycosylation by SGD. The HYSs discovered allow an unprecedented opportunity to study how enzymes achieve stereochemical control of reactive intermediates.

Equilibration of strictosidine aglycon in solution in theory could lead to AJM, 19-EA, THA, and rauniticine. THAS1 could potentially be selectively reducing only the pro-THA iminium form of the substrate pool in solution, or could actively be catalysing the cathenamine-to-pro-THA iminium equilibration in its active site. Likewise, HYS could be promiscuous and could accept all three iminum precursors into the active site for catalysis. In order to answer these questions and to better understand the reduction mechanism a more detailed investigation of the structure/function relationship of the enzymes and the catalysis mechanism is necessary. This is discussed in Chapter 3.

Most of the results described in this chapter are published in Stavrinides et al. (2015) (Annex 1).

2.5 Materials and Methods

2.5.1 RNA extraction from *C. roseus* and cDNA synthesis

Young leaf material (40 mg) of *C. roseus* cv. SunStorm® Apricot was ground in a mortar and pestle and RNA was extracted using the RNEasy plant mini kit (Qiagen, UK) according to manufacturer's instructions. The genomic DNA was degraded using RNase-free DNase (Roche Diagnostics, UK) according to manufacturer's instructions. The RNA quality was verified by running an aliquot on a 1% agarose gel at 100 V for 30 min and checking for presence of the Ribosomal RNA bands and absence of a smear that would indicated carry-over of genomic DNA. Good-quality RNA was subjected to reverse-transcription for synthesis of the first cDNA strand using the SuperScript III Reverse Transcriptase (Thermo-Fisher Scientific, UK), according to manufacturer's instructions using poly-A primers to specifically amplify expressed gene transcripts. The produced cDNA was used fresh or aliquoted in 4 µL volumes and stored at -20 ˚C until use.

2.5.2 Identification of candidate gene

The *C. roseus* transcriptome (Góngora-Castillo et al., 2012) was analysed for highly-expressed Alcohol Dehydrogenase/Reductase genes that were expressed at non-negligible levels (>1 fragments per kilobase of transcript per million mapped reads, fpkm) in young and fully developed leaves and were also upregulated by Methyl-jasmonate at the seedling stage.

In parallel the ADHs identified from the transcriptome were located in the whole genome assembly of *C. roseus* (Kellner et al., 2015) in order to determine their genomic context. THAS1 (transcript Cr024553) was found in close genomic proximity to a paralog of the Secologanin Synthase gene (SLS3, Cr024556).

2.5.3 Polymerase chain reaction

Cloning from cDNA was routinely performed using 1 µL of cDNA, 0.5 µM of each primer and the high-fidelity DNA polymerase KOD Hot Start (Merck Millipore, Herfordshire, UK) according to the manufacturer's instructions for 30 cycles. The annealing temperature was set at 2-3 ˚C lower than the lowest melting temperature of the primer pair used. Screening of positive colonies was done as follows: Four to eight colonies were picked using a sterile toothpick and placed in single wells of a 96-well sterile plate containing 150 µL of LB media supplemented with appropriate antibiotics. The cultures were placed at 37 ˚C for 3-4 hours, or until they became cloudy. PCR was done using 1 µL of the culture, 0.5 µM of each primer, and the HotStart Taq polymerase according to manufacturer's instructions (Qiagen, UK). The first step of the colony PCR was a minimum of 5 min at 92 ˚C to break open the *E. coli* cells, and the annealing temperature was set to at least 2-3 ˚C lower than the lowest melting temperature of the primer pair used.

2.5.4 Vectors

For expression of proteins the InFusion system was used; mainly the vectors pOPINF (high-copy, AmpR, N-terminal His6-tag), and pOPINA (low-copy, KanR, C-terminal His_6-tag). These vectors are designed for directional cloning and possess a T7 promoter and a T7 terminator. For simple cloning of PCR fragments the vector system pJET (high-copy, AmpR, CloneJet PCR cloning kit, Thermo Fisher Scientific) was used. This does not have directional cloning and does not require the manufacturer's kit but rather any T4 ligase can be used.

2.5.5 Cloning THAS1

The gene coding THAS1 was amplified from *C. roseus* leaf cDNA using primers designed based on the expression sequence reads alignment. The open reading frame of the gene was amplified using the primer pair detailed in table 5 which contain overhangs for directional InFusion cloning (Berrow et al., 2007). PCR was performed as detailed above and the amplified gene fragment was gel-purified from 1% agarose gel (Promega, UK) and cloned into the *E. coli* expression vector pOPINF (Berrow et al., 2007) following the manufacturer's instructions for ligation. Vectors were transformed into chemically competent *E. coli* Top10 cells by heat shock at 42˚C for 30 seconds and then spread-plated onto LB+agar plates supplemented with carbenicillin (100 µg/mL). After a night of growth at 37 ˚C positive clones were identified by PCR using the gene-specific primers above and the HotStart Taq polymerase (Qiagen, UK). Positive colonies were grown overnight in 4 mL of LB media supplemented with the appropriate antibiotics at 37 ˚C. The following day plasmids were isolated from the cultures using a miniprep kit (Qiagen) according to manufacturer's instructions. Vectors were verified visually by running on a 1% agarose gel at 100 V and the identity of the inserted sequence was confirmed by Sanger sequencing.

2.5.6 Cloning of other candidate MDRs showing homology to THAS1

The protein sequence of THAS1 was blasted against the transcriptome of *C. roseus* (Góngora-Castillo et al., 2012). The hits which were represented by a full-length gene were considered as good candidates only if their expression levels were acceptable (>1fpkm) in leaves. In order to have representative MDRs which were not likely to function as heteroyohimbine synthases the protein sequence of *Populus tremuloides* SAD was also blasted against the *C. roseus* transcriptome and the top 5 hits of this search were also included in the screen. Furthermore, some *C. roseus* MDR candidates which had been uploaded to NCBI but had not been functionally characterised were also included in this screen.

Primers were designed based on the open reading frame of transcripts after manually verifying that the predicted ORF was full length by submitting the predicted amino acid sequence to BLAST against the NCBI protein database using blastp and checking that the protein was not truncated on either end. Primers used are detailed in table 5. PCR was carried out as described above and

the resulting PCR product was gel purified and cloned into pJET according to manufacturer's instructions (Thermo Fisher Scientific). The resulting vector was transformed into *E. coli* Top10 competent cells by heat shock at 42 ˚C for 30 seconds. Cells were spread plated onto LB plates containing carbenicillin (100 µg/mL) and grown overnight at 37 ˚C. Multiple colonies were verified by colony PCR and two positive clones were verified by sequencing. After verification that the PCR fragment cloned coded for a full-length MDR (by Sanger sequencing) the PCR fragment with the InFusion overhangs was cloned into pOPINF for expression. Positive inserts were verified in two ways, first by miniprep and analysis of the purified plasmid by gel electrophoresis, and secondly by PCR by using the vector specific T7 forward primer and the gene-specific reverse primer.

Table 5: Primers used for cloning the MDR candidates into the expression vector pOPINF

Medium Chain Reductases	Forward primer	Reverse primer
THAS1 Cr024553	AAGTTCTGTTTCAGGGCCCG GCAATGGCTTCAAA	ATGGTCTAGAAAGCTTTAATT TGATTTCAGAGTGTTC
THAS2 Cr021691	AAGTTCTGTTTCAGGGCCCGTCTTCAA AATCAGCAAAACCA	ATGGTCTAGAAAGCTTTAAGCAGATTTCAAT GTGTTTTCTATGTC
THAS3 Cr010119	AAGTTCTGTTTCAGGGCCCGGCAGTTCC ATCGGCAGAAACAG	ATGGTCTAGAAAGCTTTAAACAGATCCCAAA GAATTTTCTATATC
THAS4 Cr032583a	AAGTTCTGTTTCAGGGCCCGGCTGCAAA GTCACCTGAAAATG	ATGGTCTAGAAAGCTTTAGAAAGATGGGGAT TTGAGAGTGTTTCCTACG
HYS Cr032583b	AAGTTCTGTTTCAGGGCCCGGCTGCAAA GTCACCTGAAAATG	ATGGTCTAGAAAGCTTTAGAAAGATGGGGAT TTGAGAGTGTTTCCTACG
Cr027234	AAGTTCTGTTTCAGGGCCCGGCTGGA GAAACAAC	ATGGTCTAGAAAGCTTTATTCCTCAAATTTCA ATGTATT
Cr033062	AAGTTCTGTTTCAGGGCCCGGCCAGA AAATCACCAGAAGATGAAC	ATGGTCTAGAAAGCTTTACACCTCTGATGGA AGAGTGAGAG
Cr017994	AAGTTCTGTTTCAGGGCCCGGCTCATA AGAATTGCTTGAATTTTCTT	ATGGTCTAGAAAGCTTTAGATTATGCATTCTT TCTTGAGAGTGTTC
Cr011702	AAGTTCTGTTTCAGGGCCCGGGGAGC TTGGAAGAAGCAG	ATGGTCTAGAAAGCTTTAGTGGTCAACAAGA AGGTTGCTGCC
Cr022770	AAGTTCTGTTTCAGGGCCCGGCTGGG AAATCACCAGAAG	ATGGTCTAGAAAGCTTTAAGACTCCGGTGGA GGAGTTAAAGTG
Cr030442	AAGTTCTGTTTCAGGGCCCGGCTCAAA CAACTCCAAACCATAC	ATGGTCTAGAAAGCTTTAAAGATTAGATGAT TTGGAAGCTATATCGATC
Cr006840	AAGTTCTGTTTCAGGGCCCGGCAAAG ACACCAGAAACAGAGC	ATGGTCTAGAAAGCTTTATGGACTGGATAAT GAGTTCGCC
Cr033537	AAGTTCTGTTTCAGGGCCCGGCCGGA AAATCAGCAGAAGAAGAA	ATGGTCTAGAAAGCTTTATAACTCTGACGGA GGAGTCAAGGTATT
Cr2141	AAGTTCTGTTTCAGGGCCCGGCCGGA AAATCACCAGAAGAG	ATGGTCTAGAAAGCTTTAAGGAGCTTTCAAG GTCTTTGCAACG
T3R	AAGTTCTGTTTCAGGGCCCGGCTGCAA AGTCAGTGAAGGC	ATGGTCTAGAAAGCTTTAAAAATAGATAGGG TGATTTGAAAGTGTTTCC

2.5.7 THAS1 protein expression and purification

The THAS1 gene was expressed in Rosetta 2 pLysS *E. coli* cells (Novagen®, Merck Millipore, Massachusetts, USA). A starter culture was grown overnight at 37 ˚C with 200 rpm shaking in 20 mL of LB media supplemented with carbenicillin and chloramphenicol (100 µg/mL and 34 µg/mL respectively). The culture was diluted 1:100 in fresh LB media supplemented with antibiotics and was grown at 37 ˚C until an OD_{600} of 0.6. The cultures were cooled on ice and then induced by addition of 0.1 mM of IPTG and placed at 18 ˚C with 200 rpm shaking and allowed to express the protein for 16 h. The cells were then collected by centrifugation and resuspended in 50 mL of Buffer A (50mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 20 mM imidazole) containing a tablet of protease inhibitor cocktail (Roche Diagnostics Ltd) and lysozyme (0.2 mg/mL). Cells were lysed by sonication on ice with an effective sonication time of 3 minutes with 2 seconds ON and 3 seconds OFF. The cell debris was pelleted by centrifugation at 17 x 1000g at 4 ˚C. All purification steps were performed at 4 ˚C on an ÄKTAxpress purifier (GE Healthcare). A HisTrap FF 5 mL column (GE Healthcare) column was equilibrated with Buffer A and the lysate was loaded at a flow rate of 4 mL/min and step-eluted with Buffer B (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 500mM imidazole). The eluate was further purified on a Superdex Hiload 26/60 S75 gel filtration column (GE Healthcare) at a flow rate of 3.2 mL/min using Buffer C (20 mM Hepes pH 7.5, 150 mM NaCl). The fractions were analysed by SDS-PAGE and those containing only THAS1 were pooled and concentrated in a 10 KDa cutoff Millipore filter (Merck Millipore) and the concentration was measured using a BCA assay (Thermo Fisher Scientific Inc., USA).

2.5.8 Small scale protein expression and purification

The pOPINF vectors harbouring the WT genes were transformed into SoluBL21 *E. coli* cells for expression (Novagen®, Merck Millipore, Massachusetts, USA). Positive clones were identified by PCR using the gene specific primers (Table 5). A starter culture was grown overnight at 37˚C in 50 mL of LB media supplemented with the antibiotic carbenicillin (100 µg/mL), and was then diluted 1:100 in 100 mL of fresh 2x YT media supplemented with carbenicillin and allowed to grow to an OD_{600} of 0.6 before induction of expression with 0.1 mM IPTG. The cultures were grown at 18 ˚C for 16 h, with 200rpm shaking. Cells were collected by centrifugation and resuspended in 10 mL Buffer A (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 20 mM imidazole) containing a tablet of protease inhibitor (Roche Diagnostics Ltd) and lysozyme (0.2 mg/mL). Cells were lysed on ice using sonication for an effective time of 1.5 minutes with 2s pulse ON and 3s pulse OFF. Cell debris was pelleted by centrifugation at 17 x 1000g and the supernatant was incubated for 1 h at 4 ˚C with 200 µL of Ni-NTA agarose beads (Qiagen GmbH, Germany) pre-equilibrated with Buffer A. The total was centrifuged at 1000 rpm for 1 minute to pellet the Ni-NTA agarose and the supernatant was discarded. The Ni-NTA agarose was washed three times with 500 µL of Buffer A to remove non-specifically bound proteins, each time centrifuging for 1 min at 1000 rpm to pellet the Ni-NTA agarose. Elution was done by incubating the Ni-NTA with two volumes of 300 µL of Buffer B (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 500 mM imidazole) and centrifuging for 1 min at 1000 rpm.

The eluates were transferred to a Durapore centrifugal filter (PVDF 0.1µm, Merck Millipore) to remove any residual Ni-NTA agarose. An aliquot of eluates was analysed by SDS-PAGE to verify the purity and the molecular weight of the purified proteins. Protein-containing eluates were concentrated to approximately 250 µL and buffer exchanged to Buffer D (50 mM Phosphate pH 7.6, 100 mM NaCl) in a 30 KDa cutoff membrane filter Millipore filter (Merck Millipore). Protein concentration was measured with Bradford reagent (Sigma-Aldrich) according to the manufacturer's instructions. Purified proteins were aliquoted in 20 µL aliquots and snap-frozen in liquid nitrogen before being stored at -20 ˚C.

2.5.9 Enzyme assays

Purified candidate proteins and purified SGD were used in all assays. Strictosidine aglycon was generated in situ prior to addition of reductases by incubation of strictosidine (production is described below in section 2.5.14) and SGD at room temperature in the appropriate solution for ten minutes, at which time strictosidine was completely converted to the aglycon. Strictosidine was routinely used at 300 µM and NADPH at 500 µM. During initial screening the candidate MDRs were used at a concentration of 1 µM. For ease of pipetting and to minimise error all constituents of the reaction, including strictosidine, were mixed as a Master Mix and allowed to incubate at room temperature for at least 10 minutes before aliquoting the appropriate volume and adding the reductases and NADPH. SGD has been shown to have a pH optimum between 6 and 8.5 (Luijendijk et al., 1998), therefore in enzyme assays with a pH between these values strictosidine was deglycosylated by SGD in the buffer used. In reactions where the pH was below 6 then strictosidine was deglycosylated by SGD before addition of the buffer to ensure SGD was functional.

2.5.10 Steady state kinetics for THAS1

Purified THAS1, SGD, and strictosidine were used for all enzymatic assays. Strictosidine was deglycosylated in situ prior to the addition of the reductase enzyme by incubating strictosidine with SGD as described above. Steady state kinetics were performed in 50 mM phosphate buffer (pH 7.5), with 6 nM SGD, 50 nM of THAS1, 200 µM NADPH (Sigma Aldrich), and 50 µM of the internal standard (caffeine).

The kinetics were done as follows: varying concentrations of strictosidine was placed in the wells of a 96-well plate with 50 mM of phosphate buffer, followed by addition of 6 nM of SGD and the necessary volume of MilliQ water to standardize the volume and caffeine (50 µM) was added to this mix as an internal standard. Another set of wells were prepared which contained pre-mixed solutions of 50 nM of THAS and 200 µM of NADPH. At ten minutes the strictosidine mix was added to the THAS+NADPH mix and mixed by pipetting several times. At 0.5 minutes, one minute and two minutes, a 10 µL aliquot of the reaction was placed in 80 µL of H_2O + 0.1% formic acid premixed with 10 µL of methanol, for a 10-fold final dilution of the sample. The presence of 10% methanol was required to keep the THAS product soluble. The 96-well plate was centrifuged at 4000 rpm for 10 minutes to pellet the enzymes and then analysed by LC-MS. A similar setup was

used for determination of the K_m for NADPH, using 300 µM strictosidine and varying the NADPH concentration. The initial rate of the reaction (V_0) was calculated by fitting a linear regression through the points of product formation plotted against time. Michaelis-Menten plots were done using SigmaPlot (Systat Software Inc.).

2.5.11 Isothermic Titration Calorimetry

Isothermic titration calorimetry on a MicroCal iTC200 System (GE Healthcare Life Sciences) was used to determine the binding constant (K_d) for the cofactor, NADPH. For determination of the dissociation constant of THAS1 and NADPH, the purified THAS1 was dialysed overnight against 2 L of Buffer C (20 mM Hepes pH 6.8, 100 mM NaCl). The protein concentration was adjusted with dialysis buffer before the experiment to 40 µM of dimer (80 µM of monomer). Titration was carried out by injecting 2 µL of titrant (1 mM NADPH, dissolved in dialysis buffer) at 750 rpm with a propeller stirrer at intervals of 110 seconds. The first injection was 0.5 µL and was not used for data analysis. Titrations were carried out in triplicate and a control titration was carried out in which Buffer C was injected into the cell containing THAS1 in order to determine the background dilution and mixing heat. This background was then subtracted from the analysis of the NADPH titrations. Data analysis was done using Origin 7.0 software (MicroCal) by fitting a single site model to the data obtained. The dissociation constant (K_d), binding enthalpy (ΔH), and entropy (ΔS) were determined for each replicate.

Titration was also attempted in triplicate for determination of the K_d for cathenamine. THAS1 was premixed with 1 mM of NADPH and prepared cathenamine reaction (1 mM, prepared in Buffer C) was injected using same conditions as above for NADPH injection. Unfortunately substrate or product precipitation interfered with the calorimetry and therefore did not yield any meaningful data. Titration was attempted with cathenamine only (no NADPH cofactor) but did not give any significant binding heat.

2.5.12 Secologanin purification from *Symphoricarpos albus*

Berries of *S. albus* were picked in the autumn when ripe and frozen at -20 ˚C until use. One kilogram of berries was ground using a mortar and pestle in liquid nitrogen until a fine powder. The powder was transferred to a 1 L bottle and 500 mL of MeOH was added and the solution was sonicated for 15 minutes in a water bath and then filtered into a round bottom flask. The berry debris was replaced in the bottle and another 500 mL of MeOH was added and sonication and filtration was repeated twice. The combined methanolic solution was concentrated on a rotary evaporator until reduced to a volume of approximately 10 mL.

The resulting syrupy concentrate was diluted 1:1 in MeOH and injected on a prep-HPLC for secologanin purification. This was performed on a Dionex Ultimate 3000 (Thermo Scientific) preparative HPLC coupled to a multiwavelenth UV-vis detector. Separation was performed on a Phenomenex Luna C18 column (250 x 30 mm) with 5 µm particles. The flowrate was set to 24 mL/minute and started with 99 % Solvent A (H_2O + 0.1 % TFA) and 1 % Solvent B (100% ACN). Solvent

B gradually increased to 40 % at 12 minutes, followed by a rapid increase to 100% at 14 minutes, holding there for 2 minutes and then decreasing to 1 % again at 17 minutes. Column equlibration was done for a further 6 minutes, until 23 minutes, at 1 % Solvent B. The fractions containing secologanin were verified by LC-MS as described above and were combined and concentrated on a rotary evaporator. The concentrate was rediluted in a small volume of H_2O (2-5 mL) and freeze dried overnight and the dry powder was stored at -20 ˚C. The purified secologanin was also verified by ^{1}H-NMR on a Bruker Avance III 400 NMR spectrometer, operating at 400 MHz for ^{1}H and 100 MHz for 13C.

2.5.13 Strictosidine synthase production

The gene coding Strictosidine synthase (STR) was cloned from *C. roseus* leaf cDNA into the expression vector pOPINA (Berrow et al., 2007) using the primers forward 5'AGGAGATATACCATGTCACCAATTTTGAAAAAGATTTTTATTGAAAGCCC and reverse 5'-GTGGTGGTGGTGTTTGCTAGAAACATAAGAATTTCCC. Primers were designed in order to exclude the first 93 bp of the STR gene which constitute a signal peptide for vacuolar localisation of the WT CrSTR. Vectors were transformed into competent *E. coli* Top10 cells by heat shock at 42 ˚C for 30 seconds and then incubated in SOC media for 1h before spread-plating onto LB+agar plates containing Kanamycin (kan) and placing them at 37 ˚C overnight. Colonies were screened by colony PCR using the gene-specific primer pair detailed above. Vectors were verified for correct insertion by sequencing. For expression, the pOPINA construct harbouring STR was transformed into *E. coli* SoluBL21 cells by heat shock as described above. An LB overnight starter culture of 50 mL, supplemented with Kan was placed at 37 ˚C with 200 rpm shaking. This culture was diluted 1:100 in fresh 2x YT media supplemented with Kan and was grown with shaking and allowed to grow to an OD_{600} of 0.6 before induction of expression with 0.1 mM IPTG. Purification of STR was done as described above for THAS1.

2.5.14 Production and purification of strictosidine

Purified secologanin (2 mM) was incubated with tryptamine (2.5mM, Sigma Aldrich) in 50mM Phosphate Buffer (pH 7) and purified STR (the precise amount of STR was not determined). The reaction was placed in a conical flask and placed on a rotary shaker overnight at 30 ˚C. The reaction was stopped by addition of 1 volume of MeOH and the total was concentrated on a rotary evaporator to a syrupy consistency. After filtration through a 0.2 µm pore filter the sample was injected in 2 mL volumes on a preparative HP-LC and purified as described above for secologanin, but collecting the strictosidine elution fractions.

2.5.15 Determination of pH optimum of THAS1

To determine the pH optimum for THAS1 the rate of the reaction in presence of different pH was measured. Phosphate buffer (100 mM) was varied from 5.5 to 8.0 in 1/2 pH increments. All assays were done in triplicate and to minimise the error due to pipetting Master Mixes were used everywhere possible. To stabilise the ionic charge 200 mM NaCl was also included in all

assays. Purified SGD (12 nM) was incubated with strictosidine (200 µM final concentration) in a 50 mM pH 7 buffer solution to deglycosylated strictosidine before the addition of THAS1. This was incubated at 25 ˚C for 30 minutes before aliquoting into clean tubes. The new buffer was added to each (100 mM of the new pH buffer to counter the pH 7 buffer in the SGD incubation pre-mix), as well as the NaCl solution, and the NADPH (200 µM). Purified THAS1 (10 nM) was added and mixed immediately by pipetting. Aliquots were taken at 1, 2, and 3 minutes and prepared for LC-MS measurement as described above.

2.5.16 Determination of temperature optimum of THAS1

Rate of product formation was done using a thermocycler (AB Applied Biosystems Verity 96-well Thermal Cycler) with the capability to adjust to different temperatures across the heating block. The strips were used at 25, 30, 35, 40, 45, and 50 ˚C. Strictosidine (50 µM) was incubated with purified SGD (6 nM) in Phosphate buffer (pH 7.5, 50 mM) at 37 ˚C for 10 minutes, and then was aliquoted into sterile tubes for assaying. Purified THAS1 (10 nM) was pre-mixed in another set of tubes with NADPH (200 µM). After the 10 minutes incubation of the substrate the thermocycler was adjusted at 15 ˚C for 30 sec and the enzyme/cofactor mix was added and mixed. The thermocycler was then brought to the necessary temperature and incubated for 5 minutes. An aliquot of each assay was taken at 2 and at 3 minutes and prepared for LC-MS measurement as described above. This assay was done in duplicate.

As the 50 ˚C assay had a higher rate of activity than the others. Therefore the assay was repeated another day in triplicate with the temperature range set from 40 to 65 ˚C with 5 ˚C increments. LC-MS analysis and quantification was done as described in the section “LC-MS measurement”.

2.5.17 Analysis of pH and temperature optima experiments

To determine the best condition for THAS1 activity the assays in which pH or temperature was varied were analysed in Microsoft Excel. The quantified THA concentrations were corrected for dilution (multiplied by the dilution factor, 10). For the temperature optimum the rate of the reaction was calculated by comparing the product formed at 2 minutes with that measured for 3 minutes. The equation used is

$$\frac{(y2-y1)}{(x2-x1)}$$

where y1 and y2 are the amounts of THA measured for timepoints 2 and 3 minutes respectively, and x1 and x2 correspond to the timepoints 2 and 3 minutes respectively.

For the pH optimum the rate of the reaction was calculated by plotting all three measurements (timepoints 1, 2, and 3 minutes) and calculating a linear regression through the three points using the equation

$$y=ax-b$$

where y is the THA produced, a is the slope of the correlation, x is the timepoint, and b is a constant.

The rates were plotted against pH or temperature to determine the best conditions for THAS1. Student's t-test was used to determine if two measurement means were statistically different from each other.

2.5.18 Determination of pH optimum of HYS

To determine the optimum pH/buffer combination for use during HYS reactions, a small panel of buffers was assayed. Strictosidine (500 µM) was incubated with purified SGD in deionised water for 15 minutes at room temperature. Then the reaction was split into eight tubes containing different buffers. The buffers used are (in order of increasing pH): MES pH 5.5, Phosphate pH 6.0, Citrate pH 6.0, MES pH 6.5, Phosphate pH 7.0, HEPES pH 7.0, Phosphate pH 7.5, and HEPES pH 7.5. All buffers were used at a concentration of 50 mM. NADPH (1 mM) was added, followed by 50 nM of purified His-tag cleaved HYS. The reactions were mixed and an aliquot was taken at 1 minutes and at 2 minutes and prepared as usual for analysis by LC-MS.

2.5.19 LC-MS measurements

All measurements were done on an AQUITY UPLC with a Xevo TQ-S Mass-spec equipped with a BEH Sheild RP18 1.7 µm column (Waters). Two different methods were employed for analysis of the samples. The first was a quick method (5 minutes) which allowed separation of strictosidine, strictosidine aglycon and heteroyohimbines, but did not provide separation of the different heteroyohimbine diastereomers. The second LC-MS method was developed to separate the diastereomers but was only used for assays where it was necessary to determine the ratios of the products as this method was substantially longer (23 minutes).

For the first "fast method" the solvents used were Water + 0.1% Formic Acid as solvent A1 and 100% Acetonitrile as solvent B1, with a flow rate of 0.6 mL/minute and the column was held at 35 ˚C. Two µL of the sample was injected on a gradient starting with 5% solvent B1 increasing to 35% for 3.5 minutes, changing to 100% at 3.75 minutes, held for 1 minute to wash the column, then back to 5% solvent B1 for 1 minute to re-equilibrate the column.

For the second "separation method", developed based on the method in Sun et al. (2011), the solvent A2 was Water + 0.1% NH_4OH and solvent B2 was 0.1% NH_4OH in acetonitrile. A linear gradient from 0% to 65% B2 in 17.5 minutes was applied for separation of the compounds followed by an increase to 100% B2 at 18 minutes, a 2-minute wash step and a re-equilibration at 0% B2 for 3 minutes before the next injection. The column was held at 60 °C throughout the analysis and the flow rate was 0.6 ml/minute.

All MS detections were done in positive ESI mode. Capillary voltage was 3.0 kV; the source was kept at 150 °C; desolvation temperature was 500 °C; cone gas flow, 50 L h-1 and desolvation

gas flow, 800 L h-1. Unit resolution was applied to each quadrupole. Targeted methods for each compound were developed using commercial standards (caffeine, AJM (Sigma-Aldrich Co Ltd, Dorset, UK)) or enzymatically produced compounds (strictosidine and strictosidine aglycon, Table 6). Multiple Reaction Monitoring (MRM) signals were used for detection and quantification of the heteroyohimbine alkaloids (table 6).

Table 6: Multiple Reactions Monitoring conditions

Molecule	Parent ion	Daughter ion	Cone Voltage	Collision Voltage
Strictosidine aglycone	351	170.22	28	22
	351	144.16	28	24
Strictosidine	531	352.25	32	24
Heteroyohimbines	353	117.19	50	40
	353	144.16	50	26

For each compound the parent ion was surveyed and fragmented using cone voltages and collision voltages optimised for the standards. Quantification of product formed was done using TargetLynx software by integrating the detected peak after two iterations of Savitzky-Golay smoothing (MassLynx™, Waters).

2.5.20 Large-scale THAS1 reaction and product purification

For NMR characterisation of the THAS1 product a large scale reaction was setup to produce enough purified product. Strictosidine (4.3 mg) was diluted in 20 mL of distilled water to give a final concentration of 400 µM and the pH was adjusted to 7.5 with 50 mM phosphate buffer. Purified SGD (3 nM), 500 µM of NADPH, and 100 nM of THAS1 were added to the solution. An NADPH regeneration system was employed by addition of 20 units of Glucose-6-Phosphate Dehydrogenase (Roche Diagnostics) together with 1 mM of glucose-6-phosphate. The reaction was incubated with gentle shaking at room temperature and the reaction progress was followed by subjecting aliquots to LC-MS analysis in order to verify strictosidine aglycon was consumed.

The reaction was deemed complete after 5 hours and was quenched by addition of two volumes of MeOH. The total was concentrated by evaporating to dryness under vacuum with the help of a rotary evaporator. The dry precipitate was extracted with MeOH and the supernatant was transferred to a clean round-bottom flask. This was repeated multiple times to yield an approximate volume of MeOH of 15 mL. The remaining precipitate was re-extracted by addition of 2 mL of distilled water, followed by addition of 2 mL of ethyl acetate (HPLC grade). The two phases were mixed gently and then allowed to separate. The organic fraction was transferred to a clean round-bottom flask and the extraction of the water phase was repeated 5 times with ethyl acetate. The ethyl acetate and MeOH fractions were dried under vacuum and were each resuspended in 50 µL of ethyl acetate. A preparative silica TLC plate (UNIPLATE, Analtech™) was

pre-soaked with trimethylamine (TEA). The extracts were loaded onto the TLC plate and the plate was developed in ethyl acetate : hexanes : TEA (24 : 75 : 1), twice, allowing the plate to air dry between runs. The bands were visualised using UV (254 nm). Aliquots of commercially available standards of AJM, THA, and serpentine were also loaded on the TLC as reference.

The major product of the reaction was excised from the TLC plate using a scalpel and the silica was extracted by crushing in ethyl acetate multiple times (for a total volume of approx. 20 mL). The ethyl acetate was filtered into a clean round-bottom flask through a millipore filter to remove the silica. The extract was evaporated to dryness under vacuum and the product was resuspended in 200 µL of fresh ethyl acetate and passed through an SPE column that was pre-equilibrated with hexanes (HPLC grade). The product was eluted using increasing amounts of ethyl acetate in hexanes and the first three elution fractions (corresponding to 25 – 75 % ethyl acetate) were pooled and dried under vacuum. The product was dried overnight on a high-vacuum pump and was resuspended in 300 µL of $CDCl_3$ (Sigma) and transferred to a Shigemi tube. NMR spectra (^{1}H NMR, ^{13}C NMR) were acquired using a Bruker Avance III 400 NMR spectrometer, operating at 400 MHz for ^{1}H and 100 MHz for ^{13}C. The residual ^{1}H- and ^{13}C NMR signals of CD_3Cl (δ 7.26 for ^{1}H and δ 77.36 for ^{13}C) were used as internal chemical shift references.

The minor products were purified in the same manner as detailed above, but there was not sufficient signal in the ^{1}H NMR spectra to determine the nature of the isolated compounds.

2.5.21 Pull down of THAS1 with SGD

Purified THAS1 was prepared for pull-down assay by cleaving the His-tag using 3C protease. THAS1 (200 µg) was incubated with 3C protease (1 µg) overnight at 4 ˚C, then with an additional 1 µg of fresh 3C protease was added and the reaction was allowed to progress for another thirty minutes at room temperature.

To purify the cleaved THAS1 the reaction was passed through a 0.5 mL Ni-NTi Agarose slurry (Qiagen Ltd., Manchester, UK) pre-equilibrated with Buffer E (20 mM Hepes, pH 7.5, 150 mM NaCl). The flow-through was collected and an aliquot was analysed by SDS-page gel to verify the shift in molecular weight. This cleaved THAS1 was concentrated using a Millipore filter unit with a 10 KDa cutoff and its concentration measured using a BCA assay (Thermo-Fisher). A glass chromatography column (0.5 cm x 10 cm) was loaded with 0.5 mL of Ni-NTi Agarose slurry (Qiagen) which was washed and equilibrated with 15 mL of Buffer E. His-tagged SGD (380 µg) was loaded onto the column and 1 mL fractions were collected. The column was then washed with 5 mL of Buffer C, followed by loading of 100 µg of THAS1 which was premixed with 0.5 mM of NADPH. The column was washed with 5 mL of Buffer E, and then elution was carried out with 3 mL of Buffer F (20 mM Hepes pH 7.5, 150 mM NaCl, 250 mM imidazole). A 20 µL aliquot of each fraction was analysed by SDS-page gel and stained using InstantBlue (Expedeon Ltd, Cambridgeshire, UK).

2.5.22 Alignment of protein sequences

The amino acid sequence of the cloned candidates were aligned using the program Seaview (PRABI-Doua, (Gouy et al., 2010)) using the ClustalOmega algorithm (Sievers et al., 2011). The alignment was visualised using BioEdit software (Hall, 2011).

2.6 References

AI, J., DEKERMENDJIAN, K., NIELSEN, M. & WITT, M. R. 1997. The heteroyohimbine mayumbine binds with high affinity to rat brain benzodiazepine receptors *in vitro*. *Natural Product Letters*, 11, 73-76.

BERROW, N. S., ALDERTON, D., SAINSBURY, S., NETTLESHIP, J., ASSENBERG, R., RAHMAN, N., STUART, D. I. & OWENS, R. J. 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research*, 35.

BOMATI, E. K. & NOEL, J. P. 2005. Structural and kinetic basis for substrate selectivity in Populus tremuloides sinapyl alcohol dehydrogenase. *Plant Cell*, 17, 1598-1611.

CHEN, Z., WANG, Y., ZHAI, Y. F., SONG, J. & ZHANG, Z. 2013. ZincExplorer: an accurate hybrid method to improve the prediction of zinc-binding sites from protein sequences. *Mol Biosyst*, 9, 2213-22.

COSTA-CAMPOS, L., DASSOLER, S. C., RIGO, A. P., IWU, M. & ELISABETSKY, E. 2004. Anxiolytic properties of the antipsychotic alkaloid alstonine. *Pharmacology Biochemistry and Behavior*, 77, 481-489.

ELISABETSKY, E. & COSTA-CAMPOS, L. 2006. The alkaloid alstonine: A review of its pharmacological properties. *Evidence-Based Complementary and Alternative Medicine*, 3, 39-48.

GEU-FLORES, F., SHERDEN, N. H., COURDAVAULT, V., BURLAT, V., GLENN, W. S., WU, C., NIMS, E., CUI, Y. & O'CONNOR, S. E. 2012. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature*, 492, 138-142.

GÓNGORA-CASTILLO, E., CHILDS, K. L., FEDEWA, G., HAMILTON, J. P., LISCOMBE, D. K., MAGALLANES-LUNDBACK, M., MANDADI, K. K., NIMS, E., RUNGUPHAN, W., VAILLANCOURT, B., VARBANOVA-HERDE, M., DELLAPENNA, D., MCKNIGHT, T. D., O'CONNOR, S. & BUELL, C. R. 2012. Development of Transcriptomic Resources for Interrogating the Biosynthesis of Monoterpene Indole Alkaloids in Medicinal Plant Species. *PLoS ONE*, 7, e52506.

GOUY, M., GUINDON, S. & GASCUEL, O. 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*, 27, 221-224.

GUIRIMAND, G., COURDAVAULT, V., LANOUE, A., MAHROUG, S., GUIHUR, A., BLANC, N., GIGLIOLI-GUIVARC'H, N., ST-PIERRE, B. & BURLAT, V. 2010. Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? *Bmc Plant Biology*, 10.

HALL, T. 2011. BioEdit: an important software for molecular biology. *GERF Bull Biosci*, 2, 6.

HEMSCHEIDT, T. & ZENK, M. H. 1985. Partial purification and characterization of a NADPH dependent tetrahydroalstonine synthase from *Catharanthus roseus* cell suspension cultures. *Plant Cell Reports*, 4, 216-219.

KELLNER, F., KIM, J., CLAVIJO, B. J., HAMILTON, J. P., CHILDS, K. L., VAILLANCOURT, B., CEPELA, J., HABERMANN, M., STEUERNAGEL, B., CLISSOLD, L., MCLAY, K., BUELL, C. R. & O'CONNOR, S. E. 2015. Genome-guided investigation of plant natural product biosynthesis. *The Plant Journal*, 82, 680-692.

KOSUGI, S., HASEBE, M., MATSUMURA, N., TAKASHIMA, H., MIYAMOTO-SATO, E., TOMITA, M. & YANAGAWA, H. 2009. Six Classes of Nuclear Localization Signals Specific to Different Binding Grooves of Importin alpha. *Journal of Biological Chemistry*, 284, 478-485.

LEE, C., BEDGAR, D. L., DAVIN, L. B. & LEWIS, N. G. 2013. Assessment of a putative proton relay in Arabidopsis cinnamyl alcohol dehydrogenase catalysis. *Organic & Biomolecular*

Chemistry, 11, 1127-1134.
LI, L. G., CHENG, X. F., LESHKEVICH, J., UMEZAWA, T., HARDING, S. A. & CHIANG, V. L. 2001. The last step of syringyl monolignol biosynthesis in angiosperms is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. *Plant Cell*, 13, 1567-1585.
LI, S., LONG, J., MA, Z., XU, Z., LI, J. & ZHANG, Z. 2004. Assessment of the therapeutic activity of a combination of almitrine and raubasine on functional rehabilitation following ischaemic stroke. *Current medical research and opinion*, 20, 409-415.
LOUNASMAA, M. & KAN, S.-K. 1980. A 400 mhz ^{1}H NMR study of the eight basic heteroyohimbine alkaloids. *Tetrahedron*, 36, 1607-1611.
LUIJENDIJK, T. J. C., STEVENS, L. H. & VERPOORTE, R. 1998. Purification and characterisation of strictosidine beta-D-glucosidase from *Catharanthus roseus* cell suspension cultures. *Plant Physiology and Biochemistry*, 36, 419-425.
MANSELL, R. L., GROSS, G. G., STÖCKIGT, J., FRANKE, H. & ZENK, M. H. 1974. Purification and properties of cinnamyl alcohol dehydrogenase from higher plants involved in lignin biosynthesis. *Phytochemistry*, 13, 2427-2435.
MCKNIGHT, T. D., ROESSNER, C. A., DEVAGUPTA, R., SCOTT, A. I. & NESSLER, C. L. 1990. NUCLEOTIDE-SEQUENCE OF A CDNA-ENCODING THE VACUOLAR PROTEIN STRICTOSIDINE SYNTHASE FROM *CATHARANTHUS-ROSEUS*. *Nucleic Acids Research*, 18, 4939-4939.
MIETTINEN, K., DONG, L., NAVROT, N., SCHNEIDER, T., BURLAT, V., POLLIER, J., WOITTIEZ, L., VAN DER KROL, S., LUGAN, R., ILC, T., VERPOORTE, R., OKSMAN-CALDENTEY, K.-M., MARTINOIA, E., BOUWMEESTER, H., GOOSSENS, A., MEMELINK, J. & WERCK-REICHHART, D. 2014. The seco-iridoid pathway from *Catharanthus roseus*. *Nature Communications*, 5, 3606.
QU, Y., EASSON, M. L. A. E., FROESE, J., SIMIONESCU, R., HUDLICKY, T. & DE LUCA, V. 2015. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proceedings of the National Academy of Sciences*, 112, 6224-6229.
RIVEROS-ROSAS, H., JULIÁN-SÁNCHEZ, A., VILLALOBOS-MOLINA, R., PARDO, J. P. & PIÑA, E. 2003. Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily. *European Journal of Biochemistry*, 270, 3309-3334.
ROQUEBERT, J. & DEMICHEL, P. 1984. Inhibition of the α 1-and α 2-adrenoceptor-mediated pressor response in pithed rats by raubasine, tetrahydroalstonine and akuammigine. *European journal of pharmacology*, 106, 203-205.
SCOTT, A. I., REICHARDT, P. B., SLAYTOR, M. B. & SWEENY, J. G. 1971. Mechanisms of indole alkaloid biosynthesis. Recognition of intermediacy and sequence by short-term incubation. *Bioorganic Chemistry*, 1, 157-173.
SHU, N., ZHOU, T. & HOVMÖLLER, S. 2008. Prediction of zinc-binding sites in proteins from sequence. *Bioinformatics*, 24, 775-782.
SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M., SÖDING, J., THOMPSON, J. D. & HIGGINS, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 539-539.
STAVRINIDES, A., TATSIS, E. C., CAPUTI, L., FOUREAU, E., STEVENSON, C. E. M., LAWSON, D. M., COURDAVAULT, V. & O'CONNOR, S. E. 2016. Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity. *Nat Commun*, 7.
STAVRINIDES, A., TATSIS, EVANGELOS C., FOUREAU, E., CAPUTI, L., KELLNER, F., COURDAVAULT, V. & O'CONNOR, SARAH E. 2015. Unlocking the Diversity of Alkaloids in *Catharanthus roseus*: Nuclear Localization Suggests Metabolic Channeling in Secondary Metabolism. *Chemistry & Biology*.
STÖCKIGT, J., HÖFLE, G. & PFITZNER, A. 1980. Mechanism of the biosynthetic conversion of geissoschizine to 19-epi-ajmalicine in *Catharanthus roseus*. *Tetrahedron Letters*, 21, 1925-1926.

STÖCKIGT, J., TREIMER, J. & ZENK, M. H. 1976. Synthesis of ajmalicine and related indole alkaloids by cell free extracts of *Catharanthus roseus* cell suspension cultures. *FEBS Letters*, 70, 267-270.

SUN, J., BAKER, A. & CHEN, P. 2011. Profiling the indole alkaloids in yohimbe bark with ultra-performance liquid chromatography coupled with ion mobility quadrupole time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 25, 2591-2602.

TATUSOV, R. L., KOONIN, E. V. & LIPMAN, D. J. 1997. A Genomic Perspective on Protein Families. *Science*, 278, 631-637.

WYRAMBIK, D. & GRISEBACH, H. 1975. Purification and Properties of Isoenzymes of Cinnamyl-Alcohol Dehydrogenase from Soybean-Cell-Suspension Cultures. *European Journal of Biochemistry*, 59, 9-15.

WYRAMBIK, D. & GRISEBACH, H. 1979. Enzymic Synthesis of Lignin Precursors. *European Journal of Biochemistry*, 97, 503-509.

Chapter 3

Achieving Stereoselectivity in Heteroyohimbine Synthases

3.1 Introduction

The remarkable chemical diversity found in many Apocynaceae plant species, including *C. roseus*, is due to the reactivity and instability of strictosidine aglycon, a key biosynthetic intermediate for many alkaloid natural products. The heteroyohimbine family of these alkaloids can form after the ring E closure (fig. 42) and reduction of the double bond of ring D. Although theoretically all sixteen stereoisomers (derived both from strictosidine with 3-(*S*) and vincoside with 3-(*R*)) are possible, only three have been reported to be found in the plant *C. roseus*, all derived from strictosidine: ajmalicine [AJM, 19(*S*), 20(*R*)], tetrahydroalstonine [THA, 19(*S*), 20(*S*)], and 19-epiajmalicine [19-EA, 19(*R*), 20(*R*)].

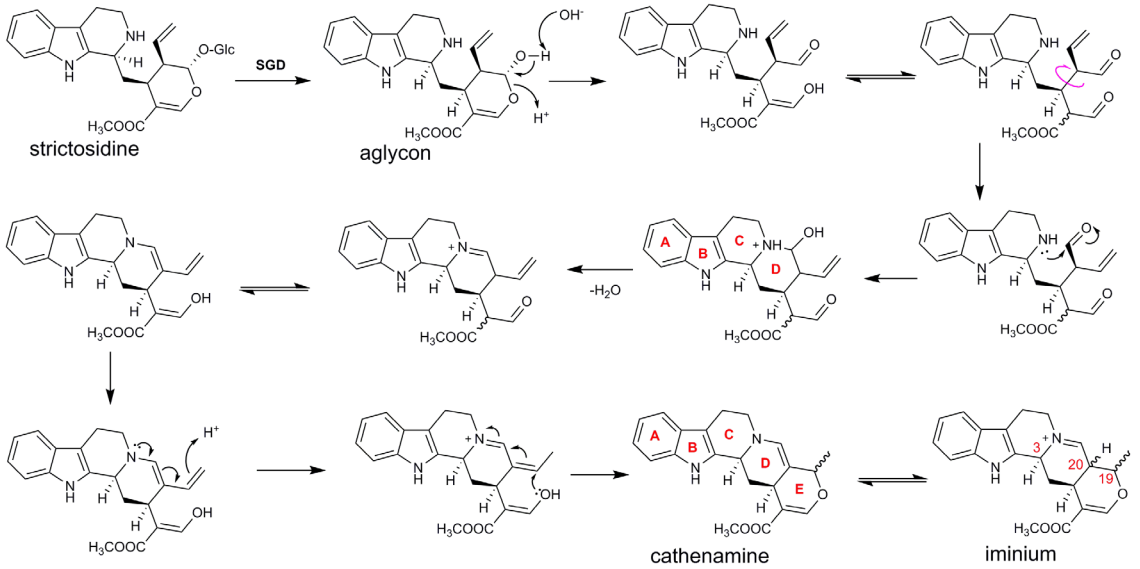


Figure 42: Deglycosylation of strictosidine and equilibration into cathenamine (and its isomer 19-epicathenamine) and generation of the iminium form through tautomerization. The rings formed are indicated in red A-E.

In Chapter 2 I described the discovery of 5 HYSs. Some of the enzymes (THAS1-4) displayed a specific THA-producing profile, but one enzyme, HYS, appeared to be promiscuous and was capable of producing all three heteroyohimbines found in *C. roseus*. Although the enzymes were found by similarity to THAS1 they are quite diverged, except for THAS4 and HYS that share approximately 94 % identity at the amino acid level. The mechanism of reduction would in theory be similar for all five, but the numerous differences in the active site and the loop sections near the predicted active site complicate interpretation of the protein alignment.

The discovery of the heteroyohimbine synthases described in Chapter 2 has demonstrated that the MDR family of ADH enzymes is capable of evolving an imine-reducing activity. However, it is not yet clear if these enzymes are truly imine reducing or if they reduce the carbon-carbon double

bond of the enamine (cathenamine) (fig. 43), but we believe the iminium would be the most likely substrate. Furthermore, the mechanism by which the enzymes control this stereoselective reduction remains unclear.

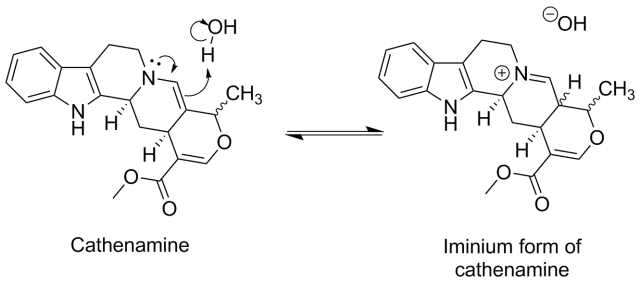


Figure 43: Equilibration of cathenamine and the iminium form of cathenamine in water.

A better understanding of the mechanism of enzymatic stereoselectivity at this key biosynthetic pathway branchpoint is necessary to understand how the other scaffolds stemming from strictosidine might be generated. To this end mechanistic studies have been carried out using deuterium labelling, and mutations to the active site of THAS1 and HYS. Crystal structures of these two enzymes, as well as of THAS2 (done by collaborators) were also obtained to aid in understanding the active site of the enzymes and for *in silico* docking of the substrate.

3.2 Results

3.2.1 Crystallisation of THAS1

THAS1 protein for crystallisation was purified by nickel resin and by gel-filtration chromatography (see fig. 40 A for representative SDS-PAGE gel). Crystallisation screens were initially set up using commercially available screens (PEG, JCSG+, and PACT) using both apo protein and protein pre-mixed with the cofactor $NADP^+$. Initial screening resulted in crystals forming under a number of different conditions, many of which were sent for X-ray crystallography to the Diamond Light Source synchrotron. The best quality holo crystal was obtained from a solution containing 0.2 M potassium/sodium tartrate with 20 % w/v PEG 3350. THAS1 apo crystals lacking the $NADP^+$ cofactor did not diffract to below 2 Å resolution, and therefore an optimisation screen was set up using the best conditions identified in the initial screens. THAS1 without the His_6-tag was also crystallised in case this produced better quality crystals. THAS1 apo crystals were obtained from protein in which the His_6-tag was removed. A high quality apo crystal was obtained from a condition containing 0.1 M MES buffer, pH 6.5, with 15 % w/v PEG 2000.

3.2.2 Crystallisation of HYS

Heterologous expression of HYS produced a dimer enzyme with approximately 38 kDa molecular weight for each subunit (as analysed on SDS-PAGE). Purification only on His-trap was not sufficient to produce protein of high enough purity for crystallisation assays (fig. 44, left). Therefore further purification by size exclusion gel filtration was needed, and resulted in a higher purity protein (fig. 44, right). However, some contaminants are still visible by SDS-PAGE. To remove these last

contaminants anion exchange was attempted, but was largely unsuccessful due to the neutral PI of HYS (7.38) and therefore was not attempted again.

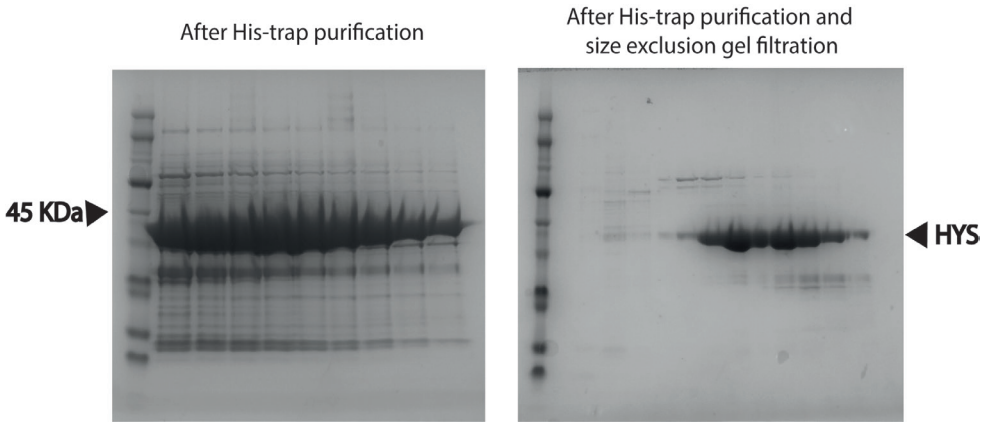


Figure 44: Images of a typical SDS-PAGE gel analysis of purified HYS after only His-trap purification (left) and after both His-trap and size exclusion gel-filtration (right).

The His_6-tag was removed before crystallisation and the protein was further cleaned by passing through a nickel column once again, resulting in a higher purity protein. Crystallisation of HYS was attempted as done for THAS1, using commercially available screens in 96-well sitting drop format but crystallisation of this enzyme proved challenging. Multiple attempts resulted in low quality crystal structures lacking the resolution necessary to clearly define some flexible loop sections, in particular the loop (loop2) over the active site. Crystallisation of the protein in presence of $NADP^+$ was difficult; the few crystals that did form had very low resolution and did not result in any useful dataset. A condition in one of the screens which gave apo crystals was selected for optimisation (fig. 45-48). The MMT Buffer is not available commercially and was prepared according to the recipe used by Molecular Dimensions.

PLATE 1

PEG %	12%	14%	16%	18%	20%	22%	PEG
0.2 mM MMT Buffer pH 5.0							1500
	*	*					3350
							4000
							6000

Figure 45: Optimisation screen for HYS apo crystals with 0.2 mM MMT Buffer (pH 5.0) with varying PEG sizes and concentrations. Asterisks mark the wells with the best crystals.

PLATE 2

PEG %	10%	15%	17.5%	20%	25%	30%	PEG
0.2 mM Sodium Nitrate							400
							1500
							3550
							6000

Figure 46: Optimisation screen for HYS apo crystals with 0.2 mM Sodium Nitrate with varying PEG sizes and concentrations. No usable crystals formed on this plate.

PLATE 3

PEG 1500 %	10%	15%	20%	25%	30%	35%	pH
0.1 mM MMT Buffer							4
							6
							7
							8

Figure 47: Optimisation screen for HYS apo crystals with 0.1 mM MMT Buffer with varying pH and PEG concentrations. No usable crystals formed on this plate.

PLATE 4

PEG %	10%	15%	20%	25%	30%	35%	PEG
0.1 mM MMT Buffer pH 5.0							400
							1500
		*	*				3550
							6000

Figure 48: Optimisation screen for HYS apo crystals with 0.1 mM MMT Buffer (pH 5.0) with varying PEG sizes and concentrations. Asterisks mark the wells with the best crystals.

The crystals that resulted from this optimisation were analysed at Diamond Light Source and the best dataset was obtained from a crystal grown in plate 4 (fig. 48) 0.1 M MMT buffer, pH 5, with 15% PEG 3350. No crystals of good quality were obtained in conditions with $NADP^+$. Due to the low quality of the crystal nucleation, seeding was attempted using crystals from conditions which crystallised more readily. Drops containing $NADP^+$ microcrystals and also drops with left over high-quality apo crystals were used. Optimisation screens were setup using seeds from a total of 6 different conditions using the best condition as precipitant (0.1 mM MMT buffer pH 5.0, PEG 3350). Optimisation screens were setup using 1 mL 24-well plates which were setup manually.

PLATE 5

Seed:	1	2	3	4	5	6	PEG 3350 %
0.1 mM MMT Buffer		*	*				8%
		*	*				
		*					
							20%

Figure 49: Optimisation seeding screen for HYS holo crystals with 0.1 mM MMT Buffer (pH 5.0) with varying PEG concentrations. Asterisks mark the wells with the best crystals.

The seeding optimisation for HYS holo crystals (fig. 49) appeared to be successful, generating multiple drops with crystals in presence of $NADP^+$. Unfortunately, these crystals were revealed to be made up of apo HYS after solving the X-ray diffraction and none of the datasets obtained showed any density for the cofactor in the active site. It appears although the crystals formed, they did not form with the cofactor bound. Therefore attempting to crystallise HYS in presence of the cofactor was not re-attempted.

After multiple failed attempts to improve the resolution, an aliquot of HYS was sent for analysis by thermofluor. This technique uses the temperature-mediated unfolding of the protein to quantify its stability in various buffers or in presence of different additives or ligands.

HYS was tested against a panel of additives, buffers, pH conditions, detergents, and also cofactors. All results of the screens are found in Appendices 2-4. Dissociating reagents, such as UREA and DMSO, appear to destabilise the protein in a concentration-dependent manner. Addition of salts, such as Na_2SO_4, did not have an effect on the protein stability. However, the addition of multivalent and metal ions had a destabilising effect, most notably the addition of 1 mM $ZnCl_2$ which caused the Tm to shift from 61 ˚C to 45 ˚C. Reducing agents, such as DTT and TCEP had a destabilising effect which was concentration dependent. Various crowding agents had a stabilising effect on the protein, notably the addition of glycerol (5-20% w/v) increased the Tm to 64 ˚C (fig. 50). PEG addition did not have the same effect. Interestingly, the addition of imidazole had a detrimental effect on the Tm of HYS in a concentration-dependent manner. The same concentration of imidazole (500 mM) used in Buffer B for protein elution from the Ni-NTA column reduced the Tm by nearly 30 ˚C reducing it to just under 35 ˚C. Some amino acids (L-Arg and L-His) destabilised the protein to some degree, but detergents such as CHAPS did not have an effect. Some cofactors (ATP, NADH, cAMP, etc) were also tested for their effect on HYS but none showed any significant effect on the Tm.

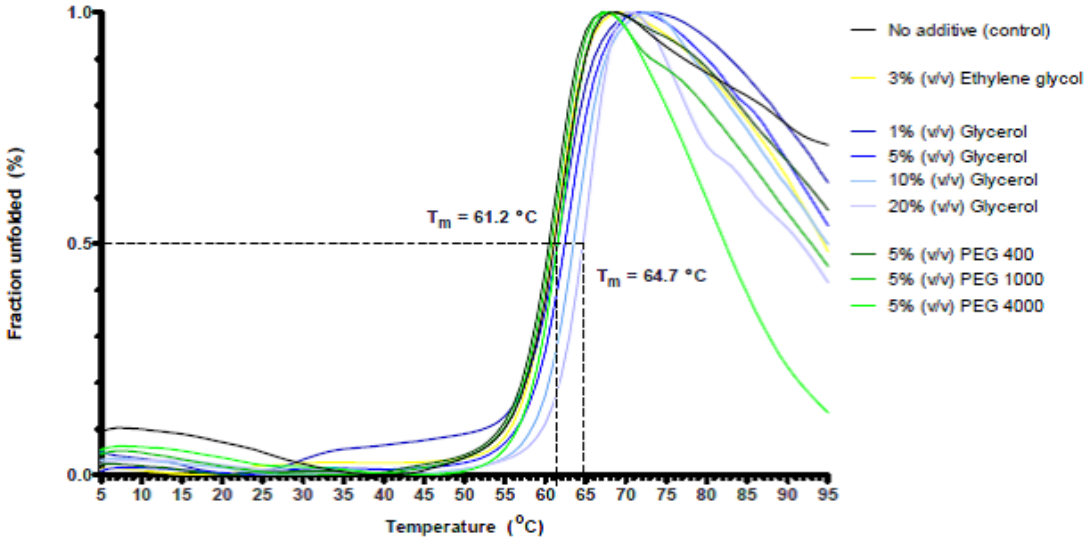


Figure 50: Thermofluor assay of HYS with addition of ethylene glycol, glycerol, or PEG at differing concentrations.

The cofactors included in the general screen did not include NADPH or $NADP^+$ and for that reason an extra screen was requested in order to understand the effect of the cofactor. HYS appeared to have a melting temperature of 60.3 ˚C in 50 mM HEPES buffer pH 7.5 in the absence of any cofactor (fig. 51). After the addition of $NADP^+$ (2 or 5 mM) the melting temperature increased slightly to about 63 ˚C. However, after addition of NADPH (2 or 5 mM) the melting temperature increased by almost 10 ˚C to 70 ˚C. Therefore it is evident this enzyme is greatly stabilised by the addition of NADPH but not by $NADP^+$. This effect probably reflects the effect the cofactor binding has on the tertiary structure of the enzyme.

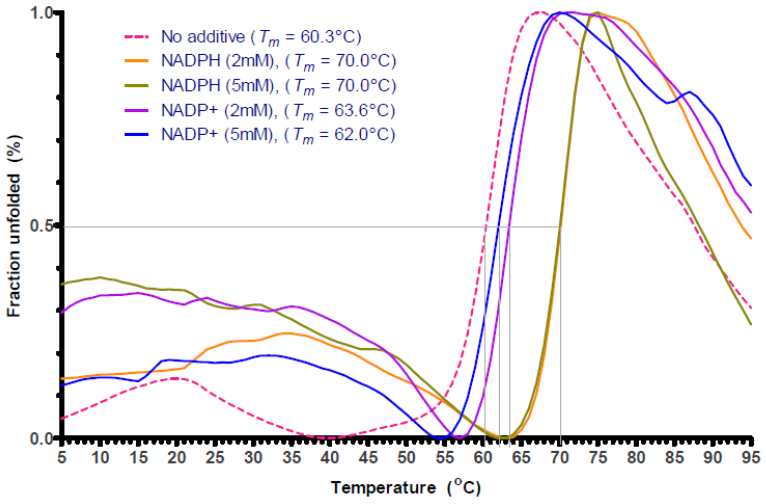


Figure 51: Thermofluor result with HYS and cofactor at varying concentrations. Protein was used at approximately 25 µM.

HYS showed slight variation of stability at different pH values, when tested against a range of different pH buffers (succinic acid/phosphate/glycine 2:7:7) it appeared to be unfolded at pH below 5.0 and above 9.0 (Appendix 2). At pH between 5.6 and 9.0 its Tm was between 57

and 62 ˚C. The concentration of the buffers HEPES, $NaPO_4$, and Tris were tested between 10 to 250 mM at pH 7.5 (8.0 for Tris) but did not seem to have any effect, negative or positive on the stability of the protein. The addition of NaCl however seemed to increase the stability of the protein in HEPES buffer (50 mM, pH 7.5) and in Tris buffer (50 mM, pH 8.0); this increase was seen even at concentrations of 1000 mM NaCl.

Based on the results of the thermofluor assay HYS was stored in a buffer designed for maximum stability (HEPES 50 mM, NaCl 150 mM, glycerol 10%, TCEP 0.5 mM, pH 7.06). The crystallisation was reattempted with HYS purified in optimal conditions with included TCEP, but no crystals could be grown of better quality than the ones identified in the previous screen.

3.2.3 Comparison of crystal structures

The crystal structure for the THAS1-holo enzyme was solved using the strong anomalous signal of the four Zn^{2+} ions contained in each enzyme dimer. After data collection a structure solution was automatically obtained by single wavelength anomalous dispersion phasing and the initial density map obtained was already of sufficient quality (fig. 52) and completeness to contain 94% of the residues expected in the dimer enzyme. The refined structure of THAS1-holo was used to solve the corresponding apo structure and also the apo structure of HYS. THAS1-holo was solved to a resolution of 1.05 Å and the apo structure was solved at 2.25 Å; HYS-apo was solved at 2.25 Å (fig. 53 and 58).

All structures have a similar tertiary organisation, with 13 α-helices and 18 β-sheets. The enzyme is an elongated homodimer with two sections to each monomer: a substrate-binding and a cofactor-binding domain. The dimerization is achieved through the β-sheets 15 and 16 in the cofactor-binding domain (fig. 57). RMSD values between apo and holo THAS1 structures are 1.21 Å and 1.54 Å for the monomer and dimer respectively. Between HYS apo and THAS1 apo the RMSD are 0.72 Å and 1.48 Å for the monomer and dimer respectively. These structures are similar to the crystal structure of the similar MDRs PtSAD (PDB accession codes 1YQX and 1YQD) and also to AtCAD (PDB accession codes 2CF5 and 2CF6) (Bomati and Noel, 2005, Youn et al., 2006). These enzymes are implicated in the lignin biosynthesis of poplar and Arabidopsis respectively (see Chapter 1 for a more detailed description).

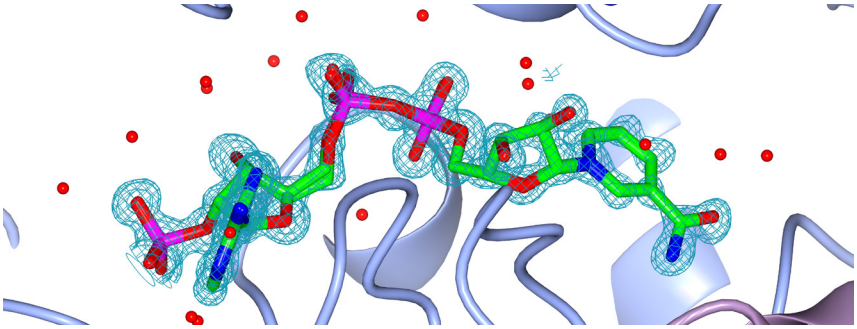


Figure 52: $NADP^+$ depicted in the electron density (turquoise mesh) which was experimentally phased after solving the structure and before any further refinement. Surrounding water molecules are illustrated as red spheres and the THAS1 chain A and B backbones are illustrated in light blue and lilac cartoon representation (Chain B and A respectively).

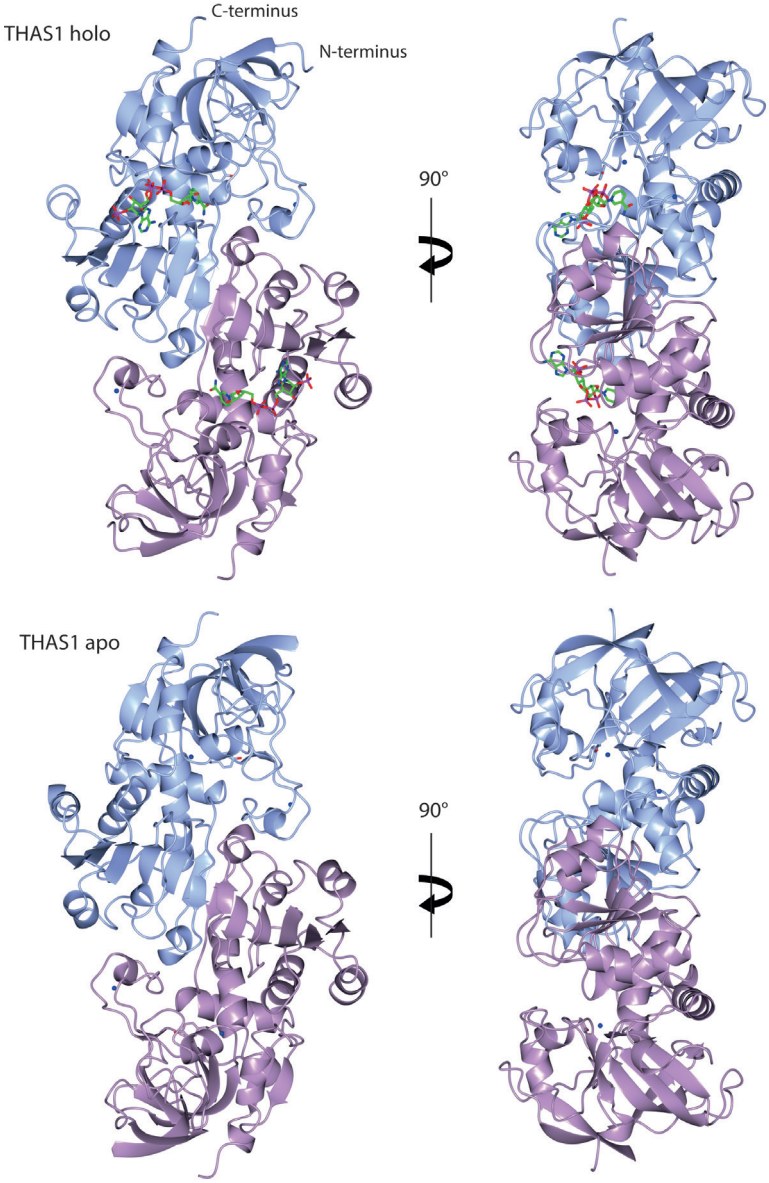


Figure 53: Crystal structure of THAS1 holo (top) and apo (bottom). Views from the side and from the front (left and right respectively). Each chain is coloured differently; zinc atoms are illustrated as dark blue spheres. The $NADP^+$ is illustrated with green carbons in the holo structure.

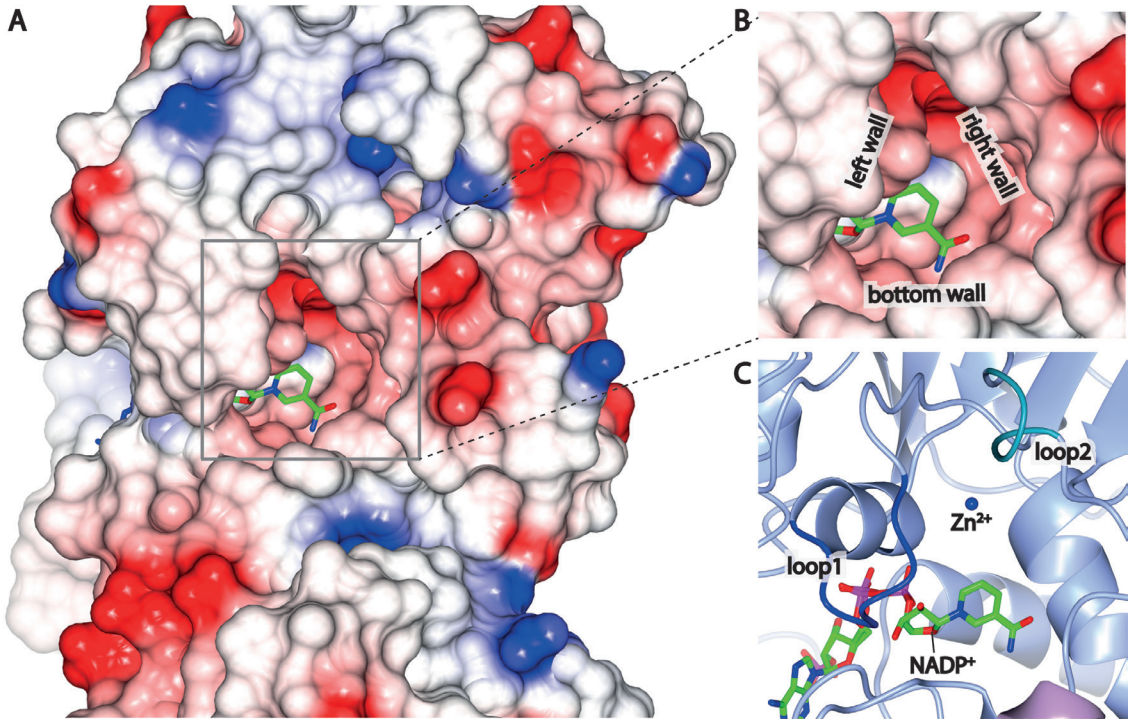


Figure 54: Overview of THAS1 active site. A: Electrostatic map of THAS1, focused on the active site; B: Zoomed in section of area in grey box of panel A, with the right, left, and bottom walls of the active site cavity illustrated; C: Same section as B in cartoon representation with loop1 and loop2 drawn in dark and light blue respectively. The $NADP^+$ is illustrated with green carbons and the Zn^{2+} ion is illustrated as a blue sphere in panel C.

The active site cavity of THAS1 is framed on the left wall by the helix α2, the "catalytic" zinc coordination sphere towards the back, and loop 1 towards the entrance and loop 2 above the active site (fig. 54). The bottom of the active site is responsible for the co-factor binding and the right wall of the active site consists mainly of turns between α-helices and β-sheets.

The $NADP^+$ is held in position by a network of hydrogen bonds with water molecules and residue side chains and backbone amide and carbonyl groups (fig. 55 and 56). Residues from both the catalytic domain of the protein (Met1 residue to Leu152 and Ala299 to Asn356) and the nucleotide binding domain (Pro153 residue to Thr298) take part in this network. The nicotinamide ring of $NADP^+$ is positioned towards the floor of the central cavity which forms the active site of the enzyme. Above this ring there is space which in this crystal structure is occupied by multiple water molecules (3) with more water molecules towards the entrance of the active site. The Tyr56 of THAS1 is positioned above the nicotinamide ring with approximately 6 Å distance between the side chain ring and the cofactor. The ribose hydroxyl groups interact with the carboxylic acid of Glu59 in THAS1-holo; this amino acid is conserved in HYS. Glu59 probably plays a similar role to that played by His55 in PtSAD although in PtSAD His55 interacts with the 3' OH of the NADPH cofactor only.

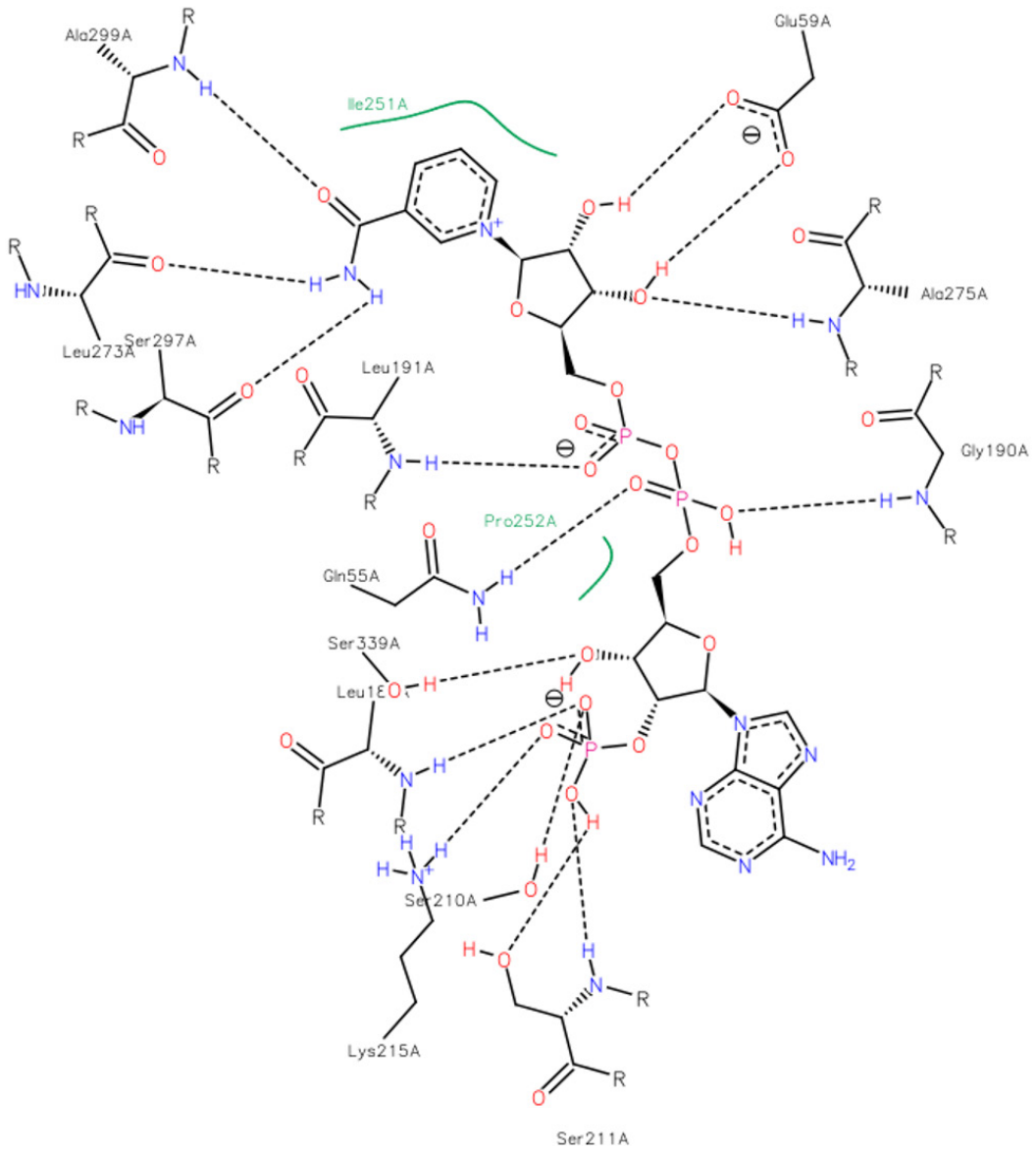


Figure 55: 2-D figure of coordination of $NADP^+$ in THAS1 chain A. Black dashed lines represent hydrogen bonds; green lines represent hydrophobic interactions. Image produced with the software PoseView (Stierand and Rarey, 2010).

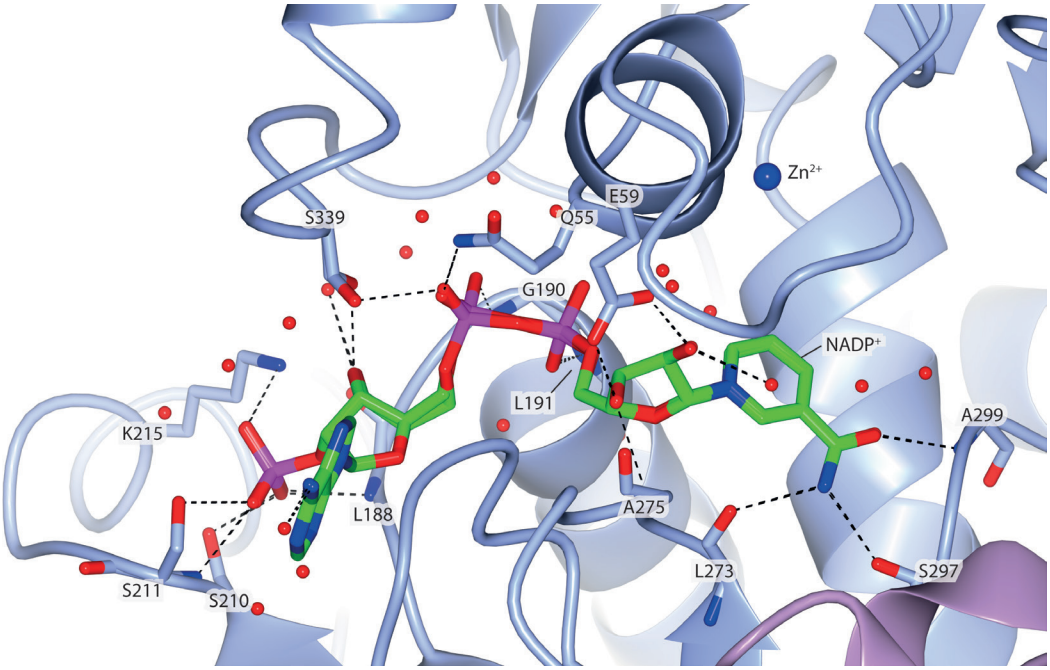


Figure 56: Interactions between NADP⁺ and THAS1. NADP⁺ is illustrated as cylinders with green carbons; THAS1 chain B is illustrated in cartoon with gray carbons and chain A is coloured lilac. The residues interacting with NADP⁺ are illustrated either only as the backbone if the interaction is achieved through the backbone oxygen or nitrogen, or illustrated only with the side chain if the interaction is only achieved through the side chain functional groups. The "active site" zinc ion is illustrated as a blue sphere and water molecules within 4 Å of NADP⁺ are illustrated as red spheres.

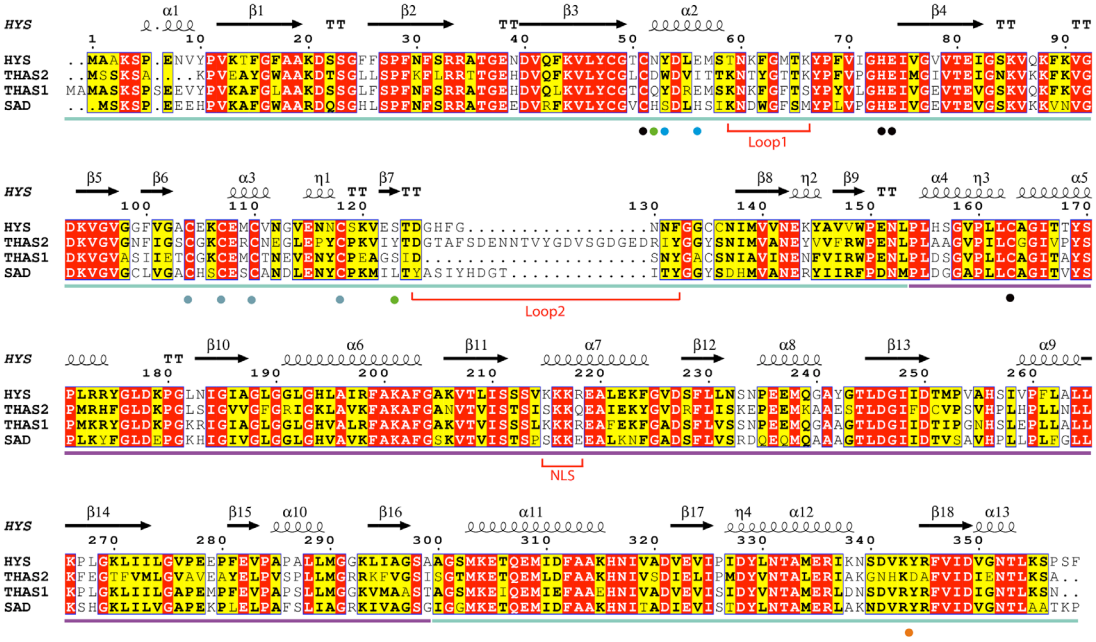


Figure 57: Sequence alignment of *C. roseus* THAS1, THAS2, HYS, and *P. tremuloides* SAD. Numbering corresponds to HYS. Identical and similar amino acids are highlighted in red and yellow, respectively. Secondary structure elements of the HYS apo crystal structure are displayed. THAS1- and HYS-active site amino acids (Tyr56/53 and Glu59/56) are indicated by blue dots, and THAS2 active site amino acids (Tyr120 and Asp49) are indicated by green dots. Ligands for catalytic and structural zinc ions are highlighted by black and grey dots, respectively. The nuclear localization signal of (THAS1 and HYS) and loops 1 and 2, respectively, are indicated in red. A

non-proline cis-peptide bond that is observed in THAS1 holo, in one subunit of THAS1 apo, in HYS apo, and not at all in THAS2 is indicated with an orange dot. The substrate-binding domain and the cofactor-binding domain are indicated by blue and purple bars, respectively.

The HYS active site cavity is in general very similar to that of THAS1 (fig. 56) with the exception of some amino acid differences on loop1 and the extended loop2 which drapes over the active site. Loop1 in THAS1 has two phenylalanines which project into the active site in the crystal structure, one of which does not have enough electron density to model confidently. These phenylalanines (Phe65 and Phe67) could be involved in substrate binding by interacting with its indole moiety. HYS also has a phenylalanine at the equivalent position to Phe65 (Phe62) but has a methionine rather than a phenylalanine at the second position (Met64). Like in the THAS1 structure this residue does not have enough electron density to model confidently.

HYS and THAS1 present a very similar tertiary structure (fig. 58). Alignment of the apo HYS with apo THAS1 gave an RMSD value of just 1.48. However, HYS has many active site differences to THAS1, most notably the loop2 over the active site is extended in HYS (D125-GHFGNN-F132) compared to THAS1 (D128-SN-Y131). This loop2 in the crystal structure of HYS appears to dip down and frame the active site from above, but the loop in THAS1 is shorter and therefore does not drape down into the active site.

AJM is a minor byproduct of the THAS enzymes discovered thus far. Extensive mutagenesis has not been attempted on any THAS homolog with the exception of THAS1. Mutations to the active site of THAS1 did not result in a significant AJM production with the exception of the E59 mutants (see section 3.2.7). Mutation of the E59 could result in the cofactor binding in a slightly different orientation in the active site resulting in more flexibility in the iminium formed (fig. 67).

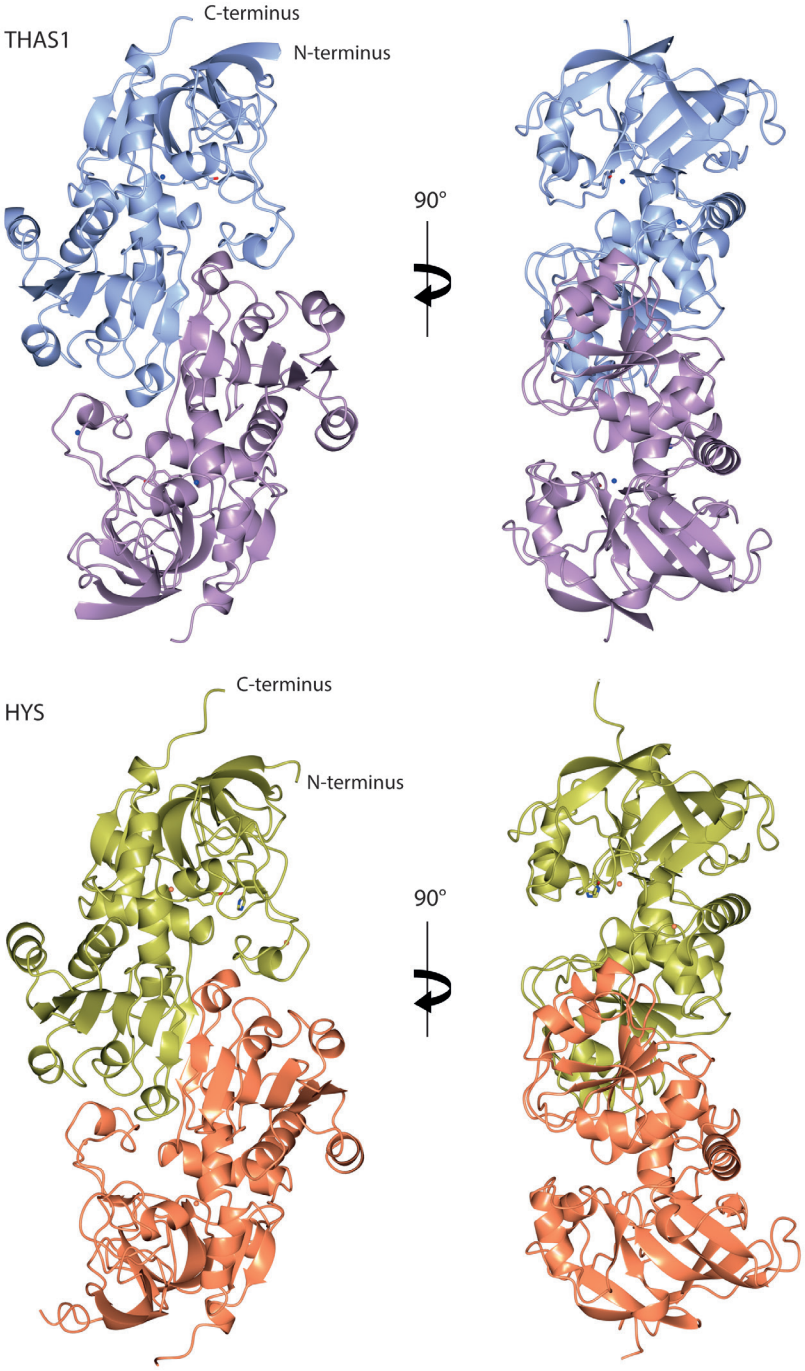


Figure 58: Overview of THAS1 (top) and HYS (bottom) crystal structures.

Two Zn^{2+} ions are found in each subunit of the enzymes, one above the active site termed “active site” zinc, and one more distal to the active site that is a “structural” zinc (fig. 59). In both THAS1 and HYS the active site zinc is coordinated by two cysteines, one histidine, and one glutamate. The structural zinc is coordinated in both structures by four cysteine residues.

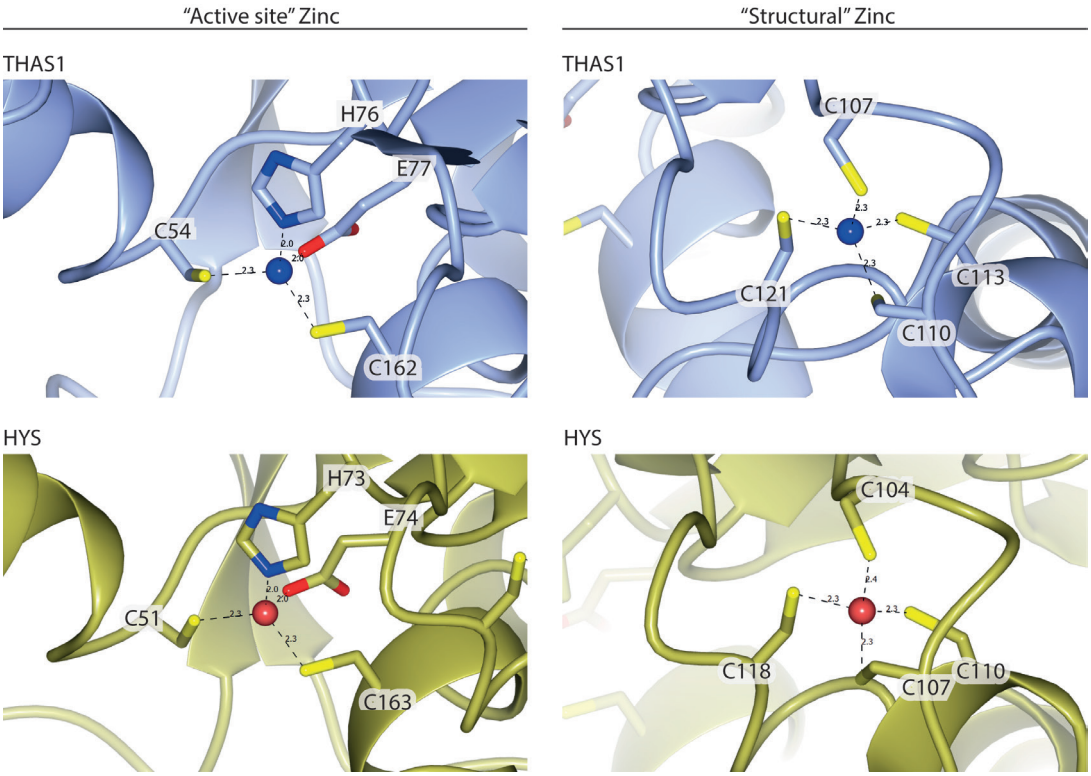


Figure 59: Zinc coordination spheres of THAS1 (upper panels) and HYS (lower panels). The amino acids coordination the "active site" and the "structural" zinc ions are illustrated as cylinders and labelled. Zinc ions are illustrated in blue for THAS1 and in coral for HYS.

3.2.4 Enzymatic labelling of products with ^{2}H

To investigate the mechanism of THAS1 and HYS, a pro-*R*-NADPD (deuterated NADPH) was used in large-scale reactions with the strictosidine aglycon substrate. The reaction products (AJM, THA, and 19-EA from the HYS reaction and THA from the THAS1 reaction) were purified by preparative TLC and analysed by ^{1}H-NMR and compared to the spectra obtained for authentic standards (fig. 60-62). In all cases one hydrogen atom was replaced by a deuterium and the effect was evident on both the spectra of H-21β and on the spectra of H-20. All the purified products lacked a signal for the H-21α consistent with this being deuterated (fig. 63).

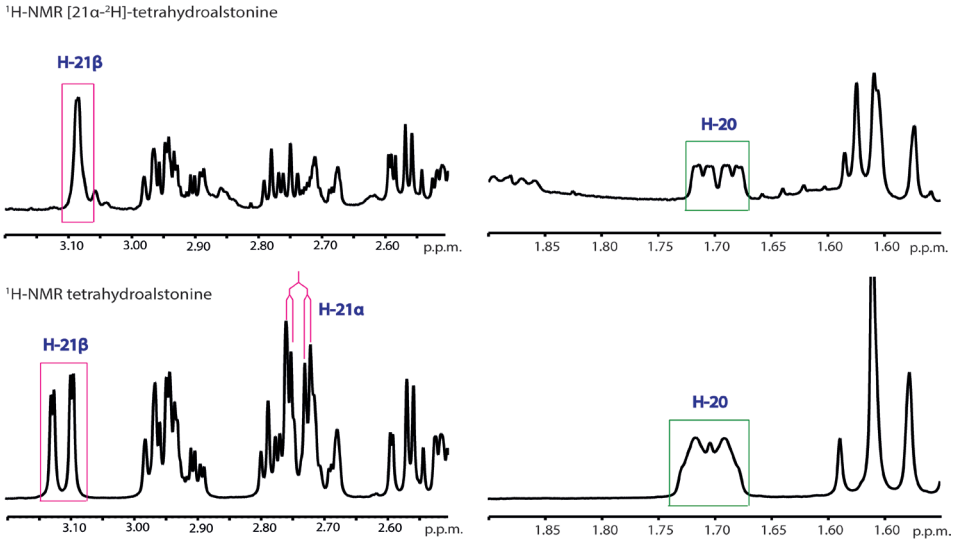


Figure 60: ^{1}H-NMR spectra of [21α-^{2}H]-THAS1 and HYS products. H-21 and H-20 are boxed in pink and green, respectively, for clarity.

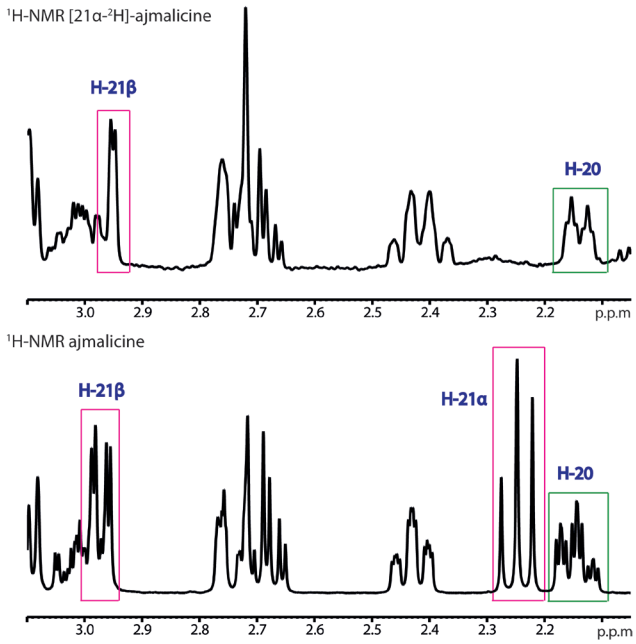


Figure 61: ^{1}H-NMR spectra of [21α-^{2}H]-ajmalicine from large-scale HYS reaction. H-21 and H-20 are boxed in pink and green, respectively, for clarity.

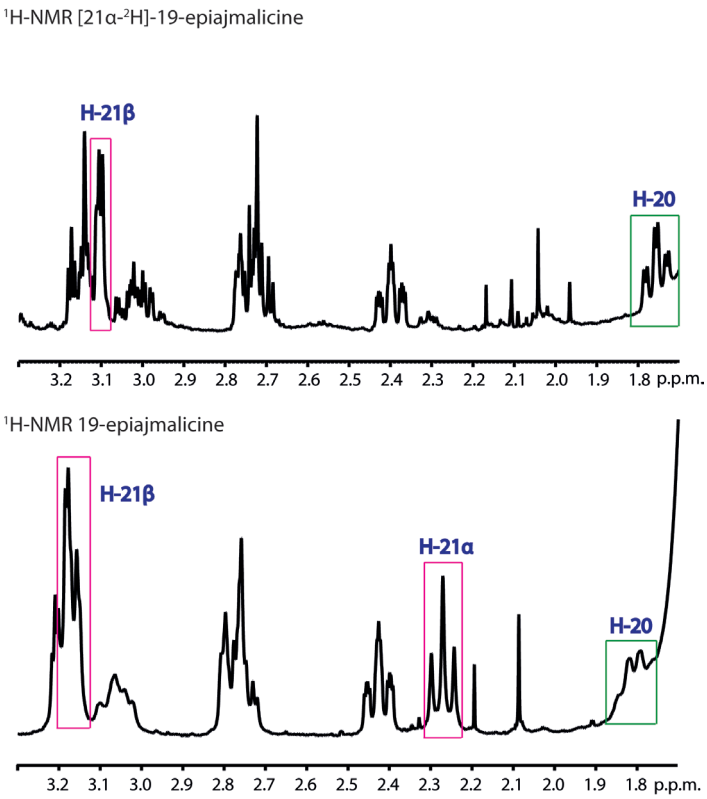


Figure 62: [1]H-NMR spectra of [21α-[2]H]-19-epiajmalicine from large-scale HYS reaction. H-21 and H-20 are boxed in pink and green, respectively, for clarity.

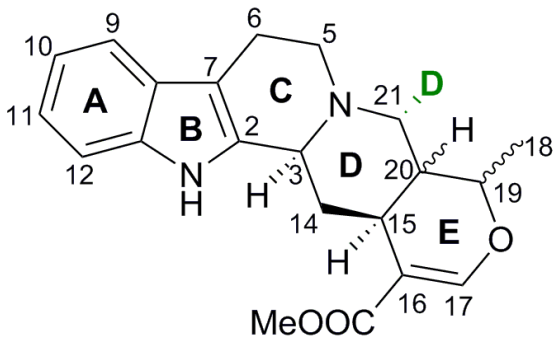


Figure 63: Heteroyohimbines with deuterium at α-position on carbon 21.

Deuterium labelling has shown that the hydrogen from NADPH is transferred to the α-position of all the substrates' C-21. This is in agreement with the hydride transfer described using crude enzyme purifications from *C. roseus* (Stoeckigt et al., 1983). Therefore the substrate must orient itself in the same general manner in the active site of both enzymes.

3.2.5 pH effect on product profile

A small panel of different pH buffers was used to test whether it is possible to modify the product profile of HYS and THAS1 in a pH-dependent manner. The enzyme profiles were compared to the product profile of chemical reduction with NaBH4 (fig. 64). Enzyme profiles were monitored by a highly optimised LCMS method that separated all diastereomers. The results indicate that chemical reduction at all four pH values produces primarily THA, and only trace amounts

(< 0.5 µg) of AJM and 19-EA (fig. 64, panels A, B, C, F). THAS1 has a product profile similar to that of chemical reduction. Both THAS1 and chemical reduction were able to produce larger amounts of THA at the lower pH values. HYS however was able to produce more AJM at all the pH values tested and produced a maximum of 2.5 µg of AJM at pH 6. HYS was also able to produce larger amounts of 19-EA than the other methods but the 19-EA amounts produced were lower than for the other compounds.

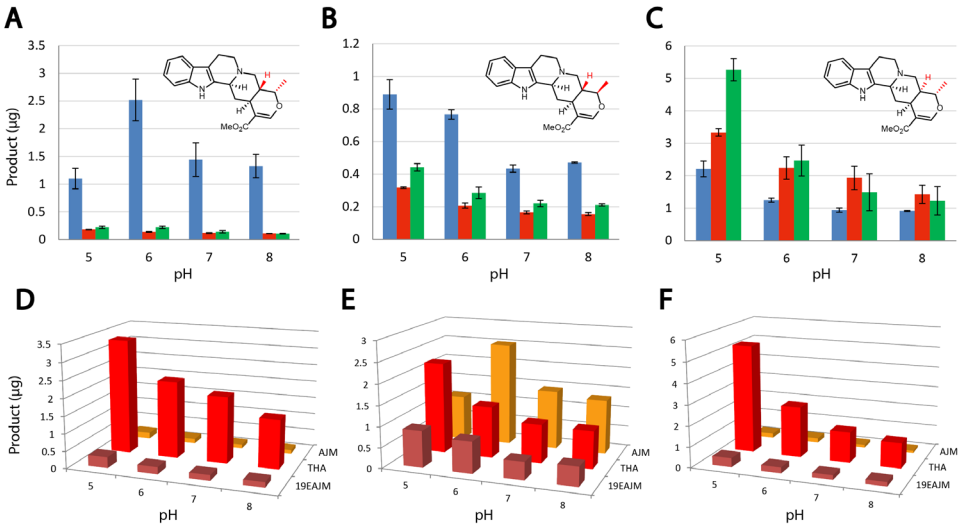


Figure 64: Product profile of THAS1, HYS, and $NaBH_4$ dependence on pH from 5 to 8. Top row: Ajmalicine (A), 19-epiajmalicine (B), and tetrahydroalstonine (C) produced by THAS1 (red), HYS (blue), or $NaBH_4$ (green) per 15.9 µg of strictosidine in 10 min. Bottom row: 3D graph of products of THAS1 (D), HYS (E), and NaBH4 (F). Error bars are the standard error of the mean of three replicates.

It is interesting that HYS is capable of producing a larger amount of AJM than other compounds at pH 6 (fig. 64, panel E). This suggests an amino acid in the active site of HYS is in the catalytically-active conformation at pH 6. His127 on loop2 has a pKa near the AJM-producing maximum of HYS (pH 6), however in the absence of detailed pH dependent kinetics it is not possible to assign with certainty that His127 is responsible for the increased AJM activity at this pH.

3.2.6 Docking of cathenamine into THAS1 active site

Co-crystallisation of THAS1 with THA had been attempted, but no crystals were ever obtained in those conditions. Soaking of some crystals with deglycosylated strictosidine had been attempted but those crystals did not show any density in the active site that would suggest the substrate had entered. Therefore, molecular *in silico* docking was used to visualise the position of strictosidine aglycon in THAS1. Cathenamine was used for the docking as collaborators identified that as the predominant strictosidine aglycon isomer in solution by using ^{15}N-strictosidine aglycon and ^{15}N-HMBC NMR as described in Stavrinides et al. (2016).

Docking was done using the THAS1-holo structure with 100 runs. This resulted in just two distinct docking clusters, which constituted 98% and 2% of the resultant poses, respectively. The latter

cluster placed the indole moiety such that the N1 atom was closest to the cofactor. This cluster was eliminated since it was inconsistent with the results of the deuterium labelling experiments (fig. 63). The poses contained within the major cluster were all deemed to be “productive” since they placed the indole moiety of cathenamine towards the entrance of the active site and the C-20 and C-21 3.3 Å above the nicotinamide C-4 atom (fig. 65). The indole moiety in this second cluster was positioned near the entrance to the active site. The top ranked pose had an estimated free energy of binding = -8.76 kcal mol-1, and is used in the structures illustrated here.

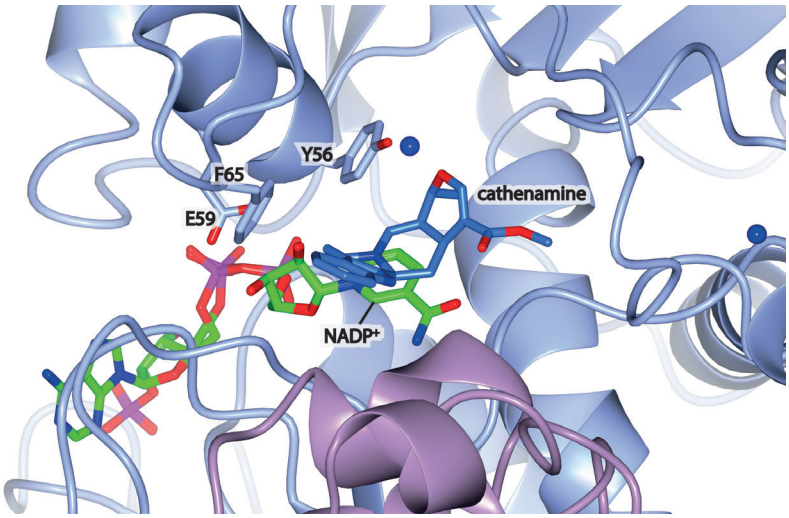


Figure 65: Cathenamine docked in the active site of THAS1. THAS1 is illustrated as cartoon with the amino acids Y56, E59, and F65 illustrated as cylinders; $NADP^+$ is illustrated as cylinders with green carbons; cathenamine is illustrated as cylinders with blue carbons; the Zn^{2+} ions are illustrated as dark blue spheres.

The substrate is positioned in the space between the nicotinamide ring and the active site Tyr56. This orientation is consistent with the labelling studies using NADPD which showed that the hydride from C-4 of the nicotinamide ring of the cofactor is transferred to the bottom face of the C-21 of the substrate to generate the final reduced product. In this orientation of the substrate it is reasonable to suggest that Phe65 or Phe67 could interact with the indole moiety of the substrate and aid in correct orientation in the active site. In HYS the substrate is expected to dock in the same manner, although in silico docking with HYS was not attempted as we were not able to obtain a HYS structure with the cofactor bound. Tyr56 and Phe65 are conserved in HYS, while Phe67 is replaced by a methionine. The role of this residue is not clear but both methionine and phenylalanine are hydrophobic, so this residue may play a similar role in both THAS1 and HYS.

3.2.7 Mutagenesis of THAS1 and HYS

Mutations to the active site of THAS1 were carried out to probe the role of the residues in catalysis of the reduction. Based on sequence homology with PtSAD the equivalent active site amino acids Tyr56 and Glu59 were chosen for mutation. Amino acids which were predicted to frame the active site, Phe65 and Phe67, were also chosen for mutation due to their possible ability of interacting with the hydrophobic indole ring of the substrate. Such an interaction is seen in STR where two aromatic residue side chains sandwich the indole and help in the correct

orientation in the active site (Loris et al., 2007). The amino acid Ser102 was also mutated due its possible equivalence with the PtSAD active site serine.

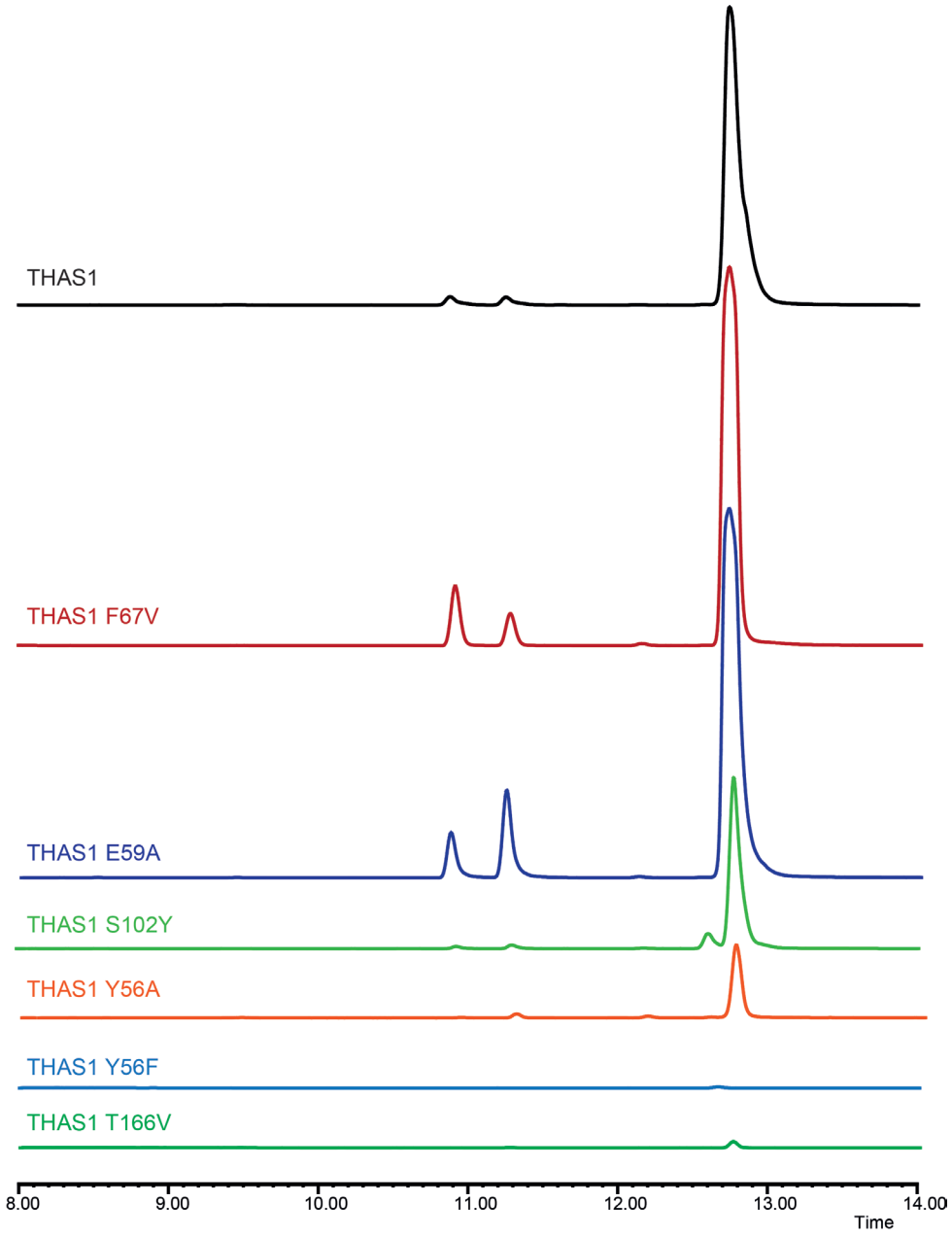


Figure 66: Selected representative traces of THAS1 mutants with interesting effect on the product profile of the enzyme.

Mutant THAS1 were compared by LC-MS analysis both with the "fast method" and with the "separation method" to determine the product ratio. In the fast method, where the heteroyohimbine diastereomers do not separate, the overall activity of the enzyme mutants was measured. The separation method was then used to determine whether the ratio of products had been altered by the mutation (fig. 66). Most mutations did not abolish activity of THAS1, reflecting the resilience and plasticity of the active site (table 7). Some mutations, such as Tyr56 mutated to Phe, resulted in an increase of AJM and 19-EA, but this mutant showed very little total conversion compared to the WT enzyme (0.72%). Mutating Tyr56 to Ser did not impact the total

conversion very substantially (21.86% conversion compared to WT) which demonstrates that an enzyme with an active site more similar to that of PtSAD which has a Ser in place of Tyr could still carry out this reaction. Interestingly, mutating Tyr56 to Ala did not abolish activity though mutation of Tyr56 to Phe did (fig. 66); although the two mutations are aliphatic the smaller residue was much better supported than the larger, more sterically similar Phe. Mutating Tyr56 to Glu resulted in a large reduction in conversion which may be the result of the introduction of a negative charge in the active site cavity.

Ser102, which is positioned in the active site and may be implicated in the mechanism (fig. 67), was mutated as well. Most mutations to this residue (Ala, Tyr, Thr) did not impact the amount of substrate conversion compared to the WT enzyme and did not have any significant effect on the product profile either (table 7). However, mutating Ser102 to Asp resulted in a reduced conversion (2.84%) and an increased promiscuity for AJM production (13.73% of total product formed). Ser102 therefore does not appear to be an important amino acid for the reduction of the substrate but the introduction of a negative charge in the active site disrupted the activity.

Table 7: List of THAS1 mutants that were screened for heteroyohimbine synthase activity against strictosidine aglycon. The total conversion refers to the yield of reduction compared to the starting concentration of strictosidine aglycon. The measured concentrations of the three identified heteroroyohimbines (ajmalicine, 19-epi-ajmalicine, tetrahydroasltonine) define the product ratio; values are rounded to the nearest integer. The activity of all enzymes was compared to activity of THAS1 wild type.

Mutants	Total conversion	Product ratio			Conversion compared to wild type
		AJM	19-EA	THA	
THAS1 Y56A	2%	4%	5%	92%	12%
THAS1 Y56E	0%	0%	29%	71%	0.3%
THAS1 Y56T	1%	8%	8%	84%	5%
THAS1 Y56S	4%	2%	6%	92%	22%
THAS1 Y56F	0%	54%	11%	35%	1%
THAS1 E59A	21%	4%	11%	85%	113%
THAS1 S102A	12%	1%	2%	97%	65%
THAS1 S102Y	6%	2%	2%	96%	29%
THAS1 S102T	13%	2%	1%	97%	68%
THAS1 S102D	1%	14%	4%	82%	3%
THAS1 T166S	1%	13%	5%	82%	3%
THAS1 T166V	0%	24%	7%	69%	2%
THAS1 F67V	19%	6%	6%	87%	100%

Mutating the Glu59 (which coordinates with the ribose OH groups of the cofactor, fig. 67) to alanine does not abolish activity, but slightly affects it (fig. 66). An increase in 19-EA biosynthesis is observed with this mutant compared to WT THAS1.

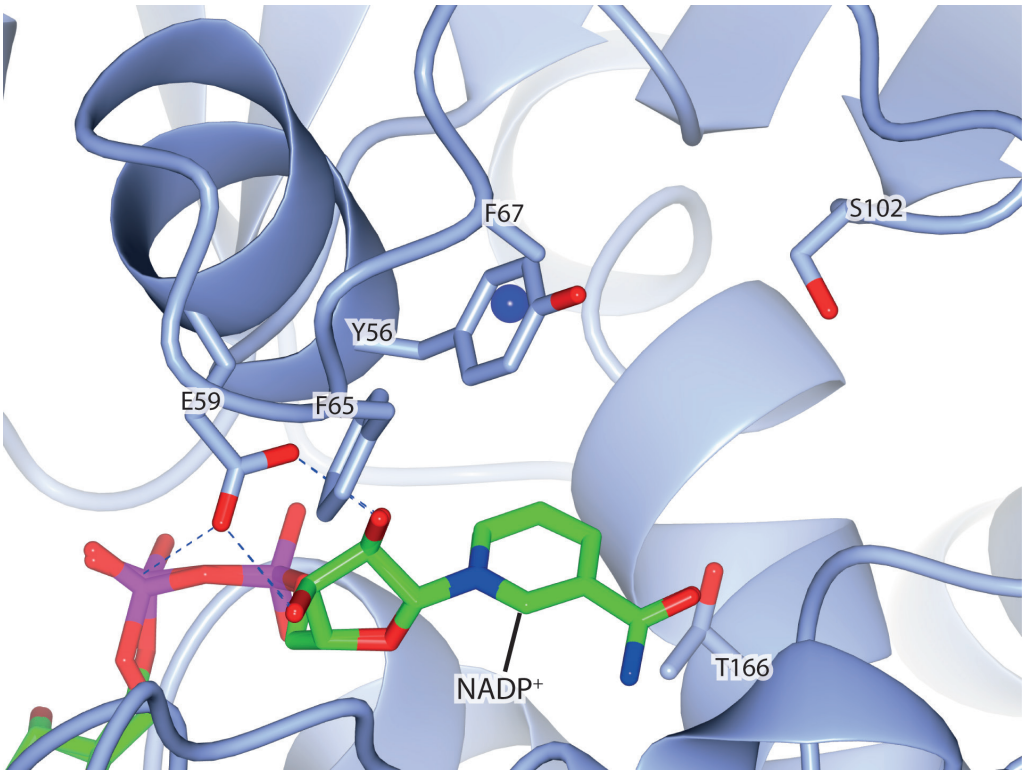


Figure 67: Mutations to THAS1 active site. Crystal structure of THAS1 with side chains of mutated residues illustrated in grey carbons and $NADP^+$ in green carbons. The side chain of Phe67 is not resolved well in the crystal structure beyond Cβ. The active site Zn^{2+} is pictured as a blue sphere and the hydrogen bonds of Glu59 illustrated by dashed lines.

Thr166 is positioned towards the back wall of the active site cavity, below the plane of the nicotinamide ring (fig. 67). This residue could potentially coordinate a water molecule which could serve as a source of protons for protonation of C-20 of cathenamine. This residue was mutated to valine and serine to probe its role in the catalysis. Both mutations reduced activity and increased the promiscuity. Notably, Thr166 to valine increased the amount of AJM and 19-EA produced compared to WT, suggesting that this residue could be coordinating a water molecule in the right position to carry out the protonation of cathenamine. However, drawing firm conclusions would only be possible after thorough investigation of the kinetic properties of this mutant, which was not done due to the complexities of this two-enzyme system, as described in Chapter 2.

The phenylalanine residues (Phe65 and Phe67, fig. 67) on loop1 were both mutated to valine. Only Phe67 is listed in table 7 as the Phe65 to valine mutant did not express well and precipitated during purification. The Phe67Val mutant had poor expression as well but remained soluble and was therefore possible to assay. Mutating Phe67 to valine did not have a negative effect on activity but did slightly increase the product promiscuity of THAS1. This could be either because

of modification of the secondary structure of the loop1, but could also signify that Phe67 may play only a minor role in correct orientation of the substrate into the active site.

3.2.8 Loop swaps and mutations to HYS loop2

The major structural difference between THAS1 and HYS is the extended loop over the HYS active site, loop2 (fig. 68). Loop1, which is near the active site and frames the entrance to the active site on the left side, also exhibits some differences between the two enzymes (THAS1: Lys62-NKFGFT-Ser69; HYS: Thr59-NKFGMT-Lys66). To investigate whether either of these loops control product stereoselectivity, and whether it is possible to switch stereoselectivity between THAS1 and HYS, the loops 1 and 2 were swapped between these two enzymes.

When THAS1 loop1 was replaced with loop1 from HYS, the resulting mutant was not expressed as well as wild type enzymes. The same was observed when Phe67 on loop1 was mutated individually in THAS1. This indicates that mutations to this section of THAS1 are not very well tolerated. When the loop2 from HYS was inserted into THAS1, the resulting mutant did not show an increase in AJM production and neither did the double loop (loop1 + loop2) swap mutant. HYS with the THAS1 loop1 swap did not show a difference in product profile, but its activity was negatively impacted. Conversely, when the loop2 of HYS was replaced with the short loop2 found in THAS1, the resulting mutant HYS enzyme produced a product profile resembling that of THAS enzymes (fig 68, table 8).

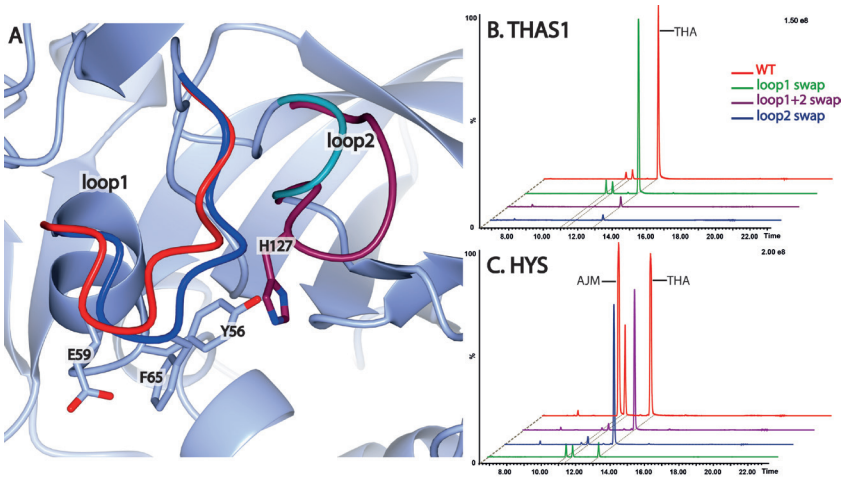


Figure 68: Product profile of THAS1 and HYS loop swap mutants. A: Crystal structure of THAS1 chain A illustrated in gold cartoon representation with Y56, E59 and F65 illustrated as sticks for reference. THAS1 and HYS loops 1 (blue or red) and loops 2 (light blue or purple) are illustrated for clarity. B: UPLC-MS chromatograms of THAS1 loop swap mutants; C: chromatograms of HYS loop swap mutants. Legend for colours used is inset in panel B.

The striking result of the loop2 swap in HYS indicates that loop2 is responsible for the increase in promiscuity of HYS. Circular dichroism of the WT enzymes and the loop swap mutants was done to verify that the swaps did not cause a significant modification of the tertiary structure of the enzymes (fig. 69). The CD spectra of all enzymes have a similar profile, however there are some

slight modifications to the THAS1 structure, but there do not appear to be major changes to the structure as a result of the loop swaps.

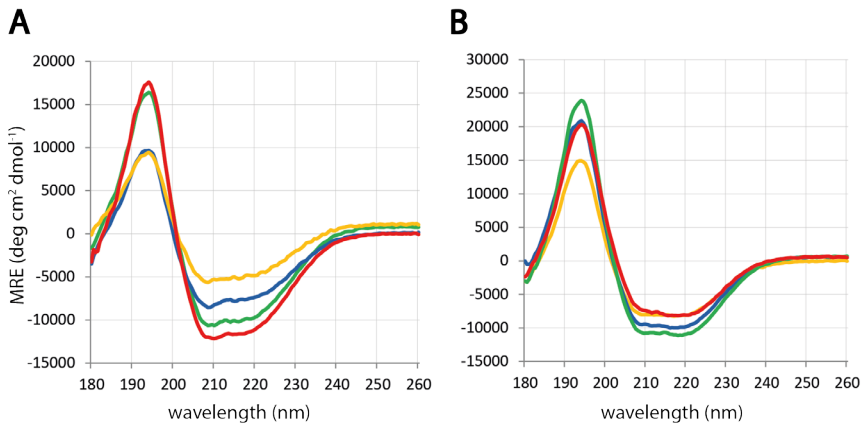


Figure 69: Circular dichroism of THAS1, HYS, and the corresponding loop swap mutants. A. WT and loop swap mutants of THAS1. B. WT and loop swap mutants of HYS. Wild-type enzymes are plotted in red, loop 1 swap mutants in green, loop 2 swap mutants are in blue and the double loop 1 and 2 swap mutants are in yellow. The secondary structure of the mutants is not substantially different from that of the wild type proteins.

Table 8: List of THAS1 and HYS loop swap mutants screened for heteroyohimbine synthase activity against strictosidine aglycon. The total conversion refers to the yield of reduction compared to the starting concentration of strictosidine aglycon. The measured concentrations of the three identified heteroroyohimbines (AJM, 19-EA, THA) define the product ratio; values are rounded to the nearest integer. The activity of all enzymes was compared to activity of the corresponding wild type.

Mutants	Total conversion	Product ratio			Conversion compared to wild type
		AJM	19-EA	THA	
THAS1 loop1 swap	11%	4%	5%	91%	59%
THAS1 loop2 swap	0.5%	19%	9%	72%	2%
THAS1 loop1 + loop2 swap	1%	12%	8%	80%	3%
HYS loop1 swap	2%	26%	33%	41%	9%
HYS loop2 swap	10%	2%	5%	93%	43%
HYS loop1 + loop2 swap	1%	20%	58%	22%	3%

3.2.9 Point mutations of HYS

Mutations were made to His127 to investigate its role in reduction of cathenamine, an isomer of strictosidine aglycon. Phe128 was also mutated to investigate whether this amino acid could be responsible for promoting a different binding mode of the substrate to the active site which would give rise to the pro-ajmalicine iminium species. However, this seemed unlikely, since deuterium labelling studies described above suggest that the substrate binds in the same position and is

labeled at the same carbon regardless of whether the THA or the AJM product is formed. The Phe128Ala and Phe128Tyr mutants did not show a modified product profile but rather showed a decrease in product formed. The His127Ala and His127Asn mutants both displayed a product profile that resembles that of THAS (fig. 70, table 9) which supports the hypothesis that AJM is formed due to the presence of this histidine on loop2.

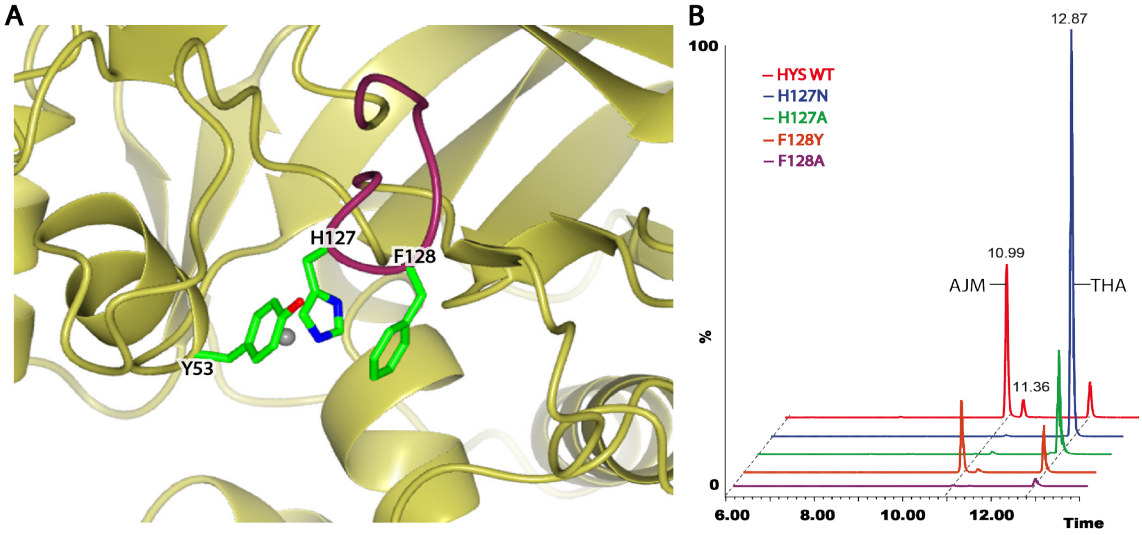


Figure 70: Results of point mutations on HYS active site amino acids. A: crystal structure of HYS chain B, active site cavity with H127 and F128 drawn in stick representation and the active site tyrosine (Y53) for reference (green carbons). Loop 2 is illustrated in purple, grey sphere is the “active site” zinc. B: UPLC-MS chromatograms of the point mutations of HYS (all traces are 1.16e8). Ajmalicine elutes at 10.99 min, 19-epi ajmalicine at 11.36, and tetrahydroalstonine at 12.87.

Table 9: List of HYS point mutants that were screened for heteroyohimbine synthase activity against strictosidine aglycon. The total conversion refers to the yield of reduction compared to the starting concentration of strictosidine aglycon. The measured concentrations of the three identified heteroroyohimbines (ajmalicine, 19-epi-ajmalicine, tetrahydroalstonine) define the product ratio; values are rounded to the nearest integer. The activity of all enzymes was compared to activity of HYS wild type.

Mutants	Total conversion	Product ratio			Conversion compared to wild type
		AJM	19-EA	THA	
HYS H127A	1%	3%	4%	93%	4%
HYS H127N	2%	1%	1%	98%	11%
HYS F128A	0%	34%	8%	57%	0.5%
HYS F128Y	0.5%	48%	6%	46%	3%

3.2.10 Mechanism proposal for tetrahydroalstonine biosynthesis in THAS enzymes

The cathenamine substrate most likely binds into the active site with the indole moiety oriented towards the phenylalanine residues on loop1 and above the NADPH nicotinamide ring (fig. 65). We propose equilibration to the iminium species is achieved by a proton addition at C-20 from

a water molecule positioned behind and below the molecule. This produces an (*S*) orientation at C-20. The crystal structure of THAS1 displays many water molecules in the active site (fig. 56); therefore it is not unlikely that a water molecule coordinated in the active site could be responsible for the iminium generation. The only amino acid residue that is positioned to possibly perform this protonation is Thr166, which is positioned towards the back wall of the active site cavity, in proximity to the location of the substrate C-20. Mutation of Thr166 to a valine did not show any significant reduction effect to the product profile. Therefore it is likely that a water molecule and not Thr166 could be responsible for providing the proton.

Once the pro-THA iminium is formed, the hydride from the pro-*R* face of NADPH attacks the C-21 to produce THA. The active site tyrosine can stabilise the transition state through the N-4 (fig. 71). Mutation of the THAS1 Tyr56 to a phenylalanine resulted in a much reduced activity which supports this hypothesis.

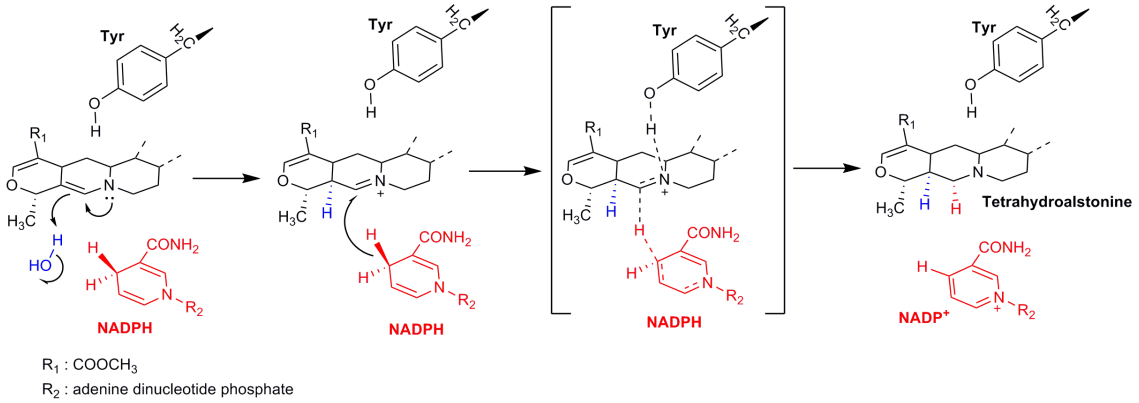


Figure 71: Reduction mechanism of Tetrahydroalstonine Synthases and Heteroyohimbine Synthase for production of tetrahydroalstonine. The substrate, cathenamine, is shown with the indole moiety truncated for clarity. The cofactor and the hydride originating from it are illustrated in red and the hydride originating from water (on C-20) is illustrated in blue.

3.2.11 Mechanism proposal for ajmalicine biosynthesis in HYS

Based on the deuterium labelling results cathenamine binds to HYS with the C-20 to C-21 bond in proximity to the NADPH cofactor, in the same orientation as it does for THA production. I hypothesize that generation of the pro-AJM iminium form of the substrate is achieved by abstraction of an exchangeable proton of the His127 side chain by C-20 at the pro-*R* position and corresponding tautomerisation to the iminium. His127 is coordinated in the crystal structure by the active site Tyr53 phenolic group. Once the pro-AJM iminium is formed the reaction proceeds in the same fashion as THAS. The NADPH hydride attacks the C-21 at the pro-*R* position and the transition state is stabilised by the Tyr53 hydroxyl (fig. 72).

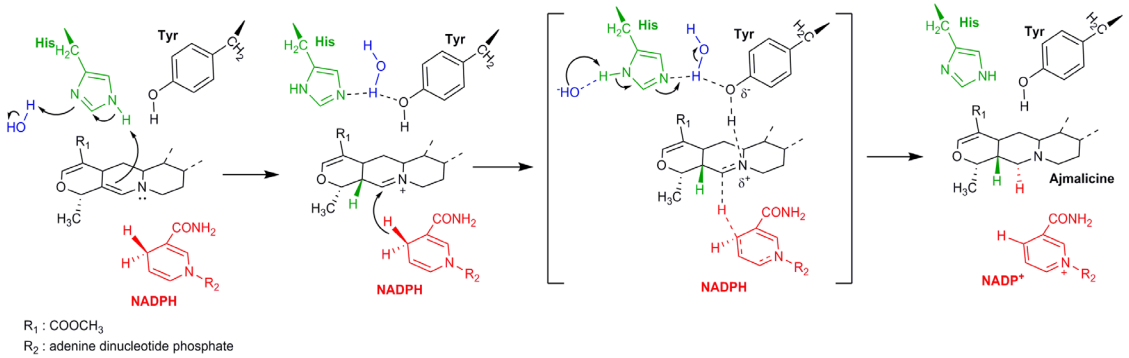


Figure 72: Reduction mechanism of Heteroyohimbine Synthase for production of ajmalicine. The substrate, cathenamine, is shown with the indole moiety truncated for clarity. The active site His127 and the hydride originating from it are illustrated in green. The cofactor is illustrated in red and the hydride originating from water (on C-20) is illustrated in blue.

3.2.12 Removal of zinc and functional assays

MDRs usually contain two zinc ions per subunit (Auld and Bergman, 2008), one “structural”, found in a flexible loop area near the surface of the MDRs, and one “active site” zinc found near the active site cavity. The latter is implicated in the catalytic mechanism in some MDRs and is therefore termed “catalytic” in those MDRs. However, in the HYSs there is no evidence that the zinc is necessary for catalysis and is therefore termed simply “active site” zinc in this thesis. The “structural” zinc is coordinated by four cysteine residues and the “active site” zinc in most MDRs is coordinated by two cysteines and one histidine with the fourth position usually occupied by a water molecule. In the case of the crystallised heteroyohimbine synthases, THAS1, THAS2 (performed by a collaborator and not described in this thesis), and HYS, the “active site” zinc is positioned further back in the active site and is coordinated in the fourth position by a glutamate residue (fig. 59).

To determine whether zinc is required for HYS activity, THAS1 and HYS were subjected to treatment with the chelator EDTA. THAS1 and HYS did not degrade after the overnight dialysis against EDTA-containing buffer, based on SDS-PAGE analysis which showed the expected molecular weight. However, after the second overnight dialysis to remove the EDTA most of THAS1 had precipitated. HYS did not show any degradation. The smell of sulfur was noted when the dialysis container was opened after the EDTA dialysis, as noticed by Hoagstrom et al. (1969). The removal of the zinc was verified visually by reacting an aliquot of Zn-holo and Zn-apo protein with a commercial zinc detection kit. The assay did not show any difference when the proteins were incubated “as is” and therefore the enzymes were boiled and pelleted before assaying again (fig. 73). The results were clearer with the boiled protein; visual inspection showed that there were very low levels of Zinc in the Zn-apo conditions, confirming that the EDTA dialysis was successful.

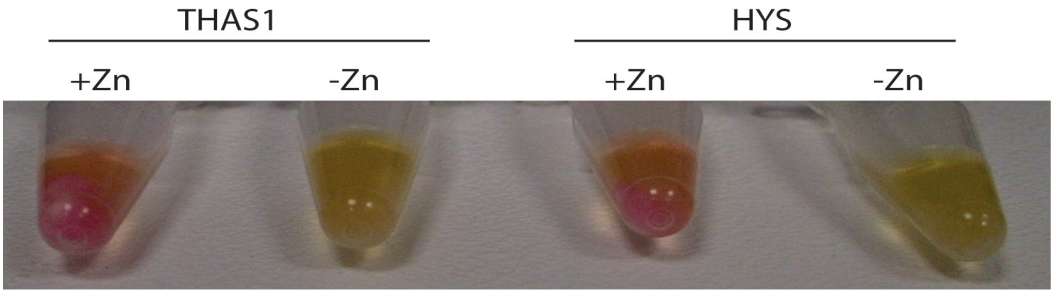


Figure 73: Photos of zinc detection assay conducted on Zn-apo and Zn-holo THAS1 and HYS. The Zinc detection assay was done after boiling of the protein. The pellets of protein in the +Zn conditions show strong pink colouring whereas the pellets in the –Zn conditions are not pink.

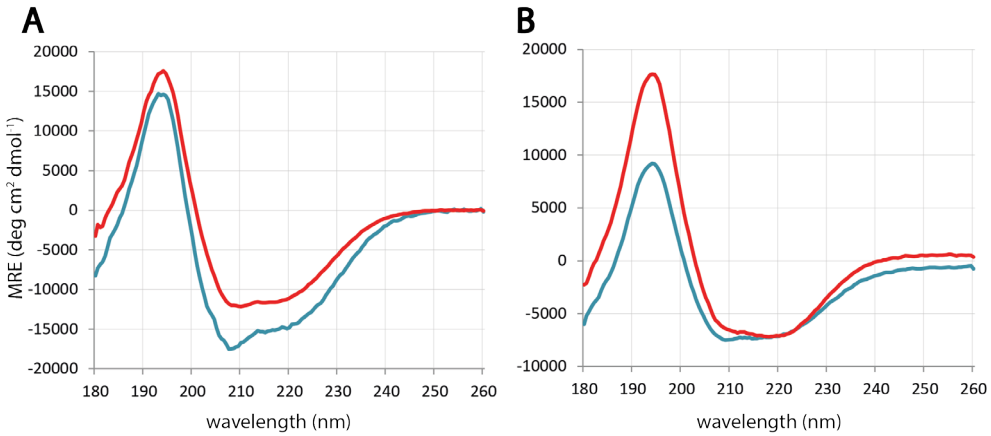


Figure 74: Circular dichroism spectra of non-treated (red) and zinc-apo (blue) THAS1 (A) and HYS (B). The secondary structure of the Zn-apo proteins is not significantly different from that of the Zn-holo proteins.

After recuperating the soluble fraction of each protein an SDS-PAGE gel analysis did not indicate degradation in either protein. CD-spectra analysis of both proteins did not suggest the soluble proteins had substantially changed in structure (fig 74). This suggests that the zinc ions are not necessary for conservation of the tertiary or quaternary structure of these enzymes.

Analysis of protein assays by UPLC-MS using the “fast method” showed that the Zn-apo enzymes were able to catalyse the reaction (fig. 75). However, the slight decrease of conversion in the THAS1-Zn-apo assays means the result of this experiment is not clear-cut. It is not possible to exclude the presence of trace amounts of “active site” zinc left over in some active sites of THAS1. It is also not possible to rule out the possibility that the EDTA treatment of THAS1 resulted in a slightly misfolded protein (as evidenced by the high protein precipitation after overnight dialysis) which possibly was not as catalytically active.

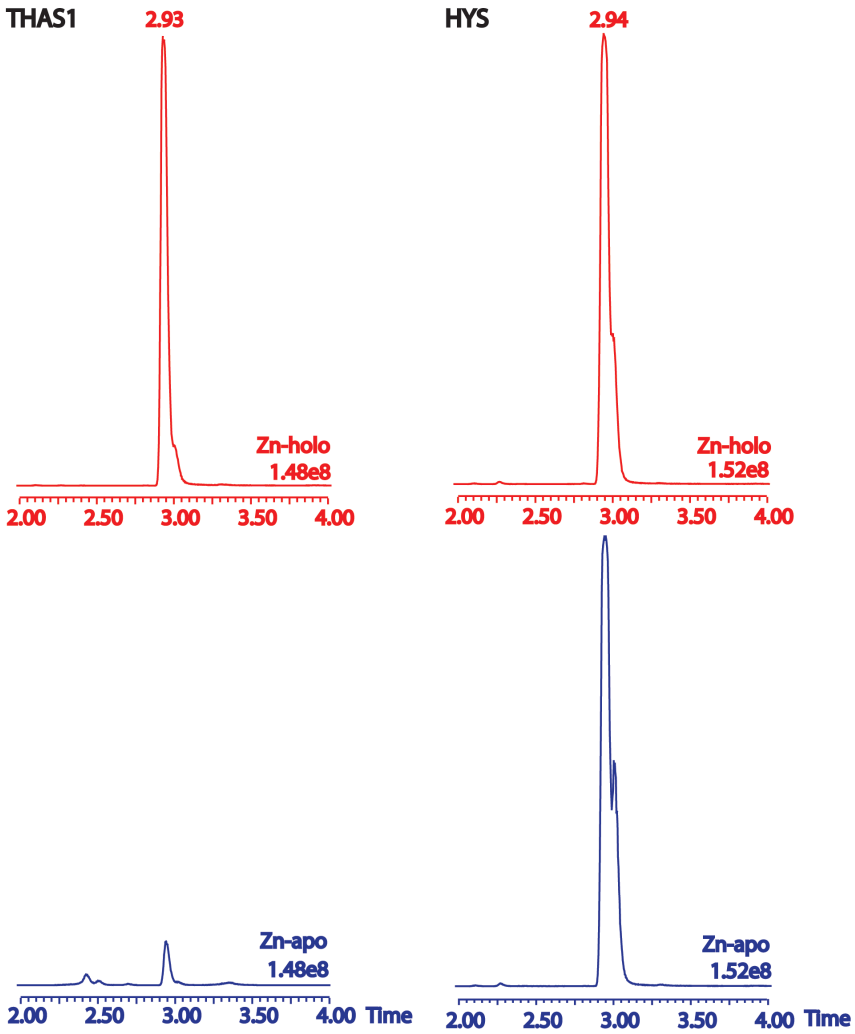


Figure 75: LC-MS chromatogram traces of THAS1-Zn-apo and HYS-Zn-apo protein assays compared to untreated control reactions.

Although these results strongly suggest that the "active site" zinc does not play a role in the mechanism detailed here, the catalytic role of zinc cannot be definitively eliminated. The distance of the zinc from the bound cofactor in the active site ($NADP^+$) is not a clear indication of its role in catalysis, as indicated by studies done on other members of the MDR family. For example, crystallisation of a Glucose Dehydrogenase from an archaeon with different combinations of substrates and products revealed that the zinc atom could swap ligands and could be displaced by as much as 1.4 Å within the active site cavity (Baker et al., 2009).

3.3 Discussion

3.3.1 Protein crystal structures

Crystallisation of THAS1 and HYS was carried out to gain a better understanding of the structure of the proteins and in particular of the active site. THAS1 crystallised well in the holo form and quite readily in the apo form. As with many other MDR enzymes, crystals in the holo form were of better resolution than the apo crystals, perhaps indicating some inherent instability of the apo enzymes, which present a more open structure between the substrate-binding and the cofactor-

binding domains. This is consistent with the Isothermic Titration Calorimetry (ITC) results for THAS1 (Chapter 2, Results 2.2.6) which indicated that the binding of the cofactor to the active site of THAS1 is exothermic.

The crystal structure of the THAS1-holo enzyme was initially going to be solved by using molecular replacement with the crystal structure of PtSAD (PDB: 1YQD (Bomati and Noel, 2005)). However, due to the very strong signal by the 4 Zn^{2+} ions contained in each enzyme dimer the anomalous signal was used successfully to generate an experimentally phased density map. The high resolution of the crystal electron density allowed for a high resolution and high confidence protein structure to be built in it.

Although the thermofluor analysis of HYS showed it was a relatively stable enzyme with a high melting temperature, especially in the presence of NADPH, it proved difficult to crystallise. An apo structure of HYS was obtained which provided the necessary resolution and details to compare it directly to the crystal structures of THAS1, along with another THAS homologue solved by a collaborator, THAS2. A holo crystal of HYS with $NADP^+$ could not be obtained.

The structures of these *C. roseus* enzymes are similar, and also similar to the structures of other MDRs. The most notable differences between them are the loop2 sections above the active sites. These loops are variable in length between THAS1 and HYS and other THAS homologues (fig. 76).

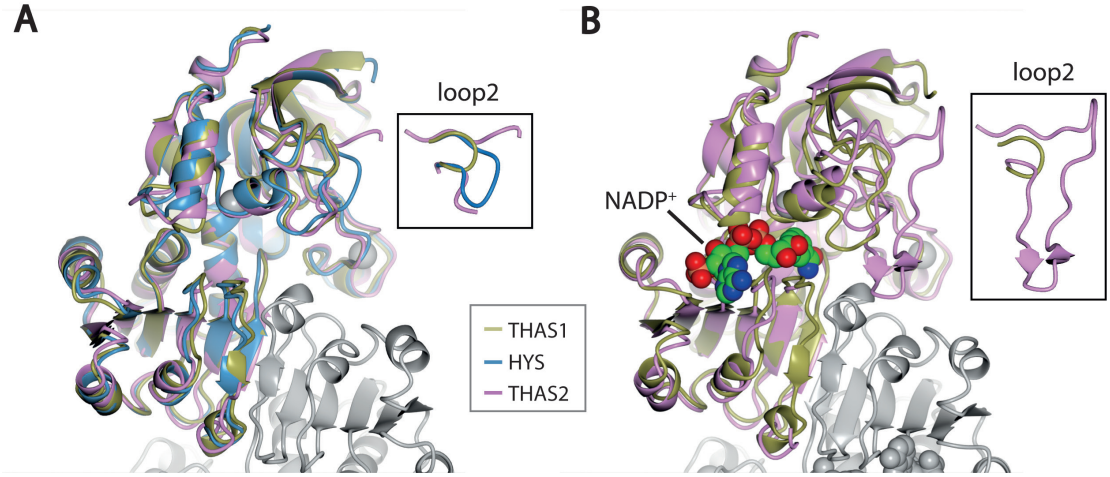


Figure 76: Crystal structures of THAS1, THAS2, and HYS with the loop2 sections highlighted. A: apo THAS1, THAS2, and HYS displayed in cartoon, with the loop2 section inset in black box.B: holo THAS1 and THAS2 displayed in cartoon, with the loop2 section inset in black box.

THAS1 possesses a very short loop2, HYS an intermediate length, and THAS2 a very long loop2 which covers the active site entrance. The crystal structures confirmed that a tyrosine is positioned in the active site in an equivalent position to the serine or threonine that is necessary for SAD or CAD enzymes to catalyse reduction of their substrate.

3.3.2 Steric course of hydrogen transfer

Strictosidine aglycon is present as a mixture of isomers in solution due to its reactive nature and inherent instability (see fig. 7 in Chapter 1). In solution the predominant species is cathenamine, as illustrated by collaborators, which is an enamine containing a double bond between C-20 and C-21. Enzymatic labelling of the products of these enzymes was the most reliable technique to identify which hydrogen is originating from the cofactor.

Deuterium labelling with NADPD and subsequent characterisation of the enzymatic products showed that in all cases the hydride from NADPH is transferred to C-21 from the bottom face of the substrate to yield the product. Labelling studies using NADPD conducted by Stöckigt et al. (1980) and Stoeckigt et al. (1983) with crude protein extracts from *C. roseus* cell suspension cultures indicated that the hydride from the cofactor is transferred onto the α-position at C-21 of the substrate, which is in agreement with the results obtained here. In their experiments, the workers used a crude enzyme preparation from an unpublished cell suspension line; therefore it is not possible to verify which heteroyohimbine synthases these authors were working with. It is also possible that the heteroyohimbine synthases described here were not expressed in the cell line that was used.

This experiment illustrates that the substrate(s) must be oriented in the same manner in both THAS1 and HYS and therefore the enzymes must follow a similar catalytic mechanism. The increased AJM production of HYS cannot be explained simply as the substrate orienting itself differently in the active site. There is no reason to suspect that orientation in other HYSs would be any different to the results presented here for these two enzymes.

In silico docking done using cathenamine and the holo crystal structure of THAS1 were consistent with the deuterium labelling experiments. The cathenamine was docked in the active site cavity in an orientation that brings the C-20-C-21 bond very close to the cofactor C-4 (which carries the hydride used in reduction). Furthermore, cathenamine was docked in an orientation which would result in the hydride being added on the bottom face of the substrate resulting in the same pattern observed with deuterium labelling. We assume that the cofactor reduces the iminium form of the substrate, which is formed by equilibration of cathenamine by addition of a proton to the C-20.

3.3.3 Effect of pH on product profile

The pH of a solution can of course have a large effect on the enzymatic reaction (as seen in Chapter 2, section 2.2.4). On the other hand, the pH could have an effect on the equilibration of strictosidine aglycon to the different isomer heteroyohimbines substrates. Stöckigt et al. (1977) chemically reduced strictosidine aglycon with $NaBH_4$ in MeOH and demonstrated that the major isomer in solution is cathenamine which then gives rise to THA. These conditions did not lead to AJM or 19-EA, suggesting they are not favoured in the non-enzymatic reaction. When, instead, the workers reduced the strictosidine aglycon with a crude enzyme preparation from *C. roseus*

they found AJM and 19-EA being produced as well. They propose the synthesis of AJM and 19-EA is entirely enzyme dependent, perhaps relying on an enzyme-driven isomerisation of cathenamine to 19-epicathenamine for generation of the correct stereocentre at C-19 for 19-EA. To investigate if the pH has an effect on the product profile of HYSs, THAS1 and HYS were tested on a small panel of buffers with different pH values.

The results of the experiment indicated that in total greater amounts of heteroyohimbines can be produced at lower pH (fig. 64). Both the enzyme assays and the control reaction using the chemical reducing agent NaBH4 displayed an increase of heteroyohimbines which was pH dependent. The product profile of THAS1 and of the chemical reduction looked very similar, with predominantly THA being produced. Interestingly, HYS responded differently to the varying pH and produced an increasing amount of AJM at lower pH at the expense of THA. The pH at which this enzyme produced the maximum amount of AJM was pH 6.0.

The absence of this effect in the chemical reduction of strictosidine aglycon suggests this increase in AJM is not due to a modification of the properties or equilibration of the substrates with the change in pH. Instead it suggests that the pH has a significant effect on the enzyme activity and allows HYS to carry out the reduction, or equilibration into pro-AJM iminium, more easily.

3.3.4 Mutations to THAS1

A better understanding of the role of the active site amino acids in THAS1 is needed to propose a reduction mechanism for generation of heteroyohimbines. The crystal structures, as well as alignments to other characterised MDRs, indicated the likely active site of THAS1. The conserved Ser or Thr of most SADs and CADs was not present in the HYSs discovered from *C. roseus*, and instead were replaced by other, often hydrophobic, residues. In THAS1 the active site residue in this position is Tyr56. A serine is present in the active site of THAS1 but it is positioned on the opposite side of the active site compared to Tyr56 or the SAD and CAD serine or threonine catalytic residues. Another notable difference between the SAD/CADs and HYSs is the histidine in SAD/CAD that is coordinated to the cofactor ribose and responsible for the electron transfer during catalysis. This residue in HYSs is replaced by an acidic amino acid, such as glutamate in THAS1 and HYS and aspartate in THAS2. The entrance of the active site also has two hydrophobic residues in THAS1 (Phe65 and Phe67) which were intriguing given the hydrophobic nature of the substrate and the product.

Mutating Tyr56 has some effect on THAS1 activity. Mutating it to Phe greatly reduces the activity and the stereoselectivity indicating that the hydroxyl group of Tyr could be implicated in the reduction mechanism. Introduction of Phe in the active site pocket, assuming the enzyme achieves correct folding and adopts the correct tertiary structure, would increase the hydrophobicity of the environment which could inhibit or modify the correct orientation of the substrate. Mutation of Tyr56 to Ala did not have the same effect on enzyme activity, perhaps due to the less disruptive size of Ala. Its small size would not increase the hydrophobicity of the active site as much as Phe would. By the same logic Y56T or Y56S would favour a more hydrophilic pocket, thus allowing the

substrate to orient and dock correctly, or allowing any water molecules necessary for activity to take up their ideal positions. The activity of these latter two mutants is relatively high, thereby confirming the important role of the hydroxyl group in the THAS1 active site. Mutating Tyr56 to Glu seems to have greatly reduced activity, most probably due to the introduction of a negative charge in the active site. This reduction in activity could also be because the α2 helix which carries both Tyr56 and Glu59 is not anchored in place (in fact it can move out of place in the absence of the cofactor as evidenced by the differences between the holo and apo structures) and the mutated residue could be competing with Glu59 for binding of the cofactor.

Mutating THAS1 Glu59 (which is coordinating with the ribose OH groups of the cofactor, fig. 67) to alanine does not abolish activity, but decreases it and also causes the product profile to be less specific. This could be because the cofactor does not dock correctly into the structure when this is mutated to alanine. A residue further back on the helix, Asp55, might compensate, but in so doing this might have an effect on the general structure of the area.

Mutating Ser102 (which lies on the right wall of the active site pocket) affects the activity slightly but does not affect stereoselectivity. Mutating Ser102 to Tyr surprisingly does not have a large effect on the enzyme; this could be due to the fact that there is enough space for such a large residue to swing out of the way of the active site cavity and not interfere with the reaction.

Mutations to the loop1 phenylalanines had detrimental effects on protein stability; one mutant (Phe65 to valine) was not stable in solution and precipitated over the course of 3-4 hours. This residue is conserved in HYS and could be important for the stability of the loop1 section of these enzymes.

Finally, Thr166 is a conserved residue in these HYSs, which is positioned at the bottom of the active site cavity. This residue seems ideally situated to coordinate a water molecule which could be the source for the proton needed during equilibration of cathenamine to the iminium forms. Mutations to this residue did affect the enzyme product profile slightly, but not enough to clearly distinguish the role of this amino acid in catalysis. Detailed kinetic studies would be needed to determine what its role is, but these are not feasible because of the two-enzyme system and the inherent instability and reactivity of strictosidine aglycon.

Although all mutations to the THAS1 active site had some slight effect on the enzyme's product profile and end-point conversion only Tyr56 appears to have a very significant effect. Therefore this residue is proposed to be taking part in the reduction of cathenamine to yield the heteroyohimbines as illustrated in figure 71.

3.3.5 Loop swaps between THAS1 and HYS

Gaining or losing the ability to synthesise one diastereomer is a difficult matter in enzyme engineering. THAS1 and HYS presented an excellent opportunity to attempt to switch stereoselectivities of two enzymes. THAS1, which has a short loop2 over the active site cavity,

produces a simple product profile of primarily THA and only trace amounts of AJM and 19-EA. HYS, however, is able to produce three heteroyohimbines. HYS and THAS1 are very similar enzymes (76.3% identity), but present a large difference in their loop2 sections. To test whether this loop could be the key to unlocking the potential of HYS to producing AJM and 19-EA, the loops were swapped between the two enzymes. Removal of the loop2 from HYS and its replacement with the loop2 of THAS1 resulted in a product profile resembling that of THAS. Unfortunately, replacing the short loops of THAS1 with the loop2 from HYS did not allow THAS1 to produce a product profile like that of HYS. Nevertheless, from these results it is apparent that the loop2 section of HYS is promoting the functional promiscuity of HYS.

Point mutations to HYS loop2 were also carried out and tested enzymatically. HYS possesses one ionisable amino acid on the loop2, His127. This could be providing a proton during equilibration of cathenamine to the pro-AJM or pro-19-EA iminium. Mutating this amino acid, and the phenylalanine next to it, confirmed the results obtained with the loop swaps and demonstrated that His127 is necessary for HYS to produce AJM and 19-EA.

3.3.6 Reduction mechanism of HYSs

The results obtained with the mutational screening and the deuterium labelling, as well as the in silico docking of cathenamine, have allowed us to propose a reduction mechanism for generation of heteroyohimbines by *C. roseus* HYSs. Cathenamine, the most abundant strictosidine aglycon isomer present in solution, enters the active site and stacks above the nicotinamide ring with the indole moiety pointing outwards towards the solvent. The cathenamine equilibrates into the iminium form in one of two ways. The first leads to the pro-THA iminium by protonation of C-20 from a water molecule to generate a tertiary carbon with S stereochemistry (fig. 77 top row). The second way to equilibration involves protonation of C-20 from the other face of the substrate from His127 to generate the tertiary carbon with R stereochemistry (fig. 77 bottom row). Depending on the orientation of the C-18 methyl group the substrate is either cathenamine (leading to AJM and THA) or 19-epicathenamine (leading to 19-EA and rauniticine). There is no evidence that any of the HYS produce rauniticine, therefore it is reasonable to hypothesise that these enzymes are not capable of protonating 19-epicathenamine from the bottom face. Whether or not cathenamine can equilibrate, through ring opening and re-closing, into 19-epicathenamine in the active site of the enzyme is still an open question.

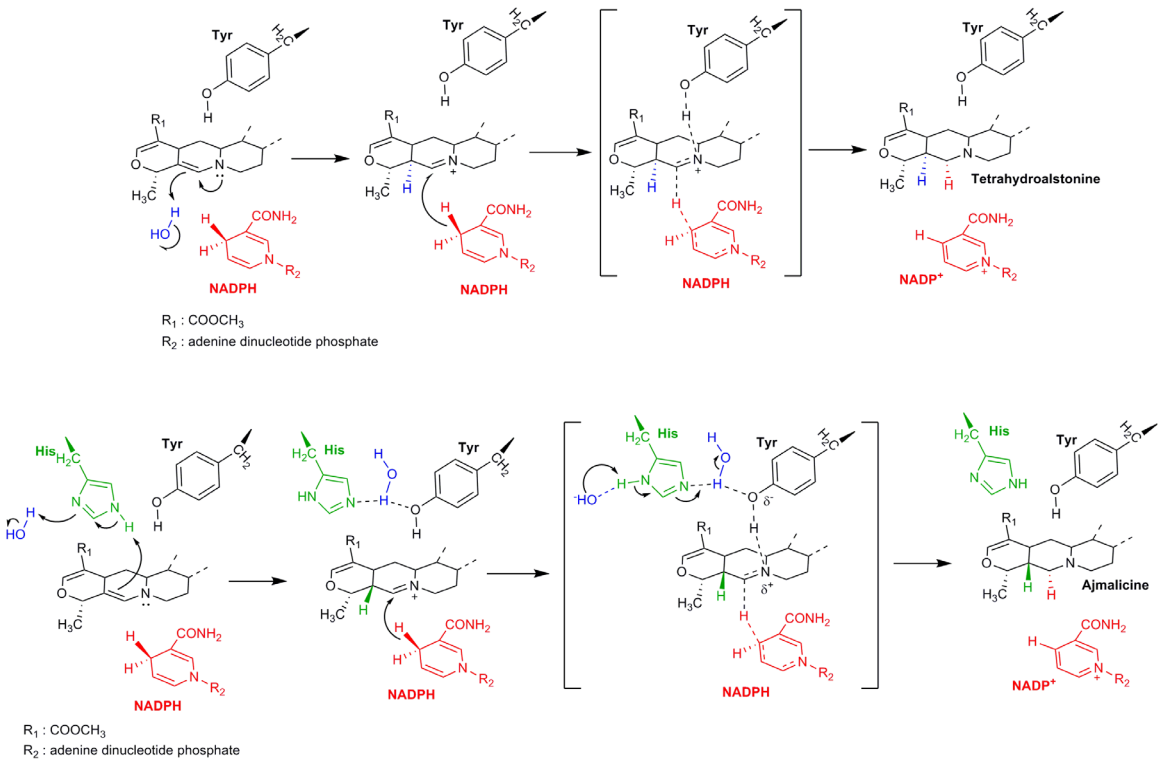


Figure 77: Mechanism proposal for reduction and production of tetrahydroalstonine (top row) and ajmalicine (bottom row). NADPH is illustrated in red; water and the hydrogens originating from it are illustrated in blue; the loop2 histidine residue of HYS and the hydrogen originating from it are illustrated in green.

After the generation of the iminium with the appropriate stereochemistry at C-20, the two reduction mechanisms are carried out in the same manner (fig. 77). The pro-*R*-hydride from the cofactor attacks the C-21 from the bottom face to generate a secondary carbon. The transition state during movement of the electron pair onto the N-4 is stabilised by the partial charges network facilitate by the active site tyrosine. The product thus formed can exit the active site before the enzyme also dissociates with the cofactor.

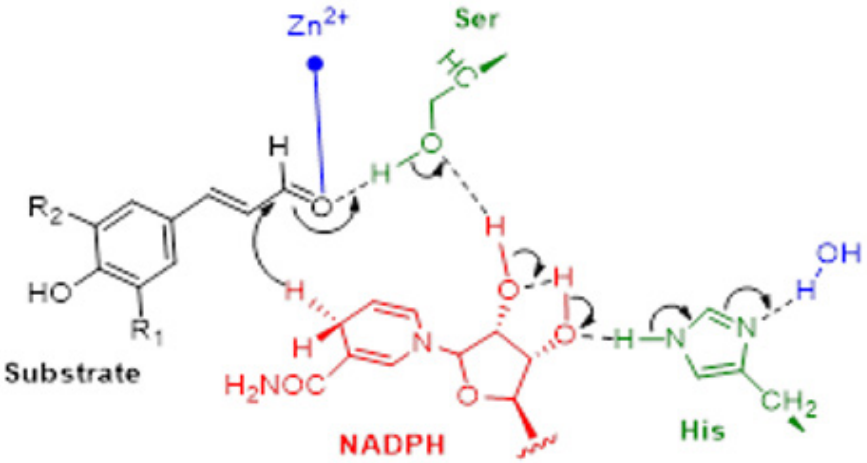


Figure 78: Mechanism of cinnamaldehyde reduction by AtCAD, redrawn from Youn et al. (2006). The substrate is illustrated in black; NADPH in red; enzyme residues in green; and Zn^{2+} and water molecule in blue.

The mechanism proposed here for HYSs differs to the mechanism proposed for SAD and CAD enzymes during reduction of their aldehyde substrates. The mechanisms proposed for those enzymes rely on the presence of the "active site" zinc to activate the carbon-oxygen double bond of the substrate (fig. 78). Once the bond is activated the carbon atom can receive the NADPH hydride and the oxygen can abstract a proton from the active site threonine (or serine in PtSAD) (Youn et al., 2006, Bomati and Noel, 2005). Based on the orientation of the residues in the active sites of AtCAD and PtSAD the electron transfer during reduction is proposed to be shuttled to the solvent through the active site threonine (or serine in the case of PtSAD) and then onto the hydroxyl groups of the NADPH ribose, following transfer onto the active site histidine before being transferred to the neighbouring aspartate and finally onto a water molecule (fig. 78).

A major difference is the role of the "active site" zinc in HYSs. Chelation studies suggest that zinc is not essential for catalysis, and there may be no need for activation of the carbon-nitrogen bond as it is easily reduced. Although the "active site" zinc might not be necessary for catalysis in HYSs it could still be carrying out a structural role. The coordinating amino acids are conserved among all the HYS from *C. roseus*. Notably, zinc is absent from certain MDRs and are apparently not necessary for their activity. For example, the quinone oxidoreductase (QOR, PDB accession code: 1QOR) from *E. coli* has lost both the structural and the "active site" zinc (Edwards et al., 1996). The "active site" zinc coordination sphere has been replaced by amino acids incapable of coordinating zinc (such as leucine and asparagine) and the loop section which is responsible for binding the structural zinc is entirely absent from the sequence. The need to boil THAS1 and HYS to successfully carry out a zinc detection assay suggests that the Zn^{2+} ions are very tightly bound to the protein and not easily accessed by other reagents. This raises the question of whether the structural zinc has a function in MDRs and could these MDRs evolve or be engineered to lose the zinc binding loop without compromising protein stability and function?

3.4 Conclusions

Crystallisation of these heteroyohimbine synthases has facilitated the study of the mechanism of heteroyohimbine synthesis from strictosidine aglycon. Extensive mutagenesis and deuterium labelling has allowed the proposal of a catalytic mechanism. Some aspects of the mechanism are not entirely clear: for example how cathenamine and 19-epicathenamine equilibrate. However, not all mechanistic aspects could be investigated in this enzymatic system.

These enzymes have revealed the potential of biosynthetic machinery to generate stereochemical variation. Even similar enzymes, such as THAS1 and HYS (76.3 % amino acid similarity), can display different stereoselectivity. Flexible loop regions can be the key to unlocking chemical diversity. As demonstrated here, mutating the extended loop2 over the HYS active site (fig. 68) impacts stereochemical outcome. The MDRs discovered from *C. roseus* show large variability in this section of their sequence. This active site loop of the MDRs could potentially be harnessed in protein engineering efforts to generate novel catalytic activity.

The existence of four (at least) redundant, and highly diverged, THAS enzymes in *C. roseus* indicates firstly that this branch of the pathway is relatively ancient and secondly that this reaction is possible with many different variations of the active site. Whether or not there is selective pressure to maintain or eliminate the promiscuous HYS enzyme remains to be determined. The function of the various heteroyohimbine stereoisomers *in planta* still remains cryptic.

The results detailed in the chapter have been published in Stavrinides et al. (2016).

3.5 Materials and Methods

3.5.1 Site directed mutagenesis of THAS1 and HYS

THAS1 mutants were generated by overlap extension PCR. Briefly, the codon to be mutated was selected and two primers, one reverse and one forward (table 10), were designed to overlap and introduce the mutation. Two PCR steps were necessary to create the mutations. A first PCR was carried out using the reverse mutant primer and the 5’ forward gene-specific primer (Chapter 2 table 5), thus generating the 5’ half of the gene carrying the mutation. In parallel, the 3’ half of the mutated gene was generated by PCR using the forward mutant primer and the 3’ reverse gene-specific primer (Chapter 2 table 5). The PCR products were gel purified and used for the second PCR overlap reaction for generation of the full-length mutated gene. To do this the 5’ and 3’ halves of the mutated genes were mixed in a PCR reaction in equimolar amounts (approx. 100 ng per fragment) and 5 cycles of PCR were carried out without including primers in order for the fragments to act as primers for each other. After 5 overlap PCR cycles the forward and reverse gene-specific primers were added to the mix and a further 30 cycles were performed using the annealing temperature for the primers. Full-length PCR products were gel purified, ligated into pOPINF expression vector and transformed into competent *E. coli* Stellar strain cells (Clontech Takara) as described in the Chapter 2 methods section. Due to repeated issues with

generating the HYS point mutants, these were instead obtained as gene fragments (Integrated DNA Technologies, Belgium) with the H127 or F128 codons mutated (H127A CAT -> GCA; H127N CAT -> AAC; F128A TTT -> GCT; F128Y TTT -> TAC) and the pOPINF overhangs included at the 3' and 5' extremities.

3.5.2 Loop swap mutants

Loop swap forward and reverse primers were designed by removing the codons for the loop amino acids and substituting those for the loop of the other enzyme (table 10). First the loop1 mutants were generated by overlap extension PCR following the same procedure detailed above for the point mutants and then the second loop2 swap was introduced. Mutant constructs were sequenced by Sanger sequencing to verify the mutant gene sequence and correct insertion.

Table 10: List of primer pairs for mutagenesis of THAS1 and HYS.

Mutant	Forward primer	Reverse primer
THAS1 mutants		
Y56A	GGACTTGCCAAGCTGACAGGGAAATG AGCAAAAACAAATTTGG	CCAAATTTGTTTTTGCTCATTTCCCTGTCAGCTT GGCAAGTCC
Y56E	GACAGGGAAATGAGCAAAAACAAATT TGG	CCAAATTTGTTTTTGCTCATTTCCCTGTCTTCT TGGCAAGTCCC
Y56T	GACAGGGAAATGAGCAAAAACAAATT TGG	CCAAATTTGTTTTTGCTCATTTCCCTGTCAGTT TGGCAAGTCCC
Y56S	GACAGGGAAATGAGCAAAAACAAATT TGG	CCAAATTTGTTTTTGCTCATTTCCCTGTCACTT TGGCAAGTCCC
Y56F	GACAGGGAAATGAGCAAAAACAAATT TGG	CCAAATTTGTTTTTGCTCATTTCCCTGTCAAAT TGGCAAGTCCC
E59A	GGACTTGCCAATATGACAGGGCAATG AGCAAAAACAAATTTGG	CCAAATTTGTTTTTGCTCATTGCCCTGTCATAT TGGCAAGTCC
S102A	CGGGGACAAAGTGGGCGTAGCAGCC ATAATTGAAACTTGTGG	GCTACGCCCACTTTGTCCCCG
S102Y	CGGGGACAAAGTGGGCGTAGCATACA TAATTGAAACTTGTGG	GCTACGCCCACTTTGTCCCCG
S102T	CGGGGACAAAGTGGGCGTAGCAACCA TAATTGAAACTTGTGG	GCTACGCCCACTTTGTCCCCG
S102D	CGGGGACAAAGTGGGCGTAGCAGAC ATAATTGAAACTTGTGG	GCTACGCCCACTTTGTCCCCG
T166S	GCAGGAATCTCGGCTTATAGTCCC	GGGACTATAAGCCGAGATTCCTGC
T166V	GCAGGAATCGTGGCTTATAGTCCC	GGGACTATAAGCCACGATTCCTGC
F67V	CAAGCTATCCTTATGTTTTAGGGC	GCCCTAAAACATAAGGATAGCTTGTAACTCC AAATTTGTTTTTGC
THAS1 loop swaps		
loop1 swap	CACAAACAAATTTGGGATGACAAAGT ATCCTTATGTTTTAGGGC	CTTTGTCATCCCAAATTTGTTTGTGCTCATT TCCCTGTCATATTGG
loop2 swap	GATGGCCATTTTGGAAATAATTTCGGG GCATGTTCAAATATAGC	GAAATTATTTCCAAAATGGCCATCTATTGATC CTGCTTCTGGACAG
HYS loop swaps		

Table 10 (cont.)

Mutant	Forward primer	Reverse primer
loop1 swap	AAGAACAAATTTGGATTTACAAGCTATC CTTTTGTTATAGGGC	GCTTGTAAATCCAAATTTGTTCTTGCTCATTTCC AAGTCATAGTTGC
loop2 swap	GAATCAACAGACAGCAATTACGGTGG ATGTTGTAATATAATGG	CATCCACCGTAATTGCTGTCTGTTGATTCTACTT TTGAACAG

The mutated codons are underlined.

3.5.3 Expression of mutant THAS1 and HYS

The pOPINF vectors harbouring the gene constructs were transformed into SoluBL21 (DE3) *E. coli* cells. A starting culture was grown overnight at 37˚C in 50 or 100 mL of LB media supplemented with carbenicillin (100 µg/mL) and was then diluted 1:100 in 100mL 2xYT media supplemented with carbenicillin (100 µg/mL) and allowed to grow to an OD_{600} between 0.8 and 1 before induction with 0.1 IPTG. The cultures were grown at 18˚C for 16h with 200 rpm shaking. Cells were collected by centrifugation and resuspended in 6 mL Buffer A (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 20 mM imidazole) along with EDTA-free protease inhibitor (Roche Diagnostics Ltd.). Cells were lysed using sonication for 4 minutes on ice using 2 s pulses with a 5-minute pause after 2 minutes on ice. Cell debris was pelleted by centrifugation at 17 x 1000 g and the supernatant was incubated for 1h at 4˚C with 200 µL of Ni-NTA agarose beads (Qiagen GmbH, Germany) pre-equilibrated with Buffer A. The total was centrifuged at 1000rpm for 1minute to pellet the Ni-NTA agarose and the supernatant was discarded. The Ni-NTA agarose was washed three times with 500 µL of Buffer A to remove non-specifically bound proteins, each time centrifuging for 1min at 1000 rpm to pellet the Ni-NTA agarose. Elution was done by washing the Ni-NTA with two volumes of 300 µL of Buffer B (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 500 mM imidazole). The eluates were transferred to a Durapore centrifugal filter (PVDF 0.1 µm, Merck Millipore) to remove any residual Ni-NTA agarose. An aliquot of eluates was analysed by SDS-PAGE to verify the purity and the molecular weight of the purified proteins. Protein containing eluates were concentrated to approximately 250 µL and buffer exchanged to Buffer C (50 mM Phosphate pH 7.6, 100 mM NaCl) in a 30 KDa membrane filter Millipore filter (Merck Millipore). Protein concentration was measured with Bradford reagent (Sigma-Aldrich) according to the manufacturer's instructions. Purified proteins were aliquoted in 20 µL aliquots and fast-frozen in liquid nitrogen before being stored at -20˚C.

3.5.4 SGD expression and purification

SGD expression and purification was done in large scale as described for THAS1 above using the expression system as described previously (Yerkes et al., 2008).

3.5.5 Expression and crystallisation of THAS1

Cultures of SoluBL21 *E. coli* expressing THAS1 (2L of 2xYT) were prepared as described above for small-scale purification. The cell pellet was resuspended in 50 mL of Buffer A and sonication

was done as described above and the cell debris was pelleted by centrifugation at 17 x 1000 g for 20 min. All large-scale purification steps were done at 4˚C on an ÄKTAxpress purifier (GE Healthcare). His-tagged THAS1 was purified using a HisTrap FF 5 mL column (GE Healthcare) equilibrated with Buffer A. The sample was loaded at a flow rate of 4 mL/min and step-eluted using Buffer B. Eluted protein was subjected to further purification on a Superdex Hiload 26/60 S75 gel filtration column (GE Healthcare) at a flow rate of 3.2 mL/min using Buffer D (20 mM Hepes, pH 7.5 150 mM NaCl) and collected by a fractionator into 8 mL fractions. The fractions were analysed by SDS-PAGE and those containing no traces of other contaminating proteins were pooled and concentrated in a 10 KDa cutoff Millipore filter (Merck Millipore) and concentration was measured using a BCA assay (Thermo Fisher Scientific Inc, USA). A fraction of the purified THAS1 was subjected to His_6-tag cleavage. Purified THAS1 was incubated overnight at 4 ˚C with 3C protease. The sample was then incubated with 1 mL of pre-equilibrated Ni-NTA column and then the total mixture was filtered and the flow-through was checked visually on SDS-PAGE for purity.

Crystallization screens were set up with purified protein in sitting-drop vapor diffusion in MRC2 96-well crystallization plates (Swissci) with a mixture of 0.3 µl well solution from the PEGs (Qiagen), PACT (Qiagen) and JCSG (Molecular Dimensions) suites and 0.3 µl protein solution. $NADP^+$ (Sigma Aldrich) was added to a final concentration of 1 mM for co-crystallization studies. Solutions were dispensed by an OryxNano (Douglas Instruments). THAS1 $NADP^+$ crystals were obtained from a solution containing 0.2 M potassium/sodium tartrate with 20% w/v PEG 3350, and THAS1 apo crystals were obtained from His_6-tag cleaved THAS1 (3C protease) in a solution containing 0.1 M MES, pH 6.5, 15% w/v PEG 2000.

3.5.6 Expression and crystallisation of HYS

Protein for crystallisation was expressed and purified as detailed above, but with the addition of 1 mM DTT to all purification buffers. The His_6-tag was cleaved as described above for THAS1. HYS was concentrated using a 30 KDa cutoff filter (Millipore) and was buffer exchanged into Buffer E (50 mM HEPES, pH 7.06, 150 mM NaCl, 10% (w/v) glycerol, 0.5 mM TCEP) for storage and crystallisation.

Crystals were grown from a hanging drop containing an optimised condition (0.1 M MMT buffer, pH 5 and 15% PEG 3350). Purified HYS (10.75 mg/mL) was mixed in a 1:1 ratio with the precipitant and incubated at 19˚C for two weeks before harvesting of crystals. Crystals were cryo-protected with well solution containing 25% ethylene glycol for 1 minute and frozen in liquid nitrogen.

Seeding experiments were carried out using droplets containing crystals which were pipetted repeatedly to crush up the crystals and then diluted 1:30 in well solution and then briefly vortexing the sample. A 48-well plate was setup by mixing protein and seeding solution in a ratio of 3:1:2 with the precipitant. $NADP^+$ was premixed with the protein for a final concentration of 1 mM. To determine the best storage and buffer conditions for HYS a thermofluor assay was done at the SPC facility (EMBL, Hamburg, Germany).

3.5.7 Data collection

X-ray datasets were recorded on one of three beamlines at the Diamond Light Source (Oxfordshire, UK) (5FI3, I04; 5FI5, I03; 5H83, I04-1) at wavelengths of 0.9000-0.976 Å (5FI3, 0.900 Å; 5FI5, 0.976 Å; 5H83, 0.920 Å) using either a Pilatus 6M or 2M detector (Dectris) with the crystals maintained at 100 K by a Cryojet cryocooler (Oxford Instruments). Diffraction data were integrated using XDS (Kabsch, 2010) and scaled and merged using AIMLESS (Evans, 2006) via the XIA2 expert system (Winter, 2010); data collection statistics are summarized in Stavrinides et al. (2016). Initially the THAS1 $NADP^+$ dataset was automatically processed at the beamline by fast_dp45 to 1.12 Å resolution and a structure solution was automatically obtained by single wavelength anomalous dispersion phasing using the SHELX suite (Sheldrick, 2008) via the fast_ep pipeline (Winter, manuscript in preparation). Despite being collected at a wavelength somewhat remote from the zinc K X-ray absorption edge (theoretical wavelength 1.284 Å) the anomalous signal was sufficient for fast_ep to locate four zinc sites and calculate a very clear experimentally phased electron density map (fig. 52). This was available to view at the beamline in the ISPyB database (Delagenière et al., 2011) via the SynchWeb interface (Fisher et al., 2015) within a few minutes of completing the data collection. The map was of sufficient quality to enable 94% of the residues expected for a THAS1 homodimer to be automatically fitted using BUCCANEER (Cowtan, 2006). The model was finalised by manual rebuilding in COOT (Emsley et al., 2010) and restrained refinement using anisotropic thermal parameters in REFMAC5 (Winn et al., 2003) against the same dataset reprocessed to a resolution of 1.05 Å as described above, and contained 97% of the expected residues, with one $NADP^+$ molecule and two zinc ions per subunit. All the remaining structures were solved by molecular replacement using PHASER (McCoy et al., 2007). In each case, the asymmetric unit corresponded to the biological dimer and the preliminary models were obtained by searching for two copies of a monomer template. For THAS1 apo and HYS apo a THAS1 $NADP^+$ protein only monomer model was used as the basis for the template, although in the latter two cases a homology model of the target structure was generated from the THAS1 template using the Phyre2 server (Kelley et al., 2015) (http://www.sbg.bio.ic.ac.uk/~phyre2) before running PHASER. In contrast to THAS1 $NADP^+$, the THAS1 apo and HYS structures were refined in REFMAC5 with isotropic thermal parameters and TLS group definitions obtained from the TLS-MD server (Painter and Merritt, 2006). Model geometries were validated with the MOLPROBITY (Chen et al., 2010) tool before submission to the PDB. The statistics of the final models are summarized in Stavrinides et al. (2016) in Annex 2. Additional statistics for Rpim: 5FI3, 0.020 (0.517); 5FI5, 0.038 (0.600); 5H83, 0.068 (0.664) and CC½: 5FI3, 0.999 (0.510); 5FI5, 0.999 (0.523); 5H83, 0.996 (0.510) (where values in parentheses are for highest-resolution shell) were also noted. Ramachandran statistics (favored/allowed/outlier (%)) are 5FI3, 96.8/3.2/0.0; 5FI5, 96.0/4.0/0.0; 5H83, 96.6/3.1/0.3. All structural figures were prepared using CCP4mg (McNicholas et al., 2011).

3.5.8 UPLC-MS and NMR analysis

UPLC-MS analysis was carried out on a UPLC (Waters) equipped with an Acquity BEH C18 1.7 µm 2.1 x 50 mm column connected to Xevo TQS (Waters). For fast dereplication of active enzymes

and mutants, a linear gradient method ("fast method") was used at a flow rate of 0.6 ml min-1 using a binary solvent system in which solvent A1 was 0.1% formic acid in water and solvent B1 was acetonitrile. The gradient profile was: 0 min, 5% B1; from 0 to 3.5 min, linear gradient to 35% B1; from 3.5 to 3.75 min, linear gradient to 100% B1; from 3.75 to 4 min, wash at 100% B1; back to the initial conditions of 5% B1 and equilibration for 1 min before the next injection. Column temperature was held at 30 °C. The injection volume for both the solutions of standard compounds and the samples was 1 µl. Samples were kept at 10 °C during the analysis.

For separation of the different heteroyohimbines, a different chromatographic method was applied that was adapted from the work of Sun J. et al. (Sun et al., 2011). In this method (Method 2) solvent A2 was 0.1% NH_4OH and solvent B2 was 0.1% NH_4OH in acetonitrile. A linear gradient from 0% to 65% B2 in 17.5 min was applied for separation of the compounds followed by an increase to 100% B2 at 18 min, a 2 min wash step and a re-equilibration at 0% B2 for 3 min before the next injection. The column was kept at 60 °C throughout the analysis and the flow rate was 0.6 ml min-1.

MS detection was performed in positive ESI. Capillary voltage was 3.0 kV; the source was kept at 150 °C; desolvation temperature was 500 °C; cone gas flow, 50 L h-1 and desolvation gas flow, 800 L h-1. Unit resolution was applied to each quadrupole. Multiple Reactions Monitoring (MRM) signals were used for detection and quantification of the compounds of interest (details of conditions in table 11).

Table 11: Multiple Reactions Monitoring conditions

Molecule	Parent ion	Daughter ion	Cone Voltage	Collision Voltage
Strictosidine aglycon	351	170.22	28	22
	351	144.16	28	24
Strictosidine	531	352.25	32	24
Heteroyohimbines	353	117.19	50	40
	353	144.16	50	26

NMR spectra were acquired using a Bruker Advance NMR instrument operating at 400 MHz for ^1H equipped with a BBFO plus 5 mm probe.

3.5.9 ^2H labelling experiments

Deuterated pro-*R*-NADPD was regenerated in solution by *Thermoanaerobacter brockii* alcohol dehydrogenase (50 units, Sigma) using 400 µM $NADP^+$ and 1% v/v [$^2H6$]-isopropanol (CIL). The NADPD regeneration was monitored by UV spectroscopy at 340 nm. Strictosidine (19.9 mg)

was incubated with 1.27 nM SGD in 94 ml of 50 mM phosphate buffer (pH 6.5). THAS1 enzyme was added to the reaction (final concentration of 1.65 µM) and the mixture was incubated at 35 ˚C with shaking. The reaction was monitored for completeness by LC-MS and after 5 h no strictosidine or deglycosylated strictosidine was observed. The reaction was stopped by addition of 100 ml of methanol and reaction mixture was concentrated to dryness. The dried reaction mixture was resuspended in 15 ml H_2O and extracted with 3 x 15 ml of ethyl acetate and the EtOAc fraction was dried. [21α-^{2}H1]-THA was isolated by preparative TLC separation on a nano-silica plate (Sigma-Aldrich), as described in Chapter 2. The band of [21α-^{2}H1]-THA was excised from the plate, silica was crushed to powder and THA was extracted with EtOAc multiple times, (total volume 40 ml). The EtOAc fraction was filtered and dried using a high-vacuum pump overnight. The [21α-^{2}H1]-THA was dissolved in 600 µl of $CDCl_3$ and ^{1}H-NMR was measured.

Strictosidine (39.3 mg) was incubated with 1 nM of SGD and 500 µM $NADP^+$ with 50 units of *T. brockii* ADH and 1% v/v [^{2}H6]-isopropanol in a total volume of 148 ml of 50 mM HEPES buffer (pH 7.5). HYS was added (final concentration 1.71 µM) and the reaction was incubated at 37 ˚C with shaking and monitored for completeness by LC-MS. After 6 h the reaction was complete and was stopped by addition of 150 ml of methanol. The reaction mixture was concentrated to dryness and then was resuspended in 50 ml H_2O, basified with 2 ml triethylamine and extracted with 5 x 20 ml of ethyl acetate. [21α-^{2}H1]-THA, [21α-^{2}H1]-AJM and [21α-^{2}H1]-19EA were isolated by preparative TLC and ^{1}H-NMR spectra measured as described above. ^{1}H-NMR spectra of deuterated compounds were compared with those of corresponding standards dissolved in the same solvent.

3.5.10 pH effect on product profile

Strictosidine was deglycosylated using purified SGD for 25 minutes at room temperature using assay conditions as described above. Strictosidine aglycon was then incubated at a final concentration of 300 µM at pH 5, 6, 7 and 8 in a buffer mix to avoid buffer ingredient effect on activity (50 mM Phosphate buffer, 50 mM citric acid, 50 mM HEPES, adjusted with NaOH or HCl).

At time zero the enzyme, either THAS1 or HYS (1 µM final concentration), premixed with NADPH (500 µM) was added to the substrate solution. In parallel, a chemical reducing agent, NaBH4 (3 mM final concentration), was added to deglycosylated strictosidine as a control reaction. All reactions were carried out in triplicate. An end-point sample (10 µl) was taken from each assay and prepared for UPLC-MS analysis by addition of 10 µl of 100% MeOH to stop the reaction, and then diluted 1 in 5 with H_2O, and centrifuged for 10 minutes at 4000 rpm. UPLC-MS and data collection and quantification were performed as described above using the “separation method”.

3.5.11 CD spectra and analysis

Far ultraviolet (UV) CD spectra of the wild-type enzymes THAS1 and HYS, as well as the loop mutants of THAS1 and HYS were recorded on a Chirascan Plus spectropolarimeter (Applied Photophysics) at 20°C in 10 mM potassium phosphate buffer pH 7.0. Samples were analysed from

180 nm to 260 nm using a 0.5 nm step at a speed of 1 s per step. Four replicate measurements were performed on each sample and baseline correction was applied to all data. Spectra are presented as the CD absorption coefficient calculated on a mean residue ellipticity (MRE) basis.

Melting curves of HYS and the HYS loop2 swap mutant were also acquired by CD. The samples were subjected to temperature ramping at the rate of 1˚C min-1 from 20 ˚C to 90 ˚C. Data collection was done from 260 nm to 201 nm using a 1 nm step and 0.75 s time per point. Data were analysed using the Global 3 software. HYS melting point was measured as 61.0 ± 0.1 ˚C; enthalpy 351.5 ± 3.6 KJ/mol. HYS loop2 swap melting point was measured at 62.0 ± 0.1 ˚C; enthalpy 535.8 ± 4.5 KJ/mol.

3.5.12 Docking of cathenamine in THAS1 holo structure

Cathenamine was docked into the THAS1-$NADP^+$ crystal structure using Autodock 4.2 (Morris, 2009). The ligand (cathenamine) was prepared with 2 torsions at the C16, the rest of the molecule being rigid, and the receptor consisted of the desolvated high-resolution crystal structure. The search space was defined by a 40x40x40 Å box with a 0.375 Å grid spacing, centred between the nicotinamide ring and the side chain of Tyr56, and encompassed the entire active site cavity. Searches were performed using the Lamarckian Genetic Algorithm, consisting of 100 runs with a population size of 150 and 2,500,000 energy evaluations. A total of 27,000 generations were analysed and clustered with an RMS tolerance of 2 Å per cluster. Results were visualised using the Autodock 4.2 user interface.

3.5.13 Testing of Zinc-apo THAS1 and HYS

Zinc was removed from THAS and HYS by dialysis against EDTA for 24h at 4˚C following a published protocol (Hoagstrom et al., 1969). EDTA was removed by overnight dialysis at 4˚C using the same buffer without the addition of EDTA. THAS1 had heavy precipitate following the second dialysis and only the soluble fraction was used for subsequent assays. Protein aliquots were analysed by SDS-PAGE to verify that there was no degradation. The activity against deglycosylated strictosidine and the product profile was determined using the same conditions detailed in enzyme reaction testing.

Zinc removal was verified by following the manufacturer's instructions using the Zinc analysis kit (Sigma Aldrich). Boiling for 10 min was necessary to denature the proteins in order for the reagent to access the zinc.

3.6 References

AULD, D. S. & BERGMAN, T. 2008. The role of zinc for alcohol dehydrogenase structure and function. *Cellular and Molecular Life Sciences*, 65, 3961-3970.

BAKER, P. J., BRITTON, K. L., FISHER, M., ESCLAPEZ, J., PIRE, C., BONETE, M. J., FERRER, J. & RICE, D. W. 2009. Active site dynamics in the zinc-dependent medium chain alcohol dehydrogenase superfamily. *Proceedings of the National Academy of Sciences*, 106, 779-784.

BOMATI, E. K. & NOEL, J. P. 2005. Structural and kinetic basis for substrate selectivity in Populus tremuloides sinapyl alcohol dehydrogenase. *Plant Cell*, 17, 1598-1611.

CHEN, V. B., ARENDALL, W. B., HEADD, J. J., KEEDY, D. A., IMMORMINO, R. M., KAPRAL, G. J., MURRAY, L. W., RICHARDSON, J. S. & RICHARDSON, D. C. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66, 12-21.

COWTAN, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography*, 62, 1002-1011.

DELAGENIÈRE, S., BRENCHEREAU, P., LAUNER, L., ASHTON, A. W., LEAL, R., VEYRIER, S., GABADINHO, J., GORDON, E. J., JONES, S. D. & LEVIK, K. E. 2011. ISPyB: an information management system for synchrotron macromolecular crystallography. *Bioinformatics*, 27, 3186-3192.

EDWARDS, K. J., BARTON, J. D., ROSSJOHN, J., THORN, J. M., TAYLOR, G. L. & OLLIS, D. L. 1996. Structural and Sequence Comparisons of Quinone Oxidoreductase, ζ-Crystallin, and Glucose and Alcohol Dehydrogenases. *Archives of Biochemistry and Biophysics*, 328, 173-183.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66, 486-501.

EVANS, P. 2006. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography*, 62, 72-82.

FISHER, S., LEVIK, K., WILLIAMS, M., ASHTON, A. & MCAULEY, K. 2015. SynchWeb: a modern interface for ISPyB. *Journal of applied crystallography*, 48, 927-932.

HOAGSTROM, C. W., IWEIBO, I. & WEINER, H. 1969. Interaction of Coenzyme with Differently Prepared Zinc-free (apo) Horse Liver Alcohol Dehydrogenases. *Journal of Biological Chemistry*, 244, 5967-&.

KABSCH, W. 2010. Xds. Acta Crystallographica Section D: Biological Crystallography, 66, 125-132.

KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols*, 10, 845-858.

LORIS, E. A., PANJIKAR, S., RUPPERT, M., BARLEBEN, L., UNGER, M., SCHÜBEL, H. & STÖCKIGT, J. 2007. Structure-Based Engineering of Strictosidine Synthase: Auxiliary for Alkaloid Libraries. *Chemistry & Biology*, 14, 979-985.

MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *Journal of applied crystallography*, 40, 658-674.

MCNICHOLAS, S., POTTERTON, E., WILSON, K. & NOBLE, M. 2011. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography*, 67, 386-394.

MORRIS, G. M. E. A. 2009. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, 30, 2785-2791.

PAINTER, J. & MERRITT, E. A. 2006. TLSMD web server for the generation of multi-group TLS models. *Journal of Applied Crystallography*, 39, 109-111.

SHELDRICK, G. M. 2008. A short history of SHELX. *Acta Crystallographica Section A: Foundations of Crystallography*, 64, 112-122.

STAVRINIDES, A., TATSIS, E. C., CAPUTI, L., FOUREAU, E., STEVENSON, C. E. M., LAWSON, D. M., COURDAVAULT, V. & O'CONNOR, S. E. 2016. Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity. *Nat*

Commun, 7.
STIERAND, K. & RAREY, M. 2010. Drawing the PDB: Protein−Ligand Complexes in Two Dimensions. *ACS Medicinal Chemistry Letters*, 1, 540-545.
STÖCKIGT, J., HÖFLE, G. & PFITZNER, A. 1980. Mechanism of the biosynthetic conversion of geissoschizine to 19-epi-ajmalicine in *Catharanthus roseus*. *Tetrahedron Letters*, 21, 1925-1926.
STÖCKIGT, J., HUSSON, H., KAN-FAN, C. & ZENK, M. 1977. Cathenamine, a central intermediate in the cell free biosynthesis of ajmalicine and related indole alkaloids. *Journal of the Chemical Society, Chemical Communications*, 164-166.
STOECKIGT, J., HEMSCHEIDT, T., HOEFLE, G., HEINSTEIN, P. & FORMACEK, V. 1983. Steric course of hydrogen transfer during enzymatic formation of 3.alpha.-heteroyohimbine alkaloids. *Biochemistry*, 22, 3448-3452.
SUN, J., BAKER, A. & CHEN, P. 2011. Profiling the indole alkaloids in yohimbe bark with ultra-performance liquid chromatography coupled with ion mobility quadrupole time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom*., 25, 2591-2602.
WINN, M. D., MURSHUDOV, G. N. & PAPIZ, M. Z. 2003. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods in enzymology*, 374, 300-321.
WINTER, G. 2010. xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography*, 43, 186-190.
YOUN, B., CAMACHO, R., MOINUDDIN, S. G. A., LEE, C., DAVIN, L. B., LEWIS, N. G. & KANG, C. 2006. Crystal structures and catalytic mechanism of the Arabidopsis cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Organic & Biomolecular Chemistry*, 4, 1687-1697.

Chapter 4

Discovery and characterisation of a Short-Chain Dehydrogenase/Reductase capable of reducing strictosidine aglycon

4.1 Introduction

4.1.1 Short-chain Dehydrogenase/Reductase family

The Short Chain Dehydrogenase/Reductase (SDR) family is a member of the Alcohol Dehydrogenase Superfamily, like the MDRs described in Chapters 2 and 3. These enzymes, as their name suggests, are usually shorter in length than the MDRs (Jörnvall et al., 1999) but, just like the MDRs, they have a Rossmann fold for NAD(P) binding. This enzyme family is of very old origin, and it is found throughout the kingdoms of life, even in viruses (Jörnvall et al., 1999, Kavanagh et al., 2008, Jörnvall et al., 2010), and are found in large numbers in plants (Moummou et al., 2012). In fact SDRs are thought to be the progenitors of the MDRs as they typically display earlier diversification than the MDRs (Strommer, 2011). The active site is characterised by a conserved catalytic tyrosine, which is usually accompanied by a lysine to give the conserved motif YxxxK (Jörnvall et al., 1995). The first characterised SDRs from plants were the tropinone reductases I and II from *Datura stramonium* (Portsteffen et al., 1992).

4.1.2 SDRs in *C. roseus*

The terpene moiety in the MIA pathway of *C. roseus* is cyclised and modified to yield secologanin, which is used to produce strictosidine. During the biosynthesis of secologanin, 10-oxogeranial undergoes a reductive cyclisation to yield nepetalactol and the enzyme responsible for the cyclisation is not a terpene synthase as would normally be expected, but instead is a member of the SDR family (Iridoid Synthase, ISY, Geu-Flores et al. (2012)). This enzyme is similar to the progesterone-5β-reductases which reduce the carbon-carbon double bond of progesterone (Gavidia et al., 2007)(fig. 79). *C. roseus* contains 5 ISY homologs, three of which have also been shown to reduce the ISY substrate, 10-oxogeranial (Munkert et al., 2015). It was demonstrated that many members of the progesterone-5β-reductase family have the ability to reduce and cyclise 10-oxogeranial even though the plants they are found in do not produce iridoids (Munkert et al., 2015).

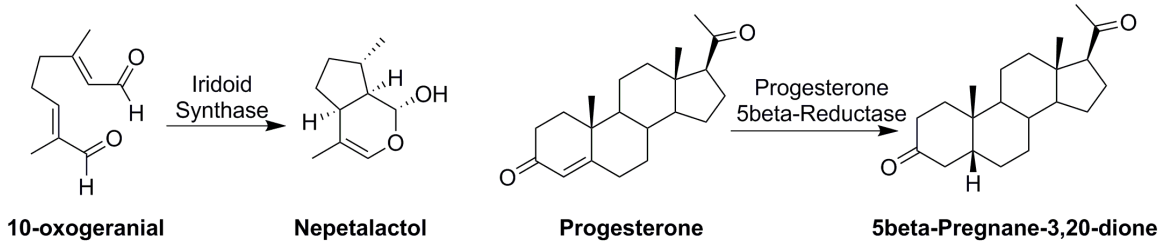


Figure 79: Reactions catalysed by Iridoid synthase and Progesterone-5β-reductase.

The identification of one SDR that catalyses a critical step in MIA biosynthesis in *C. roseus* suggests there may be other SDRs participating in the pathway. Just as there are multiple MDRs in *C. roseus* which catalyse reactions in MIA biosynthesis (10HGO, HYSs, T3R) there may also be more than one SDR involved in this pathway. To investigate this hypothesis, gene candidates that were co-regulated with known MIA biosynthetic genes were selected for screening in *in vitro* biochemical assays with various pathway intermediates as substrates. One candidate, Cro013448, was tested against 10-oxogeranial and was shown to have low activity against this substrate (the activity was not detectable by TLC whereas the activity of other candidates was detectable by TLC (Glenn, 2013)). This SDR enzyme is not a member of the progesterone-5β-reductase family and was therefore intriguing. The low activity against this 10-oxogeranial terpene moiety indicated it is probably not its native substrate. This enzyme was tested against deglycosylated strictosidine, which is the central intermediate of the MIA pathway, and showed high activity, producing a molecule of unknown structure, which did not correspond to any MIA present in *C. roseus* leaf extracts. This chapter details the characterisation of this enzyme and its product and attempts to explain its role in the plant.

4.2 Results

4.2.1 ADH10 cloning

As part of a screen to explore the activity of highly MeJa-upregulated reductases against the central MIA biosynthetic intermediate, strictosidine aglycon, the *C. roseus* transcriptome was searched for SDRs with high expression in MeJa-treated tissues. The gene coding for ADH10 was found to be relatively highly expressed in the *C. roseus* tissues analysed (fig. 80). It is located on a large whole genome contig (WGC), along with eight other SDRs or SDR fragments (Cr030915, Cr022958, Cr022957, Cr030568, Cr003140, Cr016747, Cr016749, Cr013447, and ADH10 Cr013448). Interestingly there are multiple predicted cytochrome P450s from various families on this same contig (Cr028046, Cr023307, Cr013485, Cr003361, Cr016728, and Cr016743) as well as two 2-oxoglutarate and Fe(II)-dependent oxygenase superfamily proteins (Cr019418 and Cr027425), an alpha-beta hydrolase superfamily protein (Cr010307), and an SAM-dependent methyltransferase superfamily protein (Cr016723).

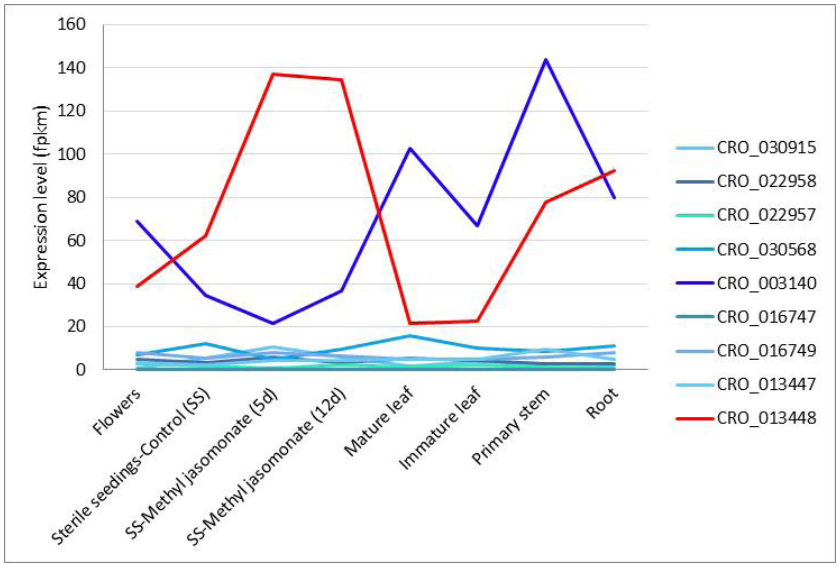


Figure 80: Expression profile of ADH10 (in red) and other ADH's on the same whole genome contig (WGC 27). In blue is illustrated the expression profile of Cr003140, another highly expressed SDR found on the same contig.

ADH10 appears to be very highly expressed in MeJa-induced seedlings (>100 fpkm) and has a higher expression in roots and stem than it does in leaves or flowers. The SDR that is next to it on the genome has a very low expression in all tissues (maximum 9.7 fpkm in the stem). All other SDRs on this contig have low expression except one, Cr003140, which has an expression profile considerably different to that of ADH10.

The ADH10 gene cloned from cDNA differed from that predicted for transcript Cr013448. The predicted protein consisted of 312 amino acids whereas the ORF cloned from cDNA coded for 299 amino acids. There were some discrepancies in the amino acid sequence of the predicted protein sequences (fig. 81). This resulted in an amino acid identity of only 80 %. It is not clear if these differences are due to problems with the assembly or because a different cultivar of *C. roseus* was used for cloning.

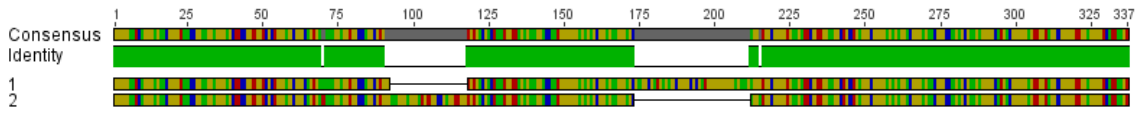


Figure 81: Alignment of predicted protein sequence of Cro013448 and the amino acid sequence of ADH10. Sequence 1: Cro013448; 2: ADH10.

4.2.2 Heterologous expression

ADH10 was expressed as a His_6-tagged fusion protein in *E. coli*. Expression was tested for ADH10 at various expression temperatures and also with various IPTG concentrations (fig. 82). After expression, the cells were lysed using the BugBuster kit and the soluble fraction was analysed by SDS-PAGE and visualised by staining with Coomasie Blue. The results indicate that expression of ADH10 is higher at 37 ˚C and very low at 10 ˚C. The concentration of IPTG used does not appear to have any significant effect.

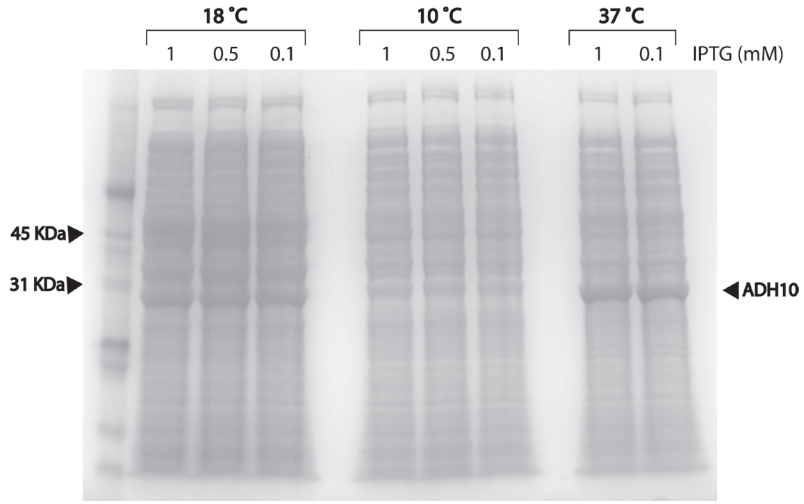


Figure 82: SDS-PAGE gel of expression trial of ADH10 with different temperatures and IPTG concentrations stained with Coomasie Blue. SoluBL21 *E. coli* cells harbouring ADH10 are induced with various concentrations of IPTG (0.1 to 1 mM) and allowed to express the protein at different temperatures (18: lanes 2-4, 10: lanes 5-7, and 37 ˚C: lanes 8-9). Lane one contains a protein molecular marker.

Large scale expression was done at 37 ˚C as that condition appeared to produce relatively high levels of protein. After purification by nickel column followed by gel filtration the protein was relatively pure (fig. 83). This was stored in Buffer D supplemented with the reducing agent TCEP and was used both for protein crystallisation and enzymatic assays.

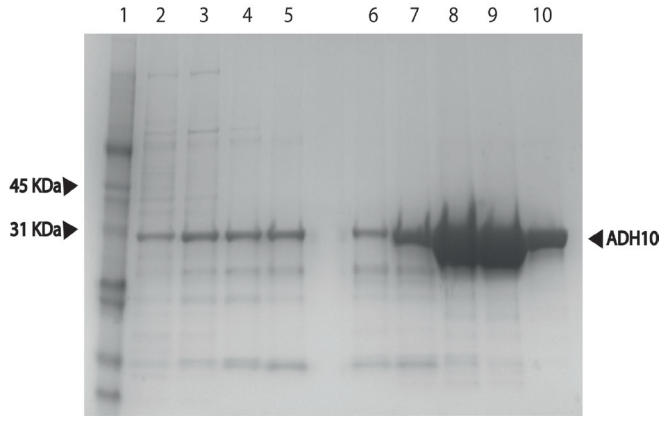


Figure 83: SDS-PAGE gel of ADH10 large-scale purification. Lane 1: Protein molecular marker; lanes 2-10: elution fractions after 2D purification by His-trap and gel-filtration of ADH10 expressed in *E. coli*.

4.2.3 pH assay with ADH10

ADH10 reactions with strictosidine aglycon were carried out at different pH values to test the effect of the pH on the reaction. Strictosidine was deglycosylated *in situ* and then the cofactor, NADPH, and an aliquot of ADH10 was added to the reaction. A small panel of different buffers was used to test the activity of ADH10 by LC-MS. The reaction was sampled at 1 minute and 10 minutes (table 12). The results of this assay suggest a larger quantity of product is formed when using citrate buffer at pH 6.0. All assays above pH 6.0 performed worse than all assays below

pH 6.0 in this endpoint assay, suggesting a lower pH is conductive for more product formation. As with the HYSs it is unclear if the increased product formation at low pH is due to the enzyme becoming more active or due to the increase in product equilibration into the isomer available for reduction by the enzyme, as described in Chapter 3.

Table 12: ADH10 product formed in different pH buffers in endpoint assays. The highest performance was recorded in Citrate pH 6.0 buffer, highlighted in red.

	Peak response area	
Buffer used	1 min	10 min
Phosphate pH 5.0	300048	546515
MES pH 5.5	213559	425088
Phosphate pH 6.0	251409	540379
Citrate pH 6.0	391654	651093
MES pH 6.5	178430	410661
Phosphate pH 7.0	111245	329108
Phosphate pH 8.0	52270	144047
HEPES pH 8.0	39150	131662

4.2.4 Kinetic studies of ADH10 with strictosidine aglycon

Spectrophotometric assays using a strictosidine concentration above 100 µM is not feasible as strictosidine aglycon is insoluble in water and forms precipitates which interfere with the absorbance reading. Therefore ADH10 could not be tested with starting concentrations of strictosidine above 100 µM. The initial velocity was highest at a substrate concentration of 100 µM and decreased with the lower concentrations tested (fig. 84, A). As with many enzymes that use the reactive strictosidine aglycon as a substrate, large fluctuation and variability was observed during these assays and therefore no concrete conclusions were drawn from these experiments and results in triplicate are not reported. ADH10 (50 nM) was also tested at different pH values (6.0 to 7.5) with 100 µM of substrate (fig. 84, B). The results resembled those obtained when product formed was monitored by LC-MS (table 12) with the assay at pH 6.0 performing better than the reactions at higher pH.

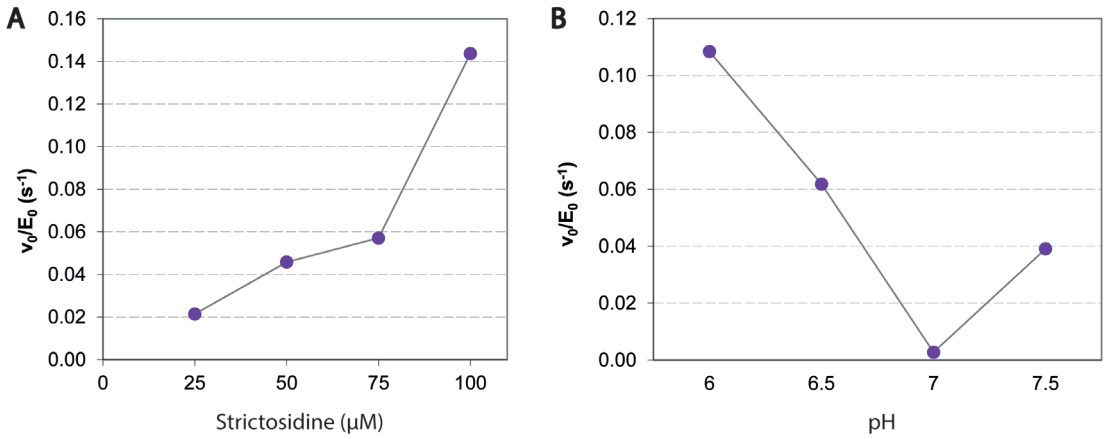


Figure 84: Activity assay of ADH10 with strictosidine aglycon. Consumption of NADPH was followed spectrophotometrically at 340 nm at four different concentrations of strictosidine. A: activity of ADH10 at different concentrations of substrate; B: activity of ADH10 at different pH with 100 µM substrate.

4.2.5 High resolution Mass-spectrometry and fragmentation

The LC-MS chromatogram of the enzymatic product did not correspond to authentic standards of any available MIAs. The purified reaction product was initially analysed by high-resolution MS to determine its exact mass and molecular formula (fig. 85). It was determined the product had a mass of 371.1967 *m/z*. This mass is very close (0.5 ppm error) to the calculated mass of 353 + H_2O (371.1965 *m/z*). The fragmentation pattern of the 371 product indicated that it is an indole with the classic 144 fragment observed for other indole alkaloids (Sagi et al., 2015). The other abundant fragments are 353, and 342. The removal of hydroxyl group as water is responsible for the fragmentation into 353. The 342 fragment could potentially arise from the loss of a CH_2O functional group.

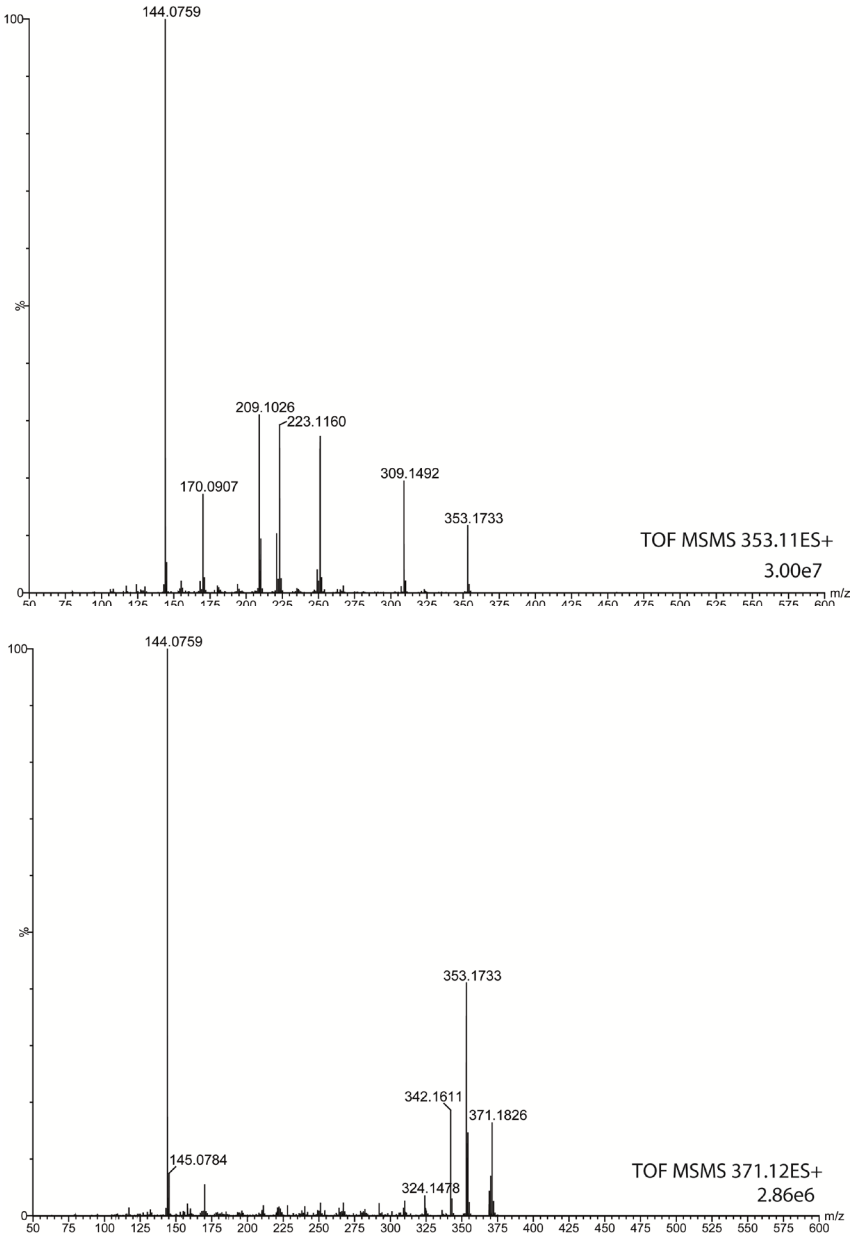


Figure 85: High-resolution MS-MS of the ADH10 product. The fragmentation of the product produces the fragments 353, 342, 324, and 144.

4.2.6 NMR characterisation of the ADH10 product

A large-scale reaction and NMR characterisation was necessary for determining the structure of the ADH10 product. Purification was carried out by TLC as opposed to preparative HPLC chromatography because the product was discovered to be unstable in the acidic chromatography conditions used in HPLC. The product mixture purified from the reaction presented multiple bands when analysed on TLC (fig. 86). The largest bands were initially tested using MS to determine which was the major product observed in the enzyme assays. It was found that the uppermost band (most hydrophobic) corresponded to the ADH10 product based on elution time and fragmentation pattern. Although there were some contaminants they were very low in concentration and therefore it was not necessary to further purify the product.

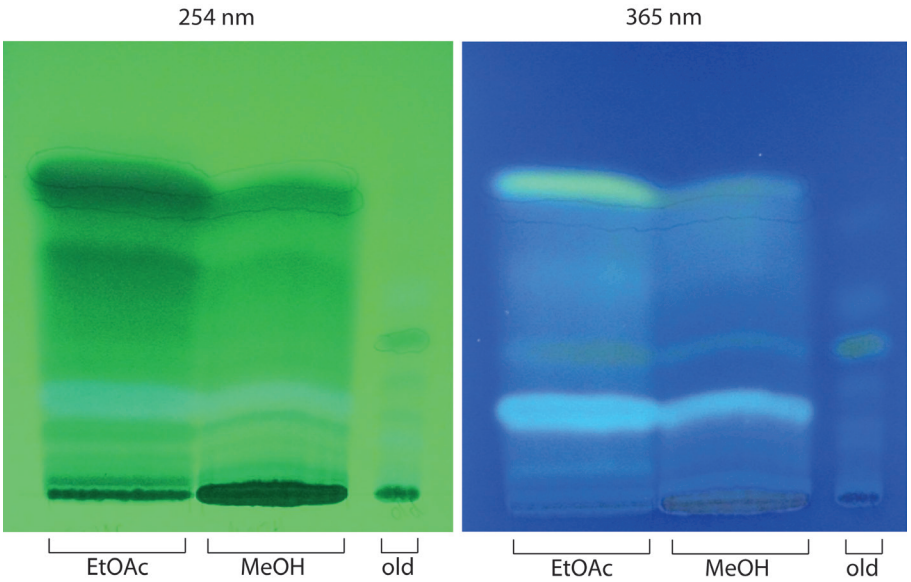


Figure 86: Large-scale purification of ADH10 product. The EtOAc-soluble compounds were spotted on the left and the MeOH-soluble compounds were spotted on the right. An old sample of purified ADH10 product which had degraded was also ran for comparison.

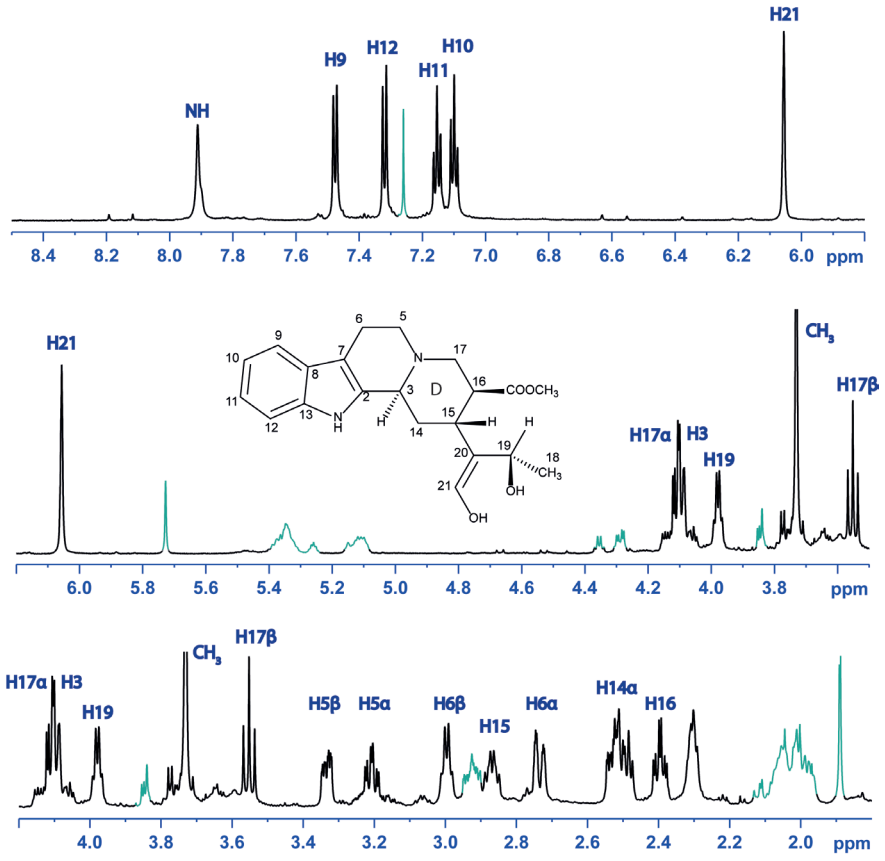


Figure 87: ^{1}H-NMR of ADH10 product. Peaks determined to be due to contaminants are shaded teal. Numbering is done following strictosidine and vallesiachotamine numbering. 18-CH_3 (at 1.36 ppm) is not included in this figure. The ring D is indicated on the proposed chemical structure.

The ADH10 product was dissolved in $CDCl_3$ and analysed exhaustively by different NMR methods including ^{1}H-NMR (fig. 87), COSY, HMBC (fig. 92), H2BC, and ROESY (fig. 88). The combination of

these measurements allowed the assignment of the hydrogens to the corresponding carbons and also the assignment of the stereochemistry at all the carbons. Correlations between carbons and hydrogens are illustrated in figures 90 and 91.

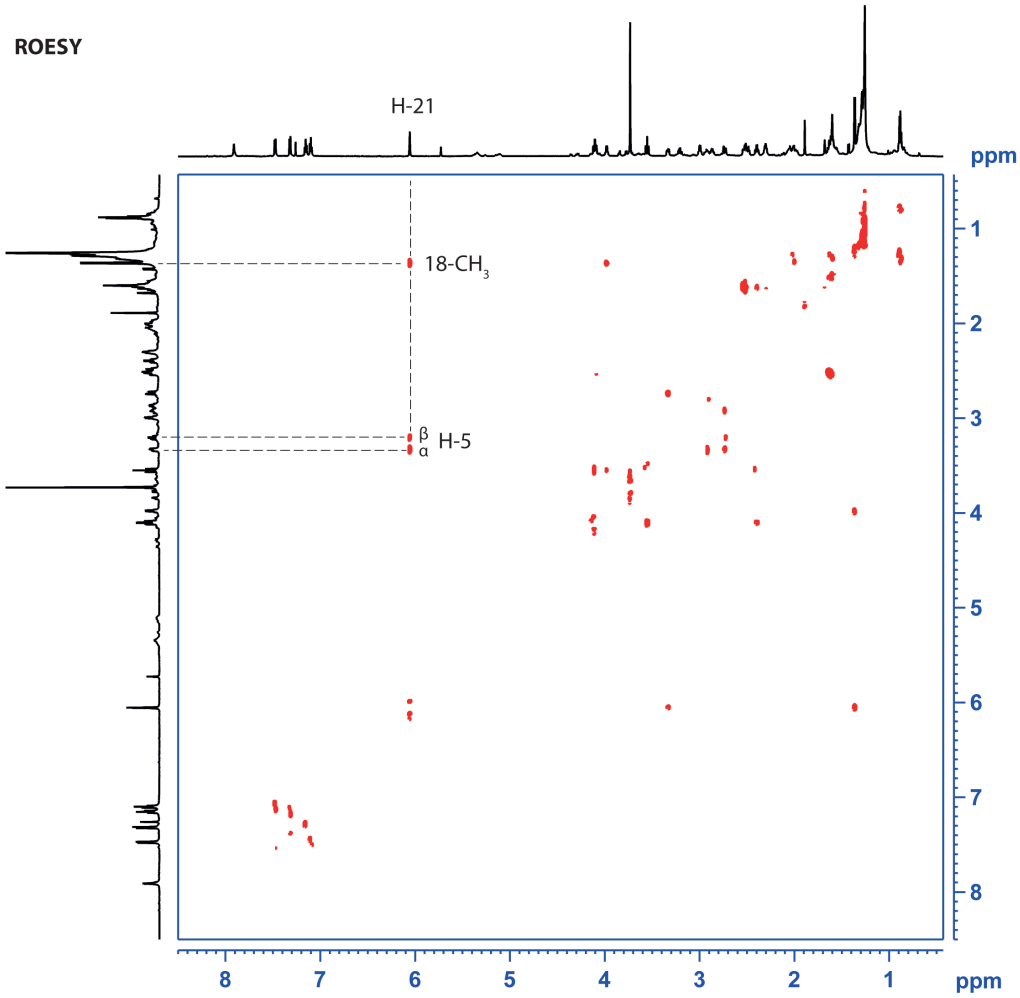


Figure 88: ROESY spectra of ADH10 product. The H-21 is picked out and the correlations with 18-CH_3 and the two H-5 are illustrated in dotted line.

ROESY is able to detect correlations in space up to 5 Å of distance. Interestingly, the ROESY indicates that the conformation of the ADH10 product brings the H-21 into proximity with the C-5 hydrogens. Therefore the C-18-19-20-21 group must be positioned below the ring D of the molecule. This is reminiscent of the proposed orientation of the predominant conformation of deformyl-E-geissoschizine in which the two H-16 are in close spatial proximity to the Hβ-6 (Jokela et al., 1993) (fig. 89) and therefore there is precedence for such a correlation between distantly connected hydrogens.

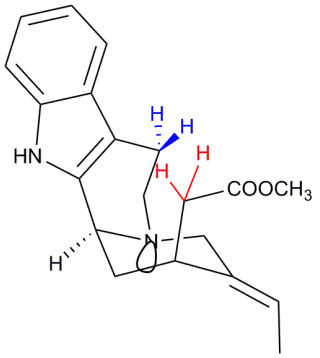


Figure 89: Deformyl-E-geissoschizine conformation. 6-C hydrogens are illustrated in blue and the 16-C hydrogens are in red. The electron doublet of N-4 is also illustrated.

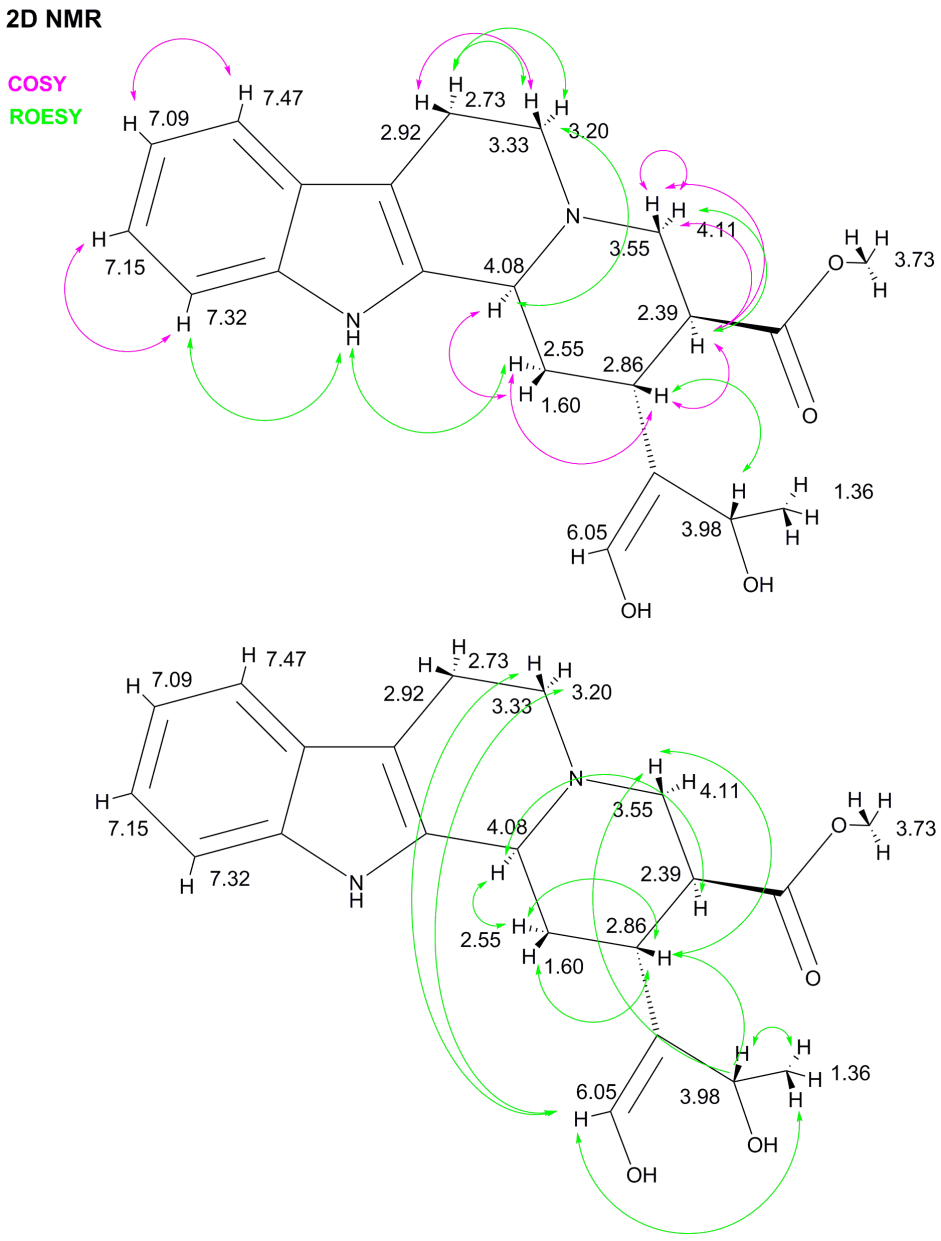


Figure 90: Correlations between carbons and hydrogens of the ADH10 product, as measured by NMR in $CDCl_3$.

2D NMR

HMBC

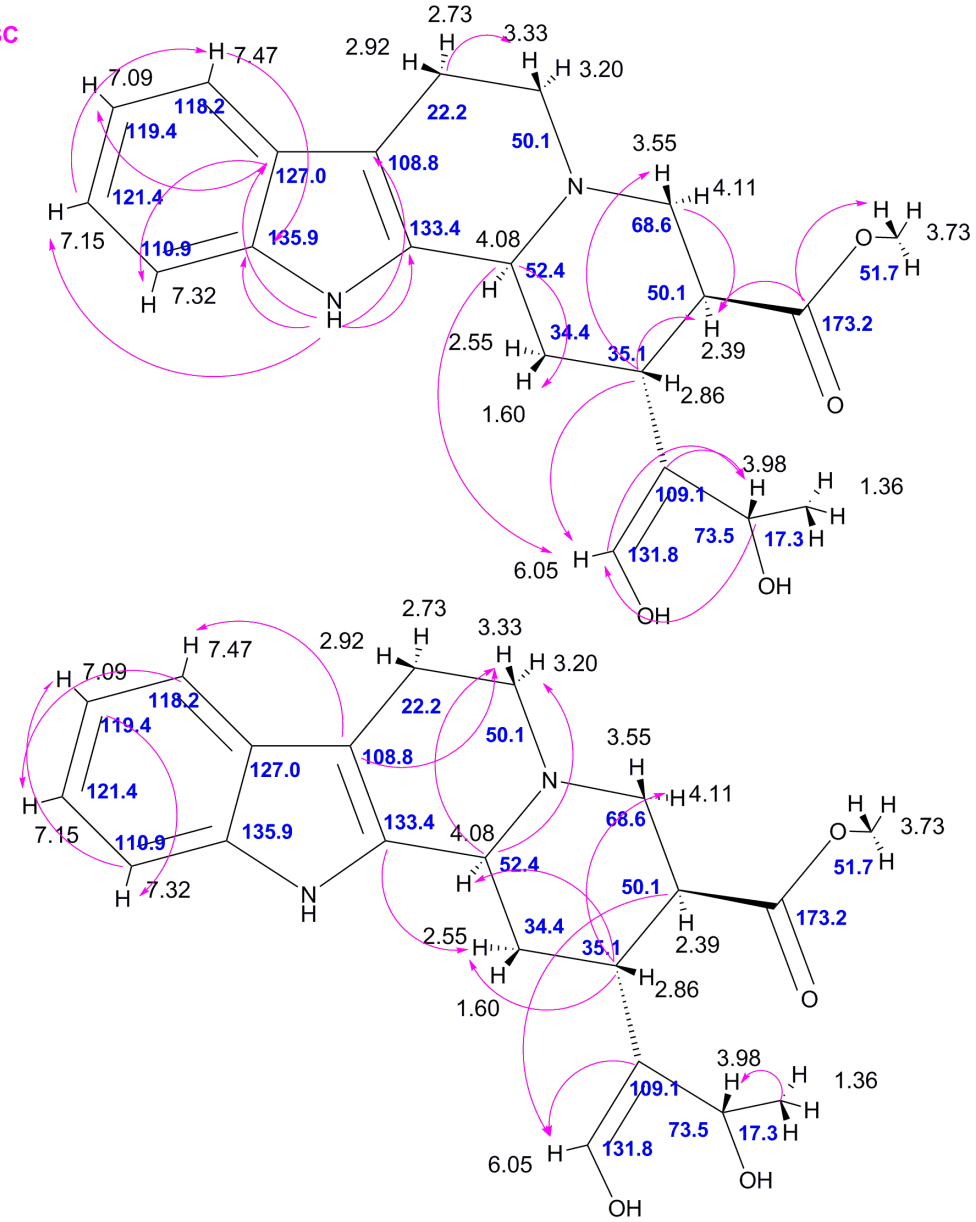


Figure 91: HMBC correlations between carbons and hydrogens of the ADH10 product, as measured by NMR in $CDCl_3$.

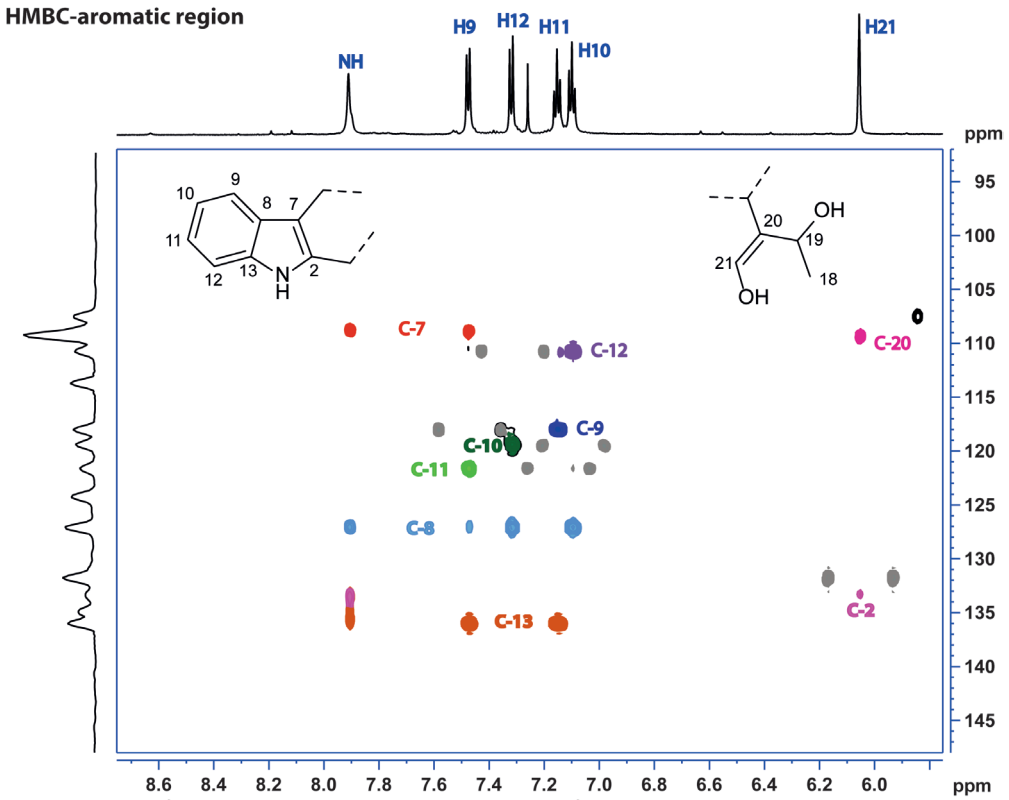
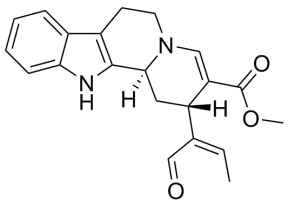


Figure 92: HMBC of ADH10 product. Display is only of the aromatic region, including H-21. The satellite peaks of carbons are illustrated in grey for clarity. The carbons with correlations to the hydrogens are each coloured differently.

Correlations between the indole moiety and the remainder of the molecule were apparent through the C-2 and the C-7. By decreasing the contour level in the HMBC we observe a cross-peak between H-21 and C-2 which is also correlated with the NH, thus confirming the two parts of the molecule are connected through the C-2 and C-7 (fig. 92).

Overall the correlations for this product suggest a structure that is similar to the natural MIA products vallesiachotamine and antirhine (fig. 93). Vallesiachotamine differs from the ADH10 product at the C-21 and C-19 positions; C-21 of vallesiachotamine has an aldehyde and has a vinyl function at position C-19. Furthermore, vallesiachotamine is oxidised at the C-17-C-16 bond whereas the ADH10 product is reduced. Antirhine, although it is reduced, lacks the $COOCH_3$ function on C-16. The ADH10 product is α-β unsaturated at position C-20-C-19 and has a hydroxyl on C-19. The ADH10 product was given the name 19,21-dihydro-17-dehydrovallesiachotamine, and the trivial name vitrosamine. Neither vallesiachotamine or antirhine are available commercially to compare spectra. As mentioned earlier the product does not seem to correspond to any product found in *C. roseus* leaf extracts, and has not been reported to have been isolated from any *C. roseus* tissue. The ADH10 product may be further derivatised by other enzymes, which is why it is not observable in *C. roseus*.



Vallesiachotamine

Figure 93: Vallesiachotamine

4.2.7 Results of deuterium labelling

Given the product structure it was apparent there were multiple bonds which could have been reduced by the action of ADH10. To pinpoint which bond was reduced, deuterium labelling was carried out using pro-*S* deuterium labelled NADPD. ADH10 is a member of the SDR family of ADHs; these typically transfer the pro-*S* hydride of the cofactor. The ADH10 product generated with this labelled cofactor was determined by LC-MS to have a mass of 372 (*m/z*) which corresponds to one hydrogen atom being replaced by a deuterium atom. This excludes the possibility that ADH10 could be performing a second reaction on the substrate by reducing an aldehyde moiety to produce a hydroxyl. The ^{1}H-NMR analysis indicated that the H-17β (3.55) had disappeared and the *J* coupling constants of H-17α (10.9) shifted to 7.1 (fig. 94). COSY analysis of the product showed only a cross peak at 4.11 and the cross peak at 3.55 was missing. Taken together these data indicate that the cofactor hydride is added to the C-17 at the β position (above the ring D).

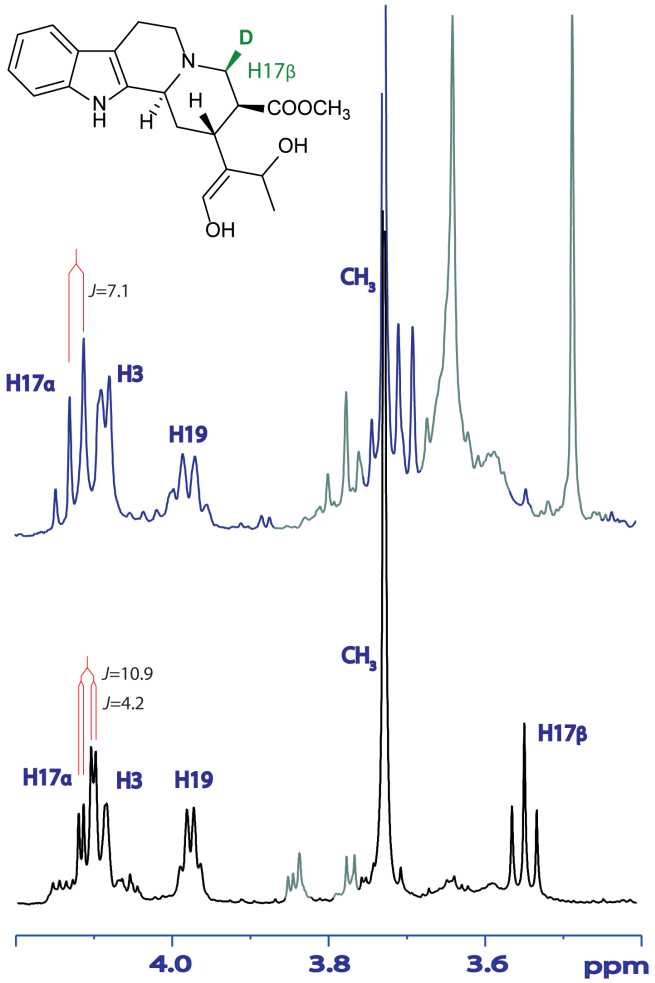


Figure 94: NMR spectra of deuterated ADH10 product (top) compared to non-deuterated ADH10 product (bottom). Signal from contaminants is coloured grey.

4.2.8 Crystallisation

Crystallisation of the enzyme was attempted using commercially available screens and also with optimised solutions. ADH10 produced crystals overnight in certain conditions of the JCSG+ suite. These first hits were collected and analysed at the Diamond Light Source Synchrotron but unfortunately none diffracted to good resolution (below 2.5 Å). Therefore optimisation screens were setup with sitting drop in MRC 2-drop 96-well plates (fig. 95-97). Various conditions were tested and constituents of those conditions are detailed in table 13. It appears PEG is necessary for ADH10 crystallisation as well as a slightly acidic pH.

Table 13: Crystallisation hit conditions used in optimisation screens

Condition	Buffer	pH	Salt	Precipitant
1	0.1 M MMT	4.0	-	25% PEG 1500
2	0.1 M MES	6.5	-	25% PEG 8000
3	0.1 M Bis-tris	5.5	0.1 M MgCl2	25% PEG 3350

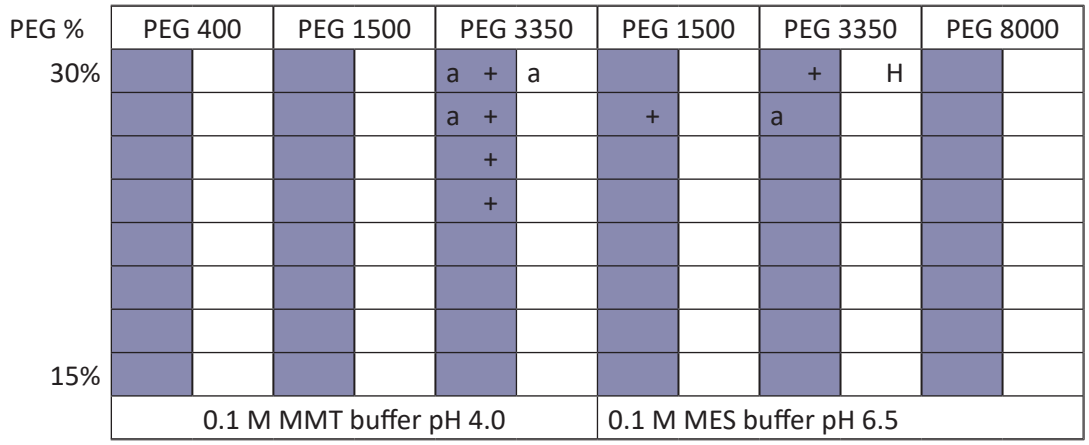


Figure 95: Optimisation screen using condition 1 and 2 from table 13. Columns containing drops with apo or $NADP^+$ protein are indicated in gray. Columns containing drops with apo or NADPH protein are indicated in white. The PEG used in each column is indicated above the columns and the percentage of that PEG is indicated at the left of the figure. The buffer content is indicated below the column at the bottom of the figure. a : crystals in apo drops; + : crystals in $NADP^+$-containing drops; H : crystals in NADPH-containing drops.

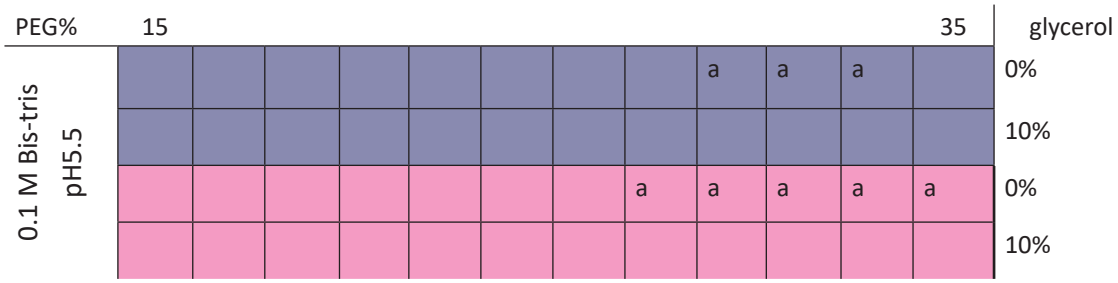


Figure 96: First crystallisation optimisation screen using condition 3 from table 13. Rows shaded in gray and pink contain 0.1 M and 0.2 M MgCl2 respectively. PEG 3350 is used throughout the screen. a : crystals in apo drops.

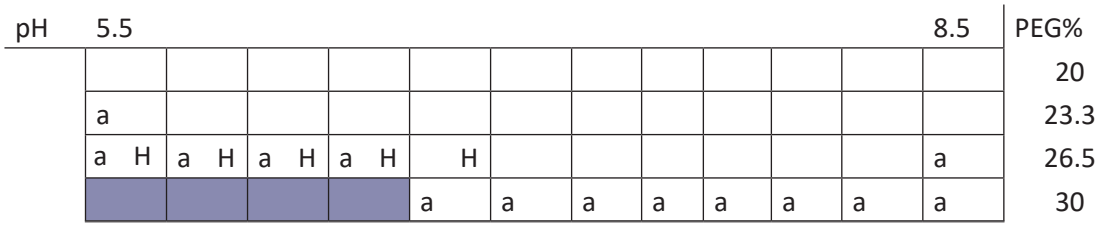


Figure 97: Second crystallisation optimisation screen using condition 3 from table 13. A pH gradient was generated by mixing 1 M Bis-tris at pH 5.5 with 1 M Bis-tris propane at pH 8.5 in different amounts for a final concentration of 0.1 M buffer. Each well contained 0.2 M MgCl2 and PEG 3350 at different concentrations. a : crystals in apo drops; H : crystals in NADPH-containing drops. The wells shaded in gray contained showers of microcrystals.

The highest resolution crystal was obtained from the MMT buffer optimised screen (table 13) in 0.1 M MMT with 24 % PEG 3350 (fig. 95). Unfortunately, the crystals that appeared in the NADPH-containing drops were very fragile and appeared to deflate when touched with a loop. The few crystals that were successfully mounted did not diffract to high resolution and were not analysed further. Interestingly, the apo crystals which diffracted were found to have a high occupancy of $NADP^+$. This indicates that ADH10 is purified from *E. coli* already bound to $NADP^+$ or

NADPH. For this reason, and because of the low resolution obtained with true apo crystals, the effort to produce better resolution apo crystals was abandoned.

4.2.9 Crystal structure of ADH10

ADH10 is not a metalloenzyme and it was not soaked with heavy metals, therefore molecular replacement was used to phase the crystal structure. When submitted to Swiss Model (Arnold et al., 2006, Guex et al., 2009, Kiefer et al., 2009) the closest match is the opium poppy enzyme Salutaridine reductase (PsSalR, fig. 98, Higashi et al. (2011)). PsSalR has only 53.2 % identity with ADH10 and therefore it is not an ideal model for this enzyme. However, at the time of dataset collection all other characterised and crystallised proteins had lower similarity to ADH10.

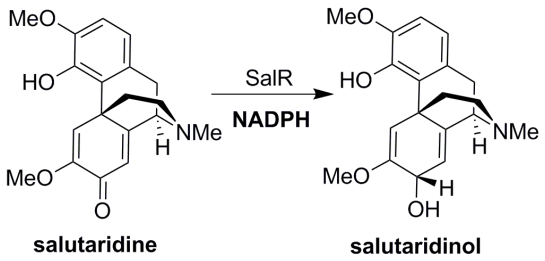


Figure 98: PsSalR reduction reaction on salutaridine.

ADH10 crystallised as a globular dimer in the asymmetric unit. In solution ADH10 is present as a monomer (determined during gel filtration purification for crystallisation, results not shown). The structure is characterised by a core of seven parallel β-sheets framed on either side by α-helices. The active site cavities face each other and are framed by the flexible loop extending over the active site, here called “lid”. This lid extends from residue 100 to residue 128 (100-VITDVEAVKKLNPAEDPADVDFSKIYKET-128) and is largely disordered in the crystal structure (fig. 99). The electron density in the lid area is too diffuse for accurate model building of the lid in this crystal structure. A lid reconstruction was attempted to coax more information from the data but extra density was not recovered after subsequent refinement using Refmac. Therefore, the lid section was truncated to the last residues which presented sufficient electron density (Val104 and Ala117). A version of the protein with the maximum amount of residues built into the density is used for illustration purposes here. The lid appears to wrap itself around the lid of the other monomer of ADH10 in the asymmetric unit in a head-to-head orientation (fig. 99). This brings the two active sites in proximity, across from each other (distance between the two $NADP^+$ C4 is 17.2 Å).

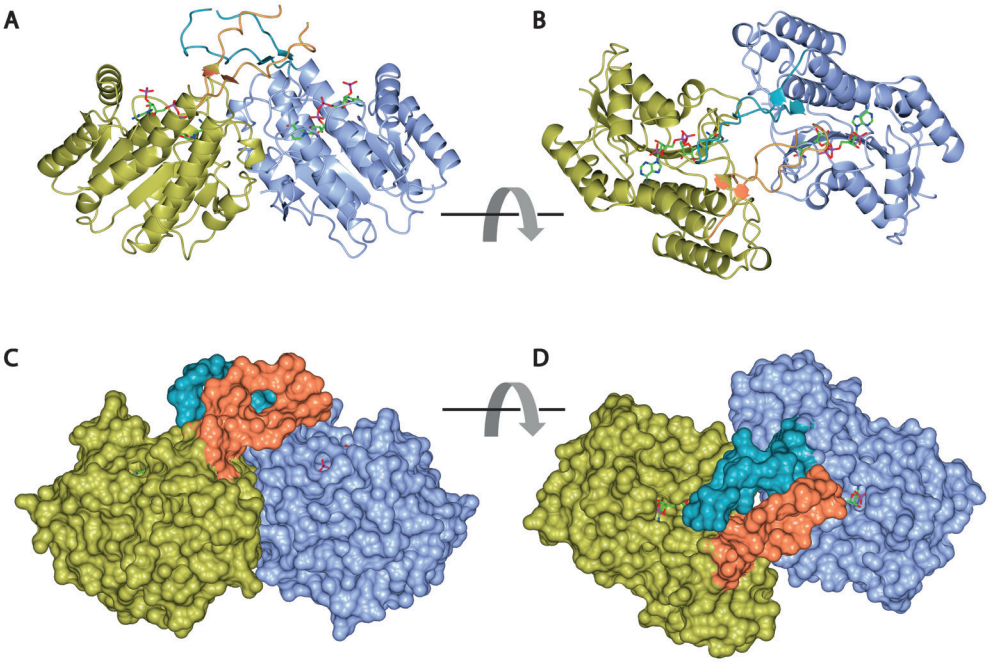


Figure 99: Overview of crystal structure of ADH10. Chain A is displayed in light blue and teal, chain B is displayed in gold and coral (body of the enzyme and lid, respectively). A and C: front view of the two chains in ribbon representation and space-filling respectively; B and D: top view of the two chains in ribbon representation and space-filling respectively. The $NADP^+$ is visible as cylinders with green bonds in the ribbon representations.

An extensive network of amino acid side chains and backbone amides holds the cofactor in place in the enzyme active site (fig. 100). The residues Ile20, Val68, Asn94, and Ala254 interact either with the carbonyl oxygen and/or nitrogen on the protein backbone. There are also a large number of water molecules coordinated to the cofactor and to the residue network holding the cofactor in place (fig. 101). The predicted active site residues, Tyr223 and Lys227, coordinate the ribose hydroxyls of the cofactor through hydrogen bonding. These residues are predicted to be part of the classic SDR catalytic triad Ser-Tyr-Lys (Filling et al., 2002, Jörnvall et al., 1995).

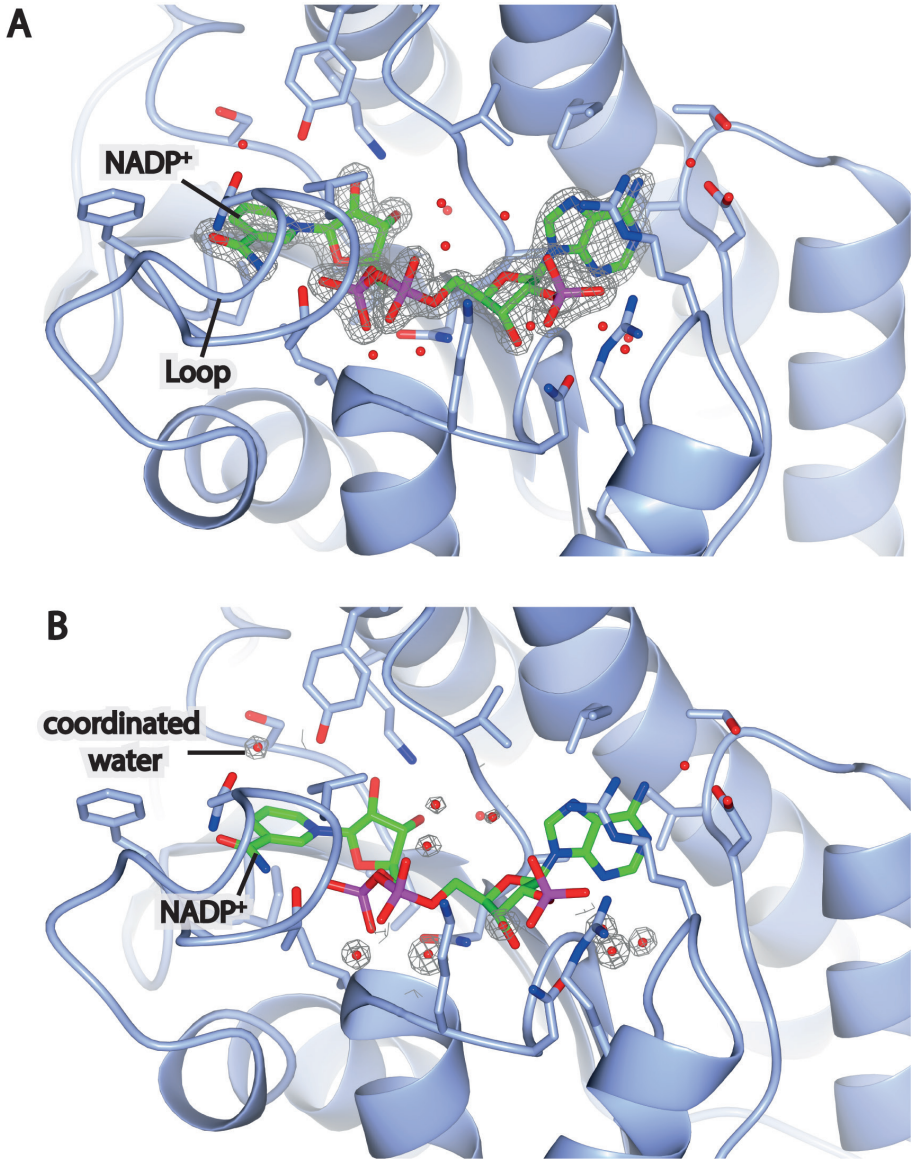


Figure 100: Electron density in the active site of ADH10. A: electron density around the cofactor, contoured at 0.55 Å of distance, the loop which lies over the cofactor is indicated; B: electron density around the water molecules within 2.5 Å of the cofactor, the water molecule coordinated to the active site tyrosine is indicated.

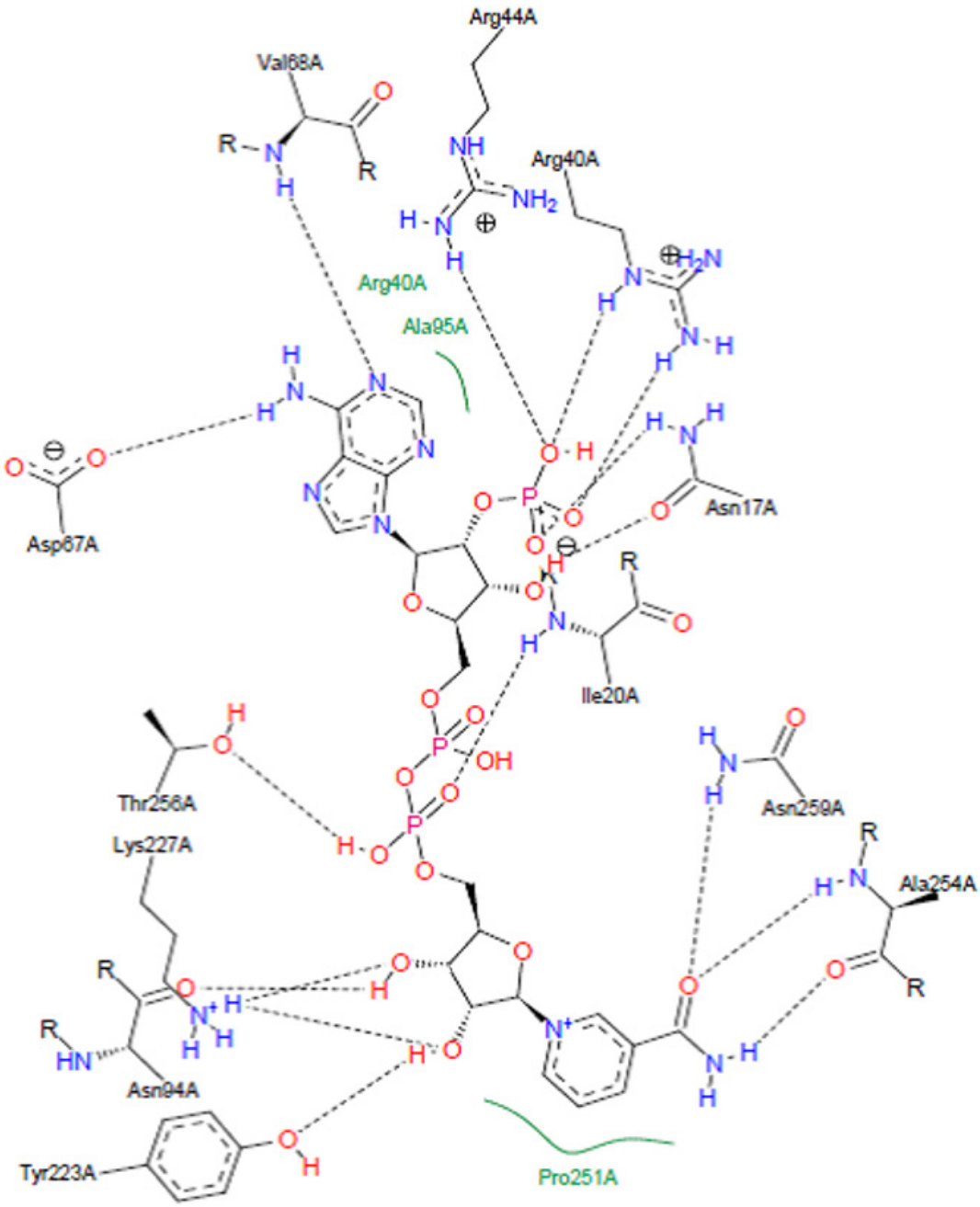


Figure 101: Network of ADH10 residues holding the NADP(H) cofactor in place. The image is calculated using Poseview (Stierand and Rarey, 2010). The active site residues Tyr223 and Lys227 also interact with the cofactor ribose moiety (bottom left). Hydrophobic interactions are illustrated as green lines; hydrogen bonds are illustrated as dashed black lines.

The active site appears to be easily accessible to the bulk solvent but appears to be largely hydrophobic as evidenced by the electrostatic potential map of the proteins' surface (fig. 102). This hydrophobic nature of the active site could be assisting the correct orientation of the substrate in preparation for catalysis. The aglycon substrate iminium would likely occupy the space occupied by the Tyr23-coordinated water molecule.

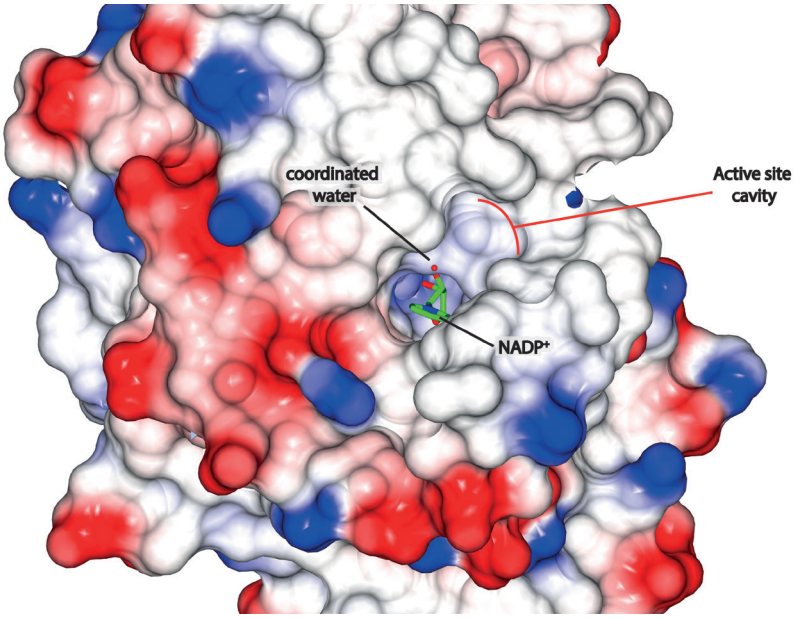


Figure 102: Electrostatic potential map of the surface of ADH10. The $NADP^+$ is illustrated as cylinders with green carbons and the Tyr223-coordinated water molecule is illustrated as a red sphere. The active site cavity is indicated by a red line.

4.2.10 Comparison to Salutaridine Reductase

One of the top hits when subjecting the ADH10 amino acid sequence to SwissModel (Kiefer et al., 2009) is the BIA biosynthetic enzyme PsSalR (Higashi et al., 2011). The other two top hits are the recently published crystal structures of menthone neomenthol reductase and isopiperitenone reductase (MMR and IPR, described in Chapter 1) from *Mentha piperita* (Lygidakis et al., 2016), but these enzymes act on a smaller, non-alkaloid, substrate. Furthermore, MMR is not stereoselective for one product, which is different to both ADH10 and PsSalR which are both stereoselective, and IPR does not have a classic catalytic triad, it instead has a glutamate in place of the tyrosine. For these reasons the ADH10 crystal structure and active site are compared here to PsSalR. Pairwise sequence alignment of ADH10 and PsSalR protein sequences were done with ClustalW in Geneious v8 using a BLOSUM matrix. The two proteins share 53.2 % identity and the identity is spread out over the entire sequence (fig. 103).

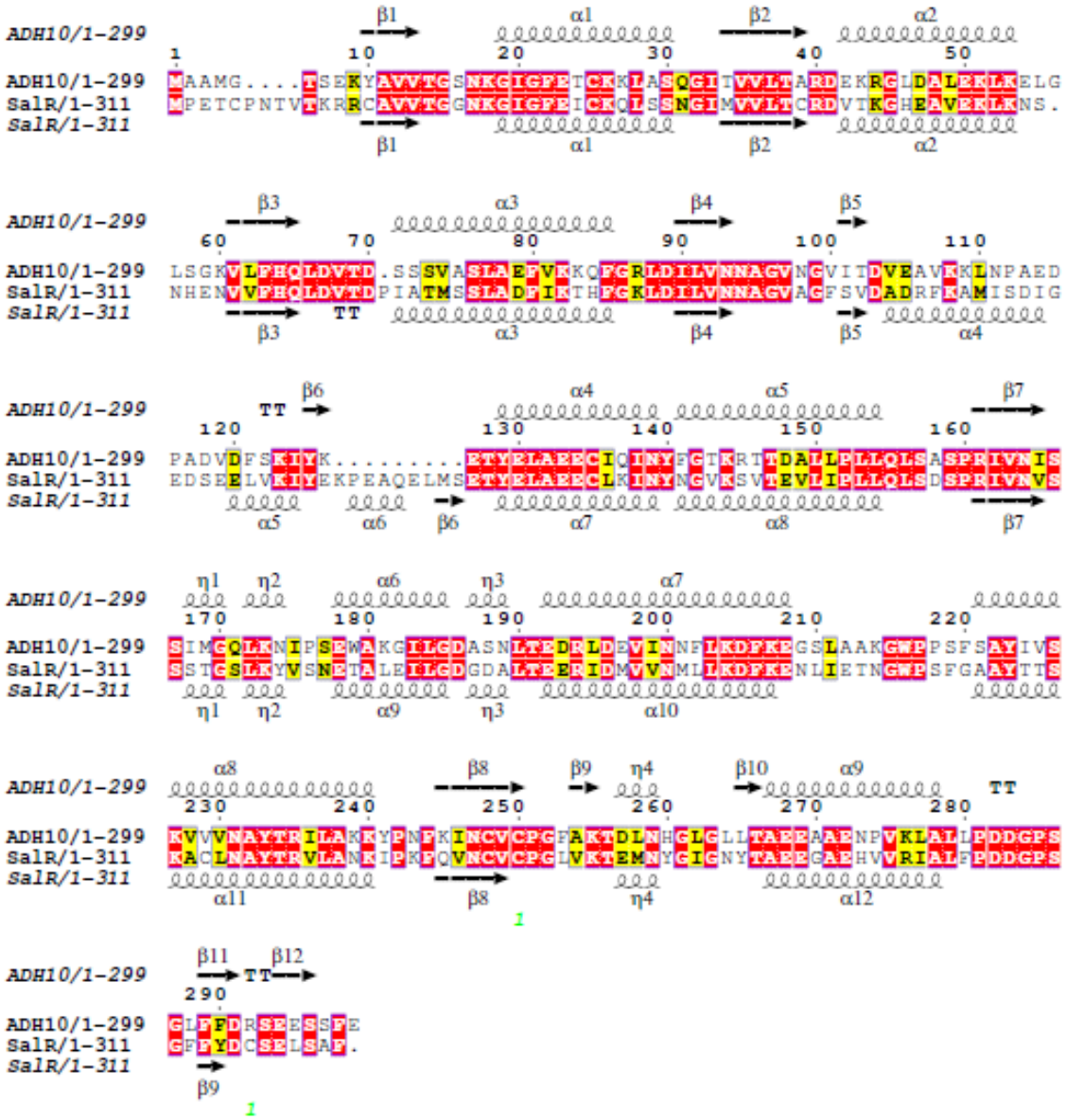


Figure 103: Alignment of ADH10 and PsSalR with secondary structures annotated. Sequence numbering is based on ADH10. Identical sites are illustrated in red and similar sites in yellow. The secondary structure is shown as arrows for the β-sheets and as helices for the α-helices. A disulphide bond in PsSalR is shown by a green '1'. The lid sections do not share very many similarities (residue 100 to 128 in ADH10).

An alignment of the ADH10 crystal structure and of PsSalR (PDB accession code: 3O26, Higashi et al. (2011)) from opium poppy (*Papaver somniferum*) was carried out in PyMol (v 1.1, DeLano Scientific LLC). The two structures are globally very similar with an overall RMSD of 0.618 Å (221 residues aligned). The main differences were found in the lid sections of the two proteins (fig. 104): in PsSalR the lid is mainly composed of α-helices and folds over the active site, whereas in ADH10 it appears disordered and interacts with the other monomer lid. PsSalR was crystallised as a monomer and therefore the lid is not interacting with another monomer. The lid of PsSalR has adopted a helical structure which folds over the active site (fig. 104, panel A). Given the high

structural similarity of ADH10 and PsSalR the lid of ADH10 could also adopt a helical conformation when in monomer form, similar to that seen in PsSalR.

The catalytic triad (Ser-Tyr-Lys) is positioned in nearly identical orientation in PsSalR and in ADH10. Therefore the substrate specificity must arise from the substrate binding in the active site. Amino acids which are different in ADH10 compared to PsSalR are Asn98 (Ala102 in PsSalR), Met169 (PsSalR: Thr182), Phe253 (PsSalR: Leu266), and Leu258 (PsSalR: Met271) (fig. 104). The water molecule coordinated to the active site tyrosine is present in both structures at the same location. As in PsSalR ADH10 has a loop which lies over the cofactor in the active site (265-279 in PsSalR, 252-266 in ADH10, fig. 104, panel A). This loop in fact carries both the Phe253 and the Leu258 of ADH10 which are Leu and Met in PsSalR.

Interestingly, the cofactor is quite buried in both these proteins, as evidenced by the space-filling model (fig. 104, panels C and D). As suggested by Higashi et al. (2011), the mobility of the lid could be for easy loading and release of the cofactor during catalysis. Concerted conformational changes could be responsible for the binding and release of the cofactor and those would involve movement of the lid and also the loop section.

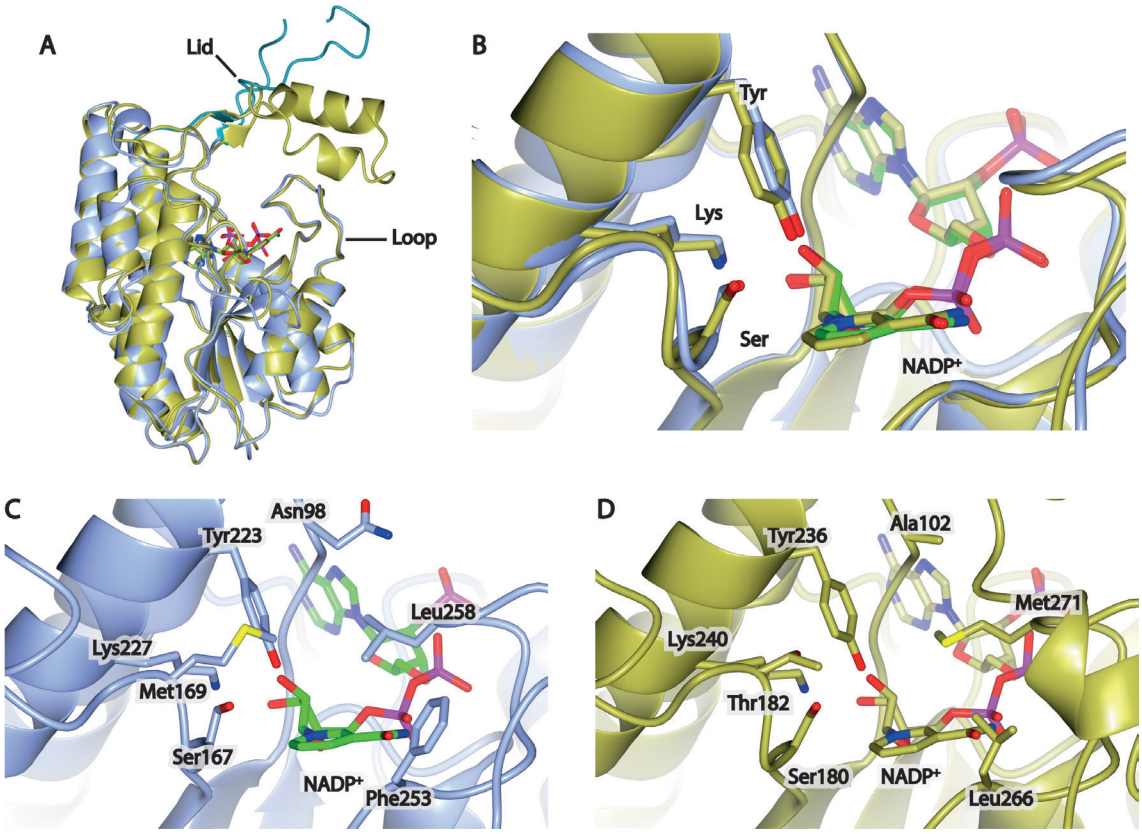


Figure 104: Comparison of SalR and ADH10 superimposed crystal structures. ADH10 is illustrated in light gray and SalR is illustrated in gold in all panels. A: general overview of crystal structures, $NADP^+$ is visible in the active site cleft; B: catalytic triad in the enzymes active sites illustrated as cylidners; C: ADH10 active site with catalytic triad (Ser167, Tyr223, and Lys227) and amino aicds which differ (Asn98, Met169, Phe253, and Leu258) illustrated as cylinders; D: PsSalR active

site with catalytic triad (Ser180, Tyr236, and Lys240) and different amino acids (Ala102, Thr182, Leu226, and Met271) illustrated as cylinders.

4.2.11 Mutations to ADH10 active site

Mutations to PsSalR Met271 and Asn272 had a detrimental effect to the activity of PsSalR. Mutations to Leu266 had an effect on the K_m of the enzyme but not so much on the k_{cat}, and mutations to Thr182 did not have a strong effect (Higashi et al., 2011).

To explore the mechanism of action of ADH10, mutations were also made in this protein. The amino acids Lys227, Ser167, and Tyr223 are positioned in the active site near the nicotinamide ring of the cofactor (fig. 104, panel B). Each amino acid was mutated to an alanine to probe its role in the catalysis of reduction. Each mutant was expressed as a His_6-tag fusion in *E. coli* and purified by Ni-NTA column. Expression was at similar levels for all mutants, and similar to the expression achieved when expressing the WT enzyme (fig. 105).

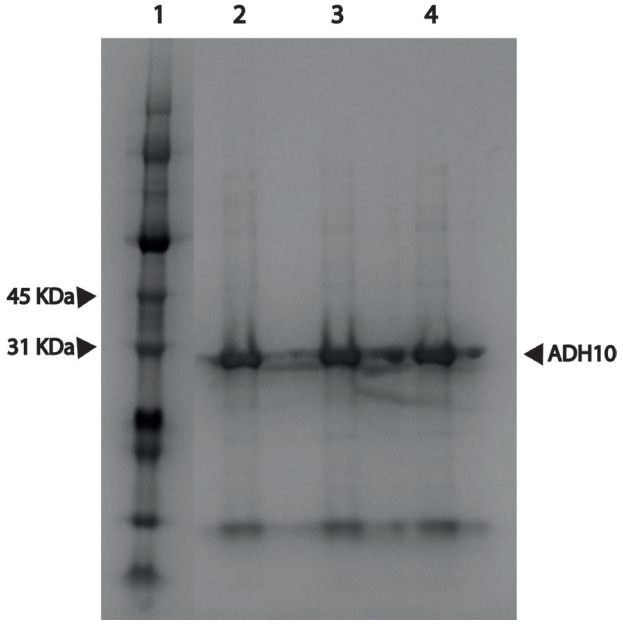


Figure 105: SDS-PAGE analysis of ADH10 mutants expression. Lane 1: Molecular marker; 2: ADH10 K227A; 3: ADH10 S167A; 4: ADH10 Y223A.

Assays conducted with the mutant ADH10 indicate that Tyr223 is necessary for catalysis. When mutated to alanine the major product (371 and 353) is not produced but the minor product at *m/z* 353 is still produced (fig. 106). This is typical for SDR enzymes (Jörnvall et al., 1995). Mutation to the other active site amino acids Lys227 and Ser167 did not have a dramatic impact on the product profile, although they both qualitatively reduced the enzyme activity based on endpoint assays. Both mutants were functional and a product of mass 371 and also of 353 was detected at the expected retention time (fig. 106).

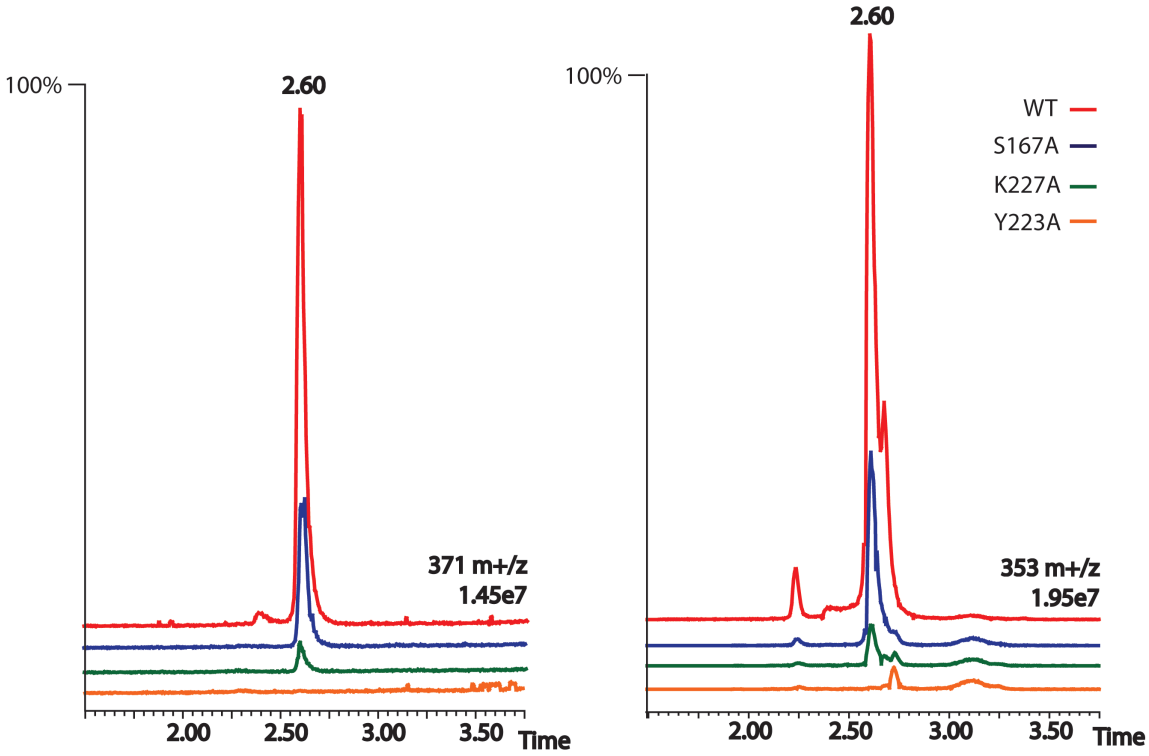


Figure 106: LC-MS trace of ADH10 and its mutants. Y223A (orange), WT (red), S167A (blue), K227A (green).

Analysis of the enzyme assays using a longer LC-MS method was carried out to determine the nature of the reaction by-products. The chromatograms were compared to an enzymatic reaction performed with HYS in the same conditions. The by-product visible in some reaction traces (fig. 107) coelutes with THA. The very low amount of THA could be a result of background reduction of cathenamine in solution. However, it is not possible to exclude the possibility that ADH10 is capable of reducing the pro-THA iminium form of cathenamine. There is no evidence that ADH10 has the ability to produce AJM or 19EA. No other heteroyohimbines are visible in these enzymatic reactions.

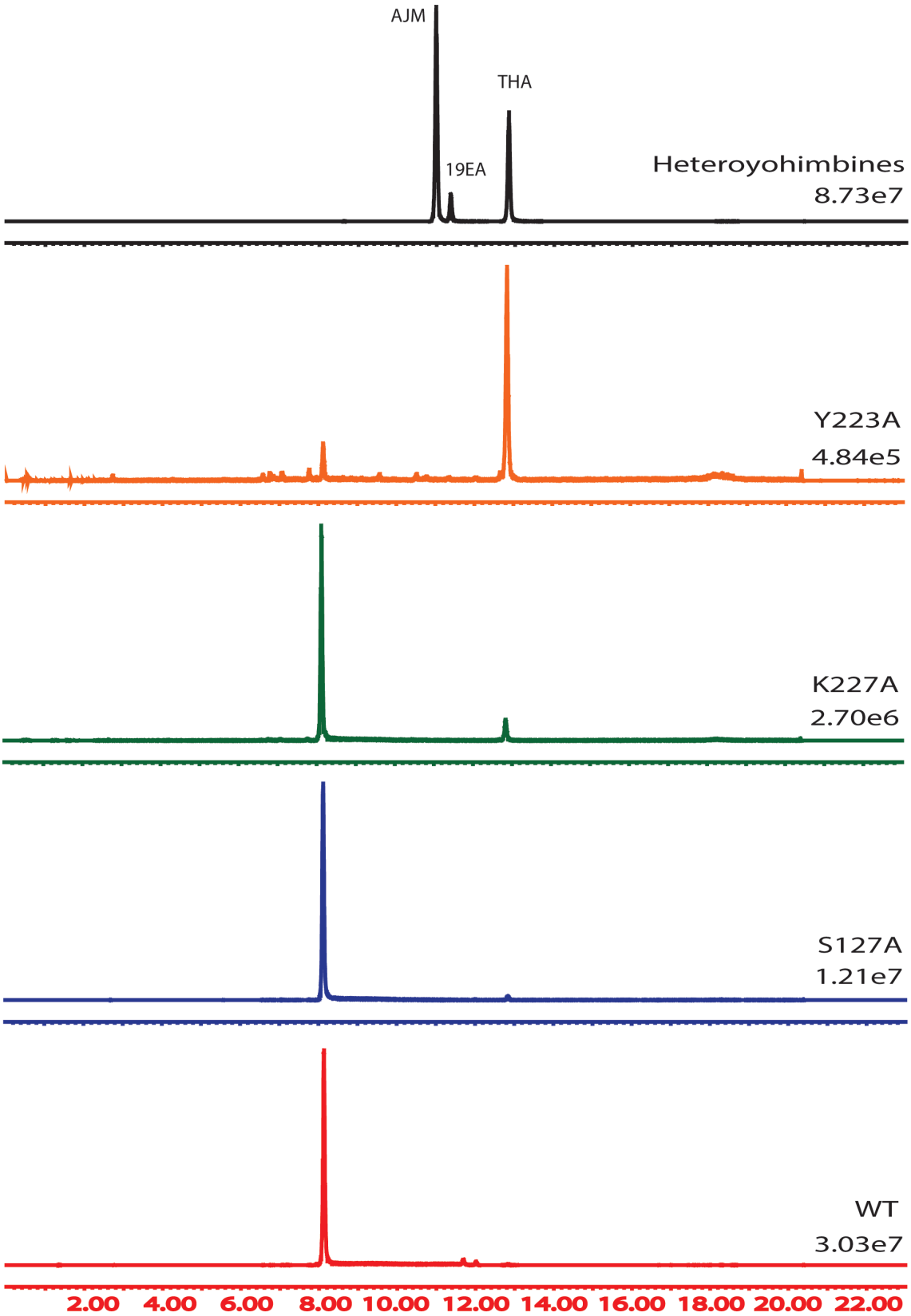


Figure 107: LC-MS trace of enzyme assays with ADH10 and its mutants compared to a HYS reaction. Assays analysed by LC-MS using the "separation method" and monitored using an MRM developed for heteroyohimbines (353 *m+/z*). Black trace: HYS reaction, producing AJM, 19EA, and THA; orange: ADH10 Y223A mutant; green: ADH10 K227A; blue: ADH10 S127A; red: ADH10 WT.

4.2.12 Proposed mechanism for ADH10 product formation

While HYSs act on cathenamine or the iminium tautomer of cathenamine to form a heteroyohimbine (Chapters 2 and 3), ADH10 generates a completely different backbone. Therefore, ADH10 uses a different rearrangement product of strictosidine aglycon, one in which reaction of the aldehyde at C-17 (not C-21 as in cathenamine) cyclises to form the D ring. If ADH10 binds the enamine it could be actively catalysing the equilibration to the iminium through providing a water molecule in close proximity to the C-16 (as discussed for the HYSs mechanism in Chapter 3). It could also bind the correct isomer of the iminium if it is present in solution.

Reduction would occur in a concerted manner, with the NADPH C-4 hydride attacking the C-17 from the top face of the substrate, thus forcing the electrons to move onto N-4 and then to abstract the proton from the Tyr223 hydroxyl (fig. 108). This is possible because of the basic side chain of Lys227 in very close proximity, which could lower the pKa of Tyr223 through the hydrogen bonding network with the cofactor ribose moiety. The Tyr223 abstracts a hydrogen from the ribose hydroxyl which in turn takes a hydrogen from Lys227. Lys227 can either take a proton from the solvent at this stage (there are many water molecules in the active site in close proximity) or could take one from the other hydroxyl of the cofactor ribose moiety to which it is also hydrogen bonded. The resulting product has the correct stereochemistry at C-17 and is released from the active site. The role of the active site serine, which is not essential according to mutagenesis studies, is not entirely clear.

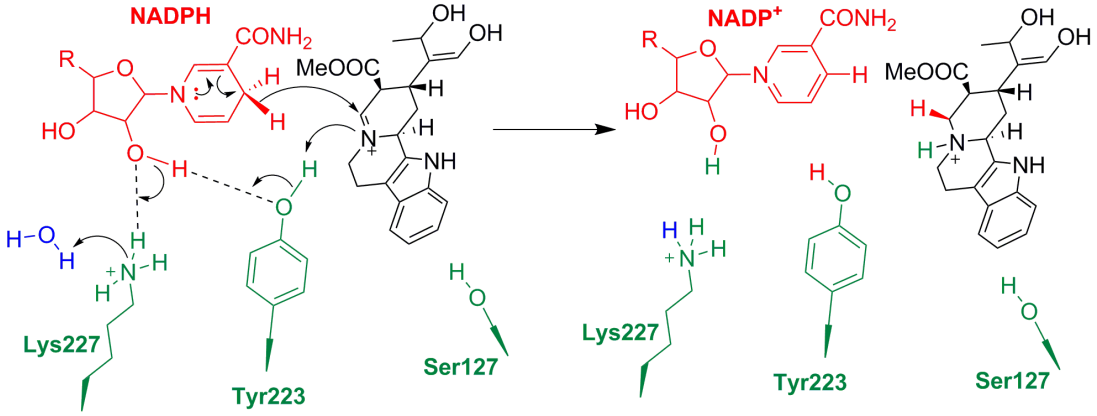


Figure 108: Mechanism proposal for reduction of strictosidine aglycon by ADH10. The cofactor and the hydrogens originating from it are illustrated in red; the active site residues and the hydrogens originating from them are illustrated in green; water molecule and the hydrogen originating from it are illustrated in blue.

4.2.13 Abundance of the pro-vitrosamine substrate in solution

An attempt was carried out to measure the kinetic constants of ADH10 with spectrophotometric methods, but it was quickly realised that the pro-vitrosamine substrate is in such low concentration that it was not possible to carry out any meaningful kinetics.

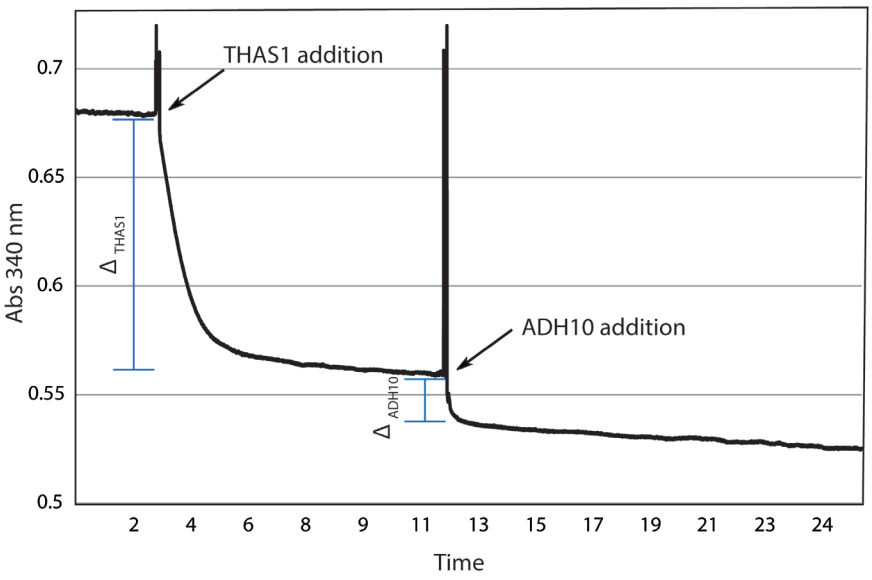


Figure 109: Spectra of absorbance at 340nm of THAS1 and ADH10 reaction at pH 7 with strictosidine aglycon at 100 µM. Addition of the ADH10 was done after the THAS1 reaction had reached a plateau.

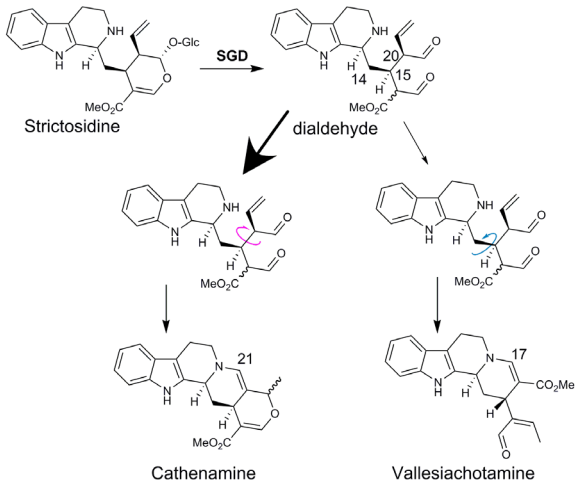


Figure 110: Equilibrium between the C-21 cyclised and C-17 cyclised aglycon is most likely not equal. The left part of the diagram represents the cyclisation between the N-4 and C-21 which is more abundant than the cyclisation between N-4 and C-17 on the right part of the diagram.

However, it was possible to calculate the relative abundance of the ADH10 substrate compared to the HYS substrates. The addition of THAS1 to a reaction mixture first allowed the determination of the abundance of cathenamine, the THAS1 substrate, in solution. THAS1 consumed the available cathenamine within 5 minutes. Addition of ADH10 after (or before) the THAS1 resulted in consumption of substrate as well, indicating that these two enzymes do not compete for the same substrate (fig. 109). The NADPH consumed in these experiments corresponded to the NADPH consumption when the enzymes acted in absence of the other.

The difference in absorbance from the beginning to the plateau (Δ_{THAS1}) was calculated in triplicate and was determined to be 16.5 ± 2.5 % of the total strictosidine aglycon pool. The

same was done for ADH10 and it was determined that the pro-vitrosamine substrate of ADH10 is present in 3.1 ± 0.7 % of the total strictosidine pool. That leaves approximately 80% of the substrate unavailable to either enzyme. This difference in equilibrium between the cathenamine and the vallesiachotamine forms is illustrated in fig. 110.

4.3 Discussion

4.3.1 Vallesiachotamine-type alkaloids

The alkaloid vallesiachotamine was first characterised by Djerassi et al. (1966)(fig. 111). This natural product has been shown to have some cytotoxic activity against human cancer cell lines with low-µM IC_{50} values (Feng et al., 2010). It was originally purified from bark of the South-American plant *Vallesia glabra* (synonym *V. dichotoma*, Apocynaceae) and mass spectrometry and elemental analysis revealed it has the chemical structure $C_{21}H_{22}N_2O_3$ (Walser and Djerassi, 1965). The authors noted that the compound is relatively unstable in presence of air or light, something which was also noted during this work with the product of ADH10. They propose that the precursor to vallesiachotamine is a corynantheine alkaloid, such as corynantheine. This precursor would undergo hydroxylation at the carbon adjacent to the N4 and could then equilibrate into an open form and ring closure could occur again after rotation of half the molecule around the C-15-C-14 bond (fig. 112).

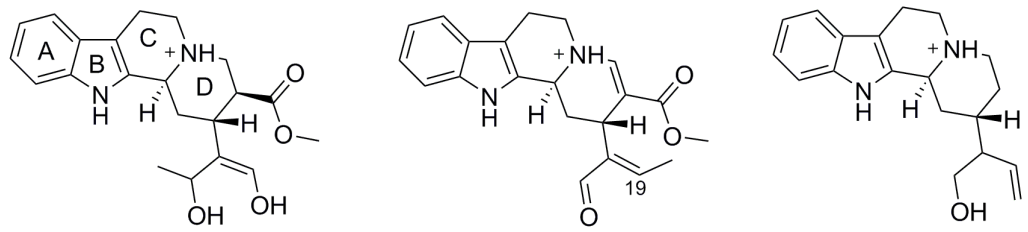


Figure 111: Vitrosamine compared to the MIA vallesiachotamine and antirhine. Isovallesiachotamine has the Z configuration at C-18,19 while vallesiachotamine has the E configuration. The ring numbering is indicated in the ADH10 product structure.

Djerassi et al. (1966) attempted to solve the structure of vallesiachotamine by reducing it with chemical reducing agents. They demonstrated that reduction of the enamine bond of vallesiachotamine with sodium borohydride in ethanolic solution is not possible, but is possible in a glacial acetic acid solution, which in theory lowers the pH to a value that allows equilibration to the iminium.

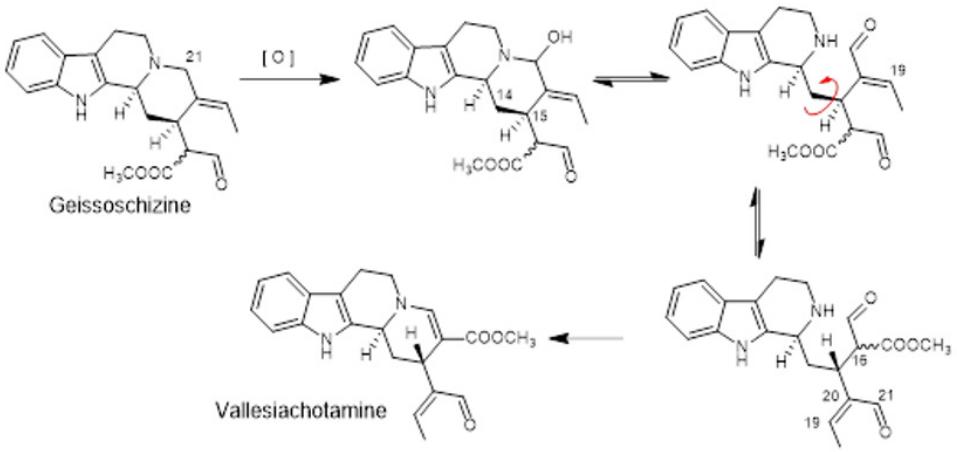


Figure 112: Proposed generation of vallesiachotamine from geissoschizine. Redrawn from Djerassi et al. (1966).

Smith (1968) had observed that hydrolysis of strictosidine led to the formation of the two diastereomers of vallesiachotamine (at C-19). It is not clear why vallesiachotamine and not cathenamine would be formed under the conditions used, but perhaps cathenamine was produced but not stable enough to be characterised by Smith (1968). Kan-Fan and Husson (1979) also found that 4,21-dehydrogeissoschizine yielded a mixture of the vallesiachotamine isomers when incubated in a buffered solution at pH 4 for 24 h. However, how the authors were able to purify and use pure strictosidine aglycon isomers with such certainty is incomprehensible as we have observed a large mixture of interconverting isomers in solution with cathenamine being the most stable one (i.e. the most abundant) (Stavrinides et al., 2016). Shen et al. (1998) incubated purified strictosidine with a bacterial culture of *Staphylococcus aureus* at pH 7.2 and observed that after 24 h incubation a yellow precipitate had formed (bacteria possess non-specific glucosidases capable of deglycosylating strictosidine). Upon purification of this precipitate the workers discovered that it consisted of vallesiachotamine and isovallesiachotamine in a 2:1 ratio. Interestingly, the yield for this reaction was rather low (15 mg (5.25 moles) of yellow precipitate from 220 mg (116.6 moles) of initial strictosidine, or 4.5% molar yield).

4.3.2 Characterisation of ADH10

ADH10 is capable of reducing, in the presence of NADPH, strictosidine aglycon to yield a product of *m/z* 371 which is relatively unstable. A large-scale reaction and purification allowed a thorough characterisation of the product with NMR. 2D-NMR characterisation of the product allowed assignment of all carbons and hydrogens. The product, given the name vitrosamine, appears to have the same stereochemistry at C-3 as strictosidine but ring D has been cyclised in a manner similar to vallesiachotamine, rather than cathenamine (fig. 113 below). All previously reported alcohol dehydrogenases found in the MIA pathway proceed via the intermediate cathenamine. However, the structure of the vitrosamine clearly shows that an alternative isomer of strictosidine aglycon is reduced (fig. 113). The iminium is formed by equilibration of the precursor by proton addition at C-16 to yield a carbon with (*S*) stereochemistry. Reduction by ADH10 with the help of the cofactor NADPH at C-17 produces the final product with a *m/z* of 371.

Mutations to the active site showed that Tyr223 is necessary for generation of the main WT product (371 and 353). Mutations to Lys227 did not have such a detrimental effect on the activity of ADH10 compared to the Tyr223 mutation, and mutation to Ser127 did not have much effect on the final product amount of the reaction. This is in line with previous reports of the active site tyrosine being critical for catalysis (Chen et al., 1993, Filling et al., 2002). Therefore this enzyme consists of a classic Ser-Tyr-Lys triad, with the tyrosine being responsible for the catalysis. This is also evident in the crystal structure of ADH10 where we observe Tyr223 and Lys227 in direct contact with the cofactor ribose hydroxyl groups. Furthermore, Tyr223 is coordinating a water molecule in the active site in close proximity to the nicotinamide C-4, which provides the hydride during reduction of the substrate. The substrate appears to undergo hydroxylation at the C-19 position to yield the second hydroxyl group (fig. 114 in section 4.3.4, in lilac). It is not clear if this hydration is enzymatically catalysed or spontaneous.

4.3.3 Comparison to heteroyohimbine biosynthesis

After deglycosylation, strictosidine aglycon can rearrange to produce different backbones. Dialdehyde formation is followed by dehydration and ring closure. While most alkaloids form from cathenamine or dehydrogeissoschizine, the work presented here demonstrates that alkaloids can also form from an alternative ring closure event (fig. 113). In the case of cathenamine the ring closure happens from the aldehyde adjacent to C-21. In the case of vallesiachotamine, antirhine, and vitrosamine the ring D closure happens from the aldehyde adjacent to C-17. This alternative cyclisation gives rise to more chemical diversity and results in a backbone with different features to those found on the yohimbine and heteroyohimbine molecules.

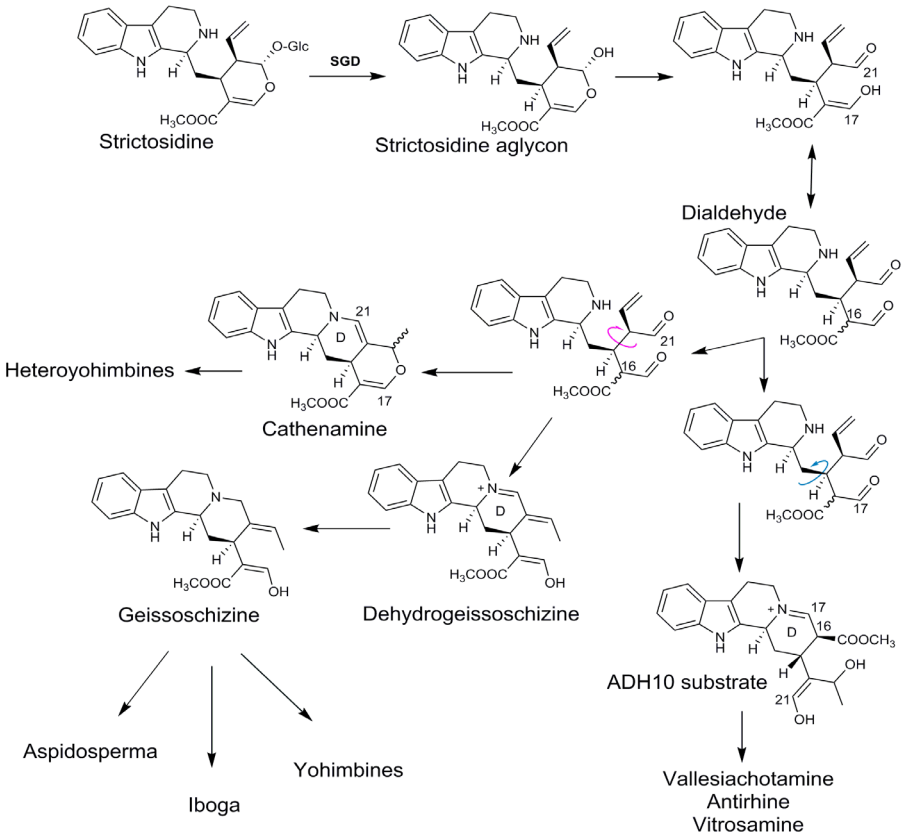


Figure 113: Backbone rearrangement of strictosidine aglycon. Strictosidine aglycon can rearrange to give rise to different backbones. Rotation around the C-15-C-20 bond (pink arrow) and subsequent cyclisation of ring D can give rise to the cathenamine and geissoschizine backbones. On the other hand rotation around the C-14-C-15 bond (blue arrow) and subsequent cyclisation of ring D can give rise to the substrate of ADH10, leading to vallesiachotamine and antirhine.

Interestingly, the major product of ADH10 is the stereoisomer with (*S*) stereochemistry at C-16. The enzyme might be selective for that stereoisomer that spontaneously forms in solution, or as with the HYS enzymes, ADH10 could directly generate this stereocentre by controlling the enamine/iminium tautomerisation.

4.3.4 Mechanism of ADH10 product formation

Equilibration of strictosidine aglycon into the ADH10 pro-vitrosamine substrate can occur in solution. First, the dialdehyde intermediate rotates around the C-14-C-15 bond (fig. 114). Cyclisation by condensation of the amine and C-17 aldehyde occurs and the stereocentre at C-16 can be set at any point thereafter by equilibration between the imine and enamine form of the molecule, just as with cathenamine. Based on the configuration of C-16 of the ADH10 product it appears that either the enzyme or the pro-vitrosamine substrate favours one steric configuration over another, as is observed with the preference of HYS for cathenamine over epi-cathenamine. The ADH10 appears to selectively accept the C-17-N-4 cyclised iminium intermediate from solution rather than cathenamine or 19-epicathenamine. Despite the structural information available for ADH10, it is still unclear what molecular features of the active site are responsible for this substrate selectivity.

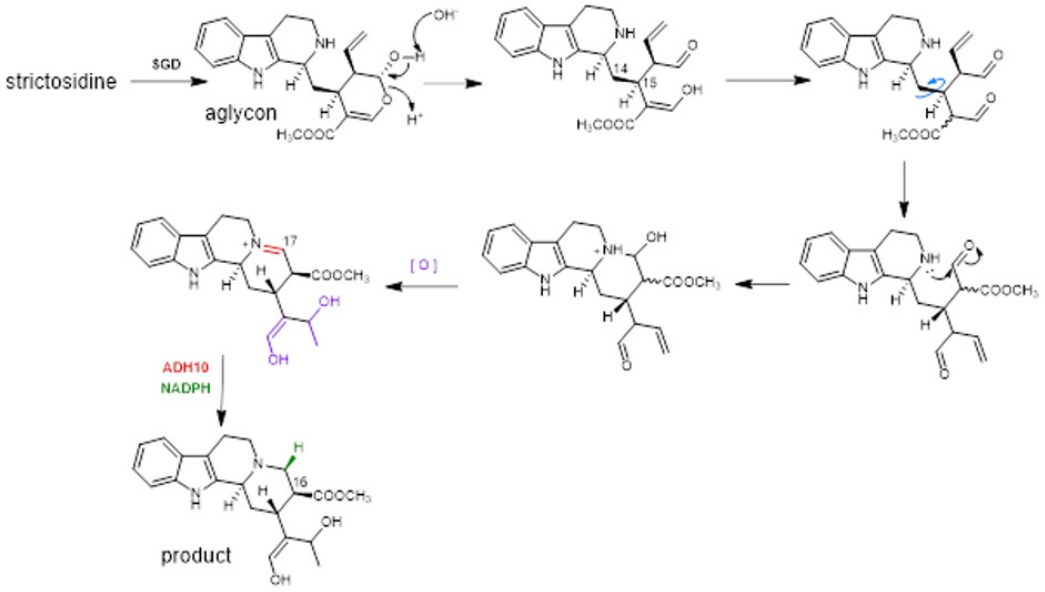
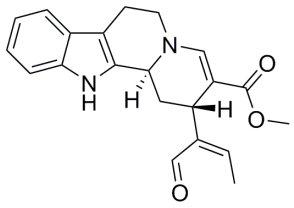


Figure 114: Hypothetical mechanism for generation of the ADH10 product from strictosidine. Strictosidine aglycon is illustrated equilibrating to the dialdehyde form and then rotation of the C-14 and C-15 bond occurs, bringing the aldehyde into proximity of the nitrogen. Cyclisation occurs and loss of water follows which gives rise to the iminium. Hydration of the alkene at C-19 generates the hydroxyl group. The timing of the reaction illustrated in lilac is not known. ADH10 finally catalyses the reduction of the iminium and gives rise to the final product, vitrosamine.



Vallesiachotamine

Figure 115: Vallesiachotamine

Here I propose that vallesiachotamine (fig. 115) is naturally occurring in plants which produce deglycosylated strictosidine, and can be produced from strictosidine aglycon without any need for enzyme activity. The product of ADH10 is a direct reduction product of the deglycosylated strictosidine intermediate which is also the precursor to vallesiachotamine. It is not clear if ADH10 can also reduce vallesiachotamine; to determine that an assay with purified vallesiachotamine would be necessary and unfortunately this compound is not available.

It is not clear if the role of ADH10 *in planta* is to produce the product that is observed *in vitro*. Colleagues attempted to silence ADH10 by Virus Induced Gene Silencing, but no clear metabolic changes were observed. Given the high levels of ADH10 expression and the lack of observable vitrosamine it is reasonable to suggest that reduction of strictosidine aglycon is not the primary role of ADH10 and the activity observed is a spurious discovery. Nevertheless, the discovery of this enzyme enriches our understanding of the complexity of biosynthetic pathways in plants and adds a new enzyme to our biosynthetic toolkit.

4.3.5 Comparison to HYSs of *C. roseus*

ADH10 provides a parallel to the heteroyohimbine synthases discovered and described in the previous chapters. Although it is a member of the SDR family while the HYSs are members of the MDR family, ADH10 also reduces an iminium of deglycosylated strictosidine, as do the HYSs. The lack of formation of any significant amount of heteroyohimbine in the ADH10 reaction indicates it is selective for the alternatively cyclised strictosidine aglycon substrate in its active site. Efforts to dock the supposed substrate into the active site of the ADH10 crystal structure did not result in an unambiguous docked substrate orientation (results not shown). The residues framing the active site, either on the loop or the flap could be interacting with the substrate in a way that is not easy to foresee. Therefore it is difficult to understand how the active site of ADH10 is stereoselective for the alternatively cyclised substrate.

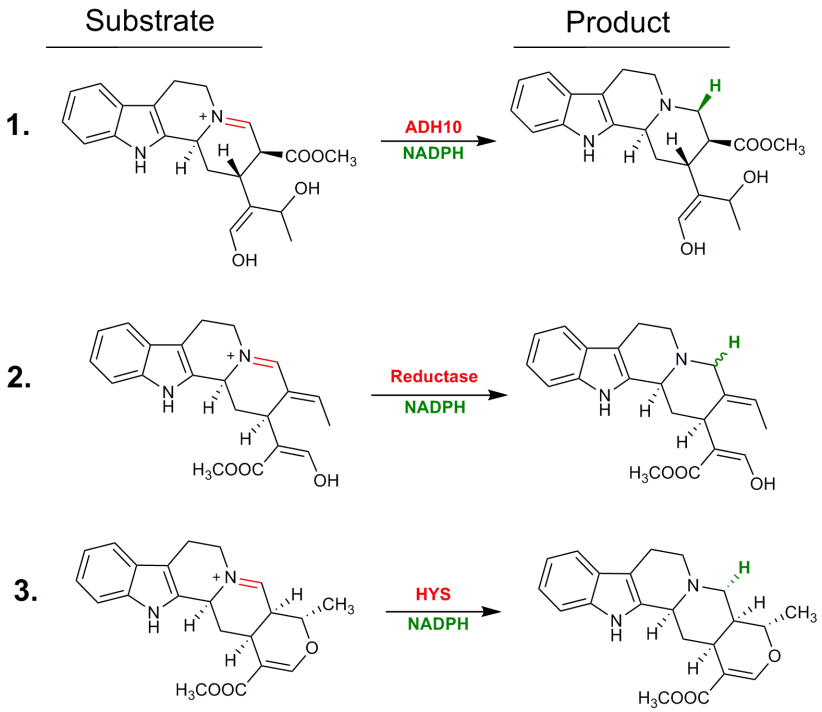


Figure 116: Reduction parallels between ADH10 reaction (1), the proposed (enzyme not yet discovered) 4,21-dehydrogeissoschizine reduction (2), and the HYS reaction (3). The double bond (probably) reduced is indicated in red and the proton originating from NADPH is indicated in green.

ADH10 reduces pro-vitrosamine iminium from the opposite face than the HYSs using the pro-*S* hydride of the NADPH cofactor rather than the pro-*R*. This is a result of the orientation in which the substrate docks in the active site. This could be a guide to determining which type of enzyme could do similar reduction reactions in *C. roseus* and related plants. As the substrate is so similar in all these three cases (fig. 116) it is reasonable to suggest a similar enzyme could be reducing 4,21-dehydrogeissoschizine into geissoschizine (the proposed precursor to many downstream MIA) as well. To find that enzyme a future researcher could start by incubating a crude *C. roseus* protein mixture with pro-*R* or pro-*S* labelled NADPD to determine which type of enzyme carries out the reduction. SDRs typically dock the cofactor in a way that the pro-*S* hydrogen is involved in reduction, and MDRs typically dock the cofactor in the opposite orientation. As demonstrated

here the substrates of these two enzymes also dock in opposite orientations in ADH10 and in HYSs, displaying the opposite face of the molecule to the cofactor. If 4,21-dehydrogeissoschizine (the substrate in reaction 2) binds to the active site in the same manner then the labelling pattern would be indicative of the family of enzyme responsible for the reaction.

4.3.6 Similarities of ADH10 and Salutaridine Reductase

Although the enzymes structures are very similar, ADH10 and PsSalR catalyse different reactions. The high overlap in the active site cavities raises questions about the specificity of the enzymes and whether it is possible that ADH10 could also reduce salutaridine, or if PsSalR could also reduce strictosidine aglycon. Assays on substrate specificity of PsSalR done by Geissler et al. (2007) showed that the enzyme was specific for salutaridine, and stereospecifically produced 7-(*S*)-salutaridinol (fig. 117). Other substrates (intermediates from the morphine biosynthesis pathway) were not reduced by PsSalR, and neither were other plant SDR substrates such as tropinone and menthone. Strictosidine aglycon was not tested as substrate, but given the enzymes preference for salutaridine and disfavour of other morphine biosynthesis intermediates it would be very surprising if it accepted the substrate and catalysed the imine reduction. Salutaridine (or any other morphine biosynthesis intermediate) was not tested as substrate for ADH10; therefore it is not possible to draw any conclusions on its substrate specificity.

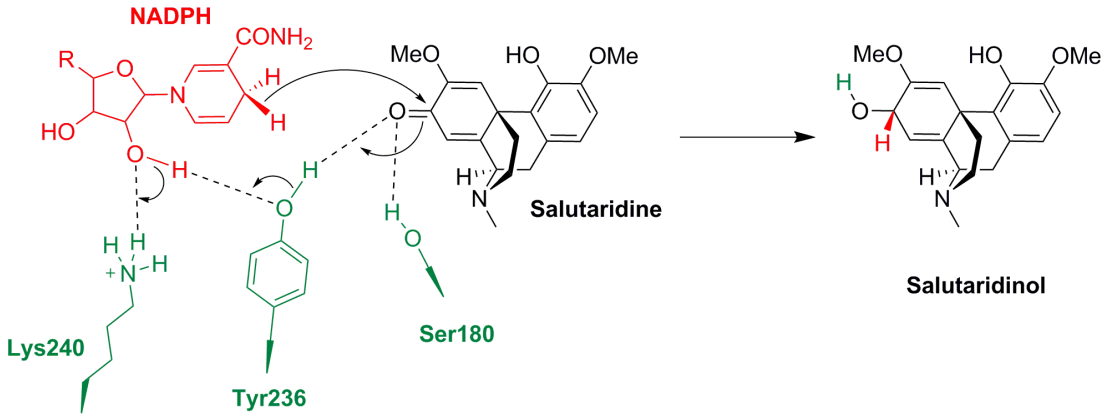


Figure 117: Proposed mechanism of salutaridine reduction by PsSalR. Figure redrawn from Geissler et al. (2007). The NADPH cofactor is drawn in red; the active site triad residues in green; substrate and product in black. The colour of the hydrogens of the product reflect where they originate from.

4.3.7 Precedence for SDR imine reductases

Reductases have great importance in the chemical synthesis industry, especially for generation of chiral products. There is interest in development of stereoselective imine reductases as biocatalysts (Scheller et al., 2014). Extensive studies have been conducted on imine reductases (IREDs) of *Streptomyces* sp. which are also members of the SDR enzyme family. Two IREDs with opposite stereoselectivity were discovered in two different *Streptomyces* strains, an (*S*)-IRED (Mitsukura et al., 2013), and an (*R*)-IRED (Mitsukura et al., 2011). Screening of the two enzymes against a panel of imine molecules revealed that they are promiscuous, and both have varying

amounts of preference for the different substrates and also different degrees of stereospecificity for the final products (Leipold et al., 2013, Hussain et al., 2015). Interestingly, for substrates tested for both IRED enzymes, the resulting products were usually of opposite configuration, suggesting that the enzymes were binding the substrates consistently in opposite orientations.

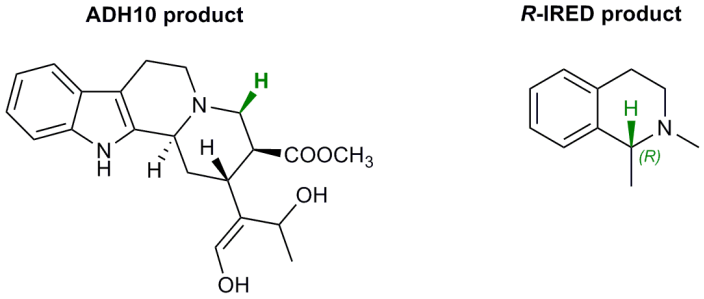


Figure 118: Comparison of orientation and position of hydride in ADH10 and (*R*)-IRED products. The hydride originating from NADPH is indicated in green in each case. The ADH10 catalyses the reduction to yield a pro-*R* hydride, as does (*R*)-IRED.

Hussain et al. (2015) were able to demonstrate that (*R*)-IRED is able to reduce an imine of a molecule similar to the MIA (fig. 118, right). This enzyme was capable of producing the (*R*) isomer with 74 % enantiomeric excess. However, during analysis of the ADH10 reaction products by LC-MS and NMR only the pro-*R* isomer has been detected. This suggests the binding constraints on the ADH10 substrate are larger than the constraints posed on (*R*)-IRED and the tested substrate. Binding of the ADH10 substrate in an opposite orientation would be non-conductive to successful reduction.

The Imine Reductase database (www.ired.biocatnet.de (Scheller et al., 2014, Fademrecht et al., 2016)) lists 12 sequences of IREDs originating from Eukaryotes. All the 12 sequences are from Eurotiomycetidae, more specifically from the families Aspergillaceae and Ajellomycetaceae. A BLAST search using ADH10 with a generous cutoff value of e-10 did not result in any hits. *C. roseus* ADH10 shares very low amino acid identity with these eukaryotic IREDs (maximum of 15.6 % with *Aspergillus oryzae* IRED). This could indicate that ADH10 has evolved its imine reducing capability independently of these fungal IREDs. As only a handful of eukaryotic IREDs (other than the HYSs and the ADH10 described here) have been described so far it is not possible to draw any concrete conclusions on this subject yet.

Some IREDs have been crystallised (for example PDB: 4OQY (Huber et al., 2014), and PDB: 3ZHB (Rodríguez-Mata et al., 2013)) and although they share some similarities with SDRs they have diverged from the SDRs quite considerably. Apart from the NADPH-binding domain, which contains the Rossmann fold, the IREDs possess a distinct dimerization domain at the C-terminal end composed mainly of α-helices. This is not found in the *C. roseus* ADH10 structure, although the “flap” appears to be promoting an enzyme dimerization in the crystal structure.

Interestingly, there is an SDR enzyme in the BIA biosynthetic pathway which appears to reduce an iminium moiety. Sanguinarine Reductase (SanR) from California poppy (*Eschscholzia californica*)

reduces sanguinarine to produce dihydrosanguinarine (Weiss et al., 2006) (fig. 119). This enzyme was not crystalized but instead a homology model was built based on the closest homolog (PDB accession code: 1XQ6, an enzyme from *Arabidopsis thaliana* of unknown function)(Vogel et al., 2010). The alkanolamine form of the substrate (fig. 120) was docked into the structure, but it is not clear if docking of sanguinarine was also attempted but unsuccessful. The authors therefore propose an unusual mechanism of reduction that involves the alkanolamide form of the substrate and the serine of the catalytic triad (fig. 120).

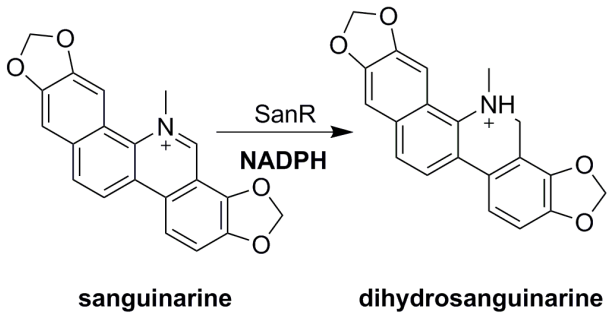


Figure 119: Reduction catalysed by SanR.

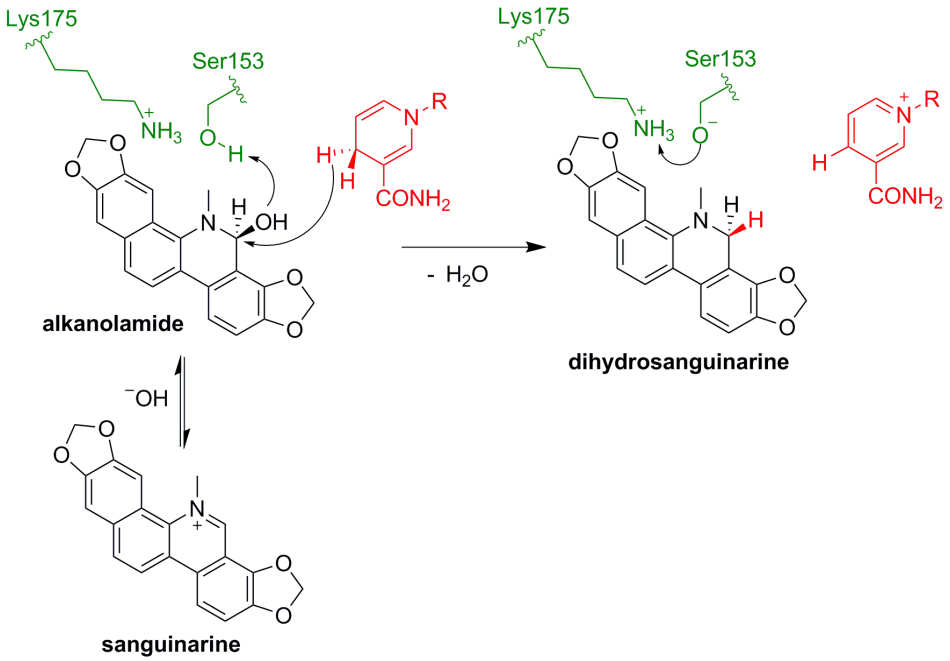


Figure 120: Proposed mechanism of SanR. Redrawn from Vogel et al. (2010). The active site residues are illustrated in green, the NADPH and the hydrogen originating from it are in red, and the substrate and product in black.

It is not reported if the workers used labelling to determine which cofactor hydride was transferred to the product (the pro-*S* or the pro-*R*), but given their molecular model it is more likely that the pro-*S* hydride is transferred, and therefore their mechanism proposal was probably mistaken. Likewise, it is not reported if labelling of the product was attempted. Consequently, it is not possible to say with any certainty without any labelling studies which face of the molecule the

hydride is oriented towards. If labelling had been carried out it would provide more evidence for the accurateness of the docking.

A mutagenesis was carried out on EcSanR to determine which residues are necessary for catalysis and substrate binding (Vogel et al., 2010). Of the catalytic triad only the serine was mutated (to alanine) but it is not clear why the workers did not mutate the other residues of the catalytic triad. Mutation to the serine (Ser153Ala) resulted in a reduction of catalytic efficiency (4 % V_{max}/K_m compared to 100 % of WT). Without studying the effect of mutations on the other two residues of the catalytic triad it is not possible to say with certainty that Ser153 is the residue responsible for catalysis. Such a mutation could have an effect on the active site dynamics, such as charge and substrate docking, and therefore the reduction in catalytic efficiency could be due to different factors. It is possible that EcSanR is in fact an imine reductase, and that it uses either the active site serine or tyrosine to catalyse the reaction in the same manner that ADH10 does.

A protozoan SDR enzyme, Pteridine Reductase (PTR1) from the parasites *Leishmania* and *Trypanosoma* (PTR1, (Gourley et al., 2001)), is an interesting detoxification enzyme which reduces an imine. This enzyme is involved in the resistance to dihydrofolates which are used as inhibitors for the essential metabolism of folates in these parasites. PTR1 is capable of reducing its substrate in two distinct ways (fig. 121). The first reaction is a reduction of an imine involving the classic SDR tyrosine-based catalysis. The second reduction, also of an imine, involves a coordinated water molecule and an arginine present on the opposite side of the active site to the tyrosine. In the first reduction the proton originates from tyrosine, but in the second reduction the two protons originate from the coordinated water molecule and the coordinated cofactor phosphate group.

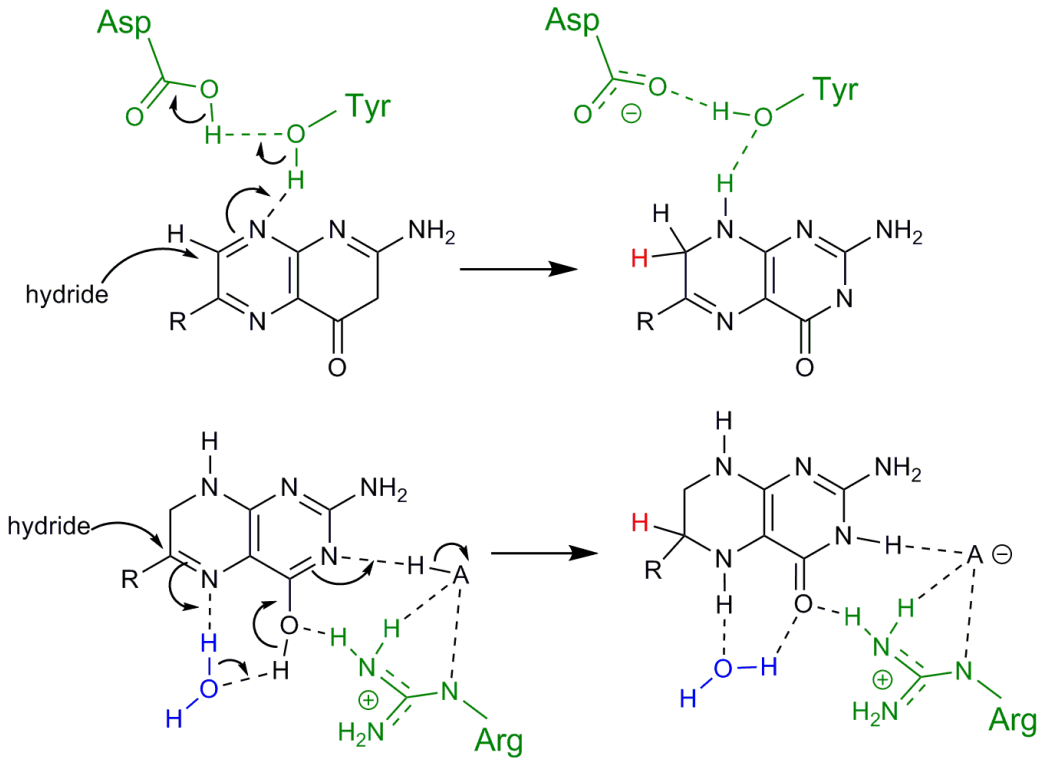


Figure 121: Double reduction of dihydrofolate by PTR1. Redrawn from Gourley et al. (2001). The hydride originating from the NADPH cofactor is indicated in red; the water molecule implicated in the 2nd reduction is illustrated in blue; the active site residues implicated in the catalysis are illustrated in green; A represents the cofactor phosphate that interacts with the substrate.

PTR1 was crystallised as a complex with the cofactor and substrate and it appears to dock in a parallel orientation to the cofactor nicotinamide ring (Gourley et al., 2001). This enzyme does not have a "flap" like ADH10 or PsSalR and therefore direct parallels cannot be drawn between these. However, the mechanism proposed for the first reduction is similar to the proposed mechanism for reduction of pro-vitrosamine by ADH10.

4.4 Conclusions

The discovery of ADH10 from *C. roseus* indicates that an SDR can reduce strictosidine aglycon, and therefore should also be included in screening efforts for discovery of biosynthetic enzymes implicated in MIA pathways both in *C. roseus* and also in other MIA-producing plants. The product of ADH10, vitrosamine, is not detected in tissues of *C. roseus*, and has in fact never been described in the literature. There could be different factors for this, but an obvious explanation is that it is an "unnatural natural product" in that the product is not formed normally *in vivo* but can be produced *in vitro*.

ADH10 could be responsible for the reduction of another substrate *in vivo*, but Virus Induced Gene Silencing of ADH10 done by collaborators in the past did not show any significant decrease in any of the major MIA of *C. roseus*. ADH10 could be acting upon a much less abundant substrate, or its product could be transient and degraded or metabolised very fast *in vivo*. What's more, this *in vitro* product could be the result of a non-specific reaction and ADH10's true substrate could

be a similar but unstable product of another enzyme (such as a P450, or another reductase) acting on strictosidine aglycon. This phenomenon has been observed previously with secondary metabolism steps in MIA biosynthesis (Krithika et al., 2015, Qu et al., 2015). On the other hand, ADH10 could be part of a strictosidine aglycon detoxification suite of enzymes, as discussed for the HYSs in Chapter 3. Enzyme redundancy (as seen with the 5 HYSs) could mask any effect the Virus Induced Gene Silencing might have produced. There are no highly similar enzymes to ADH10 in the *C. roseus* genome, but even a low identity of ~53 % could not exclude the possibility that another enzyme could act as a candidate for the same reaction. This is what is seen with the HYSs, which do display a relatively low percentage of protein identity.

The similarity between ADH10 and the benzylisoquinoline biosynthetic enzyme PsSalR from the distantly related opium poppy is intriguing. This hints that perhaps this enzyme, or something similar, could be found in other organisms throughout the plant kingdom. A further investigation into other plants could reveal the extent to which enzymes with the lid and loop features are ubiquitous or not.

Why this branch of strictosidine-derived MIA is not as well developed as that of the corynantheine, iboga, and aspidosperma branches could stem from the fact that the substrate is not present in high enough amounts. As the experiments described in this chapter show, cathenamine forms approximately 14-20 % of the strictosidine aglycon pool whereas the ADH10 substrate forms just 2.4-4 % of the pool. The inherent reactivity and equilibration speed of the different forms of strictosidine aglycon governs the more abundant species and backbones which the plant can then utilise and build upon.

The discovery of this enzyme capable of reducing a different form of strictosidine aglycon opens the door to molecular engineering. This enzyme adds to the biosynthetic toolkit available for heterologous expression and production of “unnatural natural products”. It also expands the already large chemical diversity possible from the central MIA precursor, strictosidine.

4.5 Materials and Methods

4.5.1 General molecular biology techniques used

Polymerase chain reaction, description of vectors used, RNA extraction and cDNA production, and details of the *E. coli* culture media used are detailed in Chapter 2.

4.5.2 Cloning of ADH10

Primers for amplification of the gene identified by Glenn (2013) from *C. roseus* cDNA were designed based on the ORF of transcript Cro013448 (ADH10) with pOPINF (Berrow et al., 2007) overhangs 5’ AAGTTCTGTTTCAGGGCCCGGCCGCCATGGGTACC and 3’ ATGGTCTAGAAAGCTTTATTCAAACGATGACTCCTCGC for directional cloning of ADH10 into the pOPINF vector. PCR was performed as detailed in Chapter 2 and the amplified gene fragment

was gel-purified from 1% agarose gel (Promega, UK) and cloned into the *E. coli* expression vector pOPINF following the manufacturer's instructions for ligation. Vectors were transformed into chemically competent *E. coli* Top10 cells by heat shock at 42˚C for 30 seconds and then spread-plated onto LB+agar plates supplemented with carbenicillin (100 µg/mL). After a night of growth at 37 ˚C positive clones were identified by PCR using the gene-specific primers above and the HotStart Taq polymerase (Qiagen, UK). Positive colonies were grown overnight in 4 mL of LB media supplemented with the appropriate antibiotics at 37 ˚C. The following day plasmids were isolated from the cultures using the miniprep kit (Qiagen) according to manufacturer's instructions. Vectors were verified visually by running on a 1% agarose gel at 100 V and the identity of the inserted sequence was confirmed by Sanger sequencing.

4.5.3 ADH10 expression assay

Chemically competent SoluBL21 *E. coli* cells (Novagen®, Merck Millipore, Massachusetts, USA) were transformed by heat shock at 42 ˚C with pOPINF plasmid containing the ADH10 coding sequence in frame with the N-terminal His_6-tag. Expression of His_6-tagged ADH10 protein in *E. coli* was tested in small scale at various temperatures and with different IPTG concentrations. Transformed cells were added to 50 mL starter culture of liquid LB media supplemented with carbenicillin (100 µg/mL) and incubated at 37 ˚C overnight with 200 rpm shaking. The cells were diluted 1:100 in 100 mL of 2 x YT media supplemented with carbenicillin (100 µg/mL) and allowed to grow at 37 ˚C with 200 rpm shaking. Two cultures were grown at 37 ˚C until an OD_{600} of 1.2 and were induced by 1 and 0.1 mM IPTG respectively and allowed to express protein at 37 ˚C for 3.5 h. Cultures for expression at 18 or 10 ˚C were allowed to reach an OD_{600} of approximately 0.6 and then chilled on ice before addition of 1, 0.5, or 0.1 mM IPTG. The cultures at lower temperature were incubated overnight at the respective temperatures for protein expression. The cultures at 37 ˚C were removed after 3.5 h and cells were pelleted by centrifugation at 4000 rpm and the supernatant removed. Cell pellets were stored at -80 ˚C until testing the next day. Cells were analysed for protein expression using the BugBuster® kit (Merck Millipore, UK) following manufacturer's instructions. An aliquot of each soluble fraction was analysed by SDS-PAGE.

4.5.4 ADH10 large-scale expression and purification

Cultures of SoluBL21 *E. coli* harbouring the pOPINF vector with ADH10 (2 L of 2 x YT) were prepared as described above. These were allowed to grow to an OD_{600} of 1.2 at 37 ˚C before inducing with 0.1 mM IPTG. Expression was done at 37 ˚C with 200 rpm shaking for 3.5 h, after which time the cultures were chilled on ice and then spun down at 4000 rpm for 10 min to pellet the cells. The pellet was resuspended in 200 mL of Phosphate Buffered Saline solution and pelleted again by centrifugation at 4000 rpm for 10 min. The supernatant was removed and the cell pellet was stored at -80 ˚C overnight and the next day they were thawed and resuspended in 100 mL Buffer A (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 20 mM imidazole) supplemented with 2 mM of 2-β-mercaptoethanol (Sigma) along with EDTA-free protease inhibitor (Roche Diagnostics Ltd.). Cells were lysed using sonication for 4 minutes on ice

using 2 second pulses with a 5 minute pause after 2 minutes on ice. Cell debris was pelleted by centrifugation at 17 x 1000 g for 20 min and all large-scale purification steps were done at 4 ˚C.

His_6-tagged ADH10 was purified on an ÄKTAxpress purifier (GE Healthcare) using a HisTrap FF 5 mL column (GE Healthcare) equilibrated with Buffer A. The sample was loaded at a flow rate of 4 mL/min and step-eluted using Buffer B (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 500 mM imidazole) supplemented with 2 mM of 2-β-mercaptoethanol (Sigma). Eluted protein was subjected to further purification on a Superdex Hiload 26/60 S75 gel filtration column (GE Healthcare) at a flow rate of 3.2 mL/min using Buffer D (20mM Hepes, pH 7.5 150mM NaCl) also supplemented with 2 mM of 2-β-mercaptoethanol (Sigma) and collected by a fractionator into 8 mL fractions. The fractions were analysed by SDS-PAGE and those containing no traces of other contaminating proteins were pooled. The protein was concentrated in a 10 KDa cutoff Millipore filter (Merck Millipore) and buffer exchanged into Buffer D supplemented with 0.5 mM TCEP. The concentration was measured spectroscopically using the calculated MW and extinction coefficient (21680 L $mol^{-1}cm^{-1}$).

4.5.5 Crystallisation and data collection

Purified protein at a concentration of 10.76 mg/mL was used for crystallisation assays in the commercial suites PEG (Qiagen), PACT (Qiagen), and JCSG+ (Molecular Dimensions). Crystallisation screens were conducted by sitting-drop vapour diffusion in MRC2 96-well crystallization plates (Swissci) with a mixture of 0.3 µL well solution from the suites and 0.3 µL protein solution. One drop was setup without cofactor (apo) and one drop was setup to contain $NADP^+$ (Sigma Aldrich) at a final concentration of 1 mM for co-crystallization studies. Solutions were dispensed by an OryxNano robot (Douglas Instruments). Prepared screens were placed at 19 ˚C for crystal formation. Optimisation screens were setup using the plate dispensing function of the Oryx8 robot (Douglas Instruments). Crystals were cryo-protected with well solution containing 25% ethylene glycol for 1 minute and frozen in liquid nitrogen.

Data collection was carried out essentially as described in Chapter 3 with some modifications. The structure was solved by molecular replacement with the crystal structure of PsSalR (3O26, Higashi et al. (2011)) using PHASER (McCoy et al., 2007). The asymmetric unit corresponded to two monomers in a back-to back orientation, but after consideration one of the chains was placed in the symmetry position to produce the head-to-head orientation described here which had a larger contact area. The structure was refined in REFMAC5 with TLS refinement (Winn et al., 2003).

4.5.6 Accurate mass and fragment determination

A sample of purified ADH10 product was direct-injected on the Synapt G2 HDMS mass spectrometer (Waters). The sample was infused at 5-10 µL min-1 using a Harvard Apparatus syringe pump. The instrument was calibrated using a sodium formate solution. The sample was analysed for 2 min with a scan time 1 sec in the mass range of 50-600 *m/z*. Capilary voltage was

3.5 V, cone voltage 40 V, source temperature 120 ˚C, desolvation temperature 350 ˚C desolvation gas flow 800 L/h. Leu-enkephaline peptide (1 ng/µL) was used to generate a dual lock-mass calibration with [M + H]= 556.2766 and *m/z*=278.1135 measured every 10 sec. For MS2, the precursor ion of *m/z* 371 was selected and fragmented with collision energy of 20. Spectra were generated in Masslynx 4.1 by combining a number of scans, and peaks were centred using automatic peak detection with lock mass correction. Accurate mass of product was found to be 353.1862 *m/z* and 371.1967 *m/z* for the two major species present in solution.

4.5.7 Large scale reaction with deglycosylated strictosidine and product purification

Strictosidine (30 mg) was incubated in 100 mL (final concentration 566 µM) with 50 mM citrate buffer pH 6, and 40 µL of SGD (10 nM). NADPH and ADH10 were added (final concentration 700 µM and 3 µM respectively) and the reaction incubated at 37˚ C with gentle shaking. Reaction completion was followed by mass-spectrometry, and at 2 h extra ADH10 was added (concentration brought to 6 µM).

The reaction was basified to stop the reaction at 4 h by addition of NaOH to pH of approximately 9.5 and the product was extracted multiple times with a total of 100 mL of ethyl acetate (EtOAc). The EtOAc fraction was dried under vacuum and resuspended in 100 µL of EtOAc. A nano-silica TLC was basified with tri-ethylamine (TEA) and the EtOAc product was loaded onto the plate. Some residue was still visible in the product vial and therefore it was resuspended in 100 µL of methanol and loaded onto the TLC as well. The TLC was ran in 50:50:1 EtOAc:Hexanes:TEA.

The product band was excised from the silica with a scalpel and the silica was crushed in the presence of 50 mL EtOAc and the solution was filtered and dried under vacuum overnight. The precipitate was resuspended in 300 µL of $CDCl_3$ and a ^{1}H-NMR was recorded. For unambiguous assignment of the structure ^{1}H-H correlated spectroscopy (COSY), heteronuclear multiple quantum coherence experiment via direct coupling (HMQC), and heteronuclear multiple bond correlation spectrum (HMBC) were recorded by collaborators on a Bruker Avance III HD 700 NMR Spectrometer (16.4 T , ^{1}H operating frequency 700 MHz) equipped with TCI H-C/N-D 1.7 mm microcryoprobe.

4.5.8 Deuterium labelling of ADH10 product

Deuterated ADH10 product was produced by using deuterated cofactor (NADPD). Strictosidine (15 mg) was incubated in HEPES buffer (50 mM, pH 7.0) with 200 µM $NADP^+$ in a total volume of 141 mL. NADPD was generated *in situ* by using 1-D deuterated D-glucose (250 µM Cambridge Isotope Laboratories Inc., USA) and Glucose dehydrogenase (800 U, from *Pseudomonas sp*., Sigma) as an NADPH regeneration system. Purified SGD and ADH10 (1 µM final concentration) were added and the reaction was incubated at 31 ˚C for 16 hours with 70 rpm shaking.

The reaction was basified to approximately pH 9 by addition of saturated NaOH and alkaloids were extracted multiple times in a total volume of 120 mL of EtOAc. The extract was dried using

a rotary evaporator. A high-resolution silica TLC plate (nano-silica plate, Sigma-Aldrich) was pre-soaked with TEA and the sample was loaded and ran using 50:50:1 EtOAc:Hex:TEA. The highest band was excised with a scalpel and crushed multiple times in a total volume of 50 mL MeOH and filtered into a clean flask. The extract was dried using a rotary evaporator and then placed on a high vacuum pump overnight as it was noted that the extracted compound is hygroscopic. The sample was resuspended in 500 µL of $CDCl_3$ and the ^1H and COSY were measured on a Bruker Advance NMR instrument operating at 400 MHz for ^1H equipped with a BBFO plus 5 mm probe.

4.5.9 pH assays of ADH10 with strictosidine aglycon

To determine the best pH/buffer combination for use during ADH10 reactions, a small panel of buffers was assayed. Strictosidine (300 µM) was incubated with purified SGD in deionised water for 15 min at room temperature. Then the reaction was split into eight tubes containing different buffers. The buffers used are (in order of increasing pH): Phosphate pH 5.0, MES pH 5.5, Phosphate pH 6.0, Citrate pH 6.0, MES pH 6.5, Phosphate pH 7.0, HEPES pH 8.0, and Phosphate pH 8.0. All buffers were used at a concentration of 50 mM. NADPH (500 µM) was added, followed by 1 µM of purified ADH10. The reactions were mixed and an aliquot was taken at 1 min and at 10 min and prepared as usual for analysis by LC-MS.

The analysis was carried out on a UPLC (Waters) equipped with an Aquity BEH C18 1.7 µm 2.1 x 50 mm column connected to Xevo TQS (Waters). The capillary cone volatage, the source temperature, cone and desolvation gas flows were all the same as described in Chapter 2. The LC-MS analysis was done as described in Chapter 2 using both the “fast method” and the “separation method”. A calibration curve specific for the ADH10 product was not made as the product is not available commercially. Therefore absolute quantification of the product in each sample is not possible. However, as the product has a signal at 353 *m/z* and it fragments into the same fragment as THA and AJM it is detected by the same MRM used to detect the heteroyohimbines. A manual integration of the peaks was done using the MassLynx software (Waters) using 200 peak to peak amplitude for noise reduction.

4.5.10 Spectroscopic assays

Strictosidine (100, 75, 50 or 25 µM) was deglycosylated by purified SGD (10 nM) for 30 minutes at 30 ˚C in a spectrophotometer cuvette in a total volume of 800 µL of citrate buffer pH 6.0 (50 mM). The completion of the reaction was verified by mass-spectrometry. NADPH (100 µM) was added to the reaction and mixed by pipetting. The reaction was monitored at 340 nm on a spectrophotometer (Cary 50 Bio, Varian) at room temperature. After verifying the NADPH absorbance was stable (approximately 3-5 minutes) ADH10 (100 nM) was added and mixed by pipetting. The initial velocity was recorded using the least squares regression of the Cary WinUV Kinetics Application v. 3.00(182) (Varian).

4.5.11 Combination of ADH10 and THAS1

A spectrophotometric assay was done in similar conditions to those described above. Strictosidine (100 µM) was deglycosylated by SGD at 30 ˚C in a spectrophotometer cuvette for 30 minutes in citrate buffer (50mM, pH 6.0). NADPH (100 µM) was added to the reaction and mixed by pipetting. After verifying that NADPH absorbance was stable, purified THAS1 (1 µM) was added and mixed. The reaction was allowed to progress until the plateau was reached (approx. 7 minutes), and then ADH10 (1 µM) was added to the reaction and mixed and the reaction was allowed to reach the steady-state. The difference in NADPH (ΔABS_{340}) was recorded before addition of each enzyme, and at the point the reaction reached the plateau. The assay was done in triplicate. The inverse reaction (ADH10 added first, then THAS1) was done once to verify the reverse is also true.

4.5.12 Active site mutations

The crystal structure revealed several amino acids in the active site cavity that could potentially be part of the catalytic triad. To verify their role in catalysis these three residues were each mutated to alanine by mutating the codon (Ser167->Ala167: TCC->GCC; Tyr223->Ala223: TAT->GCT; Lys227->Ala 227: AAA->GCA). The mutant ADH10 gene fragments were obtained from IDT with the pOPINF overhangs included. Ligation into pOPINF and clone selection was done as described above. Small-scale expression of mutants was done at 37 ˚C as described above as well, but purified on small-scale as described for the candidate enzymes in Chapter 2.

4.6 References

ARNOLD, K., BORDOLI, L., KOPP, J. & SCHWEDE, T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22, 195-201.

BERROW, N. S., ALDERTON, D., SAINSBURY, S., NETTLESHIP, J., ASSENBERG, R., RAHMAN, N., STUART, D. I. & OWENS, R. J. 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research*, 35.

CHEN, Z., JIANG, J. C., LIN, Z. G., LEE, W. R., BAKER, M. E. & CHANG, S. H. 1993. Site-specific mutagenesis of Drosophila alcohol dehydrogenase: Evidence for involvement of tyrosine-152 and lysine-156 in catalysis. *Biochemistry*, 32, 3342-3346.

DJERASSI, C., MONTEIRO, H. J., WALSER, A. & DURHAM, L. J. 1966. Alkaloid Studies. LVI.1 The Constitution of Vallesiachotamine. *Journal of the American Chemical Society*, 88, 1792-1798.

FADEMRECHT, S., SCHELLER, P. N., NESTL, B. M., HAUER, B. & PLEISS, J. 2016. Identification of imine reductase-specific sequence motifs. *Proteins: Structure, Function, and Bioinformatics*, 84, 600-610.

FENG, T., CAI, X.-H., LIU, Y.-P., LI, Y., WANG, Y.-Y. & LUO, X.-D. 2010. Melodinines A−G, Monoterpenoid Indole Alkaloids from Melodinus henryi. *Journal of Natural Products*, 73, 22-26.

FILLING, C., BERNDT, K. D., BENACH, J., KNAPP, S., PROZOROVSKI, T., NORDLING, E., LADENSTEIN, R., JÖRNVALL, H. & OPPERMANN, U. 2002. Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *Journal of Biological Chemistry*, 277, 25677-25684.

GAVIDIA, I., TARRÍO, R., RODRÍGUEZ-TRELLES, F., PÉREZ-BERMÚDEZ, P. & SEITZ, H. U. 2007. Plant progesterone 5β-reductase is not homologous to the animal enzyme. Molecular evolutionary characterization of P5βR from Digitalis purpurea. *Phytochemistry*, 68, 853-864.

GEISSLER, R., BRANDT, W. & ZIEGLER, J. 2007. Molecular Modeling and Site-Directed Mutagenesis Reveal the Benzylisoquinoline Binding Site of the Short-Chain Dehydrogenase/Reductase Salutaridine Reductase. *Plant Physiology*, 143, 1493-1503.

GEU-FLORES, F., SHERDEN, N. H., COURDAVAULT, V., BURLAT, V., GLENN, W. S., WU, C., NIMS, E., CUI, Y. & O'CONNOR, S. E. 2012. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature*, 492, 138-142.

GLENN, W. 2013. Understanding and manipulating alkaloid biosynthesis, Chapter 2: Discovery of 10-hydroxygeraniol Oxidoreductase Activity in *C. roseus*. Massachusetts Institute of Technology.

GOURLEY, D. G., SCHUTTELKOPF, A. W., LEONARD, G. A., LUBA, J., HARDY, L. W., BEVERLEY, S. M. & HUNTER, W. N. 2001. Pteridine reductase mechanism correlates pterin metabolism with drug resistance in *trypanosoma*tid parasites. *Nature Structural Biology*, 8, 521-525.

GUEX, N., PEITSCH, M. C. & SCHWEDE, T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis, 30, S162-S173.

HIGASHI, Y., KUTCHAN, T. M. & SMITH, T. J. 2011. Atomic Structure of Salutaridine Reductase from the Opium Poppy (Papaver somniferum). *Journal of Biological Chemistry*, 286, 6532-6541.

HUBER, T., SCHNEIDER, L., PRÄG, A., GERHARDT, S., EINSLE, O. & MÜLLER, M. 2014. Direct Reductive Amination of Ketones: Structure and Activity of S-Selective Imine Reductases from Streptomyces. *ChemCatChem*, 6, 2248-2252.

HUSSAIN, S., LEIPOLD, F., MAN, H., WELLS, E., FRANCE, S. P., MULHOLLAND, K. R., GROGAN, G. & TURNER, N. J. 2015. An (R)-Imine Reductase Biocatalyst for the Asymmetric Reduction of Cyclic Imines. *ChemCatChem*, 7, 579-583.

JOKELA, R., HALONEN, M. & MAURI, L. 1993. Predominant conformations of Na-boc-deformyl-Z-

and Na-boc-deformyl-E-geissochizine, the latter a possible synthetic intermediate in the preparation of sarpagan and ajmalan ring systems. *Tetrahedron*, 49, 2567-2576.
JÖRNVALL, H., HEDLUND, J., BERGMAN, T., OPPERMANN, U. & PERSSON, B. 2010. Superfamilies SDR and MDR: From early ancestry to present forms. Emergence of three lines, a Zn-metalloenzyme, and distinct variabilities. *Biochemical and Biophysical Research Communications*, 396, 125-130.
JÖRNVALL, H., HÖÖG, J.-O. & PERSSON, B. 1999. SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Letters*, 445, 261-264.
JÖRNVALL, H., PERSSON, B., KROOK, M., ATRIAN, S., GONZALEZ-DUARTE, R., JEFFERY, J. & GHOSH, D. 1995. Short-chain dehydrogenases/reductases (SDR). *Biochemistry*, 34, 6003-6013.
KAN-FAN, C. & HUSSON, H.-P. 1979. Isolation and biomimetic conversion of 4,21-dehydrogeissoschizine. *Journal of the Chemical Society, Chemical Communications*, 1015-1016.
KAVANAGH, K. L., JÖRNVALL, H., PERSSON, B. & OPPERMANN, U. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families: The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cellular and Molecular Life Sciences*, 65, 3895-3906.
KIEFER, F., ARNOLD, K., KUNZLI, M., BORDOLI, L. & SCHWEDE, T. 2009. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*, 37, D387-D392.
KRITHIKA, R., SRIVASTAVA, P. L., RANI, B., KOLET, S. P., CHOPADE, M., SONIYA, M. & THULASIRAM, H. V. 2015. Characterization of 10-Hydroxygeraniol Dehydrogenase from Catharanthus roseus Reveals Cascaded Enzymatic Activity in Iridoid Biosynthesis. *Scientific Reports*, 5, 8258.
LEIPOLD, F., HUSSAIN, S., GHISLIERI, D. & TURNER, N. J. 2013. Asymmetric Reduction of Cyclic Imines Catalyzed by a Whole-Cell Biocatalyst Containing an (S)-Imine Reductase. *ChemCatChem*, 5, 3505-3508.
LYGIDAKIS, A., KARUPPIAH, V., HOEVEN, R., NÍ CHEALLAIGH, A., LEYS, D., GARDINER, J. M., TOOGOOD, H. S. & SCRUTTON, N. S. 2016. Pinpointing a Mechanistic Switch Between Ketoreduction and "Ene" Reduction in Short-Chain Dehydrogenases/Reductases. *Angewandte Chemie International Edition*, 55, 9596-9600.
MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *Journal of applied crystallography*, 40, 658-674.
MITSUKURA, K., KURAMOTO, T., YOSHIDA, T., KIMOTO, N., YAMAMOTO, H. & NAGASAWA, T. 2013. A NADPH-dependent (S)-imine reductase (SIR) from Streptomyces sp. GF3546 for asymmetric synthesis of optically active amines: purification, characterization, gene cloning, and expression. *Applied Microbiology and Biotechnology*, 97, 8079-8086.
MITSUKURA, K., SUZUKI, M., SHINODA, S., KURAMOTO, T., YOSHIDA, T. & NAGASAWA, T. 2011. Purification and Characterization of a Novel (R)-Imine Reductase from Streptomyces sp GF3587. *Bioscience Biotechnology and Biochemistry*, 75, 1778-1782.
MOUMMOU, H., KALLBERG, Y., TONFACK, L., PERSSON, B. & REST, B. 2012. The Plant Short-Chain Dehydrogenase (SDR) superfamily: genome-wide inventory and diversification patterns. *BMC Plant Biology*, 12, 1-17.
MUNKERT, J., POLLIER, J., MIETTINEN, K., VAN MOERKERCKE, A., PAYNE, R., MÜLLER-URI, F., BURLAT, V., O'CONNOR, SARAH E., MEMELINK, J., KREIS, W. & GOOSSENS, A. 2015. Iridoid Synthase Activity Is Common among the Plant Progesterone 5β-Reductase Family. *Molecular Plant*, 8, 136-152.
PORTSTEFFEN, A., DRAEGER, B. & NAHRSTEDT, A. 1992. Two tropinone reducing enzymes from Datura stramonium transformed root cultures. *Phytochemistry*, 31, 1135-1138.
QU, Y., EASSON, M. L. A. E., FROESE, J., SIMIONESCU, R., HUDLICKY, T. & DE LUCA, V. 2015. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proceedings of the National Academy of Sciences*, 112, 6224-6229.

RODRÍGUEZ-MATA, M., FRANK, A., WELLS, E., LEIPOLD, F., TURNER, N. J., HART, S., TURKENBURG, J. P. & GROGAN, G. 2013. Structure and Activity of NADPH-Dependent Reductase Q1EQE0 from Streptomyces kanamyceticus, which Catalyses the R-Selective Reduction of an Imine Substrate. *ChemBioChem*, 14, 1372-1379.

SAGI, S., AVULA, B., WANG, Y.-H. & KHAN, I. A. 2015. Quantification and characterization of alkaloids from roots of Rauwolfia serpentina using ultra-high performance liquid chromatography-photo diode array-mass spectrometry. *Analytical and Bioanalytical Chemistry*, 408, 177-190.

SCHELLER, P. N., FADEMRECHT, S., HOFELZER, S., PLEISS, J., LEIPOLD, F., TURNER, N. J., NESTL, B. M. & HAUER, B. 2014. Enzyme Toolbox: Novel Enantiocomplementary Imine Reductases. *ChemBioChem*, 15, 2201-2204.

SHEN, Z., EISENREICH, W. & KUTCHAN, T. M. 1998. Bacterial biotransformation of 3α(S)-Strictosidine to the monoterpenoid indole alkaloid vallesiachotamine. *Phytochemistry*, 48, 293-296.

SMITH, G. N. 1968. Strictosidine: a key intermediate in the biogenesis of indole alkaloids. *Chemical Communications* (London), 912-914.

STAVRINIDES, A., TATSIS, E. C., CAPUTI, L., FOUREAU, E., STEVENSON, C. E. M., LAWSON, D. M., COURDAVAULT, V. & O'CONNOR, S. E. 2016. Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity. *Nat Commun*, 7.

STIERAND, K. & RAREY, M. 2010. Drawing the PDB: Protein−Ligand Complexes in Two Dimensions. *ACS Medicinal Chemistry Letters*, 1, 540-545.

STÖCKIGT, J., HUSSON, H., KAN-FAN, C. & ZENK, M. 1977. Cathenamine, a central intermediate in the cell free biosynthesis of ajmalicine and related indole alkaloids. Journal of the Chemical Society, *Chemical Communications*, 164-166.

STROMMER, J. 2011. The plant ADH gene family. *The Plant Journal*, 66, 128-142.

VOGEL, M., LAWSON, M., SIPPL, W., CONRAD, U. & ROOS, W. 2010. Structure and Mechanism of Sanguinarine Reductase, an Enzyme of Alkaloid Detoxification. *Journal of Biological Chemistry*, 285, 18397-18406.

WALSER, A. & DJERASSI, C. 1965. Alkaloid-Studien LII. Die Alkaloide aus Vallesia dichotoma RUIZet PAV. *Helvetica Chimica Acta*, 48, 391-404.

WEISS, D., BAUMERT, A., VOGEL, M. & ROOS, W. 2006. Sanguinarine reductase, a key enzyme of benzophenanthridine detoxification. *Plant, Cell & Environment*, 29, 291-302.

WINN, M. D., MURSHUDOV, G. N. & PAPIZ, M. Z. 2003. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods in enzymology*, 374, 300-321.

Chapter 5

Gene duplication in *C. roseus* has allowed the development of MIA chemical diversity through neofunctionalisation

5.1 Introduction

C. roseus belongs to the Apocynaceae family, which is one of the largest angiosperm families with approximately 5000 species (Straub et al., 2014). It is an economically important family comprising many species with secondary metabolites used as drugs (Aslam et al., 2010, Singh, 2008). Many members of the Apocynaceae family produce MIA, the chemical diversity of which relies on the elaborate enzymatic modification of the branch-point biosynthetic intermediate strictosidine (Rueffer et al., 1978). Deglycosylation of strictosidine by the action of SGD generates a reactive intermediate which can rearrange to generate a plethora of different carbon skeletons.

However, the evolution of SGD poses a trade-off for the strictosidine-producing plant. Strictosidine aglycon can be beneficial for the plant because it can provide some defence against herbivory (Guirimand et al., 2010, Sudžuković et al., 2016, Luijendijk et al., 1996). The reactivity of this compound means that it can cross link amines and nucleotides, making this a highly toxic compound. At the same time however, this compound can potentially endanger the host cell itself.

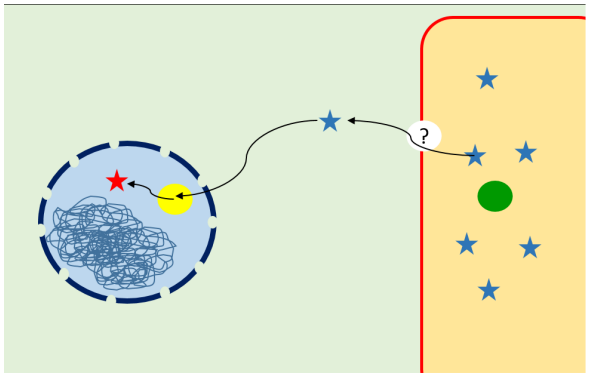


Figure 122: Strictosidine and SGD have a different subcellular localisation. STR (represented by a green circle) is located in the vacuole (orange section of the cell). It produces strictosidine inside the vacuole, represented by blue stars. Strictosidine can be transported out of the vacuole via a vacuolar tranporter (indicated as a white circle). It can diffuse into the nucleus, represented by a blue circle, where it can be deglycosylated by SGD (represented by a yellow circle). The reactive strictosidine aglycon, represented by a red star, can attack the cell constituents and cause damage if not neutralised.

The tight regulation of compartmentation of the strictosidine pool (in the vacuole) and SGD (in the nucleus, fig. 122) circumvents this problem until wounding by herbivore causes the cell to rupture and the two to mix (Guirimand et al., 2010). The strictosidine, which is exported from the vacuole, can find its way to the nucleus and be deglycosylated by SGD. Enzymes that can neutralise this reactive intermediate fast and efficiently could allow the plant to bypass the toxic

effect of background strictosidine aglycon production. It is not unreasonable to suggest that when strictosidine and strictosidine glucosidase had first evolved and started producing strictosidine aglycon, a positive selection pressure started acting on those plants to favour enzymes capable of neutralising the reactive intermediate. After the evolution of the first enzymes able to modify or reduce the reactive intermediate a host of other pathways could evolve which could take the new compounds and build on them, thus expanding the chemodiversity currently found in today's MIA-producing plants.

The previous three chapters of this thesis have detailed the discovery and characterisation of enzymes that can act upon this reactive intermediate. For example, the different heteroyohimbine synthases (HYSs) detailed in Chapters 2 and 3 can take strictosidine aglycon and reduce it to a mixture of heteroyohimbines. The discovery of the related HYSs, and the recent discovery of related MDRs that take part in other parts of the pathway (10HGO (Miettinen et al., 2014), T3R (Qu et al., 2015)), gives us an unprecedented opportunity to study the evolution at this branchpoint (fig. 123).

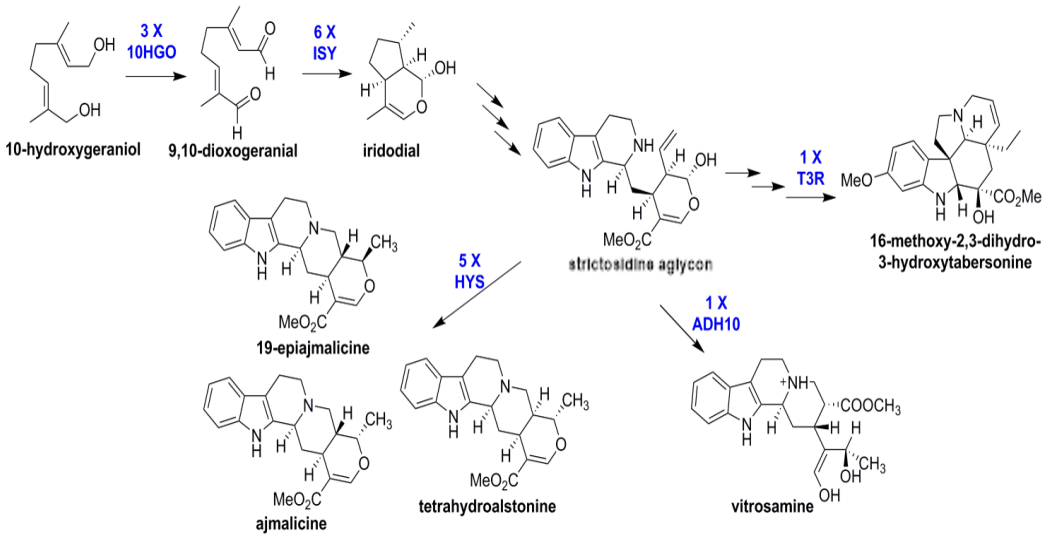


Figure 123: Abundance of reductases in the MIA biosynthetic pathways of *C. roseus*. The reductases implicated in the various reactions and their predicted copy number are indicated in blue.

5.1.1 The birth of new enzymes through gene duplication

Gene families arise as a result of gene duplications, followed by divergence and differentiation of these gene copies (fig. 124). Through mutation, a few of these duplicated genes can eventually give rise to new functions. If the new function imparts a selective advantage, the mutated gene is retained by the organism.

It is interesting to look at lineage-specific duplication events because they could be the manifestation of different evolutionary forces (such as abiotic or biotic stress) and the adaptations to those stresses that the lineage has achieved. A duplicated protein can be retained

in an organism if the duplication somehow allows adaptation. This adaptation may come in the form of recruitment of the newly duplicated protein into a new pathway, to perform a new function, or even to be expressed in a different manner to the original protein. The recurrence of duplications in a specific gene family in a short time could indicate an increased selection pressure acting on, for example, a specific branch of a given metabolic pathway.

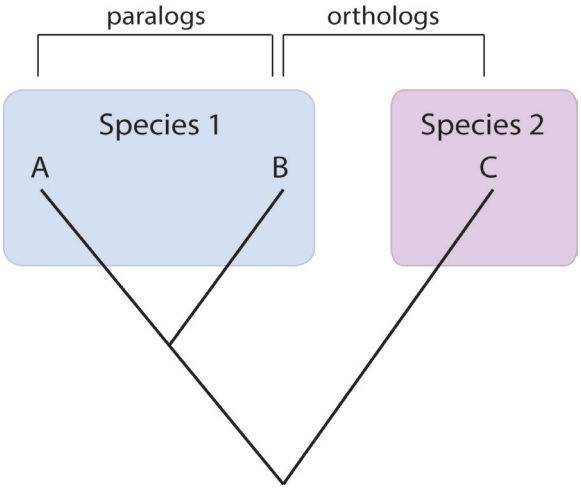


Figure 124: Two proteins are said to be paralogous if they are derived from a duplication event and orthologous if they are derived from a speciation event. Species 1 has two homologous genes, A and B, which are descendent from a duplication event. Species 2 has a gene which is an ortholog to both gene A and B from species 1 because they diverged during a speciation event. Gene C is only considered an ortholog to gene B because gene A has diverged in function since duplication whereas gene B has retained the ancestral function.

It is generally accepted that a gene can undergo mutational drift which can lead to the acquisition of a novel function. However, most theories based on this assumption posit that an enzyme would first evolve to be promiscuous through an expanded substrate recognition and later, if selection pressure is present, would tend towards increased substrate specificity (Weng et al., 2012). However, it is not known if this is always the case.

Different models of neofunctionalisation have been proposed over the years. The “innovation, amplification and divergence” (IAD) model of neofuntionalisation posits that constant selection can act on genes and still lead to the evolution of two separate activities from the same ancestor. An enzyme with a necessary primary function could have a secondary function which is beneficial to the organism. Mutants which have increased activity for the secondary function but retain the necessary levels of activity for the primary function may be difficult to achieve. Therefore, the adaptive function is achieved faster by duplication of the gene. By amplifying the gene dosage, the organism can increase the secondary function dose without hampering the primary function. The new copy is now free from selection for the original function and can undergo mutagenesis which can lead to increased efficiency in the secondary function (Bergthorsson et al., 2007).

A similar, but subtly different, method for neofunctionalisation, known as “escape from adaptive conflict” (EAC) posits that a gene coding for an enzyme with two contradicting functions (which

are mutually exclusive) cannot specialise for either without losing the activity for one of the functions. To circumvent this conflict the gene is duplicated, which allows each copy to undergo selection for one of the functions. However, to identify genes duplicated for EAC is difficult and involves identification of the ancestral state of the genes through phylogenetic reconstruction and extensive knowledge of the function of the orthologs in related organisms (Des Marais and Rausher, 2008).

The literature is replete with examples of duplicated and neofunctionalised genes in plants. Weng et al. (2012) offer a good overview of such cases in many different plants. Such examples include diverse enzyme families and functions such as: the Lysine/Ornithine Decarboxylases (L/OCDs) in quinolizidine alkaloid producing plants could have been derived from common plant OCDs (Bunsupa et al., 2012); terpene synthases in rice implicated in plant herbivore defence and have been subjected to positive Darwinian selection (Chen et al., 2014); the betalain biosynthetic enzymes of the Caryophylales which recently underwent neofunctionalisation after duplication (Brockington et al., 2015); and the latex-producing enzymes in the rubber tree *Hevea brasiliensis* which were found to be in tandem-duplication clusters (Lau et al., 2016).

5.1.2 Gene duplication and heteroyohimbine biosynthesis

The HYSs are most similar to the Cinnamyl Alchohol Dehydrogenases and the closely related Sinapyl Alcohol Dehydrogenases (CAD and SAD, respectively). These MDR enzymes evolved in plants (Riveros-Rosas et al., 2003) and are implicated in lignin biosynthesis. Studies of these enzymes in other plant species have revealed they are often duplicated. There are 8 CAD-like paralogs in *Arabidopsis thaliana*, some of which are present in duplicated blocks (Kim et al., 2004). The genes identified had varying amounts of identity to each other (from 45.9 to 98.4%) which indicates some copies are more ancient than others and have diverged considerably. An array of three tandem duplicated AtCAD genes was discovered on the *A. thaliana* chromosome 4 and a phylogenetic tree indicates that these 3 paralogs are the most closely related, indicating that they are the product of a recent local duplication event. In *Populus trichocarpa* 15 CAD paralogs were identified and their chromosomal location was identified (Barakat et al., 2009). Generally, it was found that many of the CAD duplications had occurred locally but whole clusters of duplicated CAD genes are found duplicated in other parts of the plants' genome and these in theory have arisen from whole genome duplications and rearrangements (Barakat et al., 2009). These CAD duplicates have been investigated in detail in *A. thaliana* and although they all displayed CAD-activity, the K_m, k_{cat}, and their relative expression in the tested tissues all varied considerably (Kim et al., 2004). It is not known if these duplicates have activity against other, non-lignin-related substrates *in planta*.

5.1.3 Overview of this chapter

The *C. roseus* transcriptome contains multiple MDR reductases expressed in different tissues and at different levels. In this chapter a study of the evolution of these reductases, and more specifically of the HYSs, reveals the interesting dynamics present at these loci. The duplication and divergence of these enzymes has profound implications for the chemical diversity found in the MIA and other secondary metabolic pathways. Heteroyohimbine synthesis appears to be possible with different MDR active sites, as evidenced by the discovery of multiple HYSs in *C. roseus* which do not have identical active sites, and in fact are quite variable (detailed in Chapter 2 and 3). This information indicates that although the HYS enzymes have probably originated from a common ancestor, there has not been any selection pressure to maintain specific amino acids in the active site or the active site loops. However, for reasons that are not well understood, there appears to be a positive selection pressure to maintain multiple copies, as evidenced by the multiple MDRs which display HYS activity. This flexibility in active site architecture allows the genes to drift over time which consequently has facilitated the appearance of genes that further elaborate on the MIA scaffold. For example, it is possible that the extensive gene duplication of the HYS has led to the emergence of T3R, a late stage MIA biosynthetic enzyme, which has lost the ability to catalyse the production of heteroyohimbines but has gained the ability to catalyse the reduction of a downstream MIA.

5.2 Results

5.2.1 Genomic organisation of the *C. roseus* ADHs

The recent sequencing of the *C. roseus* genome (Kellner et al., 2015) has allowed an investigation into the duplication of the ADHs to be carried out in this plant. Firstly, the genomic context of the *C. roseus* ADHs was investigated. Each gene annotated as an ADH was submitted to BLAST against the *C. roseus* WG assembly contigs (WGC, version 2.A, Buell and Kim, in preparation). A close inspection of the results reveals that there are 68 contigs containing ADHs (table 14). Of these, 28 have more than one ADH. The WGC numbers 12, 24, 27, 106, 126, 286, 371, and 420 each contain an ADH which has been shown experimentally to catalyse a reduction or oxidation in the MIA pathway of *C. roseus*. Interestingly, on the WGC which contain ADHs of different families these ADHs are usually separated by many other genes and are never in direct proximity. However, it is common in the WGC containing multiples of a single ADH family, such as the MDR family, that these genes are in tandem with each other, suggesting tandem duplication was responsible for the origin of the arrangement.

Table 14: *Catharanthus roseus* Whole Genome contigs containing an ADH gene.

WG contig	Approximate size (bp)¥	Number of ADHs	Number of MDR	Number of SDR	Number of other ADH
3	2453204	4	4/4 a		
6	2384693	3	1/3	1/3	1/3
7	2383327	3		3/3 a	

WG contig	Approximate size (bp)¥	Number of ADHs	Number of MDR	Number of SDR	Number of other ADH
8	2247593	2		2/2	
11	2076047	4	3/4 a	1/4	
12	***2081351***	***1***	***1/1***		
13	2021668	2		2/2	
24	***1697648***	***4***		***4/4 a***	
27	***1651656***	***6***		***6/6 a***	
34	1538188	1		1/1	
39	1408578	2	2/2 a		
44	1360256	1		1/1	
45	1340673	1		1/1	
50	1276821	2			2/2 a
51	1259040	4			4/4 a
54	1274735	1			1/1
56	1206958	3	1/3	2/3 a	
62	1164002	3			3/3 a
65	1177834	1		1/1	
67	1180532	1		1/1	
78	1117415	4	1/4	1/4	2/4 a
82	1104218	3	2/3	1/3	
87	1031804	6	6/6 a		
94	1028141	1	1/1		
101	967572	1		1/1	
106	***953033***	***5****	***4*/5* a***	***1/5****	
107	954116	1		1/1	
126	***910264***	***10***	***10/10 a***		
127	899872	5		5/5 a	
136	855973	2		2/2	
148	823799	1	1/1		
155	826119	1			1/1
171	794550	1		1/1	
172	799169	1		1/1	
189	753374	1		1/1	
191	750583	1	1/1		
210	721081	1			1/1
213	712053	2	1/2	1/2	
243	649553	1			1/1
267	557226	2		2/2	
280	591668	1			1/1
286	***560763***	***4***	***1/4***	***3/4 a***	
298	547598	1	1/1		

WG contig	Approximate size (bp)¥	Number of ADHs	Number of MDR	Number of SDR	Number of other ADH
371	***455742***	***2***	***2/2***		
387	458743	2	1/2		1/2
406	430839	1			1/1
420	***420531***	***1***			***1/1***
432	406618	1		1/1	
504	321640	1		1/1	
534	322847	2		2/2 a	
569	289805	5	3/5 a		2/5 a
584	291195	1		1/1	
602	276623	1		1/1	
603	258472	1	1/1		
617	272626	1			1/1
647	175517	1		1/1	
651	257338	2	2/2 a		
666	239967	1		1/1	
755	192998	1	1/1		
849	144742	1		1/1	
935	100892	1	1/1		
976	106701	1		1/1	
1094	77432	1	1/1		
1157	57567	1	1/1		
1309	9014	1	1/1		
1312	39141	1		1/1	
3258	2785	1	1/1		
4337	2026	1	1/1		

The table is based on the WG assembly number 2 version A. The number of ADHs present is indicated as well as the number of MDRs or SDRs. The contigs containing an ADH proven to be participating in the whole MIA pathway are in bold and italic font. The group 'other ADH' comprises examples from the Aldo-keto reductases (AKR), and the Aldehyde dehydrogenases (ALDH). Contigs in red span the entire length of the ADH gene on them, therefore it is not possible, with the current alignment, to know if there are other genes next to that ADH.

¥ the size is based on the location of the last gene aligned onto that contig

*This contig contains the HYS duplication and is not of good enough quality to determine if there are 4 or more copies of HYS.

a Denotes that two or more of the ADHs are in tandem with at most one other gene between them

ADH10, the SDR described in Chapter 4, is present on the WGC number 27 along with 5 other SDRs. The HYSs, described in Chapters 2 and 3, are distributed on different contigs; THAS1 and THAS3 are on WGC 371; THAS2 is on WGC 126, which contains 9 other MDRs, one of which

(Cr023176) has been shown to be a 10HGO (Krithika et al., 2015); HYS and THAS4 appear to be localised in a tandem duplication together with 3 other similar MDRs not found in the transcriptome (Góngora-Castillo et al., 2012) on the WGC 106.

5.2.2 Analysis of the ortholog groups present in the MDRs and SDRs of *C. roseus*

The MDRs and SDRs present in *C. roseus* could be descended from different families of MDR and SDR respectively. To determine what the distribution of these genes is in *C. roseus*, all the MDR and SDR sequences were submitted to BLAST against the Clusters of Orthologous Groups (COG, (Tatusov et al., 1997) database using the NCBI batch BLAST server (www.ncbi.nlm.nih.gov/Structure/bwrpsb/, table 15).

All *C. roseus* transcripts annotated as MDRs were identified as belonging to the Zinc-dependent ADHs. Twenty-nine of the MDRs were identified as being part of the COG1064 group which is classified based on D-arabinose 1-dehydrogenase. COG1064 also contains the CAD/SAD enzymes. Eight MDRs were best identified as belonging to the COG1062 (formaldehyde dehydrogenase and related Zn-dependent dehydrogenase) group, and one to the COG1063 (threonine dehydrogenase and related Zn-dependent dehydrogenases) group. Many of the remainders are part of the COG0604 groups which contain the quinone reductases (QOR), and two homologs are part of the COG2130 group which is represented by a curcumin reductase (CurA). One member is COG0647, which constitutes the ribonucleotide monophosphatase NagD and the haloacid dehydrogenase (HAD) superfamily.

Table 15: COG of MDRs from *C. roseus*.

Gene locus	COG	e-value	Gene locus	COG	e-value
THAS3	1064	7.74 e-88	Cro022431	1064	1.18e-92
THAS2	1064	7.63 e-95	Cro008301	1064	6.30e-98
THAS4	1064	1.76 e-92	Cro001761	1064	1.79e-112
HYS	1064	1.07 e-89	Cro024150	0647	1.30e-27
Cro017994	1064	5.17 e-97	Cr2141	1064	3.80e-132
Cro011702	1064	2.18 e-102	Cro023176	1064	2.59e-117
Cro030442	1064	1.26 e-119	Cro024340	0604	6.30e-66
Cro006840	1064	1.57 e-118	Cro016395	1062	2.09e-144
Cro022770	1064	3.86 e-96	Cro029195	2130	2.43e-122
Cro033537	1064	2.57 e-93	Cro013040	2130	7.05e-122
Cro033062	1064	5.89 e-80	Cro017213	1064	8.86e-92
Cro011226	1064	1.09e-114	Cro019170	1062	2.36e-149
Cro033830	1062	2.48e-126	Cro020524	0604	1.06e-73
Cro027234	1064	3.99e-102	Cro020525	0604	3.14e-73
Cro005375	1062	1.34e-116	Cro015403	1062	1.42e-156
Cro027079	1064	1.02e-98	Cro017625	1063	6.33e-96
Cro025489	1064	7.83e-104	Cro018598	0604	2.79e-64

Gene locus	COG	e-value	Gene locus	COG	e-value
Cro019494	1064	4.93e-116	Cro019716	1062	8.88e-158
T3R	1064	1.10e-106	Cro015629	1062	8.50e-152
Cro026235	1064	3.42e-102	Cro018442	0604	3.34e-100
Cro027235	1064	1.19e-109	THAS1	1064	5.23e-98
Cro016729	1064	4.13e-74			

The SDRs generally had a lower e-value to the COG domains in the NCBI database (table 16). This is probably reflective of their high sequence divergence (typically between 15-30 % pairwise identity (Jörnvall et al., 1999)). Most of the SDRs (31 out of 54) corresponded best to the COG1028 which corresponds to the group “Dehydrogenases with different specificities” of the SDRs. ADH10 (Cr013448) is a member of this group. Another highly represented group (16 homologs) is the COG0451, which is classified as “Nucleoside-diphosphate-sugar epimerases”, and which contains the progesterone-5β-reductases. The early MIA biosynthetic enzyme ISY and its homologs belong to this COG. The COG0300, corresponding to a general SDR prediction of unknown function, has 3 homologs in the *C. roseus* genome.

Table 16: COG of SDRs from *C. roseus*.

Gene locus	COG	e-value	Gene locus	COG	e-value
Cr013448	1028	1.64E-40	Cr028099	1028	1.75E-62
Cr033739	1028	1.43E-30	Cr027571	0300	5.06E-55
Cr022864	1028	1.04E-33	Cr027322	1087	1.37E-156
Cr013184	1028	7.86E-63	Cr001335	0300	7.11E-52
Cr013447	1028	1.60E-32	Cr016749	1091	9.21E-48
Cr027095	0451	3.93E-40	Cr016747	1028	1.19E-33
Cr003140	1028	9.85E-39	Cr022212	1028	1.50E-52
Cr028100	1028	2.13E-62	Cr010887	1028	4.71E-61
Cr017503	0451	8.89E-09	Cr001750	1028	1.75E-57
Cr017502	0451	8.13E-11	Cr017031	1028	2.44E-26
Cr030915	0451	1.15E-33	Cr018552	1028	6.19E-60
Cr023278	1028	1.66E-35	Cr022484	0451	2.83E-39
Cr023367	1028	9.77E-49	Cr002309	1028	3.51E-27
Cr010996	1028	1.14E-75	Cr018601	0451	6.67E-35
Cr014890	1028	1.37E-68	Cr008727	0451	5.31E-36
Cr023217	1028	1.63E-47	Cr001235	1028	1.35E-63
Cr006167	0451	1.72E-38	Cr019769	0451	8.34E-45
Cr022002	1028	4.89E-33	AIW09146.1	0451	1.24E-08
Cr001031	0451	1.55E-13	Cr025915	1028	5.97E-33
CrISY	0451	1.82E-06	Cr003619	0451	3.39E-43
Cr033366	1028	7.00E-34	AIW09148.1	0451	1.25E-07
Cr033093	1028	1.94E-55	Cr011094	0300	1.43E-51
Cr000253	1087	1.02E-144	Cr023179	1090	2.25E-107

Gene locus	COG	e-value	Gene locus	COG	e-value
Cr019499	1028	4.48E-30	Cr008631	1028	1.82E-28
Cr028501	1028	1.33E-78	Cr004988	1028	8.54E-58
Cr028094	1028	5.63E-60	Cr002470	0451	1.61E-74
Cr028096	1028	3.84E-66	Cr011896	0451	4.60E-39

These results suggest that in both enzyme families there is high divergence which is consistent with findings from other published surveys of these families in genomes (Jörnvall et al., 1999, Moummou et al., 2012).

5.2.3 Phylogeny of the *C. roseus* MDRs

Studies of large gene families, such as that of the MDRs, are facilitated by phylogenetic analysis. This method of analysis can identify distantly related paralogs and can illustrate the evolutionary relationship between two enzymes which might perform different functions or exhibit a similar function but with a different spatio-temporal distribution. A phylogenetic analysis of the MDRs contained in the *C. roseus* transcriptome and genome was undertaken in order to better understand the dynamics of the duplications and neofunctionalisation. Alignment using fsa resulted in 99.5 % of sites with gaps, compared to mafft which produced 75.6 % and muscle which produced 78.4 %. PRANK, although it went through multiple iterations comparing the alignments to a parsimony tree it had generated, had only managed 84.2 % sites with gaps. ClustalOmega aligned the sequences with 77.4 % gaps (fig. 125, and appendix 5) and was overall similar to the muscle alignment, but tended to avoid opening gaps to account for proteins with very divergent sequences.

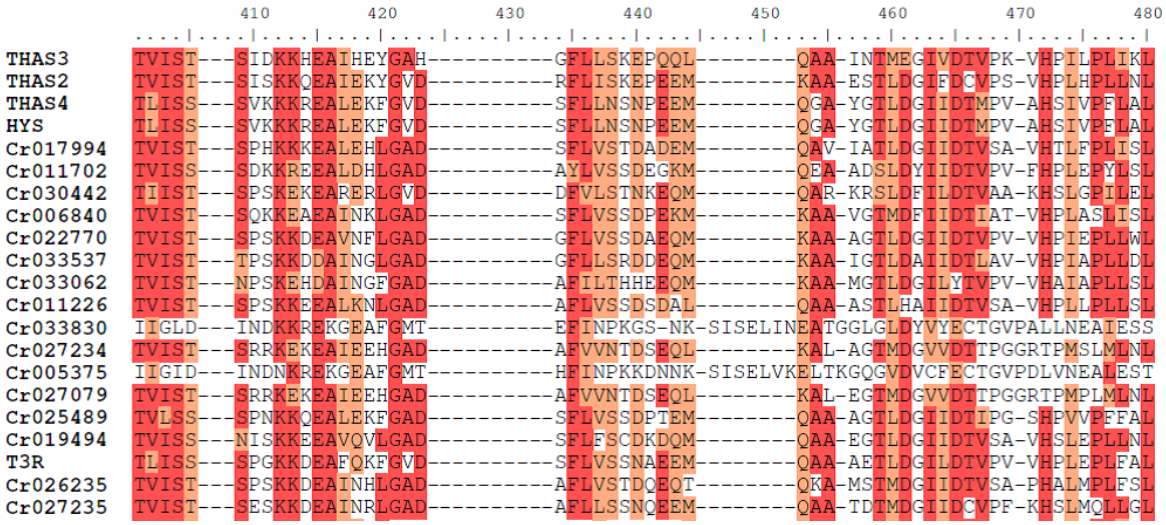


Figure 125: Example section of MDR ClustalOmega alignment. For full alignment please look at Appendix 5. The identical sites are shaded red and the similar sites are shaded orange. The shading threshold used is 50 %. Gaps in alignment are illustrated by a '-'.

The MDR protein sequences, aligned with ClustalOmega (Sievers et al., 2011), were analysed by Bayesian inference using the program MrBayes (Ronquist et al., 2012). Two independent runs of

Multiple-chain Markov chain Monte Carlo (MC3) were executed to test all amino acid models. The first run converged to model 5 (WAG) quite fast, which is similar to the result of Maximum Likelihood analyses. The exchanges between chains were low and therefore chains were heated by 0.125 (default is 0.10). The result was a better sampling across the model space; the average effective sample size was above 2000 for the heated runs compared to just above 900 for the default runs (see appendix 1 fig. 2 for convergence graphs). The gamma shape parameter, which allows the alignment sites to evolve at different rates for each other, was tested as a fixed state across all sites (mixed model), or allowed to vary across the alignment (mixed model + gamma). Stepping-stone simulation with 1 000 000 iterations was also carried out in MrBayes to verify the likelihood of the tree produced by Bayesian inference.

A phylogenetic tree was produced from the alignment generated above using a Maximum Likelihood approach as well. This was done primarily to verify the reproducibility of the topology and to gauge the branch support through a different method. The PhyML online server (http://www.atgc-montpellier.fr/phyml/)(Guindon et al., 2010) was used for this analysis, with the automatic model selection enabled, and the gamma shape parameter was also allowed to vary, as in the Bayesian inference described above. The resulting tree length was estimated to be 26.5371 which is almost identical to the tree size estimated with Bayesian inference (26.2916). The likelihood for the produced tree was similar to the estimation with Bayesian inference and the stepping stone sampling (-23866.80, -23967.83, and -26194.55 respectively). The estimation of the gamma shape parameter (2.280) was very similar to that estimated with Bayesian inference as well (2.211).

The resulting phylogenetic trees were in general all very similar (fig. 126, 127). The most striking feature of the trees is that two groups of MDRs appear to be evolving at different rates. The first group, characterised by long branches and few duplications, contains 17 MDRs. The length of the branches reflects the divergence these enzymes have undergone since the last common ancestor. Conversely, the other group, comprising 27 MDRs (including the HYSs), is characterised by short branches and repeated branching. This is reflective of numerous recent duplications.

The main differences between the phylogenetic tree constructed with Bayesian inference or Maximum Likelihood (PhyML) were in the relationship between the two 10HGO enzymes (Cr033830 and Cr005375, orange clade in fig. 126-128). Both of these relationships are well supported by relatively high posterior probabilities. The section of the trees encompassing the HYSs (fig. 126-128) is well supported, but differs slightly in the origin of Cr027234 and Cr027079 (green clade fig. 126-128). In the tree produced with PhyML, these enzymes appear to have evolved after the separation of THAS2 and THAS3 (fig. 126, blue and green clades), but this node is not supported by high Bootstrap support (40). However, in the Bayesian inference analysis the three branches are present as a trifurcation. The posterior probability is still low (53), but higher than the node Bootstrap support in the PhyML tree, and therefore the scenario presented in the MrBayes tree (fig. 126) is more conservative. This indicates there is not enough information in the data to resolve the relationships at this node. As Cr027234 has been shown by collaborators to

also act on strictosidine aglycon, but to yield a different product (unpublished), there is interest in resolving these relationships. However, given that half of the enzymes in that clade (fig. 126, blue clade) have been shown to be HYSs, it is reasonable to suggest that the progenitor enzyme was a HYS, and that the function of Cr027234 is novel and specific to that branch. Similarly, T3R (fig. 126 purple clade), the MDR shown to catalyse the reduction of a downstream MIA in *C. roseus* (Qu et al., 2015), is also included in the HYS clade (fig. 126, green clade). This strongly suggests that T3R has evolved from a duplication of an ancestral HYS copy which also gave rise to the THAS1-THAS4-HYS clade.

The *C. roseus* MDRs which show homology to the *Rauvolfia serpentina* Vomilenine Reductase 2 (RsVR2, Geissler et al. (2015)) all cluster together inside a sister clade to the HYS clade (fig. 126 and 127 in red). RsVR2 catalyses the reduction of vomilenine, a downstream MIA also produced from strictosidine in *R. serpentina* (fig. 14, Chapter 1). It is not known if any of these homologs (Cr022770, Cr008301, Cr017213, and Cr006068) are able to catalyse the same reduction as RsVR2 but that clade could represent another group of MDRs capable of reducing MIA substrates.

Since all the HYSs discovered in *C. roseus* are grouped together, it can be concluded that these genes likely arose as a duplication of an ancestral gene which was probably a HYS. Cr023176, a member of a sister clade (Cr019494, Cr016729, and Cr023176) to the HYS clade has been shown to be a 10HGO (Krithika et al., 2015). This suggests the ancestor to this and the HYS clade could have been either a HYS or a 10HGO. However, it is not possible to determine what activity the ancestral enzyme possessed without ancestral reconstruction or extensive sampling of enzyme activity of that clade in both *C. roseus* and related species. The ancestral enzyme could also have displayed a CAD/SAD activity; Cr2141, present on another sister clade to the HYS clade, has previously been shown in our group to be an SAD enzyme (Yerkes, 2010). Although many of the MDRs included in this phylogenetic tree have not been experimentally assayed, the sampling of candidates described in Chapter 2 suggests HYS activity is found primarily in this clade (fig. 126, blue clade).

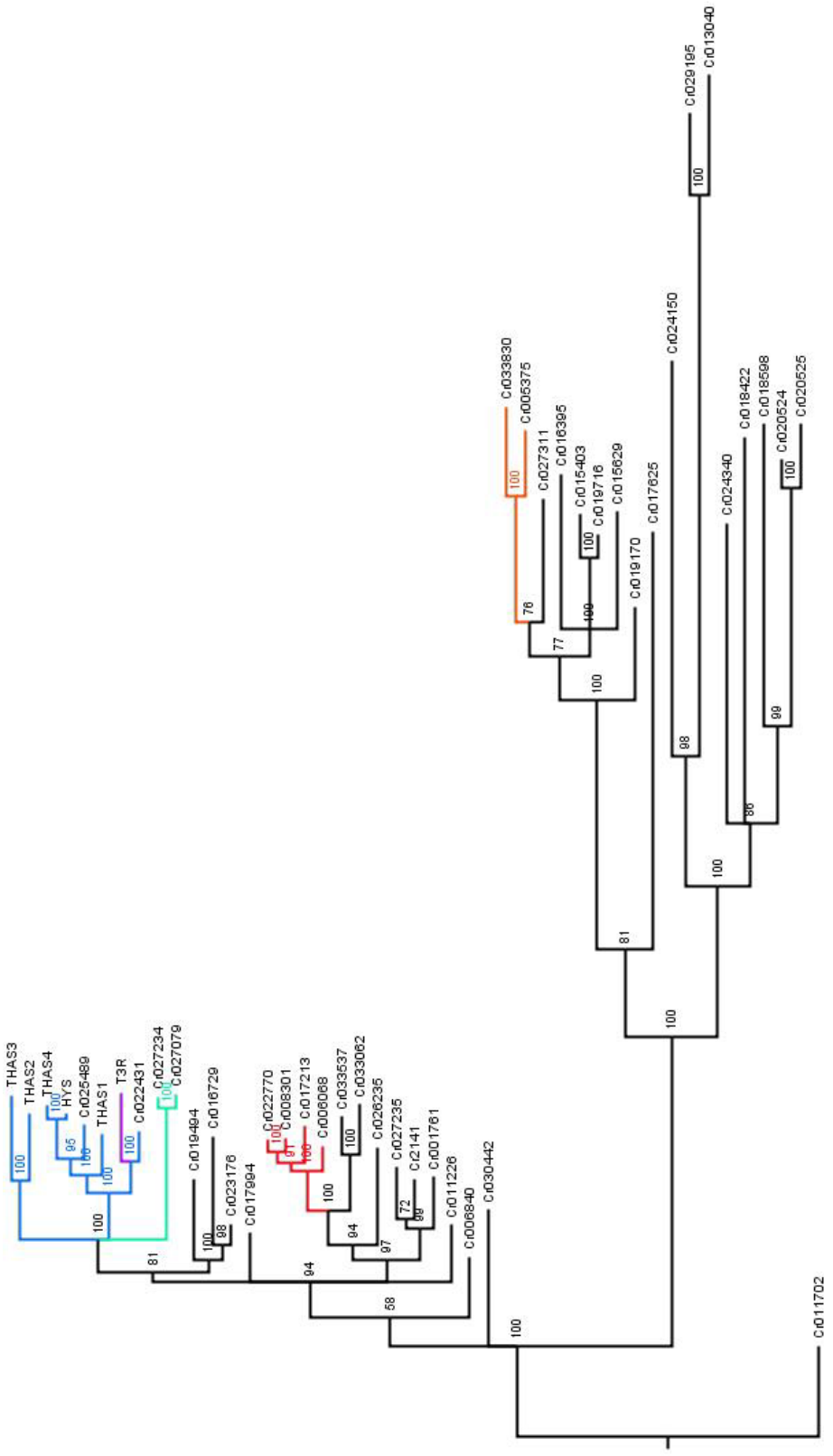


Figure 126: Bayesian phylogenetic tree of MDRs. This tree is based on a mixed amino acid model + gamma. The posterior probability of each node is indicated next to each node. The HYS clade is coloured in blue; T3R is purple; the Cr027234 clade in green; the VR2-homologous clade is illustrated in red; and the 10HGO clade is in orange; the bar indicates the changes per site.

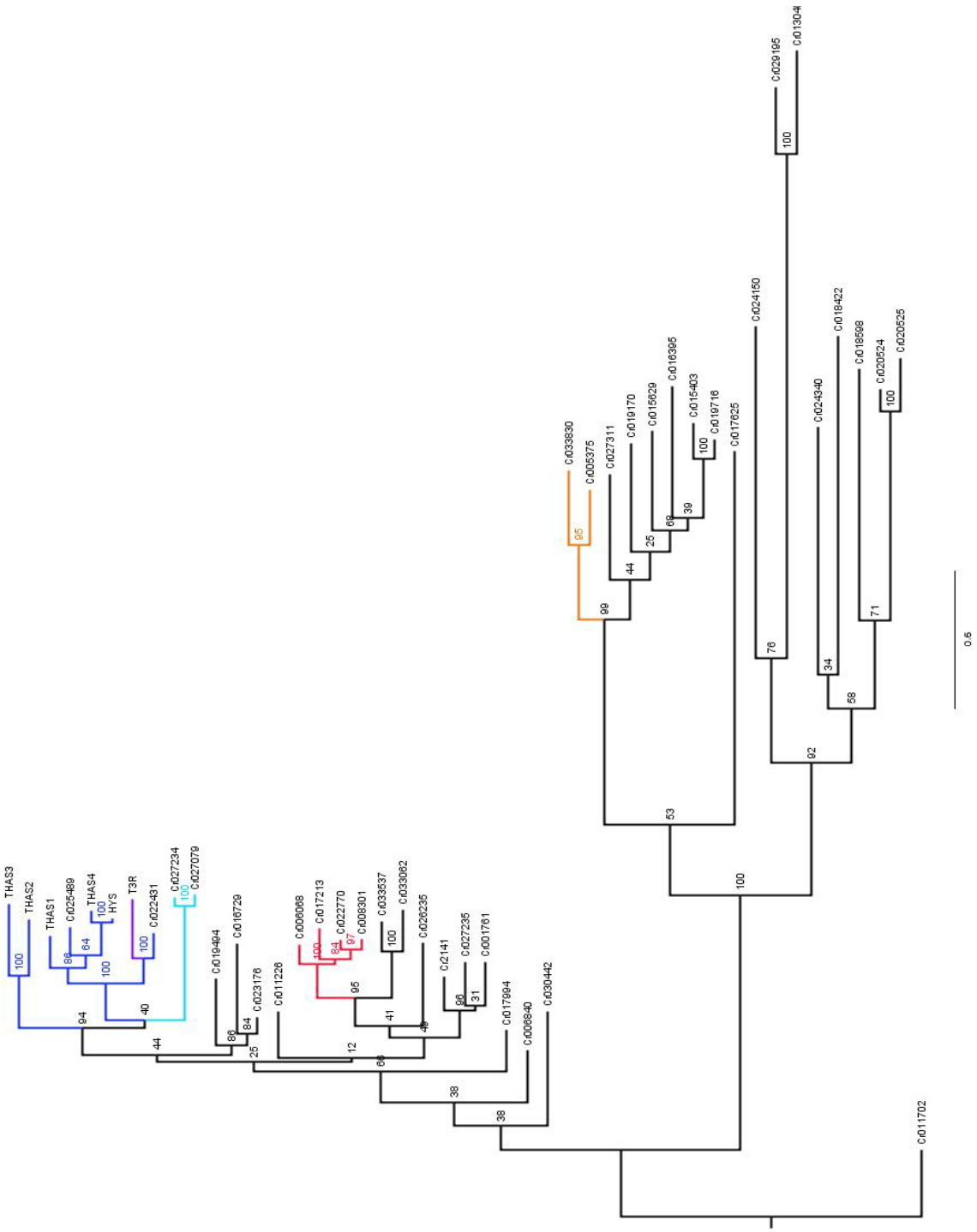


Figure 127: Maximum Likelihood phylogenetic tree of MDRs. Tree produced through online server of PhyML. Bootstrap support values are printed next to each node. The HYS clade is coloured in blue; T3R is purple; the Cr027234 clade in green; the RsVR2-homologous clade is illustrated in red; and the 10HGO clade is in orange; the bar indicates the changes per site.

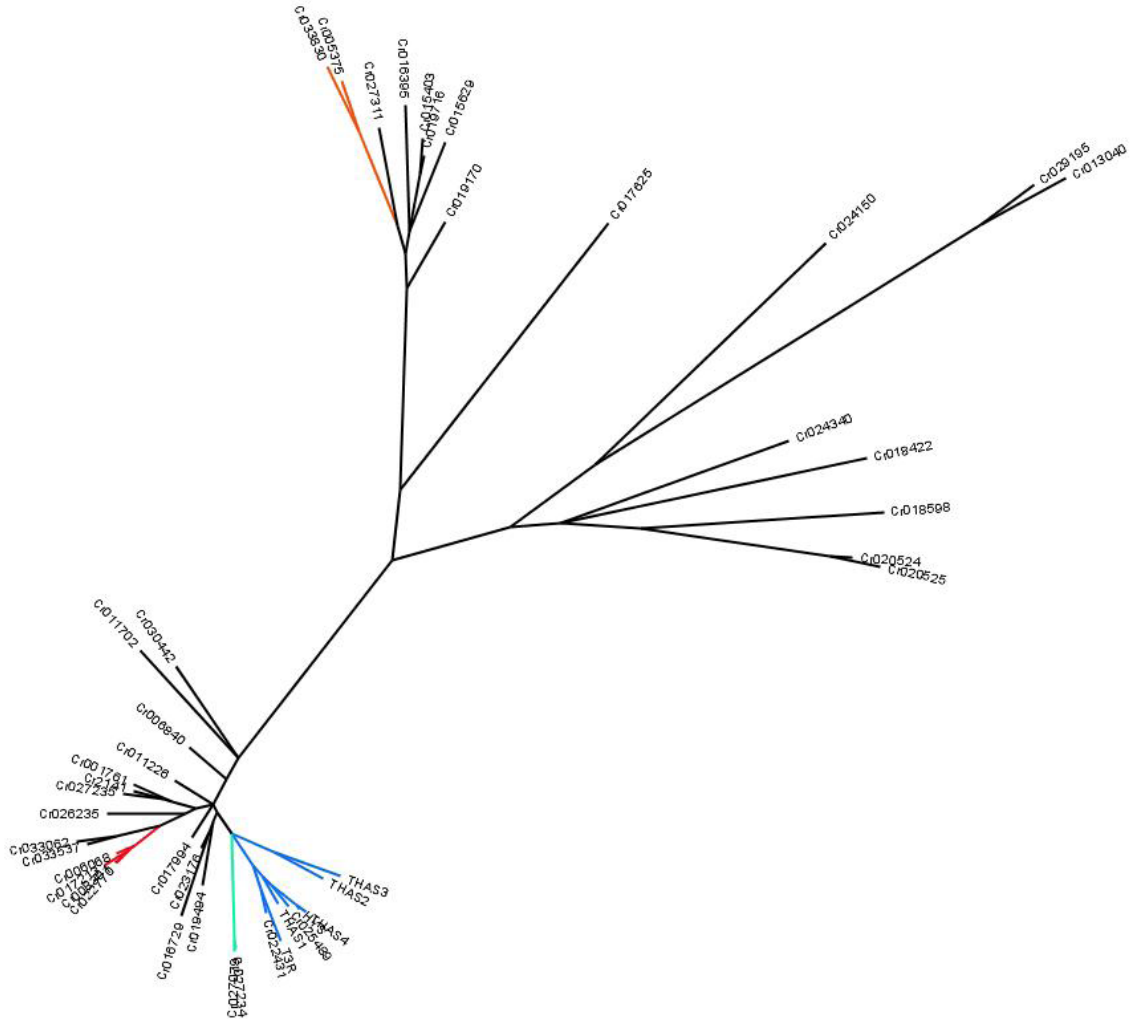


Figure 128: Phylogenetic tree of MDRs in *C. roseus* in radiating representation. The MDR clades are coloured as in fig. 126 and 127: HYS clade: blue; T3R: purple; Cr027234 clade: green; RsVR2-homologous clade: red; and 10HGO clade: orange; the bar indicates 0.5 changes per site. This is the same tree as the one displayed in fig. 126.

The radiating representation puts into perspective how recently some of these MDR duplications occurred, particularly for the genes in the bottom part of the tree. It is apparent from this tree that a group of MDRs has been duplicated throughout the evolutionary history of *C. roseus* and gave rise to the enzymes (HYSs, T3R) capable of catalysing a reduction of a MIA. Two of the 10HGO (Cr033830 and Cr005375, orange clade in fig. 126-128) on the other hand have not arisen from recent radial expansion but rather appear to be offshoots of a smaller clade of MDRs. It is unknown if the *C. roseus* RsVR2-homologous enzymes (red clade in fig. 126-128) are capable of catalysing the same reduction as RsVR2, but the substrate (vomilenine) is not found in *C. roseus*.

5.2.4 Phylogenetic analysis of the *C. roseus* SDRs

A phylogenetic analysis of the SDR homologs in *C. roseus* was undertaken to better understand the evolution of this family in this plant. This family includes the critical enzyme of iridoid biosynthesis, ISY (Geu-Flores et al., 2012), and also ADH10, which, as described in Chapter 4, can reduce

strictosidine aglycone to produce a newly discovered molecule similar to vallesiachotamine. Interestingly, ADH10 is present in a large WGC, number 27 (table 14), which contains at least 6 SDRs. An investigation into the phylogeny of this group and into the duplication abundance was undertaken to compare to the duplication history of the MDRs, and more specifically the HYSs, with the SDRs.

Alignment of the protein sequences was done using ClustalOmega. The high sequence divergence between the proteins was evident in the alignment as many sections (particularly the C-terminal region) were highly different and alignment in those sections presented mostly indels. This is to be expected because, as discussed in Chapter 4, SDRs can vary considerably in the length and sequence of the C-terminal domain. Indeed, the minimum sequence identity at sites without an indel was just 5.81 % and the largest gap had a length of 130 amino acids (full alignment can be found in Appendix 6). Maximum likelihood was used to infer phylogenetic trees through PhyML. Bayesian inference was not attempted due to computational limitation.

PhyML using the WAG model for amino acid substitution with the site rate allowed to vary (4 gamma states) resulted in a tree of larger length than that of the MDRs (50.67 and 26.53 respectively). The gamma shape parameter was found to have a value of 2.672, which is similar to that found for MDRs. The analysis using the PhyML online server gave very similar results; tree length was 50.70 and the gamma shape parameter was determined to be 2.627. The preferred model in this case was the WAG + gamma + F (empirical equilibrium frequency). The log-likelihood of this tree was found to be - 39976.83 which is very similar to the log-likelihood of the tree determined using the command-line PhyML (-39977.67). The tree built using the online PhyML server is shown below (fig. 129) but is in general very similar to that obtained through the command line PhyML (Appendix fig. 3).

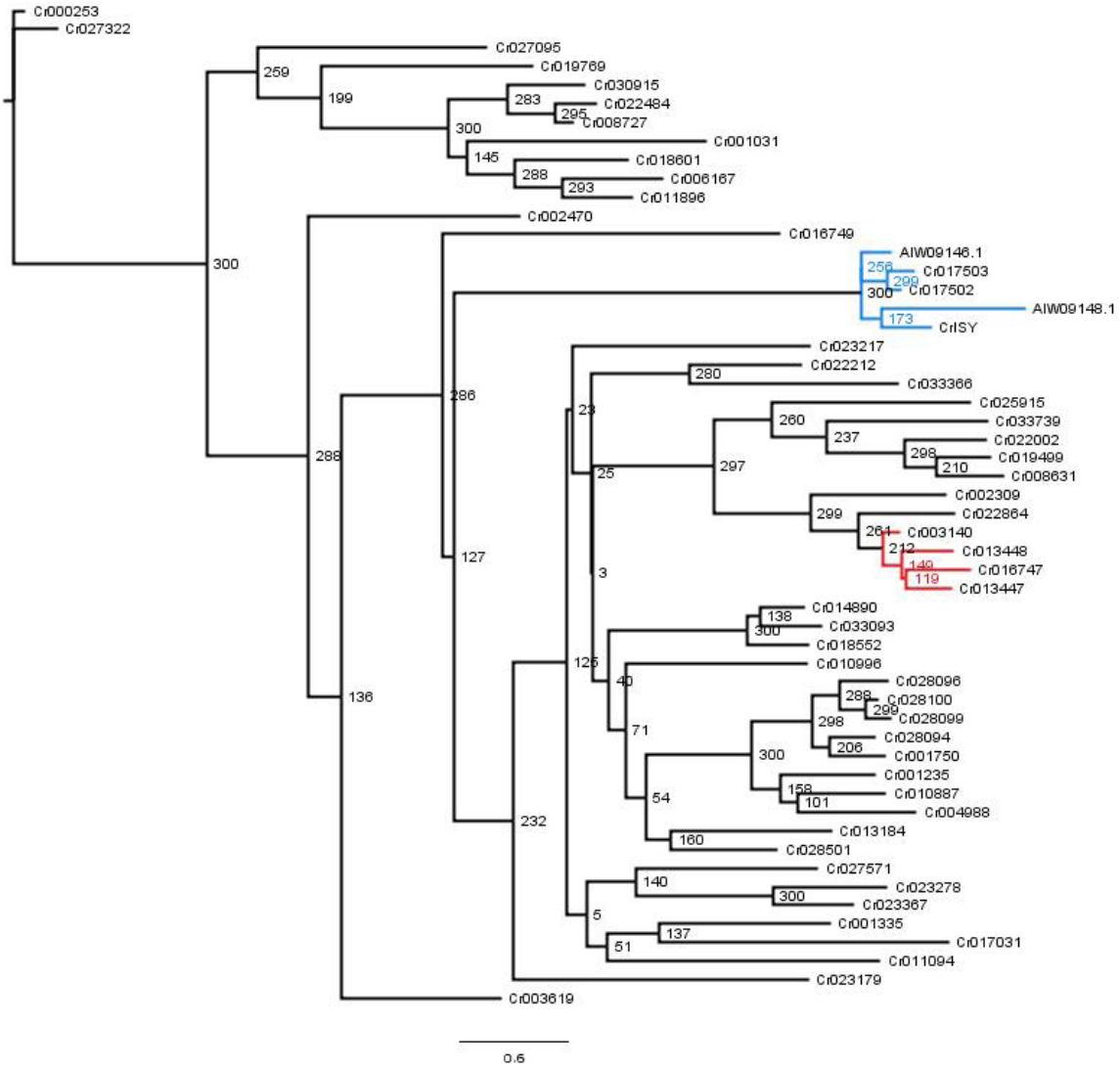


Figure 129: Phylogenetic tree of the SDR superfamily in *C. roseus*. Tree was constructed using PhyML; Bootstrap support (number of times the node was returned out of 300 iterations) is printed next to the nodes; ADH10-like clade is illustrated in red; ISY-like clade is in blue. ADH10: Cr013448. The bar at the bottom represents the distance to 0.6 changes per site.

ADH10 is part of a small clade (fig. 129, red) which is separated from the ISY clade (fig. 129, blue). This supports the COG analysis that determined that these two enzymes were members of different SDR groups. The tree suggests that ADH10 is part of a block of relatively recently duplicated SDRs. It is not known what function the other SDRs in the ADH10 clade carry out. ISY is also part of a recent duplication and the other members of that clade (fig. 129, blue) have been shown to also display ISY activity (Munkert et al., 2015). A better sampling of the SDRs in *C. roseus* would allow a more detailed understanding of the SDR phylogenetics, and how this relates to secondary metabolism.

An in-depth characterisation of the SDRs of plants found that some subfamilies of the SDRs had evolved after the split of vascular plants from Bryophytes (Moummou et al., 2012). Interestingly, PsSalR, the enzyme used for molecular replacement when solving the crystal structure of

ADH10 (Chapter 4), is a member of one of these groups, and therefore ADH10 is as well. These subfamilies had expanded through duplications in the vascular plant genomes and have given rise to many enzymes involved in secondary metabolism (Moummou et al., 2012). These data allowed the authors to suggest that the SDRs in plants were used as an adaptative character during land colonisation and development of the vascular apparatus.

5.2.5 Comparison of *C. roseus* duplication blocks to other plant genomes

The discovery that CAD and SAD enzymes are found in duplicated blocks in *Populus* (Barakat et al., 2009), rice (Tobias and Chow, 2005), and *Arabidopsis* (Kim et al., 2004) suggests that these enzymes are often duplicated and the copies are retained in the genome. The largest cluster of duplicated MDRs in *C. roseus* (WGC number 126) contains 10 MDRs (fig. 130). To test whether this block is similar to the CAD/SAD duplication blocks observed in other species, a phylogenetic alignment between the gene copies found in the WGC126 cluster and MDR copies in other plant genome clusters was attempted.

The ensembl Plant Genome Database was used for this analysis as it hosts many of the assembled plant genomes available. This database also has the advantage that it has pre-aligned syntenic blocks of genomes. Unfortunately, it is not currently possible on this database to conduct syntenic alignments of a query sequence.

Figure 130: MDR cluster arrangement on WGC 126. The arrows point in the direction of the gene ORF; the cluster is not drawn to scale.

The WGC 126 cluster member MDR Cr2141 was chosen as bait for this phylogenetic analysis because this gene robustly and repeatedly aligned onto all of the plant genomes queried. Cr2141 was submitted to BLAST against the *A. thaliana* genome assembly on the ensembl Plant Genomes Database using the TBLASTN tool. *A. thaliana* is the only plant genome in this database onto which all others have been syntenically aligned. A close inspection of *A. thaliana* chromosome 4 (region 17840497-17867094) revealed that there are 3 highly similar genes to *C. roseus* Cr2141 (AtCAD6, ELI3-1, and ELI3-2 in that order) (fig. 131). These three genes have been the subject of an investigation, and have been shown to have CAD/SAD activity (Kim et al., 2004).

Table 17: Plant species used in cluster duplication analysis

Lycopodiophyta	Bryophyta	Chlorophyta
Selaginella moellendorffii	*Physcomitrella patens*	*Chlamydomonas reinhardtii*

Asterids	Rosids	Monocots
Catharanthus roseus	*Vitis vinifera*	*Oryza sativa Japonica*
Solanum lycopersicum	*Theobroma cacao*	*Setaria italica*
Solanum tuberosum	*Prunus persica*	*Hordeum vulgare*
	Populus trichocarpa	*Musa acuminata*
	Medicago truncatula	*Brachypodium distachyon*
	Glycine max	*Sorghum bicolor*
	Brassica rapa	
	Brassica oleracea	
	Arabidopsis thaliana	

This section of the *A. thaliana* genome was then aligned to syntenic regions on the other plant genomes on the ensembl Plant Genomes Database (detailed in table 17). Not all plants were used due to redundancy; for example, *Brassica napus* and *A. lyrata* were not used because three other Brassicaceae were already used. out of a total of a total of 43 species with genomes hosted on ensembl 21 were used in this study, with most of the Rosids and all of the Asterids included.

The syntenic regions to *A. thaliana* chromosome 4 section were displayed in parallel (fig. 131, pink regions). Overall there was good conservation of syntenic blocks at this locus among the species listed except for *P. patens*, *S. moellendorffii*, and *C. reinhardtii*. The cluster in *G. max* appears to be split between the first and the second gene. The synteny among these plants strongly suggests that this genomic locus is conserved and often presents duplicated MDRs.

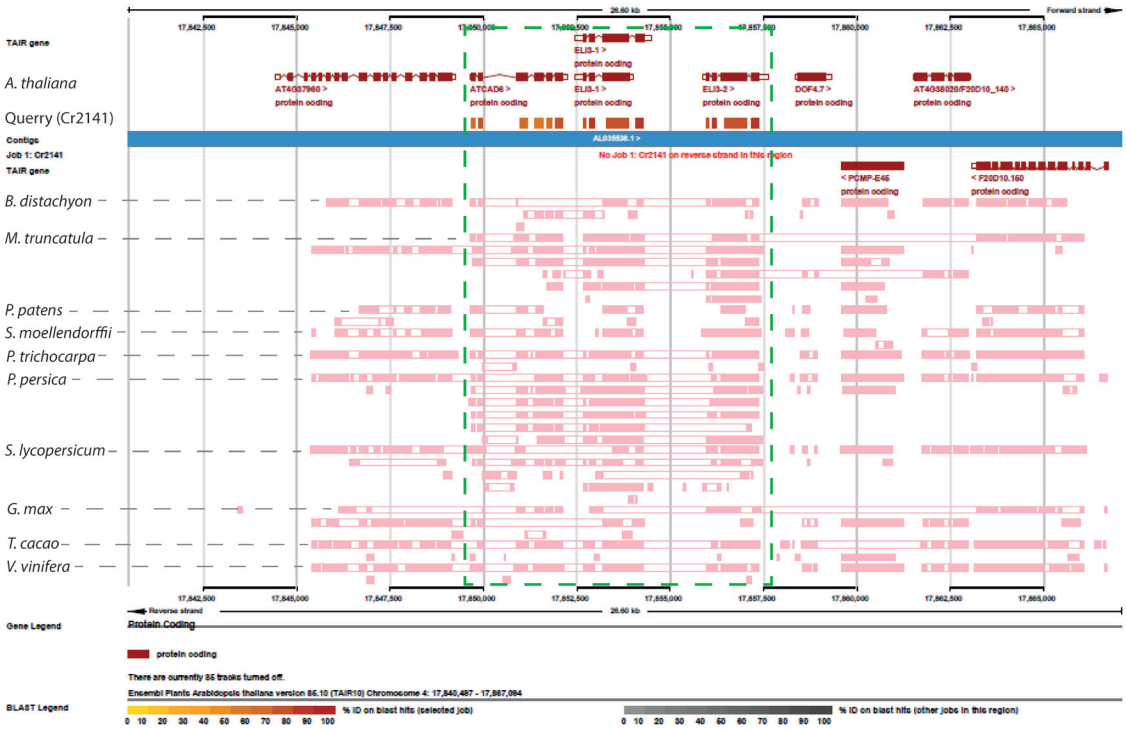


Figure 131: Syntenic regions between *A. thaliana* and other plant genomes at the Cr2141 homolog locus on chromosome 4. Figure created in ensembl Plant Genome Database. The plants are listed on the left side of the figure. The *A. thaliana* genes are illustrated at the top of the figure with exon-intron structure indicated. The positive hits of Cr2141 are illustrated under the *A. thaliana* genes in orange blocks. The syntenic blocks of other organisms are indicated as pink bars below. The area of interest is boxed in a green dashed line. The species displayed in this figure are: *Brachypodium distachyon*, *Medicago truncatula*, *Physcomitrella patens*, *Selginella moellendorffii*, *Populus trichocarpa*, *Prunus persica*, *Solanum lycopersicum*, *Glycine max*, *Theobroma cacao*, and *Vitis vinifera*.

A phylogenetic analysis of these identified syntenic genes was carried out to investigate whether the syntenic blocks of duplicated MDRs are dynamically duplicating throughout evolution or if they have been inherited as a block and conserved over evolution. The *A. thaliana* 3-MDR cluster, which comprises 3 CAD genes (Kim et al., 2007), was aligned to various other plant genomes. Most organisms presented at least 2 copies, and one plant (*P. persica*) had 8 tandem duplications. Some organisms had more than one genome block aligning to the *A. thaliana* one, potentially arising through genome duplication, or duplication of a chromosome section during whole genome duplication (Barakat et al., 2009).

5.2.6 Phylogenetic analysis of syntenic blocks

The genes of these duplication blocks were collected from 21 plant species (table 17) and were annotated using the naming system: Species/cluster number/copy number. The aligned sequences were analysed using the online PhyML server with the automatic model selection. This resulted in a tree (fig. 132) with overall good support for the branches.

In general, the monocots group together near the base of the tree, with SORBI1.1 and 3.1 (*S. bicolor*) as a sister clade to all others (fig. 132, blue). The Rosids in general group together

(fig. 132, red), as do the Asterids. SOYBN1.2 (*G. max)* is very far removed from SOYBN1.1 and does not group with any of the Rosids or Asterids. This indicates this MDR might not be truly homologous to the others as it does not group with the Fabaceae or Rosids clade (fig. 132, brown and red respectively). Another striking finding is the presence in *M. acuminata* of an MDR copy (MUSAM1.2) that does not group with the monocots. This was placed as a sister clade to the majority of the Asterids, but is supported there by a low Bootstrap value (0.298). Thus MUSAM1.2 could represent a copy of an MDR at that locus which occurred before separation of eudicots from monocots. This analysis is biased however because 5 out of 6 monocots analysed all belong to the Poales monocot clade and only one, *M. acuminata*, belongs to another monocot clade, the Zingiberales. In fact, *M. acuminata* is the only non-Poales monocot with a genome on the ensembl Plant Genomes Database. Without other monocot genome sequences from phyla outside the Poales it is not possible to draw any conclusions about the split of *M. acuminata* MDRs. Therefore this gene in *M. acuminata* could be a copy which was lost in the Poales clade, but retained in the Zingiberales. Likewise, two *C. roseus* MDRs (Cr027235 1.5 and Cr2141 1.6) are positioned outside the Asterids/Rosids clade together with some of the Brassicaceae (Rosids) MDRs (fig. 132, pink and green clade). This could represent an ancient duplication that occurred before the divergence of Rosids and Asterids which was retained in *C. roseus*, *A. thaliana*, *B. rapa*, and *B. oleracea*.

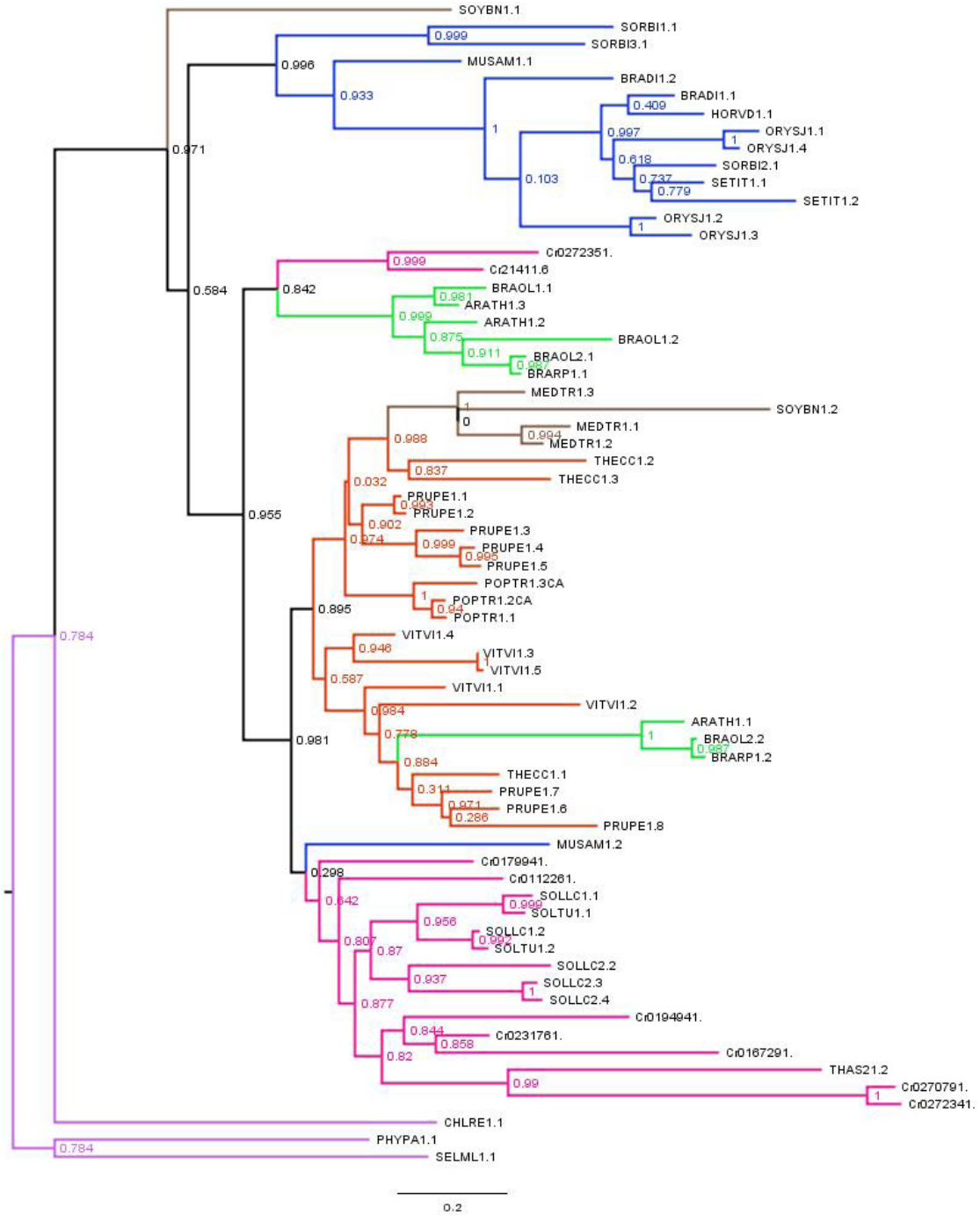


Figure 132: Phylogenetic tree of cluster MDRs from plant genomes. This tree was constructed using PhyML; bootstrap support values are printed next to each node; in lilac: the Lycopodiophyta, Bryophyta, and Chlorophyta; blue: monocots; orange: Rosids; green: Brassicaceae (also Rosids); brown: Fabaceae (also Rosids); pink: Asterids (includes *C. roseus*). Bar at the bottom represents distance of 0.2 changes per site.

The MDRs in the clusters of *A. thaliana*, *B. rapa*, and *B. oleracea*, are split into two groups. One group contains ARATH1.1, BRAOL2.2, and BRARP1.2, and the other contains ARATH1.2, ARATH1.3, BRAOL1.1, BRAOL2.1, and BRARP1.1. This signifies that in *B. oleracea* and *B. rapa* the duplicated clusters have been differentially conserved. It also illustrates that the genes have not been duplicated locally and retained since at least the divergence of these 3 plant species.

In *T. cacao* we observe two phyla containing MDRs; one which is a sister clade to the Fabaceae MDRs, and one which is positioned among the *V. vinifera*-containing clade. This suggests the cluster in this organism is composed of ancient duplications which have gone on to diverge enough so as to appear in two separate clades. *V. vinifera* and *T. cacao* group in the clade containing the other Rosids such as *P. persica* and *P. tremuloides*, and also the Fabaceae such as *M. truncatula* and *G. max*.

Eight of the *C. roseus* MDRs group with the asterids (*S. lycopersicum*, *Solanum tuberosum*) which is expected, and in general the duplications have good support. *S. lycopersicum* appears to have a duplication of the cluster which is not found in *S. tuberosum* (SOLLC2.3 and SOLC2.4 do not have homologs in *S. tuberosum*).

The grouping of two *C. roseus* MDRs with some of the Brassicaceae MDRs indicates there is a lot of information missing in this analysis. There are relatively few Asterid genomes represented in the tree, compared to Rosids and Monocots. The two Asterid genomes other than *C. roseus* are relatively closely related and therefore do not provide a great deal of novel information about the Asterids clade. More sequences of plants in the same family as *C. roseus* could help resolve some of the nodes and provide a clearer Asterid/Rosid distinction.

Finally, it is apparent from this phylogenetic analysis that the copies of the MDRs are not present in a conserved duplicated block, but rather they are dynamically duplicating throughout the evolution of these plant species. The repartition of the copies mostly in groups of related plant species indicates the duplications have occurred at a later stage after the radiation of spermatophyta into the different clades. A similar conclusion was drawn by Barakat et al. (2009) when studying the genomic organisation of the CAD/SAD genes in the available plant genomes.

5.3 Discussion

5.3.1 Genomic context of MDRs and SDRs in *C. roseus*

The whole genome sequencing and assembly of *C. roseus* has revealed that MDRs and SDRs are often present in clusters of duplicated genes. Most notable in *C. roseus* is the cluster on the WGC number 126 which contains 10 MDRs (fig. 130). One of the copies in that cluster is THAS2, another one (Cr024234) has been shown by collaborators to produce a different product from strictosidine aglycon, and a third copy in the cluster (Cr023176) has been shown to have 10HGO activity (Krithika et al., 2015). Another interesting cluster is on WGC number 27 which presents 6 copies of SDRs. These, conversely to the MDR cluster on WGC 126, are mostly not in tandem duplication. One of the SDRs in this cluster has been shown to reduce strictosidine aglycon and produce a product with yet another structural conformation (Chapter 4).

5.3.2 Ortholog groups in MDRs and SDRs

The analysis of the Clusters of Orthologous Groups (COGs) of the MDRs and SDRs indicated the scope of variation present in these enzyme families. MDRs and SDRs of *C. roseus* appear to be of diverse origin based on the predicted COG for each predicted protein. This is in accordance with whole family studies that have been conducted with MDR and SDR sequences present in online databases (Persson et al., 2008, Jörnvall et al., 1999, Riveros-Rosas et al., 2003).

5.3.3 Phylogeny of MDRs in *C. roseus*

The *C. roseus* MDRs were collected from the transcriptome and genome but it is possible some MDRs have not been included in the alignment and the phylogenetic analysis. Protein alignment showed the presence of large insertions-deletions at flexible loop regions, which presumably impact the catalytic function of these enzymes. A phylogenetic study of the *C. roseus* MDRs suggests that there are two distinct patterns of MDR evolution in this plant. The first is the largely non-duplicating MDRs placed on long branches (fig. 128, top half). These genes could be implicated in primary metabolism and are therefore under selective pressure to maintain both sequence and copy number. To verify this hypothesis, MDRs from those branches would have to be tested for their predicted activity.

The second pattern includes a more radially diverging group of MDRs that have undergone recent duplication. Ten of these highly duplicated MDRs are present in the large MDR cluster on the WGC number 126 which, in theory, arose through tandem duplications (discussed above). This supports the results of the phylogenetic analysis which places all these genes in clades near each other. Their recent divergence is evidenced from the short branches which reflect the number of changes to the protein sequence accrued since duplication. The MDRs in *C. roseus* which were shown to have activity with strictosidine aglycon (THAS1-4 and HYS) cluster on one clade of the MDR phylogeny tree (fig. 126 and 127), whereas the *C. roseus* homologs in the RsVR2 group cluster on another highly duplicated branch. Therefore, it is reasonable to suggest that this subfamily of MDRs has led, through duplication and neofunctionalisation in the Apocynaceae, to increased chemical diversity.

The tree produced here of the MDRs in *C. roseus* recalls the MDR tree constructed in Jörnvall et al. (1999) in which the plant MDRs cluster in a very highly duplicated branch. This behaviour of plant MDRs could be providing the primary material for neofunctionalisation and ultimately chemical diversity.

5.3.4 Phylogeny of SDRs in *C. roseus*

Conversely to the MDR phylogenetic tree, the SDR tree does not display any large differences in diversification rate. The tree however is about twice as long as that of the MDRs even though there are only 10 sequences more (54 protein sequences for SDRs vs. 44 for MDRs). The SDRs do not appear to have undergone a radial duplication and diversification recently, as evidenced by

the lack of highly duplicated branches such as that seen in the MDR phylogenetic tree (fig. 128). Instead, the SDR phylogenetic tree displays multiple clades with numerous duplication events in each one. The duplication of the ADH10 locus (Cr013447 and Cr016747) suggest these copies might be the material for further neofunctionalisation.

5.3.5 MDR duplication blocks in other plant genomes

To investigate whether the large duplicated MDR locus of WGC 126 was found more widely among the plant kingdom, one of the genes of this locus, Cr2141, was submitted to BLAST against the plant genome database on ensembl (Kersey et al., 2016). Cr2141 was chosen because homologs of this gene were found most widely among other plant species. A total of 21 plants were analysed, and a similar cluster of MDRs was identified through syntenic alignment in 18 plants. This intriguing result could either suggest that an ancient duplication of an ancestral MDR had occurred before divergence and speciation and that it was conserved among all these plants, or it could suggest that this locus was constantly being duplicated and presents an example of active gene birth/death.

To answer this question, a phylogenetic tree was constructed using the protein sequences of the MDR gene homologs found in these loci. The results indicate that the second scenario is more likely, since the paralogs and orthologs display a grouping among species and not among orthologs. This suggests that repeatedly during evolution, MDRs were duplicated at that location. The repeated “duplication after speciation” across almost all plant species analysed signifies that the MDRs at this locus might be highly duplicating across the angiosperms. There are undoubtedly other examples in plants of this phenomenon, perhaps even in other MDR or SDR clusters.

Jörnvall et al. (1999) observed, when constructing a phylogenetic tree of the MDRs and SDRs from published genomes, that the MDRs appear to be more often recently duplicated with subsequent neofunctionalisation than the SDRs. This is also observed in the *C. roseus* MDR dataset (fig. 128) which illustrates the recent expansion through duplication of the CAD-like MDRs.

5.4 Conclusions

This chapter investigates the role of gene duplications as drivers for chemical diversity. Duplication and redundancy has been observed for other MDRs and specifically members of the CAD subfamily. The *A. thaliana* CAD5 is the most active isoform for sinapaldehyde but its knockout does not significantly hinder production and deposition of lignin, indicating that the redundancy of CAD and SAD in *Arabidopsis* is enough to result in the eventual full deposition of lignin up to wild-type levels (Kim et al., 2004). A similar enzymatic redundancy is observed in the HYSs of *C. roseus* (Chapter 2 and 3 and detailed in Stavrinides et al. (2016)). Therefore, a phylogenetic analysis of the enzyme family in *C. roseus* was carried out to better understand the evolution of this family.

The resulting phylogenetic trees of the *C. roseus* MDRs were in general well supported. To increase the confidence levels in the trees, comparison with a structure-based alignment would be ideal. However, there is not nearly enough structural data available for this analysis. In the *C. roseus* MDRs the main differences evidenced when the protein crystal structures were aligned were found mainly in the loop regions. These regions appeared to have undergone many insertion-deletions which makes alignment of the sequences difficult without additional support from crystal structure alignments. Diversification in these loop sections, brought about through genetic drift, is likely to be key for acquisition of novel enzyme functions. As described in Chapter 3, loop sections above the active site of THAS1 and HYS played a key role in modulating the catalytic function of these enzymes.

The phylogenetic tree of the MDRs suggests that the HYSs enzymes have originated from a single ancestral enzyme that underwent several duplications. Numerous enzymes with HYS/THAS activity have been identified from *C. roseus* and it is not clear what role this enzymatic redundancy has in the plant. It is apparently a stable state, given the divergent sequences of these enzymes. Stable duplication without subsequent co-option is possible in the case that an increase in dosage does not have a negative effect on the organism. Favourable selection of this dosage increase would result in the maintenance of the doubled gene and its role in the biological system (Ganfornina and Sánchez, 1999). This could be a driving force behind maintenance of all the HYSs in the *C. roseus* genome. Furthermore, the evidence suggests that downstream pathway genes such as T3R or Cr027234 might have evolved from duplications of a HYS gene or an ancestral HYS gene. This is expected, as an active site already suited to accept a certain type of substrate is primed for catalysis of a similar reaction on a slightly modified substrate (Gerlt and Babbitt, 2001). Therefore, it is hypothesized that it is likely the same group of genes will be duplicated again in the future and give rise to more chemical and functional diversity in the pathway.

Furthermore, the phylogenetic analysis of the MDRs suggests that THAS1-4 and HYS have originated from the same ancestral MDR. This information, combined with the information that the MDR copies on the duplicated block of WGC 126 have been constantly duplicating throughout the angiosperm evolution (described above in section 5.3.5), suggests WGC 126 was the original location of their ancestral genes. Indeed, THAS2 is still present on the WGC 126. Intriguingly, the location of THAS1 and THAS3 in close proximity on the WGC 371 to a copy of the indispensable iridoid biosynthetic gene Secologanin Synthase (SLS, (Irmler et al., 2000)) suggests this location could be the seed to generation of a biosynthetic cluster such as those observed in some other plants (Field and Osbourn, 2008). As discussed in Chu et al. (2011), it appears that biosynthetic cluster genesis in plants is effectively a recruitment process in which duplicated and neofunctionalised genes are slowly recruited to a genomic locus near the other genes of that pathway. As the MIA biosynthetic pathway in *C. roseus* is not unidirectional after generation of strictosidine, but rather can follow any of a multitude of pathways for generation of hundreds of MIAs, it is not clear if such a biosynthetic cluster would be “useful” for the plant. A more detailed

genomic analysis and a more complete genome sequence of *C. roseus* could help address this question of how clusters for complex, highly branched biosynthetic pathways are formed.

Overall this work indicates that actively duplicating loci in the medicinal plant *C. roseus* could provide the material for pathway diversification in secondary metabolism. Abundant duplication of HYS could signal a positive selection pressure for more of this enzyme. These duplicated genes could be undergoing mutational drift to give rise to new functions. As discussed by Kliebenstein (2008), tandem duplications are more commonly found for genes which are not part of the primary metabolism, but secondary metabolism can display and benefit from gene duplications.

5.5 Materials and Methods

5.5.1 Reductase sequences

The transcriptome and genome annotations were searched manually for medium- and short-chain dehydrogenases/reductases by searching for the GO terms characteristic of ADHs ('cinnamyl alcohol dehydrogenase', 'Rossman-fold superfamily protein', 'elicitor-activated gene', 'zinc-binding dehydrogenase', etc.). The candidates were fetched from the transcriptome and their identity was verified by first determining the open reading frame of the gene (using the online server expasy.org (Gasteiger et al., 2003)) and then BLASTing the predicted protein sequence against the NCBI protein database to manually verify the protein family and the completeness of the sequence.

The protein sequence of each reductase was subjected to a search against the Clusters of Orthologous Groups (COG, Tatusov et al. (1997)) database v.1 of NCBI (Marchler-Bauer et al., 2015) to determine which family of MDR or SDR each protein belongs to. This was done to get a better idea of the general diversity of the MDRs and SDRs in this dataset.

The CDS, protein sequence, and genomic locus of the reductases were manually collected using the software Geneious v. 8 and the unpublished whole genome assembly version 2.A (Buell and Kim, in preparation).

5.5.2 MDR phylogenetic analysis

The MDR protein sequences were aligned using different software which employ different algorithms. Alignment was done with the software Fast Statistical Alignment (fsa, (Bradley et al., 2009)), using a gap open penalty of 1, with the software PRANK ((Whelan and Goldman, 2001), with the default parameters, using the software mafft v7.271 (Katoh and Standley, 2013) using the defaults. RAxML v8.2.8 (Stamatakis, 2014) was executed to test which model would best explain the protein alignment given empirical amino acid frequencies.

The MDR protein alignment that was produced with ClustalOmega (Sievers et al., 2011) was used as input for a phylogenetic analysis using the PhyML online server (http://www.atgc-montpellier.fr/phyml/) (Guindon et al., 2010). PhyML uses Maximum Likelihood estimation to select the most likely tree given the model and the data. The automatic model selection was used with Akaike Information Criterion as the selection criterion and NNI (Nearest Neighbor Interchange) as the tree improvement method. Bootstrap support was calculated based on 100 runs.

Bayesian inference was used in parallel to calculate a phylogeny of the MDRs protein sequences. MrBayes v. 3.2.0 (Ronquist et al., 2012) was used in an Ubuntu environment using the ClustalOmega alignment of the MDRs. Multiple-chain Markov chain Monte Carlo (MC3) is used in all analyses with 4 chains, unless otherwise specified. Initially the 4 chains were heated with the defaults. A test run was executed with two MC3 runs, setting the amino acid model to "mixed" to test which model best fit the data. The MC3 runs both converged to one model (WAG, Whelan and Goldman (2001)). Convergence was monitored by plotting the run results in Tracer (http://tree.bio.ed.ac.uk/software/tracer/).

The simple "mixed" model (1000000 generations) was compared to the more complex model which allows for variation in evolutionary rates between sites "mixed + gamma" (713000 generations). Again both runs converged on the WAG model but the fit to the data was slightly better when including the gamma factor in the model (log likelihood of -239887 for mixed + gamma compared to -243647 for simply mixed). Due to the high demands of Bayesian inference on computation time specific amino acid substitution models were not tested individually.

5.5.3 SDR phylogenetic analysis

The SDR protein sequences (54) were collected from the *C. roseus* transcriptome (Góngora-Castillo et al., 2012) and the assembled WGC version 2.A. The sequences were aligned using ClustalOmega (Sievers et al., 2011) and the alignment was verified visually with SuiteMSA (Anderson et al., 2011). The alignment was used to infer a phylogenetic tree using both the PhyML online server (www.atgc-montpellier.fr/phyml) and the command-line based PhyML version 3.0 in an Ubuntu environment (Guindon et al., 2010).

The online server was run with automatic model selection and 300 bootstrap iterations. In the command-line PhyML, where automatic model selection is not possible, the WAG amino acid substitution model (Whelan and Goldman, 2001) was selected, and branch support values were calculated using aLRT (cubic approximation, mixture of Chi2s distribution). In both cases the starting tree was inferred using BioNJ and was improved through NNI. The gamma was allowed to vary across 4 different classes to account for different substitution rates at different sites of the protein sequence. The resulting trees were visualised using Figtree (http://tree.bio.ed.ac.uk/software/figtree).

5.5.4 Phylogenetic analysis of syntenic blocks

The protein sequence of Cr2141 was submitted to BLAST against the genome of *A. thaliana* of the ensembl Plant Genome Database (http://plants.ensembl.org, release 32)(Kersey et al., 2016). The *A. thaliana* positive hit region with the highest identity was studied in more detail. A syntenic alignment of this chromosomal section of *A. thaliana* was made to identify conserved duplication regions in other plant genomes. The identified loci from other organisms were collected after verifying they were MDRs.

An alignment of the MDRs, together with the 10 *C. roseus* MDRs of the WGC 126 cluster, was made in MEGA using a ClustalW algorithm (Thompson et al., 2002) with a BLOSUM matrix. The gap open penalty was 10 and the gap extension penalty was 0.1. This was verified visually before using the alignment to construct a tree in MEGA. Tree construction was done by Maximum likelihood using the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) hierarchy method with 300 bootstrap iterations. Two different amino acid models were tested; the PAM matrix (Dayhoff) (Dayhoff and Schwartz, 1978) and the JTT matrix (Jones et al., 1992). The mutation rates were allowed to vary across protein sites by setting a Gamma distribution.

5.6 References

ANDERSON, C. L., STROPE, C. L. & MORIYAMA, E. N. 2011. SuiteMSA: visual tools for multiple sequence alignment comparison and molecular sequence simulation. *BMC Bioinformatics*, 12, 184-184.

ASLAM, J., KHAN, S. H., SIDDIQUI, Z. H., FATIMA, Z., MAQSOOD, M., BHAT, M. A., NASIM, S. A., ILAH, A., AHMAD, I. Z. & KHAN, S. A. 2010. *Catharanthus roseus* (L.) G. Don. An important drug: it's applications and production. *Pharmacie Globale (IJCP)*, 4, 1-16.

BARAKAT, A., BAGNIEWSKA-ZADWORNA, A., CHOI, A., PLAKKAT, U., DILORETO, D. S., YELLANKI, P. & CARLSON, J. E. 2009. The cinnamyl alcohol dehydrogenase gene family in Populus: phylogeny, organization, and expression. *BMC Plant Biology*, 9, 26-26.

BERGTHORSSON, U., ANDERSSON, D. I. & ROTH, J. R. 2007. Ohno's dilemma: Evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 17004-17009.

BRADLEY, R. K., ROBERTS, A., SMOOT, M., JUVEKAR, S., DO, J., DEWEY, C., HOLMES, I. & PACHTER, L. 2009. Fast Statistical Alignment. *PLoS Comput Biol*, 5, e1000392.

BROCKINGTON, S. F., YANG, Y., GANDIA-HERRERO, F., COVSHOFF, S., HIBBERD, J. M., SAGE, R. F., WONG, G. K. S., MOORE, M. J. & SMITH, S. A. 2015. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytologist*, 207, 1170-1180.

BUNSUPA, S., KATAYAMA, K., IKEURA, E., OIKAWA, A., TOYOOKA, K., SAITO, K. & YAMAZAKI, M. 2012. Lysine Decarboxylase Catalyzes the First Step of Quinolizidine Alkaloid Biosynthesis and Coevolved with Alkaloid Production in Leguminosae. *The Plant Cell Online*.

CHEN, H., LI, G. L., KOLLNER, T. G., JIA, Q. D., GERSHENZON, J. & CHEN, F. 2014. Positive Darwinian selection is a driving force for the diversification of terpenoid biosynthesis in the genus Oryza. *Bmc Plant Biology*, 14.

CHU, H. Y., WEGEL, E. & OSBOURN, A. 2011. From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. *The Plant Journal*, 66, 66-79.

DAYHOFF, M. O. & SCHWARTZ, R. M. A model of evolutionary change in proteins. In *Atlas of protein sequence and structure*, 1978. Citeseer.

DES MARAIS, D. L. & RAUSHER, M. D. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature*, 454, 762-765.

FIELD, B. & OSBOURN, A. E. 2008. Metabolic diversification - Independent assembly of operon-like gene clusters in different plants. *Science*, 320, 543-547.

GANFORNINA, M. D. & SÁNCHEZ, D. 1999. Generation of evolutionary novelty by functional shift. *BioEssays*, 21, 432-439.

GASTEIGER, E., GATTIKER, A., HOOGLAND, C., IVANYI, I., APPEL, R. D. & BAIROCH, A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31, 3784-3788.

GEISSLER, M., BURGHARD, M., VOLK, J., STANIEK, A. & WARZECHA, H. 2015. A novel cinnamyl alcohol dehydrogenase (CAD)-like reductase contributes to the structural diversity of monoterpenoid indole alkaloids in Rauvolfia. *Planta*, 1-12.

GERLT, J. A. & BABBITT, P. C. 2001. Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annual Review of Biochemistry*, 70, 209-246.

GEU-FLORES, F., SHERDEN, N. H., COURDAVAULT, V., BURLAT, V., GLENN, W. S., WU, C., NIMS, E., CUI, Y. & O'CONNOR, S. E. 2012. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature*, 492, 138-142.

GÓNGORA-CASTILLO, E., CHILDS, K. L., FEDEWA, G., HAMILTON, J. P., LISCOMBE, D. K., MAGALLANES-LUNDBACK, M., MANDADI, K. K., NIMS, E., RUNGUPHAN, W., VAILLANCOURT, B., VARBANOVA-HERDE, M., DELLAPENNA, D., MCKNIGHT, T. D., O'CONNOR, S. & BUELL, C. R. 2012. Development of Transcriptomic Resources for Interrogating the Biosynthesis of Monoterpene Indole Alkaloids in Medicinal Plant Species. *PLoS ONE*, 7, e52506.

GUINDON, S., DUFAYARD, J.-F., LEFORT, V., ANISIMOVA, M., HORDIJK, W. & GASCUEL, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59, 307-321.

GUIRIMAND, G., COURDAVAULT, V., LANOUE, A., MAHROUG, S., GUIHUR, A., BLANC, N., GIGLIOLI-GUIVARC'H, N., ST-PIERRE, B. & BURLAT, V. 2010. Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? *Bmc Plant Biology*, 10.

IRMLER, S., SCHRODER, G., ST-PIERRE, B., CROUCH, N. P., HOTZE, M., SCHMIDT, J., STRACK, D., MATERN, U. & SCHRODER, J. 2000. Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450CYP72A1 as secologanin synthase. *Plant Journal*, 24, 797-804.

JONES, D. T., TAYLOR, W. R. & THORNTON, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences: CABIOS*, 8, 275-282.

JÖRNVALL, H., HÖÖG, J.-O. & PERSSON, B. 1999. SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Letters*, 445, 261-264.

KATOH, K. & STANDLEY, D. M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30, 772-780.

KELLNER, F., KIM, J., CLAVIJO, B. J., HAMILTON, J. P., CHILDS, K. L., VAILLANCOURT, B., CEPELA, J., HABERMANN, M., STEUERNAGEL, B., CLISSOLD, L., MCLAY, K., BUELL, C. R. & O'CONNOR, S. E. 2015. Genome-guided investigation of plant natural product biosynthesis. *The Plant Journal*, 82, 680-692.

KERSEY, P. J., ALLEN, J. E., ARMEAN, I., BODDU, S., BOLT, B. J., CARVALHO-SILVA, D., CHRISTENSEN, M., DAVIS, P., FALIN, L. J., GRABMUELLER, C., HUMPHREY, J., KERHORNOU, A., KHOBOVA, J., ARANGANATHAN, N. K., LANGRIDGE, N., LOWY, E., MCDOWALL, M. D., MAHESWARI, U., NUHN, M., ONG, C. K., OVERDUIN, B., PAULINI, M., PEDRO, H., PERRY, E., SPUDICH, G., TAPANARI, E., WALTS, B., WILLIAMS, G., TELLO–RUIZ, M., STEIN, J., WEI, S., WARE, D., BOLSER, D. M., HOWE, K. L., KULESHA, E., LAWSON, D., MASLEN, G. & STAINES, D. M. 2016. Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Research*, 44, D574-D580.

KIM, S.-J., KIM, K.-W., CHO, M.-H., FRANCESCHI, V. R., DAVIN, L. B. & LEWIS, N. G. 2007. Expression of cinnamyl alcohol dehydrogenases and their putative homologues during *Arabidopsis thaliana* growth and development: Lessons for database annotations? *Phytochemistry*, 68, 1957-1974.

KIM, S.-J., KIM, M.-R., BEDGAR, D. L., MOINUDDIN, S. G. A., CARDENAS, C. L., DAVIN, L. B., KANG, C. & LEWIS, N. G. 2004. Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 1455-1460.

KLIEBENSTEIN, D. J. 2008. A Role for Gene Duplication and Natural Variation of Gene Expression in the Evolution of Metabolism. *PLoS ONE*, 3, e1838.

KRITHIKA, R., SRIVASTAVA, P. L., RANI, B., KOLET, S. P., CHOPADE, M., SONIYA, M. & THULASIRAM, H. V. 2015. Characterization of 10-Hydroxygeraniol Dehydrogenase from *Catharanthus roseus* Reveals Cascaded Enzymatic Activity in Iridoid Biosynthesis. *Scientific Reports*, 5, 8258.

LAU, N.-S., MAKITA, Y., KAWASHIMA, M., TAYLOR, T. D., KONDO, S., OTHMAN, A. S., SHU-CHIEN, A. C. & MATSUI, M. 2016. The rubber tree genome shows expansion of gene family associated with rubber biosynthesis. *Scientific Reports*, 6, 28594.

LUIJENDIJK, T. J. C., VAN DER MEIJDEN, E. & VERPOORTE, R. 1996. Involvement of strictosidine as a defensive chemical in *Catharanthus roseus*. *Journal of Chemical Ecology*, 22, 1355-1366.

MARCHLER-BAUER, A., DERBYSHIRE, M. K., GONZALES, N. R., LU, S., CHITSAZ, F., GEER, L. Y., GEER, R. C., HE, J., GWADZ, M., HURWITZ, D. I., LANCZYCKI, C. J., LU, F., MARCHLER, G. H.,

SONG, J. S., THANKI, N., WANG, Z., YAMASHITA, R. A., ZHANG, D., ZHENG, C. & BRYANT, S. H. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res*, 43, D222-6.

MIETTINEN, K., DONG, L., NAVROT, N., SCHNEIDER, T., BURLAT, V., POLLIER, J., WOITTIEZ, L., VAN DER KROL, S., LUGAN, R., ILC, T., VERPOORTE, R., OKSMAN-CALDENTEY, K.-M., MARTINOIA, E., BOUWMEESTER, H., GOOSSENS, A., MEMELINK, J. & WERCK-REICHHART, D. 2014. The seco-iridoid pathway from *Catharanthus roseus*. *Nature Communications*, 5, 3606.

MOUMMOU, H., KALLBERG, Y., TONFACK, L., PERSSON, B. & REST, B. 2012. The Plant Short-Chain Dehydrogenase (SDR) superfamily: genome-wide inventory and diversification patterns. *BMC Plant Biology*, 12, 1-17.

MUNKERT, J., POLLIER, J., MIETTINEN, K., VAN MOERKERCKE, A., PAYNE, R., MÜLLER-URI, F., BURLAT, V., O'CONNOR, SARAH E., MEMELINK, J., KREIS, W. & GOOSSENS, A. 2015. Iridoid Synthase Activity Is Common among the Plant Progesterone 5β-Reductase Family. *Molecular Plant*, 8, 136-152.

PERSSON, B., HEDLUND, J. & JÖRNVALL, H. 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cellular and Molecular Life Sciences*, 65, 3879-3894.

QU, Y., EASSON, M. L. A. E., FROESE, J., SIMIONESCU, R., HUDLICKY, T. & DE LUCA, V. 2015. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proceedings of the National Academy of Sciences*, 112, 6224-6229.

RIVEROS-ROSAS, H., JULIÁN-SÁNCHEZ, A., VILLALOBOS-MOLINA, R., PARDO, J. P. & PIÑA, E. 2003. Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily. *European Journal of Biochemistry*, 270, 3309-3334.

RONQUIST, F., TESLENKO, M., VAN DER MARK, P., AYRES, D. L., DARLING, A., HÖHNA, S., LARGET, B., LIU, L., SUCHARD, M. A. & HUELSENBECK, J. P. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61, 539-542.

RUEFFER, M., NAGAKURA, N. & ZENK, M. H. 1978. Strictosidine, the common precursor for monoterpenoid indole alkaloids with 3 α and 3 β configuration. *Tetrahedron Letters*, 19, 1593-1596.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M., SÖDING, J., THOMPSON, J. D. & HIGGINS, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 539-539.

SINGH, A. 2008. Phytochemicals of Gentianaceae: a review of pharmacological properties. *International Journal of Pharmaceutical Sciences and Nanotechnology*, 1, 33-36.

STAMATAKIS, A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*.

STAVRINIDES, A., TATSIS, E. C., CAPUTI, L., FOUREAU, E., STEVENSON, C. E. M., LAWSON, D. M., COURDAVAULT, V. & O'CONNOR, S. E. 2016. Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity. *Nat Commun*, 7.

STRAUB, S. C. K., MOORE, M. J., SOLTIS, P. S., SOLTIS, D. E., LISTON, A. & LIVSHULTZ, T. 2014. Phylogenetic signal detection from an ancient rapid radiation: Effects of noise reduction, long-branch attraction, and model selection in crown clade Apocynaceae. *Molecular Phylogenetics and Evolution*, 80, 169-185.

SUDŽUKOVIĆ, N., SCHINNERL, J. & BRECKER, L. 2016. Phytochemical meanings of tetrahydro-β-carboline moiety in strictosidine derivatives. *Bioorganic & Medicinal Chemistry*, 24, 588-595.

TATUSOV, R. L., KOONIN, E. V. & LIPMAN, D. J. 1997. A Genomic Perspective on Protein Families. *Science*, 278, 631-637.

THOMPSON, J. D., GIBSON, T. & HIGGINS, D. G. 2002. Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*, 2.3. 1-2.3. 22.

TOBIAS, C. M. & CHOW, E. K. 2005. Structure of the cinnamyl-alcohol dehydrogenase gene family in rice and promoter activity of a member associated with lignification. *Planta*, 220, 678-688.

WENG, J. K., PHILIPPE, R. N. & NOEL, J. P. 2012. The Rise of Chemodiversity in Plants. *Science*, 336, 1667-1670.

WHELAN, S. & GOLDMAN, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular biology and evolution*, 18, 691-699.

YERKES, N. 2010. Purification and substrate specificity of new *C. roseus* enzymes. PhD, Massachusetts Institute of Technology.

Chapter 6

Conclusions

There and back again: a tale of scientific discovery from chemical to genetic diversity and back again

6.1 Heteroyohimbine synthases

In the 1970s it was first demonstrated that heteroyohimbines, a type of monoterpene indole alkaloid (MIA), are the products of enzymatic reduction of strictosidine aglycon (Stöckigt et al., 1976). Forty years later, the enzymes have finally been cloned and characterised, as described in Chapters 2 and 3. This discovery was made possible through a transcriptome database of *C. roseus* (Góngora-Castillo et al., 2012), a medicinal plant that produces hundreds of MIA, inlcuding several heteroyohimbines. The chemical structural diversity of the MIA in this and related plants stems from the reactivity of the pathway's central intermediate, strictosidine aglycon. As discussed in detail in Chapters 1, 3, and 4 this intermediate can rearrange intramolecularly to yield structurally different MIA. How the reactivity of such compounds is controlled, or how it can be exploited, is of great interest to efforts in enzyme engineering, biocatalysis and genetic engineering of pathways in heterologous hosts.

The group of enzymes that catalyse the formation of the heteroyohimbines have been named heteroyohimbine synthases (HYS). Five HYSs were identified and were found to be related at the sequence and structural level. However, one enzyme produced dramatically different ratios of the three heteroyohimbines found in *C. roseus*: ajmalicine, 19-epiajmalicine, and tetrahydroalstonine (fig. 133). Site directed mutational analysis and structural characterisation allowed the identification of the active site elements that control the stereoselectivity among these enzymes (Chapter 3).

The redundancy and the localisation of some of these HYSs to the same compartment as Strictosidine glucosidase, the enzyme that generates strictosidine aglycon, suggests that the HYSs could be part of a directed overflow metabolism (Frelin et al., 2015). In this model, the strictosidine aglycon is a reactive and toxic intermediate that, while playing a role in plant defence, can also pose a threat to the plant producing it. The reaction of the HYSs serves as a mechanism to neutralise excess strictosidine aglycone into an unreactive, non-toxic product.

6.2 Enzymatic generation of the vallesiachotaman skeleton

The HYSs are alcohol dehydrogenases, specifically of the Medium-chain dehydrogenase/ reductases (MDRs) class. An alcohol dehydrogenase from the Short-chain dehydrogenase/ reductase (SDR) class was also discovered to reduce strictosidine aglycon to generate a skeleton unrelated to the heteroyohimbines. This enzyme, ADH10, acts on a different isomer

of the strictosidine aglycon substrate than the HYSs. ADH10 is the first enzyme discovered to generate the so-called vallesiachotaman skeleton of MIA (fig. 133). This has provided us with an unprecedented view into the dynamics of the strictosidine aglycon intermediate. This enzyme could form the basis of a search for similar enzymes taking part in MIA in *C. roseus* and other plants both at the central branchpoint of strictosidine aglycon and at other parts of the pathway.

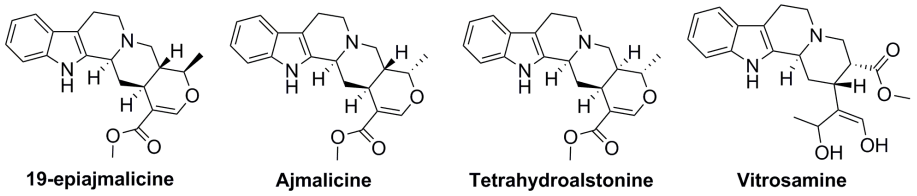


Figure 133: The heteroyohimbines of *C. roseus* and vitrosamine.

6.3 Phylogenetics reveals evolution dynamics in the HYSs

Discovery of multiple HYSs raised questions about their evolution and role in the development of the downstream pathways in MIA biosynthesis. The whole genome sequencing of *C. roseus* revealed the genomic location of many enzymes related to HYSs, allowing an understanding of the genomic context of these genes. A phylogenetic analysis of the entire MDRs enzyme family revealed that the HYSs are part of a rapidly duplicating group which also appears to have produced other MIA biosynthetic enzymes such as T3R (Qu et al., 2015) and Cr027234 (Chapter 5). This group of rapidly duplicating MDRs could be providing the primary material necessary for evolution of novel enzymes capable of acting at different parts of the pathway.

More specifically, the phylogenetics of the THAS and HYS enzymes (Chapter 5) suggests that these enzymes are descended from the same ancestor gene. The genome assembly v. 2.A places THAS1 and THAS3 in the same genomic locus as an upstream pathway gene, SLS. The distant relationship between THAS1 and THAS3 makes them unlikely to be locally duplicated. Instead, these genes appear to have been recruited by relocation to this genomic location. This could represent a novel finding in plant secondary metabolism research, as it has been posited that biosynthetic clusters in plants originate via piecewise recruitment to a common genomic locus (Chu et al., 2011). The “birth place” of THAS1 and THAS3 was most likely the highly duplicated MDR cluster on the WGC 126 (Chapter 5) as they both have similar paralogs in that cluster. The physical emergence of these clusters is not yet understood at the molecular level. The potential advantage a clustered pathway could have on the host plant (co-expression, co-inheritance) could be the driving force for the maintenance of the clusters (Chu et al., 2011).

6.4 Future directions

The MIA pathway of *C. roseus* is still not fully elucidated. The enzymes discovered throughout this project and described in this thesis can act as a basis upon which to discover the missing enzymes of this pathway. Enzymes which have evolved from the same ancestral gene could show

different activities against MIA pathway intermediates. In fact, an MDR enzyme related to THAS has recently been discovered to act on a downstream part of an unrelated (sparagan type) MIA pathway in the related plant *Rauvolfia serpentina* (Geissler et al., 2015). A detailed investigation of such plants with diverging MIA pathways will only add to the rich diversity of enzyme function.

All of the enzymes described here in this thesis have been shown to be stereoselective imine reductases. There is industrial interest in stereoselective enzymes and the results described here could be built upon to expand the enzymatic toolbox. Exon shuffling is an efficient method used in protein engineering for developing novel functions. The loop swapping in HYS and THAS1 described in Chapter 3, although not an entire exon swap, illustrates how swapping an active site loop from a specific enzyme into a promiscuous enzyme resulted in a new, stable, specific enzyme. Additional protein engineering strategies, such as DNA shuffling, could also be used to expand the substrate and product scope of these enzymes.

Finally, this work touches on the evolution of the MDRs and the related enzymes, Short-chain dehydrogenase/reductases. The phylogenetic tree of the MDRs suggests that a branch of these enzymes has been duplicated numerous times. This extensive duplication suggests that there has been a vast amount of neo-functionalisation in this group of enzymes. This branch of highly duplicating MDRs could serve as an excellent starting point when searching for reductases in other plant biosynthetic pathways.

6.5 Perspectives and outlook

The advancements of transcriptome and genome sequencing have allowed us to glean, for the first time, the mechanisms underlying adaptation and evolution in the context of secondary metabolism. It is remarkable to be able to discover and test the effects of each mutation and of each gene duplication. As it has been once said “nothing in biology makes sense except in the light of evolution” (Dobzhansky, 2013). In *C. roseus* we have now shown several examples of how the chemical diversity is a direct result of the genetic diversity. It is due to selective pressures that enzymes such as HYSs have evolved the ability to reduce strictosidine aglycone into different compounds and substrates from other parts of the MIA pathway.

The selective pressures that drive MIA evolution in *C. roseus* and other plants is still not known. It is not clear if these secondary metabolites of *C. roseus* take part in plant defence by deterring herbivore feeding or by preventing infection by pathogens. Although there have been some attempts to test the role of these alkaloids, the complexity of the mixture that is produced by the plant has greatly complicated these ecological studies. Feeding studies on insects with purified strictosidine and analogues did not show any clear effect on the lifecycle of a generalist herbivore (*Spodoptera littoralis*) but did have a slight growth-retardant effect at the young larval stage (Sudžuković et al., 2016). Similar questions have been posed about the enormous terpene chemical diversity and the role evolution has played in expanding the diversity of this family of compounds in plants. It is generally accepted that terpenes play a role in plant interactions with the environment, in particular in plant-insect interactions. The repeated appearance of increased

terpene chemical diversity in plant evolution suggests that chemical diversity is beneficial *per se* (Pichersky and Raguso, 2016). A similar driving force could well be responsible for the chemical diversity of indole alkaloids observed in the Apocynaceae.

Since the discovery of plant secondary metabolites, many hypotheses have been proposed to explain why plants would produce such compounds. An early argument that these secondary metabolites were the waste products of the plant has fallen out of favour. Instead the prevailing current theory is that the plants have evolved the metabolites as a defence against predators (Zenk and Juenger, 2007). However, the work presented in this thesis suggests an additional, more nuanced explanation. I hypothesise that the heteroyohimbines initially appeared as the product of an overflow detoxification mechanism. Whether the heteroyohimbines fulfil an additional biological role in the plant remains unknown: the role of the heteroyohimbines in the plant (signalling, storage, defensive compound etc.) is not well understood and so conclusions cannot be drawn on this point. My hypothesis however does not explain why numerous HYSs are present in the plant, and why an enzyme has evolved (HYS) to produce a variety of heteroyohimbine stereoisomers. Explaining the enormous chemical diversity that is present in many plants remains a challenge.

It is not known whether evolution favours a secondary metabolism that is promiscuous or selective. What we can analyse at any point of time is a product of the life history of the species under study; the species, and the individual, is the result of a long line of successful ancestors. Unfavourable alleles are usually lost from the population. Therefore, the collection of genes and alleles that we discover in a modern organism must not carry a significant cost to them. To carry the argument to its logical conclusion, I argue that if there is no inherent detrimental effect for the plant to possess a diverse secondary metabolism, then that is what it will tend to naturally. Mutations will occur over its evolution which will increase chemical diversity. If the compounds produced do not promote a disadvantage to the host there will be no negative selection pressure to stop producing them. Inversely, if the new chemical diversity increases the fitness of the host then the new genes or alleles will increase in frequency in the population.

6.6 References

CHU, H. Y., WEGEL, E. & OSBOURN, A. 2011. From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. The Plant Journal, 66, 66-79.

DOBZHANSKY, T. 2013. Nothing in biology makes sense except in the light of evolution. The american biology teacher, 75, 87-91.

FRELIN, O., HUANG, L. L., HASNAIN, G., JEFFRYES, J. G., ZIEMAK, M. J., ROCCA, J. R., WANG, B., RICE, J., ROJE, S., YURGEL, S. N., GREGORY, J. F., EDISON, A. S., HENRY, C. S., DE CRECY-LAGARD, V. & HANSON, A. D. 2015. A directed-overflow and damage-control N-glycosidase in riboflavin biosynthesis. Biochemical Journal, 466, 137-145.

GEISSLER, M., BURGHARD, M., VOLK, J., STANIEK, A. & WARZECHA, H. 2015. A novel cinnamyl alcohol dehydrogenase (CAD)-like reductase contributes to the structural diversity of monoterpenoid indole alkaloids in Rauvolfia. Planta, 1-12.

GÓNGORA-CASTILLO, E., CHILDS, K. L., FEDEWA, G., HAMILTON, J. P., LISCOMBE, D. K., MAGALLANES-LUNDBACK, M., MANDADI, K. K., NIMS, E., RUNGUPHAN, W., VAILLANCOURT, B., VARBANOVA-HERDE, M., DELLAPENNA, D., MCKNIGHT, T. D., O'CONNOR, S. & BUELL, C. R. 2012. Development of Transcriptomic Resources for Interrogating the Biosynthesis of Monoterpene Indole Alkaloids in Medicinal Plant Species. PLoS ONE, 7, e52506.

PICHERSKY, E. & RAGUSO, R. A. 2016. Why do plants produce so many terpenoid compounds? New Phytologist, n/a-n/a.

QU, Y., EASSON, M. L. A. E., FROESE, J., SIMIONESCU, R., HUDLICKY, T. & DE LUCA, V. 2015. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. Proceedings of the National Academy of Sciences, 112, 6224-6229.

STÖCKIGT, J., TREIMER, J. & ZENK, M. H. 1976. Synthesis of ajmalicine and related indole alkaloids by cell free extracts of *Catharanthus roseus* cell suspension cultures. FEBS Letters, 70, 267-270.

SUDŽUKOVIĆ, N., SCHINNERL, J. & BRECKER, L. 2016. Phytochemical meanings of tetrahydro-β-carboline moiety in strictosidine derivatives. Bioorganic & Medicinal Chemistry, 24, 588-595.

ZENK, M. H. & JUENGER, M. 2007. Evolution and current status of the phytochemistry of nitrogenous compounds. Phytochemistry, 68, 2757-2772.

Appendix 1

1.1 Gene and protein sequences

THAS1 nucleotide sequence:

ATGGCAATGGCTTCAAAGTCACCTTCTGAAGAAGTATATCCAGTGAAGGCATTTGGTTTGGCTGCTAAG-
GATTCTTCTGGGCTTTTCTCTCCATTCAACTTCTCAAGAAGGGCCACAGGGGAACACGATGTGCAGCT-
CAAAGTATTATACTGTGGGACTTGCCAATATGACAGGGAAATGAGCAAAAACAAATTTGGATTTA-
CAAGCTATCCTTATGTTTTAGGGCATGAAATTGTGGGTGAGGTAACTGAAGTTGGCAGCAAGGTG-
CAGAAATTCAAAGTCGGGGACAAAGTGGGCGTAGCAAGCATAATTGAAACTTGTGGCAAATGTGAAAT-
GTGTACAAATGAAGTTGAAAATTACTGTCCAGAAGCAGGATCAATAGACAGCAATTACGGGGCAT-
GTTCAAATATAGCAGTGATAAACGAGAATTTTGTCATCCGTTGGCCTGAAAATCTTCCTTTGGATTCTG-
GTGTTCCTCTTCTATGTGCAGGAATCACGGCTTATAGTCCCATGAAACGTTATGGACTTGATAAACCTG-
GAAAACGTATCGGCATAGCCGGTCTAGGAGGACTTGGACATGTAGCTCTTAGATTTGCCAAAGCTTTTG-
GGGCTAAGGTGACAGTGATTAGTTCTTCACTTAAGAAAAAACGTGAAGCCTTTGAGAAATTCGGAG-
CAGATTCTTTCTTGGTCAGCAGTAATCCAGAAGAAATGCAGGGTGCAGCAGGAACATTGGATGGGAT-
CATAGACACTATACCAGGGAATCACTCTCTTGAGCCACTCCTTGCTTTATTGAAGCCTCTTGGGAAGCT-
TATCATTTTAGGTGCACCAGAAATGCCCTTTGAGGTTCCCGCTCCTTCCCTGCTTATGGGTGGAAAAGTA-
ATGGCTGCCAGTACTGCTGGGAGTATGAAGGAAATACAAGAGATGATTGAATTTGCAGCAGAACACAA-
CATAGTAGCAGATGTGGAGGTTATCTCTATTGACTATGTGAACACTGCAATGGAGCGCCTTGATAACTCT-
GATGTGAGATATCGTTTCGTGATTGATATAGGGAACACTCTGAAATCAAATTAA

THAS1 protein sequence:

MAMASKSPSEEVYPVKAFGLAAKDSSGLFSPFNFSRRATGEHDVQLKVLYCGTCQYDREMSKNKFGFTSY-
PYVLGHEIVGEVTEVGSKVQKFKVGDKVGVASIIETCGKCEMCTNEVENYCPEAGSIDSNYGACSNIAVIN-
ENFVIRWPENLPLDSGVPLLCAGITAYSPMKRYGLDKPGKRIGIAGLGGLGHVALRFAKAFGAKVTVISSSLK-
KKREAFEKFGADSFLVSSNPEEMQGAAGTLDGIIDTIPGNHSLEPLLALLKPLGKLIILGAPEMPFEVPAPSLL-
MGGKVMAASTAGSMKEIQEMIEFAAEHNIVADVEVISIDYVNTAMERLDNSDVRYRFVIDIGNTLKSN

THAS2 coding sequence:

ATGTCTTCAAAATCAGCAAAACCAGTGGAGGCTTACGGATGGGCAGCAAAAGATACATCTGG-
GCTTCTCTCCCCCTTCAAGTTCTTAAGAAGGACCACAGGAGAACATGATGTGCAGTTCAAAG-
TATTGTATTGTGGGCTGTGCGATTGGGATGTAATTACAACCAAGAATACTTATGGCACTACTAAG-
TATCCTTTTGTTCCTGGGCATGAGATAATGGGTATTGTAACAGAGATTGGTAATAAAGTGAAGAAAT-
TCAAGGTTGGAGACAAAGTAGGTGTGGGAAACTTTATTGGATCATGTGGTAAATGTGAGAGATGTA-
ATGAAGGTCTTGAACCTTATTGTCCAAAAGTCATTTACACAGATGGAACTGCTTTTAGCGATGAAAATA-
ATACTGTCTATGGCGACGTTTCTGGAGACGGAGAAGACAGAATATACGGTGGATATTCAAATATTATG-
GTCGCCAATGAGTATGTAGTGTTCCGTTGGCCTGAAAACCTTCCTCTAGCTGCCGGTGTACCTATTTTAT-
GTGGTGGTATTGTTCCTTACAGTCCCATGAGACACTTTGGACTTGATAAACCTGGATTGAGTATTGGT-
GTGGTTGGATTTGGTCGTATTGGGAAATTAGCCGTTAAATTTGCGAAGGCTTTTGGAGCAAATGTTACTG-
TAATTAGTACATCTATTAGCAAGAAGCAAGAAGCTATTGAAAAATATGGGGTAGATAGATTCTTGATCAG-
CAAAGAACCAGAGGAGATGAAGGCGGCAGAGAGTACGCTGGATGGGATCTTTGACTGTGTCCCTAGT-
GTTCATCCTCTTCATCCATTGCTCAATTTGTTGAAGTTTGAAGGGACGTTTGTTATGCTTGGGGTGGCTGT-
GGAGGCATATGAATTGCCAGTGTCTCCACTCCTTATGGGGAGGAGAAAGTTCGTTGGCAGTATAAGCGG-
GACTATGAAGGAAACACAAGAGATGTTAGATTTTGCAGCAAAGCACAATATAGTTTCAGATATAGAGCT-
GATTCCAATGGATTATGTAAATACAGCACTGGAGCGGATTGCCAAGGGCAATCATAAAGATGCATTCGT-
CATTGACATAGAAAACACATTGAAATCTGCTTAG

THAS2 protein sequence:

MSSKSAKPVEAYGWAAKDTSGLLSPFKFLRRTTGEHDVQFKVLYCGLCDWDVITTKNTYGTTKYPFVP-

GHEIMGIVTEIGNKVKKFKVGDKVGVGNFIGSCGKCERCNEGLEPYCPKVIYTDGTAFSDENNTVYGDVS-
GDGEDRIYGGYSNIMVANEYVVFRWPENLPLAAGVPILCGGIVPYSPMRHFGLDKPGLSIGVVGFGRIGK-
LAVKFAKAFGANVTVISTSISKKQEAIEKYGVDRFLISKEPEEMKAAESTLDGIFDCVPSVHPLHPLLNLLK-
FEGTFVMLGVAVEAYELPVSPLLMGRRKFVGSISGTMKETQEMLDFAAKHNIVSDIELIPMDYVNTALERIAK-
GNHKDAFVIDIENTLKSA

THAS3 coding seuquence:

TTGGCAGTTCCATCGGCAGAAACAGGGAAGACAATCGAGGCCTATGGATGGGCAGCCAGAGACT-
CATCTGGGCTTCTCTCTCCCTTCAAGTTCCAGAGAAGGGCTACAACGGAGCATGATGTCCAGCT-
CAAAATATTGTATTGTGGGATGTGCGATTGGGATCTACATGTAGTCAAGAATTGGTTTGGCACCACCAAC-
TATCCCATTGTACCTGGGCACGAGGCAGTGGGCGTGGTGACTGAAATCGGCAACAAGGTACAGAAAT-
TCAAGATTGGGGACATAGTAGGCGTTAGTACTTACATTCGAACATGTCGGAGCTGCGAGAGATGTA-
AAGAAGGTGAAGACAGTTACTGTCCCAGCTTAATAACAGGAGATGGAACTTCATTTAGTGATGGAAAC-
GACGTATTTTTCTATGATCCAAATGATGATAATACAAAAGAGACAACAAAAACATATGGCTCATATTC-
CAATTTCACAGTTGTGGATGAATATTACGTTATTCGTTGGCCAGAAAACTTTCCTTTGGCTGCTGGAG-
TACCTCTTCTTTGTGCTGGTACAGTTCCTTATAGTCCAATGAGGCACTTTGGATTTGATAAACCTGGAAT-
TCATATTGGTGTGGTTGGATTTGGTGGGATTGGCAAATTAGTTGTTAAATTTGCTAAGGCTTTTGGAGT-
TAAAGTAACAGTGATTAGTACATCCATTGATAAGAAGCATGAAGCTATTCATGAATATGGTGCTCATG-
GATTTTTACTCAGCAAAGAACCTCAGCAGCTTCAGGCTGCTATTAATACTATGGAAGGTATAGTTGA-
TACAGTTCCTAAAGTTCACCCTATTCTTCCATTGATCAAATTGTTGAAATTCGATGGTACCCTTCTTAT-
GCTCGGAGCACCGCCGGAGCCATATGAGTTTCCAATCTCCACATTGCTTATGGGGAGGAAGAGGGTG-
GTGGGAAGTGCTGGAGCGAGCATGAAGGAAACACAAGAAATGATGGATTTTGCAGCGAAGCACAA-
CATAGTTGCAGATGTTGAGATCATTCCAATTGATTATGCAAATACAGCAATAGAGAGAATTGAGAAGG-
GAGATTTCAAGAATCGATTCGTAATTGATATAGAAAATTCTTTGGGATCTGTTTAG

THAS3 protein sequence:

MAVPSAETGKTIEAYGWAARDSSGLLSPFKFQRRATTEHDVQLKILYCGMCDWDLHVVKNWFGTTNYP-
IVPGHEAVGVVTEIGNKVQKFKIGDIVGVSTYIRTCRSCERCKEGEDSYCPSLITGDGTSFSDGNDVFFYD-
PNDDNTKETTKTYGSYSNFTVVDEYYVIRWPENFPLAAGVPLLCAGTVPYSPMRHFGFDKPGIHIGVVG-
FGGIGKLVVKFAKAFGVKVTVISTSIDKKHEAIHEYGAHGFLLSKEPQQLQAAINTMEGIVDTVPKVHPILP-
LIKLLKFDGTLLMLGAPPEPYEFPISTLLMGRKRVVGSAGASMKETQEMMDFAAKHNIVADVEIIPIDYAN-
TAIERIEKGDFKNRFVIDIENSLGSV

THAS4 coding sequence:

ATGGCTGCAAAGTCACCTGAAAATGTATACTCAGTGAAGACTTTTGGCTTCGCTGCCAAGGATTCATCT-
GGGCTTTTCTCCCCATTCAACTTCTCAAGAAGAGCCACAGGGGAAAATGATGTGCAGTTCAAAGTATTG-
TACTGTGGAACATGCTACTATGACTGGGCAATGATCACAAACAAATATGGGATGACTAATTATCCTTTC-
GTTATAGGGCACGAAATTATTGGAGTTGTAACTGAGGTTGGTAGCAAGGTAAAAAAATTTAAAGTG-
GGGGGACAAAGTTGGTGTAGGGGGCCACGTTGGAGCGTGCGAGAAGTGCGAATTGTGTATCAATG-
GAGTTGAAAATAATTGTCCAGAAGCAGAATCAACAGATGGCTTTTCCGGGAAAAATTTTGGTGGAT-
GTTCTAATATAATGGTGGTTAACGAGAAGTATGCAGTAGTTTGGCCTGAAAACCTTCCTTTGCATTCGG-
GTGTTCCTCTTTTGTGTGCAGGAATAACTACTTACAGTCCCTTGAGACGATATGGACTTGATAAACCTG-
GACTCAATATTGGCATAGCTGGTCTAGGAGGACTTGGACATTTAGCAATCAGATTTGCCAAGGCTTTTG-
GTGCTAAGGTCACACTAATCAGTTCATCTGTTAAGAAAAAGCGTGAAGCCCTCGAGAAATTTGGTG-
TAGATTCTTTCTTGCTCAACAGTAACCCCGAGGAAATGCAGGGTGCATATGGAACACTTGATGGCAT-
CATAGACACTATGCCAGTGGCTCACTCCATTGTGCCATTTCTTGCTTTATTGAAGCCCCTTGGTAAGCT-
CATCATTTTAGGGGTACCAGAGGAGCCGTTTGAGGTGCCAGCTCCTGCCCTTCTCATGGGTGGGAAAT-
TAATAGCTGGGAGTGCCGCCGGAAGTATGAAGGAAACACAAGAGATGATTGATTTTGCAGCAAAA-
CACAATATAGTAGCAGATGTTGAGGTTATTCCTATTGACTATTTAAACACCGCCATAGAACGCATTA-
AGAACTCTGATGTCAAGTATCGTTTCGTCATTGACGTAGGAAACACTCTCAAATCCCCATCTTTCTAG

THAS4 protein sequence:

MAAKSPENVYSVKTFGFAAKDSSGLFSPFNFSRRATGENDVQFKVLYCGTCYYDWAMITNKYGMT-
NYPFVIGHEIIGVVTEVGSKVKKFKVGDKVGVGGHVGACEKCELCINGVENNCPEAESTDGFSGKNFG-
GCSNIMVVNEKYAVVWPENLPLHSGVPLLCAGITTYSPLRRYGLDKPGLNIGIAGLGGLGHLAIRFAKAFGA-
KVTLISSSVKKKREALEKFGVDSFLLNSNPEEMQGAYGTLDGIIDTMPVAHSIVPFLALLKPLGKLIILGVPEEP-
FEVPAPALLMGGKLIAGSAAGSMKETQEMIDFAAKHNIVADVEVIPIDYLNTAIERIKNSDVKYRFVIDVGN-
TLKSPSF

HYS coding sequence:

ATGGCTGCAAAGTCACCTGAAAATGTATACCCAGTGAAGACTTTTGGCTTCGCTGCCAAGGATTCATCT-
GGGTTTTTCTCCCCATTCAACTTCTCAAGAAGGGCCACCGGGGAAAATGATGTGCAGTTCAAAG-
TATTATATTGTGGAACATGCAACTATGACTTGGAAATGAGCACAAACAAATTTGGGATGACAAAG-
TATCCTTTTGTTATAGGGCACGAAATTGTTGGAGTTGTAACTGAAATAGGTAGCAAGGTACAAAAATTTA-
AAGTGGGGGACAAAGTTGGTGTAGGGGGCTTCGTTGGAGCATGCGAGAAGTGCGAAATGTGTGTCAAT-
GGGGTTGAAAATAACTGTTCAAAAGTAGAATCAACAGATGGCCATTTTGGAAATAATTTTGGTGGAT-
GTTGTAATATAATGGTGGTTAACGAAAAATATGCGGTTGTTTGGCCTGAAAACCTTCCTTTGCATTCGG-
GTGTTCCTCTTTTGTGTGCAGGAATAACTACTTACAGTCCCTTGAGACGATATGGACTTGATAAACCTG-
GACTCAATATTGGCATAGCTGGTCTAGGAGGACTTGGACATTTAGCAATCAGATTTGCCAAGGCTTTTG-
GTGCTAAGGTCACACTAATCAGTTCATCTGTTAAGAAAAAGCGTGAAGCCCTCGAGAAATTTGGTG-
TAGATTCTTTCTTGCTCAACAGTAACCCCGAGGAAATGCAGGGTGCATATGGAACACTTGATGGCAT-
CATAGACACTATGCCAGTGGCTCACTCCATTGTGCCATTTCTTGCTTTATTGAAGCCCCTTGGTAAGCT-
CATCATTTTAGGGGTACCAGAGGAGCCGTTTGAGGTGCCAGCTCCTGCCCTTCTCATGGGTGGGAAAT-
TAATAGCTGGGAGTGCCGCCGGAAGTATGAAGGAAACACAAGAGATGATTGATTTTGCAGCAAAA-
CACAATATAGTAGCAGATGTTGAGGTTATTCCTATTGACTATTTAAACACCGCCATGGAACGCATTA-
AGAACTCTGATGTCAAGTATCGTTTCGTCATTGACGTAGGAAACACTCTCAAATCCCCATCTTTCTAG

HYS protein sequence:

MAAKSPENVYPVKTFGFAAKDSSGFFSPFNFSRRATGENDVQFKVLYCGTCNYDLEMSTNKFGMTKYP-
FVIGHEIVGVVTEIGSKVQKFKVGDKVGVGGFVGACEKCEMCVNGVENNCSKVESTDGHFGNNFGG-
CCNIMVVNEKYAVVWPENLPLHSGVPLLCAGITTYSPLRRYGLDKPGLNIGIAGLGGLGHLAIRFAKAFGA-
KVTLISSSVKKKREALEKFGVDSFLLNSNPEEMQGAYGTLDGIIDTMPVAHSIVPFLALLKPLGKLIILGVPEEP-
FEVPAPALLMGGKLIAGSAAGSMKETQEMIDFAAKHNIVADVEVIPIDYLNTAMERIKNSDVKYRFVIDVGN-
TLKSPSF

ADH10 coding sequence:

ATGGCCGCCATGGGTACCTCGGAAAAATATGCAGTTGTTACGGGATCAAACAAAGGTATTGGATTT-
GAAACCTGCAAGAAATTGGCTTCTCAAGGGATCACTGTGGTCCTTACTGCTAGAGATGAAAAAAGAG-
GGCTCGATGCTCTTGAGAAGCTCAAAGAATTGGGTCTCTCTGGTAAGGTGCTATTTCATCAGCTTGAT-
GTGACCGATTCATCCAGCGTTGCTTCCCTTGCAGAATTTGTCAAGAAACAATTTGGAAGACTTGA-
TATCTTGGTAAACAATGCAGGGGTTAATGGAGTGATTACTGATGTTGAAGCTGTGAAAAAGCTAAATCCT-
GCAGAAGATCCGGCCGATGTCGACTTTAGCAAGATATACAAGGAAACATATGAGTTGGCTGAAGAAT-
GCATTCAAATTAACTACTTTGGAACAAAAAGAACCACTGATGCACTTCTTCCTCTTCTCCAATTATCTG-
CATCACCAAGAATCGTAAATATTTCCTCCATCATGGGACAGTTAAAGAACATACCAAGTGAATGGGCTA-
AAGGAATCCTGGGAGATGCTAGCAACCTCACAGAAGATAGATTGGATGAGGTGATCAATAACTTCTT-
GAAGGACTTCAAGGAAGGATCCCTTGCAGCTAAGGGATGGCCTCCATCCTTTTCAGCTTATATAGTCT-
CAAAAGTTGTGGTGAATGCCTACACAAGAATTCTGGCCAAGAAGTATCCCAATTTCAAGATCAATTGT-
GTTTGTCCAGGGTTTGCCAAGACAGATTTGAATCATGGCTTAGGCTTATTAACTGCAGAAGAAGCTGCT-
GAAAACCCTGTGAAACTCGCTTTGCTGCCTGATGATGGTCCTTCGGGTTTGTTCTTTGATCGCAGCGAG-
GAGTCATCGTTTGAATGA

ADH 10 amino acid sequence:

MAAMGTSEKYAVVTGSNKGIGFETCKKLASQGITVVLTARDEKRGLDALEKLKELGLSGKVLFHQLD-
VTDSSSVASLAEFVKKQFGRLDILVNNAGVNGVITDVEAVKKLNPAEDPADVDFSKIYKETYELAEECIQINY-
FGTKRTTDALLPLLQLSASPRIVNISSIMGQLKNIPSEWAKGILGDASNLTEDRLDEVINNFLKDFKEGSLAAK-
GWPPSFSAYIVSKVVVNAYTRILAKKYPNFKINCVCPGFAKTDLNHGLGLLTAEEAAENPVKLALLPDDGPS-
GLFFDRSEESSFE

1.2 Phylogenetic data

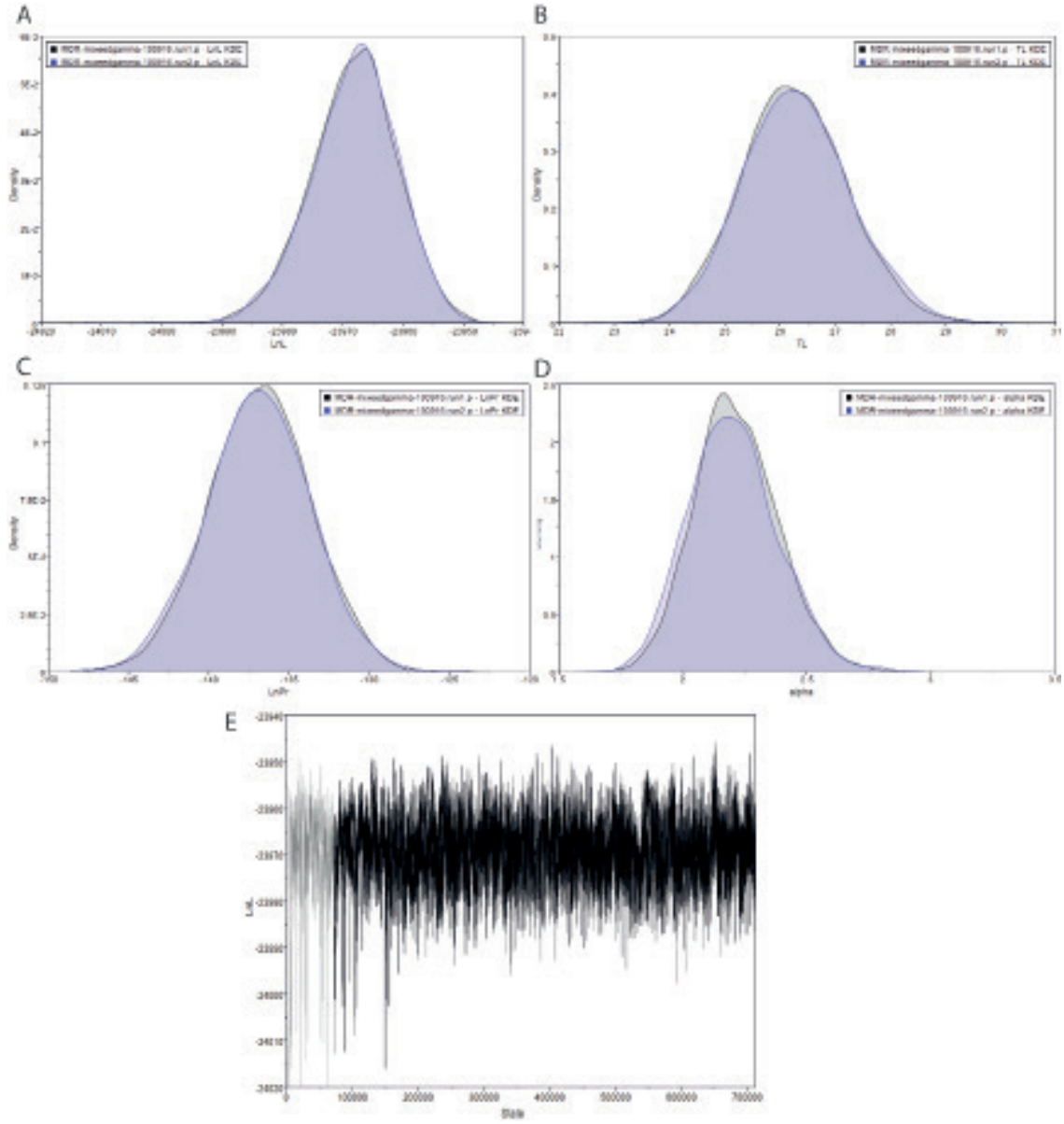


Figure 1: Convergence graphs for MC3 run with “mixed + gamma” model. A: Log Likelihood of model, LnL; B: Tree length, TL; C: Log of the prior density, LnPr; D: estimation of the gamma factor, alpha; E: Mixing graph for parameter LnL.

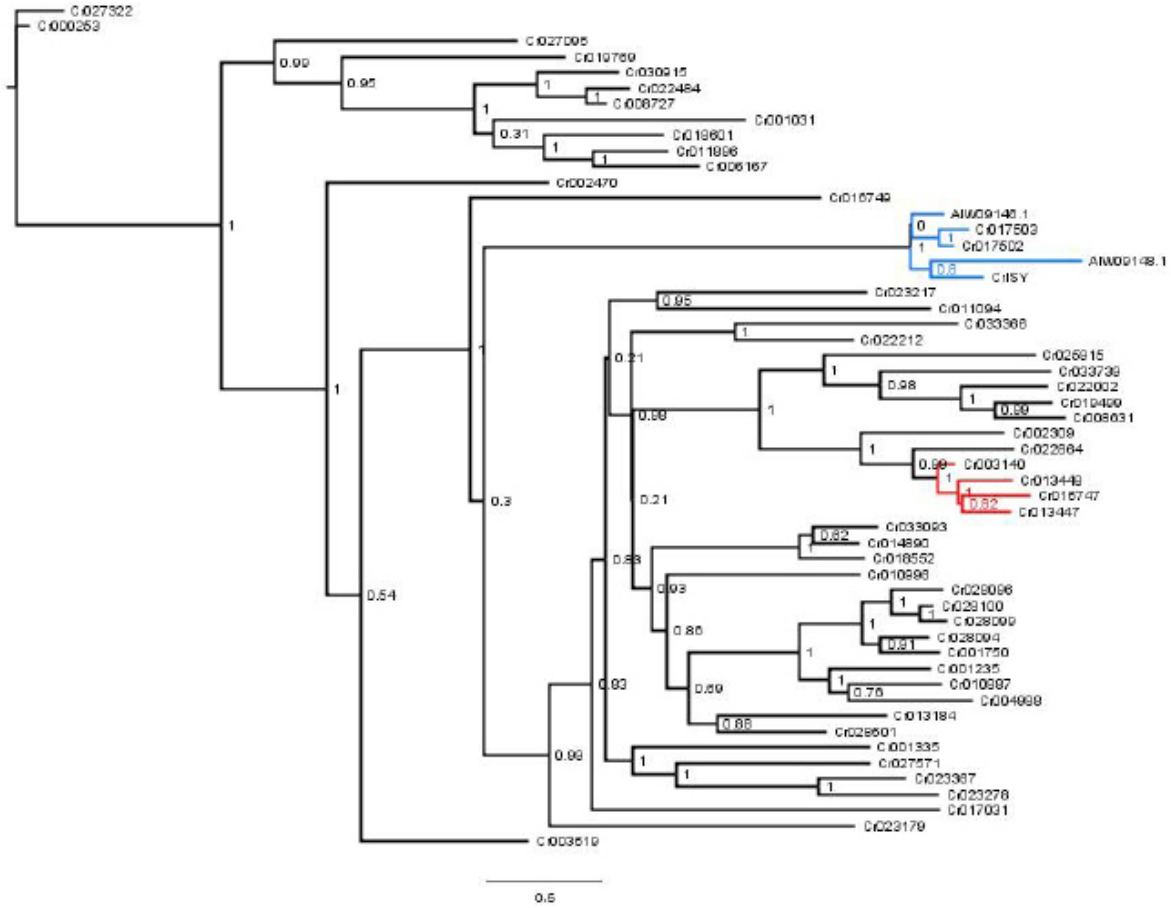


Figure 2: Phylogenetic tree of the SDR superfamily in *C. roseus*. Tree was constructed using PhyML through command-line in an Ubuntu environment; Branch support values (aLRT-based, combination of Chi2s) are printed next to the nodes; ADH10-like clade is illustrated in red; ISY-like clade is in blue. ADH10: Cr013448. The bar at the bottom represents the distance to 0.6 changes per site.

1.3 NMR data

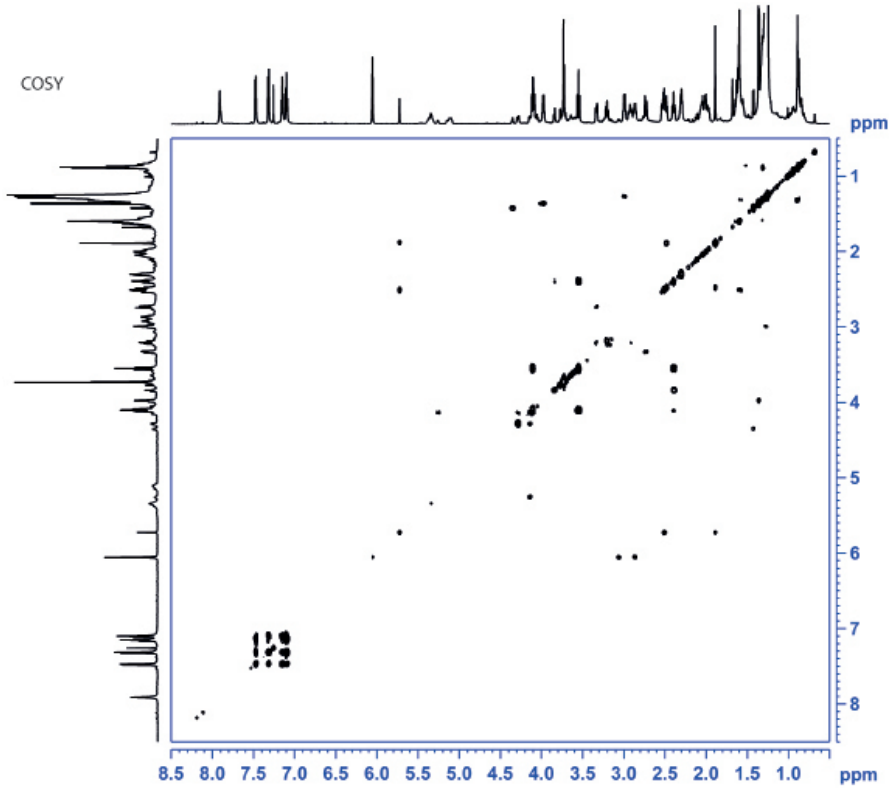


Figure 3: COSY of ADH10 product, vitrosamine.

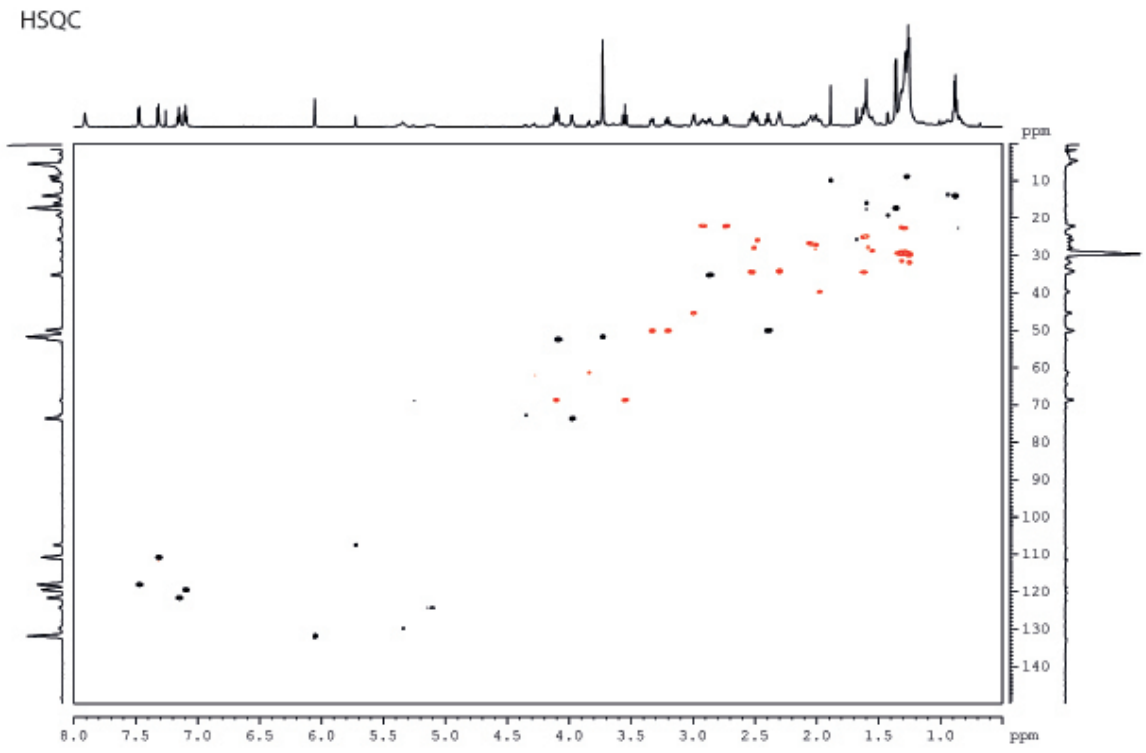


Figure 4: HSQC of ADH10 product, vitrosamine. Red cross-peaks depict to negative projections and correspond to carbons attached to an even number of hydrogens. Black cross-peaks depict positive projections and correspond to carbons attached to an odd number of hydrogens.

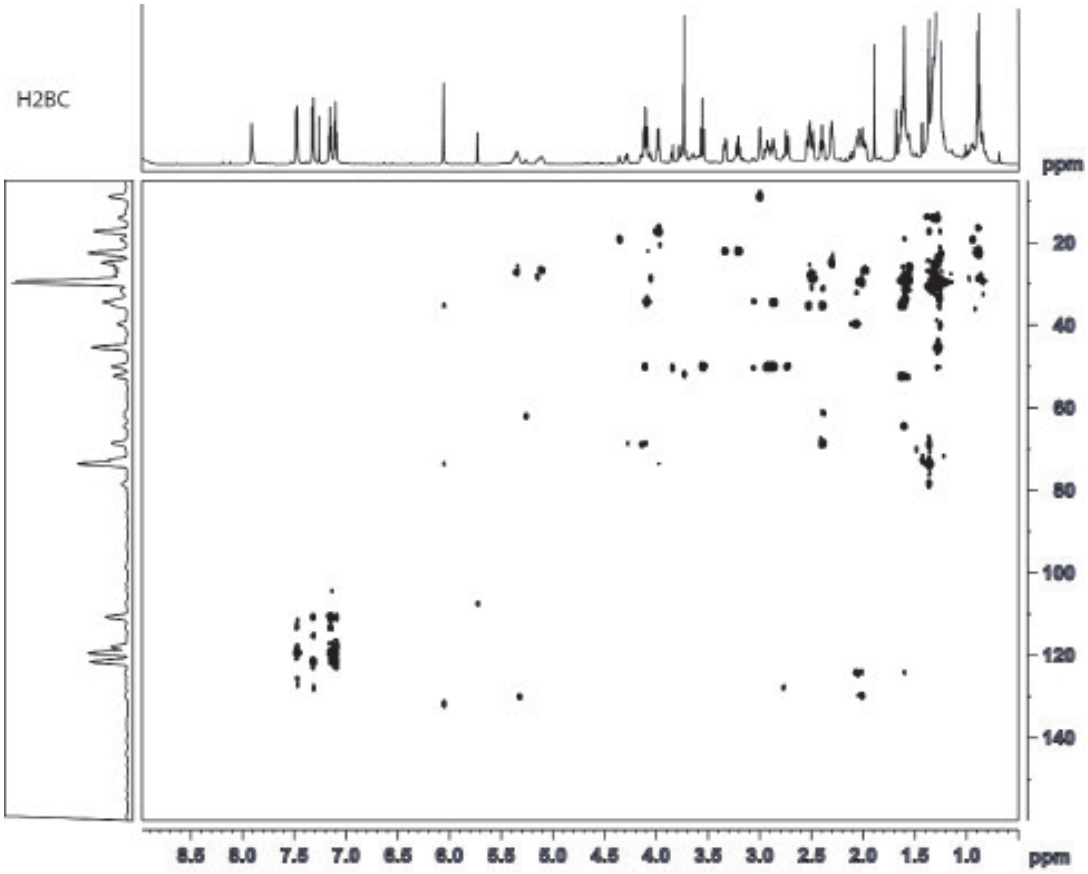


Figure 5: H2BC of ADH10 product, vitrosamine. Cross-peak correspond to proton and carbon spins separated by two covalent bonds.

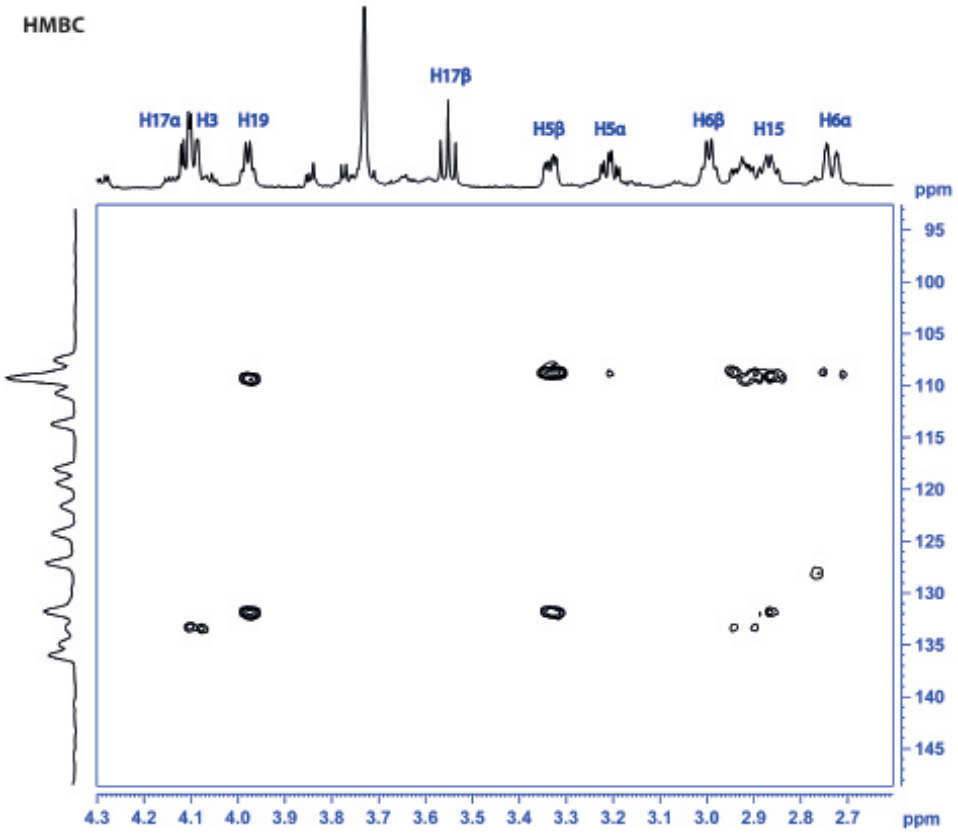


Figure 6: HMBC of aliphatic region of ADH10 product, vitrosamine.

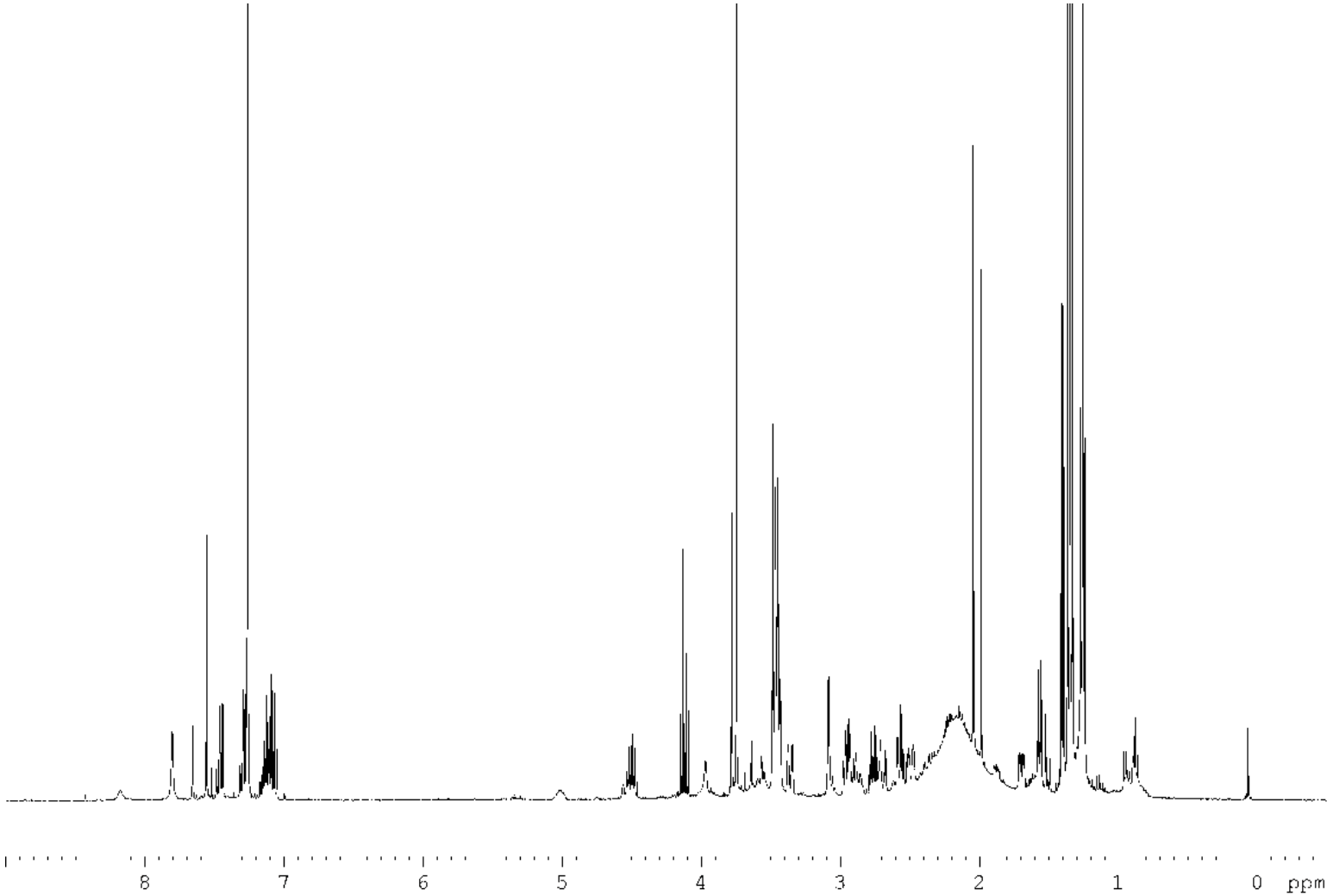


Figure 7: ^{1}H-NMR spectra of [21α-^{2}H] Tetrahydroalstonine. Purified deuterated product of THAS1 large-scale reaction.

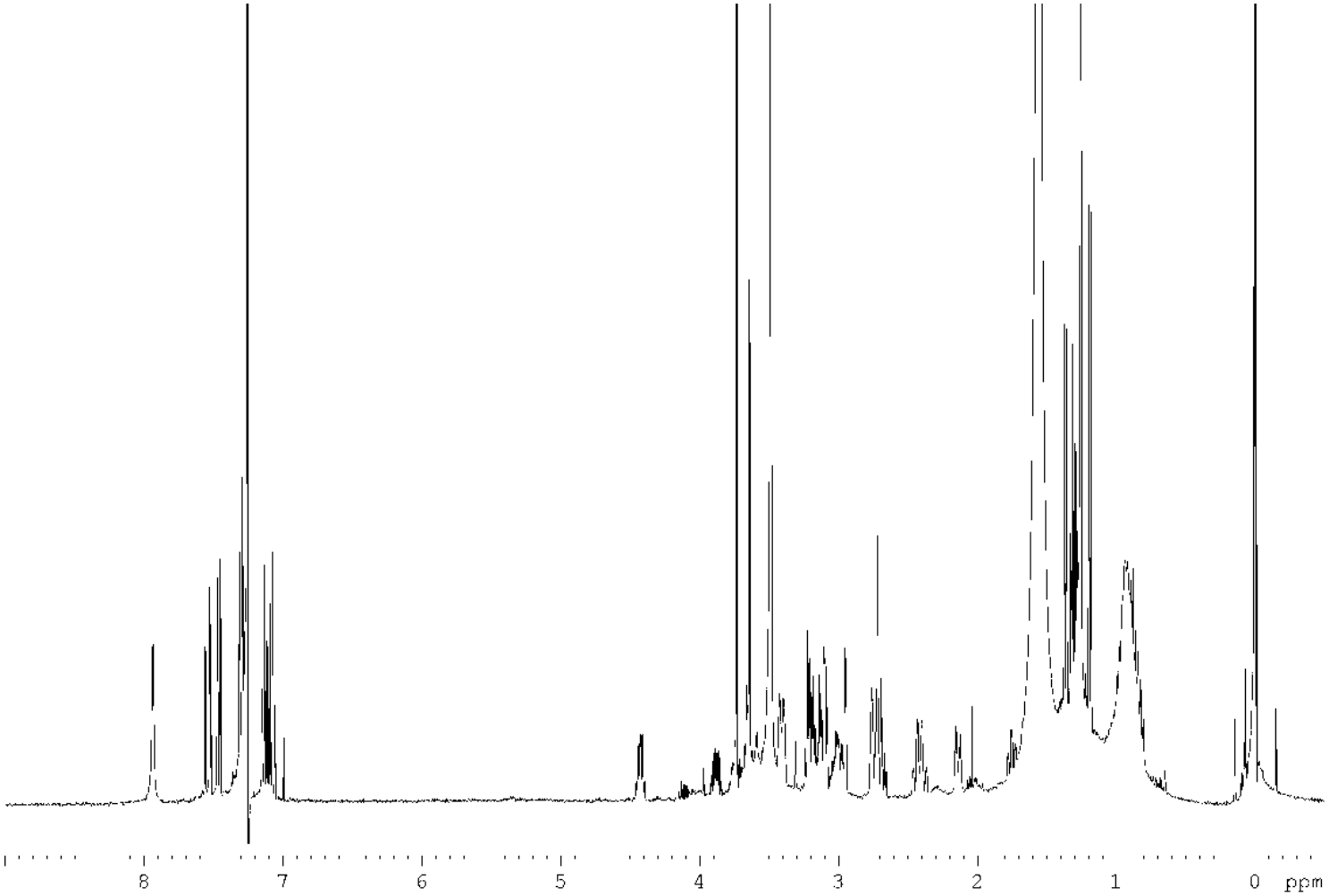


Figure 8: ^{1}H-NMR spectra of [21α-^{2}H] Ajmalicine. Purified deuterated product of HYS large-scale reaction.

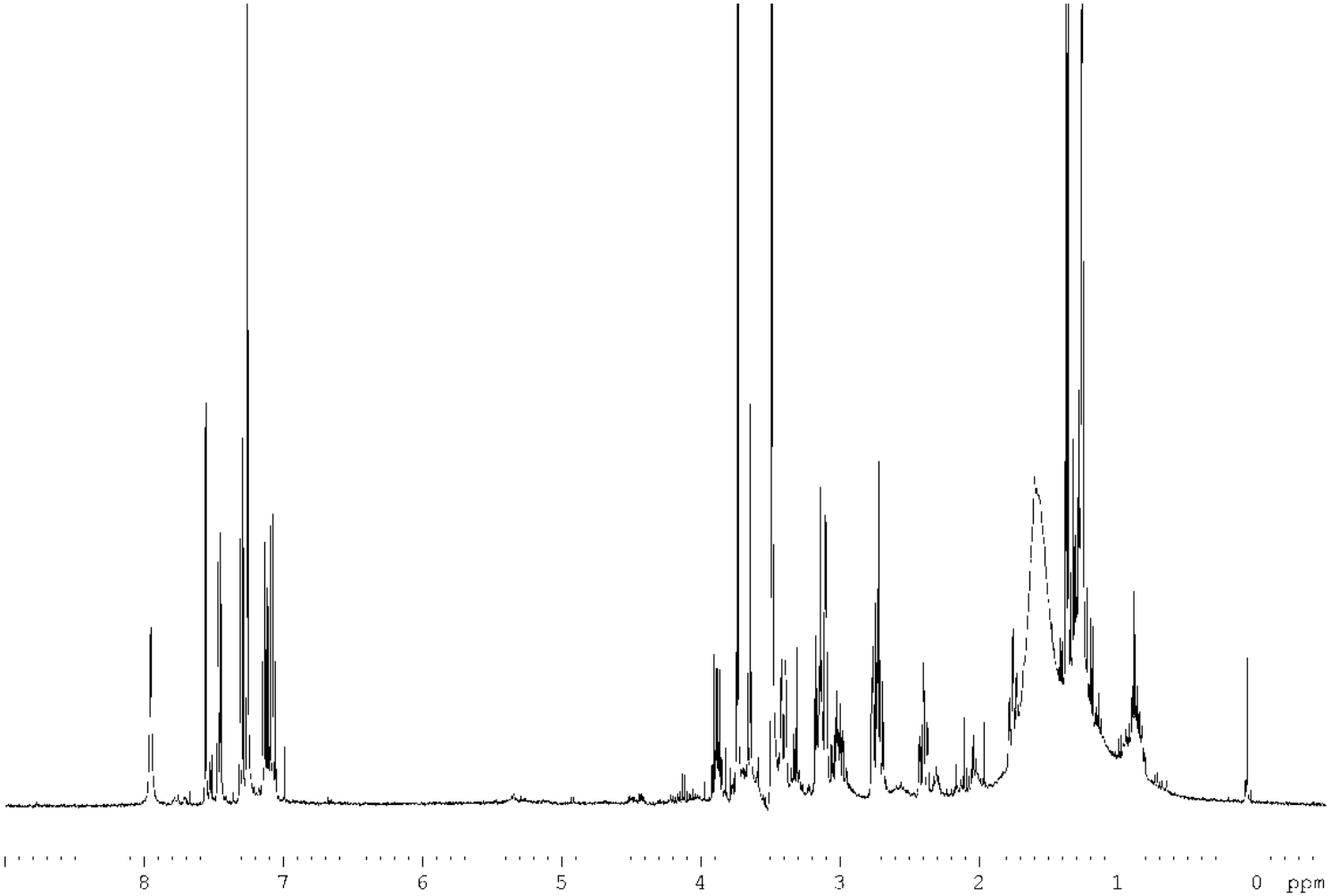


Figure 9: ^{1}H-NMR spectra of [21α-^{2}H] 19-Epiajmalicine. Purified deuterated product of HYS large-scale reaction.

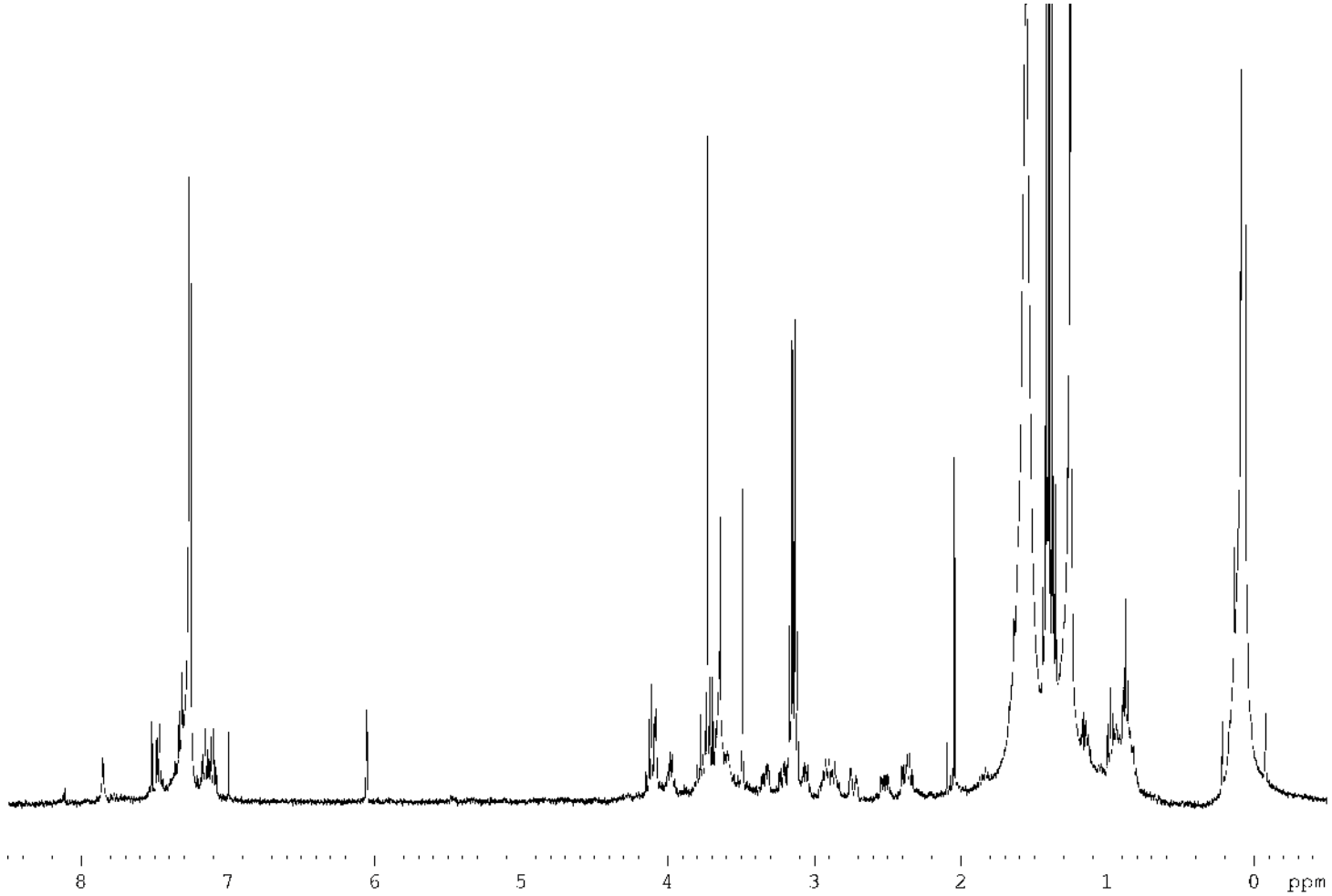


Figure 10: ^{1}H-NMR of [17β-^{2}H] ADH10 product, vitrosamine

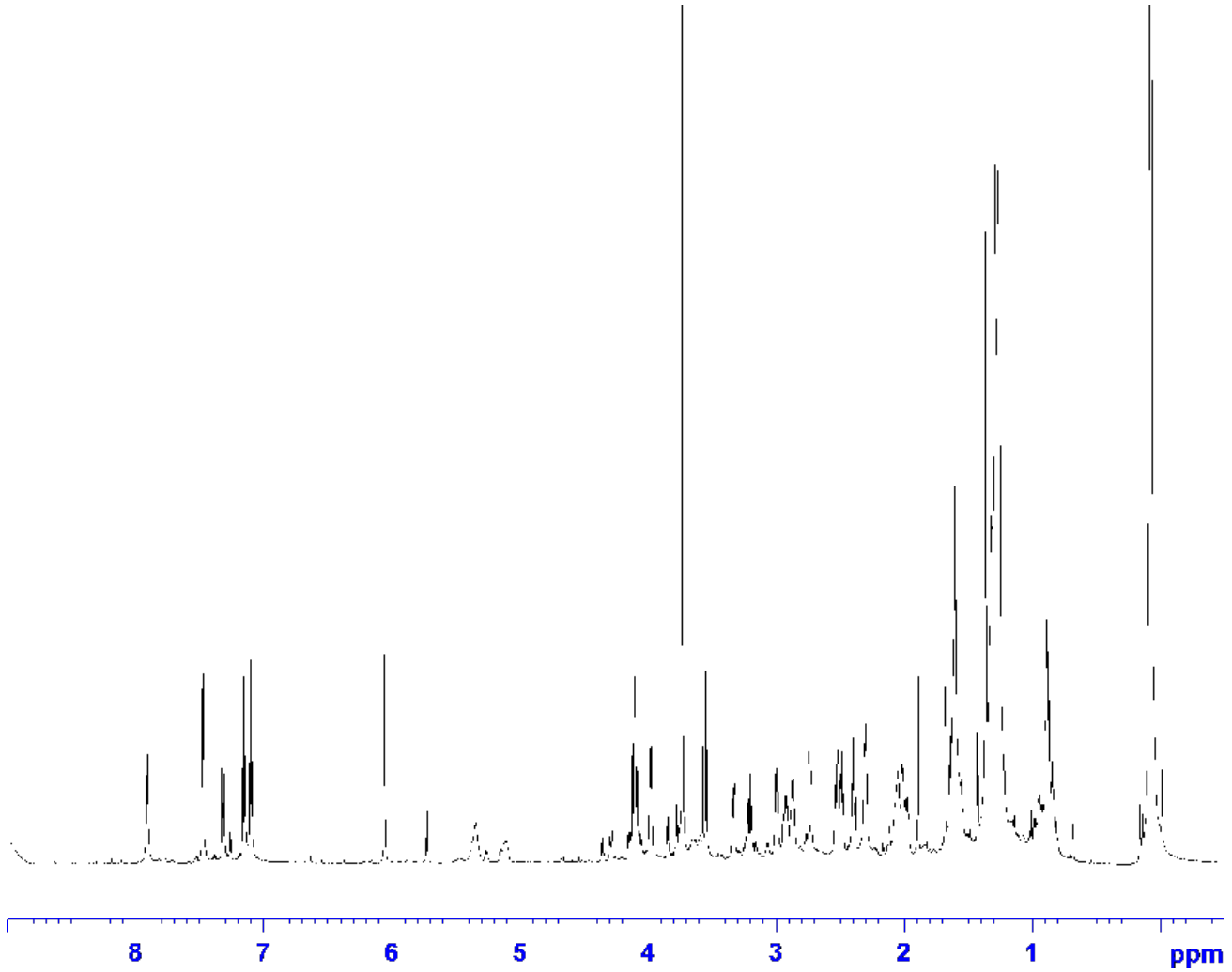


Figure 11: ^{1}H-NMR of ADH10 product, vitrosamine

Appendix 2

Report : Thermal stability assay using buffer screen 1

Sample name: Sampel 1

Date: 13-05-2015

User name : Lorenzo Caputi

Message:

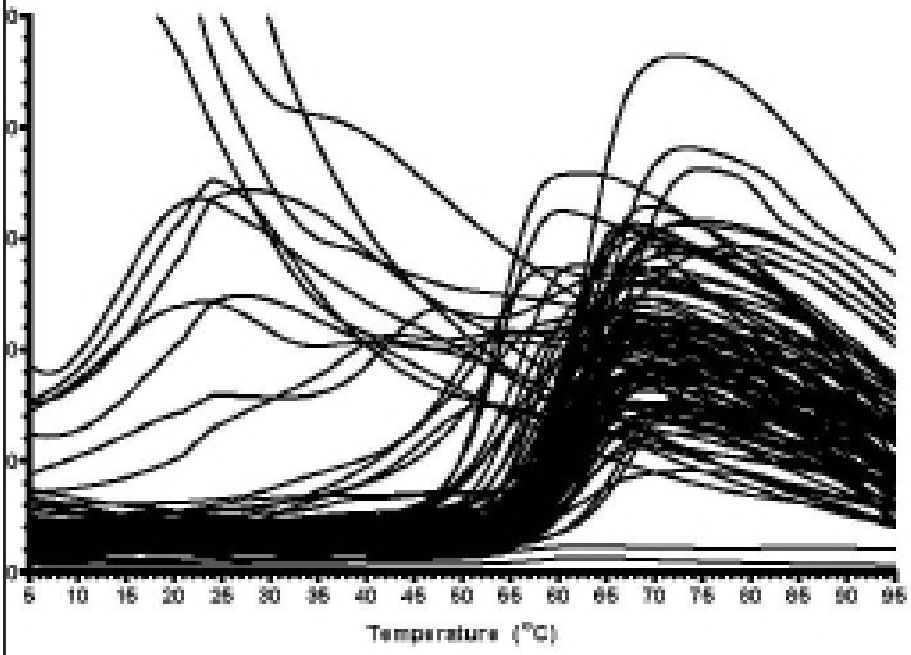
- Sample received at 10 mg/ml (MW ≈ 39 kDa, 256uM) in 20 mM HEPES, 0.5 mM TCEP.
-Sample has been diluted 1:10 in 20mM HEPES pH7.5 to obtain protein around 25uM.
-Loaded 2ul of diluted-sample into each wells
-Plot have been prepared using GraphPad v5.0.

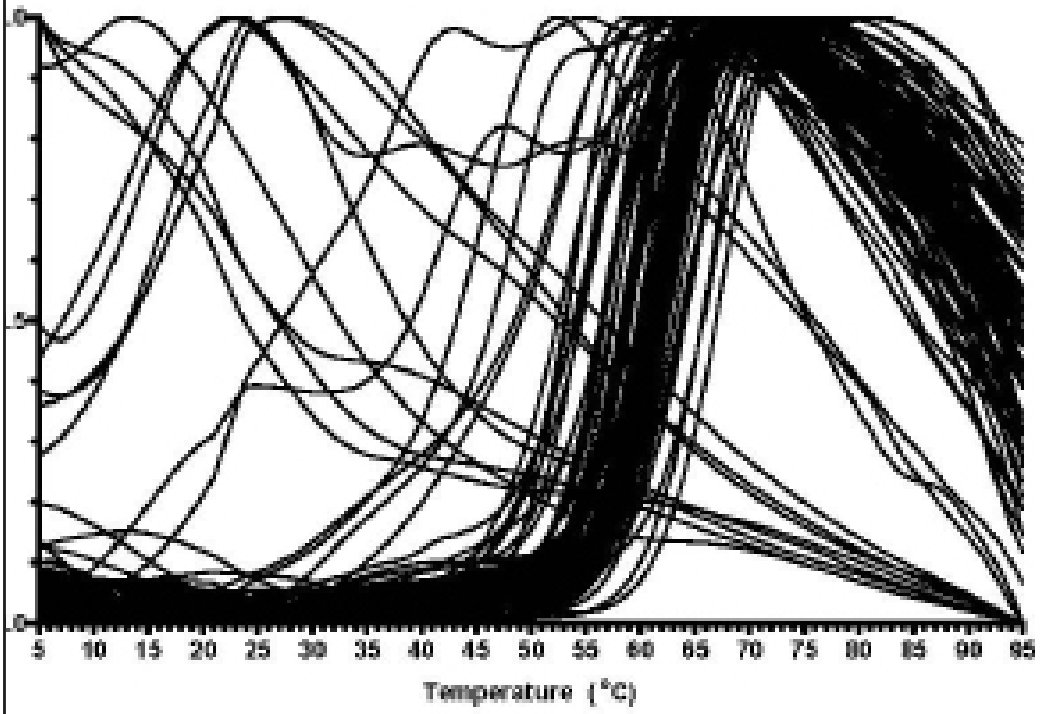
Thanks to cite:

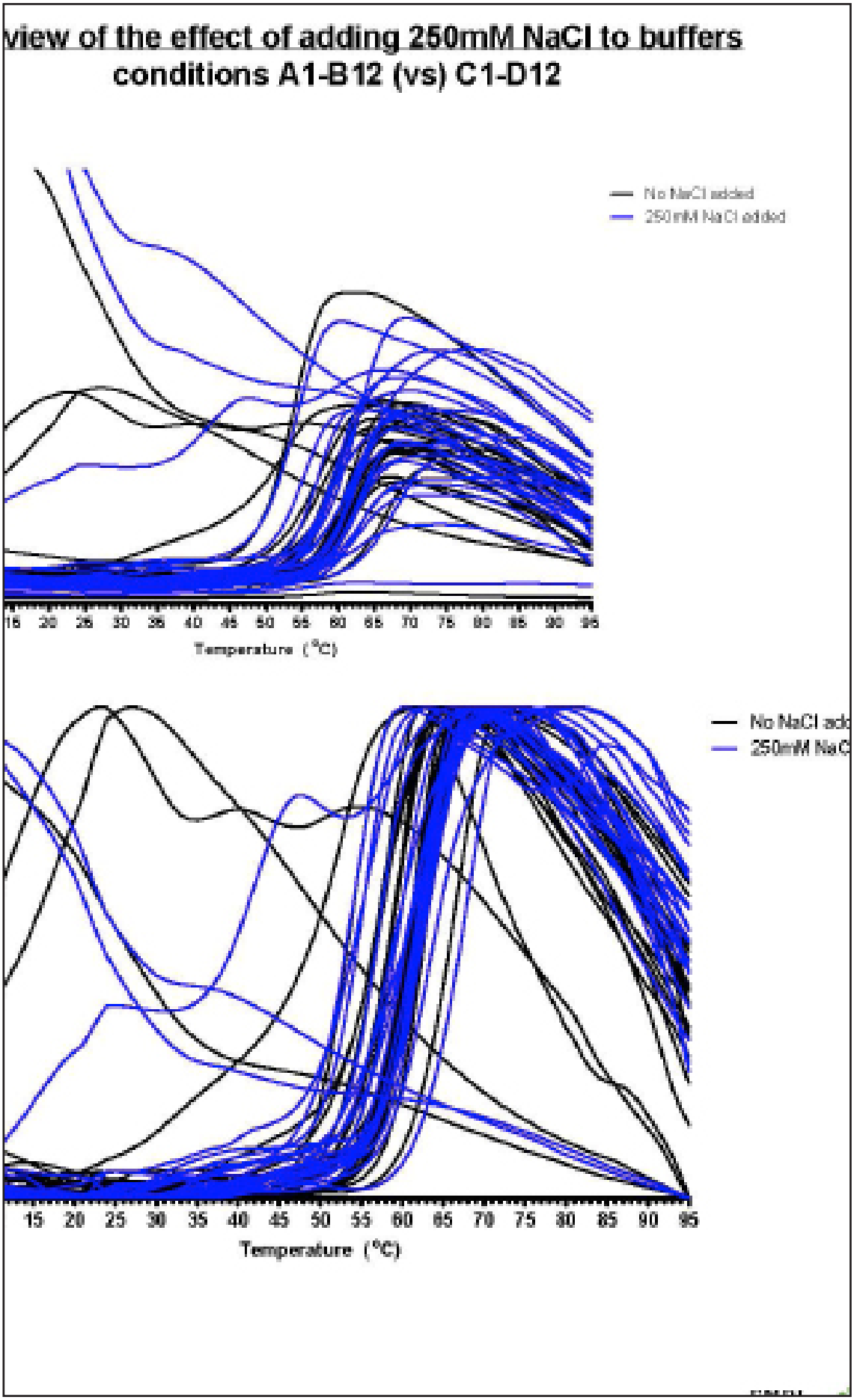
Boivin S, Kozak S, Meijers R. (2013) Optimization of protein purification and characterization using Thermofluor screens. Protein Expr Purif. 2013 Aug 12. doi:pii: S1046-5928(13)00145-9. 10.1016/j.pep.2013.08.002. [Epub ahead of print]

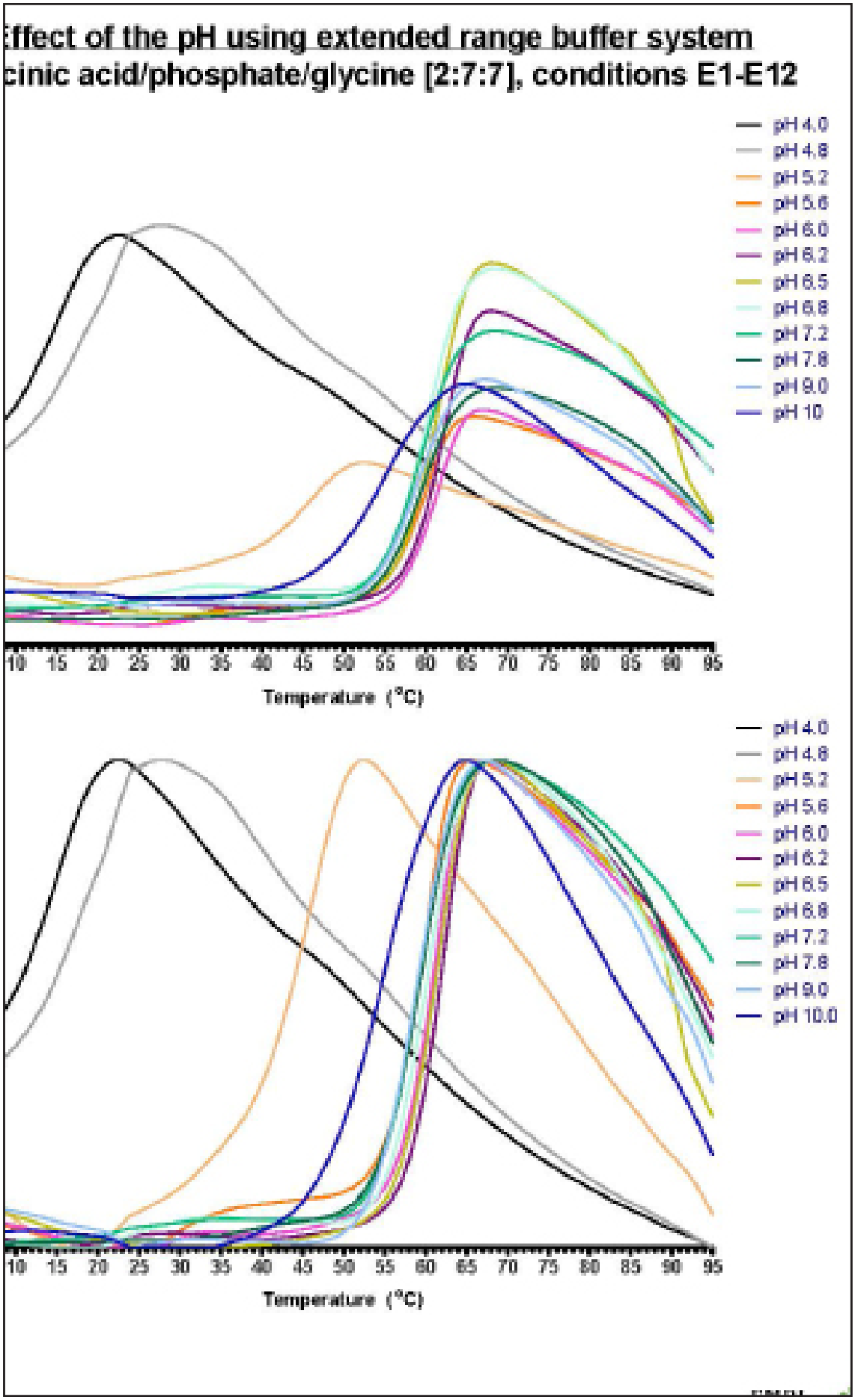
EMBL Hamburg SPC Facility

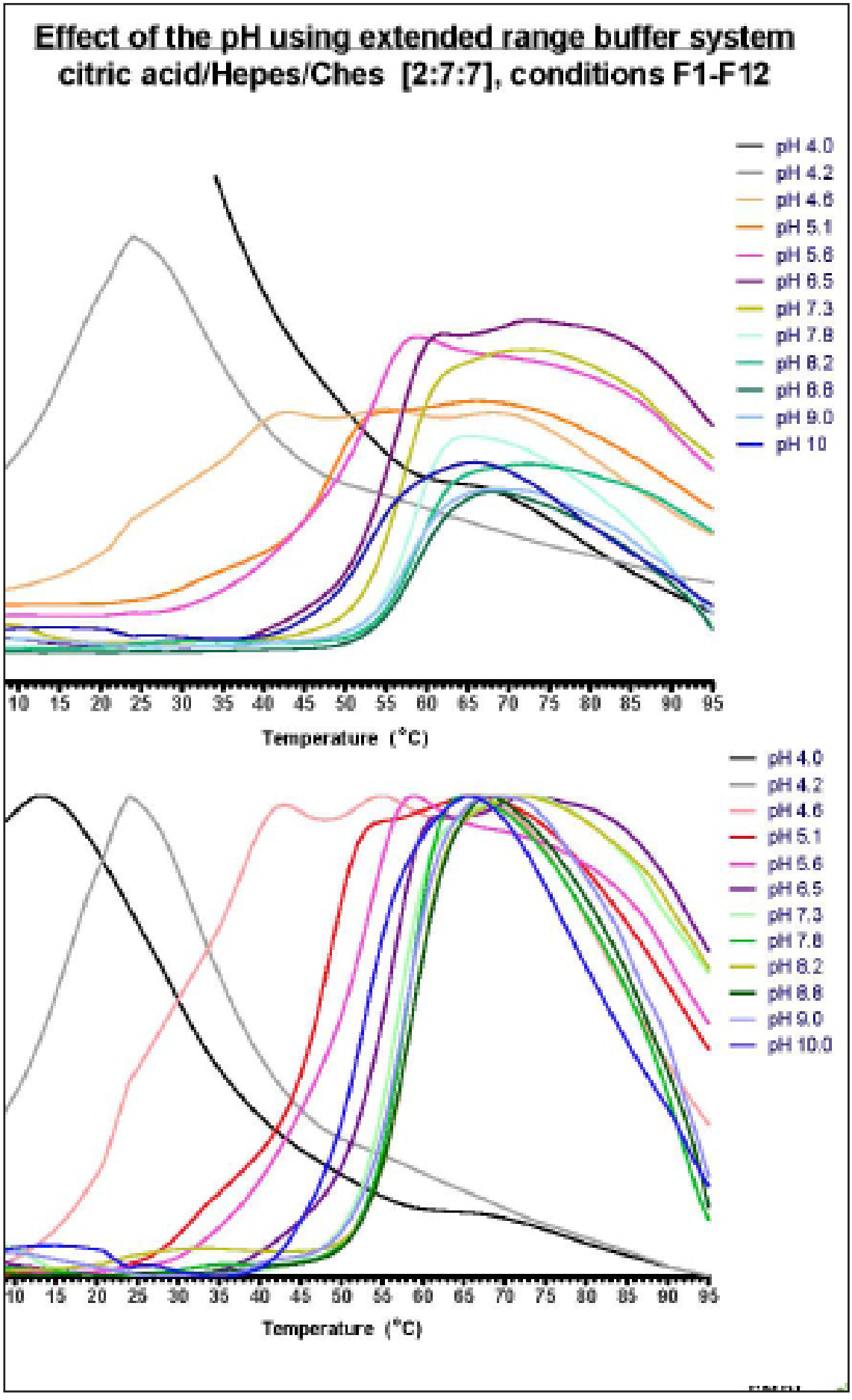
Overview of the 96 conditions
conditions A1 - H12

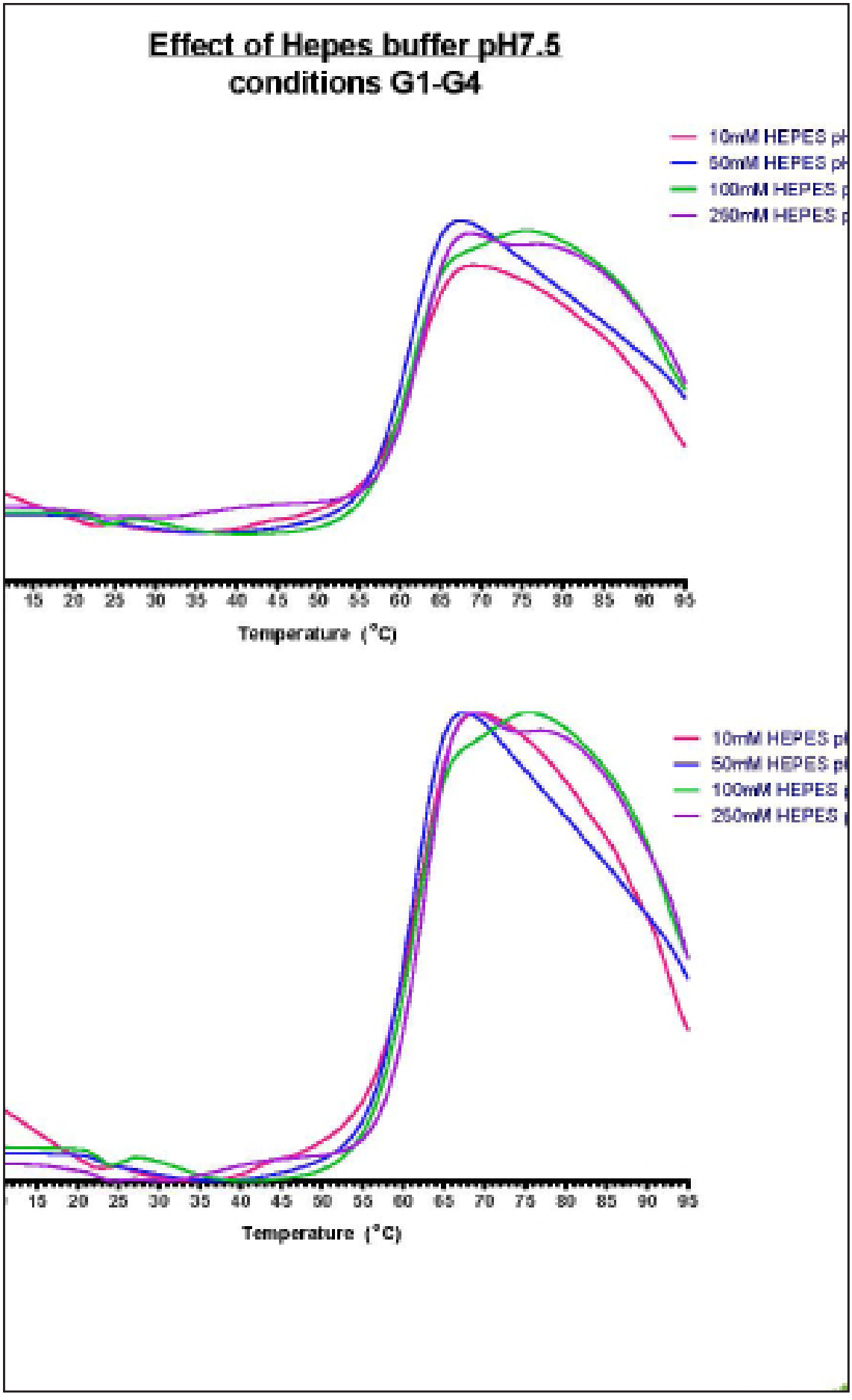


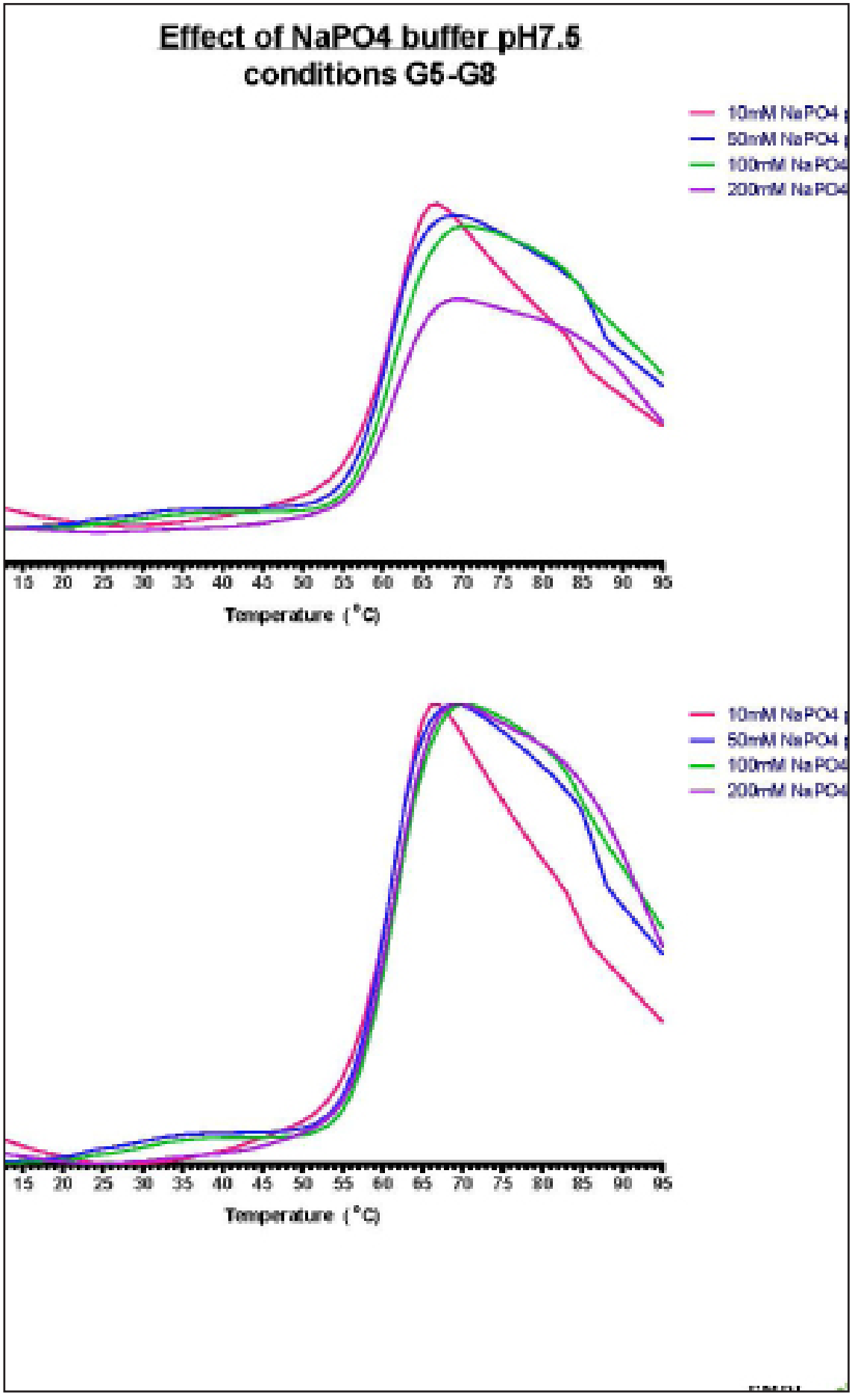


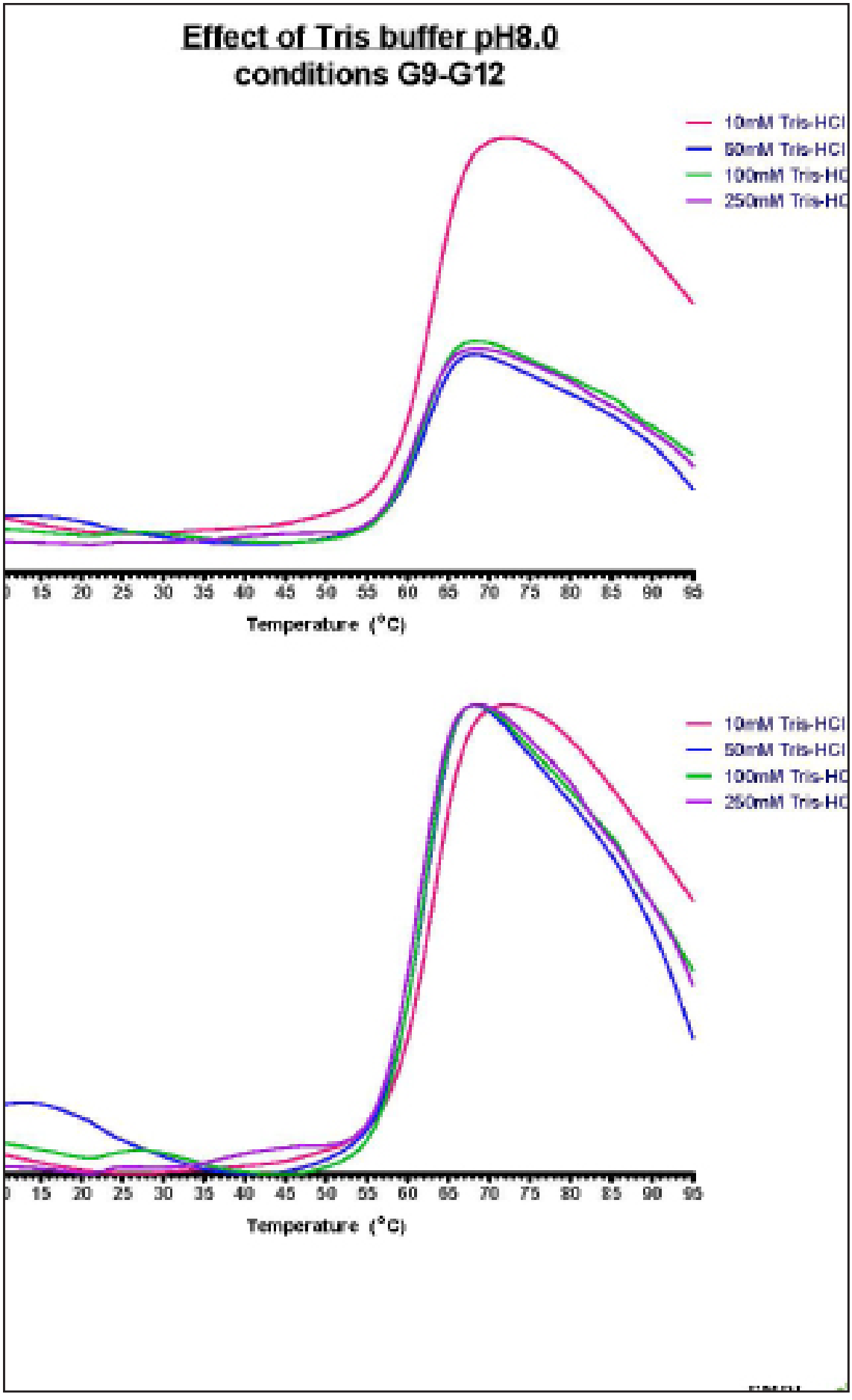

view of the effect of adding 250mM NaCl to buffers
conditions A1-B12 (vs) C1-D12
No NaCl added
250mM NaCl added
15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
No NaCl add
250mM NaC
15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)

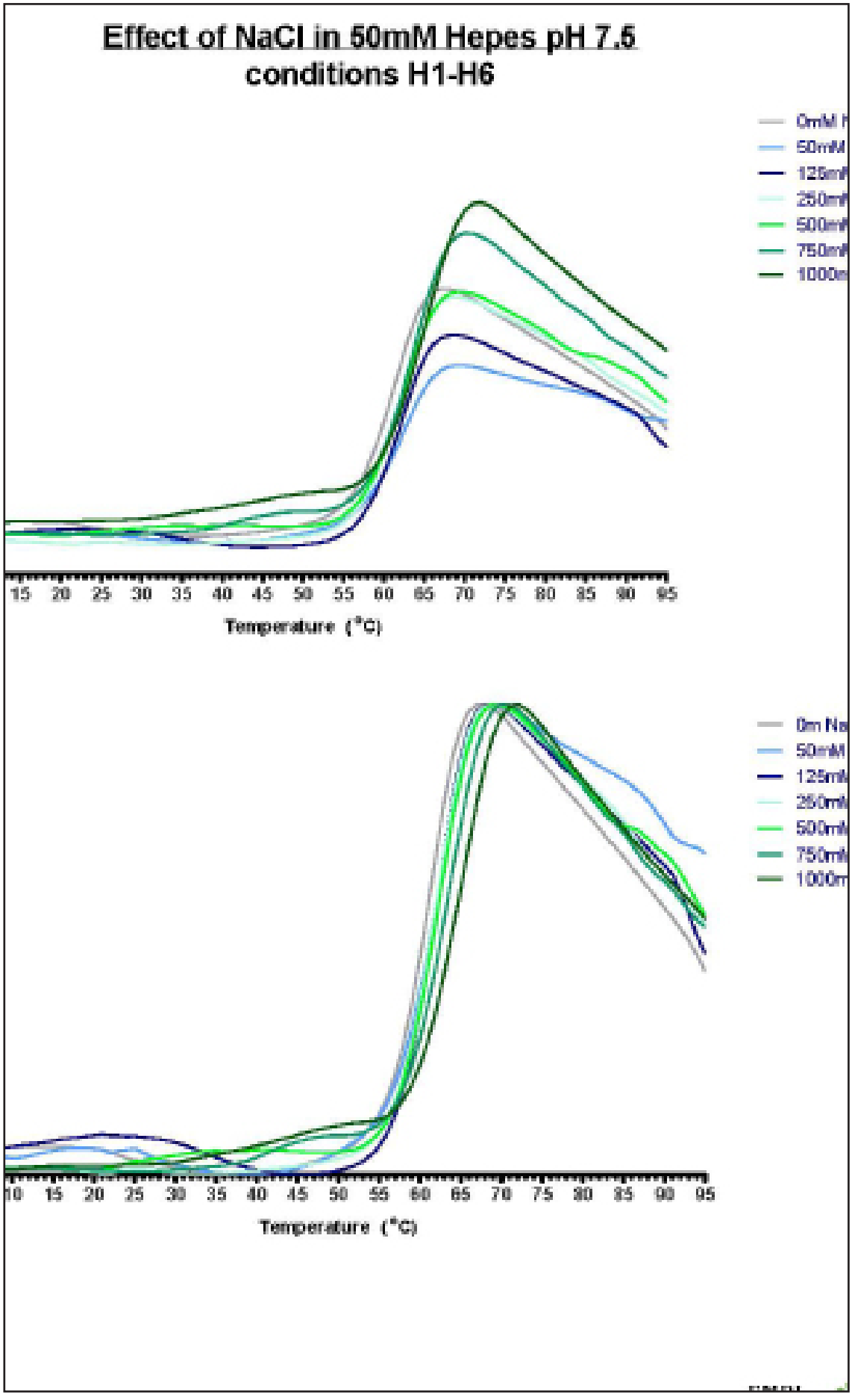

ffect of the pH using extended range buffer system
cinic acid/phosphate/glycine [2:7:7], conditions E1-E12
pH 4.0
pH 4.8
pH 5.2
pH 5.6
pH 6.0
pH 6.2
pH 6.5
pH 6.8
pH 7.2
pH 7.8
pH 9.0
pH 10
10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
pH 4.0
pH 4.8
pH 5.2
pH 5.6
pH 6.0
pH 6.2
pH 6.5
pH 6.8
pH 7.2
pH 7.8
pH 9.0
pH 10.0
10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)

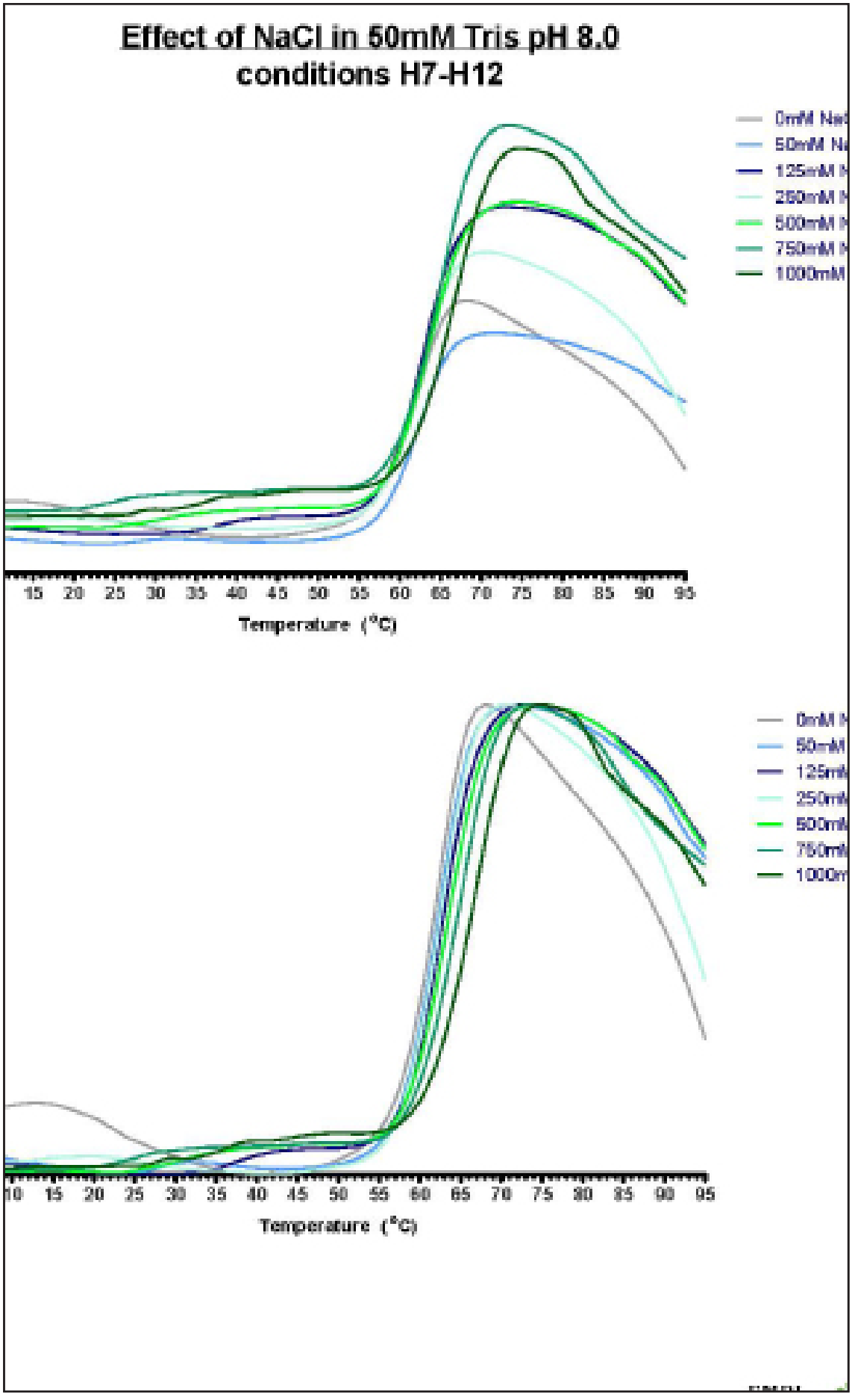

Effect of the pH using extended range buffer system
citric acid/Hepes/Ches [2:7:7], conditions F1-F12
pH 4.0
pH 4.2
pH 4.6
pH 5.1
pH 5.6
pH 6.5
pH 7.3
pH 7.8
pH 8.2
pH 8.6
pH 9.0
pH 10
10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
pH 4.0
pH 4.2
pH 4.6
pH 5.1
pH 5.6
pH 6.5
pH 7.3
pH 7.8
pH 8.2
pH 8.6
pH 9.0
pH 10.0
10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)


Effect of Hepes buffer pH7.5
conditions G1-G4
10mM HEPES pH
50mM HEPES pH
100mM HEPES p
250mM HEPES p
15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
10mM HEPES pH
50mM HEPES pH
100mM HEPES p
250mM HEPES p
15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)


Effect of NaPO4 buffer pH7.5
conditions G5-G8
10mM NaPO4
50mM NaPO4
100mM NaPO4
200mM NaPO4
15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
10mM NaPO4
50mM NaPO4
100mM NaPO4
200mM NaPO4
15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)


Effect of Tris buffer pH8.0
conditions G9-G12
10mM Tris-HCl
50mM Tris-HCl
100mM Tris-HC
250mM Tris-HC
0 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
10mM Tris-HCl
50mM Tris-HCl
100mM Tris-HC
250mM Tris-HC
0 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)


Effect of NaCl in 50mM Hepes pH 7.5
conditions H1-H6
0mM
50mM
125mM
250mM
500mM
750mM
1000m
15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
0m Na
50mM
125mM
250mM
500mM
750mM
1000m
10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)


Effect of NaCl in 50mM Tris pH 8.0
conditions H7-H12
0mM Nac
50mM Na
125mM N
250mM N
500mM N
750mM N
1000mM
15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)
0mM N
50mM
125mM
250mM
500mM
750mM
1000m
10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Temperature (°C)

Appendix 3

Report : Thermal stability assay - custom experiment

Sample name: Sampel 1

Date: 13-05-2015

User name : Lorenzo Caputi

Message:

- Sample received at 10 mg/ml (MW ≈ 39 kDa, 256uM) in 20 mM HEPES, 0.5 mM TCEP.
-Sample has been diluted 1:10 in 20mM HEPES pH7.5 to obtain protein around 25uM.
-Loaded 2ul of diluted-sample into each wells
-Plot have been prepared using GraphPad v5.0.

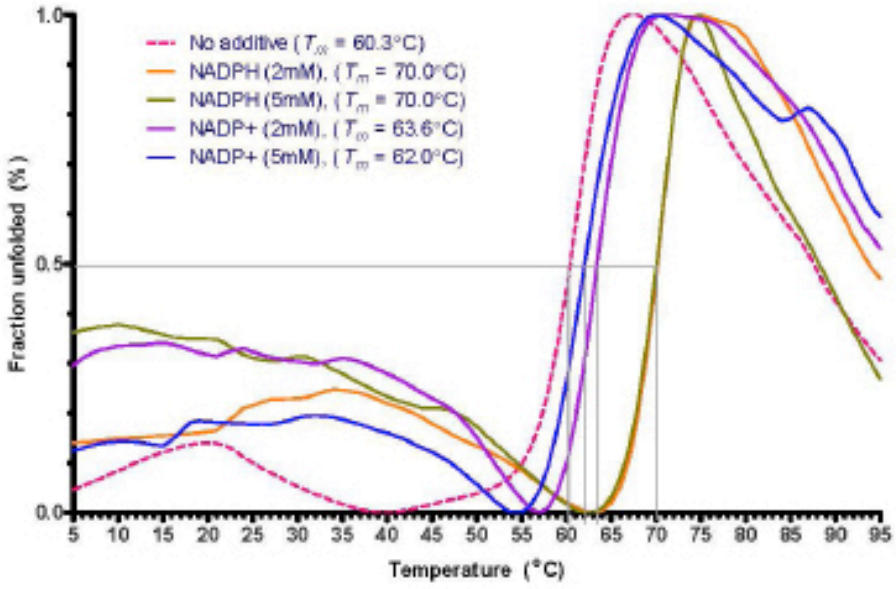
-Sigma N1630 (ß-Nicotinamide adenine dinucleotide 2'-phosphate reduced), MW 833.35
-Sigma N0505 (ß-Nicotinamide adenine dinucleotide phosphate sodium), MW 765.39
-Prepared 10mM stock solution in water
-Experiment carried out in 50mM HEPES pH 7.5 (no salt)
-Filled wells to 25ul with water
-5 minutes incubation before temprature ramping

Thanks to cite:

Boivin S, Kozak S, Meijers R. (2013) Optimization of protein purification and characterization using Thermofluor screens. Protein Expr Purif. 2013 Aug 12. doi:pii: S1046-5928(13)00145-9. 10.1016/j.pep.2013.08.002. [Epub ahead of print]

EMBL Hamburg SPC Facility

Report : Thermal stability assay using buffer screen 1



EMBL Hamburg SPC Facility

Appendix 4

Report : Thermal stability assay using additive screen 2

Sample name: ADH22b

Date: 2015-06-05

User name: Lorenzo Caputi

Message:

- Sample received at 640 uM, diluted with 20 mM HEPES, 100 mM NaCl, pH 7.5 down to 64 uM

- Screen 2 has been performed in 50 mM HEPES, pH 7.5

- 2 µl loaded in each well

- Plot have been prepared using GraphPad v5.0.

Thanks to cite:

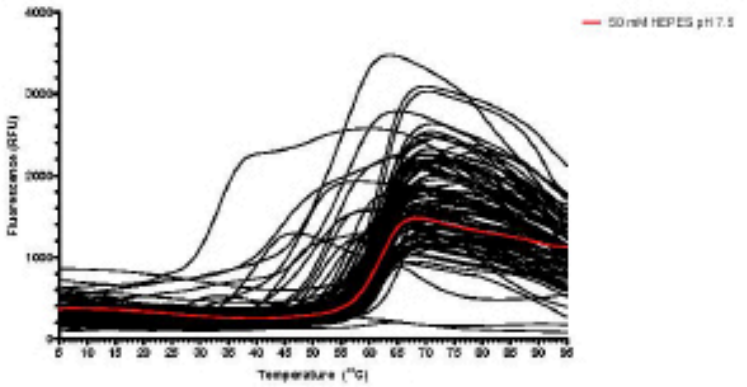
Boivin S, Kozak S, Meijers R. (2013) Optimization of protein purification and characterization using Thermofluor screens. Protein Expr Purif. 2013 Aug 12. doi:pii: S1046-5928(13)00145-9. 10.1016/j.pep.2013.08.002. [Epub ahead of print]

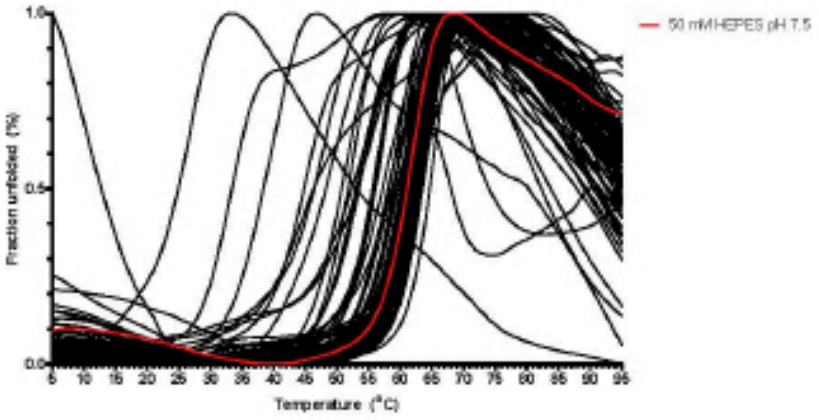
EMBL Hamburg

SPC Facility

Report : Thermal stability assay using buffer screen 2

Overview of the 96 conditions
conditions A1 - H12





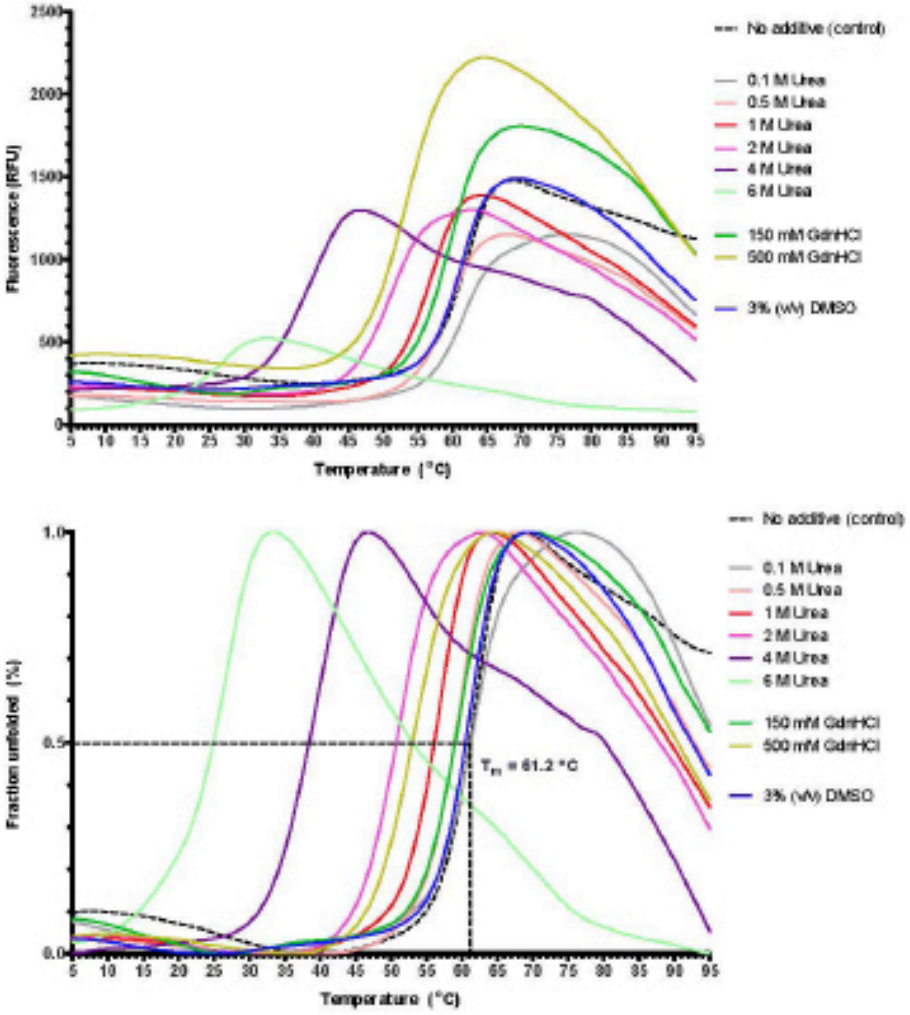
Comments:

- good signal-to-noise ratio
- sharp melting transition in majority of conditions
- some additives improve the stability of the protein and increase its T_m by few degrees in comparison to the buffer control

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2

Effect of chaotropic and dissociating reagents conditions A1-A8



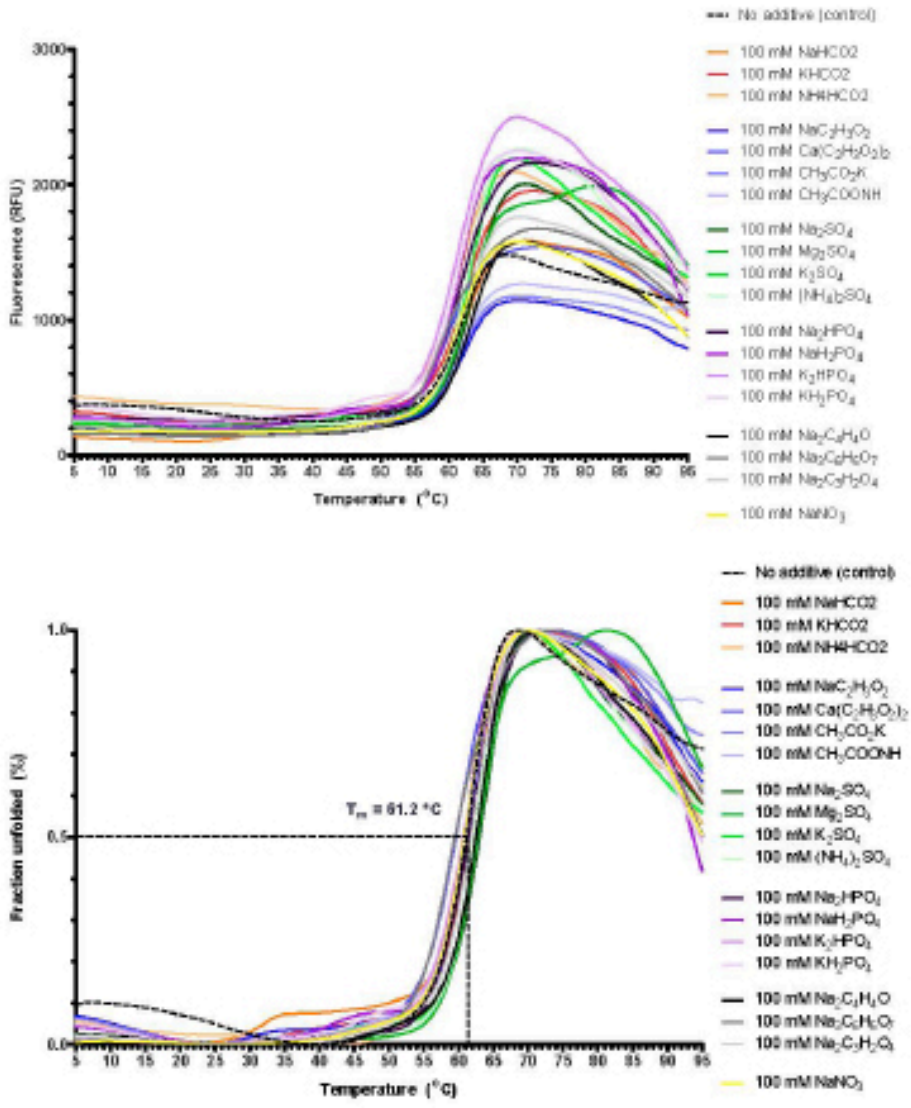
Comments:

- urea and GdnHCl destabilize the protein in a concentration-dependant manner
- 3 % DMSO does not show any significant effect

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2

Effect of 19 salts conditions A10-C4



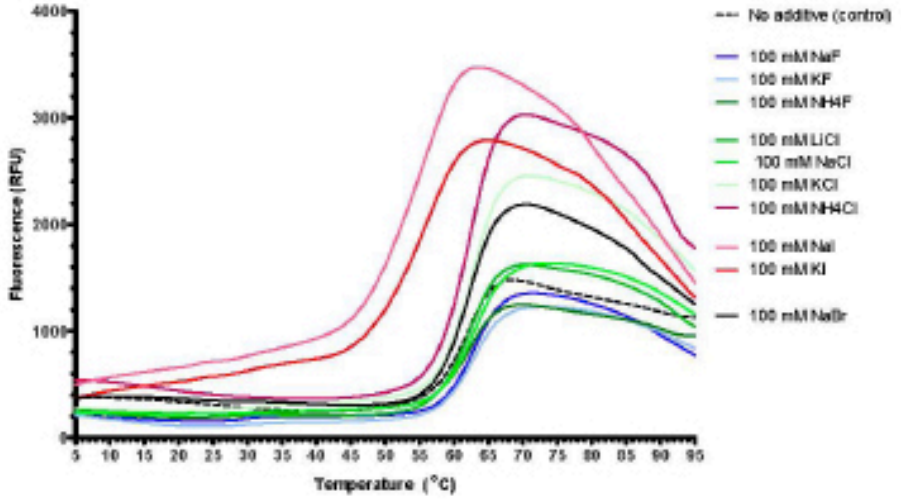
Comments:

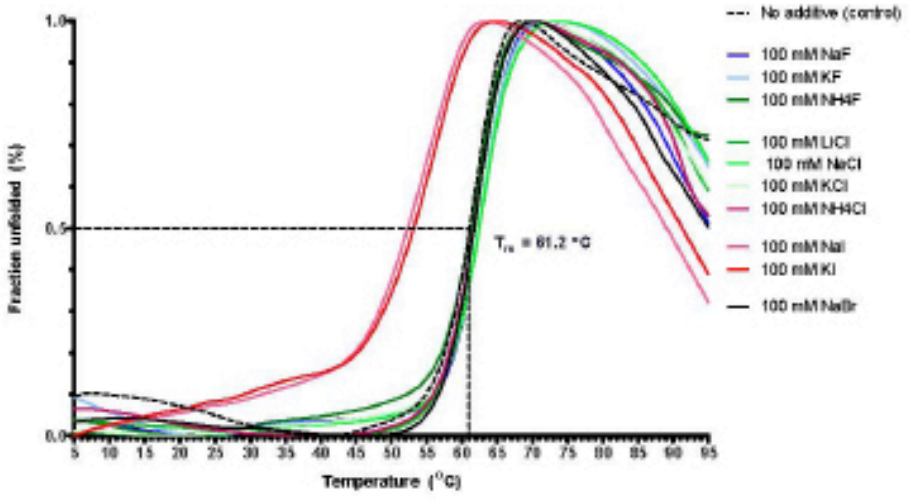
- in general, no significant effect on the protein stability
- only 100 mM $Ca(C_2H_3O_2)_2$ and 100 mM $Na_3C_6H_5O_7$ might slightly destabilize the protein

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2

Effect of monovalent ions and halogens conditons E1-E10





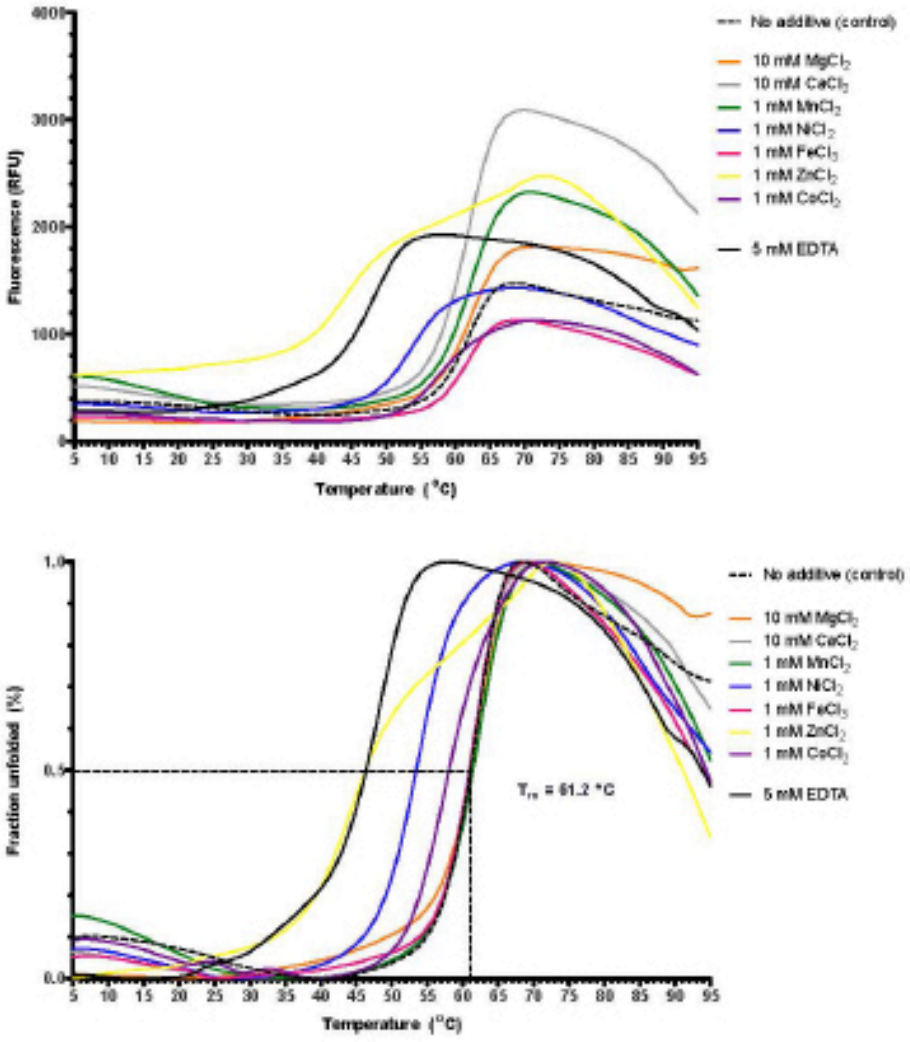
Comments:

- iodides destabilize the protein

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2

7 multivalent and metal ions conditions D6-D12



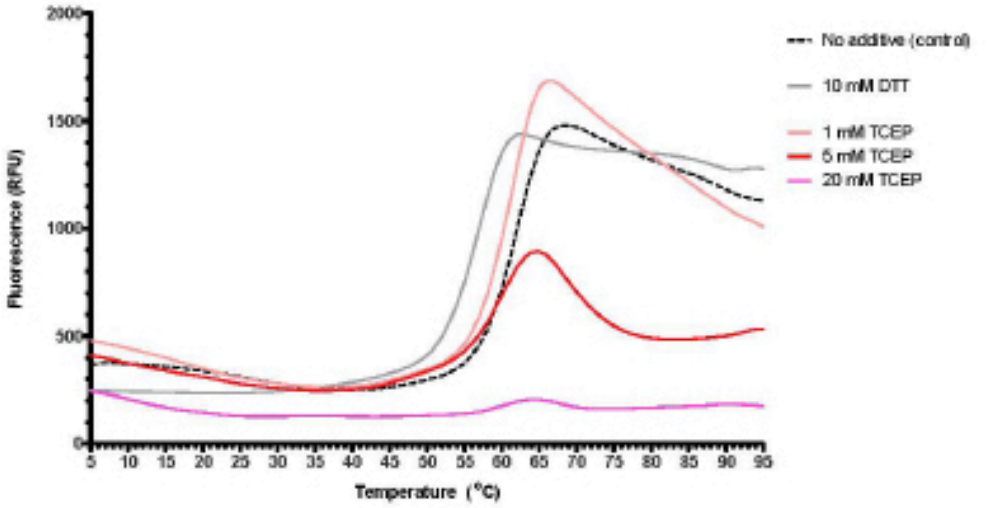
Comments:

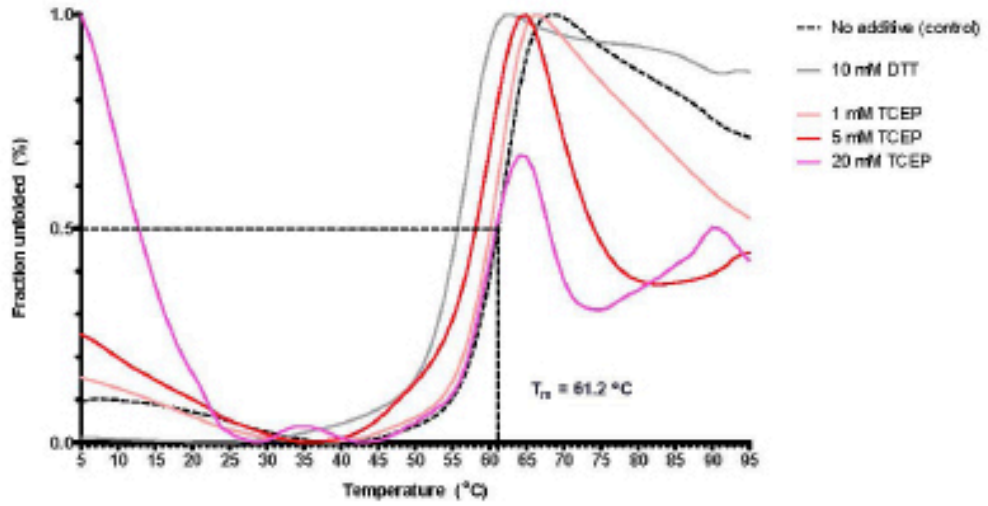
- excluding 10 mM $CaCl_2$ and 1 mM $MnCl_2$, all the multivalent ions destabilize the protein and cause its unfolding at lower temperature

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2

Effect of reducing reagents conditions C5-C8





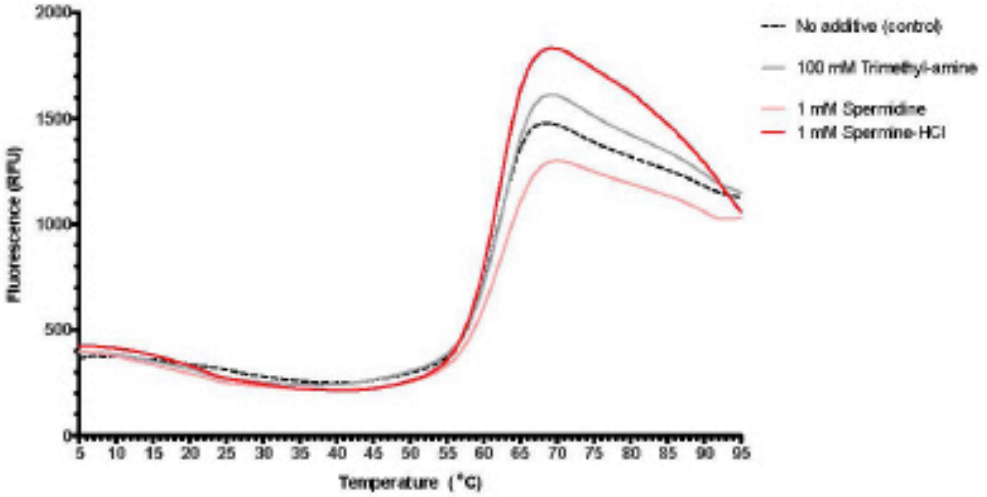
Comments:

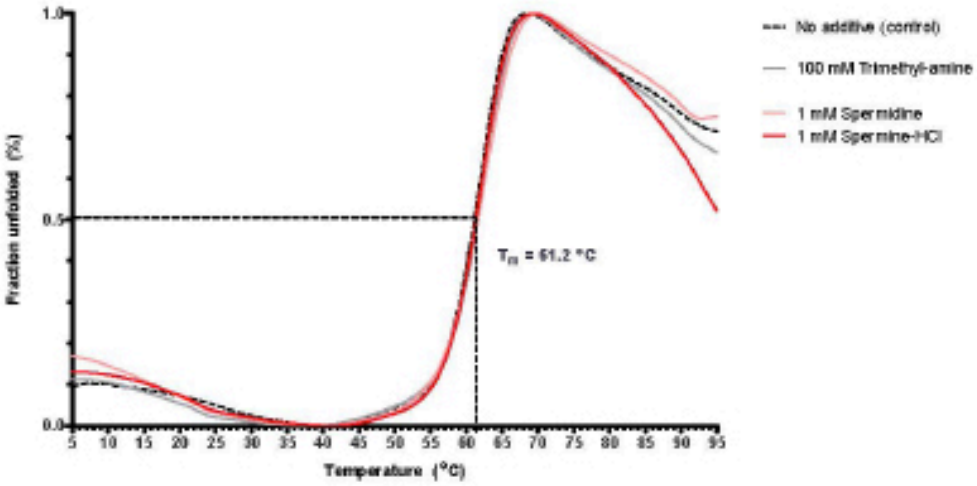
- all reducing reagents destabilize the protein
- if necessary, 1 mM TCEP might be an option to use (only slight destabilizing effect)

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2

Effect of "crowding reagents" as polyamines conditions C9-C11



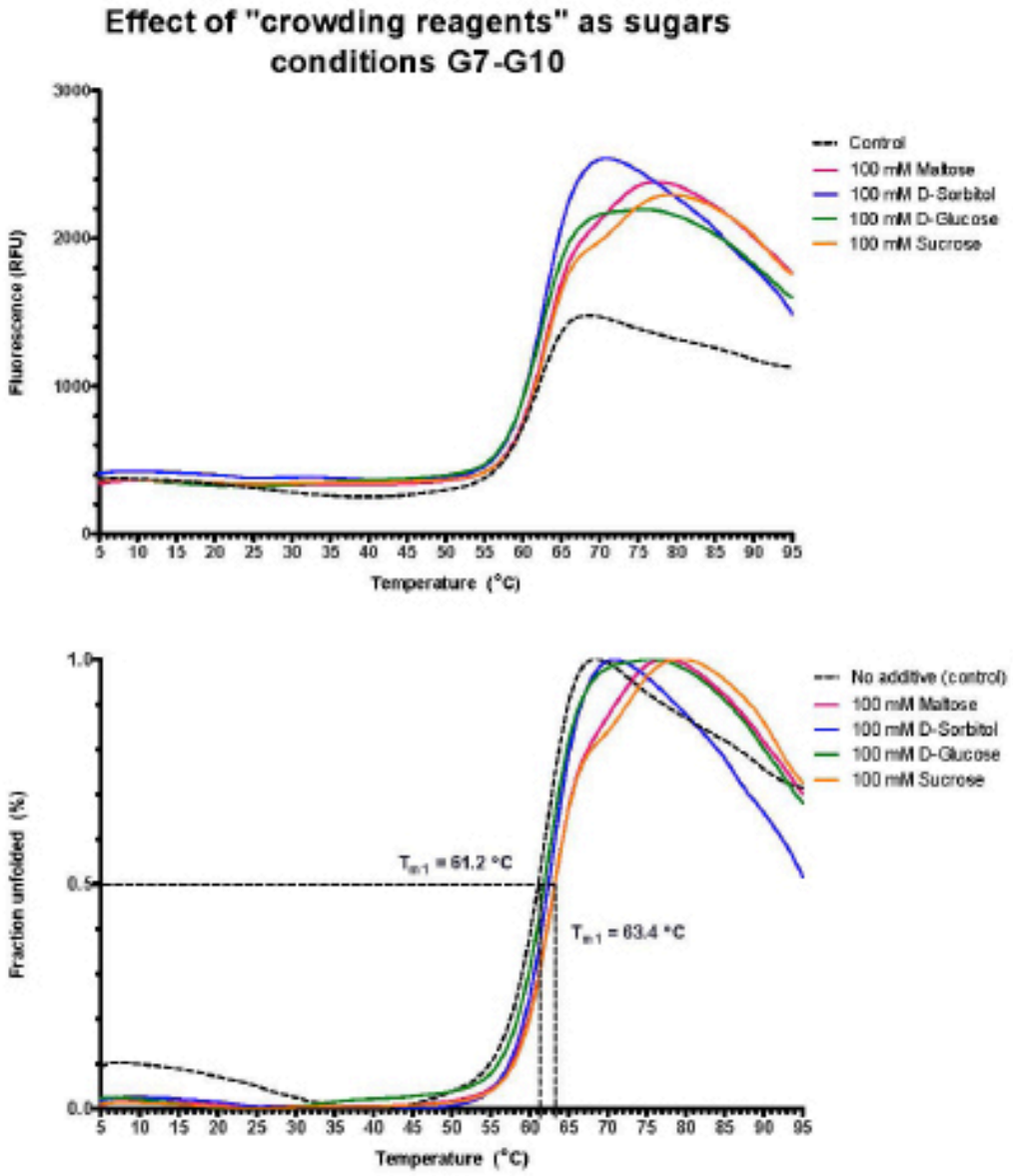


Comments:

- no significant effect

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2



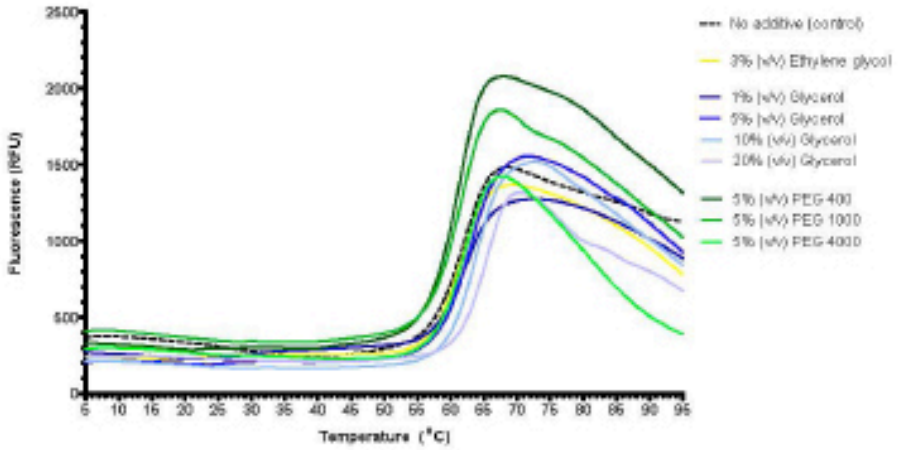
Comments:

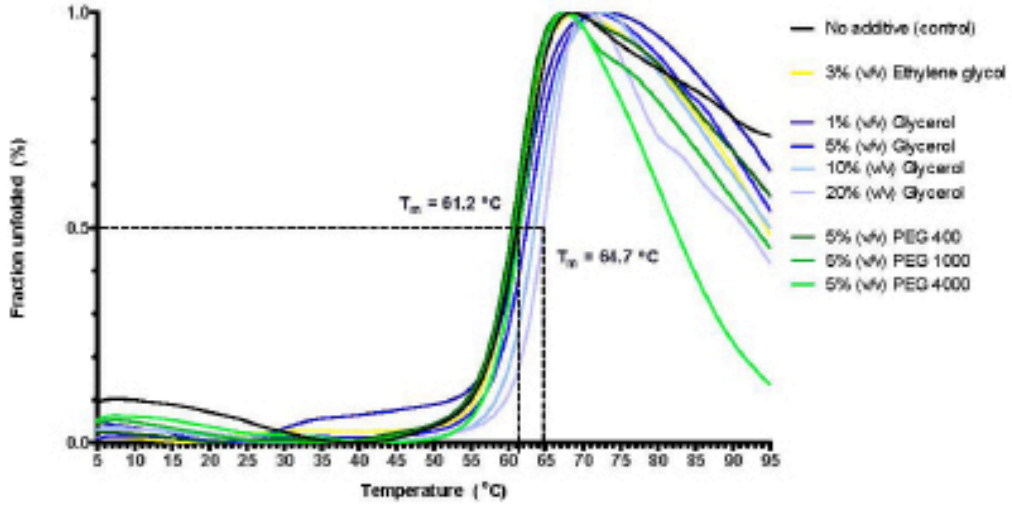
- T_m slightly increases upon addition of sugars
- it might be that the solubility of the protein is increased in presence of sugars

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2

Effect of "crowding reagents" as polyols conditions F1-F8



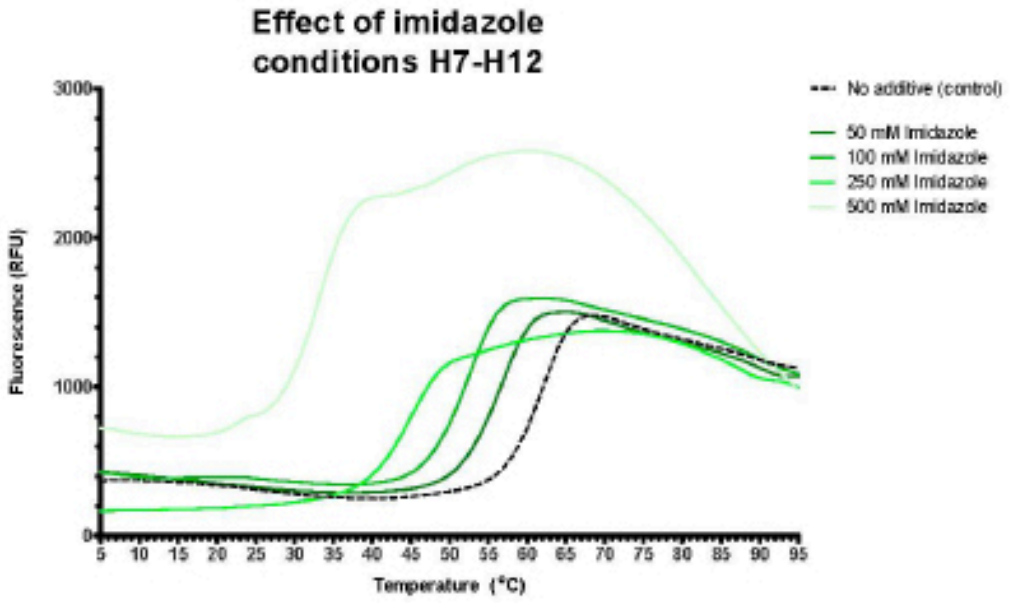


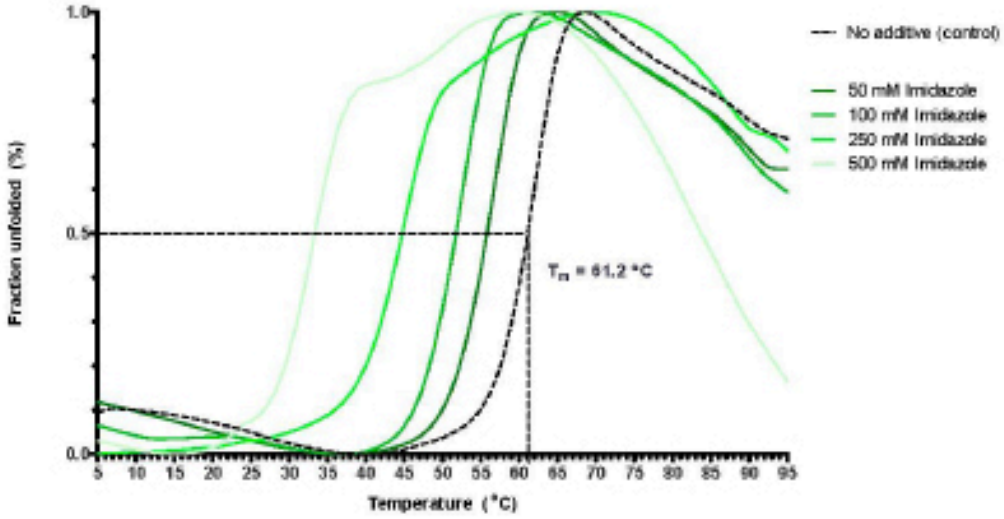
Comments:

- the addition of 5 %, 10 % and 20 % glycerol stabilizes the protein in a concentration-dependant manner (only 1 % glycerol caused baseline instabilities) and increases the T_m by 3.5 °C
- PEGs do not show any significant effect

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2



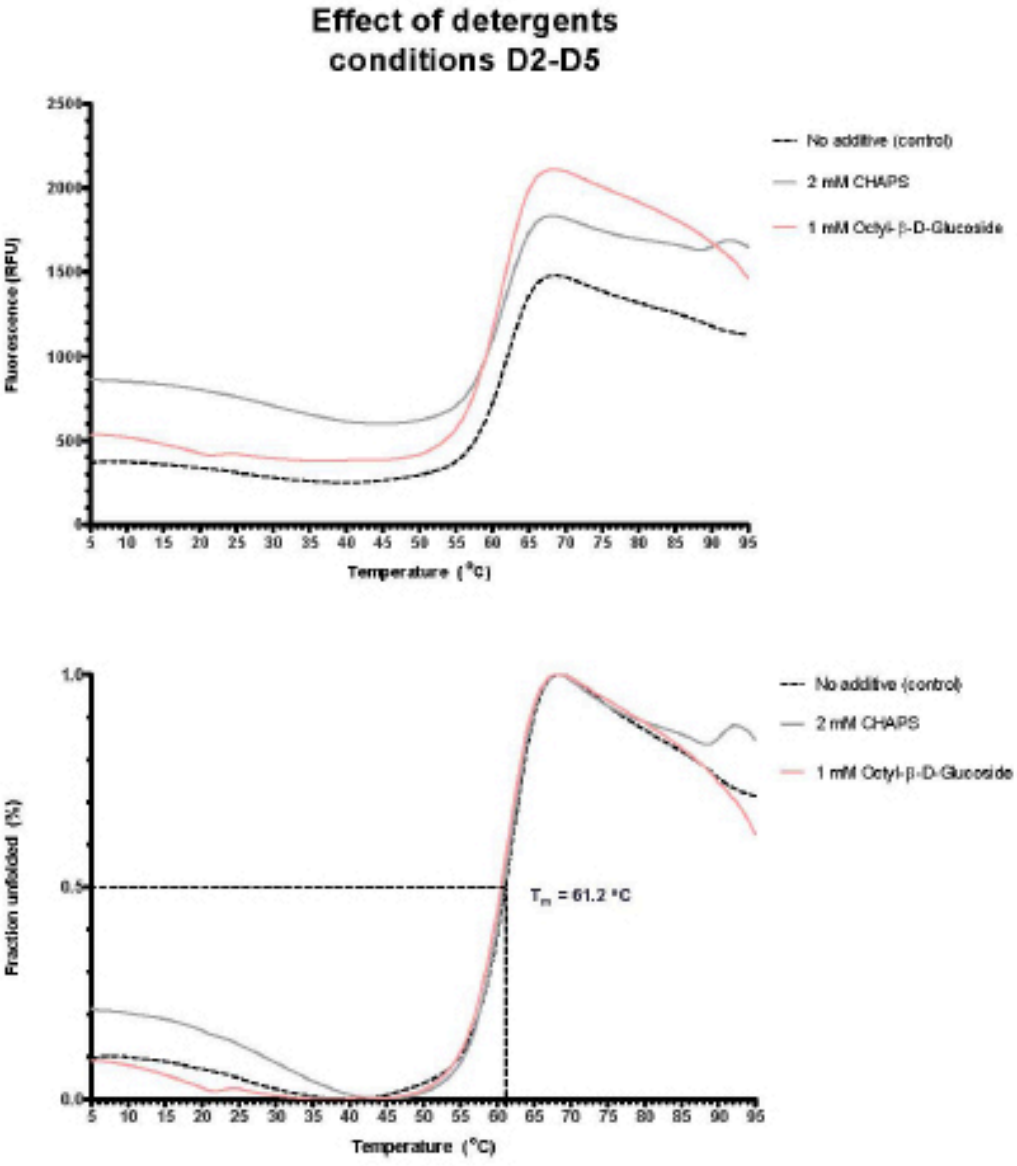


Comments:

- imidazole destabilizes the protein in a concentration-dependant manner

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2

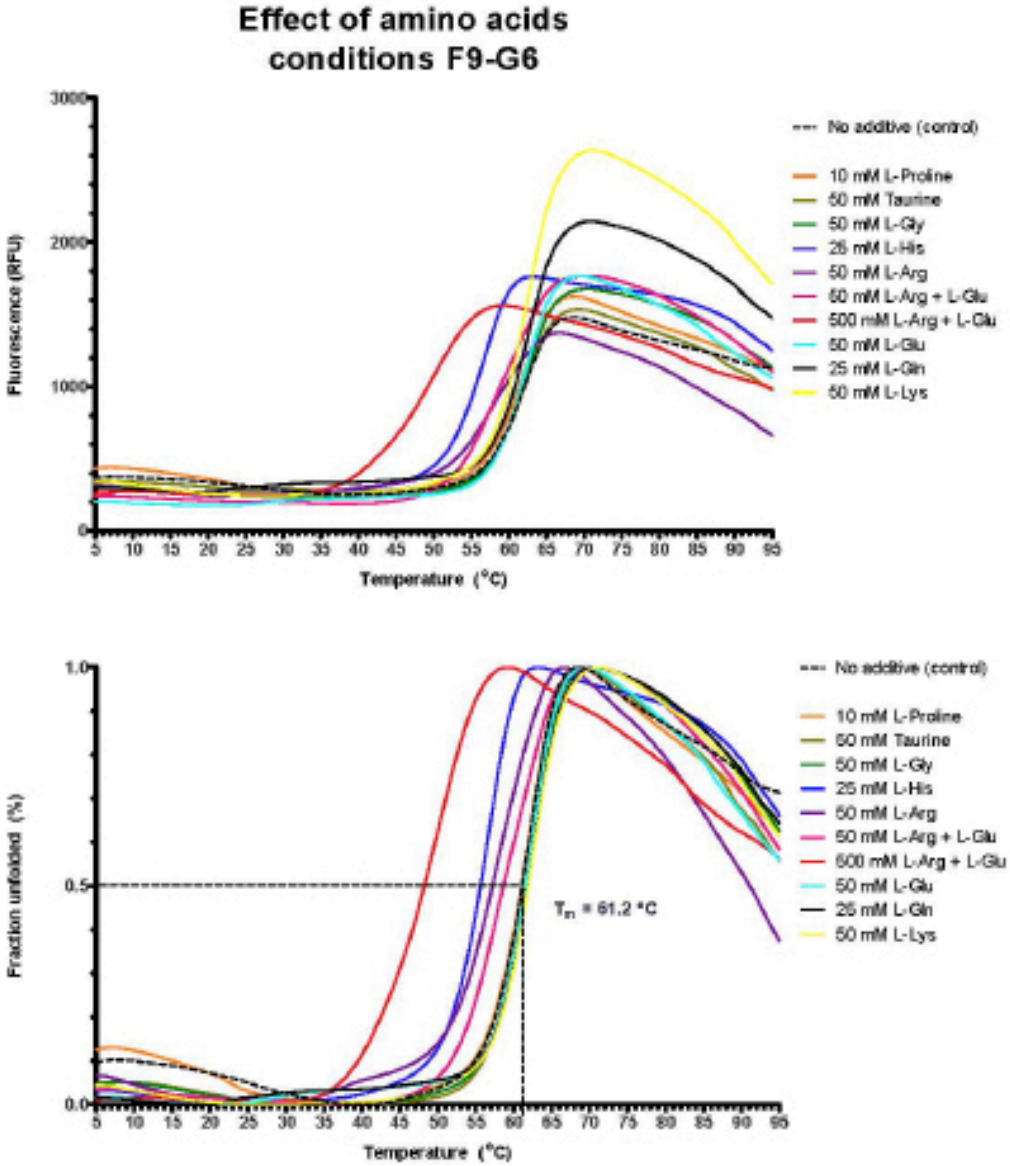


Comments:

- no significant effect on Tm value
- higher fluorescence background upon detergent addition

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2

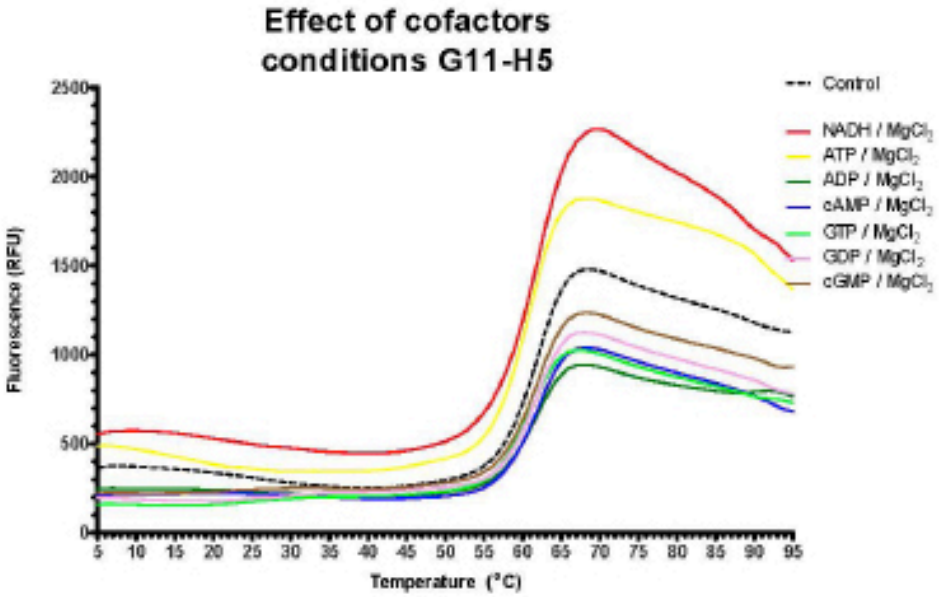


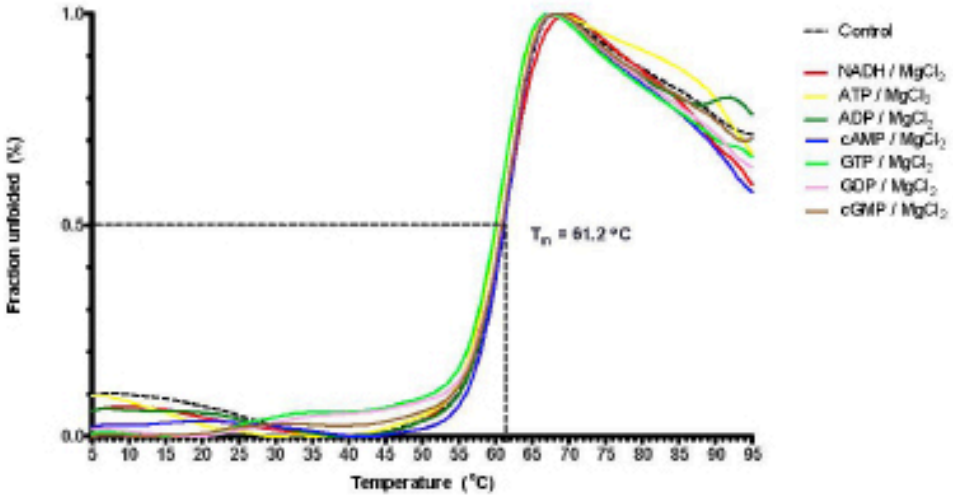
Comments:

- the addition of L-Arg (alone or with combination with another aa) and L-His destabilizes the protein

EMBL Hamburg
SPC Facility

Report : Thermal stability assay using buffer screen 2



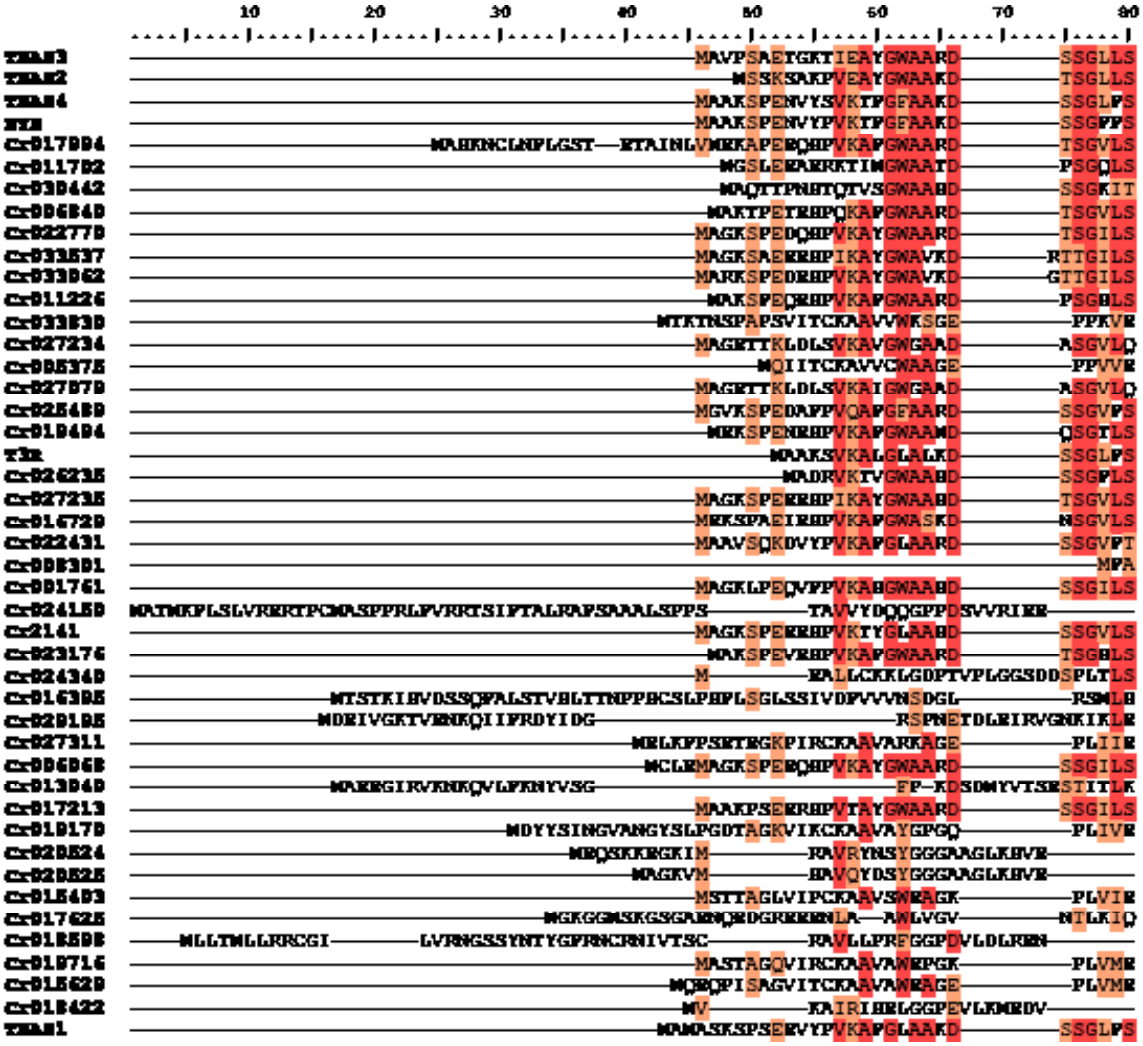


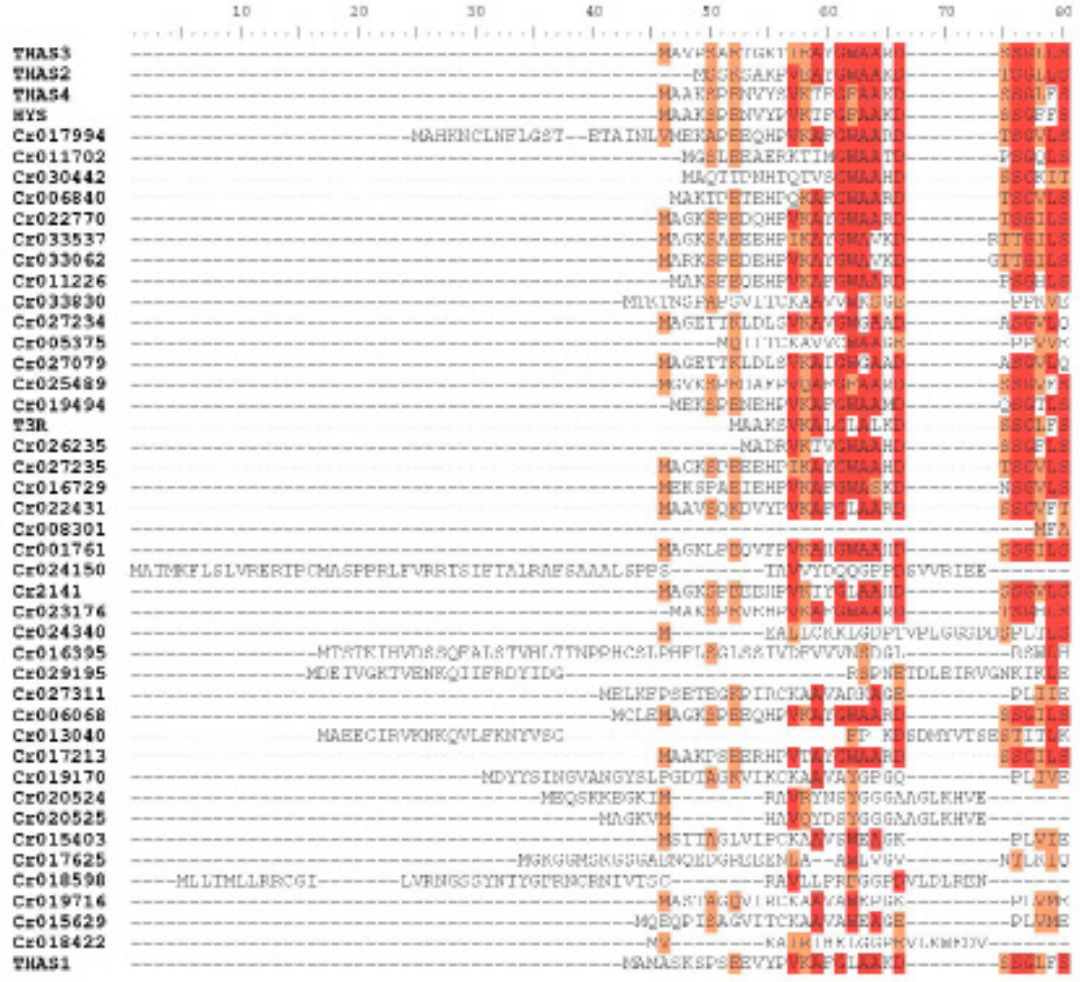
Comments:

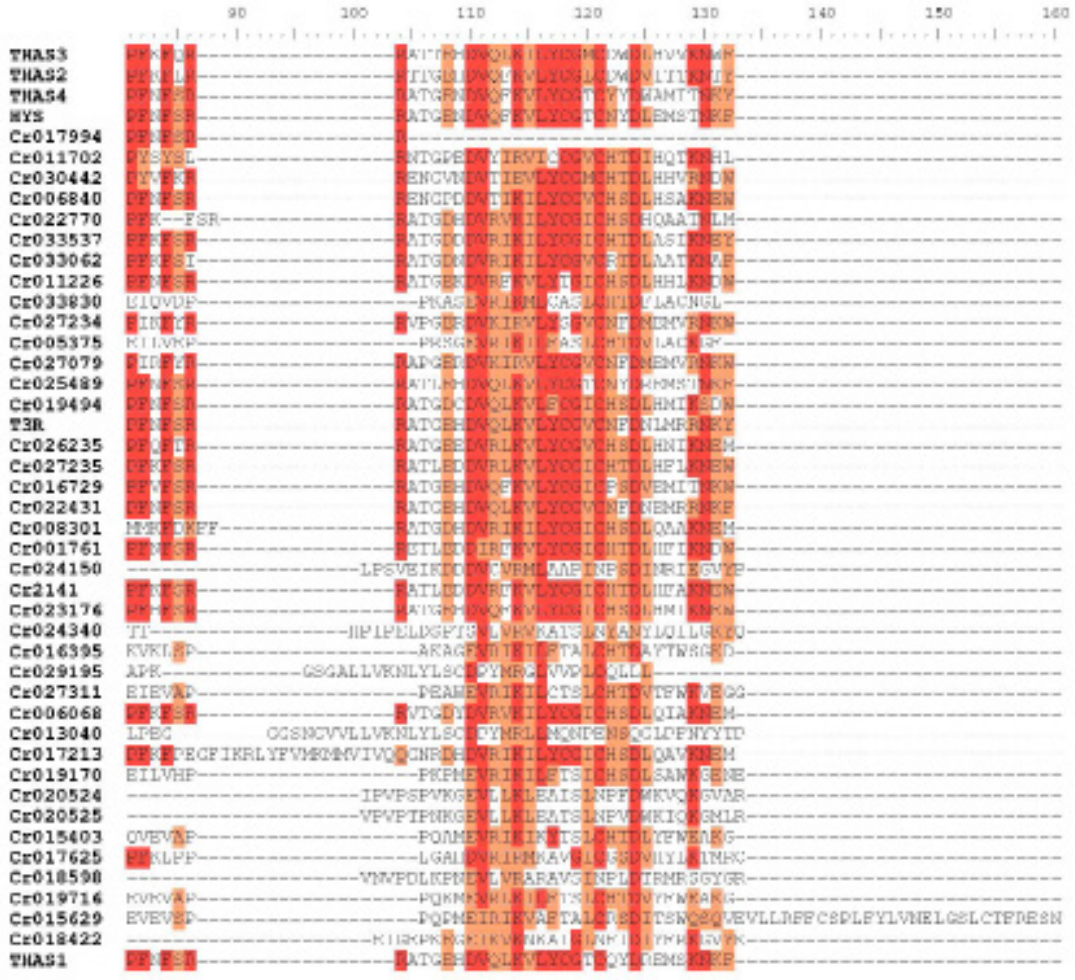
- no effect on T_m value
- small baseline instabilities when guanine additives are used

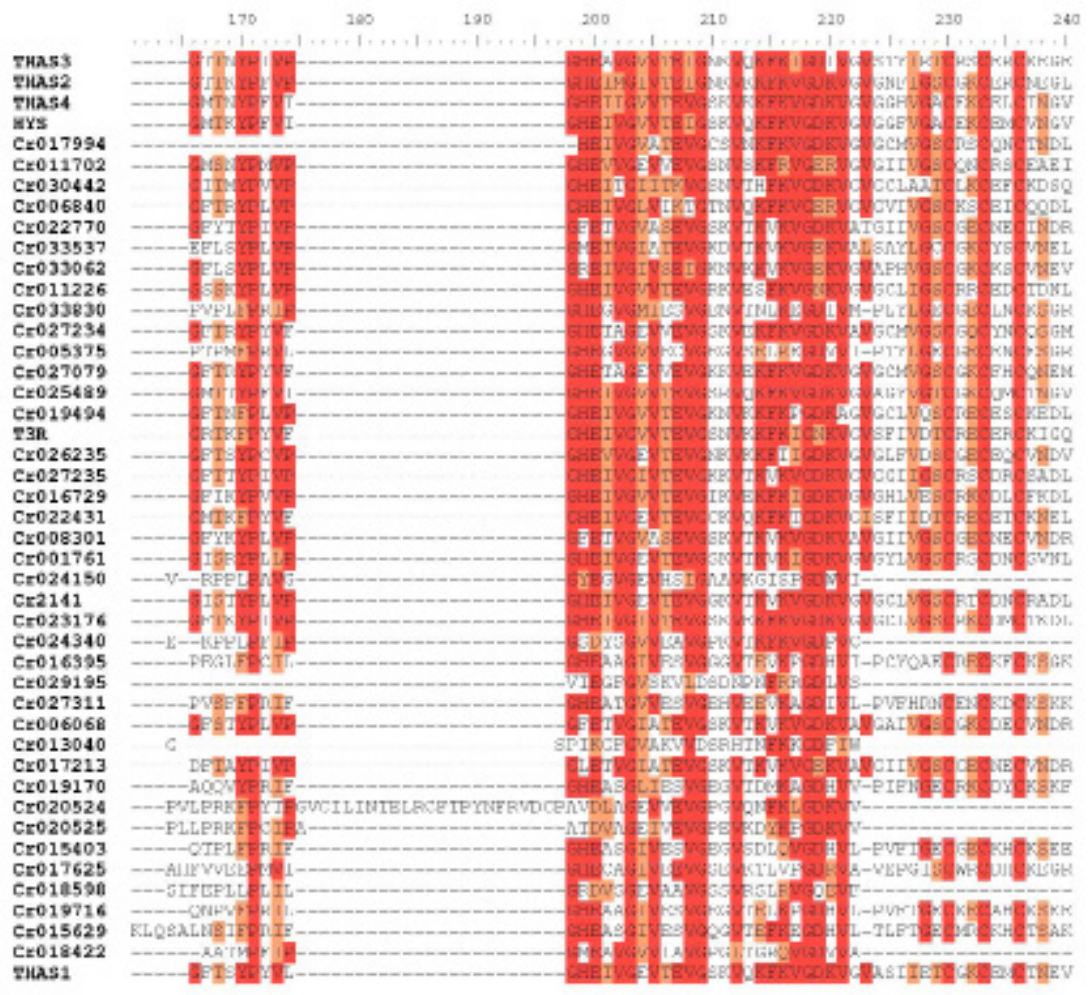
EMBL Hamburg
SPC Facility

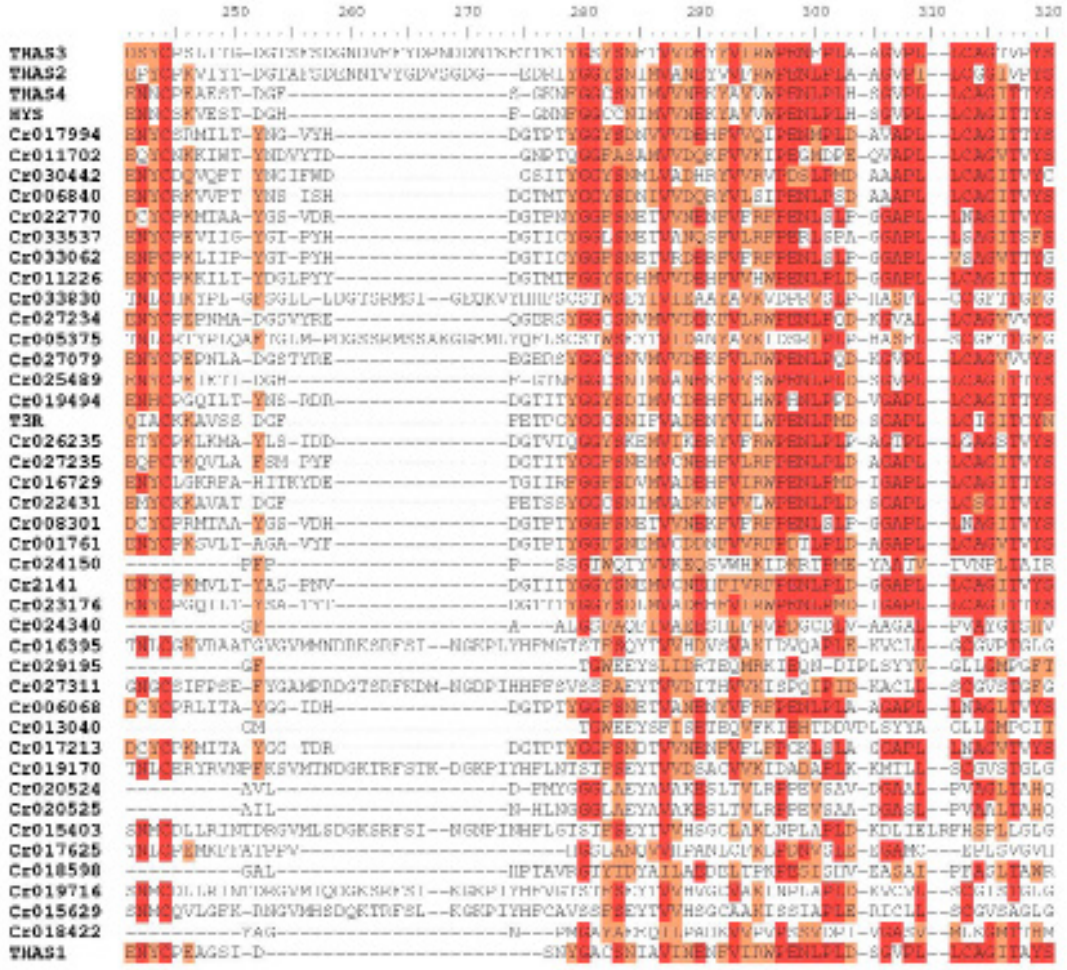
Appendix 5

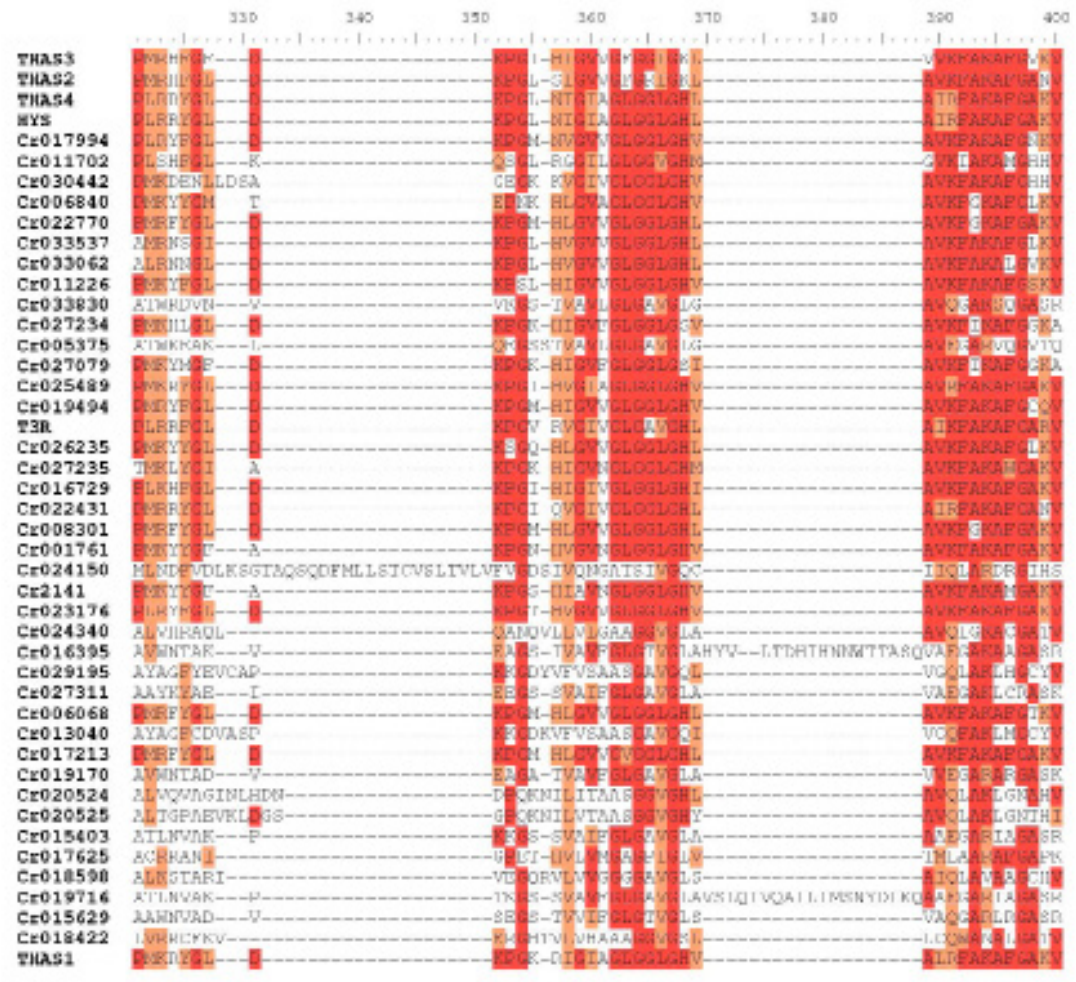


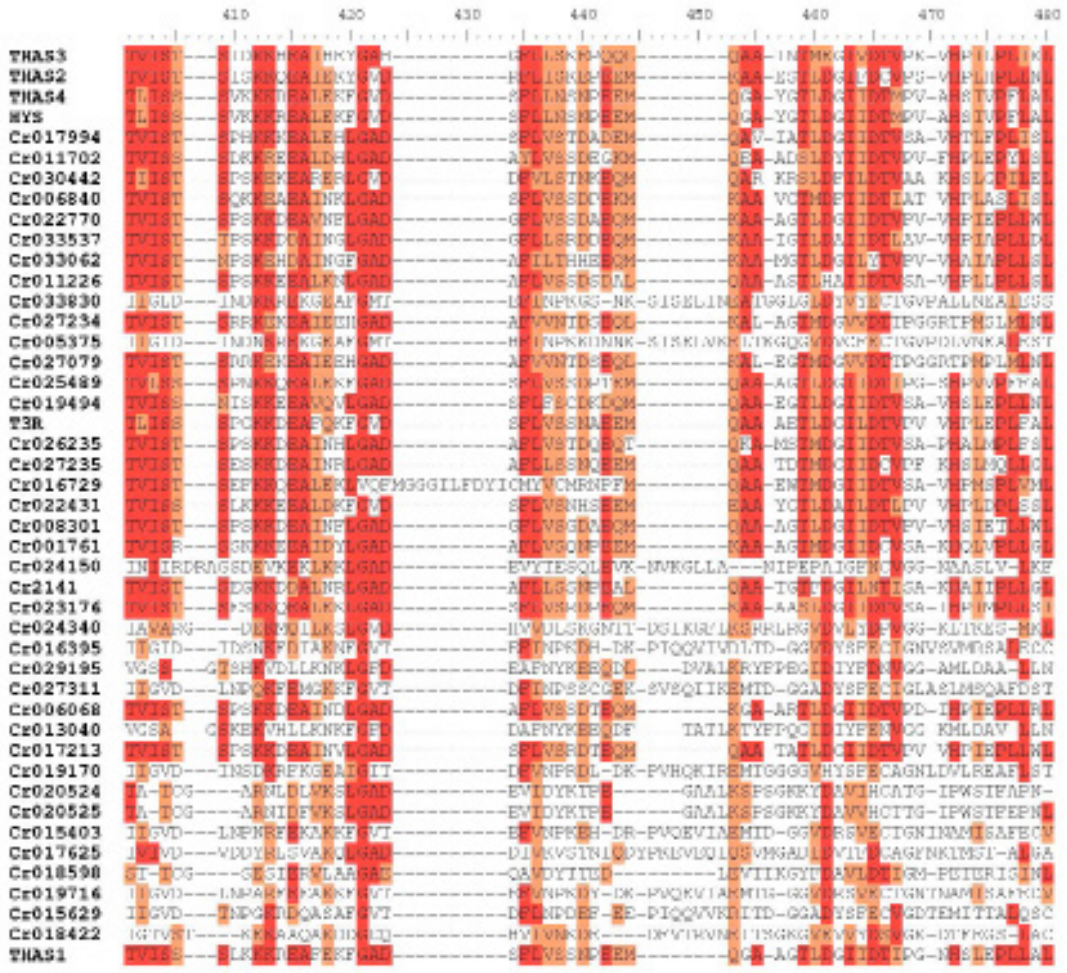


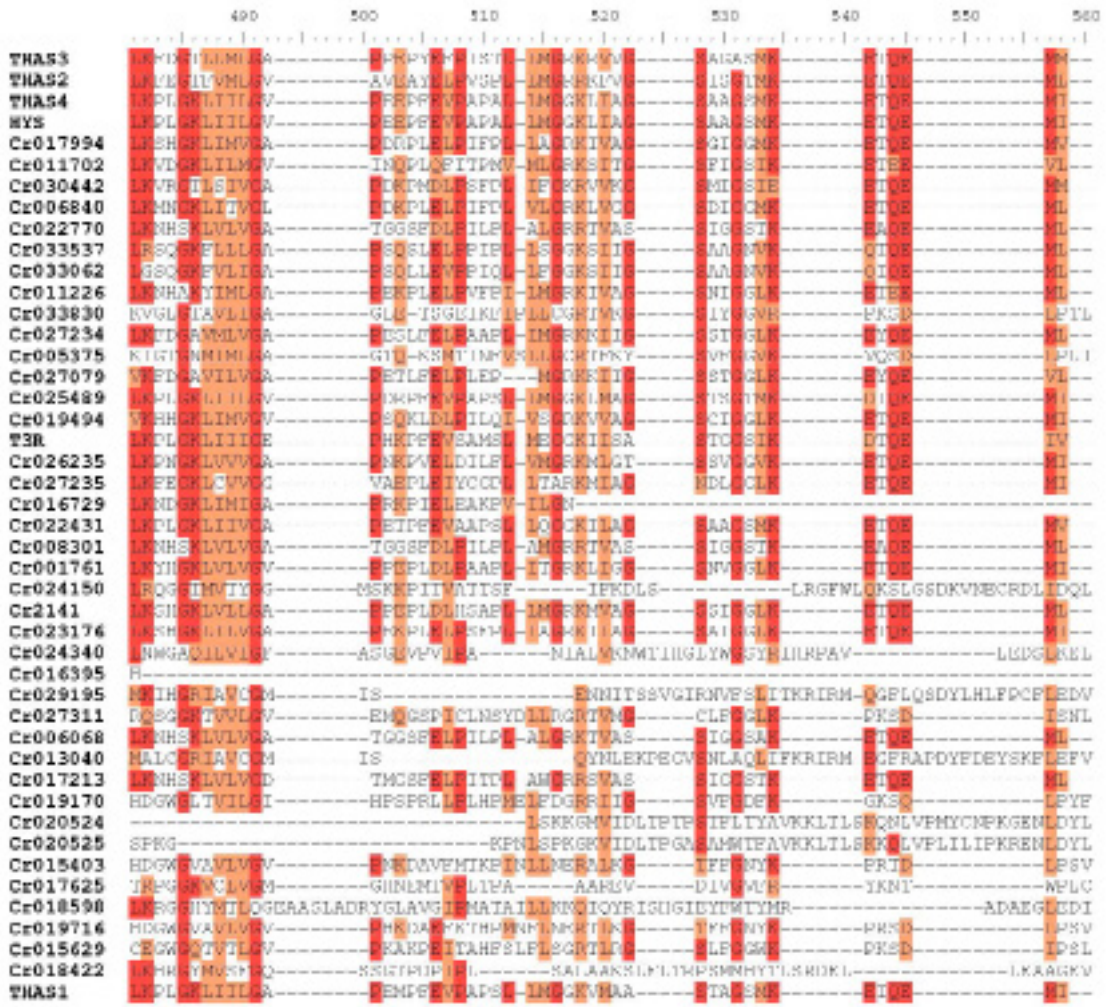

90
100
110
120
130
140
150
160
THAS3
THAS2
THAS4
HYS
Cr017994
Cr011702
Cr030442
Cr006840
Cr022770
Cr033537
Cr033062
Cr011226
Cr033830
Cr027234
Cr005375
Cr027079
Cr025489
Cr019494
T3R
Cr026235
Cr027235
Cr016729
Cr022431
Cr008301
Cr001761
Cr024150
Cr2141
Cr023176
Cr024340
Cr016395
Cr029195
Cr027311
Cr006068
Cr013040
Cr017213
Cr019170
Cr020524
Cr020525
Cr015403
Cr017625
Cr018598
Cr019716
Cr015629
Cr018422
THAS1

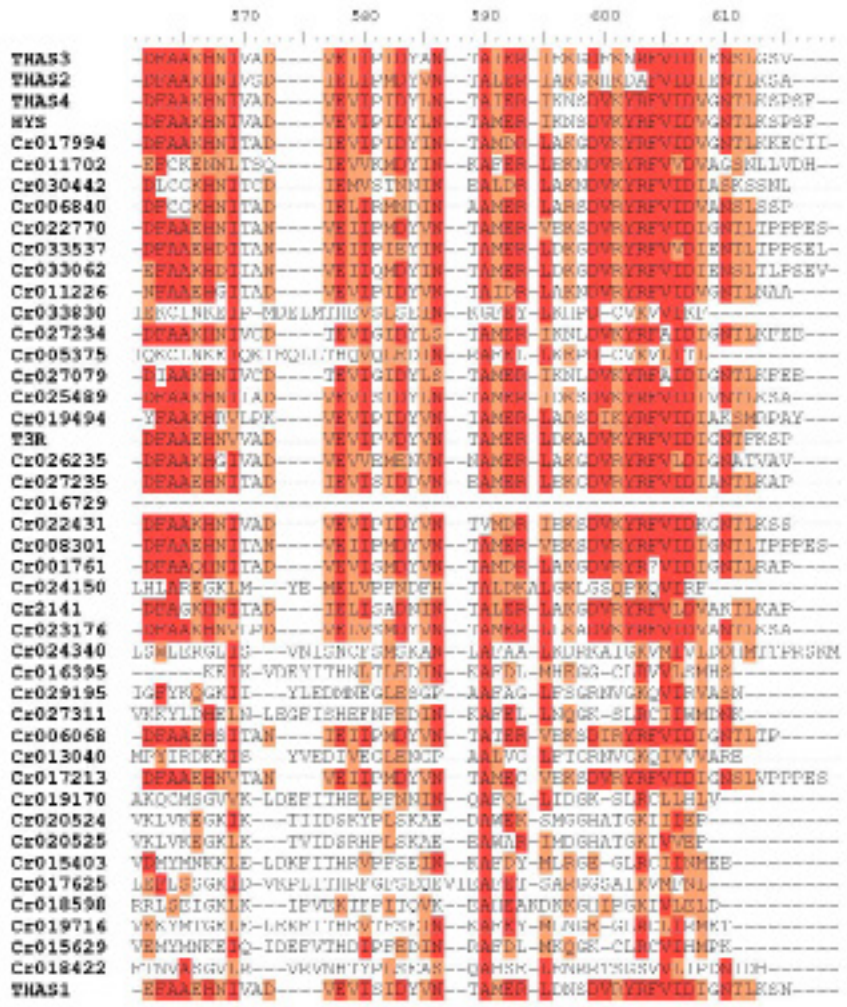

170
180
190
200
210
220
230
240
THAS3
THAS2
THAS4
HYS
Cr017994
Cr011702
Cr030442
Cr006840
Cr022770
Cr033537
Cr033062
Cr011226
Cr033830
Cr027234
Cr005375
Cr027079
Cr025489
Cr019494
T3R
Cr026235
Cr027235
Cr016729
Cr022431
Cr008301
Cr001761
Cr024150
Cr2141
Cr023176
Cr024340
Cr016395
Cr029195
Cr027311
Cr006068
Cr013040
Cr017213
Cr019170
Cr020524
Cr020525
Cr015403
Cr017625
Cr018598
Cr019716
Cr015629
Cr018422
THAS1


250
260
270
280
290
300
310
320
THAS3
THAS2
THAS4
HYS
Cr017994
Cr011702
Cr030442
Cr006840
Cr022770
Cr033537
Cr033062
Cr011226
Cr033830
Cr027234
Cr005375
Cr027079
Cr025489
Cr019494
T3R
Cr026235
Cr027235
Cr016729
Cr022431
Cr008301
Cr001761
Cr024150
Cr2141
Cr023176
Cr024340
Cr016395
Cr029195
Cr027311
Cr006068
Cr013040
Cr017213
Cr019170
Cr020524
Cr020525
Cr015403
Cr017625
Cr018590
Cr019716
Cr015629
Cr018422
THAS1

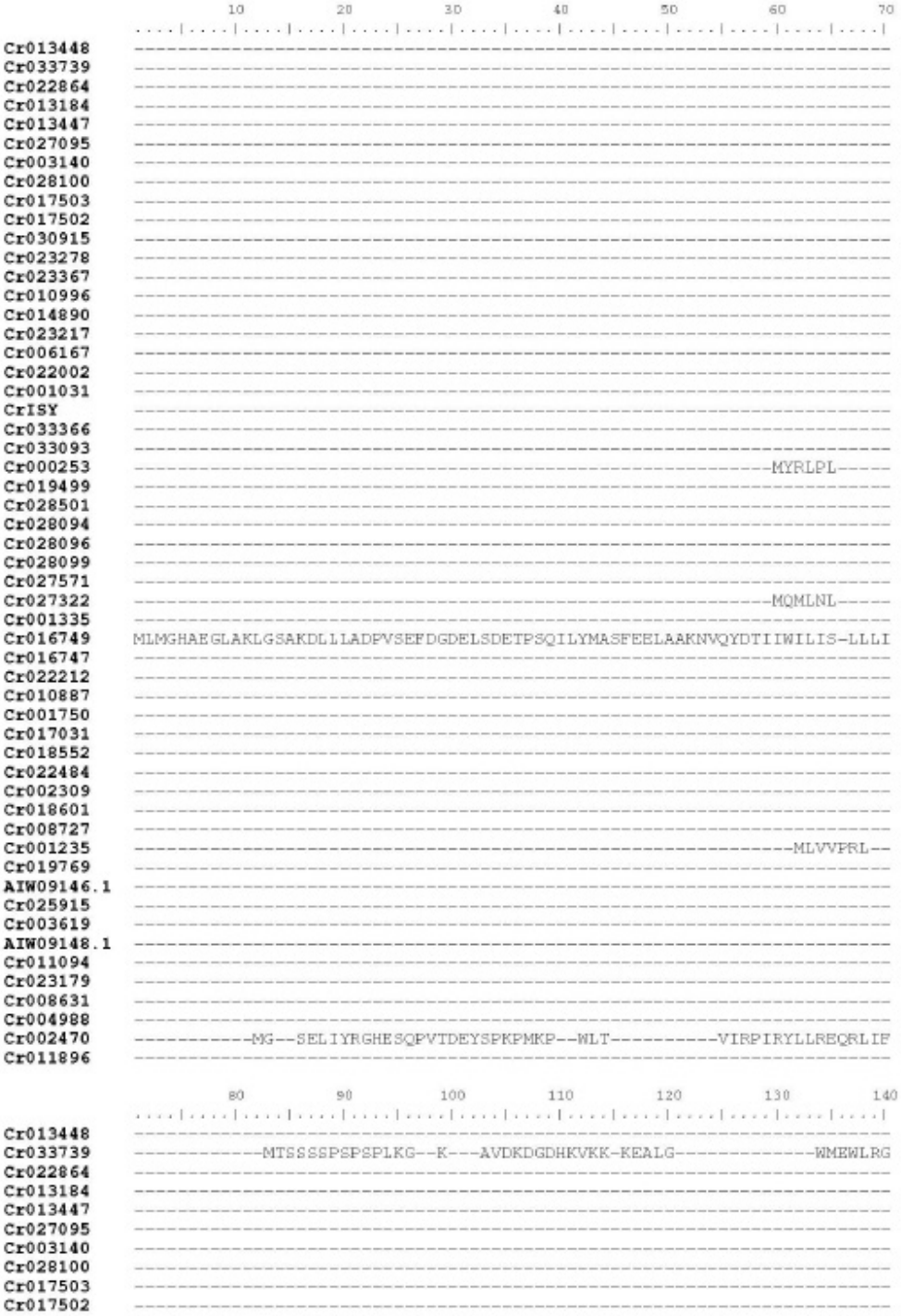

330
340
350
360
370
380
390
400
THAS3
THAS2
THAS4
HYS
Cr017994
Cr011702
Cr030442
Cr006840
Cr022770
Cr033537
Cr033062
Cr011226
Cr033830
Cr027234
Cr005375
Cr027079
Cr025489
Cr019494
T3R
Cr026235
Cr027235
Cr016729
Cr022431
Cr008301
Cr001761
Cr024150
Cr2141
Cr023176
Cr024340
Cr016395
Cr029195
Cr027311
Cr006068
Cr013040
Cr017213
Cr019170
Cr020524
Cr020525
Cr015403
Cr017625
Cr018598
Cr019716
Cr015629
Cr018422
THAS1


410
420
430
440
450
460
470
480
THAS3
THAS2
THAS4
HYS
Cr017994
Cr011702
Cr030442
Cr006840
Cr022770
Cr033537
Cr033062
Cr011226
Cr033830
Cr027234
Cr005375
Cr027079
Cr025489
Cr019494
T3R
Cr026235
Cr027235
Cr016729
Cr022431
Cr008301
Cr001761
Cr024150
Cr2141
Cr023176
Cr024340
Cr016395
Cr029195
Cr027311
Cr006068
Cr013040
Cr017213
Cr019170
Cr020524
Cr020525
Cr015403
Cr017625
Cr018598
Cr019716
Cr015629
Cr018422
THAS1

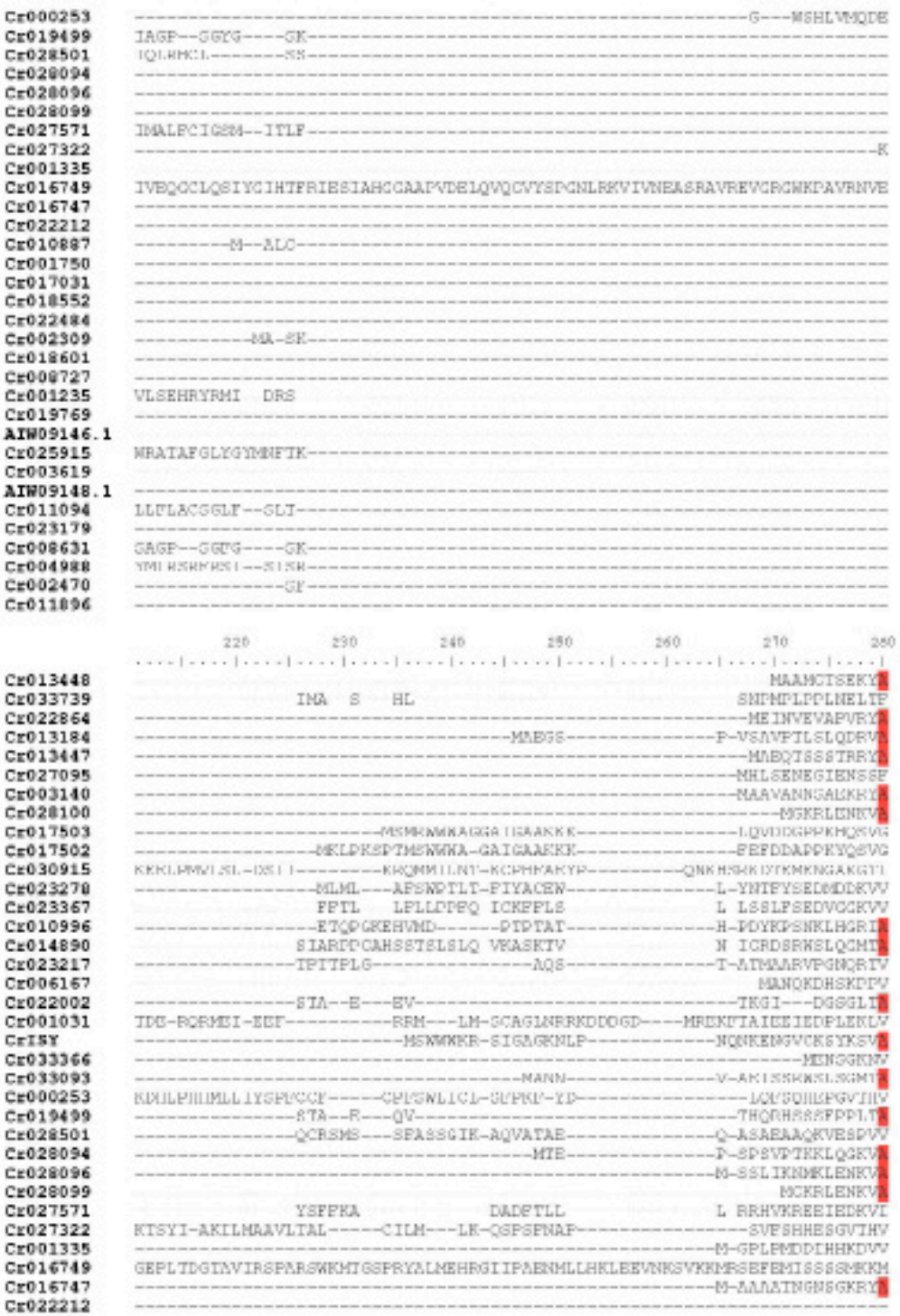

490
500
510
520
530
540
550
560
THAS3
THAS2
THAS4
HYS
Cr017994
Cr011702
Cr030442
Cr006840
Cr022770
Cr033537
Cr030062
Cr011226
Cr033830
Cr027234
Cr005375
Cr027079
Cr025489
Cr019494
T3R
Cr026235
Cr027235
Cr016729
Cr022431
Cr008301
Cr001761
Cr024150
Cr2141
Cr023176
Cr024340
Cr016395
Cr029195
Cr027311
Cr006068
Cr013040
Cr017213
Cr019170
Cr020524
Cr020525
Cr015403
Cr017625
Cr018598
Cr019716
Cr015629
Cr018422
THAS1

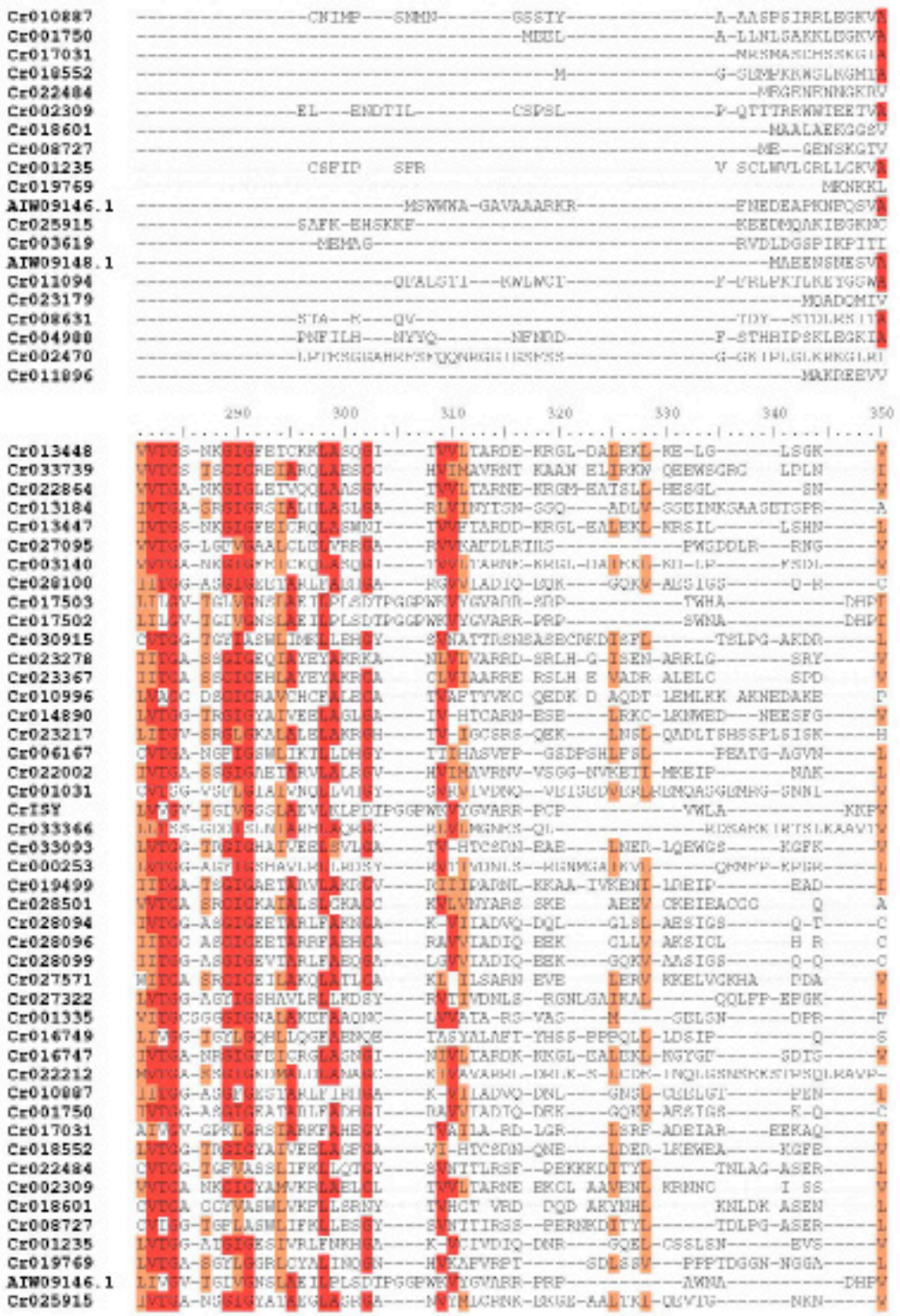

570
580
590
600
610
THAS3
THAS2
THAS4
HYS
Cr017994
Cr011702
Cr030442
Cr006840
Cr022770
Cr033537
Cr033062
Cr011226
Cr033830
Cr027234
Cr005375
Cr027079
Cr025489
Cr019494
T3R
Cr026235
Cr027235
Cr016729
Cr022431
Cr008301
Cr001761
Cr024150
Cr2141
Cr023176
Cr024340
Cr016395
Cr029195
Cr027311
Cr006068
Cr013040
Cr017213
Cr019170
Cr020524
Cr020525
Cr015403
Cr017625
Cr018598
Cr019716
Cr015629
Cr018422
THAS1

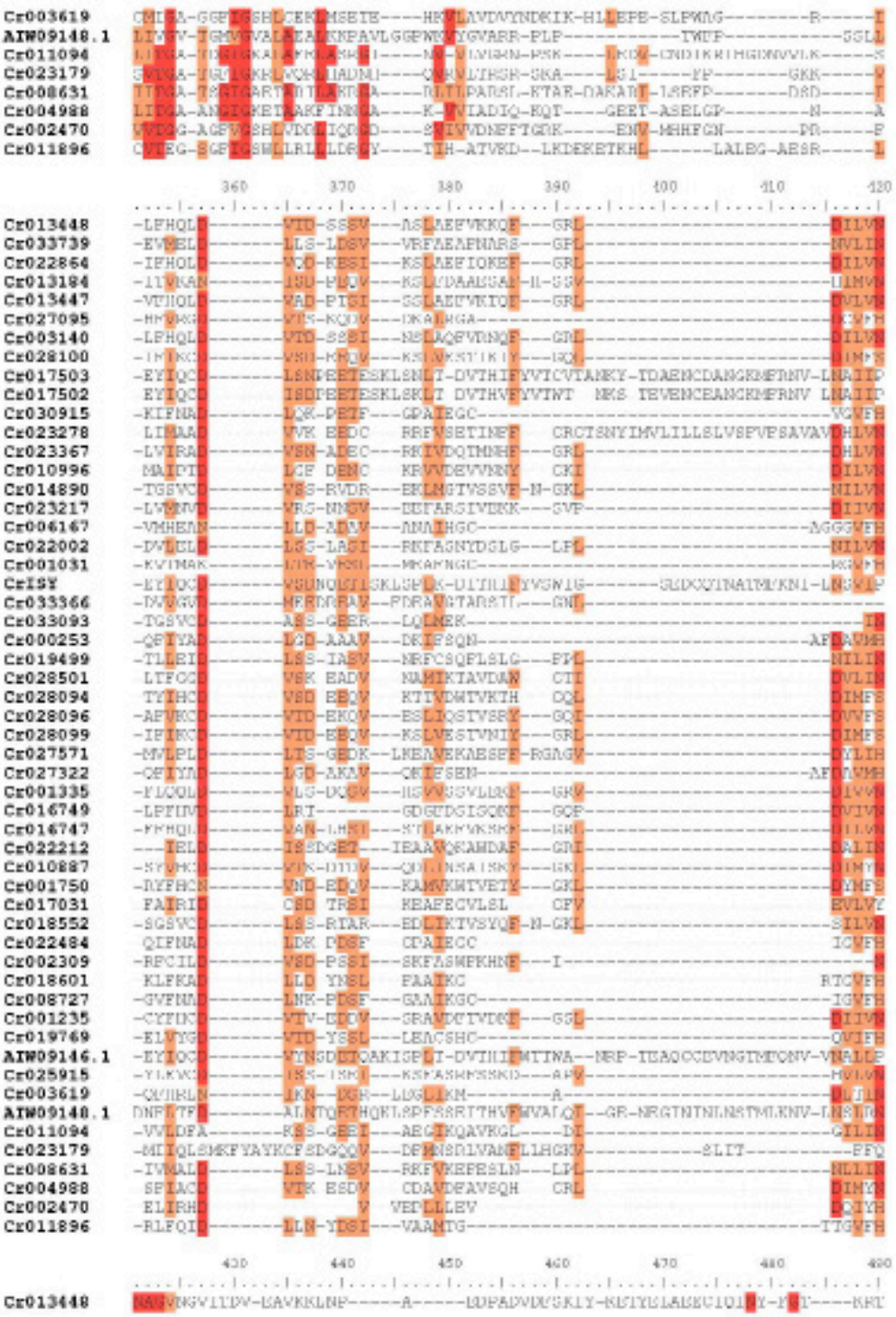
Appendix 6

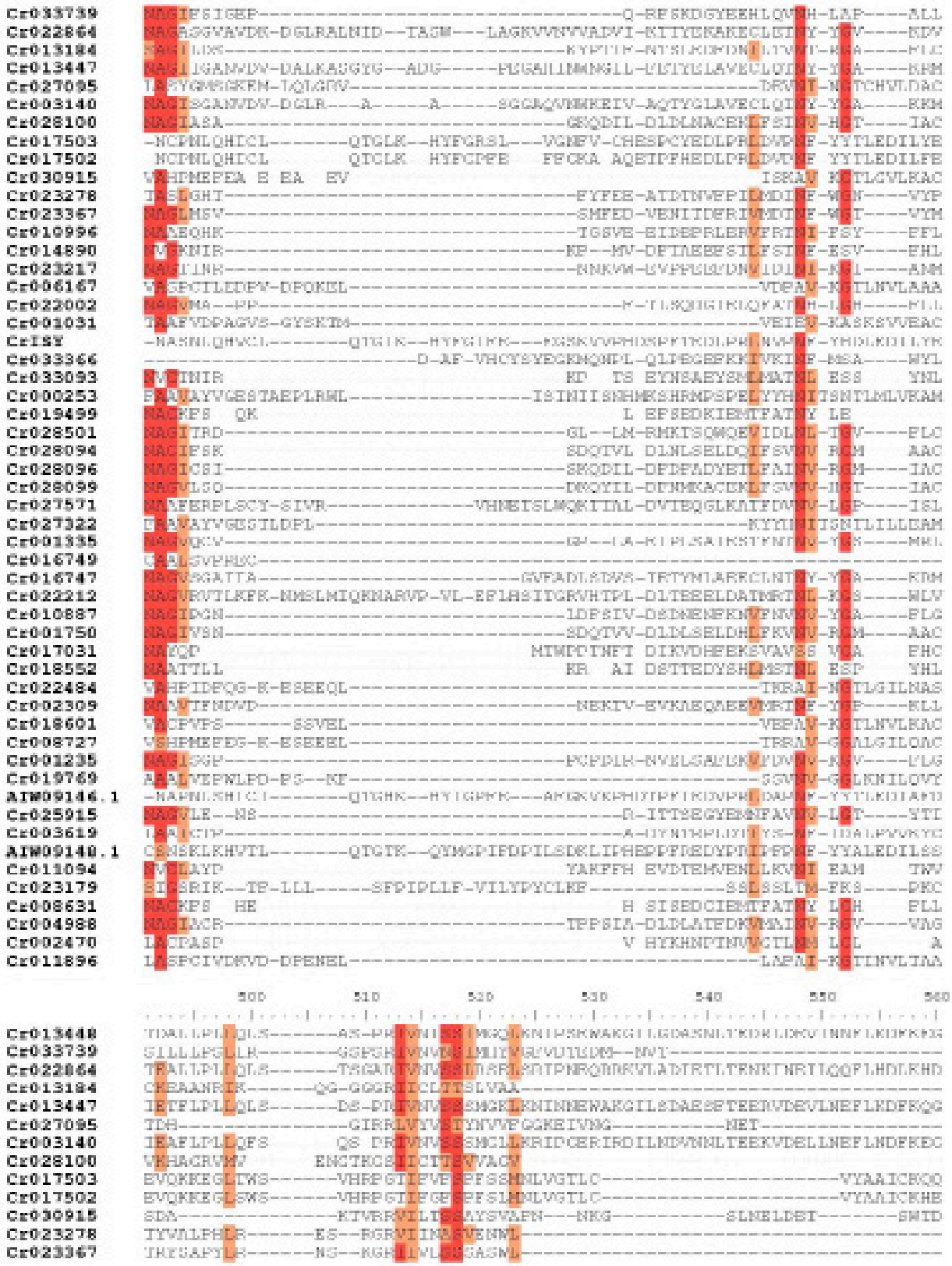

10 20 30 40 50 60 70
Cr013448
Cr033739
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502
Cr030915
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093
Cr000253 -MYRLPL-
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099
Cr027571
Cr027322 -MQMLNL-
Cr001335
Cr016749 MLMGHAEGLAKLGSAKDLLLADPVSEFDGDELSDETPSQILYMASFEELAAKNVQYDTIIWILIS-LLLI
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022484
Cr002309
Cr018601
Cr008727
Cr001235 -MLVVPRL--
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09148.1
Cr011094
Cr023179
Cr008631
Cr004988
Cr002470 ----------MG--SELIYRGHESQPVTDEYSPKPMKP--WLT----------VIRPIRYLLREQRLIF
Cr011896
80 90 100 110 120 130 140
Cr013448
Cr033739 ----------MTSSSSPSPSPLKG--K---AVDKDGDHKVKK-KEALG----------WMEWLRG
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502

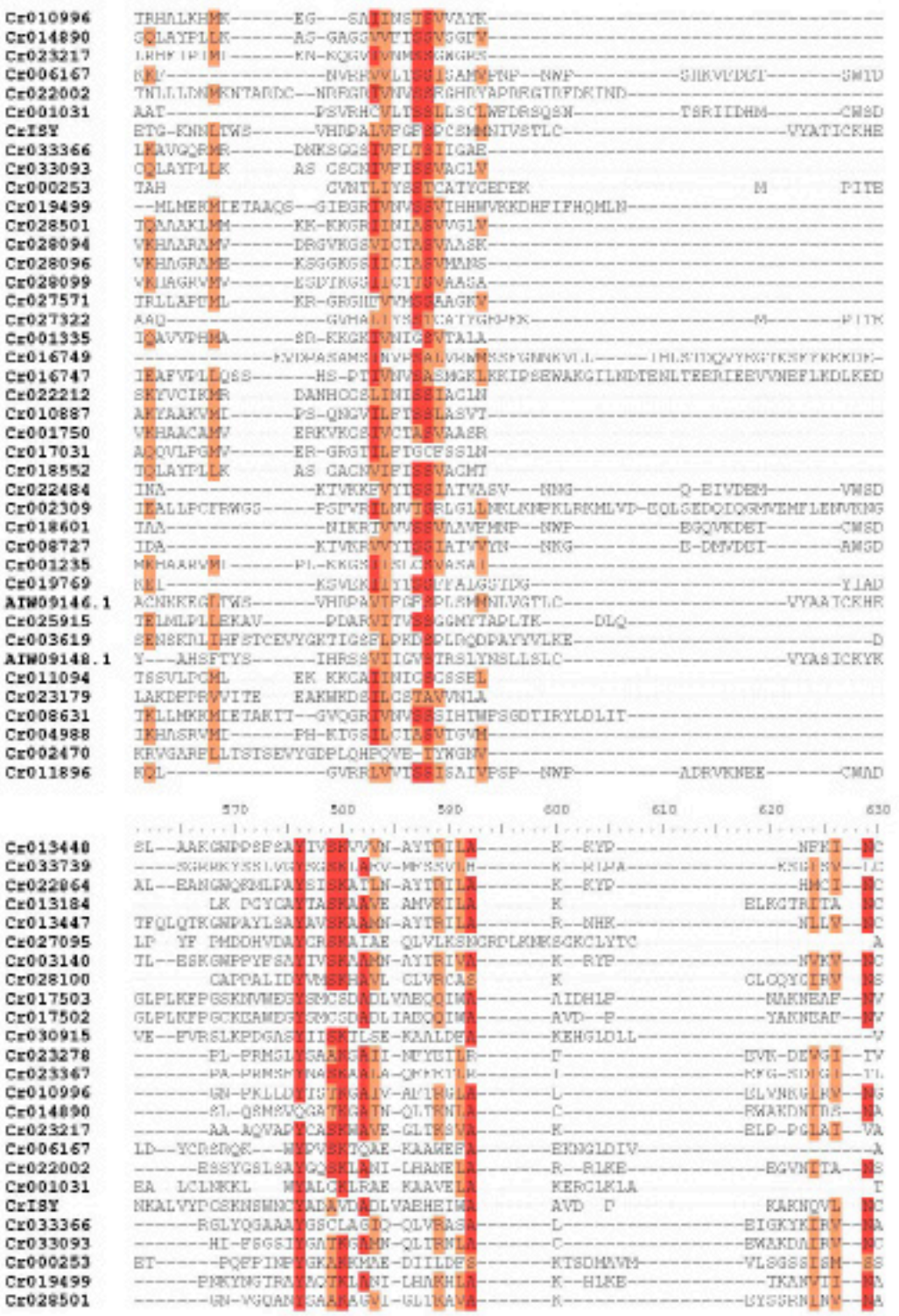

Cr030915
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099
Cr027571
Cr027322
Cr001335
Cr016749
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022484
Cr002309
Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09148.1
Cr011094
Cr023179
Cr008631
Cr004988
Cr002470
Cr011896
150
160
170
180
190
200
210
Cr013448
Cr033739
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502
Cr030915
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093

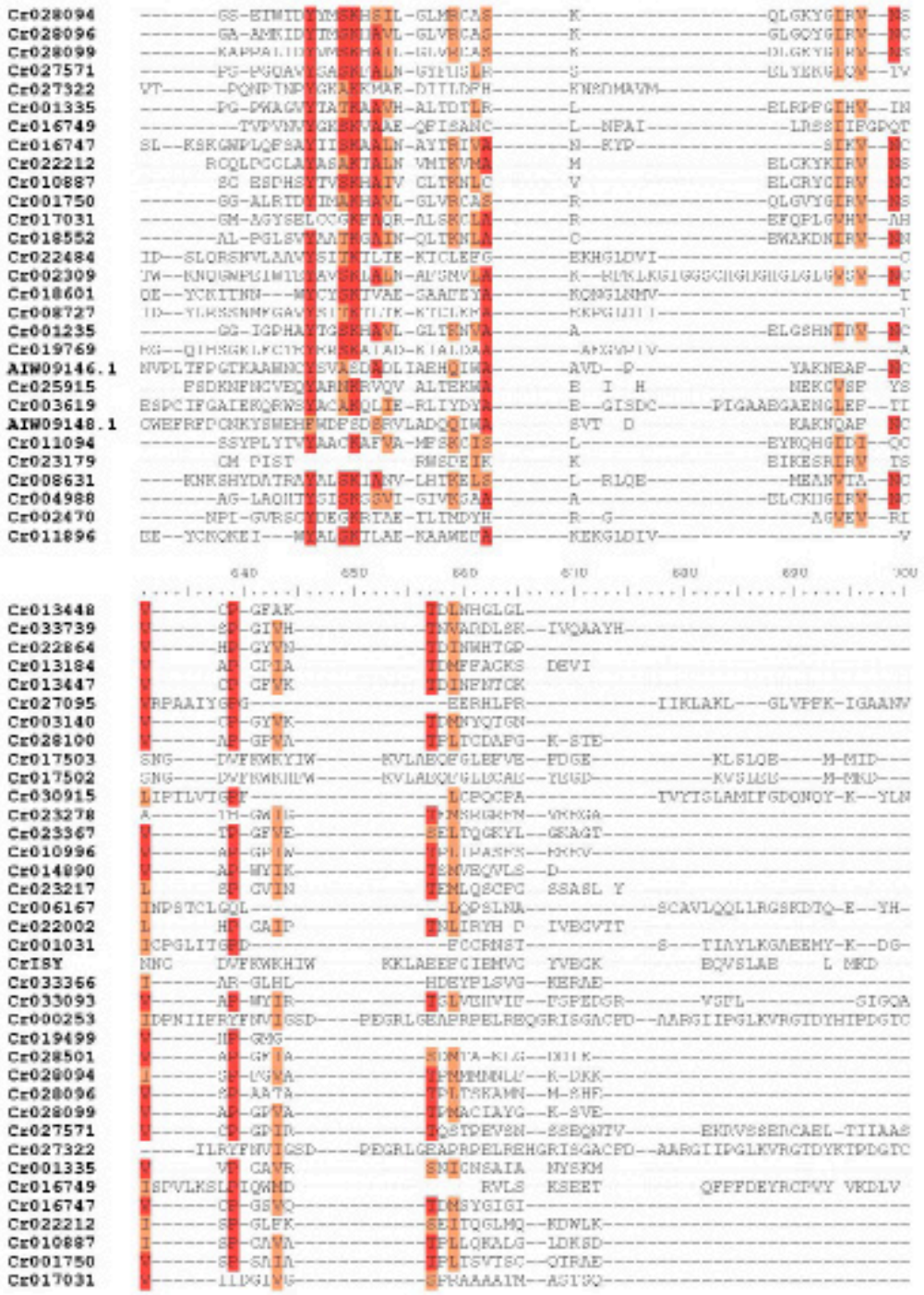


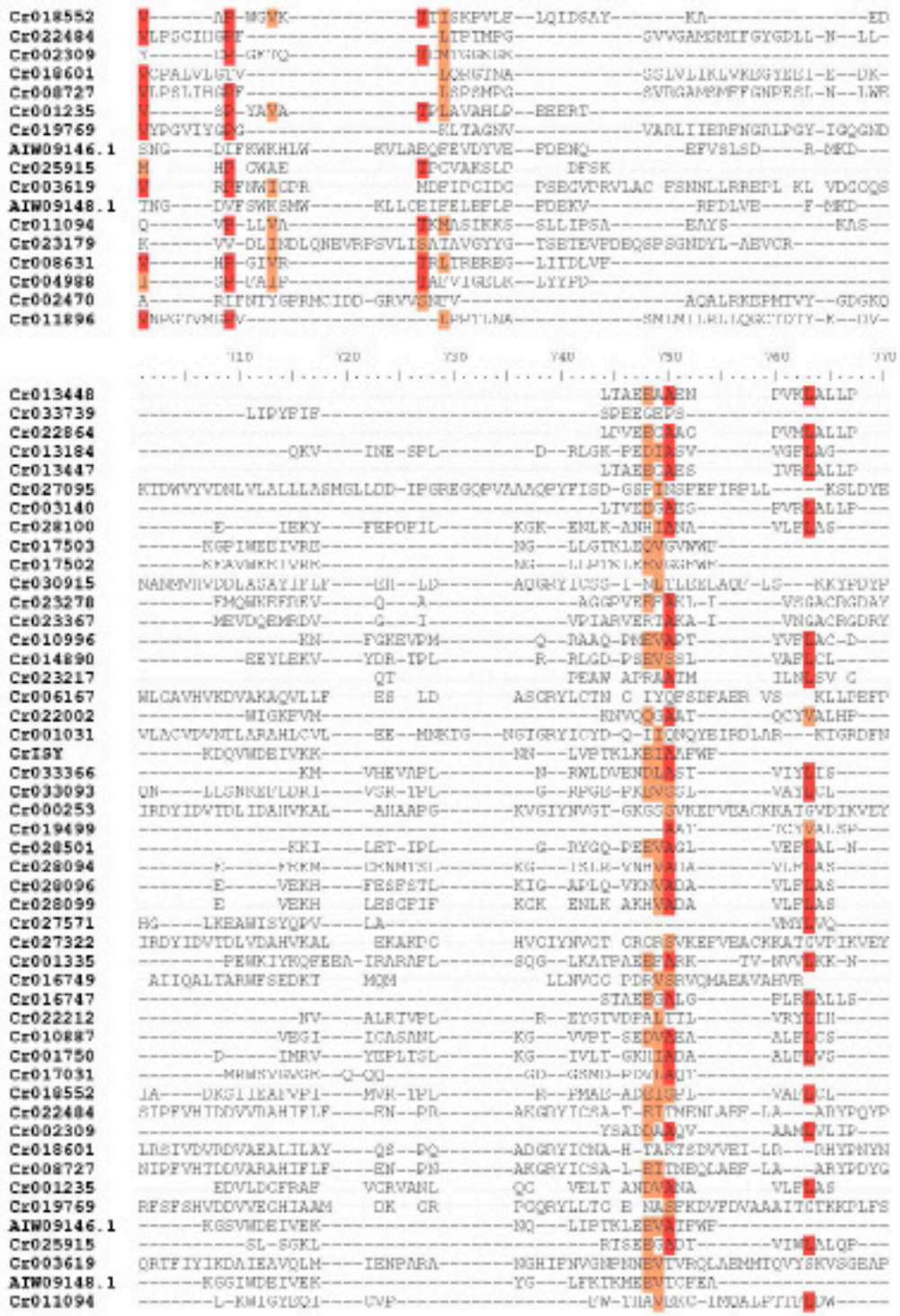

Cr010887
Cr001750
Cr017031
Cr010552
Cr022484
Cr002309
Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09148.1
Cr011094
Cr023179
Cr008631
Cr004988
Cr002470
Cr011096
290
300
310
320
330
340
350
Cr013448
Cr033739
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140
Cr020100
Cr017503
Cr017502
Cr030915
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099
Cr027571
Cr027322
Cr001335
Cr016749
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022484
Cr002309
Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915

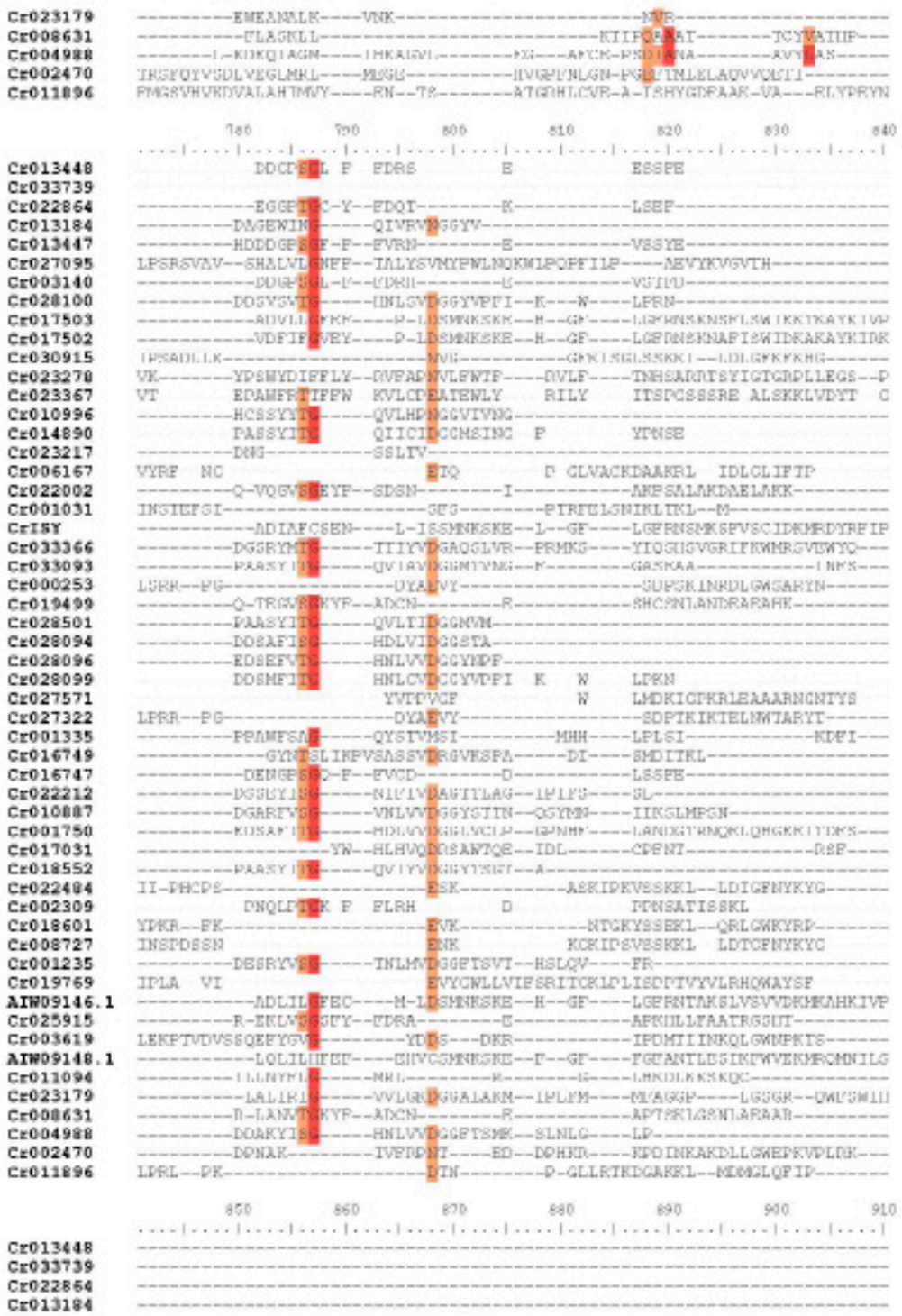



Cr033739
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502
Cr030915
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099
Cr027571
Cr027322
Cr001335
Cr016749
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022484
Cr002309
Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09140.1
Cr011094
Cr023179
Cr008631
Cr004988
Cr002470
Cr011896
500
510
520
530
540
550
560
Cr013448
Cr033739
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502
Cr030915
Cr023278
Cr023367

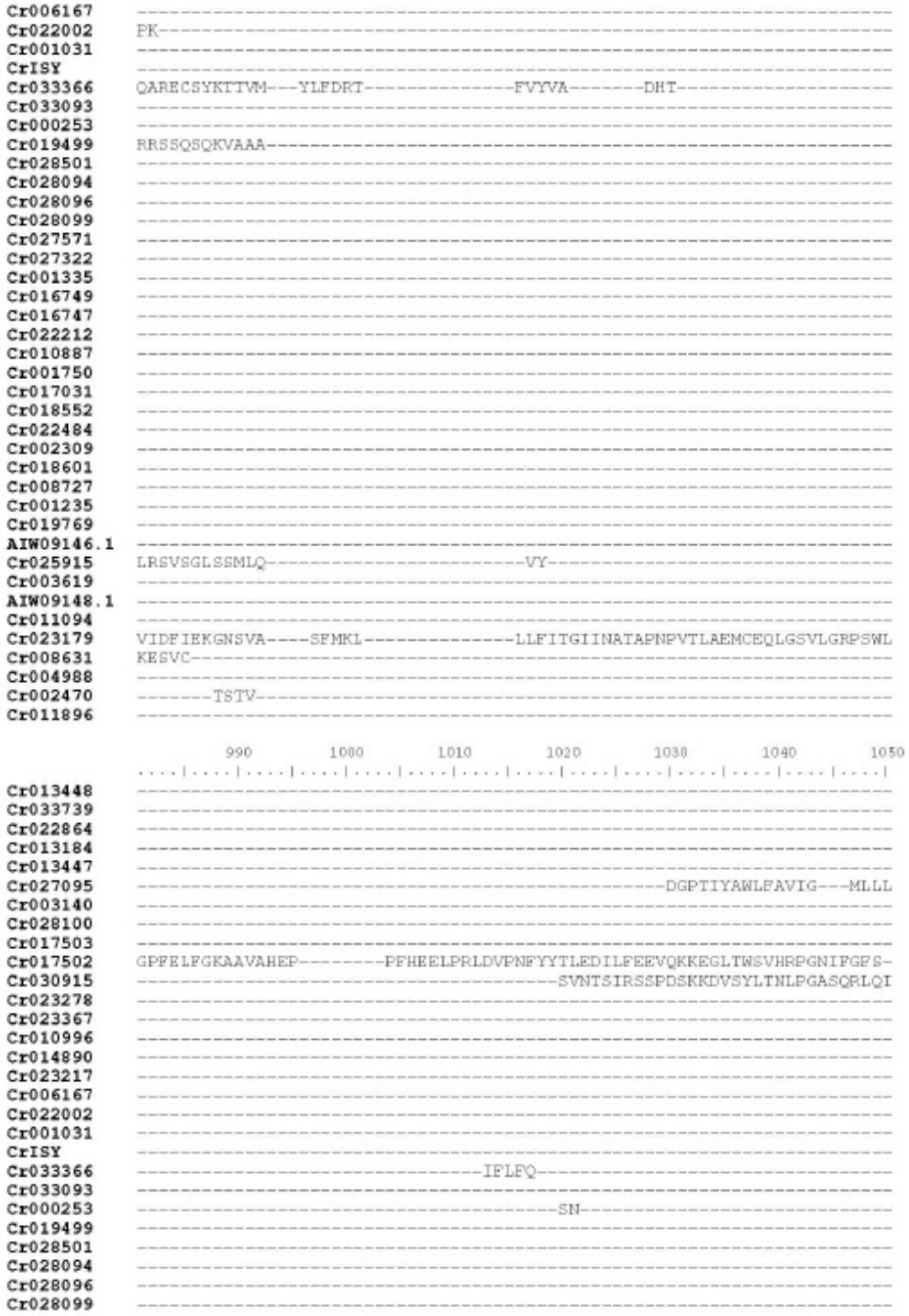

Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrIBY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099
Cr027571
Cr027322
Cr001335
Cr016749
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022484
Cr002309
Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09148.1
Cr011094
Cr023179
Cr008631
Cr004988
Cr002470
Cr011896
570
580
590
600
610
620
630
Cr013440
Cr033739
Cr022064
Cr013184
Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502
Cr030915
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrIBY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501





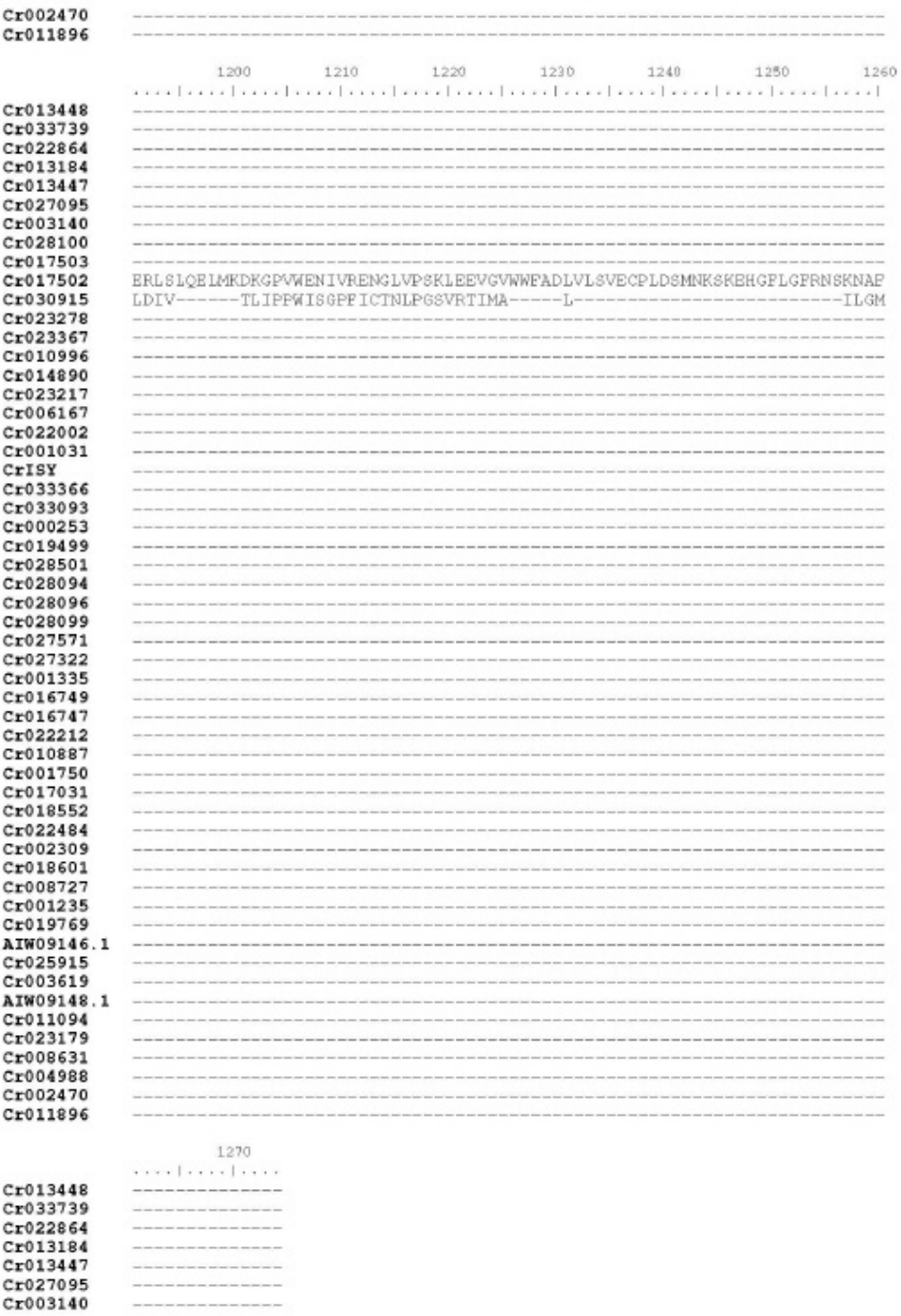

Cr023179
Cr008631
Cr004988
Cr002470
Cr011896
780
790
800
810
820
830
840
Cr013448
Cr033739
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502
Cr030915
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099
Cr027571
Cr027322
Cr001335
Cr016749
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022494
Cr002309
Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09148.1
Cr011094
Cr023179
Cr008631
Cr004988
Cr002470
Cr011896
850
860
870
880
890
900
910
Cr013448
Cr033739
Cr022864
Cr013184


Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502
Cr030915
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099
Cr027571
Cr027322
Cr001335
Cr016749
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022484
Cr002309
Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09148.1
Cr011094
Cr023179
Cr008631
Cr004988
Cr002470
Cr011896
920
930
940
950
960
970
980
Cr013448
Cr033739
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502
Cr030915
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217


Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099
Cr027571
Cr027322
Cr001335
Cr016749
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022484
Cr002309
Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09148.1
Cr011094
Cr023179
Cr008631
Cr004988
Cr002470
Cr011896
990
1000
1010
1020
1030
1040
1050
Cr013448
Cr033739
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502
Cr030915
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099


Cr027571
Cr027322
Cr001335
Cr016749
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022484
Cr002309
Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09148.1
Cr011094
Cr023179
Cr008631
Cr004988
Cr002470
Cr011896
1060
1070
1080
1090
1100
1110
1120
Cr013448
Cr033739
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502
Cr030915
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099
Cr027571
Cr027322
Cr001335
Cr016749
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022484
Cr002309


Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09148.1
Cr011094
Cr023179
Cr008631
Cr004988
Cr002470
Cr011896
1130
1140
1150
1160
1170
1180
1190
Cr013448
Cr033739
Cr022864
Cr013184
Cr013447
Cr027095 -------EAIYAWQLA------RKVDPANAKGWFWQTPA--------LGFFSLRFLLRRA--KKTSDK--
Cr003140
Cr028100
Cr017503
Cr017502 SDADLIAE-HQIWAAVDPYAKNEAFNVSNGDVFKWKHFW----------KVLAEQFGAEYAEYQG
Cr030915 LNAKTVKRFVYTSSVAAVMFDNRGLNIVD--ENMWSDLDNARKLNSLFGIYTITKTLAEKACLEFGEKHG
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099
Cr027571
Cr027322
Cr001335
Cr016749
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022484
Cr002309
Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09148.1
Cr011094
Cr023179
Cr008631
Cr004988


Cr002470
Cr011896
1200 1210 1220 1230 1240 1250 1260
Cr013448
Cr033739
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140
Cr028100
Cr017503
Cr017502 ERLSLQELMKDKGPVWENIVRENGLVPSKLEEVGVWWFADLVLSVECPLDSMNKSKEHGPLGFRNSKNAP
Cr030915 LDIV------TLIPPWISGPFICTNLPGSVRTIMA-----L-------------------------ILGM
Cr023278
Cr023367
Cr010996
Cr014890
Cr023217
Cr006167
Cr022002
Cr001031
CrISY
Cr033366
Cr033093
Cr000253
Cr019499
Cr028501
Cr028094
Cr028096
Cr028099
Cr027571
Cr027322
Cr001335
Cr016749
Cr016747
Cr022212
Cr010887
Cr001750
Cr017031
Cr018552
Cr022484
Cr002309
Cr018601
Cr008727
Cr001235
Cr019769
AIW09146.1
Cr025915
Cr003619
AIW09148.1
Cr011094
Cr023179
Cr008631
Cr004988
Cr002470
Cr011896
1270
Cr013448
Cr033739
Cr022864
Cr013184
Cr013447
Cr027095
Cr003140

```
Cr028100    ---------------
Cr017503    ---------------
Cr017502    ISWIDKAKAYKIFP
Cr030915    FSWSSTHPS-----
Cr023278    ---------------
Cr023367    ---------------
Cr010996    ---------------
Cr014890    ---------------
Cr023217    ---------------
Cr006167    ---------------
Cr022002    ---------------
Cr001031    ---------------
CrISY       ---------------
Cr033366    ---------------
Cr033093    ---------------
Cr000253    ---------------
Cr019499    ---------------
Cr028501    ---------------
Cr028094    ---------------
Cr028096    ---------------
Cr028099    ---------------
Cr027571    ---------------
Cr027322    ---------------
Cr001335    ---------------
Cr016749    ---------------
Cr016747    ---------------
Cr022212    ---------------
Cr010887    ---------------
Cr001750    ---------------
Cr017031    ---------------
Cr018552    ---------------
Cr022484    ---------------
Cr002309    ---------------
Cr018601    ---------------
Cr008727    ---------------
Cr001235    ---------------
Cr019769    ---------------
AIW09146.1  ---------------
Cr025915    ---------------
Cr003619    ---------------
AIW09148.1  ---------------
Cr011094    ---------------
Cr023179    ---------------
Cr008631    ---------------
Cr004988    ---------------
Cr002470    ---------------
Cr011896    ---------------
```

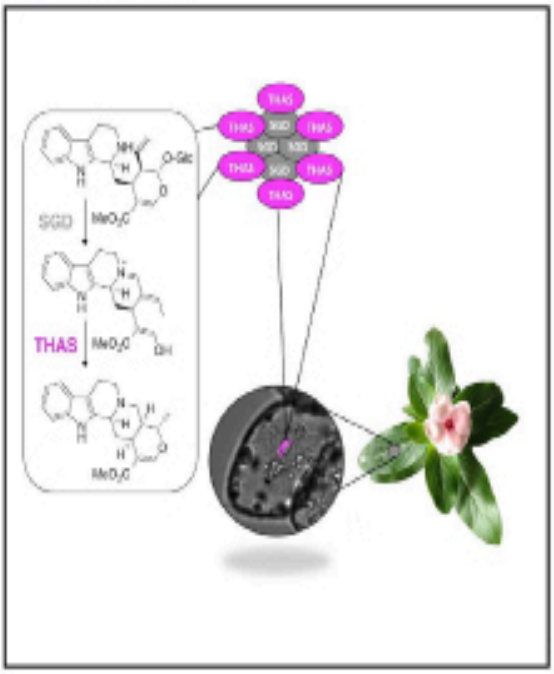
Annex 1

Brief Communication

Chemistry & Biology

Unlocking the Diversity of Alkaloids in *Catharanthus roseus*: Nuclear Localization Suggests Metabolic Channeling in Secondary Metabolism

Graphical Abstract



Authors

Anna Stavrinides,
Evangelos C. Tatsis, ...,
Vincent Courdavault,
Sarah E. O'Connor

Correspondence

sarah.oconnor@jic.ac.uk (S.E.O.),
vincent.courdavault@univ-tours.fr (V.C.)

In Brief

How plants transform the central biosynthetic intermediate strictosidine into thousands of divergent alkaloids has remained unresolved. Stavrinides et al. discover a nuclear-localized alcohol dehydrogenase homolog responsible for conversion of strictosidine aglycone to tetrahydroalstonine that appears to interact with an upstream pathway enzyme.

Highlights

- Tetrahydroalstonine synthase catalyzes the formation of a plant-derived alkaloid
- Tetrahydroalstonine synthase is localized to the nucleus
- Tetrahydroalstonine synthase and the preceding pathway enzyme interact
- Discovery of a gene controlling structural diversity of monoterpene indole alkaloids

Stavrinides et al., 2015, Chemistry & Biology 22, 1–6
March 19, 2015 ©2015 The Authors
http://dx.doi.org/10.1016/j.chembiol.2015.02.006

CellPress

Please cite this article in press as: Stavrinides et al., Unlocking the Diversity of Alkaloids in *Catharanthus roseus*: Nuclear Localization Suggests Metabolic Channeling in Secondary Metabolism, Chemistry & Biology (2015), http://dx.doi.org/10.1016/j.chembiol.2015.02.005

Chemistry & Biology

Brief Communication

Unlocking the Diversity of Alkaloids in *Catharanthus roseus*: Nuclear Localization Suggests Metabolic Channeling in Secondary Metabolism

Anna Stavrinides,[1] Evangelos C. Tatsis,[1] Emilien Foureau,[2] Lorenzo Caputi,[1] Franziska Kellner,[1] Vincent Courdavault,[2,*] and Sarah E. O'Connor[1,*]

[1]Department of Biological Chemistry, The John Innes Centre, Colney, Norwich NR4 7UH, UK

[2]Université François Rabelais de Tours, EA2106 "Biomolécules et Biotechnologies Végétales", 37200 Tours, France

*Correspondence: sarah.oconnor@jic.ac.uk (S.E.O.), vincent.courdavault@univ-tours.fr (V.C.)

http://dx.doi.org/10.1016/j.chembiol.2015.02.005

SUMMARY

The extraordinary chemical diversity of the plant-derived monoterpene indole alkaloids, which include vinblastine, quinine, and strychnine, originates from a single biosynthetic intermediate, strictosidine aglycone. Here we report for the first time the cloning of a biosynthetic gene and characterization of the corresponding enzyme that acts at this crucial branchpoint. This enzyme, an alcohol dehydrogenase homolog, converts strictosidine aglycone to the heteroyohimbine-type alkaloid tetrahydroalstonine. We also demonstrate how this enzyme, which uses a highly reactive substrate, may interact with the upstream enzyme of the pathway.

INTRODUCTION

The monoterpene indole alkaloids (MIAs) are a highly diverse family of natural products that are produced in a wide variety of medicinal plants. Over 3000 members of this natural product class, which includes compounds such as quinine, vinblastine, reserpine, and yohimbine, are derived from a common biosynthetic intermediate, strictosidine aglycone (O'Connor and Maresh, 2006). How plants transform strictosidine aglycone into divergent structural classes has remained unresolved.

The recent availability of transcriptome and genome data has dramatically accelerated the rate at which new plant biosynthetic genes are discovered. All genes that lead to strictosidine aglycone have been recently cloned from the well-characterized medicinal plant Catharanthus roseus, which produces over 100 MIAs (De Luca et al., 2014). However, gene products that act on strictosidine aglycone have not been identified in any plant, despite decades of effort. Attempts have been hampered in part by the reactivity and instability of strictosidine aglycone. In *C. roseus*, there are at least two major pathway branches derived from strictosidine aglycone (O'Connor and Maresh, 2006). One pathway is hypothesized to lead to the aspidosperma and the iboga classes to yield the precursors of vinblastine, while the other is expected to lead to alkaloids of the heteroyohimbine type (Figure 1A). These alkaloids have diverse biological activities: vinblastine is used as an anticancer agent (Kaur et al., 2014) and the heteroyohimbines have a range of pharmacological uses (Costa-Campos et al., 1998; Elisabetsky and Costa-Campos, 2006). While it is unknown how many *C. roseus* enzymes use strictosidine aglycone as a substrate, there is clearly more than one enzyme that acts at this crucial branchpoint.

The biochemical pathway leading from strictosidine aglycone to the heteroyohimbine alkaloids has been previously investigated using both crude plant extracts and biomimetic chemistry. Reduction of strictosidine aglycone with $NaBH_4$ or $NaCNBH_3$ yielded the heteroyohimbines ajmalicine (raubasine), tetrahydroalstonine, and 19-epi-ajmalicine, which differ only in the stereochemical configuration at carbons 15, 19, and 20, in various ratios (Figure 1B) (Brown et al., 1977; Kan-Fan and Husson, 1978, 1979, 1980). These three diastereomers were again observed, also in varying relative amounts, when crude *C. roseus* protein extracts were incubated with strictosidine aglycone and NADPH, but not in the absence of NADPH (Ruffer et al., 1979; Stoeckigt et al., 1976, 1977, 1983; Zenk, 1980). Collectively, these observations indicate that the heteroyohimbines result directly from the reduction of strictosidine aglycone and that an NADPH-dependent enzyme is implicated in this process. However, no gene encoding such an enzyme has been identified. Here we report the discovery of a reductase that converts strictosidine aglycone to the heteroyohimbine alkaloid tetrahydroalstonine.

RESULTS AND DISCUSSION

Given that heteroyohimbine biosynthesis likely requires reduction of an iminium present in strictosidine aglycone (Figure 1B), we used a publically available RNA-seq database that we recently generated (Gongora-Castillo et al., 2012) to search for *C. roseus* candidates displaying homology to enzyme classes known to reduce the carbonyl functional group. The alcohol dehydrogenases (ADHs), enzymes that reduce aldehydes and ketones to alcohols, were chosen as the initial focus. As part of a screen of ADHs that are upregulated in response to methyl jasmonate (Gongora-Castillo et al., 2012), a hormone known to upregulate alkaloid biosynthesis, we identified a candidate annotated as sinapyl alcohol dehydrogenase (Supplemental Information). When heterologously expressed and purified from *E. coli* (Figure S1), and assayed with strictosidine aglycone and NADPH, this candidate yielded a product with a mass consistent with a heteroyohimbine (m/z 353.186), thereby implicating this

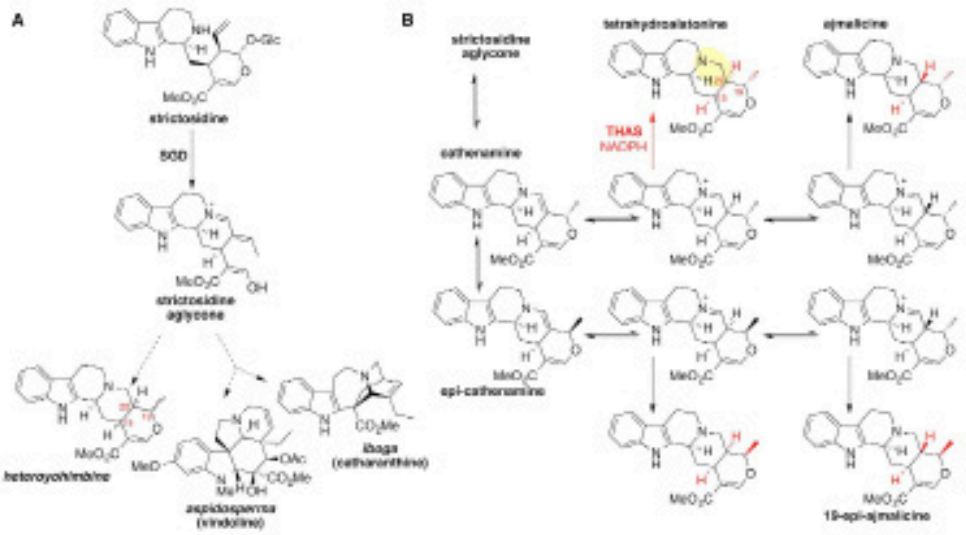


Figure 1. The Monoterpene Indole Alkaloids

(A) Representative monoterpene indole alkaloids derived from strictosidine and strictosidine aglycone found in Catharanthus roseus.

(B) Heteroyohimbine biosynthesis.

enzyme in the important structural branchpoint of the MIA biosynthetic pathway (Figure 2A).

To determine the identity of the alkaloid product, the enzyme was incubated with purified strictosidine (4.3 mg) in the presence of strictosidine glucosidase (SGD), which generated strictosidine aglycone in situ to best mimic physiologically relevant conditions. The major product (approximately 1 mg) was isolated by preparative thin-layer chromatography and exhibited an ^{1}H-NMR and ^{13}C-NMR spectrum matching an authentic standard of tetrahydroalstonine (Figure 2B; Figure S2). Hemscheidt and Zenk (1985) previously reported the isolation of an enzyme that produced tetrahydroalstonine, although this protein was purified only 35-fold from *C. roseus* cell cultures. Consistent with Hemscheidt and Zenk's (1985) nomenclature, we named this enzyme tetrahydroalstonine synthase (THAS). A minor enzymatic product was produced in yields too low for NMR characterization, but had a mass and R_t value consistent with ajmalicine, a stereoisomer of tetrahydroalstonine (Figure S2). When applied to normal phase liquid chromatography conditions, ajmalicine and tetrahydroalstonine could be resolved, indicating that the enzyme produces approximately 85% tetrahydroalstonine (Figure 3; Supplemental Information). We also silenced this gene in *C. roseus* seedlings using virus-induced gene silencing (VIGS) (Liscombe and O'Connor, 2011). LC-mass spectrometry (MS) analysis of the silenced leaf tissue showed a statistically significant decrease (approximately 50%) of a peak with a mass and retention time consistent with a heteroyohimbine, suggesting that this enzyme is involved in this biosynthetic pathway branch in vivo (Figure S2). A 50% reduction in product levels upon silencing has been observed for other physiologically relevant biosynthetic genes using the VIGS approach in both *C. roseus* (Asada et al., 2013; Geu-Flores et al., 2012) and another well-studied medicinal plant, opium poppy (Desgagné-Penix and Facchini, 2012; Chen and Facchini, 2014). Therefore, THAS is likely a major producer of tetrahydroalstonine in vivo, although additional, undiscovered *C. roseus* enzymes could also contribute to production of this compound. While we could not resolve tetrahydroalstonine and its stereoisomer ajmalicine in the silenced crude extracts, the levels of the ajmalicine-derived alkaloid serpentine remain the same, suggesting that silencing of THAS does not substantially affect ajmalicine levels and consequently that THAS does not play a major role in the biosynthesis of ajmalicine in planta.

Small-scale assays using LC-MS to monitor product formation indicated that NADPH was required for the reaction, although NADH could also be utilized (Figure S1). Efforts to accurately measure the steady state kinetic constants of this enzyme were complicated because strictosidine aglycone reacts with nucleophiles, opening the possibility that the substrate reacts with components in the reaction or the enzyme. This reactivity has already been associated with a plant defense mechanism involving strictosidine aglycone-mediated aggregation of proteins in *C. roseus* (Guinmand et al., 2010). Nevertheless, we obtained estimated K_m and k_{cat} values (Figure S1). To support these kinetic data, we also performed isothermal titration calorimetry (ITC) with THAS in the presence of NADPH and strictosidine aglycone. Titration of THAS with NADPH indicated that the co-substrate binds first with a K_d of 1.5 ± 0.1 μM (ΔH (cal/mol)

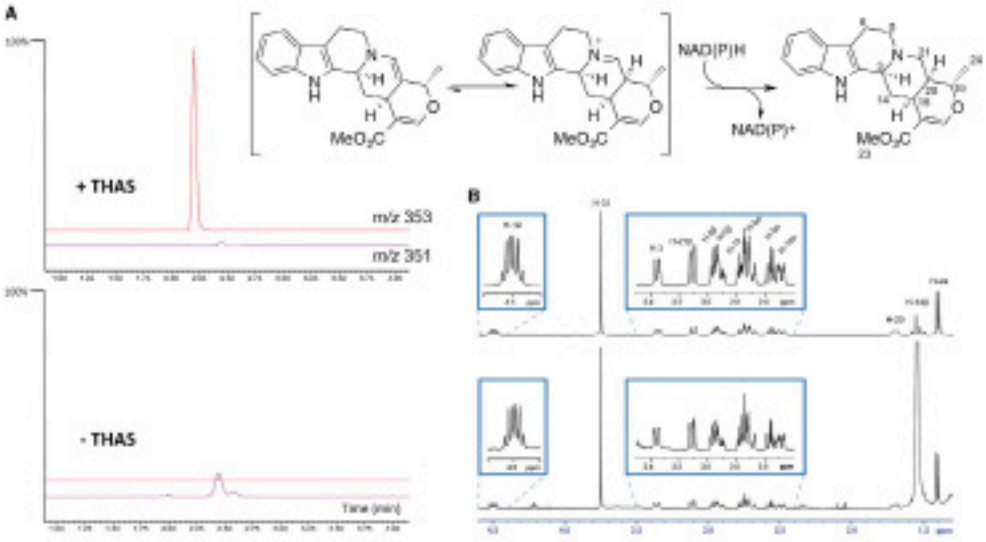


Figure 2. Activity Assays of THAS

Enzyme reactions were performed at 25°C for 30 min and assayed using a mass spectrometer in tandem with ultraperformance liquid chromatography.

(A) The total ion chromatogram for *m/z* 353 (red trace) and *m/z* 351 (purple trace) from 1 to 4 min is shown. Top trace: THAS (50 nM), SGD (6 nM), strictosidine (200 μM), NADPH (200 μM); bottom trace: same reaction in the absence of THAS. The *y* axis represents normalized ion abundance as a percentage relative to 1.00e^6 detected by selected ion monitoring at *m/z* 353 and 351.

(B) Portion of the ^{1}H-NMR spectrum of the isolated enzymatic product compared with an authentic standard of tetrahydroalstonine.

2310 ± 123.2; ΔS (cal/mol/deg) 34.2 ± 0.3) (Figure S1). The aglycone substrate does not appear to bind in the absence of NADPH, suggesting that the enzyme utilizes an ordered binding mechanism in which NADPH binds first. However, titration of the THAS-NADPH complex with strictosidine aglycone led to formation of a precipitate when concentrations of strictosidine aglycone exceeded 60 μM, preventing calculation of an accurate K_d. Collectively, the ITC data for THAS are consistent with an ordered Bi-Bi mechanism, a kinetic mechanism that has been reported for similar ADHs such as cinnamyl alcohol dehydrogenase (Chartier and Rapp, 2000; Lee et al., 2013).

The amino acid sequence of THAS was subjected to a BLAST alignment against the *C. roseus* transcriptome (Gongora-Castillo et al., 2012), as well as the NCBI (Figure S3). The closest characterized homologs of THAS are sinapyl alcohol dehydrogenase (*Populus tremuloides*, 64% amino acid identity), cinnamyl alcohol dehydrogenase (*Populus tomentosa*, 64%) and 8-hydroxygeraniol dehydrogenase (*C. roseus*, 63%), which are zinc-containing medium chain ADHs (Somati and Noel, 2005; Lee et al., 2013).

Strictosidine aglycone can rearrange into several isomers (Figure 1B), and while it has been reported that the dominant isomer is cathenamine (Gerasimenko et al., 2002; Stoeckigt et al., 1977), equilibration in solution with other isomers occurs (Brown and Leonard, 1979; Stoeckigt et al., 1983). Reduction of cathenamine or epi-cathenamine (Figure 1B) by a reductase would require reduction of the carbon-carbon double bond of an enamine; alternatively, Stoeckigt et al. (1983) and Zenk (1980) suggested that the iminium isomer is reduced (Figure 1B). THAS may catalyze the stereoselective formation of tetrahydroalstonine by selectively binding the correct isomer of the substrate for reduction, thereby relying on the inherent propensity for the enamine and imine to tautomerize under physiological conditions. Given that three diastereomers, ajmalicine, tetrahydroalstonine, and 19-epi-ajmalicine, can be obtained from chemical reduction of strictosidine aglycone, this is a chemically reasonable proposal. An alternative hypothesis is that THAS catalyzes enamine-imine tautomerization in addition to reduction. The difficulties associated with obtaining accurate kinetic data in this system, as well as the inherent reactivity of the strictosidine aglycone, make answering these questions using enzymology approaches challenging. However, identification and comparison with enzymes that generate other heteroyohimbine diastereomers will likely provide the basis for a more definitive mechanism of product specificity.

Recent research has highlighted that plant secondary metabolite biosynthetic pathways often are compartmentalized in different subcellular locations. While microscopy experiments have demonstrated that most of the early steps of monoterpene indole alkaloid biosynthesis in *C. roseus* take place in the cytosol (Courdavault et al., 2014), the enzyme that synthesizes strictosidine is located in the vacuole, and the enzyme SGD, which deglycosylates strictosidine, contains a nuclear localization signal and is in the nucleus, a highly unusual site for secondary

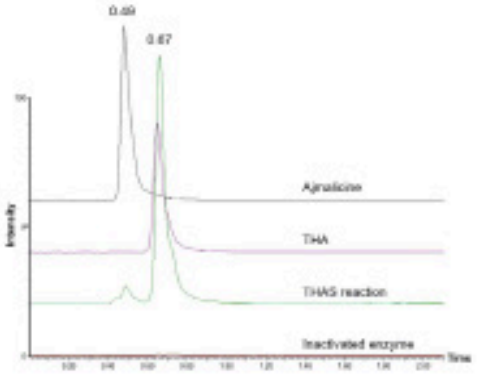


Figure 3. LC-MS Performed under Normal Phase Conditions (Hydrophilic Interaction Liquid Chromatography) Showing Separation of Ajmalicine (Retention Time of 0.49 min) and Tetrahydroalstonine (THA, Retention Time 0.67 min)

THAS produces approximately 95% of the tetrahydroalstonine (THA) diastereomer. The y axis represents normalized ion abundance as a percentage detected by selected ion monitoring at m/z 353.

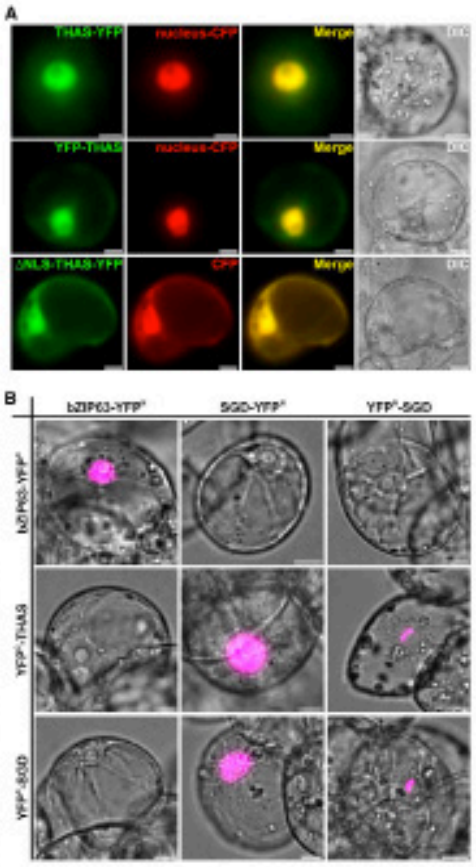


Figure 4. THAS Is Targeted to the Nucleus via a Monopartite Nuclear Localization Signal (NLS) and Interacts with SGD

(A) *C. roseus* cells were transiently cotransformed with plasmids expressing either THAS-YFP (upper row), YFP-THAS (middle row), or the NLS deleted version of THAS (lower row) and plasmids encoding the nuclear CFP marker or the nucleocytosolic CFP marker (second column). Colocalization of the fluorescence signals appears in yellow when merging the two individual (green/red) false color images (third column). Cell morphology is observed with differential interference contrast (DIC) (fourth column).

(B) THAS and SGD interactions were analyzed by BiFC in *C. roseus* cells transiently transformed by plasmids encoding fusions indicated on the top (fusion with the split YFP^N fragment) and on the left (fusion with split YFP^C fragment). bZIP63 was used as a positive BiFC control and to evaluate the specificity of THAS and SGD interactions. The images are merges of the YFP BiFC channel (magenta false color) with the DIC channel to show the nuclear localization of the interactions. Bars, 10 μm.

metabolite biosynthesis (Guirimand et al., 2010). Notably, a motif resembling a class V nuclear localization sequence (Kosugi et al., 2008) was observed in THAS ($K_{214}K_{215}K_{216}R_{217}$). Microscopy of *C. roseus* cells transformed with YFP-tagged THAS confirmed the nuclear location of this enzyme, while deletion of the KKKR sequence disrupted the localization (Figure 4A; Figure S4). This is one of the very few examples of secondary metabolism that is localized to the nucleus (Saslowsky et al., 2005).

Given the reactivity of strictosidine aglycone (Guirimand et al., 2010), metabolic channeling via a protein-protein interaction between SGD and the enzyme immediately downstream may be necessary to protect the substrate. Pull down experiments between SGD and THAS gave partially positive but inconclusive results (Figure S4). However, when we used bimolecular fluorescence complementation (BiFC) in *C. roseus* cells, we observed an interaction between SGD and THAS (Figure 4B). While this interaction generated a diffuse nuclear fluorescent signal when the C-terminal end of SGD was fused to the split-YFP fragment, a sickle-shaped signal was observed when both SGD and THAS were expressed with free C-terminal ends (YFP^N-SGD and YFP^C-THAS). Such a signal was also observed for SGD self-interactions (Guirimand et al., 2010) and likely results from the formation of SGD complexes over 1.5 MDa (Lopendijk et al., 1998). Similar experiments with SGD and an upstream MIA biosynthetic enzyme, loganic acid methyl transferase, failed to show an interaction, highlighting the specificity of this interaction (Figure S4). The fact that THAS interacts with SGD provides further support for the physiological relevance of THAS in planta. As strictosidine aglycone is reactive and most likely toxic in vivo, it has been proposed that this molecule is produced by the plant in response to attack (Guirimand et al., 2010). The nuclear localization of THAS might be an evolutionary mechanism designed to channel this molecule into a more stable product when no such defense is required. Identification of additional nuclear-localized biosynthetic enzymes in *C. roseus* and other heteroyohimbine

producing plants may provide more insight into the reasons for this unusual localization pattern.

SIGNIFICANCE

Many of the monoterpene indole alkaloid structural classes are generated at the SGD junction. Here we report the first identification of a biosynthetic gene that acts directly downstream of SGD. The enzyme, an ADH homolog, generates a heteroyohimbine alkaloid by reducing one of the isomers of strictosidine aglycone. Unusually, this enzyme is located in the nucleus and may interact with its upstream partner, SGD. The discovery of the THAS gene represents the completion of a major branch of monoterpene indole alkaloid biosynthesis, which will now allow reconstruction of heteroyohimbines and heteroyohimbine analogs in heterologous hosts. This discovery is a crucial first step in understanding how the structural diversity of MIAs is controlled.

EXPERIMENTAL PROCEDURES

The THAS gene (accession number KM974758) was cloned into pDONR and expressed in Rosetta 2 pLysS *E. coli* cells (Novagen) with induction of expression with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside. Cultures were grown at 18°C for 18 hr, with shaking at 200 rpm. His tagged THAS was purified using a HisTrap FF 5-ml column (GE Healthcare). SGD expression and purification was done as described for THAS using the expression system described previously by Yerkes et al. (2005). Purified THAS and SGD were used in all assays. Strictosidine was enzymatically synthesized from tryptamine and a crude methanol extract of secologanin (*Symphoricarpos albus*) enriched in secologanin prepared as previously described (Geerlings et al., 2001). Strictosidine aglycone was generated in situ prior to addition of THAS by incubation of strictosidine and SGD in the appropriate solution for 10 min, at which time strictosidine was completely converted to the aglycone.

Steady state kinetic analyses were performed with 50 nM THAS and 8 nM SGD, 50 m*M* sodiumphosphate buffer (pH 7.5), 200 μ*M* $NADPH$, and an internal sulfonic standard (50 μ*M*). All LC-MS measurements were performed on ACQUITY ultra-performance liquid chromatography with a Xevo TQ-S mass spectrometer.

For VIGS, a 330-bp fragment of THAS was cloned into the pTRV2u vector as described (Caro-Flores et al., 2012). The resulting pTRV2u-THAS construct was used to silence THAS in *C. roseus* seedlings essentially as described (Liscombe and O'Connor, 2011).

The subcellular localization of THAS was studied by creating fluorescent fusion proteins using the pSCA-cassette YFP plasmid (Guirimand et al., 2009, 2010). The capacity of interaction of THAS and SGD was characterized by BiFC assays using THAS PCR product cloned via SpeI into the pGPVCS(VR) plasmid (Waadt et al., 2008), which allows expression of THAS fused to the carboxy-terminal extremity of the split YFP^C fragment (YFP^C-THAS). The pSCA-SPYNE173-SGD and pSPYNE(R)173-SGD plasmids (Guirimand et al., 2010) were used to express SGD fused to the amino-terminal or carboxy-terminal extremity of the split YFP^N fragment (SGD-YFP^N and YFP^N-SGD, respectively). THAS self-interactions were analyzed via additional cloning of the THAS PCR product into the pSCA-SPYNE173 and pSCA-SPYCE(M) plasmids (Guirimand et al., 2010) to express THAS-YFP^N and THAS-YFP^C, respectively. Transient transformation of *C. roseus* cells by particle bombardment and fluorescence imaging were performed following the procedures previously described (Guirimand et al., 2009, 2010).

Complete experimental details are included in the Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and four figures, and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2015.02.006.

AUTHOR CONTRIBUTIONS

A.S. made the initial discovery of THAS activity and conducted all enzyme assays, kinetics, pulldown, and ITC; E.C.T. performed VIGS and assisted in the structural characterization of the enzyme product; E.F. performed the microscopy experiments; L.C. assisted in the purification of THAS and pulldown; F.K. provided initial genomic data that assisted in identification of the THAS candidate; V.C. conceived, initiated, and supervised all localization and BiFC experiments; S.E.O. supervised all enzymology experiments; A.S., V.C., S.E.O. wrote the manuscript.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the CNC (S11383) and from the Région Centre (France, AGISAL grant). A.S. is supported by a GSSRC DTP studentship.

Received: November 7, 2014
Revised: January 24, 2015
Accepted: February 17, 2015
Published: March 19, 2015

REFERENCES

Asada, K., Salim, V., Masada-Atsumi, S., Edmunds, E., Nagatoshi, M., Terasaka, K., Mizukami, H., and De Luca, V. (2013). A 7-deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in Madagascar periwinkle. Plant Cell *25*, 4123–4134.

Bernati, S.K., and Nové, J.P. (1995). Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. Plant Cell *7*, 1658–1611.

Brown, R.T., and Leonard, J. (1979). Biomimetic synthesis of catharanthine and 19-epicatharanthine, key intermediates to heteroyohimbine alkaloids. J. Chem. Soc. Chem. Commun. 877–879.

Brown, R.T., Leonard, J., and Sleigh, S.K. (1977). 'One pot' biomimetic synthesis of 19*S* heteroyohimbine alkaloids. J. Chem. Soc. Chem. Commun. 638–638.

Charlier, H.A., and Plapp, B.V. (2000). Kinetic cooperativity of human liver alcohol dehydrogenase γ2. J. Biol. Chem. *275*, 11569–11575.

Chen, X., and Facchini, P.J. (2014). Short-chain dehydrogenase/reductase catalyzing the final step of noscapine biosynthesis is localized to laticifers in opium poppy. Plant J. *77*, 173–184.

Costa-Campos, L., Lara, D.R., Nunes, D.S., and Elisabetsky, E. (1999). Antipsychotic-like profile of akuammine. Pharmacol. Biochem. Behav. *63*, 133–141.

Courdavault, V., Papon, N., Clastre, M., Giglioli-Guivarc'h, N., St-Pierre, B., and Burlat, V. (2014). A look inside an alkaloid multisite plant: the Catharanthus logistics. Curr. Opin. Plant Biol. *19*, 43–50.

De Luca, V., Salim, V., Thamm, A., Masada, S.A., and Yu, F. (2014). Making iridoids/secoiridoids and monoterpenoid indole alkaloids: progress on pathway elucidation. Curr. Opin. Plant Biol. *19*, 35–42.

Dogagne Puree, I., and Facchini, P.J. (2012). Systematic silencing of benzylisoquinoline alkaloid biosynthetic genes reveals the major route to papaverine in opium poppy. Plant J. *72*, 331–344.

Elisabetsky, E., and Costa Campos, L. (2006). The alkaloid akuammine: a review of its pharmacological properties. Evidence-based complementary and alternative medicine. Evid. Based Complement. Alternat. Med. *3*, 39–48.

Geerlings, A., Redondo, F.J., Contin, A., Memelink, J., van der Heijden, R., and Verpoorte, R. (2001). Biotransformation of tryptamine and secologanin into plant terpenoid indole alkaloids by transgenic yeast. Appl. Microbiol. Biotechnol. *55*, 420–424.

Gerasimenko, I., Sheludko, Y., Ma, X., and Stöckigt, J. (2002). Heterologous expression of a *Rauvolfia* cDNA encoding strictosidine glucosidase, a biosynthetic key to over 1000 monoterpenoid indole alkaloids. Eur. J. Biochem. *269*, 2204–2213.

Geu-Flores, F., Sherden, N.H., Courdavault, V., Burlat, V., Glenn, W.S., Wu, C., Nims, E., Cui, Y., and O'Connor, S.E. (2012). An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. Nature *492*, 138–142.

Góngora-Castillo, E., Childs, K.L., Fedewa, G., Hamilton, J.P., Liscombe, D.K., Magallanes-Lundback, M., Mandadi, K.K., Nims, E., Runguphan, W., Vaillancourt, B., et al. (2012). Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. PLoS One *7*, e52506.

Guirimand, G., Burlat, V., Oudin, A., Lanoue, A., St-Pierre, B., and Courdavault, V. (2009). Optimization of the transient transformation of *Catharanthus roseus* cells by particle bombardment and its application to the subcellular localization of hydroxymethylbutenyl 4-diphosphate synthase and geraniol 10-hydroxylase. Plant Cell Rep. *28*, 1215–1234.

Guirimand, G., Courdavault, V., Lanoue, A., Mahroug, S., Guihur, A., Blanc, N., Giglioli-Guivarc'h, N., St-Pierre, B., and Burlat, V. (2010). Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? BMC Plant Biol. *10*, 182.

Hemscheidt, T., and Zenk, M.H. (1985). Partial purification and characterization of a NADPH dependent tetrahydroalstonine synthase from *Catharanthus roseus* cell suspension cultures. Plant Cell Rep. *4*, 216–219.

Kan-Fan, C., and Husson, H.P. (1978). Stereochemical control in the biomimetic conversion of heteroyohimbine alkaloid precursors. Isolation of a novel key intermediate. J. Chem. Soc. Chem. Commun. 618–619.

Kan-Fan, C., and Husson, H.P. (1979). Isolation and biomimetic conversion of 4,21-dehydrogeissoschizine. J. Chem. Soc. Chem. Commun. 1015–1016.

Kan-Fan, C., and Husson, H.P. (1980). Biomimetic synthesis of yohimbine and heteroyohimbine alkaloids from 4,21-dehydrogeissoschizine. Tetrahedron Lett. *21*, 1463–1466.

Kaur, R., Kaur, G., Gill, R.K., Soni, R., and Bariwal, J. (2014). Recent developments in tubulin polymerization inhibitors: an overview. Eur. J. Med. Chem. *87C*, 89–124.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Miyo, E., Tomita, M., and Yanagawa, H. (2009). Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J. Biol. Chem. *284*, 478–485.

Lee, C., Bedgar, D.L., Davin, L.B., and Lewis, N.G. (2013). Assessment of a putative proton relay in *Arabidopsis* cinnamyl alcohol dehydrogenase catalysis. Org. Biomol. Chem. *11*, 1127–1134.

Liscombe, D.K., and O'Connor, S.E. (2011). A virus induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*. Phytochemistry *72*, 1969–1977.

Luijendijk, T.J.C., Stevens, L.H., and Verpoorte, R. (1998). Purification and characterisation of strictosidine β-D-glucosidase from *Catharanthus roseus* cell suspension cultures. Plant Physiol. Biochem. *36*, 419–425.

O'Connor, S.E., and Maresh, J.J. (2006). Chemistry and biology of monoterpene indole alkaloid biosynthesis. Nat. Prod. Rep. *23*, 532–547.

Rueffer, M., Kan-Fan, C., Husson, H.P., Stoeckigt, J., and Zenk, M.H. (1979). 4,21-Dehydrogeissoschizine, an intermediate in heteroyohimbine alkaloid biosynthesis. J. Chem. Soc. Chem. Commun. 1016–1018.

Saslowsky, D.E., Warek, U., and Winkel, B.S. (2005). Nuclear localization of flavonoid enzymes in *Arabidopsis*. J. Biol. Chem. *280*, 23735–23740.

Stoeckigt, J., Treimer, J., and Zenk, M.H. (1976). Synthesis of ajmalicine and related indole alkaloids by cell free extracts of *Catharanthus roseus* cell suspension cultures. FEBS Lett. *70*, 267–270.

Stoeckigt, J., Husson, H.P., Kan-Fan, C., and Zenk, M.H. (1977). Cathenamine, a central intermediate in the cell free biosynthesis of ajmalicine and related indole alkaloids. J. Chem. Soc. Chem. Commun. 164–166.

Stoeckigt, J., Hemscheidt, T., Hoefle, G., Heinstein, P., and Formacek, V. (1983). Steric course of hydrogen transfer during enzymatic formation of 3α-heteroyohimbine alkaloids. Biochemistry *22*, 3448–3452.

Waadt, R., Schmidt, L.K., Lohse, M., Hashimoto, K., Bock, R., and Kudla, J. (2008). Multicolor bimolecular fluorescence complementation reveals simultaneous formation of alternative CBL/CIPK complexes in planta. Plant J. *56*, 505–516.

Yerkes, N., Wu, J.X., McCoy, E., Galan, M.C., Chen, S., and O'Connor, S.E. (2008). Substrate specificity and diastereoselectivity of strictosidine glucosidase, a key enzyme in monoterpene indole alkaloid biosynthesis. Bioorg. Med. Chem. Lett. *18*, 3095–3098.

Zenk, M.H. (1980). Enzymatic synthesis of ajmalicine and related indole alkaloids. J. Nat. Prod. *43*, 438–451.

Annex 2

nature COMMUNICATIONS

ARTICLE

Received 30 Dec 2015 | Accepted 31 May 2016 | Published 15 Jul 2016

DOI: 10.1038/ncomms12116 OPEN

Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity

Anna Stavrinides[1,*], Evangelos C. Tatsis[1,*], Lorenzo Caputi[1], Emilien Foureau[2], Clare E.M Stevenson[1], David M. Lawson[1], Vincent Courdavault[2] & Sarah E. O'Connor[1]

Plants produce an enormous array of biologically active metabolites, often with stereochemical variations on the same molecular scaffold. These changes in stereochemistry dramatically impact biological activity. Notably, the stereoisomers of the heteroyohimbine alkaloids show diverse pharmacological activities. We reported a medium chain dehydrogenase/reductase (MDR) from *Catharanthus roseus* that catalyses formation of a heteroyohimbine isomer. Here we report the discovery of additional heteroyohimbine synthases (HYSs), one of which produces a mixture of diastereomers. The crystal structures for three HYSs have been solved, providing insight into the mechanism of reactivity and stereoselectivity, with mutation of one loop interconverting product specificity. Localization and gene silencing experiments provide a basis for understanding the function of these enzymes *in vivo*. This work sets the stage to explore how MDRs evolved to generate structural and biological diversity in specialized plant metabolism and opens the possibility for metabolic engineering of new compounds based on this scaffold.

[1] The John Innes Centre, Department of Biological Chemistry, Norwich NR4 7UH, UK. [2] Université François-Rabelais de Tours, EA2106 'Biomolécules et Biotechnologies Végétales', Tours 37200, France. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to V.C. (email: vincent.courdavault@univ-tours.fr) or to S.E.O'C. (email: sarah.oconnor@jic.ac.uk).

Heteroyohimbines are a prevalent subclass of the monoterpene indole alkaloids (Corynanthe type skeleton), having been isolated from many plant species, primarily from the Apocynaceae and Rubiaceae families[1]. These alkaloids exhibit numerous biological activities: ajmalicine is an α1-adrenergic receptor antagonist[2–5], and mayumbine (19-epi-ajmalicine) is a ligand for the benzodiazepine receptor (Fig. 1)[6]. Oxidized beta-carboline heteroyohimbines also exhibit potent pharmacological activity: serpentine has shown topoisomerase inhibition activity[7] and alstonine has been shown to interact with 5-HT2A/C receptors and shows promise as an anti-psychotic agent[8–13]. In addition, heteroyohimbines are biosynthetic precursors of many oxindole alkaloids, which also display a wide range of biological activities[14]. Although a total of 16 heteroyohimbine stereoisomers are possible, only 8 are reported to be found in nature, at stereocentres C3, C19, C20 (Fig. 1)[14–18]. How and why the stereoselectivity is controlled in the biosynthesis of these alkaloids remains unclear.

The medicinal plant *Catharanthus roseus* produces three of these isomers, ajmalicine (raubasine), tetrahydroalstonine and 19-epi-ajmalicine (mayumbine) (Fig. 1)[21]. These heteroyohimbines, along with the majority of monoterpene indole alkaloids, are derived from deglycosylated strictosidine (strictosidine aglycone)[22]. The removal of a glucose unit from strictosidine by strictosidine glucosidase (SGD) forms a reactive dialdehyde intermediate that can rearrange to form numerous isomers[23]. The stabilization of these isomers by enzyme-catalyzed reduction is hypothesized to be the stepping stone for the extensive chemical diversity observed in the monoterpene indole alkaloids (Fig. 1)[21,22]. We recently reported the first cloning of a biosynthetic gene encoding an enzyme that acts on strictosidine aglycone. This zinc-dependent medium chain dehydrogenase/reductase (MDR), named tetrahydroalstonine synthase (THAS), produces the heteroyohimbine tetrahydroalstonine (Fig. 1)[24]. Although these studies demonstrated that THAS is a key enzyme in heteroyohimbine biosynthesis, the mechanism by which this enzyme controls the stereoselectivity of the reduction remained unknown. Moreover, it is important to note that strictosidine aglycone serves as the precursor for many alkaloid scaffolds, and therefore represents a central branch point in the monoterpene indole alkaloid biosynthetic pathway. Therefore, we set out to identify additional heteroyohimbine synthases (HYSs) with different stereochemical product profiles that would more clearly define how structural diversity, in this case the formation of different stereoisomers, is controlled in this system.

In this study, we assayed 14 MDR homologues identified from the *C. roseus* transcriptome[25,26] that have homology to THAS (Cr_024553). This screen revealed three additional enzymes with THAS activity (Cr_010119, Cr_021691, Cr_032583a), and, importantly, an enzyme that produced a mixture of heteroyohimbine diastereomers (Cr_032583b). Crystal structures of THAS (here referred to as THAS1), a second representative THAS (Cr_021691, THAS2) and the structure of the pseudoenzymatic homologue (Cr_032583b, HYS) were solved and mutants revealed key residues that control the stereochemistry of the product profiles. Notably, analysis of the subcellular localization of some of these HYSs indicates an unusual nuclear localization pattern and an interaction with the previous enzyme, SGD. These discoveries provide insight into the mechanism and evolution of a crucial branch point in a specialized metabolic pathway with both pharmacological and evolutionary importance.

Results

Discovery of HYSs. Guided by our initial discovery of THAS1 (ref. 24) we identified candidates from the MDR protein family in the *C. roseus* transcriptome[25,26] based on amino acid similarity to this enzyme (Supplementary Table 1). Each of these candidates was cloned from *C. roseus* cDNA and expressed in *Escherichia coli*, with the exception of Cr 017994, which could not be expressed and was not considered further. The remaining candidates were assayed with the substrate strictosidine aglycone, and product formation was monitored by liquid chromatography mass spectrometry (LC-MS). Of these, four (Cr_010119, Cr_021691, Cr_032583a, Cr_032583b) reduced strictosidine aglycone to a product corresponding to one of the heteroyohimbines (Fig. 2, Supplementary Fig. 1). The products of the enzymatic reactions were identified based on LC-MS data and comparison to authentic standards (Supplementary Fig. 2). Enzymes that failed to produce a heteroyohimbine product were not studied further (Supplementary Fig. 1). Three of the enzymes (Cr_021691, THAS2; Cr_010119, THAS3; Cr_032583a, THAS4) produced tetrahydroalstonine in ~85% yield, with small amounts of 19-epi-ajmalicine (mayumbine) (<15%) also observed in these reactions, similar to the previously reported THAS1. Notably, one enzyme (Cr_032583b, HYS) produced a dramatically different product profile consisting of a mixture of ajmalicine/tetrahydroalstonine/mayumbine (55:27:18, at pH 6) (Fig. 2). The discovery of this enzyme, HYS, now provides a molecular basis to understand the generation of stereochemical diversity in this alkaloid family.

Crystallography of three HYSs. To understand the mechanism of stereochemical control at this crucial biosynthetic branch point, we crystallized three HYSs. THAS1 and THAS2, which produce predominantly tetrahydroalstonine, were both crystallized, since their amino acid sequence identity is relatively low (55%) and the predicted active sites have numerous differences (Fig. 3). HYS, which has a distinctly different product profile, was also crystallized to explore the structural basis behind this distinct stereochemical outcome. Structures (Supplementary Table 3) were obtained for 1THAS1 and 1THAS2 with $NADP^+$ bound (THAS1, 1.05 Å resolution (Fig. 4a,b; Supplementary Figs 3–5); THAS2, 2.10 Å resolution (Fig. 4d, Supplementary Fig. 4)) and in apo form (THAS1, 2.25 Å resolution (Fig. 4c, Supplementary Fig. 5); THAS2, 2.05 Å resolution (Fig. 4c)), while HYS could only be crystallized in the apo form (2.25 Å resolution, Fig. 4c).

Structural features of HYS active sites. The five HYS structures described here are similar to cinnamyl alcohol dehydrogenase (SAD; PDB accession codes 1YQX and 1YQD) and the SAD homologue cinnamyl alcohol dehydrogenase (CAD; PDB accession codes 2CF5 and 2CF6)[27,28], which reduce the aldehyde moiety of lignin precursors. Indeed, pairwise superpositions of subunits from these structures gave RMSD values of <2 Å (Supplementary Table 4). The biological unit is an elongated homodimer, with each subunit divided into a substrate and cofactor-binding domain; the latter also being responsible for forming the dimer interface (Supplementary Fig. 3). The overall structure of THAS1, with active site, cofactor and loops highlighted, is shown in Supplementary Fig. 3.

The active site cavities of the HYSs are framed by helix α2, the catalytic zinc coordination sphere, and loops 1 and 2, with the NADP(H) co-substrate binding at the base of the active site (Figs 3 and 4a,b,d, , Supplementary Fig. 4). Loop 2, which is positioned above the active site, is highly variable in length and sequence (Figs 3 and 6d). In both THAS1 and THAS2, a network of amino acids holds $NADP^+$ in place. Most notably, Glu59 of 1THAS1 anchors NADP(H) through a bidentate interaction with both ribose hydroxyls, with His60 playing a comparable role in SAD, although here the interaction is with the 3′ OH only (Fig. 4b).

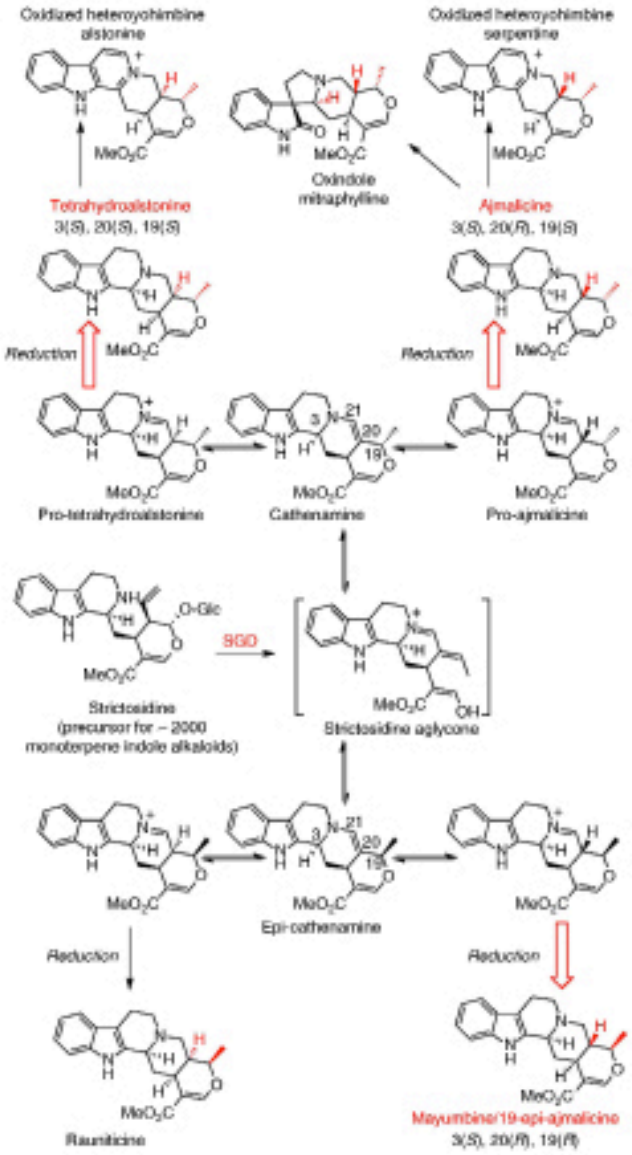


Figure 1 | Heteroyohimbine alkaloid biosynthesis. Heteroyohimbines with 3(S) stereochemistry derive from strictosidine aglycone. The three diastereomers found in *Catharanthus roseus*, are highlighted with red arrows. Alkaloids derived from heteroyohimbines are also shown.

Glu59 is conserved in HYS, but an aspartate residue (THAS3 and THAS2) or a tyrosine residue (THAS4) serves this role in other homologues. MDRs usually contain two zinc ions[30], a distal 'structural' zinc ion, which in this case is coordinated by four cysteine residues, and a proximal 'catalytic' zinc ion near the active site, which is coordinated by two cysteines, one histidine and one glutamate residue (Figs 3 and 4b)[27,31]. The proximal zinc of THAS1 is ~2 Å further away from the cofactor relative to SAD and thus may play no direct role in catalysis (Supplementary Fig. 4). However, it may have a function in maintaining the tertiary structure since three of the liganding residues are in the substrate-binding domain and the fourth is in the cofactor-binding domain. SAD/CAD utilize an active site serine that protonates the alkoxide that results from reduction of the aldehyde substrate[27,29]; this

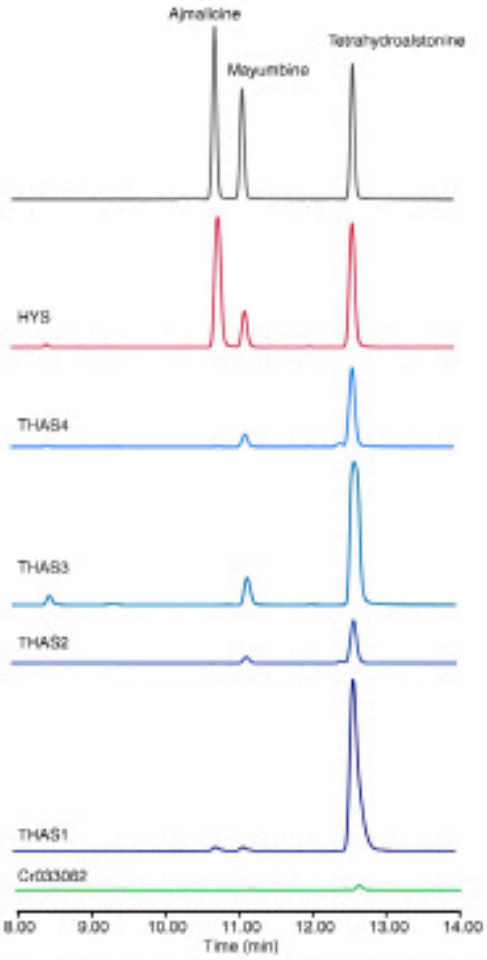


Figure 2 | LC-MS analysis of active MDR candidates against strictosidine aglycone. Cr033062 exhibited only trace activity. See Supplementary Fig. 1 for chromatograms of assays with inactive enzymes and negative controls.

serine is replaced with a tyrosine residue in THAS1 (Tyr56) and HYS (Tyr53) (Fig. 3). In THAS2, this tyrosine on helix α2 is replaced with a tryptophan residue, but a tyrosine at position 120 points into the active site. Interestingly, a non-proline *cis*-peptide is present in the THAS1-$NADP^+$ and HYS apo structures (Supplementary Fig. 5). Closer inspection of this region in THAS1 shows that when this bond is in the *trans* conformation, the side chain of Asp310 is projected into the cofactor-binding site such that it would prevent NADPH binding.

Strictosidine aglycone binding. Despite extensive efforts, both product and substrate failed to co-crystallize with any of the enzymes. Therefore, molecular docking was used to visualize the position of strictosidine aglycone in THAS1 (Fig. 4b). To ensure that the correct substrate tautomer was used for docking, we identified the most predominant strictosidine aglycone isomer that forms in solution. Although product precipitation prevented monitoring the SGD reaction *in situ* under aqueous conditions (Methods), ^1H,^{15}N-HMBC NMR showed that an enamine species was the predominant product in aqueous methanolic solution (Supplementary Fig. 6). This is consistent with literature reports that cathenamine is the major product of SGD, and is the proposed precursor of ajmalicine and tetrahydroalstonine (Fig. 1)[21,23]. *In silico* docking with THAS1 positions cathenamine between the nicotinamide of the $NADP^+$ and Tyr56, which is located on helix α2 (Fig. 4b). THAS1 loop 1 contains Phe65 that also projects into the active site and may interact with the aromatic cathenamine substrate.

Mechanism of reduction and heteroyohimbine formation. In tetrahydroalstonine biosynthesis, we hypothesize that cathenamine tautomerizes to the iminium form by protonation at C20, followed by addition of the hydride at C21. Protonation at C20 must occur from the bottom face to yield the S stereochemistry observed at this position (Fig. 5a). While there does not appear to be an appropriately positioned active site residue to perform this role, the crystal structures of these enzymes reveal the presence of numerous water molecules in the active site that could potentially protonate this carbon (Fig. 4a). ^1H,^{15}N-HMBC measurements of strictosidine aglycone at different pH values show formation of the iminium species in solution when the pH was reduced to ~3.5, indicating that this tautomer can readily form in the presence of an acidic moiety (Supplementary Fig. 6).

To elucidate the stereo and regioselectivity of reduction by NADPH, we isolated tetrahydroalstonine from reactions using THAS1 and pro-*R*-NADPD. Analysis by ^2H-NMR showed that tetrahydroalstonine is labelled with deuterium in the pro-*R* position at C21, consistent with previously reported experiments performed in crude cell extracts (Fig. 6a)[32]. It is possible that THAS1 could reduce the enamine directly, in which case hydride addition would occur at C21, followed by protonation at C20 by a water molecule as described above. The presence of mayumbine/19-epi-ajmalicine in some of the enzymatic reactions suggests that small amounts of cathenamine can open and form 19-epi-cathenamine, either in solution or in the active site.

In the case of HYS, which produces both ajmalicine (*R* C20) and tetrahydroalstonine (*S* C20), protonation must also occur from the opposite face to yield *R* stereochemistry at C20 (Fig. 5b). Products of HYS generated with pro-*R*-NADPD were also isolated and analysed by ^2H-NMR, and, as for tetrahydroalstonine from THAS1 in each case showed deuterium labelling in the pro-*R* position at C21 (Fig. 6a–c). Therefore, the stereochemical course of hydride addition is not altered in HYS compared with THAS1.

The major difference between HYS and THAS1/THAS2 appears to be the extended loop over the HYS active site (D125-GHFGNN-F132 in HYS and D128-SN-Y131 in THAS1, loop 2 in Fig. 3). The histidine residue in HYS loop 2 (His127) appears to be positioned appropriately to provide an alternative proton source for the opposite ('top') face of the substrate, which could explain the appearance of ajmalicine in the product profile of HYS. Reactions with THAS1 and HYS performed at different pH conditions (5–8) revealed that while changes in pH did not substantially impact the product profile of THAS1, HYS produced increased amounts of ajmalicine relative to tetrahydroalstonine at pH 6 compared with higher pH values (Supplementary Fig. 7). The increased level of ajmalicine in HYS at lower pH values is consistent with the pKa value of the

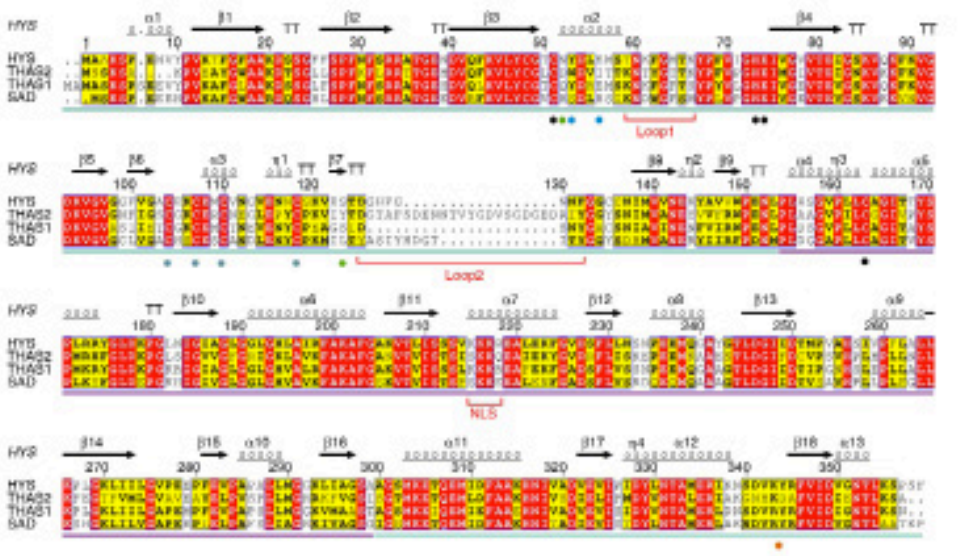


Figure 3 | Sequence alignment of *Catharanthus roseus* HYS enzymes and *Populus trenuloides* SAD Numbering corresponds to HYS. Identical and similar amino acids are highlighted in red and yellow, respectively. Secondary structure elements of the HYS apo crystal structure are displayed. THAS1- and HYS-active site amino acids (Y56/53 and E59/56) are indicated by blue dots, and THAS2 active site amino acids (Y120 and D49) are indicated by green dots. Ligands for catalytic and structural zinc ions are highlighted by black and grey dots, respectively. The nuclear localization signal of (THAS1 and HYS) and loops 1 and 2, respectively, are indicated in red. A non-proline *cis*-peptide bond that is observed in THAS1 holo, in one subunit of THAS1 apo (Supplementary Fig. 5), in HYS apo, and not at all in THAS2 is indicated with an orange dot. The substrate-binding domain and the cofactor-binding domain are indicated by blue and purple bars, respectively.

histidine side chain and supports the role of this histidine in ajmalicine biosynthesis, though attributing pH dependence to specific residues must be approached with caution[35].

Switching stereoselectivity of HYSs. Since the major sequence and structural difference between HYS and THAS1 is the extended loop over the HYS active site, loop 2 (Fig. 3), we swapped these loop regions in THAS1 and HYS to determine whether the stereochemical product profile could be switched (Fig. 7a). Loop 1, which is near the active site, was also swapped (Fig. 7a). While the THAS1 mutant containing the swaps displayed reduced activity rather than an altered product profile (Fig. 7c), the HYS mutant containing the shorter THAS1 loop 2 resulted in a product profile similar to that of THAS1 (Fig. 7b, Supplementary Fig. 8). Since His127 is the only ionizable residue in this loop, we hypothesized that this residue protonates C20, as discussed above. Therefore, we mutated this histidine to alanine or asparagine in HYS. Both of these mutants gave the same THAS-like profile, suggesting that His127 is required for producing the ajmalicine (*R* C20) stereochemistry (Supplementary Fig. 9). Mutation of other conserved ionizable residues in the THAS1 active site (Tyr56, Ser102 and Thr166) did not result in substantial changes in the distribution of products (Supplementary Table 6, Supplementary Figs 10 and 11). Mutations to Glu59, which anchors the NADPH cofactor, resulted in a slight increase in product promiscuity (Supplementary Fig. 11), perhaps by causing a shift in the cofactor position. The reactivity of the substrate[24] and precipitation at high concentrations during the assay makes obtaining accurate kinetic constants challenging[24], so end-point assays were used to assess stereoselectivity and relative activity of the mutant enzymes (Supplementary Table 6). However, the k_{cat} (observed) values could be measured for THAS1 ($1.518 \pm 0.059\,s^{-1}$), THAS2 ($0.033 \pm 0.001\,s^{-1}$), THAS3 ($0.102 \pm 0.004\,s^{-1}$), THAS4 ($0.041 \pm 0.006\,s^{-1}$), HYS ($0.083 \pm 0.005\,s^{-1}$), HYS_loop2 swap ($1.970 \pm 0.153\,s^{-1}$), THAS1 Y56S ($0.118 \pm 0.005\,s^{-1}$) and THAS1 E59A ($0.061 \pm 0.005\,s^{-1}$).

***In planta* localization of HYSs.** Plants use spatial organization on the organ, tissue and intracellular levels to control product distribution (Supplementary Fig. 12). At the subcellular level, we previously showed that THAS1 has an unusual nuclear localization pattern[24], which is also where SGD, the enzyme that synthesizes strictosidine aglycone, is localized[24]. Physical interactions using Bimolecular Fluorescence Complementation (BiFC) were observed between these two enzymes[24]. THAS2 and HYS localization were similarly investigated by expressing yellow fluorescent protein (YFP) fusions in *C. roseus* cells. Microscopy of transiently transformed cells revealed that THAS2–YFP (Fig. 8a) was located in both the cytosol and the nucleus while HYS–YFP (Fig. 8e), similar to THAS1, displayed a preferential nuclear localization (Fig. 8 and Supplementary Fig. 13). As reported for THAS1, this localization relies on the presence of a class V nuclear localization sequence in HYS (215-KKKR-218) that is absent from THAS2 (Fig. 3). BiFC assays revealed that both THAS2 and HYS are capable of self-interactions (Supplementary Fig. 14).

BiFC assays were used to determine whether THAS2 and HYS also interact with SGD (Fig. 9). C-terminal split-YFP fragment fusions of both enzymes (THAS2–YFPC and HYS–YFPC) were

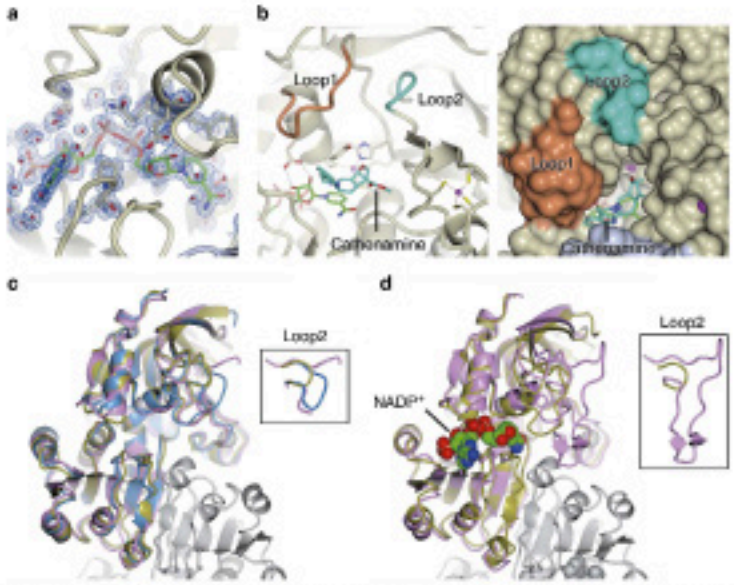


Figure 4 | Crystal structures of heteroyohimbine synthases THAS1, THAS2 and HYS. (**a**) Sample of automatically derived experimentally phased electron density from THAS1 (at 1.12 Å resolution) superimposed on the final model showing the active site region with the $NADP^+$ cofactor (green carbons) together with neighbouring residues (magnolia carbons) and water molecules (small red spheres). (**b**) THAS1 docked with catharanthine (pale blue carbons) with the protein shown in both cartoon (left) and space filling (right) modes. The $NADP^+$ cofactor is shown with green carbons; loop 1 is in orange and loop 2 is in cyan. Zinc ions are displayed as magenta spheres. The active site is largely contained within a single subunit (magnolia surface), although the mouth of the channel leading to the active site is partially bounded by the second subunit of the biological dimer (grey surface). (**c**) Superposition of the apo structures of THAS1 (gold), THAS2 (pink) and HYS (blue). (**d**) Superposition of the holo ($NADP^+$ containing) structures of THAS1 (gold) and THAS2 (pink), with the cofactor of THAS1 shown as van der Waals spheres for emphasis. For **c** and **d**, the structures were superposed onto the THAS1 structure, based on the upper subunit alone; only part of the lower subunit of the THAS1 structure is shown in grey for reference (see Supplementary Fig. 3 for images of the full THAS1 dimer). The insets emphasize the differing lengths of loop 2 between the various structures; the central portion of loop 2 in apo THAS2 was disordered.

co-transformed with SGD that was fused to a N-terminal split-YFP fragment (YFPN–SGD). The formation of a nuclear BiFC complex suggests that both of these enzymes interact with SGD in the nucleus (Fig. 9a–d). Interestingly, the emitted fluorescent signal exhibited a punctated, sickle-shaped aspect as previously observed for the THAS1/SGD interaction (Fig. 9e,f) and for SGD localization[34]. In contrast, no interactions were detected when the BiFC assay was conducted with a downstream enzyme from this biosynthetic pathway, 16-hydroxytabersonine 16-*O*-methyltransferase (16OMT), that is not expected to interact with SGD (Fig. 9g,h).

Double BiFC assays were performed to combine the study of THAS2 and HYS interactions, as well as their interactions with SGD. After transformation into plant cells (16 h), we noted the formation of a dual fluorescent signal for THAS2, both in the cytosol and as punctates in the nucleus that may correspond to the superposition of the signal observed for THAS2 self-interactions and THAS2–SGD interaction (Fig. 9i,j) as confirmed by multicolour BiFC (mBiFC) assays (Supplementary Fig. 15). Increased time of expression (36 h) progressively resulted in the disappearance of the cytosolic signal, and it is intriguing to speculate that this implies a recruitment of THAS2 by SGD (Fig. 9q,r). A similar phenomenon was observed for HYS and THAS1 (Fig. 9k–n,s–v; Supplementary Fig. 15). While self-interactions of 16OMT were detected, no nuclear signal was recovered, confirming the specificity of THAS1–, THAS2–, HYS–SGD interactions (Fig. 9o,p).

***In planta* silencing of HYSs.** The expression levels of the HYS transcripts do not suggest whether a specific HYS is the most biologically relevant (Supplementary Fig. 12). To establish whether any of these enzymes synthesize the expected metabolic product *in planta* the genes encoding active HYSs were silenced. For many medicinal plants, including *C. roseus*, virus-induced gene silencing (VIGS) is the only established method to silence genes in the whole plant. In *C. roseus*, the effect of VIGS is temporally and spatially limited to the first two leaves that emerge immediately after infection[35]. Each of the genes encoding a biochemically active enzyme (*THAS1*, *THAS2*, *THAS3*, *THAS4* and *HYS*) was subjected to VIGS in *C. roseus* seedlings and the effect on alkaloid production was monitored by MS. Since *HYS* and *THAS4* were too similar to silence separately, one common gene fragment was used for silencing both genes simultaneously. Successful silencing of the genes was confirmed by quantitative reverse transcription–PCR (qRT–PCR) (Supplementary Fig. 16). Aside from a small degree of cross-silencing between *THAS2* and THAS3 (12%), all of the target genes were silenced selectively, as measured by qRT–PCR (Supplementary Fig. 16).

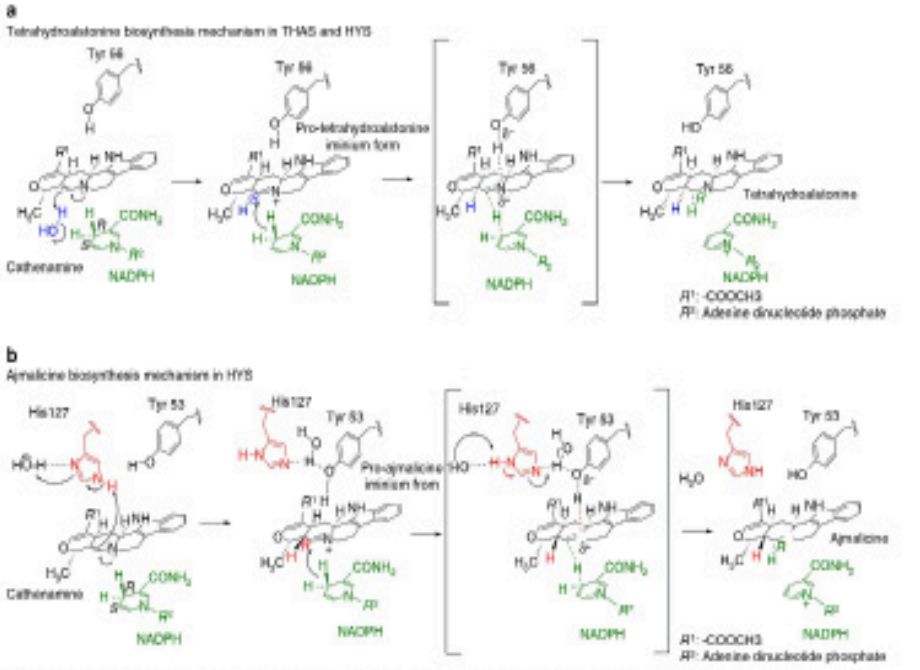


Figure 5 | Mechanistic hypothesis for heteroyohimbine synthases. (**a**) Proposed mechanism for formation of the tetrahydroalstonine (*S* C20) diastereomer. (**b**) Proposed mechanism of formation of the ajmalicine (*R* C20) diastereomer that is observed in HYS, which contains a histidine residue near the active site.

Due to the inherent reactivity of the HYS substrate (strictosidine aglycone), changes in the level of this compound *in planta* are difficult to accurately detect. Instead, the effect of silencing must be established by quantitatively measuring decreases in heteroyohimbine levels. Previously for *THAS1*, we measured the combined peak for heteroyohimbines, since the diastereomers were difficult to resolve on a reverse phase LC column under the reported conditions[24]. However, after substantial optimization (see Methods), an LC-MS method was developed to separate ajmalicine and tetrahydroalstonine in crude leaf extracts (19-epi-ajamlicine/mayumbine is not observed in *C. roseus* leaves, Supplementary Fig. 17). Ajmalicine and particularly tetrahydroalstonine are observed in low levels even in empty vector control samples compared with the major leaf monoterpene indole alkaloids vindoline and catharanthine, and in addition, heteroyohimbine composition varied substantially among individual leaves. Therefore, accurate measurement of decreases in ajmalicine and tetrahydroalstonine levels is challenging. There was no evidence for a decrease of ajmalicine or tetrahydroalstonine when *THAS2* and *THAS3* were silenced. However, for *HYS*, there was a statistically significant decrease (*t*-test 0.0275) in ajmalicine, and no change in tetrahydroalstonine levels. Surprisingly, a statistically significant decrease in ajmalicine, as well as tetrahydroalstonine (*t*-test 0.0277 and 0.0276, respectively) was also noted for *THAS1* (Supplementary Fig. 16). While the results are statistically significant, the leaf-to-leaf variability, the low level of endogenous production, and the catalytic redundancy of these enzymes make it difficult to draw firm conclusions from these VIGS data. In addition, regulatory factors that impact the ratio of ajmalicine and tetrahydroalstonine, such as transport mechanisms and/or further derivatization to other products, cannot be ruled out. Additional silencing systems, in different tissues, will be required to more firmly establish the physiological function of these enzymes. Nevertheless, we can state that silencing of *HYS* and *THAS1* impacts alkaloid production in *C. roseus* leaves.

Discussion

Here we report several medium chain dehydrogenases/reductases that produce the heteroyohimbine stereoisomers ajmalicine and/or tetrahydroalstonine, thereby providing a framework to understand the enzymatic control over stereoselectivity in this metabolic pathway. It is notable that we have identified four enzymes that generate tetrahydroalstonine, yet ajmalicine is the more abundant isomer *in planta* (Supplementary Fig. 17). Expression profile data of the genes identified in this study suggest that HYS, which produces ajmalicine, is not expressed at higher levels than the other synthases (Supplementary Fig. 12). There may be additional ajmalicine synthases that are not related to the MDR superfamily homologues identified in this study. Alternatively, tetrahydroalstonine could be shuttled into another pathway or degraded, thereby resulting in the observed lower levels that accumulate *in planta*.

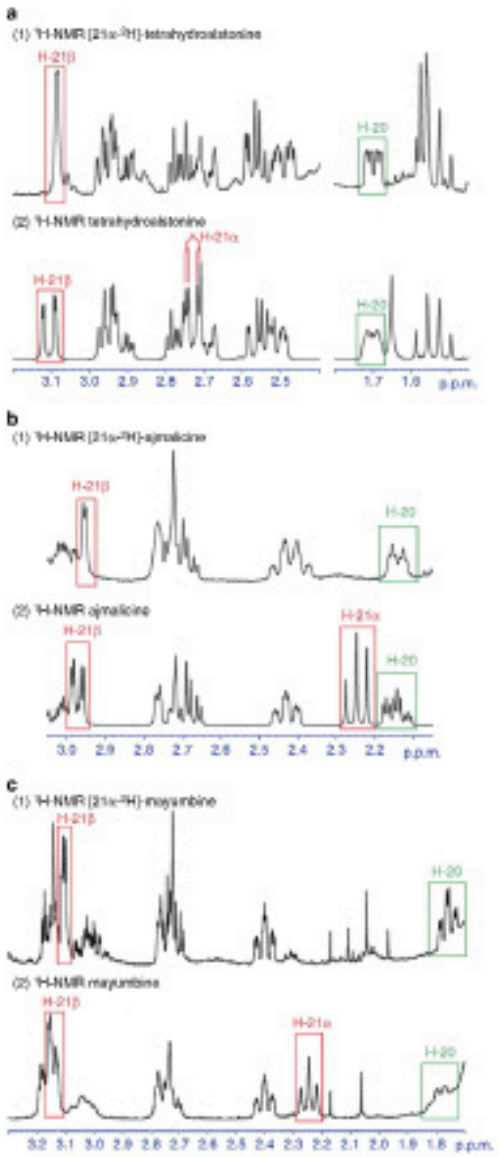


Figure 6 | Deuterium labelling of THAS1 and HYS products using pro-R-NADPD. Comparison of selected regions of ^{1}H-NMR spectra of labelled (**a**) tetrahydroalstonine, (**b**) ajmalicine, (**c**) mayumbine. The spectra indicate that C21 is labelled with deuterium in the pro-R position.

Importantly, the pharmacological activity of heteroyohimbines is impacted by the stereochemistry. Ajmalicine has recently been used in combination with almitrine in post-stroke treatments, though the side effects caused by almitrine resulted in widespread withdrawal of the drug in 2013 (ref. 1). While tetrahydroalstonine has no reported pharmacological function, its oxidized product, alstonine (Fig. 1), has recently been shown to act by a unique mechanism for modulating dopamine uptake and shows potential as an anti-psychotic drug[35]. The heteroyohimbines have excellent promise as a scaffold for pharmacological activity. The discovery of the HYSs, along with recently developed heterologous production platforms for monoterpene indole alkaloids[36], now allows the possibility of generating these alkaloids and unnatural derivatives through metabolic engineering/synthetic biology strategies.

The crystal structures of three HYSs reveal the potential of biosynthetic machinery to generate stereochemical variation. Flexible loop regions can be the key to unlocking chemical diversity: as we have demonstrated here, mutating the extended loop over the HYS active site (loop 2 in Fig. 3) impacts stereochemical outcome. Notably, the MDRs that we have identified demonstrate high variability at this region (Fig. 5). Phylogenetic analysis (Supplementary Fig. 18) suggests that these HYSs, which appear to have originated from a common ancestor, may have undergone neo-functionalization through mutation in this loop region. This loop could potentially be harnessed in protein engineering efforts to generate novel catalytic activity.

While the *in planta* function of heteroyohimbines is unknown, deglycosylated strictosidine is toxic and may act as a defense compound[34], similar to the defense roles of the aglycones of the iridoids from which stictosidine is derived[37,38]. SGD is expressed in most tissues (Supplementary Fig. 12), suggesting that the plant must have evolved mechanisms to control the levels of the toxic strictosidine aglycone. In directed overflow metabolism, excess reactive intermediates are converted into non-reactive byproducts[39]. It is intriguing to speculate that monoterpene indole alkaloid biosynthesis may have initially arisen as a mechanism for handling overflow of strictosidine aglycone. The HYSs perform a single, chemically straightforward reduction reaction that immediately neutralizes the reactivity of strictosidine aglycone/cathenamine. The co-localization and interaction of THAS1, THAS2 and HYS with SGD reinforces the hypothesis of an evolutionary mechanism deployed by strictosidine-accumulating plants to manage the reactivity of the strictosidine aglycone. It also raises the question of a possible competition between HYSs for recruitment by SGD when distinct enzymes are co-expressed in the same tissue/cells. Whether the heteroyohimbines serve an active biological function in the plant, or whether they are simply the end product of directed overflow metabolism, or both, remains to be investigated. Regardless, it is clear that MDRs play an important role in the generation of a wide variety of chemical structures. Duplication of the evolutionary dehydrogenase ancestor may have given rise to multiple HYSs, along with MDRs with other biosynthetic activities, such as tabersonine-3-reductase that is involved in the biosynthesis of the anti-cancer alkaloid vinblastine (Supplementary Fig. 18)[40].

Methods

Selection and cloning of candidate MDRs. The nucleotide and the protein sequences of THAS1 were subjected to a BLAST search against the *C. roseus* Sunstorm Apricot V1.0 Transcript sequences (http://medicinalplantgenomics.msu.edu) and the MDR sequences with the highest identity to THAS1 at the active site and which showed non-negligible expression levels in young and mature leaves were selected as candidates for cloning and expression. The protein sequence of SAD was blasted against the same database and MDRs were also selected based on their active site similarity to that of SAD. The genes coding the candidate MDRs were amplified from *C. roseus* leaf cDNA and cloned into the *E. coli* expression vector pOPINF using the In-Fusion cloning kit (Clontech Takara)[41] by using primers designed based on the transcript sequences (Supplementary Table 1).

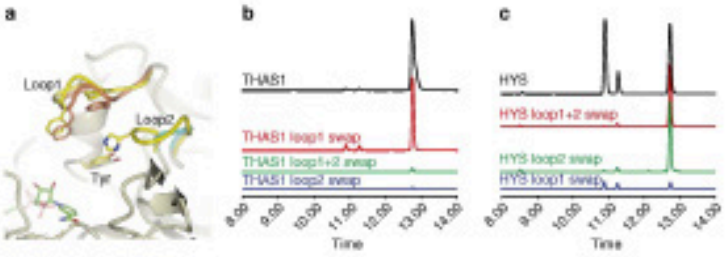


Figure 7 | Product profiles of THAS1 and HYS loop swap mutants. (**a**) Shown is the apo THAS1 structure (magnolia) with loops 1 and 2 highlighted in orange and cyan, respectively. For clarity, only the corresponding loops of the HYS apo structure are shown in yellow after superposition. Similarly, only the cofactor from the superposed holo THAS1 structure is shown for reference (green carbons). The side chains of important residues are also shown. (**b**) Representative LC-MS chromatograms of assays with THAS1 mutants in which loop 1, loop 2 or both have been swapped with the corresponding sequences from HYS. (**c**) Representative LC-MS chromatograms of assays with HYS mutants in which loop 1, loop 2 or both have been swapped with the corresponding sequences from THAS1.

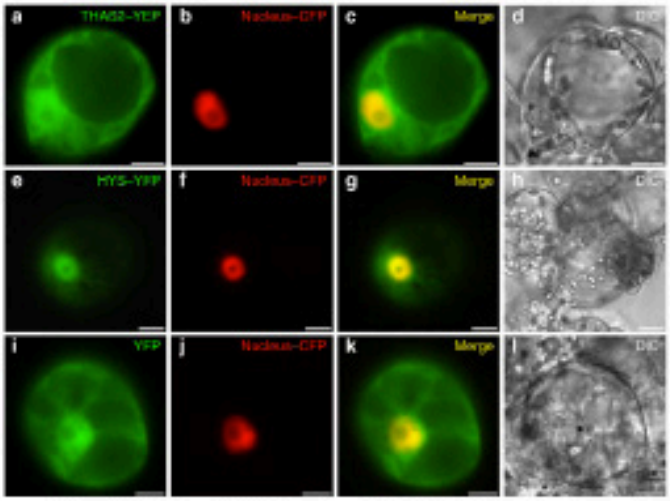


Figure 8 | THAS2 displays nucleocytosolic localization while HYS is preferentially targeted to the nucleus. *C. roseus* cells were transiently co-transformed with plasmids expressing either THAS2–YFP (**a**), HYS–YFP (**e**) or YFP (**i**) and the plasmid encoding the nuclear-CFP marker (**b**,**f**,**j**). Co-localization of the fluorescence signals appears in yellow when merging the two individual (green/red) false colour images (**c**,**g**,**k**). Cell morphology is observed with differential interference contrast (DIC) (**d**,**h**,**l**). Scale bars, 10 μm.

Site-directed mutagenesis of THAS1 and HYS. THAS1 mutants were generated by overlap extension PCR. Briefly, the codon to be mutated was selected and two primers, one reverse and one forward (Supplementary Table 4), were designed to overlap and introduce the mutation. A first PCR was carried out using the reverse mutant primer and the 5′ forward gene-specific primer (Supplementary Table 1), thus generating the 5′ half of the gene carrying the mutation. In parallel, the 3′ half of the mutated gene was generated by PCR using the forward mutant primer and the 3′ reverse gene-specific primer (Supplementary Table 4). PCR products were gel purified and used for the second PCR overlap reaction for generation of the full-length mutated gene, where the 5′ and 3′ halves of the mutated genes were mixed in a PCR reaction in equimolar amounts (~100 ng per fragment) and five cycles of PCR were carried out without including primers. After the 5 overlap PCR cycles, the forward and reverse gene-specific primers were added to the mix and further 30 cycles were performed. Full-length PCR products were gel purified, ligated into pDONR/F expression vector and transformed into competent *E. coli* Stellar strain cells (Clontech Takara). HYS point mutants were obtained as gene fragments (Integrated DNA Technologies, Belgium) with the H127 or V128 codons mutated (H127A CAT- > GCA; H127N CAT- > AAC; F128A TTT- > GCT; F128Y TTT- > TAC) and the pDONR/F overhangs included at the 3′ and 5′ extremities. The THAS1 and HYS double loop mutants were generated by first making their loop 1 mutant genes and then inserting the second loop 2 swap following the same procedure described above. Mutant constructs were sequenced to verify the mutant gene sequence and correct insertion.

Enzyme activity assays. All candidate enzymes and mutants were expressed in Sole8121 (1993) *E. coli* cells (Gevlantis) grown in 2 × YT medium. Protein production was induced by addition of 0.2 mM IPTG and the cultures were shaken at 18 °C for 16 h. Cells were collected by centrifugation, lysed by sonication in Buffer A (50 mM Tris-HCl pH 8, 50 mM glycine, 300 mM NaCl, 5% v/v glycerol, 20 mM imidazole) supplemented with EDTA-free protease inhibitor (Roche Diagnostics) and 0.5 mg ml^{-1} lysozyme. Soluble proteins were purified on Ni-NTA agarose (Qiagen) and eluted with Buffer B (50 mM Tris-HCl pH 8, 50 mM glycine, 300 mM NaCl, 5% v/v glycerol, 500 mM imidazole). Eluates were analysed

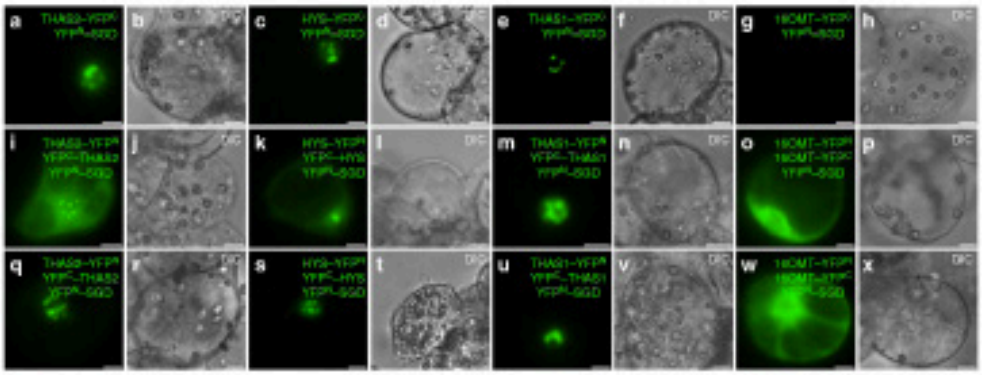


Figure 9 | THAS2 and HYS interact with SGD in the nucleus. THAS2/SGD (**a**,**i**,**q**) and HYS/SGD (**c**,**k**,**s**) interactions were analysed by BiFC in *C. roseus* cells transiently transformed by distinct combinations of plasmids encoding fusions with the two split YFP fragments, as indicated on each fluorescence picture. THAS1/SGD (**e**,**m**,**u**) and 16OMT/SGD (**g**,**o**,**w**) interactions were studied to evaluate the specificity of THAS2/SGD and HYS/SGD interactions. Single BiFC assays showing interactions with SGD (upper row) and double BiFC assays highlighting both interactions with SGD and THAS2, HYS, THAS1, 16OMT self-interactions were conducted and observed 16 h (middle row) and 36 h (lower row) post-transformation. Cell morphology is observed with differential interference contrast (DIC) (**b**,**d**,**f**,**h**,**j**,**l**,**n**,**p**,**r**,**t**,**v**,**x**). Scale bars, 10 μm.

by SDS–PAGE to verify the purity and the molecular weight of the purified proteins. All proteins were dialysed in Buffer C (50 mM phosphate pH 7.6, 100 mM NaCl) and concentrated. Protein concentration was measured with Bradford reagent (Sigma-Aldrich) according to the manufacturer's instructions. Purified proteins were divided in 20 μl aliquots, fast-frozen in liquid nitrogen and stored at −20 °C.

Candidate MDR enzymes and the selected mutants were screened for activity against dihydrocorynantheine. The substrate was generated by dihydrocorynantheine synthase (DCS) in the presence of 50 mM phosphate buffer (pH 6.5) at room temperature for 10 min. The reactions were started by the addition of MDR enzyme (1 μM) and NADPH (5 mM). Caffeine (50 μM) was used as internal standard. All reactions were performed in triplicate. Aliquots of the reaction mixtures (10 μl) were sampled 1 and 30 min after addition of MDR enzyme. The reactions were stopped by the addition of 10 μl of 100% MeOH. Samples were diluted 1:5 in mobile phase (H_2O : 0.1% formic acid) and centrifuged for 10 min at 4,000*g* before UPLC-MS injection (1 μl). The activity of MDR enzymes and mutants was measured by UPLC-MS. Enzymes exhibited a loss of activity after one freeze thaw cycle.

The initial velocity was determined for the wild-type enzymes that displayed HYS activity (THAS1-4 and HYS), as well as for the mutants THAS1 E368L E394A and H45S loop 2 swap. Reactions were monitored at 340 nm at room temperature using a (Cary 50 Bio, Varian) spectrophotometer. Strictosidine (50 μM) was incubated with purified SGD in 50 mM phosphate buffer (pH 7.0) for ~30 min at 30 °C. NADPH (100 μM) was added, mixed by pipetting and the absorbance at 340 nm was monitored until stabilized (minimum of 2 min). A predetermined amount of enzyme (10–400 nM) was then added, mixed and the reaction was monitored for a minimum of 5 min at 340 nm. The resulting slope was calculated during the linear reaction range, usually 0–60 s after addition of the enzyme. The reactions were replicated five times and the initial velocity was calculated for each replicate after accounting for the background. The concentration of substrate (50 μM strictosidine) was determined to be saturating for all wild-type enzymes under these conditions.

Protein crystallization. Proteins for crystallization were purified from 2 l cultures in 2 × YT medium. Protein expression was induced by addition of IPTG and the cultures were grown for 16 h at 18 °C. Cells were collected by centrifugation and lysed by sonication in 30 ml Buffer A supplemented with EDTA-free protease inhibitor and 30 mg of Lysozyme. Lysates were clarified by centrifugation at 17,000*g* for 20 min. Two-dimensional automated purification was performed on an ÄKTAxpress purifier (GE Healthcare). The IMAC step was performed on HisTrap HP 5 ml columns (GE Healthcare) equilibrated with Buffer A. Proteins were step-eluted with Buffer B and directly injected on a gel filtration column equilibrated with Buffer D (20 mM HEPES, 150 mM NaCl, pH 7.5). Fractions were collected and analysed by SDS–PAGE and those containing pure protein were pooled and concentrated in a 10 kDa membrane filter—Millipore filter (Merck Millipore).

Purification of HYS required the addition of 1 mM DTT to all purification buffers and dialysis in Buffer D containing 0.5 mM *tris*(2-carboxyethyl)phosphine (TCEP) before crystallization and storage.

Crystallization screens were conducted by sitting-drop vapour diffusion in MRC2 96-well crystallization plates (Swissci) with a mixture of 0.3 μl well solution (from the PEGs (Qiagen), PACT (Qiagen) and JCSG (Molecular Dimensions) suites) and 0.3 μl protein solution. Protein concentrations were adjusted to 7–10 mg ml^{-1} while $NADP^+$ (Sigma-Aldrich) was added to a final concentration of 1 mM for co-crystallization studies. Solutions were dispensed either by an OryxNano or an Oryx8 robot (Douglas Instruments).

THAS1 apo crystals were obtained from His_6-tag cleaved THAS1 (3C protease) in a solution containing 0.1 M MES, pH 6.5, 15% w/v PEG 3000. THAS1-$NADP^+$ crystals were obtained from a solution containing 0.2 M potassium/sodium tartrate with 20% w/v PEG 3350. T1AS2 crystals (with and without $NADP^+$) were obtained from a condition containing 0.2 M lithium chloride and 20% w/v PEG 3350. HYS crystals were obtained after removal of the His_6-tag (using 3C protease) in 0.1 M MMT buffer, pH 5 and 15% w/v PEG 3350. All crystals were cryoprotected by soaking in crystallization solution containing 25% v/v ethylene glycol before flash-cooling in liquid nitrogen.

Data collection and structure determination. X-ray data sets were recorded on one of three beamlines at the Diamond Light Source (Oxfordshire, UK) (I23, I04; I24, I03; I04; I04-1; I02, I04-1) at wavelengths of 0.9000–0.976 Å (I23, 0.980 Å; I24, 0.976 Å; I04, 0.918 Å; I02, 0.920 Å; I03, 0.920 Å) using either a Pilatus 6M or 2M detector (Dectris) with the crystals maintained at 100 K by a Cryojet cryocooler (Oxford Instruments). Diffraction data were integrated using XDS[42] and scaled and merged using AIMLESS[43] via the XIA2 expert system[44]; data collection statistics are summarized in Supplementary Table 3. Initially the THAS1-$NADP^+$ data set was automatically processed at the beamline by fast_dp[45] to 1.12 Å resolution and a structure solution was automatically obtained by single wavelength anomalous dispersion phasing using the SHELX suite[46] via the fast_ep pipeline (Wilson, manuscript in preparation). Despite being collected at a wavelength somewhat remote from the zinc *K* X-ray absorption edge (theoretical wavelength 1.284 Å), the anomalous signal was sufficient for fast_ep to locate four zinc sites and calculate a very clear experimentally phased electron density map (Fig. 4a). This was available to view at the beamline in the ISPyB database[47] via the SynchWeb interface[48] within a few minutes of completing the data collection. The map was of sufficient quality to enable 94% of the residues expected for a THAS1 homodimer to be automatically fitted using BUCCANEER[49]. The model was finalized by manual rebuilding in COOT[50] and restrained refinement using anisotropic thermal parameters in REFMAC5 (ref. 51) against the same data set reprocessed to a resolution of 1.06 Å as described above (Supplementary Table 3), and contained 95% of the expected residues, with one $NADP^+$ molecule and two zinc ions per subunit. All the remaining structures were solved by molecular replacement using PHASER.[52] In each case, the asymmetric unit corresponded to the biological dimer and the preliminary models were obtained by searching for two copies of a monomer template. For THAS1 apo, THAS2 $NADP^+$ and HYS apo, a THAS1-$NADP^+$ protein-only monomer model was used as the basis for the template, although in the latter two cases a homology model of the target structure was generated from the THAS1 template using the Phyre2 server[53] (http://www.sbg.bio.ic.ac.uk/~phyre2/) before running PHASER.

For solving the THAS2 apo structure, a THAS2 $NADP^+$ monomer was used as the template. In contrast to THAS1-$NADP^+$, three four-residue were refined in REFMAC5 with isotropic thermal parameters and TLS group definitions obtained from the TLS-MD server[34]. Model generation were validated with the MOLPROBITY[35] tool before submission to the PDB. The statistics of the final models are summarized in Supplementary Table 3. Additional statistics for R_{free}: 5F13, 0.020 (0.517); 5F15, 0.036 (0.680); 5H81, 0.041 (0.349); 5H82, 0.035 (0.439); 5H83, 0.008 (0.066) and $CC_{1/2}$: 5F13, 0.999 (0.310); 5F15, 0.999 (0.329); 5H81, 0.998 (0.725); 5H82, 0.999 (0.639); 5H83, 0.996 (0.316) (where values in parentheses are for highest-resolution shell) were also noted. Ramachandran statistics (favoured/allowed/outliers [%]) are: 5F13, 96.8/3.2/0.0; 5F15, 96.0/4.0/0.0; 5H81, 96.2/3.8/0.0; 5H82, 96.1/3.9/0.0; 5H83, 96.6/3.1/0.3. All structural figures were prepared using CCP4mg[36].

Simulated annealing omit maps were calculated for the active site regions of all five structures presented in this study. For all structures, selected residues bordering the active site (and the cofactors in the case of the two holoenzyme structures) were deleted from the coordinates of the final models. The resultant PDB files were used as inputs to simulated annealing refinement with PHENIX (http://www.phenix-online.org) from a starting temperature of 5,000 K after applying small random shifts to the model (‘shake’ keyword set to 0.5). The resultant mFobs–dFcalc difference electron density maps (contoured at ~2.8σ) are displayed superposed on the final coordinates, where the corresponding omitted atoms are shown in stick representation. In each case, structurally equivalent residues were omitted, with the exception of HYS where His127 from loop 2 was also omitted.

UPLC-MS analysis. Enzyme assays and plant tissue samples from VIGS experiments on *C. roseus* plants were analysed by UPLC-MS. UPLC-MS analysis was carried out on a UPLC (Waters) equipped with an Acquity BEH C18 1.7 μm 2.1 × 50 mm column connected to Xevo TQS (Waters). MS detection was performed in positive ESI. Capillary voltage was 3.8 kV; the source was kept at 150 °C; desolvation temperature was 500 °C; cone gas flow, 50 l h^{-1}; and desolvation gas flow, 800 l h^{-1}. Unit resolution was applied to each quadrupole. Multiple reaction monitoring signals were used for detection and quantification of caffeine (m/z 195 > 110, 138) and heteroyohimbine alkaloids (353 > 117, 144).

For rapid deregulation of active enzymes and mutants, a linear gradient method (Method 1) was used at a flow rate of 0.6 ml min^{-1} using a binary solvent system in which solvent A1 was 0.1% formic acid in water and solvent B1 was acetonitrile. The gradient profile was: 0 min, 5% B1; from 0 to 3.5 min, linear gradient to 35% B1; from 3.5 to 3.75 min, linear gradient to 100% B1; from 3.75 to 4 min, wash at 100% B1; back to the initial conditions of 5% B1 and equilibration for 1 min before the next injection. Column temperature was held at 30 °C. The injection volume for both the solutions of standard compounds and the samples was 1 μl. Samples were kept at 10 °C during the analysis.

For separation of the different heteroyohimbines, a different chromatographic method was applied that was adapted from the work of Sun L. *et al.*[37] In this method (Method 2) solvent A2 was 0.1% NH_4OH and solvent B2 was 0.1% NH_4OH in acetonitrile. A linear gradient from 0 to 65% B2 in 17.5 min was applied for separation of the compounds followed by an increase to 100% B2 at 18 min, a 2-min wash step and a re-equilibration at 0% B2 for 3 min before the next injection. The column was kept at 40 °C throughout the analysis and the flow rate was 0.6 ml min^{-1}.

^{2}H labelling experiments. Deuterated Pro-*R*-NADPD was regenerated in solution by *Thermoanaerobacter brockii* alcohol dehydrogenase (50 U, Sigma) using 400 μM $NADP^+$ and 1% v/v [^{2}H$_8$]-isopropanol (CIL). The NADPD regeneration was monitored by ultraviolet spectroscopy at 340 nm. Strictosidine (19.9 mg) was incubated with 1.27 nM SGD in 94 ml of 50 mM phosphate buffer (pH 6.5). THAS1 enzyme was added to the reaction (final concentration of 1.67 μM) and the mixture was incubated at 35 °C. The reaction was monitored for completeness by UPLC-MS and after 5 h no strictosidine or deglycosylated strictosidine was observed. The reaction was stopped by addition of 100 ml of methanol and reaction mixture was concentrated to dryness. The dried reaction mixture was resuspended in 15 ml H_2O and extracted with 3 × 15 ml of ethyl acetate and the EtOAc fraction was dried. [21α-^{2}H$_1$] tetrahydroalstonine was isolated by preparative TLC (mno-silica plate, Sigma-Aldrich), as previously described[38]. The band of [21α-^{2}H$_1$]-tetrahydroalstonine was excised from the plate and THA was extracted with EtOAc multiple times (total volume 40 ml). The EtOAc fraction was filtered and dried under high-vacuum overnight. The [21α-^{2}H$_1$]-tetrahydroalstonine was dissolved in 600 μl of $CDCl_3$ and ^{1}H-NMR was measured.

Strictosidine (35.3 mg) was incubated with 1 nM of SGD and 500 μM $NADP^+$ with 50 U of *T. brockii* ADH and 1% v/v [^{2}H$_8$]-isopropanol in a total volume of 140 ml of 50 mM HEPES buffer (pH 7.5). HYS was added (final concentration 1.71 μM) and the reaction was incubated at 37 °C with shaking and monitored for completeness by UPLC-MS. After 6 h, the reaction was complete and was stopped by addition of 150 ml of methanol. [21α-^{2}H$_1$]-tetrahydroalstonine, [21α-^{2}H$_1$]-epiacuteine and [21α-^{2}H$_1$]-epivincine were isolated by preparative TLC and ^{1}H-NMR spectra measured as described above.

^{15}N labelling experiments. *C. roseus* tryptophan decarboxylase (TDC) was cloned into pJOE3047 vector, expressed in *E. coli* and purified as described above for the MD6x. [α/β/ε ^{15}N] tryptophan (CIL, 50 mg) was incubated with 500 nM of TDC, 400 μM pyridoxal-5′-phosphate in 100 ml 50 mM phosphate buffer (pH 7.5) at 35 °C. The reaction was monitored by MS, continued through completion after 4 h and terminated by addition of 50 ml MeOH. [α/pi/ε-^{15}N]-tryptamine (24 mg) was isolated by preparative HPLC. The isolated [α/pi/ε-^{15}N]-tryptamine was incubated with 3 mM secologanin and 200 nM strictosidine synthase in 180 ml 50 mM phosphate buffer (pH 7.0) at 30 °C overnight. The reaction was terminated by addition of 50 ml MeOH. [ε-^{15}N]-strictosidine (62 mg) was isolated by preparative HPLC. [ε-^{15}N]-strictosidine was then reacted with SGD and the product was extracted (for ethyl acetate, dissolved in MeOH-d_4 and characterized by ^{1}H,^{15}N-HMBC as described above.

Compound characterization. High-resolution electrospray ionization MS spectra were measured with a Shimadzu IT-TOF mass spectrometer. NMR spectra were acquired using a Bruker Advance NMR instrument operating at 400 MHz for ^{1}H equipped with a BBFO plus 5-mm probe. The number of scans depended on the concentration of the sample. The ^{1}H,^{15}N HMBC experiment was acquired with a spectral width: 6,009 Hz in the F2 (^{1}H) dimension and 38,610 Hz in the F1 (^{15}N), with an acquisition time of 0.09 s and 360 scans per increment. The long range delay was optimized after a series of experiments with [ε-^{15}N]-strictosidine using a range of different mixing times and finally was adjusted for a coupling of 5 Hz. The relaxation delay was 1.5 s, the data collection matrix was 1024 × 64, the t1 dimension was zero filled to 1k real data points and a π/2 square sine bell window was applied in both dimensions. The ^{1}H-NMR spectra were compared with those of standards and literature data.

[21α-^{2}H$_1$]-*tetrahydroalstonine*. TLC (EtOAc: C_6H_{14}: Et_2N, 20: 75: 1 v/v): R_f = 0.58; HR-MS (IT-TOF): found for [M + H] + : $[C_{21}H_{21}DN_2O_3]^+$ = 354.1621; *calcd* 354.16225; ^{1}H-NMR (400 MHz, $CDCl_3$): δ 7.80 (br s, 1H), 7.56 (s, 1H), 7.45 (dd, J = 7.6 Hz, J = 1.8 Hz, 1H), 7.28 (dd, J = 7.6 Hz, J = 1.8 Hz, 1H), 7.12 (ddd, J = 7.6 Hz, J = 7.6 Hz, J = 1.8 Hz, 1H), 7.07 (ddd, J = 7.6 Hz, J = 7.6 Hz, J = 1.8 Hz, 1H), 4.36 (dq, J = 10.1 Hz, J = 6.5 Hz, 1H), 3.71 (s, 3H), 3.59 (dd, J = 12.0 Hz, J = 3.8 Hz, 1H), 3.08 (d, J = 2.0 Hz, 1H), 2.95 (ddd, J = 12.2 Hz, J = 6.3 Hz, J = 2.5 Hz, 1H), 2.69 (ddd, J = 14.3 Hz, J = 6.3 Hz, J = 2.5 Hz, 1H), 2.57 (ddd, J = 12.0 Hz, J = 4.9 Hz, J = 3.8 Hz, 1H), 2.48 (ddd, J = 14.3 Hz, J = 10.8 Hz, J = 3.8 Hz, 1H), 2.36 (ddd, J = 12.2 Hz, J = 10.4 Hz, J = 4.8 Hz, 1H), 2.19 (dddd, J = 12.0 Hz, J = 4.9 Hz, J = 3.8 Hz, 1H), 1.78 (ddd, J = 10.4 Hz, J = 4.7 Hz, J = 1.8 Hz, 1H), 1.33 (dd, J = 12.0 Hz, J = 12.0 Hz, 1 H), 1.40 (d, 6.5, 3H).

[21α-^{2}H$_1$]-*ajmalicine*. TLC (EtOAc: C_6H_{14}: Et_2N, 20: 75: 1 v/v): R_f = 0.41; HR-MS (IT-TOF): found for [M + H] + : $[C_{21}H_{21}DN_2O_3]$ + ... 354.1621; *calcd* 354.16225; ^{1}H-NMR (400 MHz, $CDCl_3$): δ 7.94 (br s, 1H), 7.53 (d, J = 1.6 Hz 1H), 7.46 (dd, J = 7.7 Hz, J = 1.5 Hz, 1H), 7.30 (dd, J = 7.7 Hz, J = 1.5 Hz, 1H), 7.15 (ddd, J = 7.7 Hz, J = 7.7 Hz, J = 1.5 Hz, 1H), 7.08 (ddd, J = 7.7 Hz, J = 7.7 Hz, J = 1.5 Hz, 1H), 4.45 (dq, J = 13.5 Hz, J = 6.7 Hz, 1H), 3.74 (s, 3H), 3.41 (ddd, J = 11.3 Hz, J = 4.3 Hz, J = 1.8 Hz, 1H), 3.22 (m, 1H), 3.18 (m, 1H), 3.00 (m, 1H), 2.95 (d, J = 3.1 Hz, 1H), 2.75 (m, 1H), 2.68 (m, 1H), 2.42 (m, 1H), 2.14 (ddd, J = 13.5 Hz, J = 11.4 Hz, J = 3.1 Hz, 1H), 1.32 (dd, J = 11.4 Hz, J = 11.4 Hz, 1H), 1.19 (d, 6.7, 3H).

[21α-^{2}H$_1$] *raucaffricine*. TLC (EtOAc: C_6H_{14}: Et_2N, 20: 75: 1 v/v): R_f = 0.20; HR-MS (IT-TOF): found for [M + H]$^+$: $[C_{27}H_{31}DN_2O_8]^+$ = 354.1921; calcd 354.19225; ^{1}H-NMR (400 MHz, $CDCl_3$): δ 7.86 (br s, 1H), 7.56 (d, J = 1.6 Hz 1H), 7.46 (dd, J = 7.6 Hz, J = 1.6 Hz, 1H), 7.30 (dd, J = 7.6 Hz, J = 1.6 Hz, 1H), 7.15 (ddd, J = 7.6 Hz, J = 7.6 Hz, J = 1.6 Hz, 1H), 7.09 (ddd, J = 7.6 Hz, J = 7.6 Hz, J = 1.6 Hz, 1H), 3.88 (dq, J = 12.0 Hz, J = 6.3 Hz, 1H), 3.73 (s, 3H), 3.40 (dd, J = 11.1 Hz, J = 4.4 Hz, 1H), 3.16 (ddd, J = 12.6, Hz J = 5.0 Hz, J = 3.0 Hz, 1H), 3.14 (dd, J = 12.0 Hz, J = 6.8 Hz, 1H), 3.10 (d, J = 3.1 Hz, 1H), 3.03 (ddd, J = 14.8 Hz, J = 10.9 Hz, J = 3.5 Hz, 1H), 2.72 (dd, J = 16.3 Hz, J = 6.7 Hz, 1H), 2.40 (ddd, J = 15.8 Hz, J = 10.5 Hz, J = 3.4 Hz, 1H), 1.76 (ddd, J = 10.2 Hz, J = 10.2 Hz, J = 3.5 Hz, 1H), 1.37 (d, 6.3, 3H), 1.28 (dd, J = 14.8 Hz, J = 11.0 Hz, 1H), 1.21 (dd, J = 16.2 Hz, J = 6.7 Hz, 1 H).

[ε-^{15}N]-*strictosidine*. HR-MS (IT-TOF): found for [M + H] $^+$: $[C_{27}H_{30}N^{15}NO_8]$ + = 532.2304; calcd 532.23074; ^{1}H-NMR (400 MHz, CD_3OD): δ 10.5 (br s, 1H), 7.77 (s, 1H), 7.41 (dd, J = 7.8 Hz, J = 1.9 Hz, 1H), 7.30 (dd, J = 8.1 Hz, J = 1.6 Hz, 1H), 7.11 (ddd, J = 8.1 Hz, J = 7.0 Hz, J = 1.9 Hz, 1H), 7.02 (ddd, J = 7.8 Hz, J = 7.0 Hz, J = 1.6 Hz, 1H), 5.84 (ddd, J = 17.5 Hz, J = 10.4 Hz, J = 7.3 Hz, 1H), 5.05 (d, J = 9.8 Hz, 1H), 5.04 (dd, J = 17.0 Hz, J = 2.7 Hz, 1H), 5.18 (dd, J = 10.6 Hz, J = 2.8 Hz, 1H), 4.79 (d, J = 7.9 Hz, 1H), 3.98 (dd J = 12.0 Hz, J = 2.2 Hz, 1H), 3.78 (s, 3H), 3.64 (dd, J = 11.0 Hz, J = 6.9 Hz, 1H), 3.42–3.30 (m, 3H), 3.13 (dd, J = 9.8 Hz, J = 8.5 Hz, 1H), 3.02 (dd, J = 9.3 Hz, J = 7.9 Hz, 1H), 3.10–3.00 (m, 2H), 2.97 (ddd, J = 15.8 Hz, J = 4.9 Hz, J = 1.3 Hz, 1H), 2.72 (ddd, J = 8.5 Hz, J = 7.0 Hz, J = 4.8 Hz, 1H), 2.27 (dd, J = 14.6 Hz, J = 11.5, 1H), 2.18 (dd, J = 11.6 Hz, J = 3.8, 1H); ^{15}N-NMR (40 MHz, CD_3OD): δ 44.8.

Subcellular localizations and analysis of protein-protein interactions by BiFC. Subcellular localization of THAS2 and HYS were studied by creating fluorescent fusion proteins using the pSCA-cassette YFP plasmid[39]. The full-length open reading frame of THAS2 was amplified using the specific primers 5′ GTGAGAA CTAGTATGTCTTCAAAATCAGCAAAACCAGTG-3′ and 5′-CTGAGAACTA GTAGCAGATTTCAATGTGTTTTCTATGTCAAT-3′, and HYS ORF with primers 5′ GTGAGAACTAGTATGGCTGCAAAGTCACCTGAAAATGT ATAC-3′ and 5′-CTGAGAACTAGTGAAAGATGGGGATTTGAGAGTG TTTCCTAC-3′, which were designed to introduce the SpeI restriction site at both

cDNA extremities. PCR products were sequenced and cloned at the 5′ end of the YFP-coding sequence to generate the THAS2–YFP^N, HYS–YFP^N fusion proteins or at the 3′ end to express the YFP^N–THAS2 and YFP^N–HYS fusions.

The interaction of THAS2 and HYS with SGD were characterized by BiFC assays using the previously amplified THAS2 and HYS PCR products cloned via SpeI into the p87YC2 (M) vector[34], which allows expression of THAS2 and HYS fused to the amino-terminal extremity of the split-YFP^C fragment (THAS2–YFP^C, HYS–YFP^C, respectively), and into the pSPYNE(R)173-SGD plasmid[34] expressing SGD fused to the carboxy-terminal extremity of the split YFP^N fragment (YFP^N–SGD). Plasmids encoding THAS8–YFP^N, THAS9–YFP^C, YFP^N–THAS1 and plasmids expressing 16OMT^1 YFP^N and 16OMT^1 YFP^C were used as controls and were constructed previously[20,28].

THAS2 and HYS self interactions were analysed via additional cloning of the THAS2 and HYS PCR products into the pSCA-SPYNE173, pSPYNE(R)173 and pSCA-SPYCE (MR) plasmids[34,35] to express THAS2–YFP^N, HYS–YFP^C and YFP^C–THAS2, YFP^C–HYS, respectively.

The capacities of THAS2 and HYS to interact with SGD were also characterized by double BiFC and mBiFC. The previously amplified THAS2 and HYS PCR products were fused to the coding sequences of the amino-terminal or carboxy-terminal of the split YFP fragments into the pSCA-SPYNE173, pSCA-SPYCE (M) and pSCA-SPYCE (MR) plasmids[34,35], allowing expression of THAS2–YFP^N, YFP^C–THAS2 HYS–YFP^N and HYS–YFP^C, respectively. SGD was subsequently fused to the carboxy-terminal extremity of the split YFP fragment (YFP^N–SGD) and the CFP^N fragment (CFP^N–SGD).

Transient transformation of *C. roseus* cells by particle bombardment and fluorescence imaging were performed following the procedures previously described[18]. Briefly, *C. roseus*-plated cells were bombarded with DNA-coated gold particles (1 μm) and 1,100 psi rupture disc at a stopping screen to target distance of 6 cm, using the Bio-Rad PDS1000/He system. Cells were cultivated for 16–24 h before being harvested and observed. The subcellular localization was determined using an Olympus BX-51 epifluorescence microscope equipped with an Olympus DP-71 digital camera and a combination of YFP and CFP filters. The pattern of localization presented in this work is representative of more 50 observed cells. The nuclear localizations of the different fusion proteins were confirmed by co-transformation experiments using a nuclear-CFP marker[34]. Such plasmid transformations were performed using 40 mg of each plasmid or 100 mg for BiFC assays.

Agrobacterium VIGS and qPCR. The THAS1, THAS2, THAS3 and THAS4-UTR silencing fragments were amplified with primers (Supplementary Table 6) and the resulting fragments were cloned into the pTRV2a vector as described[31]. Since THAS1 and HYS are ~91% identical, it was not possible to design silencing fragments to avoid cross-silencing. Therefore, a common silencing fragment for both of the two genes was designed. The resulting pTRV2a constructs were used to silence the different THASs and HYS in *C. roseus* seedlings essentially as described before[15]. Leaves from the first two pairs to emerge following inoculation were harvested from eight plants transformed with the empty pTRV2a and pTRV2a carrying the silencing fragment. The collected leaves were frozen in liquid nitrogen, powdered using a pre-chilled mortar and pestle, and subjected to LC-MS and qRT–PCR analysis. The heteroyohimbine content of silenced leaves was determined by LC-MS. Leaf powder was weighed (10–20 mg), extracted with methanol (1 ml) and vortexed for 1 min. After a 10-min centrifugation step at 17,000*g*, an aliquot of the supernatant (20 μl) was diluted to 200 μl with methanol, filtered through 0.2-μm PTFE filters and analysed on Waters Xevo TQ-MS. The chromatographic separation and MS measurements were carried out as described above (method 2). To more comprehensively assess the global effect of silencing the HYS gene by VIGS on *C. roseus* metabolism, an untargeted metabolomics analysis by LC-MS was performed as previously reported.[28] However, aside from the changes in the heteroyohimbines reported in Supplementary Fig. 16, no substantial differences in metabolic profiles were noted using this approach.

Gene silencing was confirmed by qRT–PCR. qRT–PCR was also used to check the expression of the other HYS genes to ensure that no cross silencing occurred. RNA extraction was performed using the RNeasy Plant Mini Kit (Qiagen). RNA (1 μg) was used to synthesize cDNA in 20-μl reactions using the iScript cDNA Synthesis Kit (Bio-Rad). The cDNA served as template for quantitative PCR performed using the CFX96 Real Time PCR Detection System (Bio-Rad) using the SSO Advanced SYBR Green Supermix (Bio-Rad). Each reaction was performed in a total reaction volume of 20 μl containing an equal amount of cDNA, 0.25 mM forward and reverse primers and 1 × Sso Advanced SYBR Green Supermix (Bio-Rad). The reaction was initiated by a denaturation step at 95 °C for 10 min followed by 40 cycles at 95 °C for 15 s and 60 °C for 1 min. Melting curves were used to determine the specificity of the amplifications. Relative quantification of gene expression was calculated according to the delta-delta cycle threshold method using the 40S ribosomal protein S9 (RPS9). All primer pair (Supplementary Table 7) efficiencies were between 90 and 100%, and the individual efficiency values were considered in the calculation of normalized relative expression, which was performed using the Gene Study feature of CFX Manager Software. All biological samples were measured in technical duplicates.

pH effect on product profile. Strictosidine was deglycosylated using purified SGD for 15 min at room temperature using assay conditions as described above. Strictosidine aglycone was then incubated at a final concentration of 380 μM at pH 5, 6, 7 and 8 in a buffer mix to avoid buffer ingredient effect on activity (150 mM phosphate buffer, 50 mM citric acid, 50 mM HEPES). Caffeine (50 μM) was used as an internal standard.

At time 0 the enzyme, either THAS1 or HYS (1 μM final concentration), premixed with NADPH (500 μM) was added to the substrate solution. In parallel, a chemical reducing agent, $NaBH_4$ (3 mM final concentration), was added to deglycosylated strictosidine as a control reaction. All reactions were carried out in triplicate. An end-point sample (10 μl) was taken for each assay and prepared for UPLC-MS by addition of 10 μl of 100% MeOH to stop the reaction, and then diluted 1:9 with H_2O, and centrifuged for 10 min at 5,000 r.p.m. UPLC-MS and data collection were performed as described above for heteroyohimbine separation and quantification.

CD spectra and analysis. Far ultraviolet CD spectra of the wild-type enzymes THAS1 and HYS, as well as the loop mutants of THAS1 and HYS were recorded on a Chirascan Plus spectropolarimeter (Applied Photophysics) at 20 °C in 10 mM potassium phosphate buffer pH 7.0. Samples were analysed from 180 to 280 nm, using a 0.5-nm step at a speed of 1 s per step. Four replicate measurements were performed on each sample and baseline correction was applied to all data. Spectra are presented as the CD absorption coefficient calculated on a mean residue ellipticity basis.

Melting curves of HYS and the HYS loop 2 swap mutant were also acquired by CD. The samples were subjected to temperature ramping at the rate of 1 °C min^{-1} from 20 to 90 °C. Data collection was done from 360 to 200 nm using a 1 nm step and 0.75 s time per point. Data were analysed using the Global 3 software. HYS melting point was measured as 61.0 ± 0.1 °C; enthalpy 351.5 ± 2.6 KJ mol^{-1}. HYS loop 2 swap melting point was measured at 62.0 ± 0.1 °C; enthalpy 335.6 ± 4.5 KJ mol^{-1}.

Protein sequence alignments and phylogenetic tree. Protein sequence alignment was generated using ClustalW algorithm with Geneious v.8 (http://www.geneious.com)[61,62]. The alignment was edited manually using Seaview V4 (ref. 64) and secondary structure depiction was added using ESPript V3 (http://espript.ibcp.fr)[63]. Phylogenetic analysis was performed using the neighbour-joining[65] algorithm and Bootstrap analysis with 1,000 replicates.

Docking of catharanthine in THAS1-NADP$^+$ structure. Catharanthine was docked into the THAS1-NADP$^+$ crystal structure using Autodock 4.2 (ref. 67). The ligand (catharanthine) was prepared with two torsions at the C16, the rest of the molecule being rigid, and the receptor consisted of the deselected high resolution crystal structure. The search space was defined by a 40 × 48 × 40 Å box with a 0.375-Å grid spacing, centred between the nicotinamide ring and the side chain of Tyr58, and encompassed the entire active site cavity. Searches were performed using the Lamarckian Genetic Algorithm, consisting of 100 runs with a population size of 150 and 2,500,000 energy evaluations. A total of 27,000 generations were analysed and clustered with an RMS tolerance of 2 Å per cluster. This resulted in just two distinct clusters, which constituted 98% and 2% of the resultant poses, respectively. The latter cluster placed the indole moiety such that the nitrogen atom was closest to the cofactor. Thus, this cluster was eliminated since it was inconsistent with the results of the deuterium labelling experiments (Fig. 6). The poses contained within the major cluster were all deemed to be 'productive' since they placed the indole moiety of catharanthine towards the entrance of the active site and the C20 and C21 3,3-Å above the nicotinamide C4 atom. The top ranked pose (with an estimated free energy of binding = −8.56 kcal mol^{-1}), as selected by the software, is used in the structures illustrated here (Fig. 4b).

Data availability. The atomic coordinates and structure factors of the five X-ray structures described in this manuscript have been deposited in the Protein Data Bank (http://www.pdb.org/), with accession codes 5U3J, 5T3S, 5H81, 5V8U and 5T83. Accession numbers: THAS1 (AEP92528.1); THAS2 (KU865323); THAS3 (KU865322); THAS4 (KU865324); HYS (KU865325); Cro_017994 (KU865330); Cro_011702 (KU865327); Cro_030442 (KU865328); Cro_006840 (KU865329); Cro_012778 (KU865308); Cro_030537 (KU865331); Cro_027134 (AFT68946); Cro_033861 (KU865332); Tabersonine-3-reductase (AEM11361). Data supporting the findings of this study are available within the article and its Supplementary Information files and from the corresponding author upon reasonable request.

References

1. Shimoma, M. & Ridley, J. M. The stereochemistry of the heteroyohimbine alkaloids. *J. Am. Chem. Soc.* **88**, 2507–2512 (1965).
2. Allain, H. & Bentue-Ferrer, D. Clinical efficacy of sinnitrine-raubasine. An overview. *Eur. Neurol.* **30**, 39–44 (1990).
3. Deruk, G. Pharmacological features of an sinnitrine-raubasine combination. Activity at cerebral levels. *Eur. Neurol.* **30**, 51–56 (1990).
4. Li, S. *et al.* Assessment of the therapeutic activity of a combination of sinnitrine and raubasine on functional rehabilitation following ischaemic stroke. *Curr. Med. Res. Opin.* **20**, 109–115 (2004).

5. Roquebert, J. & Demichel, P. Inhibition of the α1- and α2-adrenoceptor-mediated pressor response in pithed rats by raubasine, tetrahydroalstonine and akuammigine. *Eur. J. Pharmacol.* **106,** 203–205 (1984).
6. Ali, L., Dekermendjian, K., Nielsen, M. & Witt, M. R. The heteroyohimbine mayumbine binds with high affinity to rat brain benzodiazepine receptors in vitro. *Nat. Prod. Lett.* **11,** 73–76 (1997).
7. Dassonneville, L. et al. Stimulation of topoisomerase II-mediated DNA cleavage by three DNA-intercalating plant alkaloids: cryptolepine, matadine, and serpentine. *Biochemistry* **38,** 7719–7726 (1999).
8. Costa-Campos, L., Iwu, M. & Elisabetsky, E. Lack of pro-convulsant activity of the antipsychotic alkaloid alstonine. *J. Ethnopharmacol.* **93,** 307–310 (2004).
9. Elisabetsky, E. & Costa-Campos, L. The alkaloid alstonine: A review of its pharmacological properties. *eCAM* **3,** 39–48 (2006).
10. Herrmann, A. P. et al. Effects of the putative antipsychotic alstonine on glutamate uptake in acute hippocampal slices. *Neurochem. Int.* **61,** 1144–1150 (2012).
11. Linck, V. M. et al. Alstonine as an antipsychotic: effects on brain amines and metabolic changes. *eCAM,* 418597 (2011).
12. Linck, V. M. et al. 5-HT2A/C receptors mediate the antipsychotic-like effects of alstonine. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **36,** 29–33 (2012).
13. Linck, V. M. et al. Original mechanisms of antipsychotic action by the indole alkaloid alstonine (Picralima nitida). *Phytomedicine* **22,** 52–55 (2015).
14. Saxton, J. E. *The Chemistry of Heterocyclic Compounds, Indoles: the Monoterpenoid Indole Alkaloids* (Wiley, 2009).
15. Amer, M. A. & Court, W. E. P. Alkaloids of Rauwolfia nitida root bark. *Phytochemistry* **20,** 2569–2573 (1981).
16. Hochstein, F. A. Alkaloids of Rauwolfia sellowii. *J. Am. Chem. Soc.* **77,** 5744–5745 (1955).
17. Melchio, J., Bouquet, A., Pais, M. & Goutarel, R. Alcaloïdes indoliques CVI (1) Identité de la mayumbine et de l'épi-19 ajmalicine. L'iso-3 mayumbine, un nouvel alcaloïde extrait du Corynanthe mayumbensis (R. Good) N. Hallé. *Tetrahedron Lett.* **18,** 315–316 (1977).
18. Phillipson, J. D. & Supavita, N. Alkaloids of uncaria elliptica. *Phytochemistry* **22,** 1809–1813 (1983).
19. Ponglux, D., Supavita, T., Verpoorte, R. & Phillipson, J. D. Alkaloids of Uncaria attenuata from Thailand. *Phytochemistry* **19,** 2013–2016 (1980).
20. Robinson, R. & Thomas, A. F. The alkaloids of picralima nitida, Stapf, Th. and H. Durand. Part I. The structure of akuammigine. *J. Chem. Soc.* 3479–3482 (1954).
21. Stoeckigt, J., Husson, H. P., Kan-Fan, C. & Zenk, M. H. Cathenamine, a central intermediate in the cell free biosynthesis of ajmalicine and related indole alkaloids. *J. Chem. Soc. Chem. Commun.* 164–166 (1977).
22. O'Connor, S. E. & Maresh, J. J. Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Nat. Prod. Rep.* **23,** 532–547 (2006).
23. Gerasimenko, I., Sheludko, Y., Ma, X. & Stöckigt, J. Heterologous expression of a Rauvolfia cDNA encoding strictosidine glucosidase, a biosynthetic key to over 2000 monoterpenoid indole alkaloids. *Eur. J. Biochem.* **269,** 2204–2213 (2002).
24. Stavrinides, A. et al. Unlocking the diversity of alkaloids in Catharanthus roseus: nuclear localization suggests metabolic channeling in secondary metabolism. *Chem. Biol.* **22,** 336–341 (2015).
25. Gongora-Castillo, E. et al. Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PloS ONE* **7,** e52506 (2012).
26. Kellner, F. et al. Genome-guided investigation of plant natural product biosynthesis. *Plant J.* **82,** 680–692 (2015).
27. Bomati, E. K. & Noel, J. P. Structural and kinetic basis for substrate selectivity in Populus tremuloides sinapyl alcohol dehydrogenase. *Plant Cell* **17,** 1598–1611 (2005).
28. Pan, H. et al. Structural studies of cinnamoyl-CoA reductase and cinnamyl-alcohol dehydrogenase, key enzymes of monolignol biosynthesis. *Plant Cell* **26,** 3709–3727 (2014).
29. Youn, B. et al. Crystal structures and catalytic mechanism of the Arabidopsis cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Org. Biomol. Chem.* **4,** 687–1697 (2006).
30. Auld, D. S. & Bergman, T. Medium- and short-chain dehydrogenase/reductase gene and protein families: The role of zinc for alcohol dehydrogenase structure and function. *Cell. Mol. Life Sci.* **65,** 3961–3970 (2008).
31. Elshoud, H. & Ramaswamy, S. Medium- and short-chain dehydrogenase/reductase gene and protein families: Three-dimensional structures of MDR alcohol dehydrogenases. *Cell. Mol. Life Sci.* **65,** 3907–3917 (2008).
32. Stoeckigt, J., Hemscheidt, T., Hoefle, G., Heinstein, P. & Formacek, V. Steric course of hydrogen transfer during enzymatic formation of 3α-heteroyohimbine alkaloids. *Biochemistry* **22,** 3448–3452 (1983).
33. Knowles, J. R. & Jencks, W. P. The intrinsic pka values of functional groups in enzymes: Improper deductions from the ph-dependence of steady-state parameter. *Crit. Rev. Biochem.* **4,** 165–173 (1976).
34. Guirimand, G. et al. Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? *BMC Plant Biol.* **10,** 182 (2010).
35. Liscombe, D. K. & O'Connor, S. E. A virus-induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*. *Phytochemistry* **72,** 1969–1977 (2011).
36. Brown, S., Clastre, M., Courdavault, V. & O'Connor, S. E. De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl Acad. Sci. USA* **112,** 3205–3210 (2015).
37. Konno, K., Hirayama, C., Yasui, H. & Nakamura, M. Enzymatic activation of oleuropein: A protein crosslinker used as a chemical defense in the privet tree. *Proc. Natl Acad. Sci. USA* **96,** 9159–9164 (1999).
38. Pankoke, H., Buschmann, T. & Müller, C. Role of plant beta-glucosidases in the dual defense system of iridoid glycosides and their hydrolyzing enzymes in *Plantago lanceolata* and *Plantago major*. *Phytochemistry* **94,** 99–107 (2013).
39. Frelin, O. et al. A directed-overflow and damage-control N-glycosidase in riboflavin biosynthesis. *Biochem. J.* **466,** 137–145 (2015).
40. Qu, Y. et al. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc. Natl Acad. Sci. USA* **112,** 6224–6229 (2015).
41. Berrow, N. S. et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35,** e45 (2007).
42. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 (2010).
43. Evans, P. R. & Murshudov, G. N. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 1204–1214 (2013).
44. Winter, G. Xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43,** 186–190 (2010).
45. Winter, G. & McAuley, K. E. Automated data collection for macromolecular crystallography. *Methods* **55,** 81–93 (2011).
46. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. Sect. A* **64,** 112–122 (2008).
47. Delagenière, S. et al. ISPyB: an information management system for synchrotron macromolecular crystallography. *Bioinformatics* **27,** 3186–3192 (2011).
48. Fisher, S. J., Levik, K. E., Williams, M. A., Ashton, A. W. & McAuley, K. E. SynchWeb: a modern interface for ISPyB. *J. Appl. Crystallogr.* **48,** 927–932 (2015).
49. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D* **62,** 1002–1011 (2006).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).
51. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374,** 300–321 (2003).
52. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
53. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,** 845–858 (2015).
54. Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39,** 109–111 (2006).
55. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21 (2010).
56. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 386–394 (2011).
57. Sun, J., Baker, A. & Chen, P. Profiling the indole alkaloids in yohimbe bark with ultra-performance liquid chromatography coupled with ion mobility quadrupole time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **25,** 2591–2602 (2011).
58. Guirimand, G. et al. Optimization of the transient transformation of *Catharanthus roseus* cells by particle bombardment and its application to the subcellular localization of hydroxymethylbutenyl 4-diphosphate synthase and geraniol 10-hydroxylase. *Plant Cell Rep.* **28,** 1215–1234 (2009).
59. Guirimand, G. et al. Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *J. Plant Physiol.* **168,** 549–557 (2011).
60. Waadt, R. et al. Multicolor bimolecular fluorescence complementation reveals simultaneous formation of alternative CBL/CIPK complexes in planta. *Plant J.* **56,** 505–516 (2008).
61. Geu-Flores, F. et al. An alternative route to cyclic terpenes by reductive cyclization. *Nature* **492,** 138–142 (2012).
62. Kearse, M. et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).
63. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).
64. Gouy, M., Guindon, S. & Gascuel, O. Seaview version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27,** 221–224 (2010).

65. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323 (2003).
66. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
67. Morris, G. M. *et al.* Autodock4 and autodocktools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).

Acknowledgements
We gratefully acknowledge support from the ERC (311365) and a BBSRC Institute Strategic Programme grant (SWEET, BB/J004561/1) to S.E.O'C. and from the Region Centre (France, ARSAL grant) to V.C. A.S. is supported by a DBSRC DTP studentship. The Diamond Light Source provided access to beamlines I03, I04 and I04-1 (proposal MX8475).

Author contributions
A.S. and E.C.T. and S.E.O'C. designed the project; A.S., E.C.T. and L.C. performed molecular cloning/enzyme assays; L.C., A.S., C.R.M.S. and D.M.L. assisted with crystallization, X-ray data acquisition and structure refinement; E.C.T. and L.C. performed ViteG; E.C.T. performed NMR structural characterization; F.F. and V.C. performed all localization experiments; S.E.O'C. supervised the work; A.S. and E.C.T. and S.E.O'C. wrote the manuscript with input from all authors.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions

How to cite this article: Sarrinikolas, A. *et al.* Structural investigation of heterocephalidine alkaloid synthesis reveals active site elements that control stereoselectivity. *Nat. Commun.* 7:12116 doi: 10.1038/ncomms12116 (2016).