

**β -amyirin synthase: Investigating the
structure and function of an
oxidosqualene cyclase involved in disease
resistance in oats**

by

Melissa Dokarry

A thesis presented for the degree of Doctor of Philosophy at the
University of East Anglia.

©2010

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis, nor any information derived there from, may be published without the author's prior, written consent.

Declaration

I declare that the work contained in this thesis, submitted by me for the degree of Ph.D., is my own original work, except where due reference is made to other authors, and has not been submitted by me for a degree at this or any other university.

Melissa Dokarry

Acknowledgements

I would like to thank both of my supervisors, Dr Andrew Hemmings and Professor Anne Osbourn for all of their help and support over the last four years, through the good times and the bad! I would also like to thank all the members of the Osbourn and Hemmings labs both past and present for all their advice, tips and for making my time in the lab enjoyable! Especially Richard Hughes – for teaching me everything you could ever want to know about cloning and protein expression; Ariane Kemen – for all her wise words and invaluable advice; Sam Mugford – for keeping me sane during my western blot dramas and Amorn Owatworakit – for his good advice.

Finally, I would like to thank my family – especially my parents, sister and my boyfriend Andy – for their continuing support throughout all of my studies.

Table of Contents

Abstract.....	1
Chapter 1 - Introduction	2
1.1 <i>Cereals and disease resistance</i>	<i>2</i>
1.1.1 The Gramineae: a family of the plant kingdom	2
1.1.2 Susceptibility of cereals to pathogenic fungi	2
1.1.3 Resistance mechanisms in plants	3
1.2 <i>Secondary metabolites</i>	<i>5</i>
1.2.1 Secondary metabolites in plants	5
1.2.2 Saponins and disease resistance.....	6
1.2.3 Avenacins: Structure and function	8
1.3 <i>Pathway to avenacin production</i>	<i>9</i>
1.3.1 Primary and secondary metabolism in plants.....	9
1.3.2 Sterol and triterpene synthesis: a branchpoint between primary and secondary metabolism.....	9
1.3.3 Applications of characterising secondary metabolic pathways	13
1.3.4 Elucidating biosynthetic pathways in plants.....	14
1.3.5 Identification of genes involved in avenacin biosynthesis in oats.....	15
1.4 <i>Oxidosqualene cyclases in plants, animals and fungi</i>	<i>26</i>
1.4.1 Cloning of oxidosqualene cyclases involved in sterol production.....	26
1.4.2 Cloning of oxidosqualene cyclases involved in triterpene production.....	27
1.4.3 Alternative strategies for cloning plant oxidosqualene cyclase genes	29
1.5 <i>Structure-function relationships in plant OSCs.....</i>	<i>30</i>
1.5.1 Conserved features: vast product differences	30
1.5.2 Solving the structures of triterpenoid cyclases.....	31
1.5.3 Insights into enzyme mechanism	33
Chapter 2 - Homology modelling of AsbAS1	37
2.1 <i>Introduction</i>	<i>37</i>
2.1.1 Homology modelling of oxidosqualene cyclases	37
2.1.2 Site directed mutagenesis of oxidosqualene cyclases.....	38
2.1.3 Molecular dynamics simulations.....	38
2.1.4 Aims	39
2.2 <i>Materials and methods</i>	<i>39</i>
2.2.1 Modelling the structure of AsbAS1	39
2.2.2 Sequence and phylogenetic analysis	41
2.2.3 Molecular dynamics simulations.....	41
2.3 <i>Results and Discussion</i>	<i>42</i>
2.3.1 Modelling the structure of AsbAS1	42
2.3.2 Analysis of the initial stages of AsbAS1 mechanism.....	51
2.3.3 Analysis of the cyclisation steps in AsbAS1 mechanism.....	57

Chapter 3 - Genetic characterisation of <i>sad1</i> mutants	67
3.1 <i>Introduction</i>	67
3.1.1 Characterisation of the original <i>sad1</i> mutants.....	67
3.1.2 Identification of additional <i>sad1</i> mutants from the oat mutant collection.....	67
3.1.3 Aims	68
3.2 <i>Materials and Methods</i>	68
3.2.1 Isolation of genomic DNA from <i>sad1</i> mutants.....	68
3.2.2 PCR amplification of <i>Sad1</i> genomic DNA	68
3.2.3 Sequencing of <i>sad1</i> genomic DNA segments	69
3.2.4 Diversity Array Technology (DArT) analysis of <i>sad1</i> mutants	69
3.2.5 Transcript analysis of <i>sad1</i> mutants	70
3.3 <i>Results and Discussion</i>	71
3.3.1 Sequence analysis of <i>sad1</i> mutants.....	71
3.3.2 Transcript analysis of <i>sad1</i> mutants	80
 Chapter 4 - Protein analysis of <i>sad1</i> mutants	 87
4.1 <i>Introduction</i>	87
4.1.1 Genetic analysis of <i>sad1</i> mutants.....	87
4.1.2 Aims	87
4.2 <i>Materials and Methods</i>	87
4.2.1 Cloning of <i>Sad1</i> into pET-14B vector	87
4.2.2 Expression of pET-14b- <i>Sad1</i> in <i>Escherichia coli</i> BL21(DE3) cells	88
4.2.3 Purifying AsbAS1 from inclusion bodies	89
4.2.4 SDS-PAGE gel purification of AsbAS1 and antibody production	89
4.2.5 Protein analysis of <i>sad1</i> mutants	90
4.3 <i>Results and Discussion</i>	91
4.3.1 Generation of AsbAS1 antibody	91
4.3.2 Western blot analysis of <i>sad1</i> mutants	100
 Chapter 5 - Expression and purification of AsbAS1 in <i>Pichia pastoris</i> .	 108
5.1 <i>Introduction</i>	108
5.1.1 Cloning and purification of oxidosqualene cyclases	108
5.1.2 Purification of hOSC in <i>Pichia pastoris</i>	109
5.1.3 Challenges in purifying membrane proteins.....	110
5.1.4 Aims	111
5.2 <i>Materials and Methods</i>	111
5.2.1 Cloning of <i>Sad1</i> into pPICZB vector.....	111
5.2.2 Transformation into <i>Pichia pastoris</i>	112
5.2.3 Small scale expression trials of AsbAS1 <i>P. pastoris</i> X-33 transformants.....	112
5.2.4 SDS-PAGE analysis of protein expression.....	113
5.2.5 Western blot analysis of protein expression.....	113
5.2.6 GC-MS analysis of <i>in-vivo</i> protein activity	114
5.2.7 Large scale expression of AsbAS1	114
5.2.8 Extraction and solubilisation of AsbAS1 protein.....	115
5.2.9 Immobilised metal affinity chromatography	115
5.2.10 Gel filtration.....	116
5.2.11 Hydrophobic interaction chromatography.....	116
5.3 <i>Results and Discussion</i>	116
5.3.1 <i>Pichia pastoris</i> expression system	116

5.3.2	Design of expression constructs.....	118
5.3.3	Cloning of <i>Sad1</i> into the pPICZB vector	118
5.3.4	Expression of pPICZB-BAS1, BAS2 and BAS4 in <i>Pichia pastoris</i> X-33	120
5.3.5	In-vivo functional assay of AsbAS1 activity in <i>Pichia</i> X-33 transformants	124
5.3.6	Large scale expression of pPICZB-BAS2.1:X-33	129
5.3.7	Immobilised metal affinity chromatography of pPICZB-BAS2.1:X-33	129
5.3.8	Secondary purification of pPICZB-BAS2.1:X-33	133
5.3.9	Repeat purification problems of pPICZB-BAS2.1:X-33	136
5.3.10	Expression of pPICZB-BAS1, BAS2 and BAS4 in <i>Pichia pastoris</i> KM71H.....	137
 Chapter 6 - Conclusion		140
 Chapter 7 - Future work.....		145
 Appendices		148
<i>Appendix 1: BLAST search of AsbAS1 against the UniProt database</i>		<i>148</i>
<i>Appendix 2: BLAST search of AsbAS1 against the Protein Data Bank</i>		<i>150</i>
<i>Appendix 3: Structure of original AsbAS1 model</i>		<i>152</i>
<i>Appendix 4: Table of characterised plant oxidosqualene cyclases</i>		<i>153</i>
<i>Appendix 5: Multiple sequence alignment of characterised plant OSCs.....</i>		<i>157</i>
<i>Appendix 6: Genomic DNA sequence of Sad1</i>		<i>172</i>
<i>Appendix 7: Original data from DArT analysis of sad1 mutants</i>		<i>175</i>
<i>Appendix 8: Western blot analysis of sad1 PTC mutants</i>		<i>178</i>
 References.....		179

Abstract

Plants synthesise a diverse range of secondary metabolites that have a wide range of functions from pigmentation to plant defence. Oxidosqualene cyclases (OSCs) catalyse the first committed step in triterpenoid secondary metabolite synthesis and cyclise 2,3-oxidosqualene to a number of cyclic products. This work describes the characterisation of β -amyrin synthase from the diploid oat *Avena strigosa* (AsbAS1). AsbAS1 catalyses the first committed step in synthesis of Avenacin A-1, an antimicrobial triterpene glycoside that confers broad spectrum disease resistance. To date, functional analysis of plant OSCs has been studied via site-directed mutagenesis in yeast and inhibitor studies. This work describes the first set of OSC mutants isolated through a forward screen for loss of function.

A homology model of AsbAS1 was generated based on structural information from human lanosterol synthase (hOSC). Each step of the cyclisation process was explored in detail and AsbAS1 was found to have conserved catalytic residues that are seen in all OSCs, as well as two amino acid residues, Trp257 and Phe259, that are implicated in the formation of pentacyclic products. AsbAS1 contains catalytic regions that are unique to monocot triterpene OSCs suggesting that they have evolved separately to dicot triterpene OSCs.

Seventeen AsbAS1 (*sad1*) mutants identified from a forward screen for avenacin deficiency were characterised at the gene, transcript and protein level. All mutants contained single point mutations in the *Sad1* gene which introduced predicted premature termination of translation codons, splicing errors or amino acid substitutions into the *Sad1* coding sequence. Only two *sad1* mutants, #358 and #384/#1023, produced full-length AsbAS1 protein and the mutations, both located in the active site, are likely to affect reaction initiation and D-ring stabilisation respectively. The location of the substitution mutations implies that OSCs are very tolerant of mutations in areas of the enzyme that are not critical for function and only mutations that seriously affect enzyme expression or function result in premature termination of metabolic pathways.

To date, no triterpene OSCs have been purified on a large scale from heterologous expression. AsbAS1 was cloned and heterologously expressed in *Pichia pastoris*. An *in vitro* assay was developed to screen for functional transformants and a purification strategy using metal affinity chromatography and gel filtration enabled partial purification of tagged AsbAS1.

Chapter 1 - Introduction

1.1 Cereals and disease resistance

1.1.1 The Gramineae: a family of the plant kingdom

The family *Gramineae*, also known as *Poaceae*, is part of the kingdom *Plantae* and is found in the subclass Liliopsida (monocots) of the flowering plants. Members of the *Gramineae* family, around 10,000 species, are commonly referred to as grasses and are found across all parts of the world in every kind of habitat. This widespread occurrence of grasses is due to their ability to adapt to a range of soil types and climates, to compete successfully with other plant types and to survive high levels of predation (1). The grasses that are grown for human consumption are known as cereals. In 2005 cereals provided 46% of the total worldwide calorie consumption and most of the worldwide carbohydrate either directly via cereal grain from crops such as wheat, rice, maize, barley, oats, sorghum and millet, or indirectly through animal feed. Many parts of the world rely on cereals as their major food source, for example, rice in Asia and millet in Africa. The three most common cereals produced throughout the world are maize (*Zea mays*), wheat (*Triticum aestivum*) and rice (*Oryza sativa*) and together they account for 86% of all grain production worldwide (2-4).

1.1.2 Susceptibility of cereals to pathogenic fungi

Like all other plants, cereals are susceptible to colonisation by pathogenic fungi. For example, *Rhynchosporium secalis* causes leaf scald, *Tapesia yallundae* causes eyespot and *Claviceps purpurea* causes ergot. The fungus *Gaeumannomyces graminis* var. *tritici*, commonly known as the Take-all fungus, causes one of the most serious diseases of cereals and mainly affects wheat and barley (5). Take-all is a large problem in the United Kingdom and currently the only option for control of the disease is crop rotation, as chemical treatments are unreliable and there is no good source of genetic resistance to Take-all in cultivated wheat. In the United Kingdom it is estimated that over half of the wheat crop is affected and suffers losses of 5-20%, with up to 50% loss if the disease is severe. The estimated loss to farmers is £60 million a year (6). The plant can be infected throughout its growth and development but infection is usually most severe in the third or fourth successive cereal crop (7). The fungus begins its infection from inoculums in the soil and then spreads along the root via

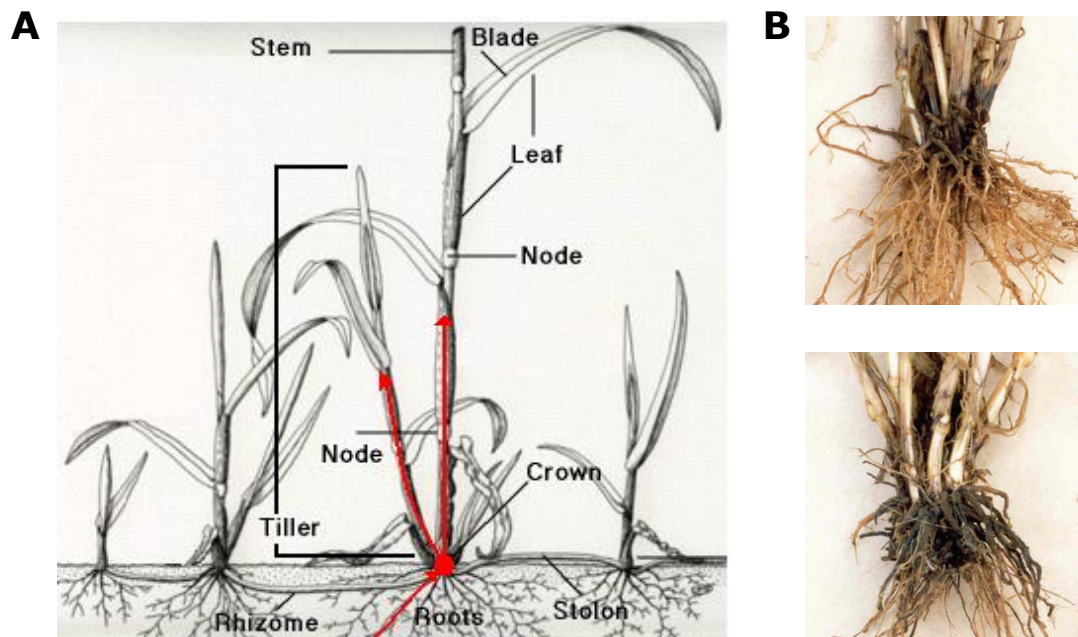


Figure 1.1: Colonisation of cereals by the Take-all fungus

A. Shows the path of Take-all infection as the fungus spreads through cereal plants. The fungus enters the plant through the cortical cells of the roots. It travels upwards through the plant until it reaches the crown and from there it spreads up the main stems and tillers. Image adapted from <http://www.ipm.ucdavis.edu>. **B.** The effect of Take-all on wheat roots. Compared to healthy roots (above), diseased roots (below) have black lesions. Images taken from Bateman *et al.* (2006) (6).

runner hyphae. It then enters the cortical cells of the root, stem and tillers of the plant, blocks the vascular system which cuts off the supply of nutrients and water causing death. Phenotypes of the disease include stunted growth, fewer tillers, premature ear formation, bleached ears, blackened brittle roots and empty spikelets (7-8).

Whilst most cereals are susceptible to Take-all, oats (*Avena spp.*) have developed extreme resistance. Oats are susceptible to the var. *avenae* strain of the fungus, but this strain is rare (6). Biochemical characterisation has shown oats to contain secondary metabolites called saponins which have potent antimicrobial activity and have been implicated with a role in plant defence. Avenacin deficient mutants of oat are susceptible to infection by *G. graminis* var. *tritici* providing evidence to support a role for these compounds in disease resistance (9-10).

1.1.3 Resistance mechanisms in plants

Despite fungi having developed many methods to invade and propagate within cereals and other plants, plants have evolved many unique strategies to evade and defend these attacks. Unlike animals, plants cannot move to escape so they have developed defence mechanisms. These defences can be split into two categories, preformed and induced. Preformed defences are the first line of

attack against pathogens, but if these defences fail then the plant can activate induced defence mechanisms in order to survive (11).

Preformed defences are both structural and chemical. The first barrier the fungus must overcome is the wax layer present on the leaf surface which together with leaf hairs prevent the formation of water droplets that are essential for fungal spore germination (11). The layers below the cuticle are formed of fatty acids and cellulose woven into a high tensile fabric making entry into the plant very difficult (12). The waxy cuticle protects against airborne spores, but spores can also enter the plant via the soil. As the root tips elongate they are vulnerable to pathogen attack, so are guarded by root border cells which secrete anthocyanins and antimicrobial antibiotics to repel pathogens (11). Plants also have a wide range of preformed antimicrobial compounds known as phytoanticipins. Phytoanticipins are low molecular weight, antimicrobial compounds that are present in plants before invasion by microorganisms or are produced after infection solely from pre-existing constituents (13-14). Phytoanticipin distribution is usually tissue specific and these compounds are often found in the outer cell layers of healthy plants, sequestered in vacuoles or organelles. Protective compounds of this kind include avenacins and other saponins, cyanogenic glycosides and phenolic compounds (14).

Induced defences can be activated by the presence of a fungus on an external plant surface whilst others are only switched on once the fungus has penetrated the cell wall. Cytoskeletal rearrangements are one of the defences activated prior to fungal penetration, resulting in a build-up of cytoplasm and plant microtubules at the site of contact. The inner surface of the cell wall is reinforced with compounds such as callose, extensin and lignin which form an insoluble barrier. Peroxidase enzymes also act to strengthen the cell wall. One of the first responses to be activated once penetration has occurred is the production of enzymes such as chitinases and glucanases which degrade fungal cell walls. Another preliminary response is the oxidative burst that results in the production of hydrogen peroxide which is converted to reactive oxygen species (ROS) and leads to activation of signalling pathways for systemic defence responses. Other chemical mediators of defence signalling pathways include salicylic acid, jasmonic acid, ethylene and nitric oxide. The spread of pathogens can be blocked by the hypersensitive response in which there is localised death of plant cells surrounding the site of pathogen entry. This restricts the spread of the fungus and prevents diffusion of fungal toxins (11-12, 14). Plants have also evolved resistance (R) genes which provide gene-for-gene resistance against pathogens with the corresponding avirulence (Avr) genes. This triggers signal

transduction pathways which activate defence responses such as the hypersensitivity response, synthesis of antimicrobial proteins and metabolites and cell wall thickening (15-16). Plants can also induce the production of antimicrobial compounds known as phytoalexins which are defined as low molecular weight, antimicrobial compounds that are both synthesised by and accumulated in plants after exposure to microorganisms (17). These compounds are synthesised by the isoprenoid and shikimic acid pathways which predominantly synthesise primary metabolic compounds, but can be diverted to form secondary metabolites phytoalexins such as terpenoids, phenolics and nitrogen and sulphur compounds (14).

1.2 Secondary metabolites

1.2.1 Secondary metabolites in plants

As mentioned in the previous section plants can synthesise a number of chemical compounds, both preformed and induced, and these may provide protection against attack from fungal pathogens. These chemical compounds belong to a family known as secondary metabolites. These compounds are not required for the growth and development of the plant and are not essential for survival, unlike primary metabolites, hence are termed "secondary". Although

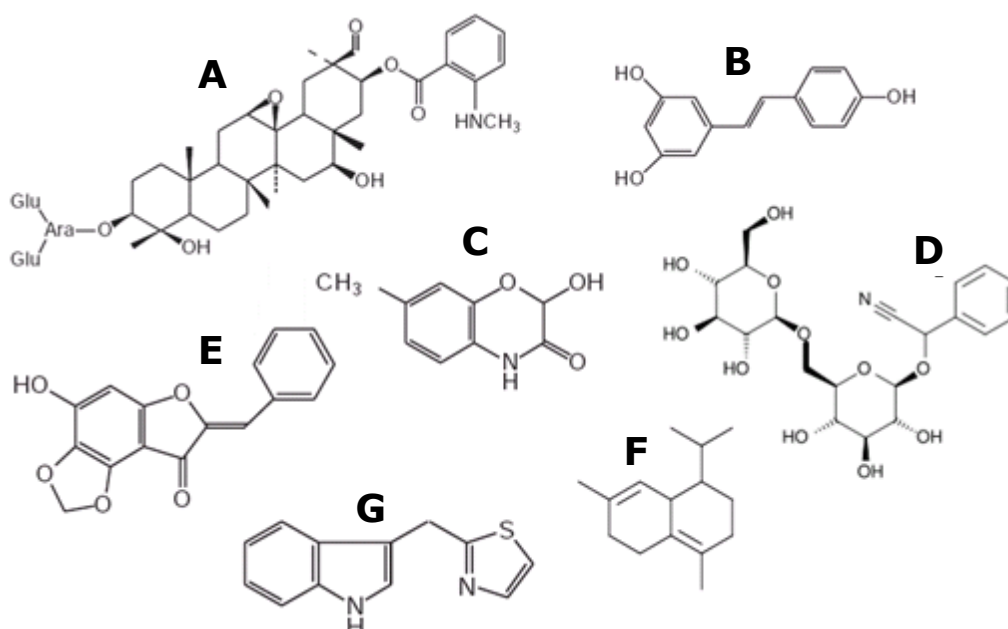


Figure 1.2: Groups of plant secondary metabolites

Chemical diversity of natural products in plants. **A.** Saponins – avenacin A-1 (*Avena strigosa*). **B.** Phenolics – resveratrol (*Vitis viniferis*). **C.** Cyclic hydroxamic acids – DIMBOA (*Zea mays*). **D.** Cyanogenic glycosides – luteone (*Lupinus albus*). **E.** Sesquiterpenes – cadinene (*Juniperus oxycedrus*). **F.** Sulphur-containing indole derivatives – camalexin (*Arabidopsis thaliana*).

not crucial for plant survival under standard growth conditions, they may confer selective advantages in return, for example, through protection against disease and repressing neighbouring plant growth. Many secondary metabolites also have antimicrobial properties so have been implicated with a role in plant defence (9, 18-20). This family of compounds is chemically diverse and includes groups such as saponins, phenolics, cyclic hydroxamic acids, cyanogenic glycosides, isoflavonoids, sesquiterpenes, sulphur-containing indole derivatives and many others (Figure 1.2) (19). The medical community has begun to refocus on natural products as sources of antimicrobial compounds. Antibiotics isolated from microorganisms are rapidly becoming obsolete due to the appearance of resistant pathogen strains. Laboratory designed antimicrobials can sometimes be toxic or have multiple side effects and therefore are not suitable for clinical use; so plant natural products are now one group of compounds at the forefront of antibiotic development. Natural products provide a large diverse resource of compounds that are proven to be effective against bacteria and other pathogens, and by chemically and enzymatically modifying them, more potent antimicrobials that can combat resistant strains are being developed (21-22).

1.2.2 Saponins and disease resistance

Saponins are a major family of secondary metabolites and are found in many plant species. These compounds have surfactant properties and form soap-like foams in aqueous solution, a characteristic that gave rise to the name saponin which is derived from the latin word for soap, *sapo* (23-24). Saponins consist of glycosylated steroids, steroidal alkaloids and triterpenoids and each are found in distinct areas of the plant, often as complex mixtures (Figure 1.3). The basic structure of these molecules consists of an aglycone, consisting of a triterpenoid 5-ring system, and an oligosaccharide chain at the C-3 position which is formed of up to five sugar molecules. The aglycone can also be further modified by glycosylation, acylation or oxidation steps (24-26). The saponin content in plants can be made up of a complex mixture and is dependent on a number of different factors – genetic background, tissue type, age, health and environment of the plant (25, 27). Saponins have many different properties and have been utilised commercially by the biochemical industry as drugs and medicines, adjuvants, foaming agents, sweeteners, taste modifiers and cosmetics (24). As mentioned previously, many saponins have antimicrobial activity so may protect plants against pathogen attack. A number of studies have shown

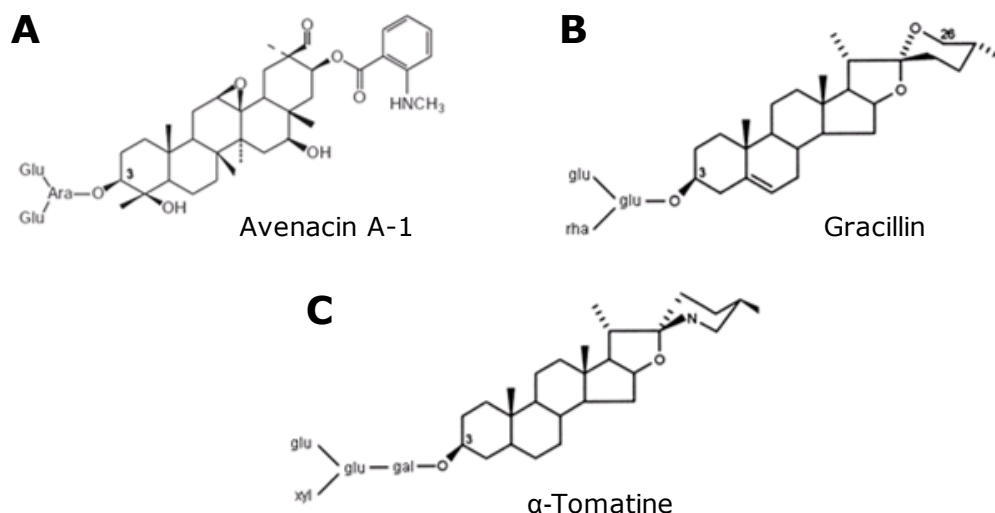


Figure 1.3: Three classes of saponins

Shows example structure from the three classes of saponins that are found in plants. All have a triterpenoid ring system that is modified by glycosylation, acylation, and oxidation steps. **A.** Triterpenoid saponins – Avenacin A-1 (from the roots of *Avena strigosa*). **B.** Glycosylated steroidal saponins – Gracillin (from the rhizomes of *Costus speciosus*). **C.** Steroidal alkaloid saponins – α -Tomatine (stems, leaves and unripe fruit of *Solanum lycopersicum*).

saponins to be biologically active against a range of plant, animal and human pathogens. Recent examples include four saponins from the roots of *Cephalaria ambrosioides* (28), two saponins from the roots of *Scrophularia ningpoensis* (29) and two saponins from extracts of *Tribulus terrestris* (30).

One example of a saponin that has been characterised and shown to have antimicrobial activity is avenacin A-1 from oat roots. Oats are immune to attack by the fungus *Gaeumannomyces graminis* var. *tritici* and this was thought to be due to the presence of the saponin avenacin in its roots. One species of oats, *Avena longiglumis* lacks avenacin and was found to be susceptible to infection by the *G. graminis*, suggesting that presence of avenacin was the determining factor in fungal resistance (20). Further evidence came from both gene-knockouts experiments in fungi and genetic analysis of oat mutants with increased susceptibility to disease (18, 31). Oats, although immune to the var. *tritici* strain of the fungus *Gaeumannomyces graminis*, are susceptible to the var. *avenae* strain. This strain is able to infect oats due to the presence of the enzyme avenacinase which detoxifies avenacin by removing the two glucose molecules from the oligosaccharide chain attached to C-3 of the aglycone. Fungal gene-knockout mutants that lacked the avenacinase enzyme lost pathogenicity on oats but were still pathogenic to wheat – which does not synthesise saponins (31). The second line of evidence came from chemical mutagenesis experiments carried out in the diploid oat *Avena strigosa*. Out of the ten independent mutants

isolated, eight lacked avenacins and the remaining two had reduced avenacin levels. When challenged with the *G. graminis* var. *tritici* fungus, all ten showed disease symptoms with the eight mutants lacking avenacins being more susceptible than the two with reduced avenacin levels (18). Together these two experiments provided a clear link between saponins and disease resistance.

1.2.3 Avenacins: Structure and function

The avenacins, avenacin A-1, B-1, A-2 and B-2 are a family of four structurally related triterpenoid saponins that are found in oat roots. Avenacin A-1 is the most abundant, comprising 70% of the total avenacin content in young oat roots (32). The structure of the four avenacins is very similar with each consisting of a pentacyclic aglycone which is modified by β -1,2- and β -1,4-linked D-glucose molecules attached via L-arabinose to the aglycone at the C-3 carbon. Avenacins A-1 and B-1 have a fluorescent *N*-methyl anthranilate group attached to the aglycone via an ester bond to C-21 whereas avenacins A-2 and B-2 have a benzoic acid in the corresponding position which conversely emits little or no fluorescence. In addition to the esterified groups avenacins A-1 and A-2 also have a hydroxide group at C-23 (24, 33) (Figure 1.4). Avenacins are localised in the epidermal cells of young oat root tips, so are in an ideal position to provide protection against soil borne pathogens (20).

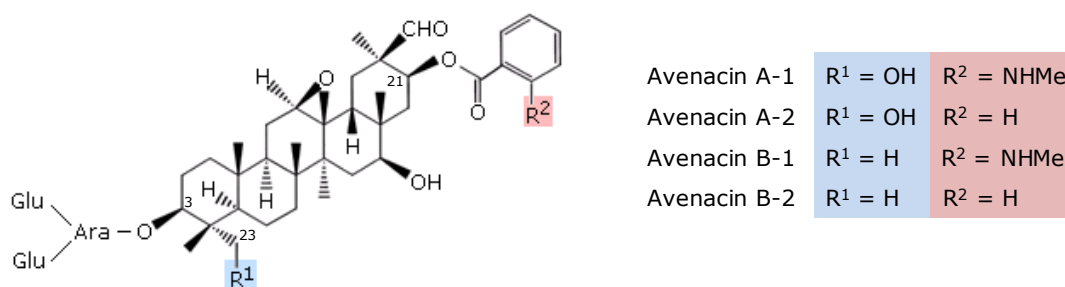


Figure 1.4: Structure of avenacin

Shows the common structural features of the four avenacins found in oat. All four avenacins have a pentacyclic aglycone modified at a number of positions by oxidation, glycosylation and acylation. All have an oligosaccharide sugar chain attached to the aglycone at C-3 and addition of an esterified group at C-21. The identity of the esterified group at C-21 (R²) and the group attached to C-23 (R¹) varies between each of the avenacins and is shown in the table above.

Avenacins confer their antifungal properties by forming complexes with sterols in fungal membranes. The aglycone portion of the avenacin molecule inserts into the membrane where it binds a cholesterol molecule. This triggers the formation of a transmembrane pore via aggregation of the sugar moieties on avenacin which initiates lipid bilayer rearrangements causing loss of membrane

integrity and death of the fungus. This process relies on the presence of a oligosaccharide chain attached to the aglycone at C-3 and removal of even a single sugar from this chain results in a large reduction in avenacin bioactivity (34). In order to protect themselves from the permeabilising affects of these molecules, oats store avenacins in vacuoles (35). Vacuole membranes of avenacin containing cells may have substituted sterols or a low sterol content to prevent the host cell from permeabilising its own membranes (24). This compartmentalisation of avenacin to protect the cell from the toxic effects of the molecule is also seen for other toxic secondary metabolites produced in other plants (36-37).

1.3 Pathway to avenacin production

1.3.1 Primary and secondary metabolism in plants

Metabolism in plants is a complex process involving many enzymes. Solar energy captured in chloroplasts is used for a wide range of metabolic processes to ensure the survival of the plant. Together the complex reactions create metabolic pathways in which the precursor is converted to the product through a series of intermediates known as metabolites (38). Plants synthesise a huge range of metabolites and one of the largest families is the isoprenoids (or terpenoids) consisting of over 30,000 members (39-40). This family is split into two classes of compounds: primary and secondary metabolites. Primary metabolites include hormones (gibberellins and brassinosteroids), photosynthetic pigments (carotenoids), membrane components (sterols) and electron carriers (chlorophylls), and are vital for the functioning of essential plant systems (41). Secondary metabolites, although not vital for plant survival, have functions such as insect attractants (essential oils and flower colours), antimicrobial compounds (phytoanticipins and phytoalexins) and defences against herbivores (antifeedants) (41).

1.3.2 Sterol and triterpene synthesis: a branchpoint between primary and secondary metabolism

The isoprenoid family is classified according to the number of isoprene units in their structure, with one isoprene unit consisting of five carbon atoms and eight hydrogen atoms. The major groups within the isoprenoid family are the monoterpenes (C_{10} - two isoprene units), sesquiterpenes (C_{15} - three isoprene units), diterpenes (C_{20} - four isoprene units) and triterpenes (C_{40} - five isoprene units) (42). Isoprenoid precursors in plants are synthesised by condensation of

isopentenyl diphosphate (IPP) with its isomer dimethylallyl diphosphate (DMAPP) and these two molecules can be synthesised via two different pathways: the mevalonate pathway in the cytosol which synthesises 2,3-oxidosqualene for sesquiterpene, triterpene and sterol formation (41, 43); and the 1-deoxy-D-xylulose-5-phosphate (DXP) pathway in the plastid which synthesises 2,3-oxidosqualene for monoterpene, diterpene, chlorophyll and carotenoid formation (44).

For triterpene and sterol biosynthesis IPP (C_5) is added in a head-to-tail fashion with DMAPP (C_5) to form geranyl diphosphate (GPP) (C_{10}). Further condensation of GPP with IPP forms the larger prenyl diphosphate, farnesyl diphosphate (FPP) (C_{15}) which then dimerises to form squalene (C_{30}) (41, 45). The enzyme squalene epoxidase adds an oxygen across the first carbon-carbon double bond of squalene to form the epoxide, 2,3-oxidosqualene (Figure 1.5) (46). It is at this stage that the sterol and triterpene biosynthetic pathways diverge, with the former producing membrane sterols and hormones (primary metabolites) and the latter producing a diverse range of triterpenes (secondary metabolites). This divergence represents the branchpoint between primary and secondary metabolism and cyclisation of 2,3-oxidosqualene to distinct products is catalysed by the oxidosqualene cyclases (OSCs) (23). OSC catalysis converts a linear molecule of 2,3-oxidosqualene to a polycyclic product.

Cyclisation of 2,3-oxidosqualene by the sterol OSCs, lanosterol synthase and cycloartenol synthase, result in production of the tetracyclic sterol precursors lanosterol and cycloartenol respectively (Figure 1.6). Lanosterol is the precursor to the membrane sterols cholesterol and ergosterol, in animals and fungi respectively; cycloartenol is the precursor to a number of phytosterols in plants, for example sitosterol, stigmasterol and campesterol. With this apparently clear distinction between animal, fungal and plant sterol precursors, it was previously thought that lanosterol synthases were restricted to animals and fungi and cycloartenol synthases to plants, but recent evidence has shown that functional lanosterol synthases also exist in plants and have arisen by convergent evolution from cycloartenol synthase (47-48). The role that lanosterol plays in plant metabolism is currently not understood but the presence of a lanosterol synthase, in addition to a cycloartenol synthase that already provides a route to membrane sterols indicates that it may have an alternative role in plants, possibly in defence responses (47-48).

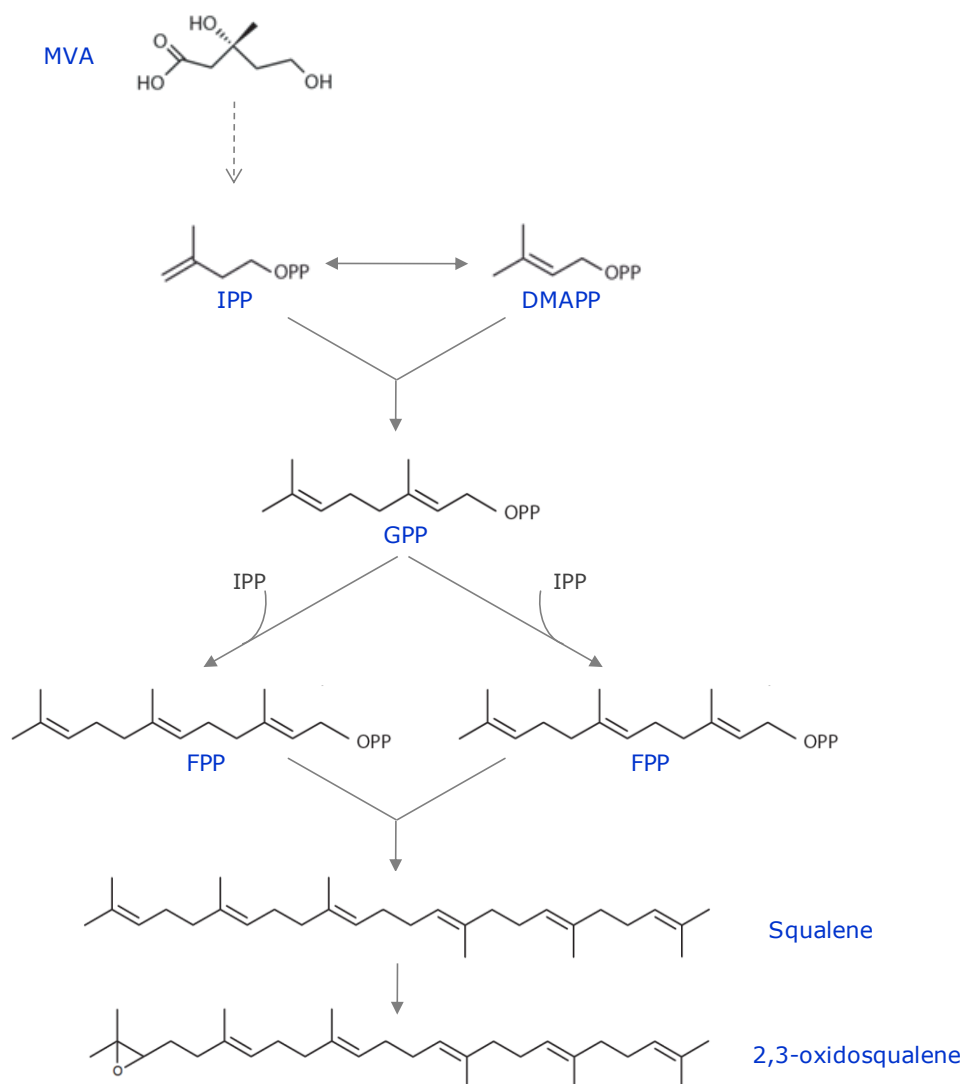


Figure 1.5: Synthesis of 2,3-oxidosqualene via the mevalonate pathway

Pathway to 2,3-oxidosqualene, the sterol and triterpene precursor, via the mevalonate pathway. Mevalonate (MVA) is converted to isopentenyl diphosphate (IPP) which reacts with its isomer dimethylallyl diphosphate (DMAPP) to form geranyl diphosphate (GPP). GPP reacts with another IPP molecule to form farnesyl diphosphate (FPP) which dimerises to form squalene. Addition of an epoxide group across a carbon-carbon double bond forms 2,3-oxidosqualene which can then be cyclised to form sterol or triterpene precursors.

The triterpene OSC family has many more members than the sterol OSC family, therefore cyclisation of 2,3-oxidosqualene by the triterpene OSCs produces a wide range of cyclic products (Figure 1.6). The two major groups of triterpene OSCs identified to date are the β -amyrin synthases and lupeol synthases and these synthesise the pentacyclic products β -amyrin and lupeol respectively. These enzymes are found throughout a wide range of monocot and dicot plants and are precursors for a number of secondary metabolites, for example, avenacin – *Avena strigosa* (49); soyasaponin I – *Lupinus angustifolius*

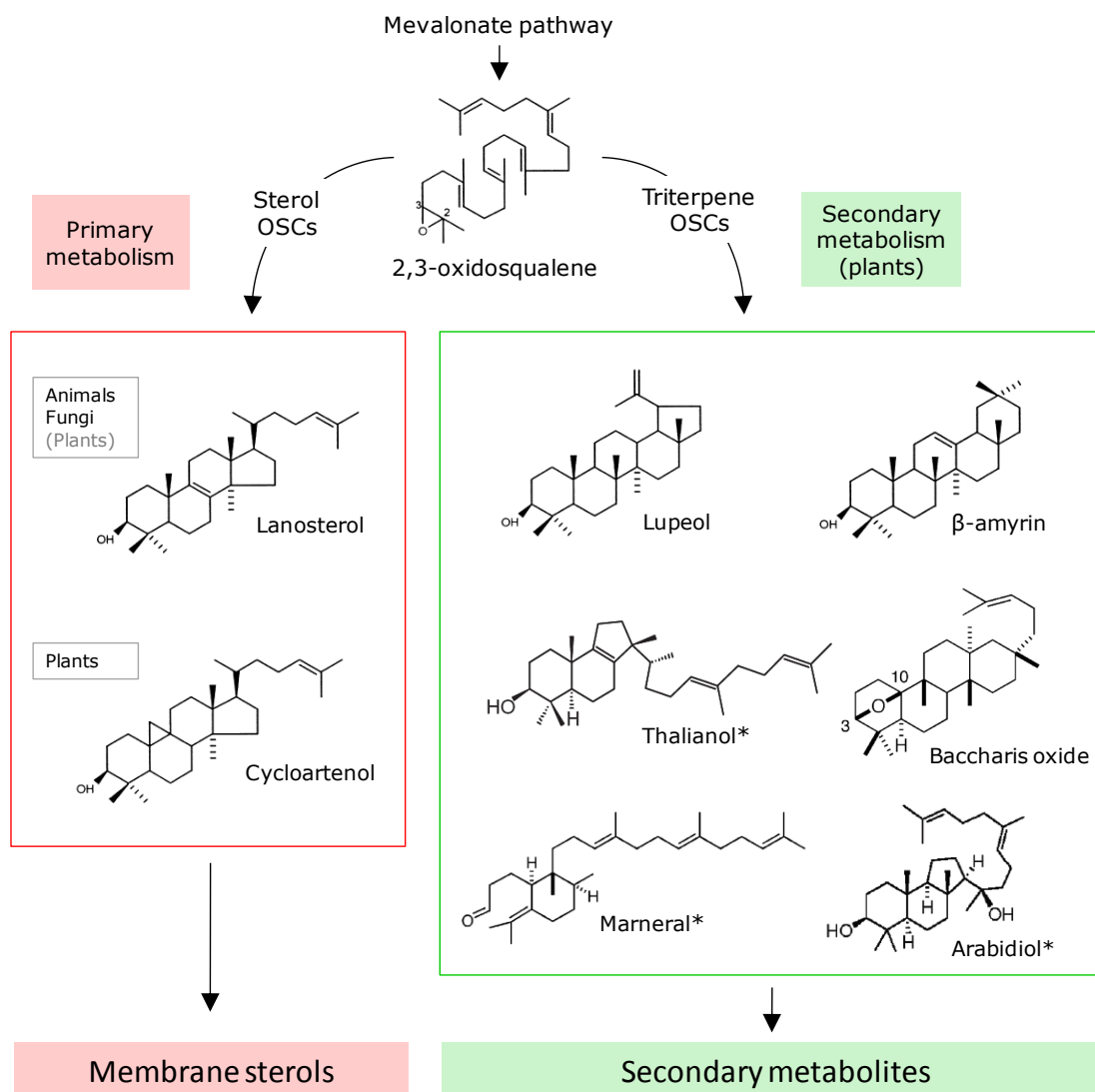


Figure 1.6: Cyclisation of 2,3-oxidosqualene to sterols and triterpenes

2,3-oxidosqualene is situated at the branch point between primary and secondary metabolism. For primary metabolism it is cyclised by the sterol oxidosqualene cyclases (OSCs) to lanosterol in animals and fungi and cycloartenol in plants. Lanosterol has been found in some plants, but it is not thought to participate in primary metabolism (47). Lanosterol and cycloartenol are then modified by a number of enzymes to form membrane sterols. For secondary metabolism, 2,3-oxidosqualene is cyclised by the triterpene OSCs to a wide range of cyclic products. Just a small number of the cyclic products that have been identified in plants are shown and the enzyme that catalyses each reaction is named as the synthase of the product formed, for example, lupeol is synthesised by lupeol synthase. Products labelled with an asterisk are those that have only been observed in *Arabidopsis thaliana*.

(50); and bourneioside A - *Lonicera bournei* (51). Compared to the sterol OSCs, product specificity for triterpene OSCs is not as strict. Many synthesise multiple triterpene products, usually one or two major triterpene products and a number of minor products (52-56). With the recent advances in whole genome sequencing and genome mining, a number of new triterpene OSCs have been identified, mainly in *Arabidopsis thaliana*, that have novel cyclic products. Some, like thalianol synthase (*Arabidopsis thaliana*) (57) are specific enzymes, but the majority are multifunctional and produce a wide range of cyclic products.

Examples include baccharis oxide synthase (*Stevia rebaudiana*), and from *Arabidopsis thaliana*, baurol synthase (58), marneral synthase (59) and arabidiol synthase (60). As more whole genome sequences for plants become available, it is likely that many more new triterpene skeletons will be discovered.

After 2,3-oxidosqualene cyclisation the triterpene backbone undergoes a series of modification steps which can include oxidation, alkylation, glycosylation, and acylation. Elaboration of the triterpene skeleton to produce the final product can take a large number of steps and the pathways enzymes involved are largely uncharacterised. One well characterised pathway is the synthesis of avenacin in *Avena strigosa*. The first step in avenacin synthesis is cyclisation of 2,3-oxidosqualene by β -amyrin synthase. The β -amyrin skeleton is then modified by hydroxylation by at least one cytochrome p450 (61), addition of sugars at C-3 by glucosyltransferases, and addition of an *N*-methyl anthranilate group by a combination of three enzymes, a glycosyltransferase, methyltransferase and a serine carboxypeptidase-like acyltransferase (62).

1.3.3 Applications of characterising secondary metabolic pathways

The mechanism of 2,3-oxidosqualene cyclisation into sterol and triterpene precursors and their subsequent modification to primary and secondary metabolites has been a source of huge interest to both biochemists and the field of industry. However the complexity of the pathway and the many enzymes that are involved have made this a difficult challenge. Currently many research programmes are trying to identify the genes involved in triterpene biosynthesis, however DNA sequence information alone can make it hard to construct complex pathways (19). The elucidation of the avenacin pathway was aided by the generation of a collection of avenacin-deficient mutants and the clustering of the pathway genes in the oat genome (63). Genes for metabolic pathways are often found in clusters in bacterial and fungal genomes but clustering of biosynthetic genes in plants is rare, so elucidation of biosynthetic pathway genes is not as simple (64). To date, only a handful of plant biosynthetic gene clusters have been characterised, but with the continual advances in genetics, identification of biosynthetic pathways is becoming easier (65-67).

As the avenacin pathway appears to be unique to oats, an improved understanding of this complex pathway and transference of pathway genes from oat into other major cereal crops such as wheat, barley, maize and rice could improve disease resistance substantially (63). Many crop plants have been subjected to many years of selective breeding and have therefore lost large numbers of natural products that still exist in their wild counterparts, so gene

transfer to provide disease resistance would appear to be a solution (68). As well as being utilised in plants for disease resistance, triterpenes are also used by pharmaceutical companies as lead compounds for antifungal and cholesterol-lowering drugs, so knowledge of the biosynthetic pathways could greatly enhance progress in this field (69). However, whilst transferring pathway genes to other crops may seem like a simple solution and has been achieved for single resistance genes (70-73), transferring multiple pathway genes between plant species represents a substantial technical challenge. The production of antimicrobial compounds relies on the coordinated involvement of many unrelated genes, so large numbers of genes would have to be transferred to ensure effective antimicrobial activity. How the pathways are regulated is also important, and to date this area is not well understood. There is also the problem of pathogens developing resistance for the introduced antimicrobial compound (19). Despite these potential stumbling blocks, elucidation of the pathway would be a major step forward in understanding the way in which plants fight disease.

1.3.4 Elucidating biosynthetic pathways in plants

Despite the technological advantages that will come from elucidation of individual steps in metabolic pathways, with the exception of the first cyclisation step, the remainder of saponin biosynthetic pathways remain uncharacterised at the molecular level. The first step in saponin and triterpenoid biosynthesis is cyclisation of 2,3-oxidosqualene which is catalysed by triterpene OSCs. These enzymes synthesise the triterpene skeleton which is modified at various positions by subsequent enzymes in the pathway. A large number of triterpene OSCs have been cloned from a wide range of plant species and this will be discussed in a later section.

The gap in knowledge comes at the latter stages of triterpenoid synthesis. A large number of triterpenoids have been isolated from plants and their structures have been determined by mass spectroscopy and NMR, which has allowed speculation about the enzymes that are involved in modifying the triterpene backbone. However very few of these enzymes have been cloned and characterised from plants in comparison to enzymes involved in fungal and bacterial biosynthetic pathways. Elucidation of pathways in fungi and bacteria has been aided by the genes for metabolic pathways being clustered in the genome, a phenomenon that is rarely seen in plants. The spread of metabolic pathway genes throughout the plant genome makes elucidation of pathways more challenging. Despite this, advances have been made in identification of non-

clustered genes involved in saponin biosynthesis in *Medicago truncatula*, *Glycine max* and *Saponaria vaccaria*.

Characterisation of the saponins found in *Medicago truncatula* found that 2,3-oxidosqualene was cyclised to β -amyrin which was further modified by cytochrome P450s to generate five triterpenoid backbones that formed the core of the 37 saponins identified in *M. truncatula*. Clustering analysis of transcript and metabolic profiles identified a candidate P450 for modification of β -amyrin and two glucosyltransferases which had activity against the five triterpenoid backbones (74-75). Other enzymes that have been cloned and characterised include a P450 hydroxylase that modifies a β -amyrin backbone in *Glycine max* (76), and a glucosyltransferase that modifies a carboxylic acid β -amyrin derivative in *Saponaria vaccaria* (77).

1.3.5 Identification of genes involved in avenacin biosynthesis in oats

As mentioned previously, clustering of biosynthetic genes in the genome has greatly aided the elucidation of metabolic pathways. Work carried out by the Osbourn group at the John Innes Centre, Norwich has been focussed on the elucidation of the pathway to avenacin in oats. As discussed previously in section 1.2.2, the triterpenoid saponin avenacin found in oat roots was thought to confer resistance to the fungus *Gaeumannomyces graminis* var. *tritici* (32). To investigate this hypothesis, avenacin-deficient mutants were generated in the diploid oat species *Avena strigosa* by mutagenising seeds with sodium azide. The fluorescent phenotype of avenacin presence was then used to screen the mutagenised oats for avenacin-deficient plants followed by TLC and HPLC analysis of root extracts to quantify avenacin levels in each of the mutants. Ten avenacin deficient mutants, termed *sad* mutants (**s**aponin-**d**eficient), were identified from the first screen. Eight of them had no avenacin or trace amounts of avenacin present and the remaining two had reduced avenacin levels (Figure 1.7A). Test crosses with the wild-type suggested that the mutations were recessive alleles of single genes and analysis of the F₂ progeny found that the ten mutants fell into at least four complementation groups, named *Sad1* to *Sad4*. When inoculated with the pathogen *G. graminis* to which wild type oats are known to be resistant, all ten *sad* mutants developed lesions on their roots whereas the wild-type showed no visible signs of infection (Figure 1.7B). The ten mutants were also susceptible to other fungal pathogens such as, *Fusarium culmorum* and *Fusarium avenaceum* which provided a clear link between saponin deficiency and compromised disease

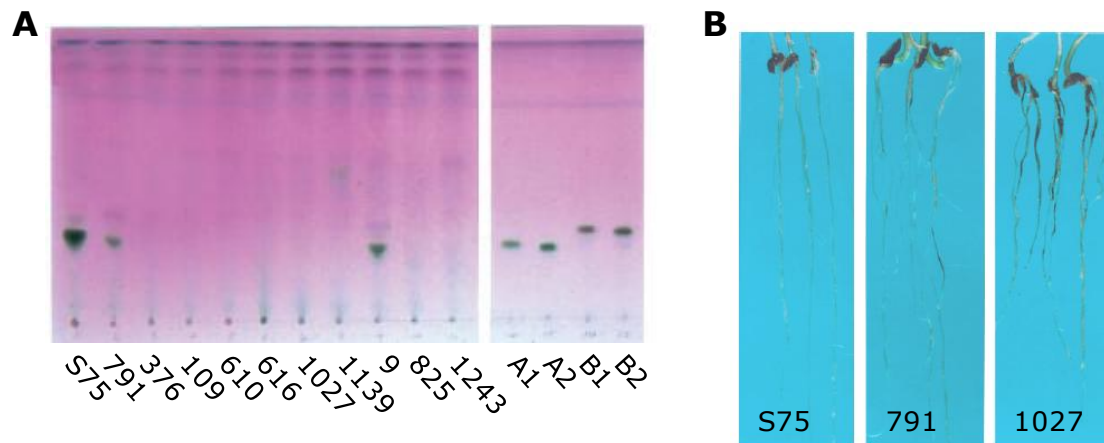


Figure 1.7: Evidence for antimicrobial activity of avenacin A-1

A. TLC analysis of partially purified methanolic root extracts from wild-type (S75) and the ten independent oat mutants using a chromogenic reagent to detect the four avenacins found in oat roots. Preparations enriched for each of the four avenacins are shown on the right. Wild-type extract shows clear presence of avenacin, mutant nos. 791 and 9 have reduced avenacin levels and the remaining eight mutants have no detectable avenacin present. **B.** Oat roots inoculated with *Gaeumannomyces graminis* var. *tritici*. Wild-type (S75) roots show no infection. Mutant no. 791 (reduced avenacin levels) has some black lesions on the roots whereas mutant no. 1027 (no avenacin) has more severe disease symptoms. Figure adapted from Papadopoulou *et al.* (1999) (18)

resistance in oats (18). Subsequent analysis of an extended collection of 92 oat mutants with reduced root fluorescence revealed an additional four loci (*sad6* to *sad9*) that were involved in avenacin biosynthesis (61).

Further genetic analysis showed genetic linkage between the *Sad1* locus and five of the six *sad* loci previously defined by mutation (18, 63). *Sad2*, 6, 7 and 8 were 0 cM from the *Sad1* locus, with the *Sad3* locus was less closely linked at 3.6 cM from the *Sad1* locus. Together these genes form a cluster around the *Sad1* locus (63). Subsequent detailed analysis of a bacterial artificial chromosome (BAC) contig that spanned the *Sad1* locus and surrounding genes revealed that the *Sad1* and *Sad2* loci were 70 kb apart and the *Sad1* and *Sad7* loci were 62 kb apart (Figure 1.8) (61-62). It is known that *Sad1*, 2, 3 and 7 are

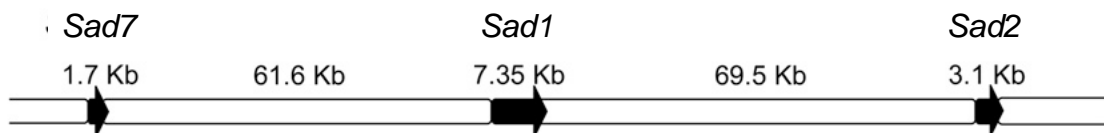


Figure 1.8: Clustering of genes required for avenacin synthesis in the oat genome

Extension of the bacterial artificial chromosome (BAC) contig upstream and downstream of the *Sad1* gene revealed loci for *Sad2* and *Sad7*, both of which are involved in the avenacin biosynthetic pathway. *Sad2* is located 69.5Kb upstream of the *Sad1* gene and *Sad7* is located 61.6Kb downstream of the *Sad1* gene. There are no open reading frames found in the regions between any of the loci. Figure adapted from Mugford *et al.* 2009 (62).

required for the synthesis of avenacin. The enzymes that they encode are distinct and these genes are not tandem duplications (63). Gene clusters of genes that share sequence relatedness are common in plants, for example, leucine rich receptor kinase genes required for specific disease resistance. Clusters which consist of unrelated genes acting in the same biosynthetic pathway, such as the *Sad* gene cluster in oats, are uncommon, but recently a number of functional gene clusters for metabolic pathways have been discovered. These are the cyclic hydroxamic acid (DIBOA) pathway in maize (66, 78-79), the triterpene thalianol pathway in *Arabidopsis* (65) and the diterpenoid momilactone and phytocassane pathways in rice (80-81). As many of the genes found in eukaryotic metabolic pathways are generally unlinked, questions have been raised as to how these gene clusters originated and persisted in the genome. There are many advantages of clustering of genes belonging to a metabolic pathway, for example, by facilitating regulation and coordinated expression of the pathway genes; and promoting the inheritance of genes that confer a selective advantage to the plant. Furthermore, disruption of the cluster may lead not only to loss of the advantageous effect of the pathway but also in accumulation of deleterious intermediates which are detrimental to plant survival (82). How these clusters have assembled in plants is currently not well understood, but is thought to have occurred through gene duplication, acquisition of new function, and genome reorganization rather than horizontal gene transfer from microbes (64-65).

Currently three of the genes in the avenacin biosynthetic cluster have been cloned and characterised, *Sad1*, *Sad2* and *Sad7*; and the *Sad3* and *Sad4* loci have also been partly characterised (35, 49, 61-62). This has allowed elucidation of the individual steps of the avenacin pathway and facilitated the search for gene clusters in cereals and other plants.

1.3.5.1 Sad1 – β -amyrin synthase (AsbAS1)

The first gene from the oat pathway to be characterised was the gene for β -amyrin synthase. Evidence showed that β -amyrin and avenacins are synthesised in the root tip and β -amyrin synthase activity was also observed there (26, 83). Two cDNA libraries were constructed from the root tips of *Avena strigosa* and sequence analysis identified two putative OSC sequences. One had high homology to cycloartenol synthases from other plants so was presumed to be the *Avena strigosa* cycloartenol synthase (AsCAS), but the other sequence was clearly distinct from AsCAS (55% homology) so was postulated to be the *Avena strigosa* β -amyrin synthase (AsbAS1) (49). The putative β -amyrin synthase

cDNA sequence was cloned into the yeast expression vector pYES2 and transformed into the yeast strain GIL77 which accumulates 2,3-oxidosqualene due to a lack of lanosterol synthase (84). HPLC analysis of the yeast cells expressing the putative AsbAS1 cDNA accumulated β -amyrin thus confirming that the cDNA encoded β -amyrin synthase. Amino acid sequence analysis showed that AsbAS1 is distinct from other β -amyrin synthases and demonstrated closer homology to lanosterol synthases from animals and fungi than to other plant cyclases. AsbAS1 was a novel enzyme and therefore defined in a new class of β -amyrin synthases (49).

Further analysis found AsbAS1 to be synonymous with the *Sad1* locus. Radiochemical feeding experiments on the original ten *sad* mutants identified two mutants (#109 and #610) that represented mutant alleles at a locus termed

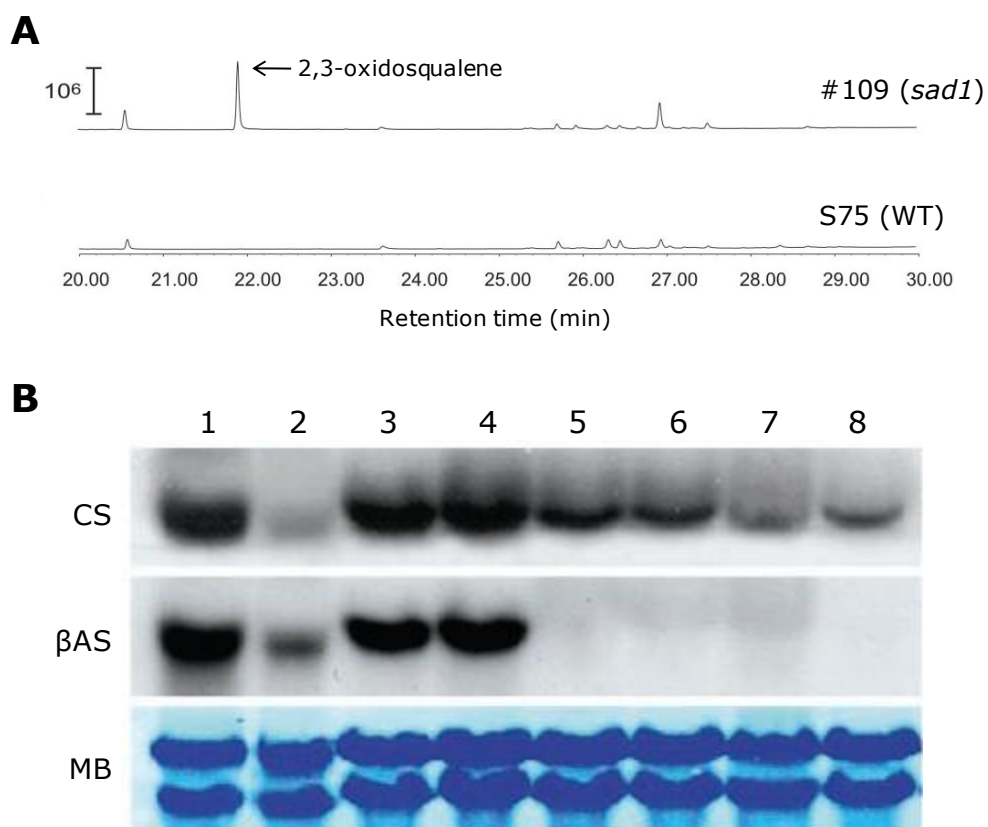


Figure 1.9: Characterisation of the *Sad1* locus

A. Identification and characterisation of *sad1* mutants by GC analysis of root extracts from the wild-type S75 and mutant #109. Mutant #109 accumulates 2,3-oxidosqualene which was identified using MS data. The vertical bar indicates relative mass abundance. Figure adapted from Qi *et al.* (2006) (61). **B.** Northern blot analysis of RNA from roots of different cereals using oat cycloartenol synthase (CS) and oat β -amyrin synthase (β AS) probes. RNA levels were monitored using methylene blue dye (MB). Lane 1, *A. strigosa* S75; lane 2, *A. longiglumis*; lane 3, *A. strigosa* accession no. CI1994; lane 4, *A. strigosa* accession no. CI13815; lane 5, *Triticum aestivum* (cultivar Riband); lane 6, *Hordeum vulgare* (cultivar Golden Promise); lane 7, *Oryza sativa* (accession M12); and lane 8, *Zea mays* (accession P10). Figure adapted from Haralampidis *et al.* (2001) (49).

Sad1. Both of these mutants accumulated 2,3-oxidosqualene and lacked β -amyirin and also had no detectable β -amyirin activity, which indicated that that they were likely to contain mutations in the β -amyirin synthase gene itself or a gene involved in regulation of its expression (Figure 1.9A). Sequencing of the β -amyirin synthase gene in both of the mutants found single point mutations which resulted in premature termination of translation codons being predicted. These results together with single nucleotide polymorphism (SNP) analysis, which showed that the *sad1* phenotype cosegregated absolutely with the mutations in the AsbAS1 gene, provided convincing evidence that *Sad1* locus encoded AsbAS1 (49, 83).

More detailed analysis of the *Sad1* locus found that *Sad1* mapped to a part of the oat genome named *Oisu441* which is found on linkage group AswC (85). This region of the oat genome has overall synteny with chromosome six in the rice genome, however high homology sequences corresponding to *Sad1* or *Oisu441* are not observed. Northern blot analysis of RNA from cereal root extracts confirmed that *Sad1* transcripts are only present in oats whereas cycloartenol synthase expression was found in all cereals tested (Figure 1.9B). Together, this data showed that the β -amyirin synthase gene is found in a region of the oat genome which is not conserved in other cereals (49).

1.3.5.2 *Sad2* – cytochrome p450 sterol demethylase (AsCYP51H10)

In 2006, the second enzyme in the avenacin pathway was cloned from the oat cDNA library and found to be a cytochrome P450 monooxygenase. The gene was identified by extension of the BAC contig surrounding the *Sad1* gene and found to be homologous with cytochrome P450 monooxygenases, specifically obtusifoliol 14 α -demethylase, and was named *AsCYP51H10* according to cytochrome P450 nomenclature. Sequence analysis placed *AsCYP51H10* in the avenacin gene cluster, 66,828 base pairs downstream of *Sad1* with no open reading frames (ORFs) between the two genes. Cytochrome P450s have been implicated in avenacin biosynthesis and as *AsCYP51H10* expression was only found in oat roots, in a similar to pattern to *Sad1* expression, this suggested a role in avenacin biosynthesis (Figure 1.10A) (61).

Further analysis found *AsCYP51H10* to be synonymous with the *Sad2* locus. The *AsCYP51H10* gene in each of the ten original *sad* mutants was sequenced to ascertain whether it corresponded to any of the loci defined by

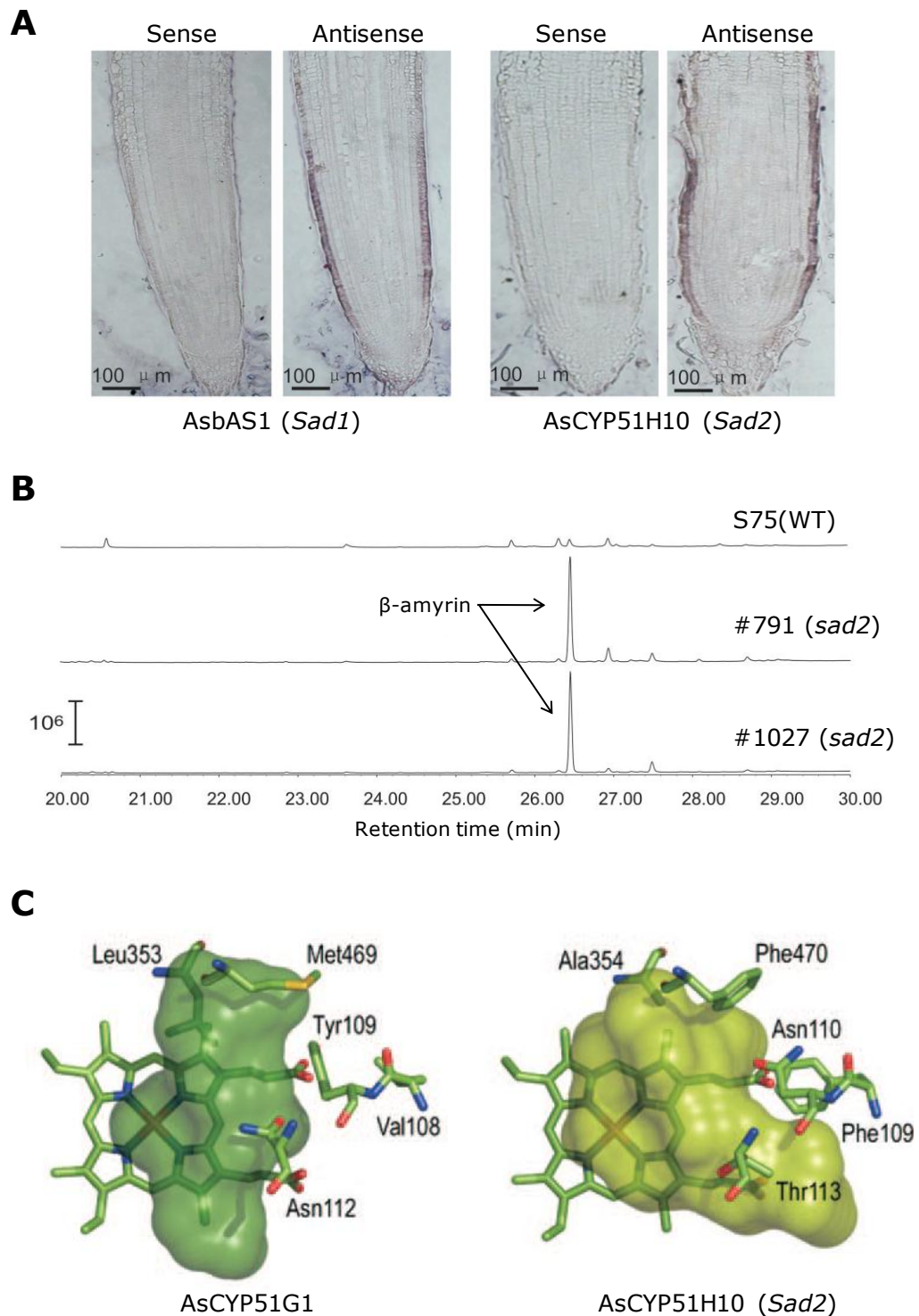


Figure 1.10: Characterisation of the *Sad2* locus

A. *In situ* mRNA analysis of *Sad1* and *Sad2* transcripts in the root tips of *A. strigosa*. Avenacin synthesis is restricted to the epidermal cell layer of the root tip and *AsCYP51H10* expression is restricted to this location indicating that it is involved in avenacin synthesis. **B.** Identification and characterisation of *sad2* mutants by GC analysis of root extracts from the wild-type S75 and mutant #791 and #1027. Mutant #791 and #1027 both accumulate β -amyrin which was identified using MS data. The vertical bar indicates relative mass abundance. **C.** Modelling of the active site cavity of the oat sterol 14 α -demethylase (left) and *AsCYP51H10* (right). *AsCYP51H10* has a larger active site cavity than other sterol demethylases indicating acquisition of a different function. All figures adapted from Qi *et al.* (2006) (61).

mutation. Single point mutations were found in the coding region in two mutants (#791 and #1027) that represented mutant alleles at the *Sad2* locus. Previous radiochemical feeding experiments and GC-MS analysis had shown both of these mutants accumulated β -amyrin indicative of a block in the pathway after the AsbAS1 enzyme (Figure 1.10B). Screening of the extended mutant collection revealed six additional *sad2* mutants which accumulated β -amyrin and also had single point mutations in the *AsCYP51H10* gene. This evidence confirmed that *AsCYP51H10* corresponded to the *Sad2* locus.

Sad2 is a member of the CYP51 sterol demethylase family, an ancient and highly conserved family thought only to be involved in sterol synthesis. Members of this family from bacteria, protozoa, fungi, animals and plants share 34 conserved amino acid residues, however *AsCYP51H10* shares only 28 of these conserved residues with the remaining six residues being divergent. These six residues are found in conserved substrate recognition sites or regions important for enzyme structure and/or function. Active site comparisons show the active site of *AsCYP51H10* to be much larger than other demethylases and this, together with the sequence information indicates that *AsCYP51H10* has acquired a different function from the other enzymes in the CYP51 class (Figure 1.10C). Phylogenetic analysis also indicates that the CYP51H family to which *AsCYP51H10* belongs has evolved from the conserved CYP51G family which is responsible for plant sterol biosynthesis (61). This evidence parallels the evolution of the *Sad1* gene which is predicted to have arisen from a primary sterol cycloartenol synthase-like gene ancestor (63). These results show a connection between primary and secondary metabolism in plants as both *Sad1* and *Sad2* have been linked with two independent steps in the primary sterol pathway. This indicates that enzymes involved in secondary metabolic pathways are likely to have been recruited from ancestral primary metabolic enzymes (61). The specific function of *AsCYP51H10* is currently under investigation in the Osbourn lab and it is known that the enzyme is involved in at least one hydroxylation step to modify the β -amyrin product following AsbAS1 catalysis. However there are a number of oxidation steps that modify the β -amyrin ring structure and it is unknown how many of these steps are catalysed by *AsCYP51H10* (86).

1.3.5.3 *Sad 7* – Serine carboxypeptidase-like acyltransferase (AsSCPL1)

In 2009, the third enzyme in the avenacin pathway was cloned and characterised. Two out of the ten original *sad* mutants (#376 and #616)

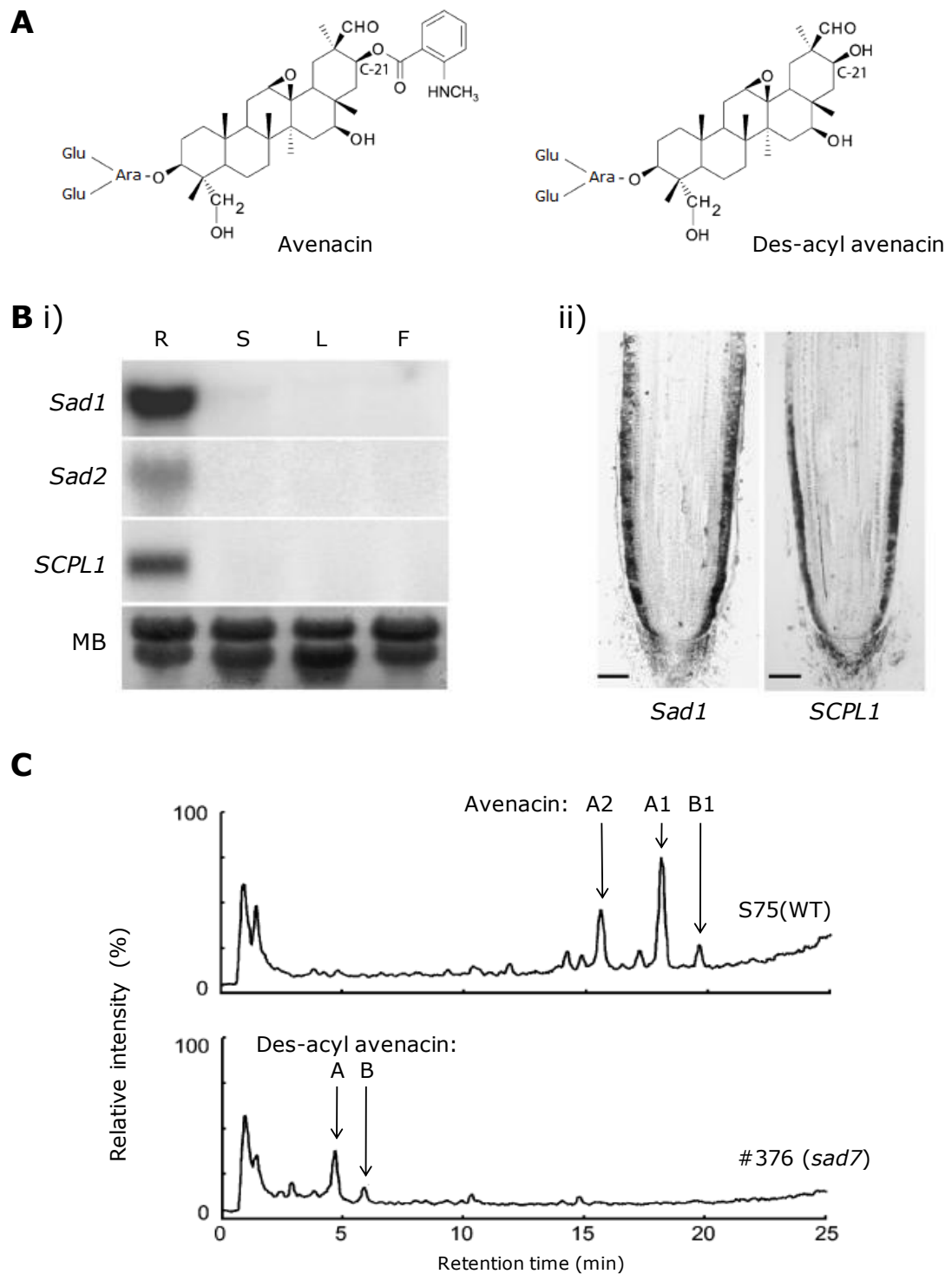


Figure 1.11: Characterisation of the *Sad7* locus

A. Structure of avenacin A-1 and des-acyl avenacin. All four avenacins are esterified with *N*-methyl anthranilate at the C-21 position of the aglycone. Des-acyl avenacins lack this esterified group and accumulate in *sad7* mutants. **B i).** RNA gel blot analysis of *Sad1*, *Sad2* and *SCPL1* transcripts in different oat tissues: root (R), shoots (S), leaves (L), flowers (F). RNA loading was monitored with methylene blue dye (MB). **ii).** *In situ* mRNA analysis of *Sad1* (left) and *SCPL1* (right) transcripts in the root tips of *A. strigosa*. In both i) and ii) *SCPL1* shows an identical expression pattern to *Sad1* and *Sad2* indicating that it is involved in avenacin synthesis. **C.** Identification and characterisation of *sad2* mutants by LC-MS analysis of root extracts from the wild-type S75 and mutant #376. The wild-type accumulates avenacins whereas mutant #376 lacks avenacins and accumulates des-acyl avenacins. All figures adapted from Mugford *et al.* (2009) (62).

were found to accumulate des-acyl avenacins (avenacin lacking the *N*-methyl anthranilate group at C-21). This indicated that they were deficient in acylation, and so represented independent mutant alleles at a locus termed *Sad7* and possibly lacked a functional acyltransferase (Figure 1.11A) (18, 63). Genetic analysis had shown the *Sad7* locus cosegregated with the *Sad1* and *Sad2* loci (63), therefore the BAC contig that spanned the *Sad1* and *Sad2* region was extended. A third gene was discovered 62 kb upstream of the *Sad1* gene and was found to encode a serine carboxypeptidase-like acyltransferase (AsSCPL1). Gel blot and in situ localisation analysis showed that the expression of AsSCPL1 was restricted to the epidermal cells of the root tip which is an identical pattern to the other two members of the avenacin biosynthetic pathway, AsBAS1 and AsCYP51H10 (Figure 1.11B). Sequencing of the AsSCPL1 gene in mutant nos. 376 and 616 revealed single point mutations that were predicted to result in amino acid substitutions. Another *sad7* mutant (#19) was identified from the extended mutant collection that also accumulated des-acyl avenacins and that had a single point mutation in the AsSCPL1 gene (Figure 1.11C). Together these data provide strong evidence that AsSCPL1 was encoded by the *Sad7* locus and that it was likely to be involved in avenacin acylation (62).

The *Sad7* gene encodes a 493 amino acid protein which has an N-terminal signal peptide, possibly for targeting to the vacuole where avenacins are known to accumulate (35). Removal of the signal peptide results in a 51.4 kDa protein which is formed of a large and small subunit, 29 kDa and 19 kDa respectively, joined by a linker region. To produce a functional protein product, the full-length protein is predicted to undergo cleavage in a two step process to remove the linker region and form a heterodimer from the large and small subunits. Analysis of AsSCPL1 protein levels in the three *sad7* mutants using an AsSCPL1 antibody show that no protein is present in mutant #616 and #19 whereas mutant #376 has reduced protein levels. The location of the mutation in mutant no. 376 was in a loop region in the active site so may affect protein activity, and the location of the mutations in mutant #616 and #19 are both in β -sheets which form part of the secondary structure of the protein so are likely to cause misfolded protein resulting in protein degradation (62).

Functional characterisation of the three *sad7* mutants found that all three accumulated *N*-methyl anthraniloyl-*O*-glucose in the roots compared to the wild-type which implied that AsSCPL1 uses an *O*-glucose ester as the acyl donor substrate. To test this hypothesis, AsSCPL1 was expressed in *Nicotiana benthamiana* leaves using a virus-based system and an AsSCPL protein preparation was incubated with methanolic extracts from the roots of *sad7*

mutants (which contain des-acyl avenacins and the glucose ester *N*-methyl anthraniloyl-*O*-glucose, the presumed substrates for AsSCPL1). The reaction products were analysed by LC-MS and resulted in production of all four avenacins (A-1, A-2, B-1 and B-2) which indicated that AsSCPL can catalyse the transfer of both *N*-methyl anthraniloyl and benzoyl groups onto des-acyl avenacins (62).

1.3.5.4 ***Sad3* and *Sad4* – glucosyltransferases?**

In addition to modification of the β -amyrin triterpene skeleton by oxidation (*Sad2*) and acylation (*Sad7*), it is known that β -amyrin is also acted on by glucosyltransferases which add the branched sugar chain onto C-3 of the aglycone and also play a part in activation of the acyl group which is added at C-21 by AsSCPL1 (18, 87). Glycosylation at C-3 of the aglycone occurs late in the pathway and is critical for the antifungal activity of avenacin, as removal of even a single sugar results in loss of potency (88). Mutants at two of the loci identified in the mutant screen, *Sad3* and *Sad4*, accumulate monodeglucosyl avenacin A-1 which lacks the β -1,4-linked D-glucose, indicating that both of these loci are involved in glucose addition possibly by encoding glucosyltransferases (GTs) or regulating their expression (Figure 1.12A) (83). GTs are a large enzyme family that catalyse the transfer of sugars to a wide variety of acceptor molecules. There are 77 GT families of which GT Family 1 is the largest and is mainly responsible for transferring sugar molecules onto low molecular weight molecules such as avenacin A-1 (89). Therefore it is likely that the GTs responsible for glycosylation of the aglycone are members of Family 1 GTs. Analysis of the expressed sequence tag (EST) library constructed from oat root tips revealed 27 sequences with similarity to Family 1 GTs. A number of these candidate cDNA sequences have been cloned for heterologous expression in bacteria and it is hoped that functional analysis of these enzymes to determine their donor and acceptor profiles will enable the discovery of GTs that are involved in avenacin synthesis (90).

Further characterisation of *sad3* and *sad4* mutants revealed that as well as being deficient in avenacin synthesis, they also have stunted roots and reduced numbers of root hairs. Further analysis using propidium iodide staining, calcofluor staining and electron microscopy revealed that membrane trafficking had been disrupted in root epidermal cells. Calcofluor staining showed that accumulation of callose and other primary cell wall components disrupted the epidermal cell layer resulting in root hair deficiency (Figure 1.12B). To ascertain

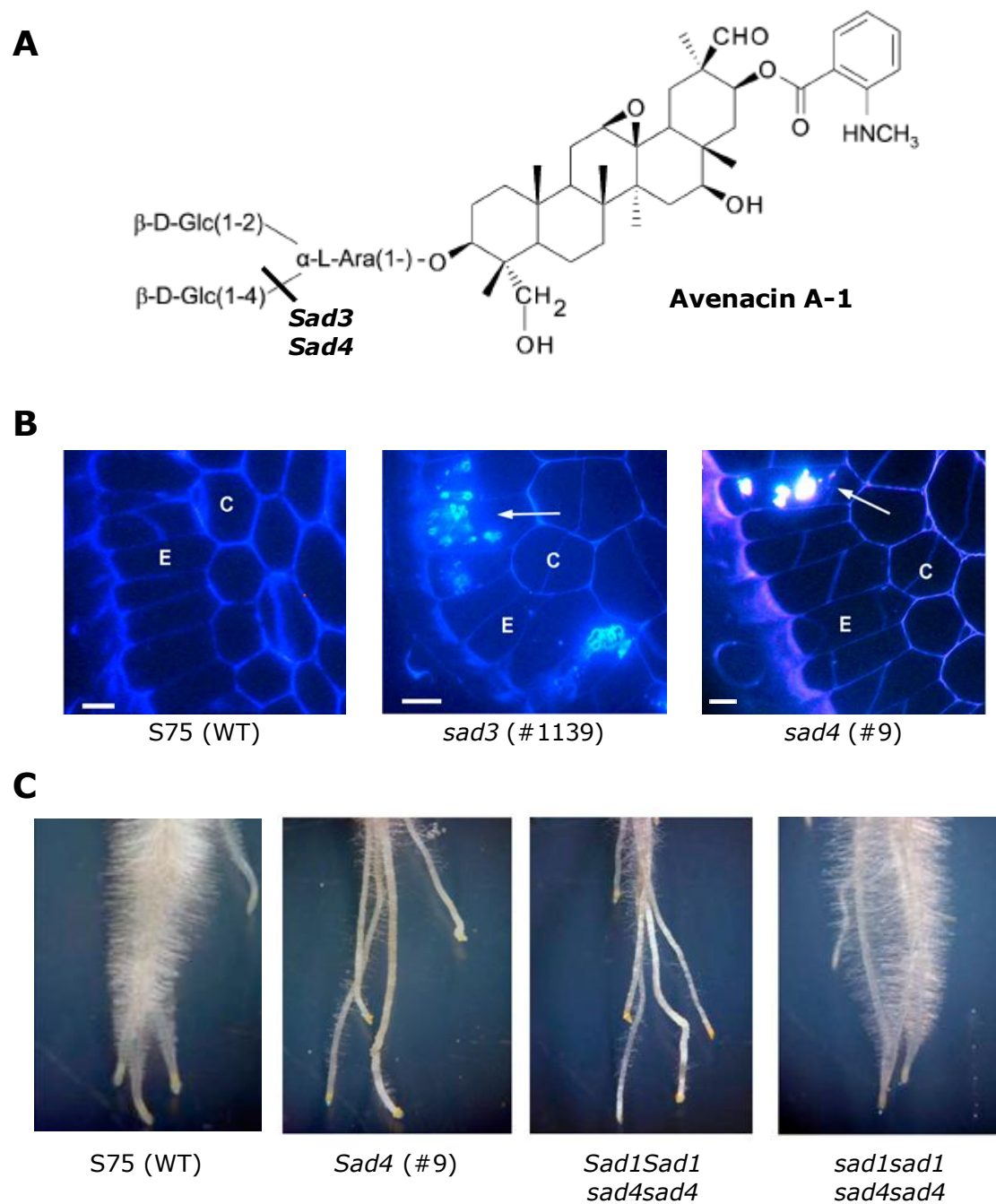


Figure 1.12: Characterisation of the *Sad3* and *Sad4* loci

A. Structure of avenacin A-1. *sad3* and *sad4* mutants accumulate monodeglucosyl avenacin A-1 which lacks the β -1,4-linked D-glucose at the position indicated by a black line. **B.** Calcofluor staining of oat meristematic root cross sections of wild type S75, *sad3* and *sad4* mutants. Calcofluor staining aggregates (arrows) are seen in both *sad3* and *sad4* mutants. C, cortex; E, epidermis. Bars = 50 μ m. **C.** Root phenotypes of 4-day old seedlings of *sad1/sad1 sad4/sad4* double mutants. *sad4* mutants in a wild type *Sad1* background have a mutant root morphology whereas in a *sad1* mutant background (where the avenacin pathway is blocked) normal root morphology is seen. Similar results were seen for *sad3* mutants. All figures adapted from Mylona *et al.* (2008) (35).

whether these root defects were a direct result of *sad3* and *sad4* mutations or as a result of additional mutations, both of the *sad3* and *sad4* mutations were introduced into a wild-type, heterozygous and homozygous *Sad1* backgrounds.

In the wild-type background, root defects are still observed, but in a homozygous *sad1* mutant (where the avenacin pathway is blocked) root defects are suppressed (Figure 1.12C). This implies that monodeglucosyl avenacin A-1, which accumulates in *sad3* and *sad4* mutants, causes the observed root defects. The results also indicated that *Sad3*, which is found in the avenacin gene cluster, has a specific role in avenacin synthesis whereas *Sad4*, which is unlinked, has a much broader function (35).

1.4 Oxidosqualene cyclases in plants, animals and fungi

Even though the avenacin pathway is restricted to oats, the initial step of 2,3-oxidosqualene cyclisation into lanosterol, cycloartenol, β -amyrin, lupeol and other sterol and triterpene precursors is observed in all plants. Studying other triterpenoid synthase pathways has given great insights into the enzymes involved in triterpenoid production. By studying homologous genes in other plant species, information relating to evolution of the pathways; enzyme structure and mechanism; product specificity and pathway regulation has enabled further understanding of triterpenoid synthesis and the origins of these fascinating and complex pathways.

1.4.1 Cloning of oxidosqualene cyclases involved in sterol production

The first sterol OSC to be cloned from any organism was lanosterol synthase from *Candida albicans* by complementation of a *Saccharomyces cerevisiae* OSC mutant (*erg7*) which lacked lanosterol synthase activity (91-92). As lanosterol synthase is crucial for membrane sterol synthesis (ergosterol in fungi), *erg7* mutants were unable to grow without exogenous ergosterol. Complementation of the *ERG7* gene by the lanosterol synthase sequence in *C. albicans* confirmed that *ERG7* encoded a lanosterol synthase (91). PCR cloning using degenerate primers complementary to homologous regions from other cloned OSCs and complementation of the *erg7* mutant has been used for cloning of seven lanosterol synthases from the fungi *Schizosaccharomyces pombe* (93), *Saccharomyces cerevisiae* (94), *Pneumocystis carinii* (95) and *Cephalosporium caerulens* (96); the parasitic protozoa *Trypanosoma cruzi* (95); animals *Homo sapiens* (97-98) and *Rattus norvegicus* (99); and the bacterium *Methylococcus capsulatus* (100). Whole genome sequencing of a number of animals and fungi has resulted in identification of a number of putative lanosterol synthase sequences which are all targets for future cloning and functional characterisation.

In plants, cycloartenol synthase converts 2,3-oxidosqualene to cycloartenol, the major precursor for sterol synthesis. Cloning of cycloartenol

synthases could not be done simply by complementation of the *erg7 S. cerevisiae* mutant as for the lanosterol synthases as yeast does not synthesise cycloartenol synthase. The *erg7* mutant was instead utilised for heterologous expression of putative cycloartenol synthase cDNA sequences and researchers took advantage of 2,3-oxidosqualene accumulation in the mutant which could be used by the expressed gene to generate a novel cyclisation product that was detected by TLC, HPLC or GC-MS analysis (101). Subsequently the first cycloartenol synthase gene was cloned from *Arabidopsis thaliana* and subsequently over 20 cycloartenol synthases have been isolated from a wide variety of plant species by sequence homology. To date, cycloartenol synthases have been cloned from dicots (102-111), monocots (49, 53, 112), a gymnosperm (*Abies magnifica*) and a fern (*Adiantum capillus-veneris*) (113). Functional cycloartenol synthases have also been cloned from a soil amoeba (*Dictyostelium discoideum*) (114) and a myxobacteria (*Stigmatella aurantiaca*) (115) which show approximately 50% homology to the plant cycloartenol synthases. This implies that cycloartenol synthases from bacteria and plants are orthologous and represent a highly conserved and important class of OSCs that predates the emergence of plants (116).

Experiments carried out by Gibbons *et al.* in 1971 provided strong evidence that the precursor of sterols in plants and other photosynthetic organisms was cycloartenol, whereas in animals and fungi lanosterol was used (117). However, studies in *Euphorbia* identified lanosterol in a non-polar fraction of the latex that could not be made via conversion of cycloartenol which suggested that plants may contain a lanosterol synthase (118-120). A candidate lanosterol synthase protein sequence was found in *Arabidopsis thaliana* with 64% identity to the cycloartenol synthase protein sequence. Complementation of a lanosterol synthase-deficient yeast and functional expression found the candidate sequence to be a functional lanosterol synthase and functional lanosterol synthases have since been cloned from *Lotus japonicus* (121) and *Panax ginseng* (120). Phylogenetic analysis has shown lanosterol synthases to be widespread amongst dicots and to have a different evolutionary origin from animal and fungal lanosterol synthases (47).

1.4.2 Cloning of oxidosqualene cyclases involved in triterpene production

For primary metabolism most organisms contain one sterol OSC, either a cycloartenol synthase or a lanosterol synthase, which synthesises a single sterol precursor, cycloartenol or lanosterol respectively. Production of triterpenes by secondary metabolism can be catalysed by numerous enzymes in a single plant

species. Compared to the sterol OSCs, collectively the triterpene OSCs produce many more cyclic products. Some OSCs are specific, producing a single product, but many are promiscuous and produce numerous different products which add to the diversity of secondary metabolites found in plants. Production of triterpenoids through secondary metabolism is a process that is mostly found in plants but is also seen in fungi and sea cucumbers (122), the only animals to produce triterpenoids. Over the last twelve years, a wide variety of triterpene OSCs have been cloned from a number of plant species and one basidiomycete.

The first plant triterpene OSC to be cloned was a β -amyrin synthase from the roots of *Panax ginseng* using homology-based PCR methods (111). A number of β -amyrin synthases have since been cloned from dicots and represent the largest group of plant triterpene OSCs (54, 56, 75, 77, 104, 109, 111, 123-128). β -amyrin synthase from the monocot *Avena strigosa* (AsbAS1) is the only triterpene OSC to be cloned from monocots to date and represents a novel class of β -amyrin synthases (49). Triterpene OSCs show far less sequence conservation than their sterol synthase counterparts (approx. 50% compared to 75%) and the large phylogenetic separation of dicot and monocot sequences suggests that the two groups evolved β -amyrin synthases separately (63).

The second largest group of triterpene OSCs are the lupeol synthases which cyclise 2,3-oxidosqualene to the pentacyclic product lupeol. The first lupeol synthases to produce solely a lupeol product were cloned from *Olea europaea* and *Taraxacum officinale* in 1999 (105). Other lupeol synthases have since been cloned from *Betula platyphylla* (104), *Glycyrrhiza glabra* (129), *Lotus japonicus* (109) and *Bruguiera gymnorrhiza* (56), with the latter forming a new branch of lupeol synthases. These enzymes are phylogenetically distinct from other OSCs, but show 74-81% sequence homology to each other (116). There are also two multifunctional triterpene OSCs which synthesise lupeol as their major product and these have been cloned from *Arabidopsis thaliana* (52, 130) and *Ricinus communis* (107).

The majority of single product enzymes that have been cloned are either β -amyrin synthases or lupeol synthase but over 90 different triterpene skeletons exist in nature so this implies that there are also more triterpene OSCs that have not yet been discovered. Currently through homology based PCR methods, four novel triterpene OSCs have been cloned that synthesise pentacyclic or tetracyclic triterpenes; isomultiflorenol synthase (*Luffa cylindrica*), cucurbitadienol synthase (*Cucurbita pepo*) and dammarenediol-II synthase (*Panax ginseng* and *Centella asiatica*). The remainder of the triterpene OSCs that have been cloned to date fall into the category of multifunctional triterpene OSCs, in that they synthesise

more than one triterpene product (53-56, 131-134). The products that they produce include β -amyrin, α -amyrin, lupeol, taraxasterol, germanicol, butyrospermol and many others.

The only non-plant triterpene OSC to be cloned to date is clavarinone cyclase from the basidiomycete *Hypholoma sublateritium* which catalyses the conversion of 2,3-oxidosqualene to clavaric acid (135).

1.4.3 Alternative strategies for cloning plant oxidosqualene cyclase genes

Most OSCs to date have either been cloned via functional complementation or sequence homology and their products identified by TLC, HPLC, GC-MS or natural product isolation. Systematic analysis of fully sequenced plant genomes is now being used to identify new enzymes that are involved in natural product synthesis. The technique requires a whole genome sequence of a particular organism from which protein-encoding regions are identified and functions are assigned to each of the regions based on their homology to other genes of known function (136-138). As more plant genomes become fully sequenced, this will enable the whole OSC gene family to be uncovered. It also allows more sensitive probing of plant genomes in order to identify new triterpenoids and by coupling this process to heterologous expression, may uncover new biosynthetic pathways and low-abundance natural products (57, 116).

Using this technique in *Arabidopsis thaliana* facilitated the cloning of a number of enzymes which generate new triterpenoid products including those with partially cyclised triterpene skeletons. These enzymes include, a multifunctional triterpene OSC (139), thalianol synthase (57), marnerial synthase (140), arabidiol synthase (60), camelliol synthase (141), baurol synthase (58) and β -amyrin synthase (128). Many of these products were thought not to occur in nature, but recently, low levels of thalianol have been detected in *Arabidopsis* roots which matched the expression pattern of thalianol synthase (AtTHAS). Analysis of the genomic region surrounding the AtTHAS gene uncovered two cytochrome P450s and an acyltransferase. Both of these enzyme classes have been implicated in secondary metabolism and mutational analysis revealed that both P450s modify the thalianol gene product and provided strong evidence for a novel biosynthetic pathway in *Arabidopsis* (65).

1.5 Structure-function relationships in plant OSCs

1.5.1 Conserved features: vast product differences

One of the great mysteries surrounding OSCs is how one substrate can be converted to so many different products using only a handful of different enzymes. One might think that these enzymes would be significantly different from each other, but sequence analysis has shown that they not only have highly homologous sequences but have also evolved from a common ancestor. They also share common conserved features that identify them as OSCs.

OSC activity is generally associated with microsomes, and so OSCs were presumed to be membrane bound, but sequence analysis found no signal sequences or transmembrane domains. Theories suggested the presence of a hydrophobic membrane-localising region which provided a different mechanism for OSCs to attach to the membrane (142). This region has been observed on both the *Alicyclobacillus acidocaldarius* squalene cyclase (AaSHC) (143) and the *Homo sapiens* lanosterol synthase (hOSC) (144) crystal structures, thereby confirming that AaSHC and hOSC are monotopic membrane proteins. Both structures have a hydrophobic surface of approximately 1.7\AA^2 which is encircled by a ring of polar residues which form salt bridges with the displaced head groups of the phospho- and sulpholipids and form the most likely passage for substrate access to the enzyme active site (143). Mutagenesis and inhibitor binding studies in a wide variety of species have identified the catalytic motif DCTAE that is found in all OSCs (145-147). Experiments with *Rattus norvegicus* lanosterol synthase and SHC showed that the aspartate in this conserved motif to be the catalytic residue required for initiation of the cyclisation reaction (146-147).

Another feature that is characteristic of all OSCs and squalene cyclases (SCs) is the QW motif. It was first identified by Karl Poralla through comparisons of the AaSHC and *Candida albicans* lanosterol synthase amino acid sequences. Upon aligning the sequences he found four to five repetitions of approximately 16 amino acids which formed the consensus motif: (R/K)-(G/A)-X₂₋₃-(F/Y/W)-(L/I/V)-X₃-**Q**-X₂₋₅-G-X-**W**. The motif, which is found in β -turns, was named QW due to the presence of glutamine and tryptophan residues at the end of the motif (148). The AaSHC crystal structure showed that these two residues stabilise the protein structure by stacking together and forming hydrogen bonds with the amino end of the adjacent outer barrel helix and the carbonyl end of the preceding outer barrel helix. This connects the outer helices of the alpha helical barrel which stabilises the protein structure (143). AaSHC has eight QW motifs whereas the OSC

enzymes have fewer motifs with only five but are presumed to have the same function as seen in the SCs (144).

In addition to the DCTAE catalytic motif and the QW motifs, SCs and both sterol and triterpene OSCs have a number of residues that are both common and exclusive to each class of cyclase. These residues play a large role in the enzymatic cyclisation mechanism and are responsible for product determination in each of the enzymes.

1.5.2 Solving the structures of triterpenoid cyclases

The question of why sterol and triterpene synthases, which have considerable sequence homology and conserved features, produce such diverse products has been at the forefront of secondary metabolite research for many years. Large advances have already been achieved through the cloning of enzymes involved in sterol and triterpene synthesis in a wide variety of plant species. Mutagenesis experiments too, have allowed the elucidation of many of the important residues required for cyclisation. Whilst this data is invaluable to deciphering the inner workings of one of the most complex reactions in nature, they cannot provide the whole story. The ability of these enzymes to synthesise different products is most likely to be due to subtle differences in amino acid residues and active site orientations. This information can easily be obtained from crystal structures and the ability to crystallise enzymes with substrates, intermediate analogues and products bound would give vital clues into the enzymatic mechanism.

Membrane proteins represent about one third of the genome but have traditionally been difficult to crystallise, with less than 1% having known crystal structures. Expressing eukaryotic membrane proteins in prokaryotes is problematic due to the different membrane constituencies of the organisms and the inability of bacterial hosts to post-translationally modify eukaryotic proteins. Protein purification is also difficult due to the hydrophobic nature of the membrane spanning regions of these proteins, and usually requires a delicate balance of detergents to solubilise the proteins. The final challenge lies in forming viable crystals, as detergents have complex phase behaviour that hinders the formation of well-ordered crystals (149). As a result, only two structures exist for triterpene cyclases, the bacterial SC (AaSHC) and the human OSC (hOSC). No structures currently exist for any of the plant OSCs.

The *Alicyclobacillus acidocaldarius* squalene cyclase (AaSHC) was the first of its kind to be cloned from bacteria. It was expressed in *Escherichia coli*,

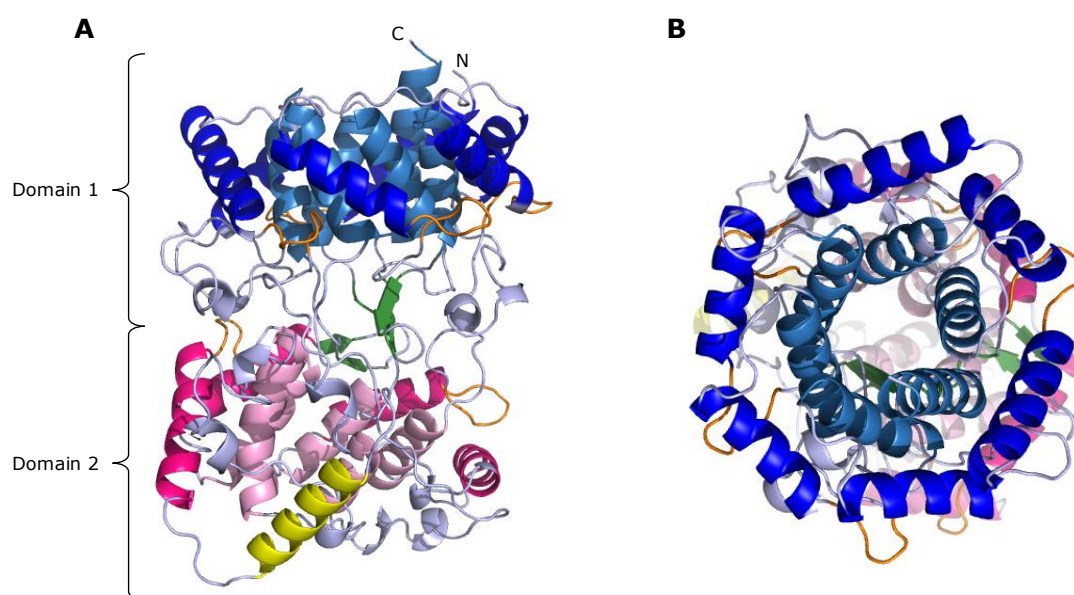


Figure 1.13: Crystal structure of *Alicyclobacillus acidocaldarius* squalene cyclase (AaSHC)

A. Stereoview of AaSHC with labelled amino (N) and carboxy (C) termini. Structure consists of two α -helical domains: Domain 1, external helices – royal blue; internal helices – light blue. Domain 2, external helices – dark pink; internal helices – light pink; β -sheet structure – green; QW motif – orange; non-polar plateau – yellow. **B.** Top view of domain 1 - α -helical barrel forms two concentric rings that are stabilised by QW motifs (orange).

purified by anion-exchange chromatography and gel filtration, and produced protein crystals that diffracted at 2.9\AA resolution (150). Analysis of the crystal structure revealed two domains, domain 1 - an α_6 - α_6 barrel consisting of two concentric rings of helices, and domain 2 - an α - α barrel, with the loops of both domains forming a β -structure which encloses a 1200\AA^3 cavity. The QW motifs are found in almost identical conformations, and the stacked glutamine and tryptophan side chains connecting several of the outer helices via hydrogen bonds which stabilise the structure (Figure 1.13). There is a non-polar gated channel leading from the membrane to the central cavity and this provides the path for substrate access to the active site (143).

The *Homo sapiens* lanosterol synthase (hOSC) was cloned in 1995 (97), and in 2004 was expressed in *Pichia pastoris*. It was purified using metal-chelate chromatography and gel filtration and crystals produced by the vapour diffraction method diffracted to 2.0\AA resolution (151). Analysis of the crystal structure revealed a similar structure to the bacterial enzyme: two α - α barrel domains connected by loops which formed three β -structures. The QW motifs were found in the same conformation as AaSHC and an additional domain was found at the amino-terminus in between the two domains and may play a role in stabilising their orientations (Figure 1.14). The active site cavity is located between the two

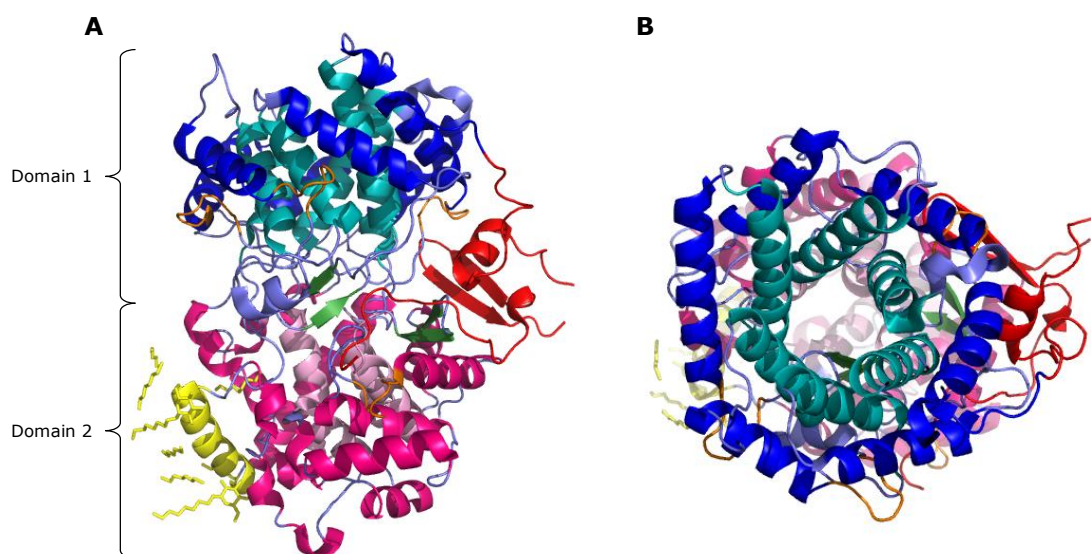


Figure 1.14: Crystal structure of *Homo sapiens* lanosterol synthase (hOSC)

A. Stereoview of hOSC. Structure consists of two α -helical domains: Domain 1, external helices – royal blue; internal helices – light blue. Domain 2, external helices – dark pink; internal helices – light pink; amino-terminal domain – red; β -sheet structure – green; QW motif – orange; non-polar plateau – yellow; detergent molecules – yellow sticks. **B.** Top view of domain 1 - α -helical barrel of domain 1 forms two concentric rings that are stabilised by QW motifs (orange).

barrel domains and like AaSHC, a non-polar gated channel leads from the active site to the membrane. Presence of bound detergent molecules indicate where the membrane-binding helix of the protein is and show that it is amphipathic, so inserts into the membrane but does not span it (144).

1.5.3 Insights into enzyme mechanism

The extraordinary way in which essentially linear squalene and 2,3-oxidosqualene are transformed into complex cyclic triterpenoids has been of huge interest to the field of science. In recent years there has been much speculation about the exact nature of these reactions which have been investigated through studies with artificial substrates, inhibitors and site-directed mutagenesis. Elucidation of the structures of two cyclases has confirmed many of these theories and shed further light on both the conserved residues and cyclase specific residues found in these enzymes. Due to the inability to isolate reaction intermediates triterpenoid synthesis was originally thought to be a concerted process in which cyclisation occurred in a single step (152). However, work in the early eighties showed that cyclisation occurred through a number of rigid carbocation intermediates and was supported by the isolation of partially cyclised triterpenoid intermediates from the Mediterranean shrub *Pistacia lentiscus* (153-154).

Structure determination of AaSHC combined with data from mutagenesis and organic experiments allowed elucidation of the first mechanism for enzymatic triterpene cyclisation (143, 152, 155-164). In the absence of an OSC crystal structure, the similarity of the bacterial and human cyclase, although having only 26% identity, allowed a model of hOSC to be determined. This gave further insight into the cyclisation differences between eukaryotes and prokaryotes at the atomic level (165). Subsequently the structure of hOSC was solved and supported the conclusions drawn by the model (144).

The mechanisms for triterpenoid cyclisations between the two enzymes are very similar, with very few differences that promote the formation of the different products. They consist of the same basic steps: binding of the substrate in a folded conformation, reaction initiation by protonation, ring formation, skeletal rearrangement and reaction termination by deprotonation (Figure 1.15). Before the reaction begins the substrates are pre-folded to promote cyclisation. Each ring is pre-folded into a chair conformation with the exception of the B-ring of 2,3-oxidosqualene which is folded into the energetically unfavourable boat conformation. The B-ring is forced into this conformation by Lys331 and Tyr98 which push the C₁₀ methyl group below the molecular plane. If cyclisation were to proceed via the favourable chair-like conformation, steric hindrance with the Tyr98 would cause deprotonation and premature termination of cyclisation, hence why the unfavourable boat conformation is formed (144).

Initiation of AaSHC and hOSC catalysis occurs via a conserved aspartate residue, Asp376 and Asp455 respectively. In AaSHC, by bonding to His541, Asp376 has increased acidity and therefore increased protonating ability. hOSC lacks an equivalent activating residue, but as epoxides such as 2,3-oxidosqualene are more readily protonated than alkenes such as squalene, this is expected. Instead in hOSC Asp455 is hydrogen bonded to two cysteines (Cys456 and Cys533) in order to become activated. The activated aspartate protonates the first carbon-carbon double bond of squalene in AaSHC, and the epoxide group of 2,3-oxidosqualene in hOSC to commence the cyclisation reaction (143-144, 156).

After initiation, ring formation occurs first with formation of the A-, B- and C-rings which occurs by a similar process in both enzymes. Conserved aromatic residues promote stabilisation of the C₆, C₁₀ and C₁₅ tertiary cations respectively via cation- π interactions. Formation of the D-ring signals the branchpoint between AaSHC and hOSC mechanisms. In AaSHC, the D-ring is formed first as a 5-membered carbon ring which upon stabilisation by surrounding aromatic residues is expanded to form a six-membered carbon ring. In hOSC, a five-

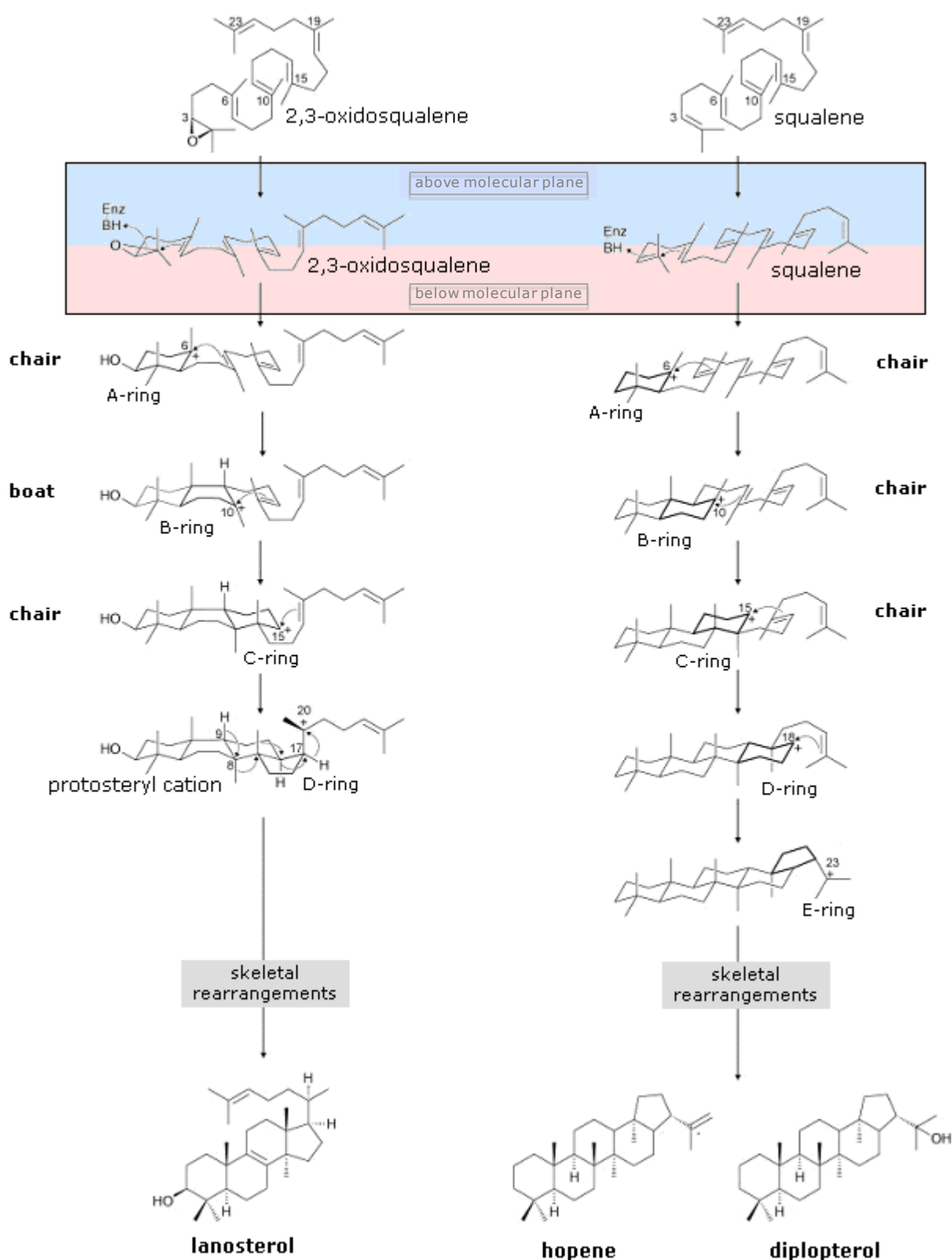


Figure 1.15: Cyclisation mechanism for hOSC and AaSHC

Scheme to show each of the cyclisation steps that convert squalene (AaSHC) and 2,3-oxidosqualene (hOSC) into the cyclic products hopene and lanosterol respectively. The reaction is initiated by protonation of the substrate. Cyclisation then takes place via the chair-chair-chair (AaSHC) or chair-boat-chair (hOSC) conformations. After ring formation is complete, skeletal rearrangements transform the cyclisation product to the final product cation which is then deprotonated to form either hopene (AaSHC) or lanosterol (hOSC). Figure adapted from Schulz-Gasch and Stahl (2003) (165).

membered carbon D-ring is also formed but it is unable to be stabilised so the cyclisation cascade ends and the tertiary protosteryl cation is formed at C₂₀. Cyclisation in AaSHC continues to generate a 5-membered carbon E-ring after which its cyclisation cascade ends. This difference in ring number is due to a variation in amino acid identity between the two enzymes. SHC contains two phenylalanine residues, Phe601 and Phe605 which stabilise the D-ring thereby facilitating the expansion of the D-ring and formation of the E-ring. hOSC only contains one of these phenylalanine residues (corresponding to Phe601 in AaSHC) and therefore cannot stabilise the cation formed upon D-ring formation so the cyclisation cascade is terminated (144, 164).

Skeletal rearrangements then occur to convert the respective cations into their final products. These rearrangements happen by 1,2-shifts of hydride and methyl groups which form the hopenyl cation in AaSHC and the lanosterol cation in hOSC. Finally the reactions are terminated by deprotonation of the final cations to yield hopene or lanosterol respectively. In AaSHC, this postulated to occur via a polarised water molecule that is found in a network of water molecules surrounding Glu45 (164). In hOSC, the proton is abstracted from the lanosterol cation by Tyr503 which is hydrogen bonded to His232 (144).

It is clear that solving the structures of both SCs and OSCs has given huge insights into the mechanisms by which these remarkable enzymes turn a single substrate into a vast variety of products. It is clear that to further understand the nature of these fascinating reactions, more crystal structures need to be solved which will provide vital clues as to how the product specificity of each enzyme is achieved.

Chapter 2 - Homology modelling of AsbAS1

2.1 Introduction

When studying a particular protein, the best tool that can be used to gain insights into its structure is X-ray crystallography and generation of a crystal structure. However, there are many obstacles that must be overcome on the way to generating a high resolution structure. First a pure protein sample in milligram quantities is needed, followed by the right solution conditions in order to generate crystals. Finally the crystals need to be of an appropriate quality to diffract and generate a high resolution image and hours of refinement are needed to generate the final structure. None of these steps are trivial and sometimes can take years and be almost impossible to achieve, therefore researchers have had to employ other quicker methods, for example homology modelling, to study the structures of their proteins.

2.1.1 Homology modelling of oxidosqualene cyclases

The *Homo sapiens* lanosterol synthase (hOSC) is one example where homology modelling was used in the absence of a crystal structure. OSCs were rapidly becoming targets for antifungal, hypocholesterolemic and phytotoxic drugs due to their involvement at the early stages of both sterol and triterpene metabolism. However, due to problems expressing and purifying the enzymes because of their membrane-bound nature, no X-ray crystal structures had been solved. The only related structure that was available was the *Alicyclobacillus acidocaldarius* squalene cyclase (AaSHC) which is involved in bacterial hopanoid synthesis, a homolog of eukaryotic sterols (143, 150). Both OSCs and SCs have closely related sequences and mechanisms so were predicted to have a similar overall fold. Although the hOSC and AaSHC sequences have only 26% identity over the whole sequence; in the active site cavity, identity rises to 50% indicating that an accurate model could be generated. The hOSC model was generated using the AaSHC crystal structure as a template and the model produced was verified by testing a number of geometric parameters. The model was then used to investigate OSC mechanism at a molecular level and the results supported data that had been gathered from mutagenesis and inhibitor studies (165). Knowledge about active site size and shape and substrate binding properties facilitated the design of novel drugs that had a more potent activity (166). Subsequently, when hOSC was crystallised in 2004 (144), the crystal structure

confirmed the mechanism hypotheses made from the homology model. This study shows that homology modelling can provide a detailed and accurate picture of enzyme structure and function in the absence of a crystal structure.

2.1.2 Site directed mutagenesis of oxidosqualene cyclases

In addition to structural investigations through homology modelling, site directed mutagenesis has also been used to identify specific amino acid residues involved in the enzyme mechanism of OSCs. The majority of these experiments have been done using *Saccharomyces cerevisiae* lanosterol synthase (ScOSC), one of the first sterol OSCs to be cloned (94). Site-directed mutagenesis experiments identified His234 (167), Phe445 (168), Trp232 (169), Tyr707 (170), Phe699 (171), Tyr510 (172) and Trp443 (173) as catalytically important residues as mutagenesis led to the accumulation of truncated cyclisation products. As well as identifying catalytically important residues, mutagenesis experiments have also identified residues that define one type of OSC from another. Studies in *Arabidopsis thaliana* cycloartenol synthase identified a single mutation, Tyr410Thr, that was sufficient to alter the product identity of the enzyme from cycloartenol to a mixture of lanosterol, parkeol and 9 β -lanosta-7,24-dien-3 β -ol (174). Subsequent experiments identified five specific mutations which enabled a cycloartenol synthase to produce lanosterol as its sole product, Tyr410Cys, Ala469Val, His477Tyr, Ile481Thr and Tyr532His (175). The only mutagenesis studies to be done in triterpene OSCs were carried out on *Panax ginseng* β -amyrin synthase (PNY) and *Olea europaea* lupeol synthase (OEW). The studies identified two residues, Tyr261 (PNY) and Trp259 (PNY)/Leu256 (OEW), which are responsible for producing pentacyclic products and altering the product identity of the enzyme from β -amyrin to lupeol respectively (176). Both of these residues play vital roles in pentacyclic triterpene formation, but in the absence of a crystal structure or homology model for a triterpene synthase the exact role that they play in the OSC mechanism is unclear.

2.1.3 Molecular dynamics simulations

Whether you determine enzyme structure from an X-ray crystal structure of a homology model the final structure will be a snapshot view of the enzyme in a single conformation. If the structure has been crystallised with the substrate, intermediate or product bound then this can provide additional insights into either the binding, active site residues or one step in the enzyme mechanism. However to see how enzyme structure changes throughout the mechanism from the native

protein to substrate binding, catalysis and finally to product release, another level of analysis is required. Molecular dynamics (MD) simulates the movements of atoms and molecules over a fixed period of time using the known laws of physics and energetics. It can be used to refine structures, simulate the positions of atoms and molecules in the structure and simulate the movement of the enzyme over the course of the whole reaction. MD simulations have been used to develop atom interpretations of experimental results and to test hypotheses for enzyme mechanism (177). MD has been used in the characterisation of a number of enzymes, for example, aquaporin 1 and the glycerol uptake facilitator which are involved in water and glycerol transport respectively (178). MD has also been used by the pharmaceutical industry to test drug targets such as protein kinases (179) and to study early events in rhodopsin activation (180).

2.1.4 Aims

Currently no crystal structures or homology models are available for any triterpene OSC, therefore in order to begin characterisation of *Avena strigosa* β -amyrin synthase (AsbAS1) a homology model needs to be generated. Reaction intermediates of 2,3-oxidosqualene cyclisation by AsbAS1 will be modelled into the active site using MD simulations which will allow detailed enzymatic mechanism of AsbAS1 to be explored. The structure of AsbAS1 and its mechanism will then be compared to hOSC as well as amino acid sequence comparisons with plant OSCs sequences from a variety of species. This information will be used to investigate the mechanistic differences between sterol and triterpene OSCs at a molecular level and lead to further understanding of how vastly different products are produced by highly homologous enzymes which all use the same substrate. Key amino acid residues will also be identified, which can be used as targets for future site-directed mutagenesis experiments.

2.2 Materials and methods

2.2.1 Modelling the structure of AsbAS1

2.2.1.1 Identification of homologous structures

An NCBI-BLAST2 search (181) was carried out using as query the AsbAS1 sequence against the set of proteins that had entries in the RCSB Protein Data Bank (PDB) (182). Homologs were selected as candidates for homology modelling according to the quality of the structural data. The validity of each candidate structure was analysed using the information available in the PDB.

Parameters that were taken into account were: sequence identity, resolution of the crystal structure, R-values, bond lengths and angles, number of outliers in the Ramachandran plot and the level of homology with the AsbAS1 sequence. Further analysis of the real space R-values were carried out for each residue of the candidate structures that had an anomalous R-value using the Uppsala Electron Density Server (EDS) (183). The steric geometry of the candidate structures was analysed using MolProbity (184). Finally, the oligomeric state of the structures was analysed using the Protein Quaternary Structure file server (PQS) (185). Only monomeric structures were carried forward to the next stage of the modelling process.

Based on this data analysis, the structures having the best homology, good steric and geometric values, the fewest number of outliers in the Ramachandran plot, good quality electron density data (i.e. lowest number of residues with real space R-values above 0.3) and were monomeric were carried forward to the next stage of the modelling process.

2.2.1.2 Sequence alignment and model generation

A structure-based sequence alignment of the AsbAS1 sequence with the selected template sequence was generated using FUGUE (186) and the resulting alignment was viewed using the CORE evaluation on the Toffee server (187). Proposed insertion and deletions to the template sequence were visualised in PyMOL (188) to ascertain whether they could be accommodated into the candidate structures.

The structure-based sequence alignment was converted into a PIR-format alignment file and any minor adjustments to the alignment were made manually. The AsbAS1 model was then built using the automated comparative protein modelling program MODELLER v9.4 (189).

2.2.1.3 Validation of ASbAS1 model

Before the structural features of the model could be analysed, the validity of the structure was checked to determine if a reliable model had been produced. The overall energy of the structure was calculated using the Prosa II web server (190) and atomic volumes and dihedral angles for each residue in the structure were calculated using PROVE and PROCHECK respectively (191-192). The coordinates of the AsbAS1 model and the template structure were submitted to each of the structure validation servers and the results were compared to determine how closely the AsbAS1 structure matched the template.

2.2.2 Sequence and phylogenetic analysis

2.2.2.1 **Multiple sequence alignment**

An NCBI-BLAST2 search (181) was carried out using as query the AsbAS1 protein sequence against a database containing non-redundant protein sequences of all known plant proteins. The output was analysed and only full-length OSC sequences were selected for use in the final sequence alignment. A multiple sequence alignment was generated using MUSCLE (**M**Ultiple **S**equence **C**omparison by **L**og-**E**xpectation) (193) and sequence features were viewed and annotated using Jalview (194).

2.2.2.2 **Phylogenetic analysis**

The MUSCLE sequence alignment of the plant OSCs was used for phylogenetic analysis. MEGA 4.0 software (195) was used to construct a neighbour-joining phylogeny for assessment of sequence diversity.

2.2.3 Molecular dynamics simulations

2.2.3.1 **Generating structures of AsbAS1 reaction intermediates**

Starting molecular structures for reaction intermediates were generated using high resolution crystal structures from the Cambridge Structural Database (CSD) (196). Structures for the following intermediates were generated: dammarenyl cation based on entry DELWAT (197); baccharenyl cation based on entry XENQOX (198); lupenyl cation based on entry ACHLUP10 (199); oleanyl cation based on XICPAB (200); and the β -amyryn cation and β -amyryn based on DABWUZ (201). Each intermediate was docked manually into the active site of the AsbAS1 model (which was superimposed onto hOSC) by alignment to the lanosterol molecule found in the crystal structure of hOSC. Force-field parameters for each of the reaction intermediates were generated using the PRODRG server (202).

2.2.3.2 **Molecular dynamics simulations**

The position of each docked intermediate was refined with molecular dynamics (MD) simulations using GROMACS 3.1 (203) using the force-field parameters determined for each intermediate. The procedure followed for the complex of the enzyme with each intermediate was essentially identical. The AsbAS1 model in complex with an intermediate was solvated in a box of water molecules, counter ions were added and then underwent 2000 cycles of conjugate

gradient energy minimisation to relieve strain. Position-restrained MD (restraining force constant $100 \text{ kcal mol}^{-1}\text{\AA}^{-2}$) was conducted over 20ps of simulation using a time-step of 1.5 fsec, followed by position-unrestrained MD conducted over 100ps of simulation with the same time-step. The final model structures were generated by energy minimising (2000 cycles, unrestrained) the structure coordinates obtained by averaging over the final 20ps of the position unrestrained simulation. The structures were the viewed in PyMOL for analysis.

2.3 Results and Discussion

2.3.1 Modelling the structure of AsbAS1

2.3.1.1 Sequence analysis

Currently, no crystal structure is available for AsbAS1, so the model was created using crystal structure data from a homologous protein. To serve as a template structure in homology modelling, any candidate protein required significant sequence homology to the target sequence and a crystal structure in the PDB. The AsbAS1 cDNA sequence was translated into a protein sequence and the reading frame which contained the longest coding sequence was selected. A preliminary BLAST analysis of this sequence showed that the AsbAS1 sequence showed 98% homology with β -amyrin synthases from other oat species and between 71-80% homology with cycloartenol synthases from other plant species (Appendix 1). However, none of these protein sequences had crystal structures, so the BLAST search was repeated and restricted to sequences which had crystal structures in the PDB (Appendix 2). The AsbAS1 sequence showed good homology with two protein structures in the PDB, the human lanosterol synthase (PDB ID: 1W6K) (144) and the bacterial squalene-hopene cyclase (PDB ID: 2SQC) (143).

2.3.1.2 Structural analysis and validation of template structures

The two candidate structures were analysed to determine their suitability to act as a template for the AsbAS1 model. A number of parameters, described in Table 2.1 were tested using data obtained from crystal structure data (182) and MolProbity analysis. Sequence identity denotes the percentage of amino acid residues that are identical between the template and target sequence. 1W6K showed much higher sequence identity to AsbAS1 than 2SQC, and also had a higher sequence homology – 56% compared with 40% for 2SQC. The 2SQC structure was solved at a slightly higher resolution than 1W6K, and also displayed

Table 2.1: Structural analysis of AsbAS1 sequence matches

Table shows a summary of the structural parameters analysed for each of the crystal structures with sequence matches to AsbAS1. Bond lengths and angle values denote the average standard deviations of all bonds in the structure from their mean value.

PDB ID	Identity (%)	Resolution of structure (Å)	R-Value (free)	Bond lengths (av. st. dev)	Bond angles (av. st. dev)	Ramachandran plot	
						Number of outliers	Residues in favoured regions (%)
1W6K	40	2.10	0.188	1.0	<0.5	0	98
2SQC	23	2.00	0.187	<0.5	<0.5	4	98

bond length and angle values which had slightly fewer standard deviations away from the mean values. However, Ramachandran plot analysis favoured 1W6K as all residues were found in allowed regions with 98% in favoured regions. Although 2SQC also had 98% of its residues in favoured regions, 4 residues were found in disallowed regions of the plot indicating steric hindrance between side-chain and main-chain atoms.

The R-value is a numerical value (expressed as a percentage) which denotes how well the refined structure fits the observed crystallographic data in terms of the observed structure factor amplitudes. A typical value for fully refined structures is ten times its resolution (i.e. 20% for a 2.0Å resolution structure). Many refinement programs now calculate the free R-value which is used as a form of validation and prevents over-refinement of the structure. The free R-value is calculated in the same way as the R-value but uses 5-10% of the structure that is set aside and not used in the refinement of the model. This value is usually 5-10% higher than the R-value. Both candidate structures show free R-factors for the overall structure well within the expected values for structures at 2.0Å resolution. Further analysis of the R-values focused on individual residues with anomalous real space R-values (those above 0.3 – three standard deviations from the mean). The real space R-value is a measure of the similarity between the electron density map calculated from the experimental X-ray diffraction data and the refined model. It allows evaluation of the goodness of fit of individual residues in the refined model and can identify specific areas of the protein that need further refinement. For residues with anomalous real space R-values, the electron density data was consulted to identify a possible reason for these anomalous values. 2SQC contained three deviant values which were all located on the protein surface in loops or turns. These regions are often flexible so can result in poor electron density. For example, the electron density map for 2SQC: tryptophan 216 (Figure 2.1) shows that the protein structure is not consistent with the electron density map and therefore results in the high real

space R-value that was observed for this residue. Unfortunately, no electron density data was available for 1W6K, but due to the similarity of its R-value with 2SQC it was assumed that any anomalous residues would be found in similar surface loop regions.

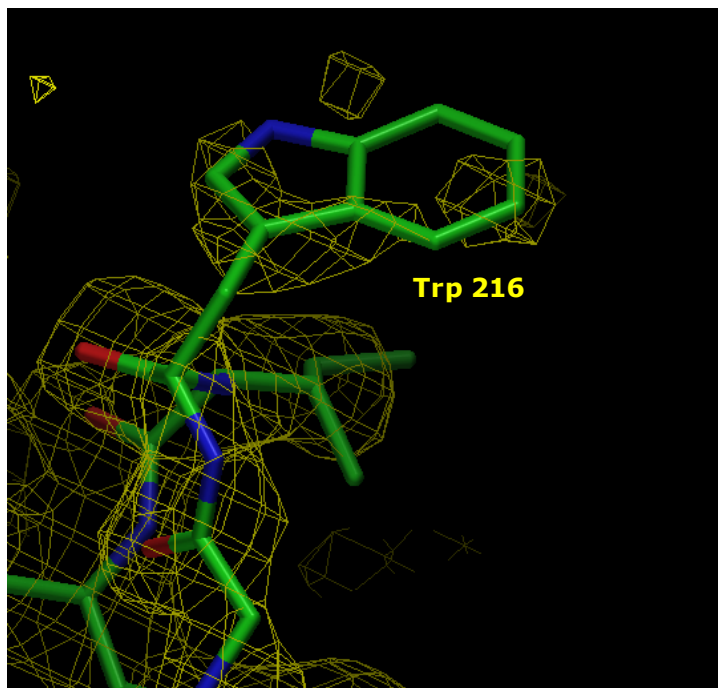


Figure 2.1: Electron density of 2SQC – Trp 216

The yellow net shows the electron density map obtained from X-ray crystallography (contoured at 1σ) and the local protein structure is shown in sticks format. The regions not surrounded by a yellow electron density map indicate a poor fit between the protein structure and the electron density data.

Finally the likely oligomeric state of the candidate structures was analysed using the PQS server. Data available from the PDB indicated that 1W6K was present as a monomer in the asymmetric unit whereas 2SQC was present as a dimer. The PQS server uses CCP4 (204) and WHATIF (205) programs to determine the association of the proteins in solution by calculating inter-chain atomic contacts. The results predicted 1W6K to be a monomer and 2SQC to be a dimer with non-specific intersubunit contacts which supported the crystallographic data.

This validation analysis showed the two candidate structures to be sufficiently reliable to serve as templates for homology modelling. However 1W6K displayed significantly higher overall sequence identity and was phylogenetically closer to AsbAS1, therefore it was chosen as the most appropriate crystal structure to model AsbAS1.

2.3.1.3 Structure based sequence alignment

The first step in the modelling process was to align the AsbAS1 sequence to the chosen template sequence using the structure based sequence alignment program, FUGUE (186). FUGUE aligns sequences based on the environment of each individual amino acid and structure-dependent gap penalty selection algorithms. Each residue requires a particular environment which restricts amino acid substitutions, so by using environment-specific substitution tables, alignments are more precise. Gap penalty algorithms take into account the position of residues in relation to secondary structure elements (SSEs). Residues in SSEs or solvent inaccessible regions are generally more conserved than those in folded or solvent accessible regions, therefore insertions and deletions are less likely to occur in these conserved areas (186). Unlike CLUSTALW (206), MUSCLE (193) or Toffee (187), the FUGUE alignment creates a sequence alignment based on SSEs which promises to be more reliable.

The FUGUE alignment was viewed using CORE evaluation on the Toffee server (Figure 2.2). This evaluates the consistency between each pair of aligned residues and places them on a coloured core index scale between 0-9 (blue-red). Residues with a core index ≥ 5 have a 90% chance of being correctly aligned (187). Areas of red indicate strong sequence homology whilst yellow and blue areas indicate average and poor homology respectively. Areas with extremely poor or no homology, such as insertions and deletions, have no colour.

Figure 2.2 shows that the majority of the residues are accurately aligned and have good sequence homology. Comparison of the Toffee alignment with a DSSP representation of secondary structure (207) revealed that areas of the alignment with poor homology corresponded to loop regions of the protein and were not present in alpha-helix or beta-sheet regions. As loop regions have more variability, in terms of the residues that they contain, regions of low sequence homology are expected. A large region containing no homology or alignment to 1W6K was observed at the N-terminus of AsbAS1. This was a 33 amino acid insertion (residues 61-93) which was presumed to form a loop region immediately following the amino terminal domain. Due to the large number of residues in this insertion it was postulated that it may have formed an additional secondary structure domain, but secondary structure predictions failed to assign any structure to this region. A sequence alignment of the AsbAS1 sequence with a selection of plant OSCs from a variety of plant species revealed that these enzymes also contained the same inserted region which was restricted to plants and not present in bacteria or other eukaryotes. This region had high sequence

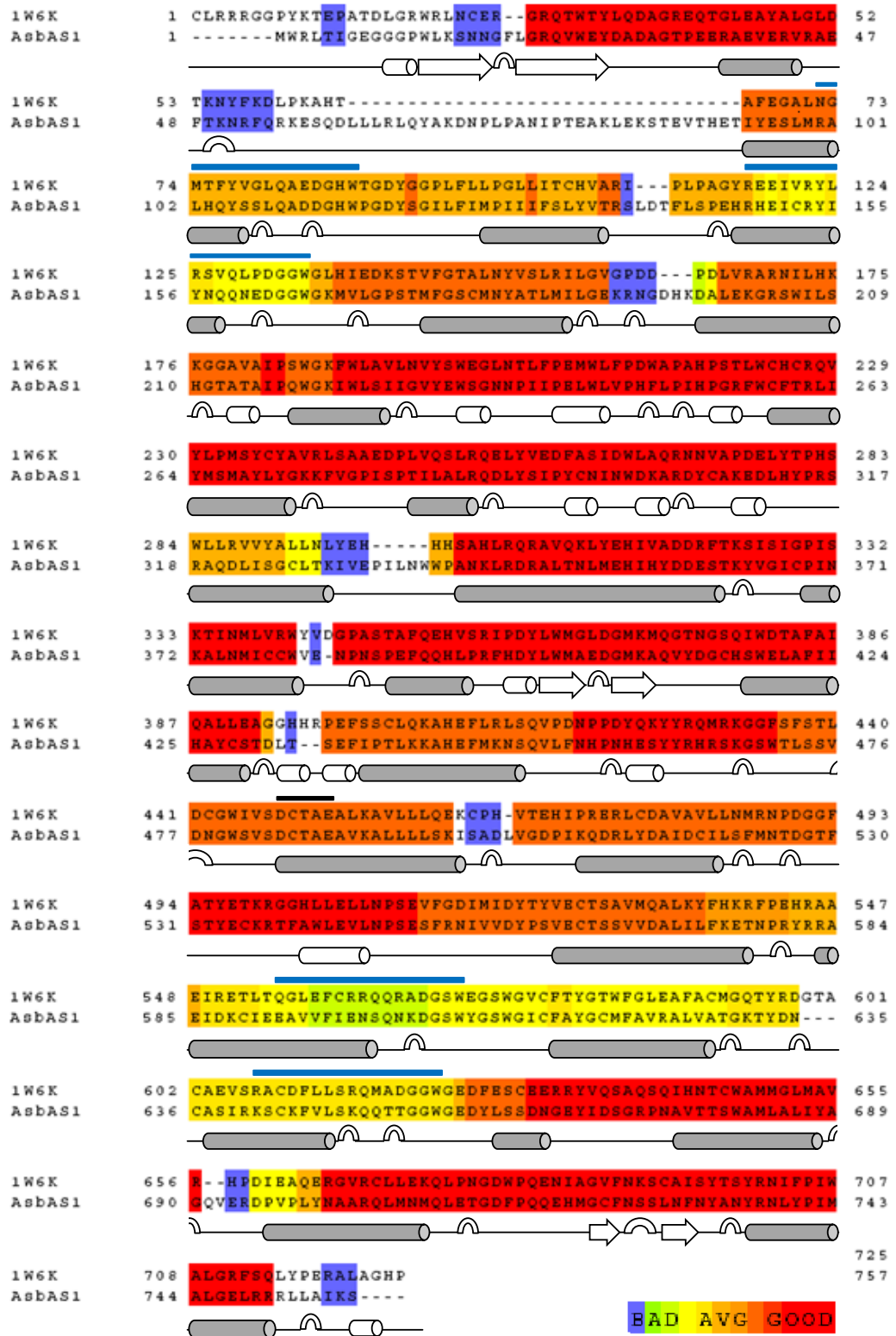


Figure 2.2: FUGUE alignment for AsbAS1 against 1W6K

FUGUE structure-based sequence alignment of AsbAS1 against the template sequence, 1W6K. The scale at the bottom denotes the strength of the alignment, with red indicating a strong alignment, orange/yellow indicating an average alignment and green/blue indicating a weak alignment. Proposed insertions to the 1W6K sequences are shown as dashed lines. The DSSP representation of secondary structure is shown below the alignment with alpha and 3-10 helices shown as grey and white cylinders respectively; beta sheets shown as arrows and hydrogen bonded turns shown as an arc. Two important conserved motifs, the catalytic DCTAE motif and QW motifs are shown above the sequence alignment in black and blue respectively.

identity in all plant species indicating that it may have a specific function in plants, possibly in additional enzyme stability or protein-protein interactions, but it is not clear.

The sequence was examined in detail to determine whether the insertions and deletions proposed by the FUGUE alignment could be accommodated into the 1W6K structure and each position was viewed using PyMOL. There were six insertions and three deletions proposed which, due to the ability of FUGUE to recognise SSEs all occurred in loop regions so could be accommodated into the 1W6K structure without disrupting secondary structure elements.

2.3.1.4 Generating a structural model for AsbAS1

An AsbAS1 homology model was generated by MODELLER. MODELLER uses comparative protein structure modelling to generate a model from an alignment of the target sequence to the template sequence. The model is assigned a number of restraints, which assume that distances between template residues will be the same as for the aligned residues in the target sequence. The molecular mechanics force-field adds additional restraints on bond lengths, bond angles, dihedral angles and non-bonded atom-atom contacts. The model then undergoes optimization which minimises the energy conflicts. The final model coordinates are exported as a pdb format file (189).

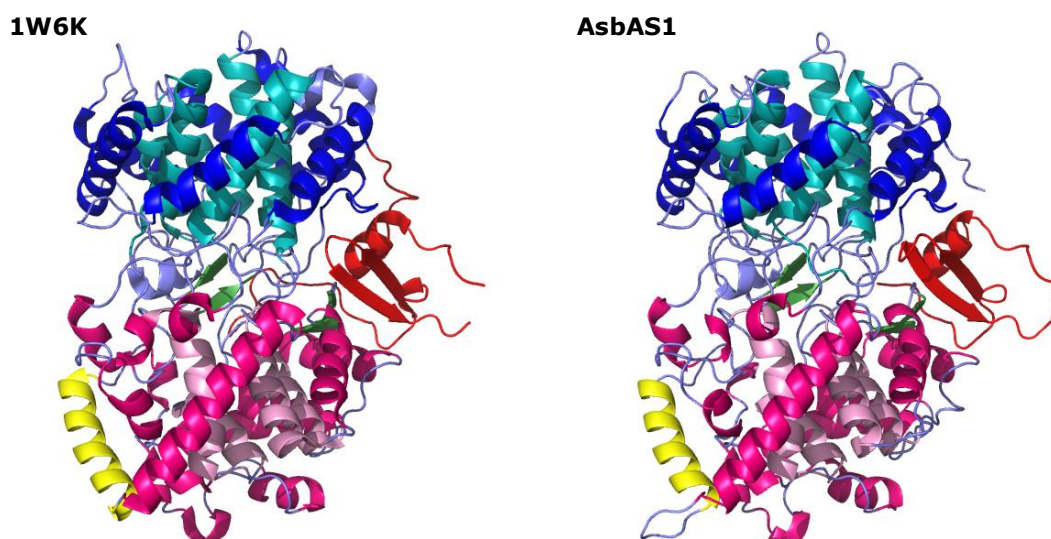


Figure 2.3: Comparison of the AsbAS1 model to the 1W6K structure

The 1W6K template structure is shown on the left and the AsbAS1 model is shown on the right. Domain 1 is a double α -helical barrel with the inner helices shown in light blue and the outer helices shown in dark blue. Domain 2 consists of 12 α -helices with the outer helices shown in dark pink and the inner helices shown in light pink. β -sheets are found between the two domains and are shown in green. The amino terminal domain is shown in red and the membrane attachment helix is shown in yellow.

The AsbAS1 model structure was viewed PyMOL and superimposed onto the 1W6K template structure. As expected from the secondary structure predictions, no secondary structure was assigned to the 33 amino acid inserted region at the N-terminus. This region was modelled as an unordered loop which extended away from the surface of the protein, therefore as it contained no secondary structure information it was removed from the final model (Appendix 3). Structural comparison of the final model against the template structure (Figure 2.3) shows that many of the secondary structure elements and motifs are conserved in the AsbAS1 model, with the overall domain structure almost identical between the two structures. This preliminary analysis indicates that 1W6K appears to be a good template for AsbAS1.

2.3.1.5 Validation of the AsbAS1 model

The AsbAS1 model was subjected to validation to determine whether a reliable model had been produced. The validation process examined the overall energy of the structure, dihedral angles and atomic volumes. This would give an indication of the stability of the model and determine whether it had the features characteristic of other protein structures solved by X-ray crystallography and refined at high resolution.

Overall protein structure was analysed using Prosa II, which calculates the energy of the whole structure using a distance-based pair potential which incorporates all forces acting between groups of atoms and the surrounding medium. This energy value is normalised to give a Z-score which is displayed in a plot containing the Z-score of all proteins currently in the PDB (208). Z-scores differ based on the length of the protein, with small proteins having high Z-scores whereas larger proteins have much lower Z-scores. 1W6K and AsbAS1 were calculated to have Z-scores of -11.31 and -10.56 respectively which places them well within the range of Z-scores observed for proteins of similar size (Figure 2.4). Due to their similar size and high sequence homology, the Z-scores of both structures were expected to have similar values and this was observed. This confirms that the overall structure of the model produced is reliable and unlikely to contain major structural errors.

The AsbAS1 model was analysed in more detail using PROCHECK, which calculates the phi (ϕ) and psi (ψ) angles for all residues and displays them on a Ramachandran plot. All residues fall into distinct areas of the Ramachandran plot depending on the secondary structure element they are located in (alpha-helix, beta-sheet, left-handed helix, beta-turns and extended chains). The majority of

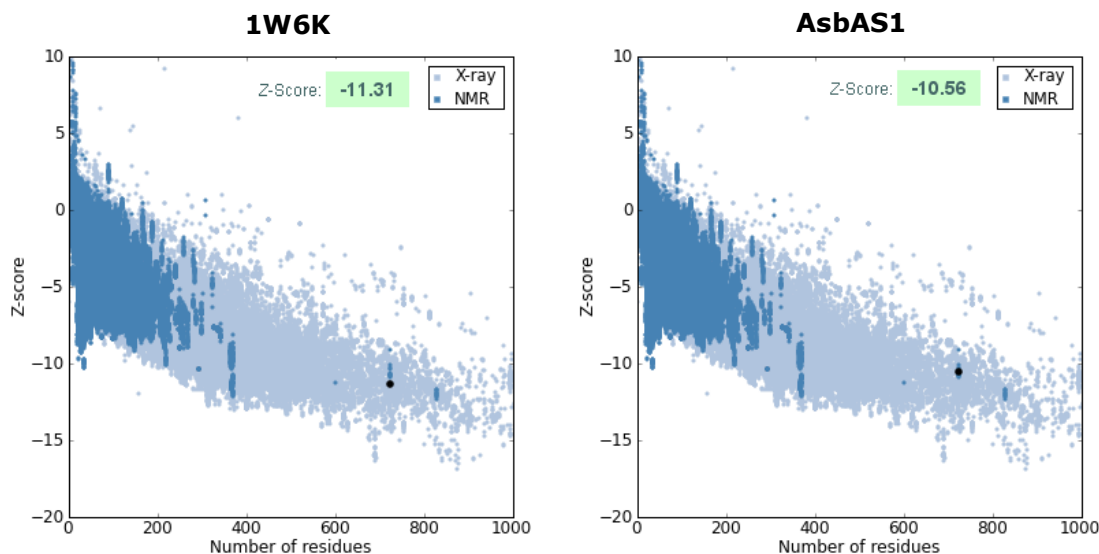


Figure 2.4: Prosa II analysis of AsbAS1 model

The overall energy (Z-score) of the AsbAS1 model (right) was calculated using Prosa II and compared to the 1W6K template structure (left). The light and dark blue dots show the Z-score of all known structures in the PDB relative to their size. The black dot shows where the analysed structure is placed. Both AsbAS1 and 1W6K have comparable Z-scores to other proteins of their size.

these residues should fall into these distinct areas, and structures solved to 2Å resolution are expected to have over 90% of residues in allowed regions. Structures that contain a significant number of residues in disallowed areas are likely to have steric irregularities therefore indicating that the model quality is low.

PROCHECK analysis of the AsbAS1 model showed that 99.3% of residues were in allowed regions (90.3% in most favoured regions, 7.8% in additionally allowed regions and 1.4% in generously allowed regions) (Figure 2.5). 1W6K also showed similar results with 99.7% of residues in allowed regions (93.3% in most favoured regions, 5.9% in additionally allowed regions and 0.5% in generously allowed regions). The AsbAS1 model had only three residues in disallowed regions, Glu459, Ser434 and Val692, all of which were present in loop regions on the protein surface. Surface loop regions are flexible and mobile, so unusual dihedral angles are commonly found. These data show that the majority of the dihedral angles in the AsbAS1 model are in allowed regions of the Ramachandran plot which further verifies that the model is reliable.

The final validation of the AsbAS1 was carried out using PROVE which analyses the atomic volumes for each of the residues in the structure. These values are compared to normal ranges for atomic and residue volumes that have been calculated from 64 highly resolved and well refined structures (191). The

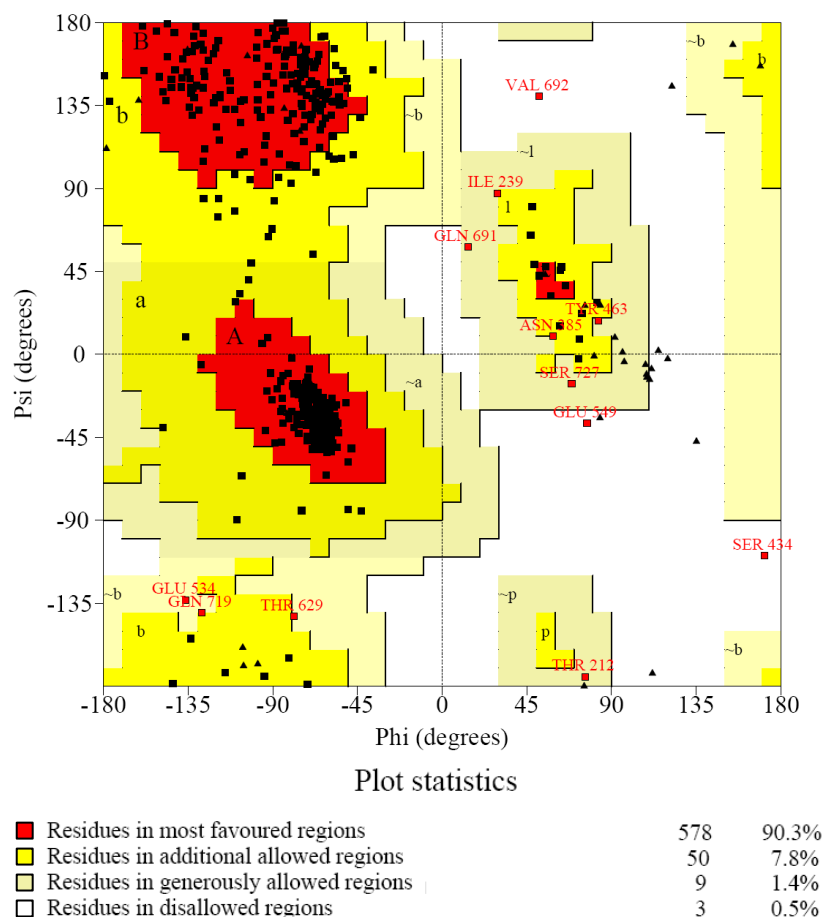


Figure 2.5: Ramachandran plot of AsbAS1 model

The dihedral angles of the AsbAS1 model were analysed by PROCHECK and the results displayed in a Ramachandran plot. The plot is split into different regions: red – most favoured, yellow – additionally allowed, beige – generously allowed and white – disallowed. The majority of the residues fall within the red and yellow areas indicating that the model is reliable.

results showed that atomic volumes throughout the AsbAS1 structure were similar to 1W6K and both scores were within the range seen for other proteins solved at the same resolution.

Together, the validation methods show that structural parameters for AsbAS1 are very similar to the template structure 1W6K and other proteins of similar size. It was therefore concluded that the model produced was reliable and could now be used for a detailed analysis of β -amylin synthase structure and mechanism.

2.3.1.6 Overall AsbAS1 structure

The AsbAS1 model structure was superimposed onto that of hOSC using PyMOL (Figure 2.6). As observed in Figure 2.3, many of the secondary structure elements and motifs are conserved in the AsbAS1 model, with the overall domain

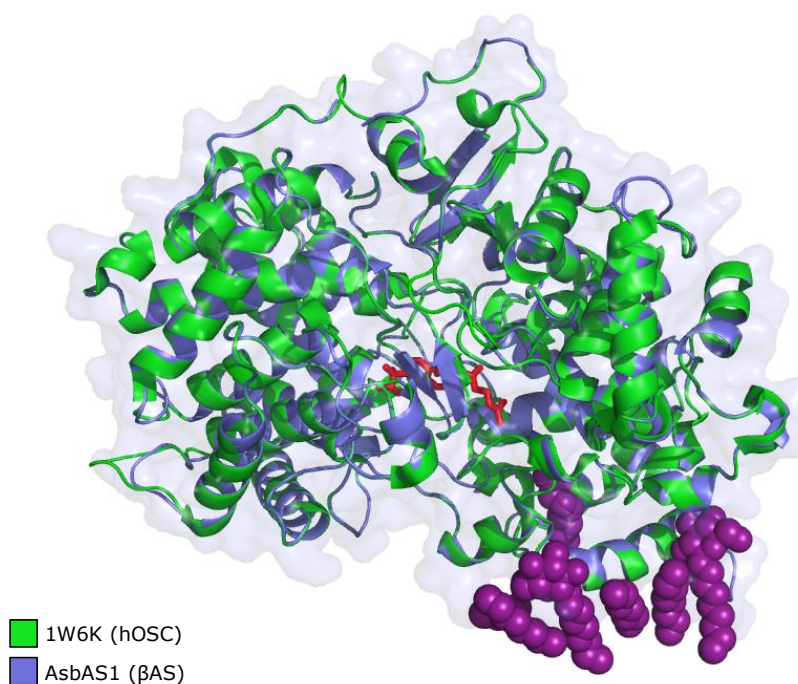


Figure 2.6: Comparison of structural features of AsbAS1 and hOSC

Superimposed image of hOSC crystal structure (green) and the AsbAS1 model (lilac). The features shown in spheres were solved from the hOSC crystal structure. The product of hOSC catalysis – lanosterol (red sticks) is shown bound in the active site. Detergent molecules (purple spheres) used for solubilisation and crystallisation of hOSC show the orientation of the enzyme at the membrane. A detergent molecule is also bound in the substrate access channel.

structure almost identical between the two structures. The superimposed image also shows that alpha helices and beta sheets are of similar length and although some loop regions are longer in AsbAS1, they are well aligned to 1W6K.

AsbAS1, like 1W6K, has two α - α barrel domains connected by loop regions which form three β -structures. The active site is found between the two domains where the product of hOSC catalysis, lanosterol, can be seen represented by red sticks. The hOSC protein was crystallized with detergent present as it is a membrane-associated protein, and these can be seen in the crystal structure represented by purple spheres (151). The location of these detergents shows how the protein is orientated in the membrane and also show position of the substrate access channel which leads from the membrane to the active site.

2.3.2 Analysis of the initial stages of AsbAS1 mechanism

The AsbAS1 structure and mechanism were analysed in detail to investigate the similarities and differences between sterol and triterpene OSCs. This would allow the mechanistic differences between the two enzyme classes to be investigated at a structural level. As human lanosterol synthase (hOSC) is the only sterol OSC to have its crystal structure solved, this structure was used as a

representative structure for sterol OSCs (144). In addition to looking at structural information, sequence and phylogenetic analysis was also used to investigate OSC mechanism. A large number of OSCs (both sterol and triterpene OSCs) from a variety of different plant species have been cloned and functionally characterised and their sequence information together with the structural information from the hOSC crystal structure and the AsbAS1 model provided additional insights into OSC mechanism. **Note that in this section, in the main text and figure legends, hOSC residues are numbered in normal text and AsbAS1 residues are numbered in italics. In figures both enzyme residues are numbered in normal text and colour coded as described.**

2.3.2.1 Phylogenetic analysis of plant oxidosqualene cyclases

To investigate the phylogeny between sterol and triterpene OSCs, a multiple sequence alignment was generated in MUSCLE (193). This included 81 amino acid sequences from all cloned and/or functionally characterised plant OSCs, together with 20 sequences for predicted OSCs from a variety of plant species. The multiple sequence alignment was then used to generate a neighbour-joining phylogenetic tree (195) (Figure 2.7).

Preliminary analysis supports previous conclusions made by Qi *et al.* (63) that AsbAS1 is more closely related to plant cycloartenol synthases than the dicot β -amyrin synthases. With the increasing availability of sequence data from other monocot species, it is clear to see that all predicted monocot triterpene synthases are found in the same clade as AsbAS1 indicating that the monocot and dicot triterpene synthases have evolved separately. The tree also shows that enzymes involved in sterol synthesis have remained constant throughout evolution, with cycloartenol synthase being the sole enzyme required for sterol synthesis in all plants. Enzymes involved in triterpene synthesis however, have undergone much more diversification in all plant species, resulting in evolution of both specific and multifunctional enzymes that produce a wide range of secondary metabolites. Some of these enzyme classes are found in clearly defined clades, for example, the dicot β -amyrin synthases and the dicot lupeol synthases. The remainder of the clades contain many enzyme classes that produce a vast range of products. With the recent advances in whole genome sequencing, as more sequences become available these clades will become more clearly defined and evolutionary relationships between triterpene synthases will become more clearly understood.

The information gathered from this comprehensive sequence alignment and phylogenetic tree was then used together with the AsbAS1 model to

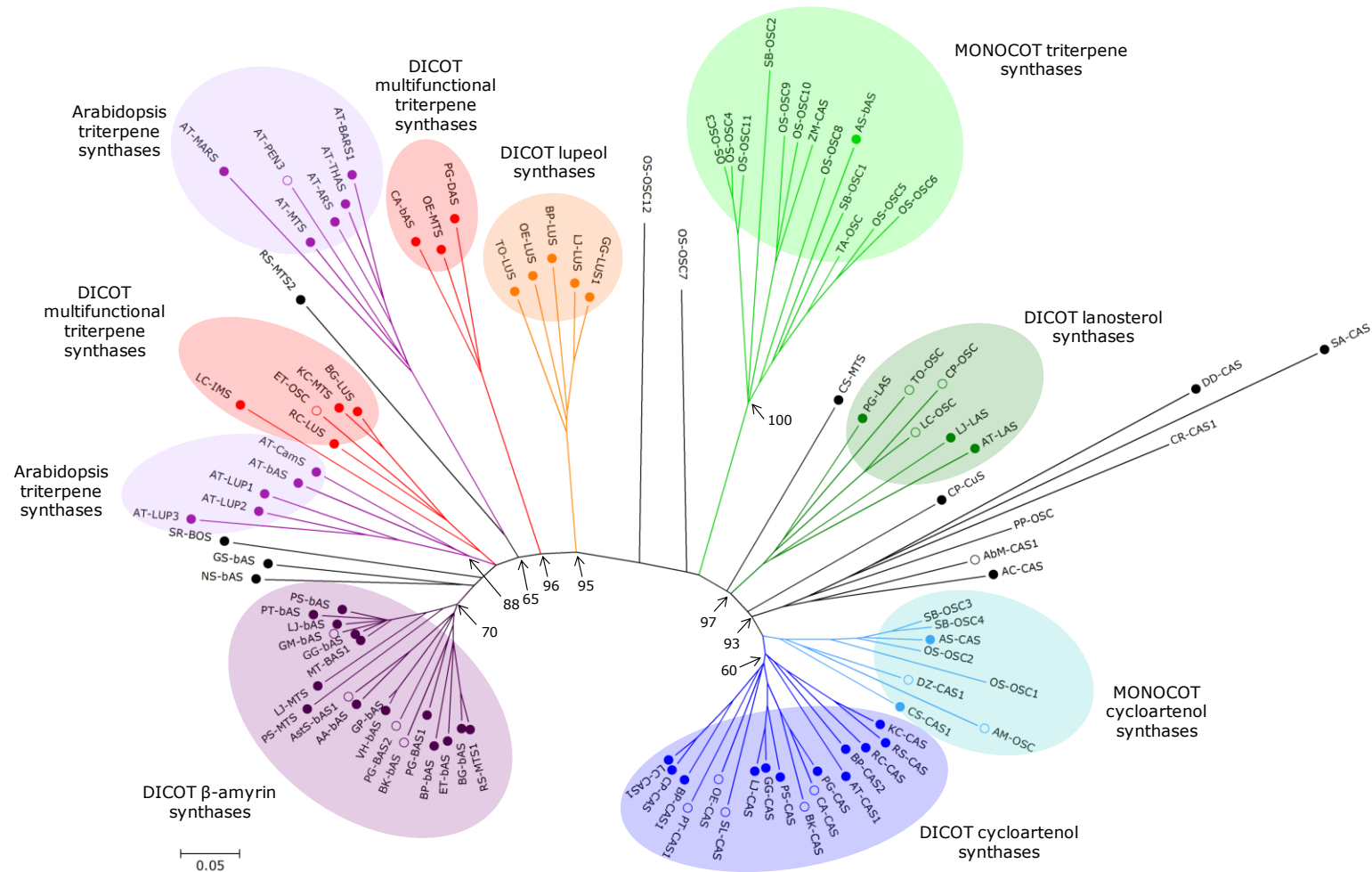


Figure 2.7: Phylogenetic analysis of plant oxidosqualene cyclases.

A neighbour-joining phylogenetic analysis of aligned protein sequences from plant OSCs is shown. Bootstrap values (percentage values from 1000 replicates) are shown for key branches. The scale bar indicates 0.05 substitutions per site. The clades are highlighted based on enzyme classification. Filled circles (●) indicate enzymes that have been functionally characterised in yeast, empty circles (○) indicate enzymes that have been cloned and the remainder that are blank are predicted sequences deduced from whole genome sequencing. Protein sequence accession numbers and the multiple sequence alignment can be found in Appendices 4 and 5 respectively

investigate AsbAS1 mechanism and the mechanistic differences between the sterol and triterpene OSC classes.

2.3.2.2 Substrate entry to the active site

The hOSC structure contains a detergent molecule bound in a channel leading from the membrane to the active site, which allows the passage of the substrate, 2,3-oxidosqualene. This channel is also observed in the AsbAS1 model. In hOSC, this channel is blocked by three residues, Tyr237, Cys233 and Ile524 (hOSC numbering), which are presumed to undergo a conformational change to allow the substrate to enter the active site (144). Examination of the same region in the AsbAS1 model revealed that two of the residues, Tyr237 and Ile524, were conserved, and the third residue, Cys233, was substituted with threonine which of a similar size (Figure 2.8). These three residues were found in the same orientation in both hOSC and AsbAS1 indicating that substrate entry into the active site occurs by the same mechanism for both sterol and triterpene enzymes. The multiple sequence alignment of all plant OSCs showed that these three residues are very highly conserved (Appendix 5), with any substitutions being for residues of similar size and charge, which provides further evidence for a single substrate entry mechanism in all OSCs.

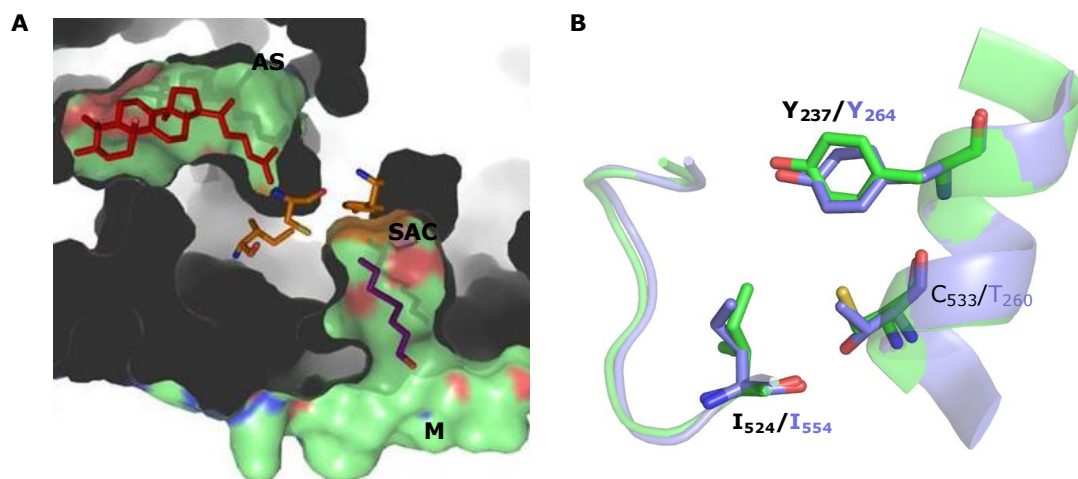


Figure 2.8: Comparison of residues blocking substrate access in AsbAS1 and hOSC

A. Image to show path of substrate entry to active site in hOSC. AS – Active site; SAC – Substrate access channel; M - Membrane attachment surface. The product, shown in red sticks, is bound in the active site. The substrate access channel leads from the membrane to the active site has a detergent molecule (purple sticks) bound. The channel is blocked by 3 residues, labelled in orange sticks, that separate the channel from the active site. They are predicted to undergo a conformational change to allow the substrate to enter the active site. **B.** Superimposition of residues blocking substrate access in hOSC (green, black labels) and AsbAS1 (purple, purple labels). Residues labelled in bold are conserved between the two enzymes.

2.3.2.3 Pre-folding of the substrate in the active site

After the 2,3-oxidosqualene substrate has entered the active site, the first three rings are pre-folded into an energetically favourable conformation for cyclisation. Each ring is capable of forming two different conformations, chair or boat, and the combination of these conformations differs between sterol and triterpene enzymes and represents one of the major differences between their mechanisms. Sterol OSCs are pre-folded into the chair-boat-chair conformation, whilst in triterpene OSCs the chair-chair-chair conformation is formed (Figure 2.9).

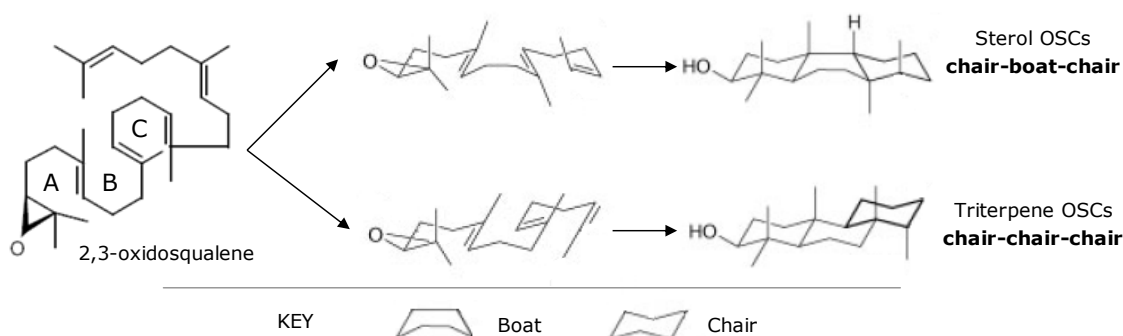


Figure 2.9: Pre-folded conformations of 2,3-oxidosqualene for sterol and triterpene cyclisation.

Diagram to show how 2,3-oxidosqualene is pre-folded in the active site of sterol and triterpene OSCs. Each ring of the substrate can be folded in two different conformations – chair and boat. In sterol OSCs, the first 3 rings of 2,3-oxidosqualene form the chair-boat-chair conformation whilst triterpene OSCs form the chair-chair-chair conformation. This difference in B-ring conformation represents one of the major differences between sterol and triterpene OSCs.

In hOSC the B-ring forms a boat conformation due to the presence of Tyr98, which pushes the C₁₀ carbon of 2,3-oxidosqualene below the molecular plane. The boat conformation is energetically unfavourable, due to the flagpole interaction between the two axial 1,4-hydrogen atoms of the carbon ring, but is enforced due to the steric pressure of the tyrosine (144, 209). Therefore in triterpene OSCs, a lack of steric pressure was expected at the equivalent Tyr98 position to allow the energetically favourable chair conformation to be formed. The multiple sequence alignment showed that dicot triterpene OSCs had a smaller residue at the equivalent position which would remove the steric pressure on the B-ring, but a tyrosine was still present in the monocot triterpene OSCs. Therefore, this region was studied in the superimposed structures of AsbAS1 and hOSC to understand how the different B-ring conformations are achieved (Figure 2.10).

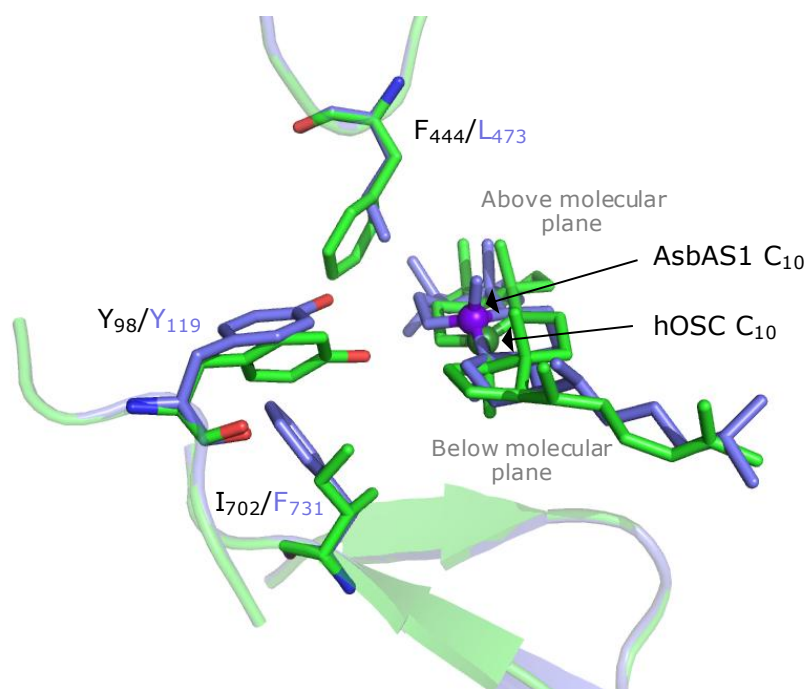


Figure 2.10: Residues enforcing B-ring conformation

Diagram to illustrate how B-ring conformation is determined in hOSC (green, black labels) and AsbAS1 (purple, purple labels). Tyr₉₈ in hOSC enforces the boat conformation by forcing the C₁₀ of 2,3-oxidosqualene below the molecular plain. Tyr₉₈ is held in place by a large residue above and a small residue below. In AsbAS1 these positioning residues are reversed which elevates the tyrosine relative to its hOSC counterpart. This removes the steric pressure on C₁₀ and allows the B-ring to adopt the energetically favourable chair conformation.

In hOSC, Tyr₉₈ is held in position by a bulky residue above, Phe₄₄₄, and a small residue below, Ile₇₀₂, which creates the steric pressure to enforce the boat conformation. In AsbAS1, these positioning residues are reversed, so the tyrosine (Tyr₁₁₉) is held in place by a small residue above, Leu₄₇₃, and a bulky residue below, Phe₇₃₁, which elevates its position in relation to the hOSC tyrosine. This elevation presumably removes the steric pressure on the substrate C₁₀ to allow it to adopt the chair conformation. The sequence and phylogenetic analysis of all plant OSCs shows that this residue reversal is restricted to the monocot triterpene OSC clade. Dicot triterpene OSCs have the same positioning residues as hOSC but have a much smaller residue, typically asparagine, cysteine or isoleucine, at the equivalent Tyr₉₈ position so the B-ring is free to adopt the favourable chair conformation. This clear difference between monocot and dicot OSCs may be a result of the divergent evolutionary origins of the two groups. Phylogenetic analysis has shown that monocot triterpene OSCs are more closely related to cycloartenol synthases, responsible for plant sterol biosynthesis, than other dicot plant triterpene OSCs, indicating the two triterpene OSC groups evolved separately (63). The more cycloartenol synthase-like character of the monocot triterpene OSCs might explain why some of the features of sterol OSCs

have been maintained and modified to allow the enzyme to function as a triterpene OSC.

2.3.2.4 Reaction initiation

Once the substrate is pre-folded in the active site, cyclisation can occur. The reaction is initiated by protonation of the epoxide group of 2,3-oxidosqualene by a highly conserved aspartate residue which induces the cyclisation reaction. The conserved aspartate is part of a highly conserved DCTAE catalytic sequence motif and is present in all OSCs, suggesting that the reaction initiation mechanism is conserved amongst all OSCs (52, 124, 210-211). In hOSC, Asp455 is the proton donor for reaction initiation and is hydrogen bonded to two cysteine residues, Cys456 and Cys533, which increase its acidity to allow protonation of the 2,3-oxidosqualene epoxide group. This aspartate (Asp484) is conserved in AsbAS1 and is hydrogen bonded to two cysteine residues, Cys485 and Cys563. All three residues were found in an identical orientation to hOSC which implies that initiation occurs by a common mechanism (Figure 2.11) (144, 212).

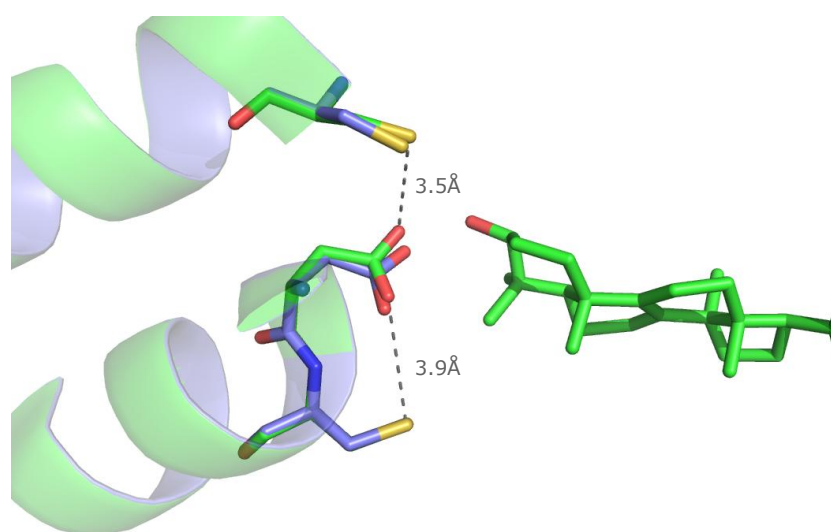


Figure 2.11: Reaction initiation

Superimposition of the residues involved in cyclisation initiation in hOSC (green, black labels) and AsbAS1 (purple, purple labels). Initiation occurs via a catalytic aspartate hydrogen bonded to two cysteine residues which increase its acidity and allow it to protonate the epoxide group of 2,3-oxidosqualene substrate. Hydrogen bonds are shown by dashed lines.

2.3.3 Analysis of the cyclisation steps in AsbAS1 mechanism

After reaction initiation, cyclisation of the substrate can proceed. Cyclisation was originally thought to be a concerted process (152), but experiments conducted by Van Tamelen *et al.* found evidence that cyclisation proceeded via a series of discrete, conformationally rigid carbocationic

intermediates (153). Rings are numbered A through to E and are formed sequentially, with each ring requiring stabilisation before the next one can be formed. Stabilisation occurs via aromatic residues in the active site cavity which stabilise the positive charge formed on the newly cyclised ring via cation- π interactions. The aromatic ring stacks against the new ring and the delocalised electrons stabilise the positive charge (164).

Ring formation represents another major difference between sterol and triterpene OSC mechanisms as sterol OSCs form tetracyclic products whereas triterpene OSCs predominantly form pentacyclic products. Mutagenesis studies have shown that ring stabilisation by active site aromatic residues plays a vital role in product determination (167-171, 174, 176). Mutagenesis of only one critical active site residue can result in cyclisation termination due to lack of stabilisation of the developing positive charge. (139, 213). Stabilisation is only achieved by aromatic residues, therefore a single substitution of a stabilising residue can change the active site environment and result in a different product being formed. Due to the high sequence similarity between sterol and triterpene OSCs, identification of subtle differences in active site residues could allow further understanding of the mechanistic differences between these two enzyme groups.

To further analyse AsbAS1 mechanism, each step of the cyclisation process was studied in detail to identify residues that are likely to be catalytically important and explore differences to the sterol OSC mechanism. Reaction intermediates, generated from structures in the Cambridge structural database, were docked into the active site of the AsbAS1 model (Figure 2.12). Molecular dynamics (MD) simulations were then performed on all docked intermediates to create a realistic representation of their position in the active site.

MD simulates the movement of atoms and molecules in protein structures. *In vivo*, all atoms and molecules are constantly moving, so MD simulates the atom movements in an aqueous environment to find an optimal, energetically favourable position for the intermediate in the active site. Despite being membrane associated, the model was simulated in an aqueous environment. Superimposition of AsbAS1 onto hOSC implies that like hOSC, AsbAS1 is membrane associated via a single helix therefore the majority of the protein is exposed to the aqueous environment of the cytoplasm. So, whilst membrane attachment may affect the low frequency movements associated with substrate entry, it is unlikely to affect the high frequency movements of the active site during catalysis. Therefore the simulation of reaction intermediates can be accurately modelled using an aqueous environment.

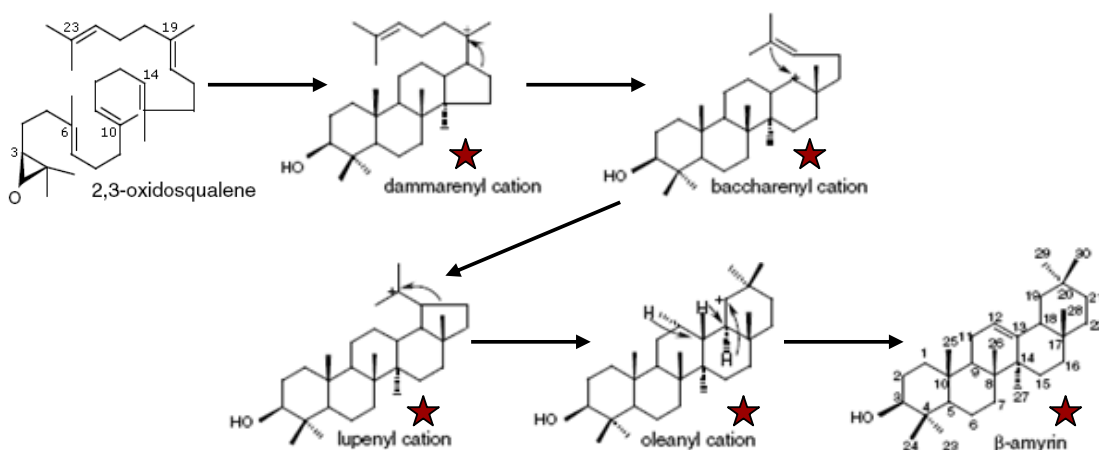


Figure 2.12: Carbocation intermediates of β -amyrin synthesis

Pathway to show intermediates formed on cyclisation of 2,3-oxidosqualene to β -amyrin. Cyclisation is concerted up to formation of the 5-membered D-ring (dammarenyl cation). Three more intermediates are formed upon D-ring expansion, E-ring formation and E-ring expansion leading to the oleanyl cation which undergoes skeletal rearrangement followed by deprotonation to form the final β -amyrin product. Red stars denote the structures docked into the AsbAS1 active site using molecular dynamics.

The final AsbAS1 model structures were analysed to identify residues involved in ring stabilisation. Stabilising residues from AsbAS1 and hOSC were compared along with information from the multiple sequence alignment to find residues responsible for product differences between sterol and triterpene OSCs.

2.3.3.1 A-ring formation

The A-ring is the first to be cyclised and ring closure forms a positive charge at C₆ (2,3-oxidosqualene numbering – Figure 2.12). This charge must be stabilised in order for cyclisation to continue, and in hOSC this is achieved by Trp581. Trp581 is located below C₆ in the active site, and forms cation- π interactions to stabilise the positive charge (144). Further stabilisation comes from the hydroxyl groups of three nearby tyrosine residues, Tyr503, Tyr704 and Tyr707, which are directed towards the active site to provide additional negative electrostatic potential (165).

Analysis of A-ring stabilisation in the AsbAS1 model was carried out with the tetracyclic dammarenyl cation docked into the active site, as the CSD did not contain a suitable template to generate a monocyclic triterpene structure. Residues for A-ring stabilisation are completely conserved between the two enzymes, with the C₆ cation stabilised from below by Trp611, and tyrosine residues Tyr533, Tyr733 and Tyr736 (AsbAS1 numbering) providing additional negative charge. The plant OSC multiple alignment shows that with the exception of Tyr533 which is replaced by a tryptophan in dicot triterpene

synthases, these residues are highly conserved in all sterol and triterpene OSCs, suggesting that A-ring formation occurs by the same mechanism.

2.3.3.2 B-ring formation

Formation of the B-ring results in a shift of the positive charge from C₆ to C₁₀ (2,3-oxidosqualene numbering – Figure 2.12). In hOSC, this charge is stabilised via cation- π interactions from above and below by Phe444 and Trp581 respectively. As for A-ring formation, the hydroxyl groups of surrounding tyrosine residues provide additional negative charge to stabilise the developing positive charge (144).

Analysis of B-ring stabilisation in the AsbAS1 model was carried out with the tetracyclic dammarenyl cation docked into the active site, as the CSD did not contain a suitable template to generate a bicyclic triterpene structure. For B-ring stabilisation, the conserved *Trp611* is present below C₁₀ to stabilise the positive charge and tyrosine residues *Tyr533*, *Tyr733* and *Tyr736* are also conserved. However, Phe444 is not conserved in AsbAS1 and other monocot triterpene OSCs and is substituted for a leucine residue (*Leu473*), which appears to play a role in *Tyr119* positioning to allow the B-ring to form a chair conformation (see section 2.3.2.3). As stated in section 2.3.2.3, *Tyr119* is positioned by a pair of residues, *Phe731* (below) and *Leu473* (above), both of which are conserved in monocot triterpene OSCs. Together they raise the

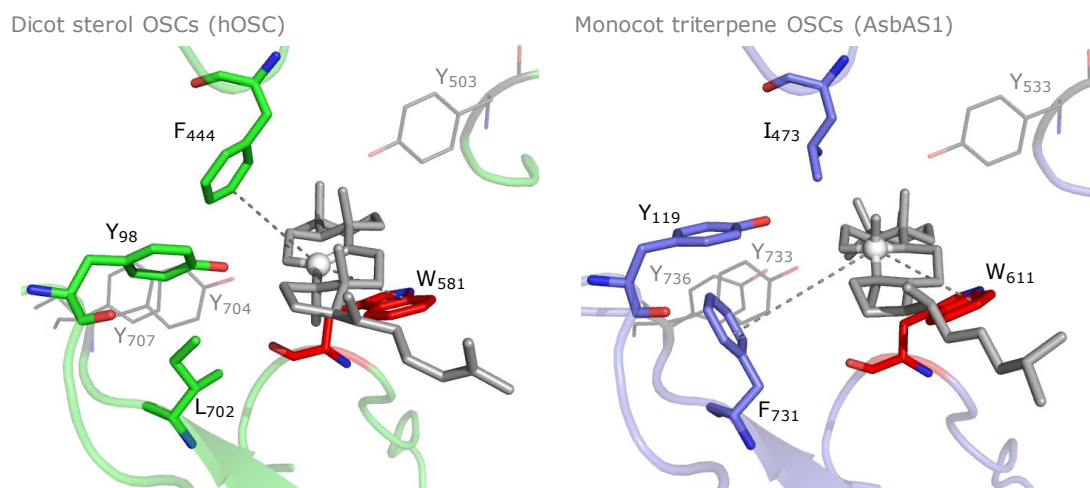


Figure 2.13: B-ring stabilisation in hOSC and AsbAS1

Diagram shows differences in B-ring stabilisation between hOSC (left) and AsbAS1 (right). The reaction intermediate is shown in grey and the C₁₀ cation is shown as a white sphere. In hOSC (and all dicot OSCs), C₁₀ stabilisation occurs via Trp581 and Phe444. In AsbAS1 C₁₀ stabilisation occurs via Trp611 (conserved amongst all OSCs) and Phe731 (unique to monocot triterpene OSCs). Residues involved in stabilisation and conformation of the B-ring are shown in green (hOSC) and purple (AsbAS1), and conserved residues are shown in red. Tyrosine residues providing additional negative potential to the active site are shown as light grey sticks.

position of *Tyr119* in relation to the positive charge at C_{10} , so that it no longer causes steric hindrance the B-ring. In hOSC and all sterol OSCs, these positioning residues (*Phe444* and *Ile702*) are reversed which lowers the position of *Tyr98* in relation to the positive charge at C_{10} , which creates the steric hindrance to enforce the boat conformation of the B-ring. *Phe731* is therefore indirectly conserved, and although it is found in a different location to *Phe444*, it is appropriately positioned to stabilise the C_{10} cation (Figure 2.13). Monocot triterpene OSCs are the only triterpene OSCs that contain an equivalent *Tyr98* residue, usually only observed in sterol OSCs. This difference in B-ring stabilisation may have arisen as a result of the separate evolution of monocot and dicot triterpene OSCs (63).

2.3.3.3 C-ring formation

Formation of the C-ring results in a shift of the positive charge from C_{10} to C_{14} (2,3-oxidosqualene numbering – Figure 2.12). In hOSC this charge is stabilised via cation- π interactions from above and below by *Phe444* and *Phe696* respectively. *His232* also interacts with the C_{14} cation to stabilise the charge (144). Mutagenesis studies of *Saccharomyces cerevisiae* lanosterol synthase have shown that deletion of *His232* or *Phe444* leads to formation of tricyclic products, indicating that these residues are vital for C-ring formation (167-168).

Analysis of C-ring stabilisation in the AsbAS1 model was carried out with the tetracyclic dammarenyl cation docked into the active site, as the CSD did not

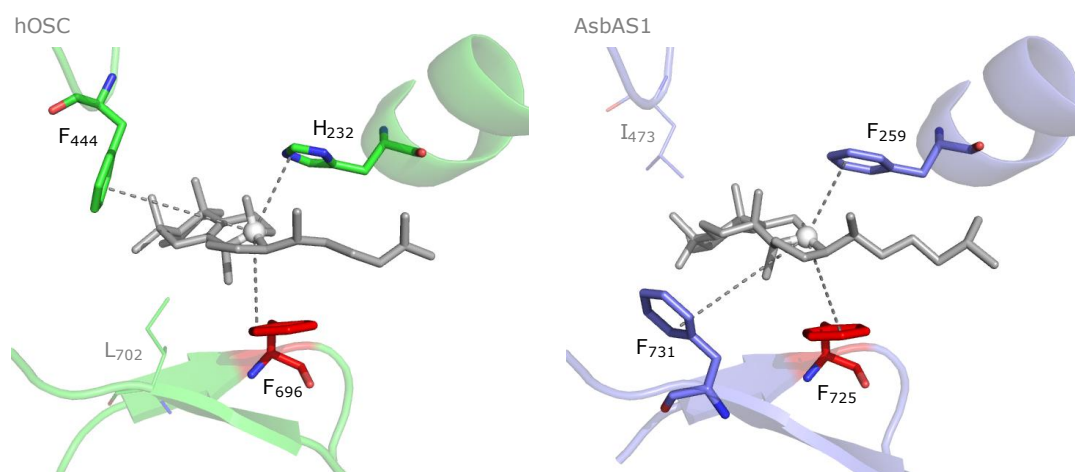


Figure 2.14: C-ring stabilisation in hOSC and AsbAS1

Diagram shows differences in C-ring stabilisation between hOSC (left) and AsbAS1 (right). The reaction intermediate is shown in grey and the C_{14} cation is shown as a white sphere. In hOSC (and all dicot OSCs), C_{14} stabilisation is mediated by *His232*, *Phe444* and *Phe696*. In AsbAS1, *Phe696* is conserved (*Phe725*) and *Phe444* is indirectly conserved (*Phe731*), but *His232*, a residue conserved in all sterol OSCs, is replaced by *Phe259*, a residue conserved in all triterpene OSCs. Residues involved in stabilisation of the C-ring are shown in green (hOSC) and purple (AsbAS1), and conserved residues are shown in red.

contain a suitable template to generate a tricyclic triterpene structure. For C-ring stabilisation, the phenylalanine below C_{14} (*Phe725*) is conserved, but *Phe444* and *His232* are not conserved (Figure 2.14). However, as *Phe444* is indirectly conserved (*Phe731*), it may assume the same role that *Phe444* has in B and C-ring stabilisation. The absence of *His232* represents one of the major differences between sterol and triterpene OSCs. The plant OSC multiple sequence alignment (Appendix 5) shows that *His232* is conserved throughout all sterol OSCs whereas all triterpene OSCs have an aromatic phenylalanine (*Phe259*) or tyrosine. *Phe259* plays a vital role in later cyclisations but for C-ring stabilisation it assumes the same role as *His232*.

2.3.3.4 D-ring formation

Formation of the 5-membered D-ring results in a shift of positive charge from C_{14} (2,3-oxidosqualene numbering – Figure 2.12) to C_{20} (lanosterol numbering), forming the protosteryl cation in hOSC and the dammarenyl cation in AsbAS1. It is here that the sterol and triterpene OSC mechanisms diverge, with D-ring formation signalling the end of sterol OSC cyclisation, whereas triterpene OSC cyclisation continues. In hOSC, the positive charge at C_{20} is coordinated from above and below by *His232* and *Phe696* respectively. However, the negative potential of these residues is not sufficient to stabilise the charge at C_{20} , which would allow expansion of the D-ring and formation of the 5-membered E-ring. Lack of stabilisation results in cyclisation termination and deprotonation takes place to yield the lanosterol product (144, 165).

Analysis of D-ring stabilisation in the AsbAS1 model was carried out with the tetracyclic dammarenyl cation and baccharenyl cations docked into the active site. In AsbAS1 the phenylalanine below C_{20} (*Phe725*) is conserved, but *His232* is not conserved and is replaced by a phenylalanine (*Phe259*). This phenylalanine appears to provide the additional negative potential required to stabilise the positive charge on the dammarenyl cation and allow expansion of the 5-membered D-ring to form the baccharenyl cation (Figure 2.15). The expansion of the D-ring transfers the positive charge to C_{18} (β -amyirin numbering – Figure 2.12), which is also stabilised by *Phe259* and *Phe725*. As mentioned previously, the plant OSC multiple sequence alignment shows that all sterol OSCs have a histidine residue at the *His232* position whereas all triterpene OSCs have either a phenylalanine (monocots) or a tyrosine (dicots). This provides a clear residue difference between sterol and triterpene enzymes at the exact cyclisation stage that forms their differing products – which indicates that *Phe259* may be the key residue required for producing pentacyclic products. Mutagenesis studies on β -

amyrin synthase from *Panax ginseng* have provided further support for this theory, as mutation of this residue to a histidine, as seen in sterol OSCs, resulted in the formation of tetracyclic dammarene-like products (176).

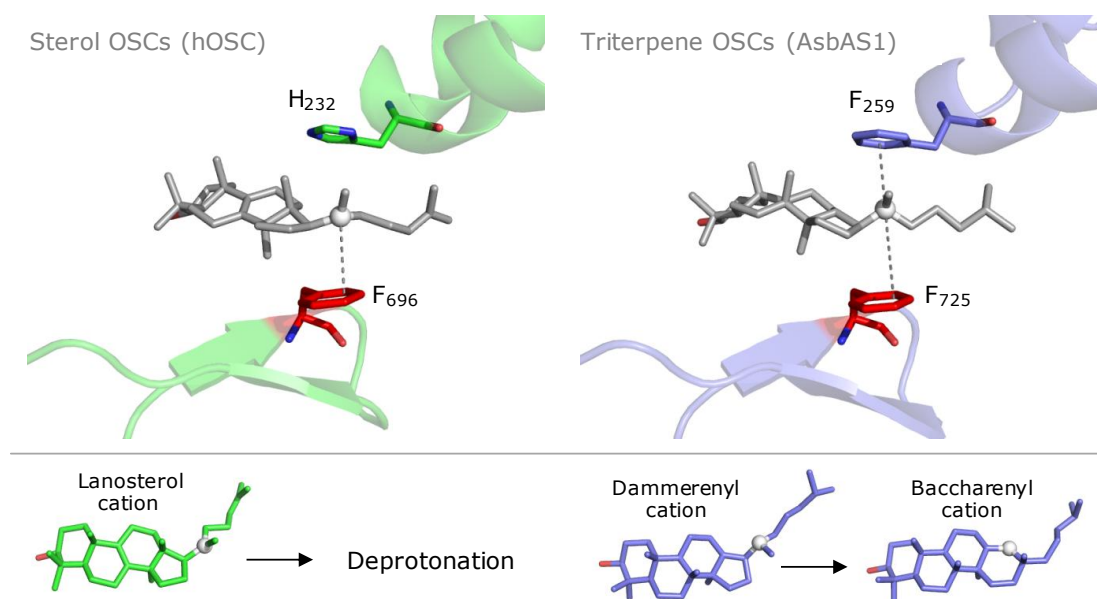


Figure 2.15: D-ring stabilisation in hOSC and AsbAS1

Diagram shows differences in D-ring stabilisation between hOSC (left) and AsbAS1 (right). The reaction intermediate is shown in grey and the C₂₀ cation is shown as a white sphere. In hOSC (and all sterol OSCs), Phe696 and His232 are not enough to stabilise the charge at C₂₀ resulting in cyclisation termination and deprotonation. In AsbAS1 (and other triterpene OSCs), His232 is replaced by an aromatic residue (Phe259), which together with Phe725 is able to stabilise the C₂₀ charge and allow cyclisation to continue by expansion of the D-ring to form the baccharenyl cation. Residues involved in stabilisation of the D-ring are shown in green (hOSC) and purple (AsbAS1), and conserved residues are shown in red.

2.3.3.5 E-ring formation

E-ring formation only occurs in triterpene OSCs, as sterol OSCs cease cyclisation at the D-ring due to lack of cation stabilisation. Formation of the 5-membered E-ring results in a shift of positive charge from C₂₀ (lanosterol numbering) to C₂₀ (β -amyrin numbering – Figure 2.12), forming the lupenyl cation. Stabilisation of the positive charge allows E-ring expansion to form the oleanyl cation. Mutagenesis and chimeric studies between lupeol synthase from *Olea europaea* (OEW) and β -amyrin synthase from *Panax ginseng* (PNY) identified a single residue, *Trp257*, that appeared to be responsible for stabilising the expanded E-ring in triterpene OSCs (176). If the E-ring expansion is not stabilised then cyclisation is halted at the lupenyl cation which is deprotonated to form lupeol. OEW has a leucine residue at the equivalent *Trp257* position, and mutation of this residue to tryptophan resulted in production of β -amyrin.

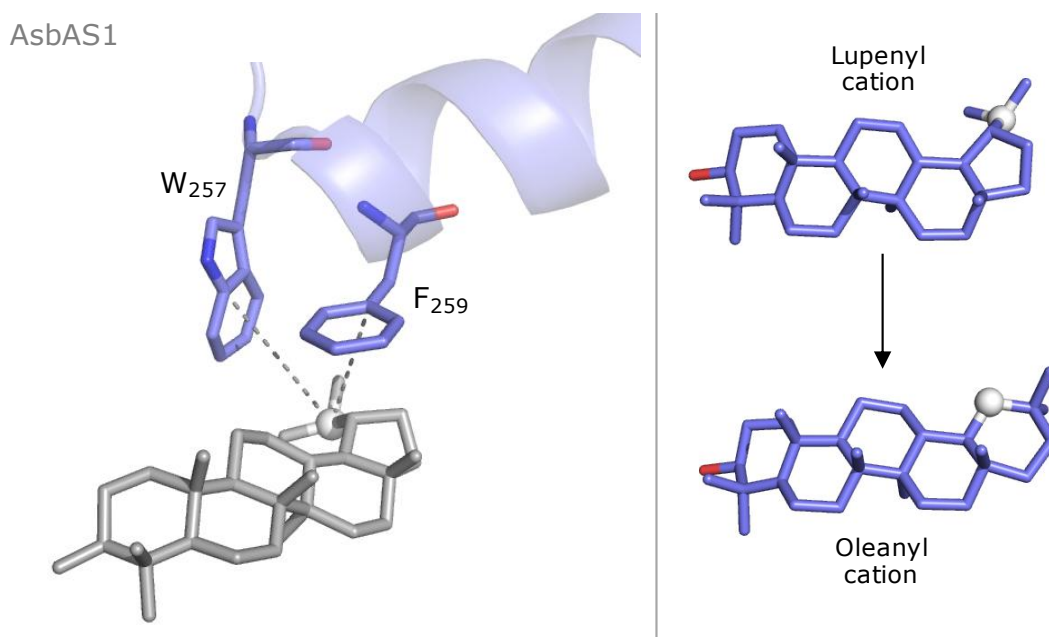


Figure 2.16: E-ring stabilisation in AsbAS1

Diagram shows how E-ring is stabilised in AsbAS1. The lupenyl cation intermediate is shown in grey and the C₂₀ cation is shown as a white sphere. The positive charge is stabilised by Trp257 and Phe259 which enables expansion of the E-ring to form the oleanyl cation. Residues involved in stabilisation are shown in purple.

Likewise, mutation of tryptophan in PNY to leucine resulted in production of lupeol (176, 214). Although mutagenesis studies of triterpene OSCs have been carried out (176, 214), AsbAS1 is the first triterpene OSC to be structurally modelled. This provides a unique tool to study the residues responsible for E-ring stabilisation in the active site.

Analysis of E-ring stabilisation in the AsbAS1 model was carried out with the pentacyclic lupenyl and oleanyl cations docked into the active site. The model suggested that the positive charge at C₂₀ could be stabilised through cation- π interactions with *Phe259* and *Trp257*, which would allow expansion of the 5-membered E-ring forming the oleanyl cation (Figure 2.16). The results observed for the mutagenesis of *Trp257* in PNY can be clearly explained by the AsbAS1 model. Mutation of the tryptophan to leucine results in the C₂₀ cation being stabilised solely by *Phe259* which does not provide sufficient stabilisation to allow E-ring expansion, therefore lupeol is formed. Consequently mutation of the equivalent leucine in OEW to tryptophan provides additional stabilisation to the C₂₀ cation allowing E-ring expansion and hence β -amyrin formation. The plant OSC multiple sequence alignment shows that all lupeol synthases have a leucine residue at the Trp257 position whereas the β -amyrin synthases have a

tryptophan residue. Together these results provide strong evidence that *Trp257* is required for E-ring expansion in triterpene OSCs.

2.3.3.6 Skeletal rearrangement and deprotonation

After cyclisation is complete, the intermediate undergoes skeletal rearrangement to form the product cation, which is deprotonated to yield the final product. Skeletal rearrangement causes a shift of the positive charge from a region of low electron density to an area of high electron density by 1,2 shifts of methyl and hydride groups. Migration is thought to be induced by the steric difficulty associated with the chair-boat/chair-chair folds (165, 213). After 2,3-oxidosqualene cyclisation by hOSC, the positive charge resides on C₂₀ (lanosterol numbering) creating the protosteryl cation. Skeletal rearrangement shifts the positive charge from C₂₀ to C_{8/9} (lanosterol numbering) which forms the lanosterol cation (144). Deprotonation of the lanosterol cation completes the cyclisation reaction. In hOSC, His232 is the only catalytic base present in the vicinity of the lanosterol cation and forms a hydrogen bond to Tyr503. It is presumed that this activated tyrosine accepts the proton from C_{8/9} (lanosterol numbering) to form the lanosterol product (144, 164-165).

Analysis of skeletal rearrangement in the AsbAS1 model was carried out with the pentacyclic β -amyrin cation docked into the active site. In AsbAS1 the

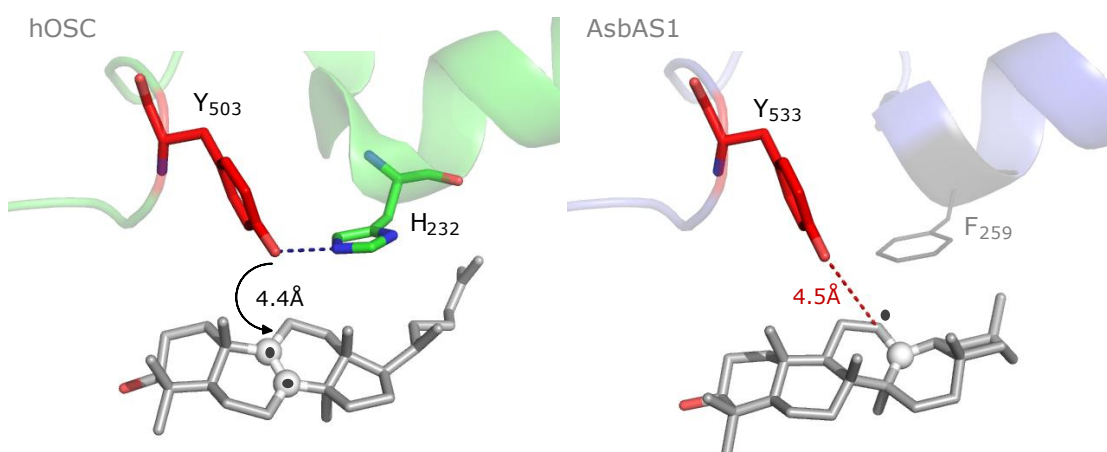


Figure 2.17: Deprotonation in hOSC and AsbAS1

Diagram shows how deprotonation occurs in hOSC (left) and a possible candidate for deprotonation in AsbAS1 (right). The product cation is shown in grey, the location of the final cation is shown by a white sphere (N.B. lanosterol cation can reside on C₈ or C₉) and the location of the proton that is removed is shown as a dark grey dot. In hOSC, His232 hydrogen bonds to Tyr503 which acts as the final proton acceptor. AsbAS1 lacks a suitable catalytic base, but may use a water molecule activated by Tyr533 as a final proton acceptor. Residues involved in deprotonation are shown in green (hOSC) and conserved residues are shown in red. Hydrogen bonds are represented by a dotted blue line, and distances are shown as a dotted red line.

positive charge shifts from C₁₉ to C₁₃ (β -amyrin numbering) to form the β -amyrin cation (163). Deprotonation occurs at C₁₂ (β -amyrin numbering), but the catalytic base is harder to determine as there is no equivalent to the His232 residue, as in hOSC, to activate the equivalent tyrosine residue, *Tyr533*. In the search for a candidate base, the location of each basic residue in the active site vicinity was determined to identify a residue close enough to C₁₂ (β -amyrin numbering) to act as a catalytic base. However the only residue in close contact with C₁₂ (β -amyrin numbering) is the conserved *Tyr533* at a distance of 4.5Å (Figure 2.17). With no other candidate present, it is postulated that in AsbAS1, the hydroxyl group of *Tyr533* may form a hydrogen bond with a water molecule to facilitate deprotonation at C₁₂ (β -amyrin numbering). The plant OSC multiple sequence alignment shows that although *Tyr533* is conserved in monocot triterpene OSCs and in all sterol OSCS, dicot triterpene OSCs have a tryptophan residue at the corresponding position, which leaves the mechanism for deprotonation in dicot OSCs unclear.

Chapter 3 - Genetic characterisation of *sad1* mutants

3.1 Introduction

The *sad* mutant collection has been an invaluable tool in the elucidation of the genes involved in the avenacin pathway. Using the mutagen sodium azide, 92 mutants have been identified in total using a reduced root fluorescence screen that exploited the fluorescent property of the major end product of the pathway, avenacin A-1. This comprehensive set of metabolic pathway mutants is the first to be isolated through a forward screen for loss of function. These mutants have defined at least seven loci involved in avenacin synthesis and characterisation of these mutants has provided insights into each individual step of the pathway.

3.1.1 Characterisation of the original *sad1* mutants

Of the ten *sad* mutants identified in the first reduced root fluorescence screen, two, #109 and #610, had single point mutations at a locus defined as *Sad1* which was later identified as the β -amyrin synthase gene. Both of these mutants lacked β -amyrin and had no detectable β -amyrin synthase activity. Sequencing of the *Sad1* gene in each of these mutants revealed single point mutations that resulted in predicted premature termination of translation codons. Further characterisation revealed that *Sad1* transcript levels were substantially reduced in both mutants, possibly as a result of nonsense-mediated mRNA decay.

3.1.2 Identification of additional *sad1* mutants from the oat mutant collection

Characterisation of mutants in the extended mutant collection has identified six additional *sad2* mutants and one additional *sad7* mutant. To identify additional *sad1* mutants in the extended mutant collection, methanolic root extracts of 6-day old oat seedlings were analysed by TLC to identify mutants that accumulated 2,3-oxidosqualene, as seen for the two original *sad1* mutants. The methanolic extracts were dissolved in 100 μ l chloroform:methanol (7:3 v/v) and analysed by TLC (hexane:acetone 80:20 v/v) which identified 15 candidate *sad1* mutants that lacked avenacin and accumulated 2,3-oxidosqualene. Root extracts were also dissolved into hexane for GC-MS analysis which confirmed the accumulation of 2,3-oxidosqualene in the original two *Sad1* mutants and the 15 candidate *sad1* mutants thereby confirming these mutants as new *sad1* mutants (215).

3.1.3 Aims

Two *sad1* mutants and 15 candidate *sad1* mutants have been identified from the extended oat mutant collection and partially characterised. The aim of this part of the project was to characterise the *sad1* mutant collection at a gene, transcript and protein level. The gene and transcript analysis of the *sad1* gene in each of the 17 mutants is discussed in this chapter and the protein analysis will be discussed in Chapter 4.

Characterisation of the *sad1* mutants at a genetic level will involve sequencing of the *Sad1* gene in each of the new *sad1* mutants to determine the nature of the mutations. The *sad1* mutants will then be characterised at the transcript level by reverse-transcription PCR and Northern blotting.

3.2 Materials and Methods

3.2.1 Isolation of genomic DNA from *sad1* mutants

Seeds of *Avena strigosa* S75 (wild-type) and *sad1* mutants were surface sterilised in 0.05% (v/v) sodium hypochlorite solution for 10 minutes and incubated on Whatman No.1 9 cm filter circles (Sartorius Stedim UK Ltd) at 4°C in the dark for 48 hours. For germination, seeds were transferred to 1% (w/v) agar plates and incubated for 6 days at 22°C with 16 hours light and 8 hours dark. Genomic DNA (gDNA) was extracted from 4 cm of leaf material using the DNeasy Plant Mini Kit (Qiagen) and analysed by 1% (w/v) TAE agarose gel electrophoresis to confirm recovery of gDNA.

3.2.2 PCR amplification of *Sad1* genomic DNA

The *Sad1* gene was amplified using the gDNA of each of the *sad1* mutants as a template. The gene was amplified in four segments and the following primers were used: Segment 1, AmyStaF1 = ACGAGTGCTTGTTTTCTCGTA and Amy610R = CATAACGACAACCATATTTTTCCCA; Segment 2, Amy610F = GTCGCTACATTTACAATCAACAGGCAT and Amy109R = TCTATACCAACCTGTGCCTTCATTCC; Segment 3, Amy109F = TATCCATTATGACGACGAATCAACC and Amy23R = TCGGAACATTGTCGTGGACTA; Segment 4, Amy12F = CTCAACCCTTCTGAGAGTTTT and AmyEndR1 = TGATGACATCGGTAGGAA. Gradient PCR was carried out using Phusion™ High Fidelity DNA polymerase (Finnzymes) for 35 cycles (program: 98°C for 10 seconds, 40-65°C for 1 min, 70°C for 1.5 mins (segments 2 and 3) or 2 mins (segments 1 and 4), and final extension at 72°C for 10 min). Fractions which contained a single band at the correct size were confirmed by 1% TAE

agarose gel electrophoresis with ethidium bromide staining. These fractions were pooled and purified using a Wizard PCR Preps Kit (Promega).

3.2.3 Sequencing of *sad1* genomic DNA segments

Sequencing reactions were carried out using BigDye® Terminator v3.1 Ready Reaction Mix (Applied Biosystems) using the program: 92°C for 40 seconds, 52°C for 15 seconds, 60°C for 4 minutes followed by 25 cycles of 95°C for 10 seconds, 52°C for 15 seconds and 60°C for 4 minutes. Purified DNA segments from each of the *sad1* mutants were sequenced using a variety of primers along the length of the *Sad1* gDNA (Table 3.1). Sequencing reaction analysis was carried out by the Genome Laboratory at the John Innes Centre, Norwich. Sequencing data was viewed using Chromas (Technelysium Pty Ltd) and the resulting sequence was aligned to the gDNA sequence from wild-type *Sad1* using BioEdit (IBS Therapeutics) (216). Each mutant was sequenced to at least 2-fold coverage using the primers listed in Table 3.1.

Table 3.1: Primers used for *sad1* mutant sequencing

Name and sequence of primers used to sequence each segment of the *Sad1* gene are shown. Primers suffixed with F or F1 are the primers for the sense DNA strand and primers suffixed with R or R1 are the primers for the antisense strand. Capital letters denote exon sequence whereas small letters denote intron sequence. The exact location of each primer on the gDNA is shown in Appendix 6.

Primer name	Primer sequence	Primer name	Primer sequence
Segment 1		Segment 3	
AmyStaF1	ACGAGTGCTTGTTTTCTCGTA	Amy109F	TATCCATTATGACGACGAATCAACC
Amy10F	gtatggatttcgtaccgtaaat	Amy05F	TGGATGTCATAGCTGGGA
Amy16F	CAGTGGGATTCTCTCATTATGC	Amy06F	TATCCGCTGACCTTGTTG
Amy18F	gtggctcatcacattgatcaca	Amy17R	AGCCTTTTGATCTGTGGCGATA
Amy26R	ATCGTCTGCTGTAGAGAGGA	Amy23R	TAGTCCACGACAATGTTCCGA
Amy25R	agttcaaccaagattttagacaac	Segment 4	
Amy610R	catacCGACAACCATATTTTTCCCC	Amy12F	CTCAACCCCTTCTGAGAGTTT
Segment 2		Amy08F	GTGGATGGGGTGAAGACT
Amy610F	GTCGCTACATTTACAATCAACAGgcat	Amy19F	TGGGCAATGTTGGCTTTAATTT
Amy11F	CTGCGACAAGACCTCTATA	Amy22R	tggacacggtcatcaaaactg
Amy24R	GTCCCAATTAATGTTGCAGTAAG	AmyEndR1	TCAGCTCTTAATCGCAAGAAGT
Amy109R	tctataccaacCTGTGCCTTCATTCC		

3.2.4 Diversity Array Technology (DART) analysis of *sad1* mutants

Genomic DNA was extracted from a single leaf of a 6 day old oat seedling as described in section 3.2.1, and analysed by 1% (w/v) TAE agarose gel electrophoresis to confirm recovery of gDNA. Two independent genomic DNA samples from each mutant were sent for Diversity Array Technology (DART) analysis (Diversity Arrays Technology Pty Ltd).

3.2.5 Transcript analysis of *sad1* mutants

Total RNA was extracted from the terminal 0.5 cm roots of 3 day old oat seedlings using TRI-REAGENT (Sigma) followed by DNase digestion at 37°C for 10 minutes by adding 2 µl of DNase I (Roche) to 50 µl RNA. DNase was removed by phenol/chloroform extraction and ethanol precipitation and the RNA concentration was determined spectrophotometrically using a NanoDrop ND-1000 (Thermo Scientific).

For RT-PCR, cDNA was synthesised using 1 µg of DNase treated RNA which was primed with 0.5 µg Oligo(dT)₁₅ primers and 0.5 mM each of dATP, dCTP, dGTP, dTTP (Invitrogen) in a final volume of 13 µl at 65°C for 5 minutes then immediately chilled on ice for 1 minute. First strand cDNA synthesis was carried out by adding the following components to the denatured RNA to yield a final concentration of 1 x First Strand Buffer (Invitrogen), 5 mM dithiothreitol (Invitrogen), 40 U RNaseOUT™ (Invitrogen) and 200 U of Superscript III Reverse Transcriptase (Invitrogen). The reverse transcription was carried out at 50°C for 1 hour and the enzyme was inactivated by incubation at 70°C for 50 minutes. cDNA was stored at -20°C and cDNA was diluted 5-fold for use in RT-PCR reactions. PCR amplification of the *Sad1* and GAPDH was carried out using the following primers: *Sad1*-5' = ATGTGGAGGCTAACAATAGG, *Sad1*-3' = TATCTCATG ACGATGTTCCG, *AsGAPDH*-5' = CGCCATGGGCAAGATTAAGATCGGAATCAACG and *AsGAPDH*-3' = GTGGCAGTGATGGCATGAACAGTKGTCATCA. PCR amplification of splice site regions in the four splicing mutants was carried out using the following primers: 110, *Amy08F* = GTGGATGGGGTGAAGACT and *AmyEndR1* = TGATGACAT CGGTAGGAA; 225 and 589, *Amy11F* = CTGCGACAAGACCTCTATA and *Amy17R* = TATCGCCACAGATCAAAAAGGCT; and 1001, *Amy05F* = TGGATGTCATAGCTGGGA and *Amy23R* = TCGGAACATTGTCGTGGACTA. PCR was carried out using Taq DNA polymerase (HT Biotechnology) using the program: 95°C for 2 minutes followed by 29 cycles of 95°C for 30 seconds, 55°C for 40 seconds, 72°C for 40 seconds and a final extension at 72°C for 3 minutes. Products were then run on a 1% TAE agarose gel and visualised with ethidium bromide staining.

For Northern blotting, 10 µg of total RNA was run on a 1.2% agarose/0.25 M formaldehyde gel at 60 V for 4 hours and RNA was visualised with ethidium bromide staining. The gel was soaked in 10x SSC for 20 minutes and transferred to a Hybond-N+ nylon membrane (Amersham) in 10x SSC overnight. After transfer was complete, RNA was immobilised on the membrane by baking at 80°C for 3 hours. cDNA probes for the 5' and 3' end of the *Sad1* gene were generated using the following primers: 5' probe, *Sad1*-5'-F = ATGTGGAGGCTAACAATAGG and *Sad1*-5'-R = TATCTCATGACGATGTTCCG; 3' probe, *Sad1*-3'-F = CGA

ATCCACGATATCGAAGAG and Sad1-3'-R = TCAGCTCTTAATCGCAAGAAG. Each probe was purified using the QIAquick PCR purification Kit (Qiagen) and eluted in TE buffer pH 8.0. Probes were labelled with ^{32}P -dCTP using the Rediprime II Random Prime Labelling Kit (Amersham). Hybridisations were carried out overnight at 65°C in 10 ml Church Buffer, 0.1 mg/ml salmon sperm DNA (Sigma) and 50 μl ^{32}P -dCTP labelled probe. Membrane was washed at 65°C in 0.1% (w/v) SDS with decreasing concentrations of SSC until radioactivity was localised to a specific region of the membrane. The membrane was exposed to a BAS-IIIS imaging plate (Fuji) overnight and imaged using a Typhoon 9200 Variable Mode Imager (Amersham). Blots were stripped by incubation with two washes at 85°C in 5 mM sodium phosphate buffer pH 7.2 and 0.1% SDS (w/v) for 10 minutes and then exposed to a phosphoimaging screen overnight to check the quality of the stripping.

3.3 Results and Discussion

3.3.1 Sequence analysis of *sad1* mutants

Fifteen potential *sad1* mutants (in addition to the two previously characterised *sad1* mutants, #109 and #610 (49)) were identified from a TLC screen of the extended collection of 92 reduced fluorescence oat mutants which looked for mutants which accumulated 2,3-oxidosqualene, the β -amyrin synthase substrate (215). These mutants were isolated by screening for reduced root fluorescence (avenacin deficiency) and shown to have undergone loss of function mutations in the *Sad1* gene via complementation tests and/or TLC analysis (18, 215). To determine the nature of the mutations in the 15 new *sad1* mutants, the *Sad1* gene in each of the 15 new *sad1* mutants, as well as the two previously characterised *sad1* mutants, was sequenced. The location of these mutations has the potential to provide valuable clues as to amino acid residues and regions of the gene that are important for function.

3.3.1.1 Gradient PCR amplification of *sad1* mutant gDNA

Sad1 is a large gene containing 7340 base pairs, 18 exons and 17 introns, therefore in order to amplify it effectively it was split into four overlapping segments of between 1.7 and 2.3kb (Figure 3.1). PCR amplification for each segment using the optimal annealing temperatures calculated for each set of primer pairs was inconsistent as not all gDNA samples produced a product at the desired size. Therefore the protocol was modified and each segment was

amplified using gradient PCR and products were analysed by 1% TAE agarose gel electrophoresis to look for those annealing temperatures which produced a single band at the correct size. Products from these temperatures were then pooled, purified and used in the sequencing reactions.

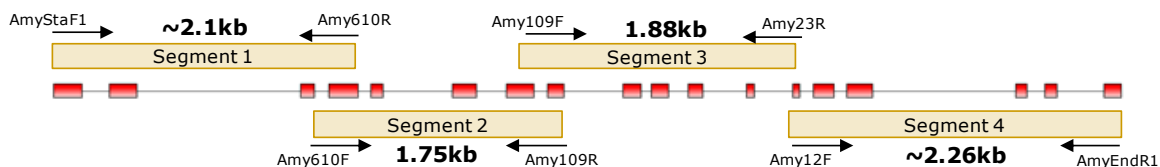


Figure 3.1: Position of segments designed for amplification of *Sad1* genomic DNA

The *Sad1* gene contains 18 exons and is 7.34kb in length. Exons are represented as red boxes and introns as lines connecting the boxes. The positions of the four segments designed for amplification of the gene are shown in yellow with the forward and reverse primers labelled at the beginning and the end of the segment respectively.

3.3.1.2 Genomic DNA sequence analysis of *sad1* mutants

The *sad* mutants were generated previously by mutagenising *Avena strigosa* S75 with sodium azide (18, 217). Sodium azide has been widely used as a mutagen in many plant and microbial species due to its ability to produce single base substitutions that result in single point mutations. Unlike other mutagens, e.g. ionizing radiation and ethyl methanesulfonate (EMS), chromosome abnormalities are not observed, which is a huge advantage as mutagenesis is more specific and efficient (218-221). Therefore when sequencing the *sad1* mutants, it was expected that single point mutations would be found in the coding regions of the genomic DNA. Each segment was sequenced with overlapping sense and antisense primers to give at least 2-fold sequencing coverage to each segment. Primers used for sequencing are shown in Table 3.1.

Sequencing chromatogram data were analysed using Chromas and yielded very clear peaks for identification of each individual base, which showed that the consensus sequence obtained was very reliable. The FASTA sequence from each sequencing run was then exported and aligned to the *Sad1* genomic DNA sequence in order to find the location of the point mutation. The use of high fidelity Taq DNA polymerase and at least 2-fold coverage of the *Sad1* genomic DNA sequence identified 5 base changes in introns 1, 8, 15 and 17 in addition to the single point mutation found in the exon. These 5 intron base changes were seen in all mutants and the re-sequenced wild-type *Sad1* gene, and as sodium azide mutagenesis does not usually cause point mutations in non-coding DNA, this indicated that the original gDNA sequence was incorrect. The *Sad1* gDNA sequence was therefore updated.

Sequence analysis of all 15 new *sad1* mutants showed that in each case all 15 mutants had single point mutations in the coding sequence of the *Sad1* gene. The mutations observed fell into three different categories: predicted premature termination of translation, splicing errors and amino acid substitutions (Table 3.2).

Table 3.2: Sequence analysis of *sad1* mutants

Mutant	Mutation event	Predicted amino acid change	Predicted Mutation type
A1	G ¹⁹¹² →A	Tyr ¹⁶⁵ →STOP ^a	Premature termination of translation
B1	G ¹⁹¹² →A	Tyr ¹⁶⁵ →STOP ^a	
109	G ³⁴¹⁷ →A	Tyr ³⁸⁰ →STOP	
610	G ¹⁹¹² →A	Tyr ¹⁶⁵ →STOP ^a	
1146	G ⁴¹⁶⁹ →A	Tyr ⁴⁷¹ →STOP	
1293	G ³⁹ →A	Tyr ¹³ →STOP	
110	G ⁶⁶⁸⁹ →A	-	Splicing error
225	G ³³⁰² →A	-	
589	G ³⁹¹⁴ →A	-	
1001	G ⁴³⁶⁵ →A	-	
297	G ³⁹³⁹ →A	Glu ⁴¹⁹ →Lys	Amino acid substitution
358	G ⁵²³⁴ →A	Cys ⁵⁶³ →Tyr	
384	C ⁷²⁴⁹ →T	Ser ⁷²⁸ →Phe ^b	
532	G ⁵⁴⁹ →A	Gly ¹²¹ →Glu	
599	G ²⁸⁰⁹ →A	Gly ²⁷⁷ →Glu	
1023	C ⁷²⁴⁹ →T	Ser ⁷²⁸ →Phe ^b	
1217	G ²⁰²⁵ →A	Gly ²⁰³ →Glu	

^a Identical mutations (G¹⁹¹² → A)

^b Identical mutations (G⁷²⁴⁹ → A)

3.3.1.2.1 *Premature termination of translation mutants*

The first group of mutants (#A1, #B1, #109, #610, #1146 and #1293) all had single point mutations which substituted a guanine for an adenine and changed the coding sequence from TGG to TGA (Figure 3.2A). These nonsense mutations resulted in predicted premature termination of translation codons (PTCs) being introduced into the coding sequence. Nonsense mutations have been shown to cause nonsense-mediated mRNA decay (NMD) in order to prevent the cell from producing C-terminally truncated proteins that have deleterious effects (222-225). If this was the case for the STOP codon mutants then the mRNA transcript would be degraded, AsbAS1 protein would not be formed and the avenacin pathway would be prematurely terminated, which would explain why these mutants have severely reduced root fluorescence.

3.3.1.2.2 *Splicing error mutants*

The second group of mutants (#110, #225, #589 and #1001) had single point mutations at intron-exon boundaries and so were likely to be affected in

pre-mRNA splicing. Mutants #110 and #225 had mutations at the 5' splice site of introns 16 and 7 respectively (Figure 3.2B). The consensus sequence at the 5' splice site in plants is AG/GTAAGT, where the / symbol represents the exon-intron boundary, and this where the U1 small nuclear ribonucleoprotein particle (snRNP) binds and initiates the formation of the major spliceosome. This consensus sequence contains the dinucleotide, GT, found at the start of the intron, which is absolutely conserved in all organisms (226-227). Mutant #110 had a mutation of the guanine preceding the conserved dinucleotide and mutant #225 had a mutation of the guanine in the conserved dinucleotide. Both of these mutations would most likely be affected in the binding of the U1 snRNP and may prevent correct splicing of the mRNA at this location. Mutants #589 and #1001 had mutations at the 3' splice site of introns 8 and 10 respectively. The consensus sequence at the 3' splice site in plants is TGCAG/G, where the / symbol represents the intron-exon boundary, and also contains the dinucleotide AG at the end of the intron which is absolutely conserved in all organisms. Both of these mutants had a mutation of the guanine in the conserved dinucleotide which is involved in the binding of the U5 snRNP and are therefore likely to have splicing errors at this location (226-227). Mutations that affect splice sites may cause an exon to be deleted, an intron to be included which may introduce a PTC, or cause splicing at an aberrant site (228).

3.3.1.2.3 Amino acid substitution mutants

The final group of mutants (#297, #358, #384, #532, #599, #1023 and #1217) had single point mutations in the coding sequence of the *sad1* gene which resulted in amino acid substitutions (Figure 3.2C). Unlike the previous two mutant groups which are likely to be affected at the post-transcriptional or translational stage, substitution of an amino acid would not be expected to have any effect on transcription or translation, therefore this group of mutants was most likely to be affected at the protein level. In each case the resulting substitution changes the size as well as the charge or polarity of the amino acid at that position. Amino acid substitutions may affect protein folding or stability leading to protein degradation or should fully folded protein be produced, substitutions may impede the catalytic mechanism. If the latter occurs, this may provide an insight into amino acid residues or regions of the protein that are critical for function.

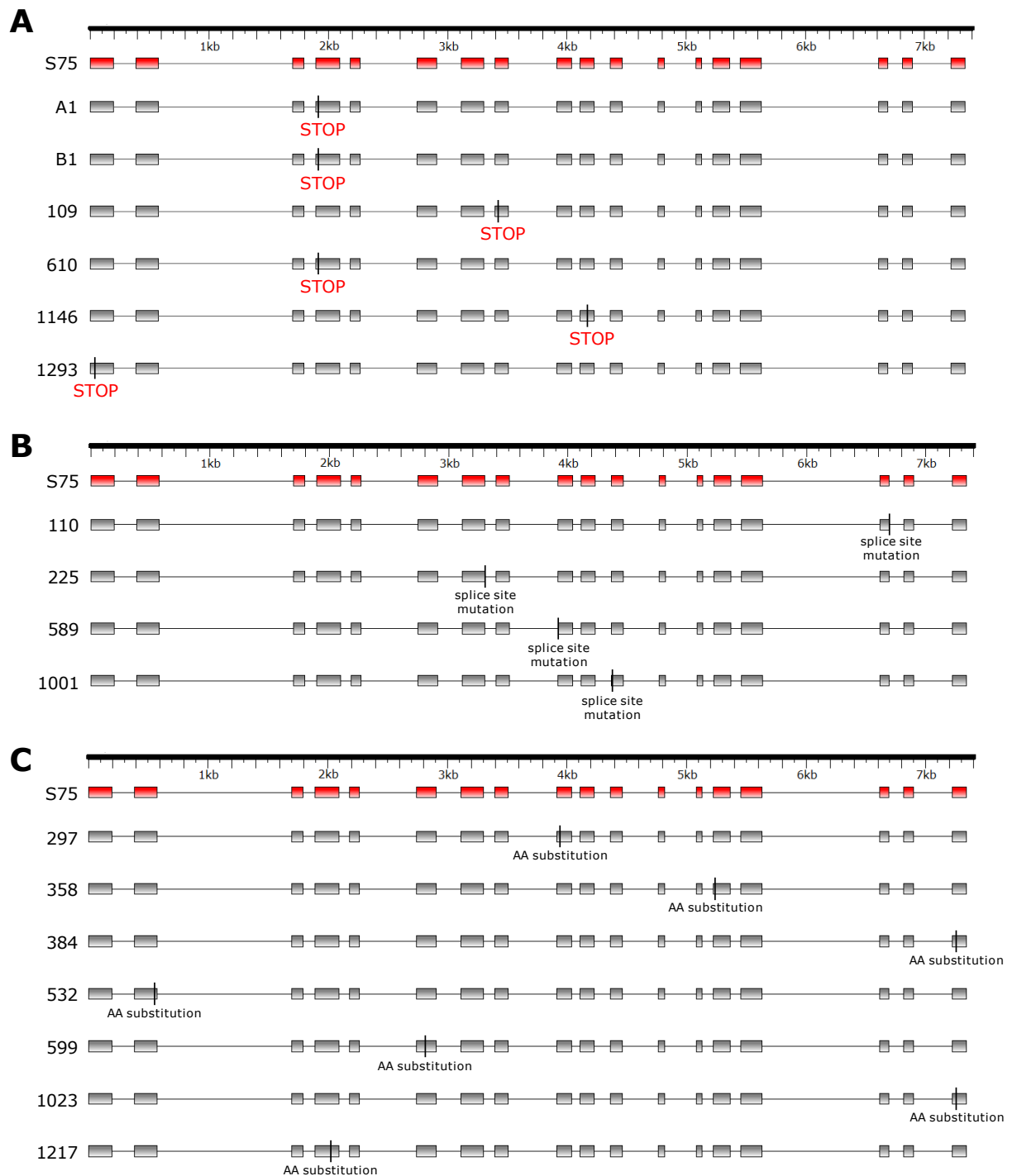


Figure 3.2: Location of single point mutations in *sad1* mutants

Schematic of the intron-exon structure of the *Sad1* gene for each of the *sad1* mutants showing the location of the different point mutations. Exons are represented by boxes and introns by lines. The wild-type S75 is shown in red and the mutants are shown in grey. The mutants are organised according to the type of mutation they contain (see Table 3.2) and the location of the mutation within the gene is shown by a vertical line. **A.** Premature termination of translation mutants. **B.** Splicing error mutants. **C.** Amino acid substitution mutants.

3.3.1.3 DArT analysis of *sad1* mutants

Sequencing of the 17 *sad1* mutants had shown that all had single point mutations in the *Sad1* gene, however closer inspection of the mutation locations revealed that there were two sets of mutants that had identical mutations. Mutants #A1, #B1 and #610 all had a mutation in base 1912 (guanine to adenine mutation) and mutants #384 and #1023 had a mutation in base 7249 (guanine to adenine) (Table 3.2). The location of each mutation was verified independently from two different gDNA samples so sampling error was ruled out a cause of the identical mutants. The mutants were generated by mutagenising oat seed with sodium azide and one bag of M₂ seed was collected from individual M₁ plants and sequentially numbered to avoid the isolation of siblings. However during seed collection and bagging there may have been cross-contamination between mutant populations which would account for the results that were obtained. It was therefore important to ascertain whether the mutants were genetically identical or independent.

As sodium azide mutagenesis results in single point mutations, it can be assumed that in addition to the single point mutation found in the *Sad1* gene of each mutant, there will be other single point mutations located throughout the genome. If the mutants are genetically independent then these additional mutations will be found in different locations, but if the mutants are genetically identical then the mutation pattern will be identical. Diversity Array Technology (DArT) is a high-throughput genotyping system that can be used to differentiate between the 17 *sad1* mutants. It is a solid-state, open-platform method for analysing DNA polymorphisms in a number of different organisms without the need for DNA sequence information. Genomic representations of the DNA sample are generated and hybridised to an array containing a library specific to the target organism. DNA polymorphisms are detected by comparing hybridisation patterns of genomic representations from different genotypes (229-230). For DArT analysis of the 17 *sad1* mutants, two independent gDNA samples were extracted from the leaf of a 6 day old oat seedling for each mutant and the wild-type S75. The subsequent DArT analysis was carried out at the DArT laboratory in Canberra, Australia.

The first step involves creating the array containing the library of oat specific markers. Genomic representations from a pool of individuals that cover the genetic diversity of the target species are generated using complexity reduction which involves restriction enzyme digestion with a rare and frequent cutter, adapter ligation and amplification. The genomic representations are

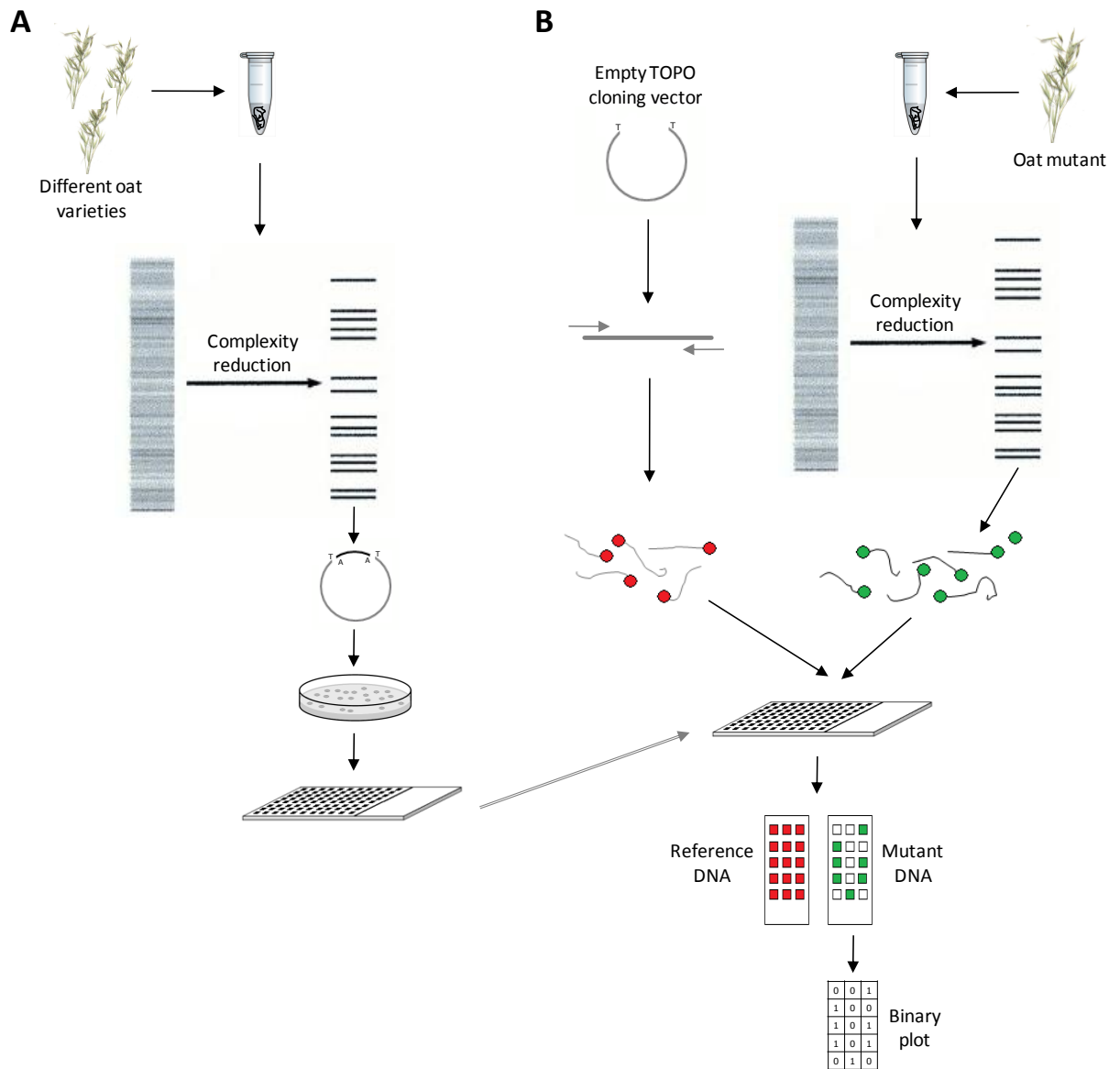


Figure 3.3: Schematic representation of DArT analysis of *sad1* mutants

A. Generation of a library of oat specific markers. Genomic DNA (gDNA) from a variety of oat species is pooled and the complexity of the sample is reduced by restriction enzyme digestion, adapter ligation and PCR amplification using primers with selective overhangs. The fragments are cloned into *E. coli* and the inserts are amplified using vector-specific inserts, purified and arrayed onto a solid support. **B.** DArT analysis of *sad1* mutants. gDNA from each *sad1* mutant undergoes complexity reduction and cloning as in A and is then labelled with a fluorescent dye. To allow for quantification of the signal, fragments of the cloning vector, which are common to all elements of the array are amplified and labelled with a different fluorescent dye (reference DNA). Both reference DNA and the mutant gDNA representation are hybridised to the library array created in A. The ratio of signal intensity between the reference and mutant DNA is measured for each position on the array and is displayed in a binary format (1 – marker present; 0 – marker absent). Each mutant gDNA representation can be hybridised to the array and polymorphic spots can be identified by comparing the binary plots between all the mutants. Figure adapted from Jaccoud *et al.* (2001) (229)

cloned into *E. coli* and arrayed onto a solid support which can then be hybridised to labelled representations from specific genotypes (Figure 3.3A). Then, a genomic representation of each of the oat mutants is generated and labelled with a fluorescent dye. This is hybridised to a library array containing the genomic representations (markers) from the target species (which in this case is oat). The arrays are also hybridised with fluorescently labelled reference DNA which binds to a fragment of the cloning vector common to all elements of the array. This allows the signal from each marker to be quantified and the relative hybridisation intensity is measured for each marker on the array. Each position is scored based on whether the marker is present (1) or absent (0) which gives each genotype a unique binary pattern (Figure 3.3B) (229-230). It was expected that the binary pattern would be identical between the two replicates from each oat mutant whereas those mutants with differing point mutations would have different binary patterns. Of the two groups of mutants with identical mutations, if any of the mutants were genetically identical then their binary patterns would be identical. If the binary patterns are different then the mutants are genetically independent but contain the same *Sad1* polymorphism.

The DArT library array used for analysis of the *sad1* mutants contained 15,000 markers, of which only 35 markers were found to be polymorphic. This showed that there was no significant source of genetic variation between the samples which allowed the subtle changes due to mutagenesis to be identified. The DArT binary data for all the *sad1* mutants is shown in Appendix 7. The binary data for each mutant was analysed using a Hamming matrix which gives a numerical score of the similarity between each of the mutants. It is calculated by dividing the number of differences (excluding missing data) between a pair of samples by the total number of active comparisons. A dendrogram was generated from the Hamming matrix using statistiXL (231) which provided a visual representation of the similarity between the mutants (Figure 3.4). The furthest neighbour method was chosen for clustering as it calculates the greatest dissimilarity between clusters therefore produces tight clusters of similar cases. As two independent samples were analysed for each mutant, this provided a negative control as both samples are genetically identical so should display a distance of zero on the Hamming matrix. Eleven out of the 17 *sad1* mutants were found to have genetically identical sample pairs and are labelled in green in Figure 3.4. However not all of the mutants had identical sample pairs, five of the mutants (110, 297, 384, 532 and 1293) had both samples present in the same branch but were not found to be identical and only the remaining mutant (384) as

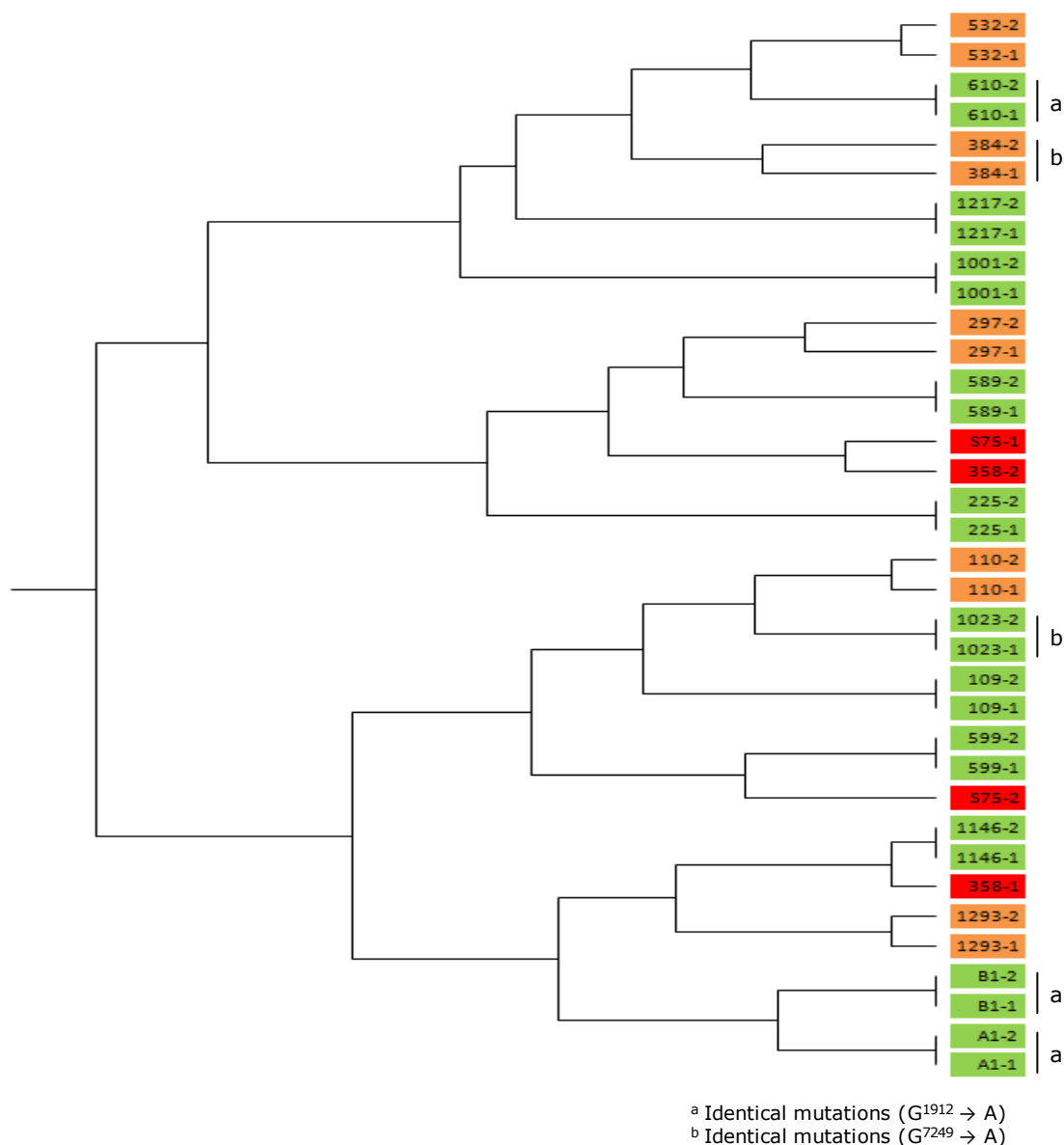


Figure 3.4: Distance dendrogram of Humming matrix data from *sad1* mutant DArT analysis

Furthest neighbour dendrogram based on hybridisation of 17 *sad1* mutant genotypes to 35 DArT markers. Two independent samples were analysed for each mutant and are labelled -1 and -2 respectively. Mutant numbers highlighted in green represent sample pairs with a genetic distance of 0. Mutant numbers highlighted in orange represent sample pairs that are closely related in the same branch but did not produce identical results. Finally mutant numbers highlighted in red represent mutant numbers whose paired sample was distantly related and is likely to be as a result of sampling error.

well as the wild-type S75 were found to be significantly different. These significant differences are most likely to be due to sampling errors or DNA contamination as the sample pairs have such large differences.

Nevertheless, in spite of the two anomalous results the data indicated that the majority of the mutants were genetically independent. Most sample pairs were shown to have identical (65%) or very similar (30%) genetic profiles so should any of the mutants with identical *sad1* polymorphisms been genetically

identical, it would have been clearly seen. For the two groups of identical mutants in question (A1, B1 and 610; 384 and 1023), #384 and #1023 were found in different branches therefore it was concluded that although this group contained the same *sad1* polymorphism, the mutants were genetically independent. Results for the other group showed #610 was genetically independent from #A1 and #B1 as it was found in a separate branch, however, #A1 and #B1 although not in the same branch were, genetically, the most similar. Of the DArT markers that were 100% reproducible #A1 and #B1 were 95% identical, but for DArT markers that were only 97% reproducible they were only 47% identical. These results indicate that #A1 and #B1 are highly similar but does not provide conclusive proof that these mutants are identical, so they will be treated as independent mutants in the subsequent analysis.

3.3.2 Transcript analysis of *sad1* mutants

Sequence analysis revealed that all of the 17 *sad1* mutants were independent (with the possible exception of #A1 and #B1) and had single point mutations in the *Sad1* gene which fell into three different categories: predicted premature termination of translation, splicing errors and amino acid substitutions. It was now of interest to study the level of *Sad1* transcript in each of the mutants in order to gain further information.

Previous analysis of the two original *sad1* mutants, #109 and #610, had revealed PTCs in the coding sequence and further characterisation revealed transcript levels were severely reduced in the roots of both mutants (49). On the basis of these results, it was expected that the other four predicted premature termination of translation mutants would all show the same transcript phenotype. It was also expected that the amino acid substitution mutants would all have transcript present as nonsynonymous mutations are not anticipated to affect transcript stability. Mutations at intron-exon boundaries can affect the mRNA transcript in different ways. If intron inclusion occurs in the splicing mutants then this may introduce a PTC into the transcript, resulting in NMD and hence severely reduced transcript levels. If exon exclusion occurs, this can also shift the reading frame of the transcript, which may also introduce a PTC. In order to determine how the intron-exon structure of the *sad1* mutants were affected, transcript levels were studied using RT-PCR and Northern blotting.

3.3.2.1 Transcript analysis of *sad1* mutants by RT-PCR

For RT-PCR analysis, RNA was extracted from the terminal 0.5 cm of root tips of 3 day old oat seedlings and 1µg was used for cDNA synthesis. The primers used for RT-PCR were designed complementary to the first 20 bases of the *Sad1* gene, starting from the ATG start codon. In order to standardise the levels of cDNA used for each RT-PCR reaction, the *Avena strigosa* glyceraldehyde 3-phosphate dehydrogenase (*AsGAPDH*) gene was used. The products of the RT-PCR reactions were run on a 1% TAE agarose gel and visualised using ethidium bromide staining.

3.3.2.1.1 Premature termination of translation mutants

As expected, when compared to the wild-type S75 all predicted premature termination of translation mutants showed a severe reduction in transcript levels (Figure 3.5A). The correlation between the presence of PTCs and decreased levels of mRNA abundance is thought to be caused by NMD (232-241). Studies in crop plants have found that mRNA decay occurs post-transcriptionally and may cause mRNA instability which leads to decay (236). The mechanism by which PTCs are identified in plants is as yet unknown, but is a different mechanism to that seen in mammals and yeast. It is thought that normal and premature codons are distinguished by measuring the spacing between the termination codon and elements present within the 3' untranslated region (3' UTR) of the gene. Unusually long 3' UTRs are thought to trigger NMD as well as the presence of introns in the 3' UTR (242-243). Once transcripts have been targeted for NMD they are degraded in a similar manner to general mRNA decay, but the mechanism used in plants has yet to be identified (243).

It is highly likely that the reduction in transcript levels in the premature termination of translation mutants is due to NMD. Without the *Sad1* transcript, these mutants cannot synthesise AsAS1, leaving the avenacin pathway halted at the first stage.

3.3.2.1.2 Splicing error mutants

RT-PCR of the four splicing mutants revealed that all had *Sad1* transcript present (Figure 3.5B). Errors at splice sites can cause intron inclusion or exon deletion, but as the primers used for RT-PCR were designed at the 5' end of the *Sad1* gene and did not span any of the regions affected by splicing errors, the results did not give any indication as to whether transcript size had been affected

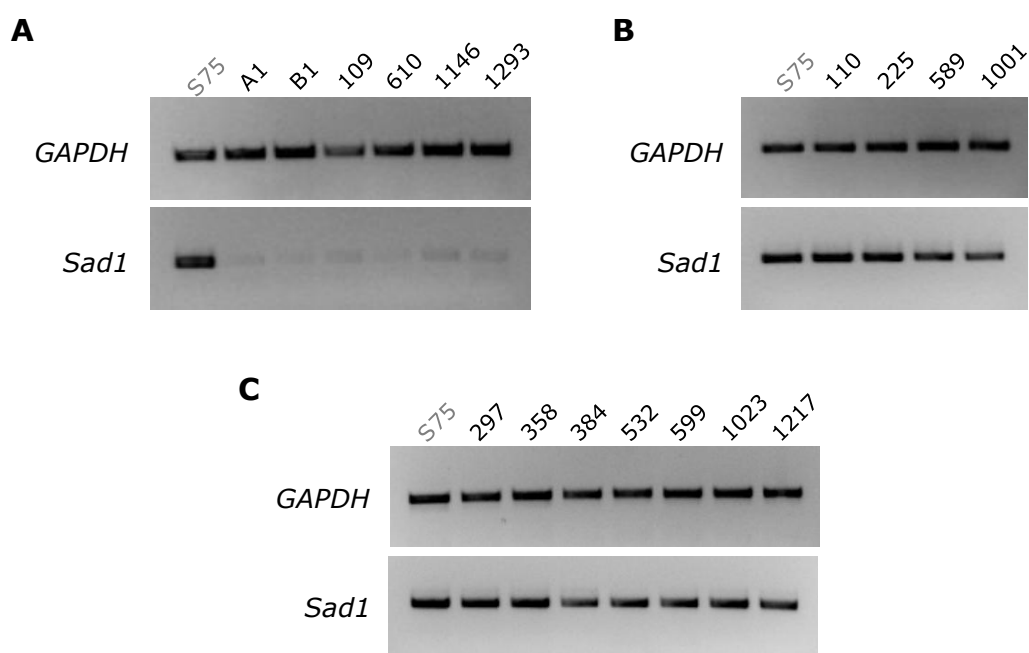


Figure 3.5: RT-PCR to investigate transcript levels in *sad1* mutants

RT-PCR using *GAPDH* and *Sad1* primers to determine transcript levels in wild-type and *sad1* mutants. *GAPDH* was used as positive control of expression and to standardise the levels of cDNA. The levels of gene expression for each of the *sad1* mutants were compared to the wild-type S75. **A.** Premature termination of translation mutants. **B.** Splicing error mutants. **C.** Amino acid substitution mutants.

by these mutations. In order to further investigate transcript size, Northern blotting and further RT-PCR reactions spanning the splice site mutation locations were carried out and these results are discussed in section 3.3.2.2 and 3.3.2.3 respectively.

3.3.2.1.3 Amino acid substitution mutants

As expected when compared to the wild-type S75, all amino acid substitution mutants had transcript present (Figure 3.5C). As these mutants had single point mutations which were predicted to result in a single amino acid substitution, it was assumed that transcript size was not affected in these mutants, therefore the cause of non-functionality in these mutants must presumably have occurred at the protein level.

3.3.2.2 Northern blot analysis of splicing error mutants

RT-PCR analysis using primers at the 5' end of the *Sad1* gene had shown that all splicing mutants had transcript present but gave no indication as to whether transcript size had been altered by the mutations. All of the splicing mutants were analysed by Northern blotting to obtain valuable information about overall transcript size, using probes for the 5' and 3' ends of the gene.

Ten micrograms of root total RNA was used for Northern blot analysis. All four splicing error mutants were analysed, together with the following controls: wild-type S75 as a positive control for expression; predicated premature termination of translation mutant 109 as a negative control; and amino acid substitution mutant 532 as a positive mutant control. All samples were run on a 1.2% agarose/0.25 M formaldehyde gel and visualised using ethidium bromide staining to check that sample loading levels were equal (Figure 3.6A). After the RNA had been transferred and immobilised on the nylon membrane, it was probed with cDNA probes complementary to the 5' and 3' end of the *Sad1* gene. Hybridisation was performed at 65°C overnight with ³²P-dCTP labelled probes. After the membrane had been thoroughly washed to remove background radiation, it was exposed to a phosphoimaging screen overnight.

Ethidium bromide staining showed that sample loading was mostly even between all samples – wild-type S75 had slightly lower levels than the mutants (Figure 3.6A). Both the 28S and 18S ribosomal RNA bands could be clearly seen

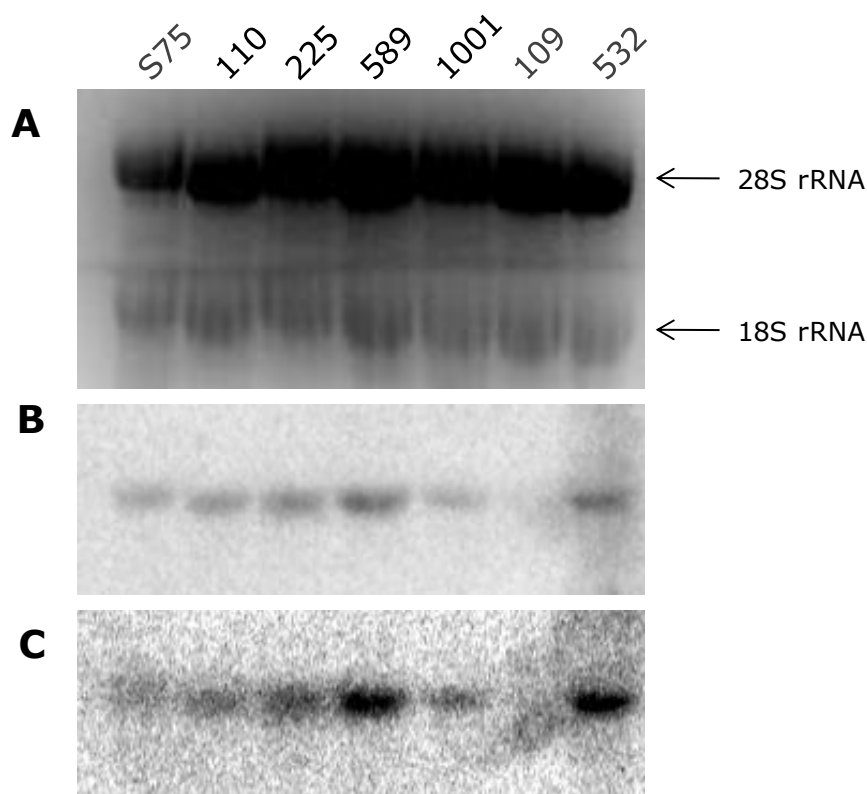


Figure 3.6: Northern blot analysis of RNA from roots of splicing error mutants

A. Ethidium bromide stained formaldehyde agarose gel of root tip RNA to ensure even loading. **B.** Membrane hybridised with *Sad1* 5' cDNA probe. **C.** Membrane hybridised with *Sad1* 3' cDNA probe. Lane 1, S75 – wild-type positive control; Lane 2-5, splicing mutants 110, 225, 289 and 1001; Lane 6, premature termination of translation mutant 109 – negative control; Lane 7, amino acid substitution mutant 532 – positive mutant control. Ten micrograms of RNA was loaded in each lane and RNA levels were monitored using ethidium bromide staining.

and no lower size bands were detected, which confirmed that the RNA was of good quality. The *Sad1* 5' cDNA probe was generated using the same primers as for the RT-PCR analysis, therefore hybridisation with this probe was expected to produce the same results. The Northern blot results confirm the results seen from RT-PCR analysis, as transcript is seen in the wild-type controls as well as all of the splicing mutants and the positive mutant control 532. As seen in RT-PCR analysis, premature termination of translation mutant 109 does not have transcript (Figure 3.6B). The *Sad1* 3' cDNA probe was generated from the terminal 21 bases of the *Sad1* gene and presence or absence of transcript would be indicative of whether mRNA decay had taken place which would provide clues as to the type of splicing error that had occurred. Results from hybridisation using the *Sad1* 3' cDNA probe are identical to those of the *Sad1* 5' cDNA probe which confirms that the *Sad1* transcript is intact (Figure 3.6C). This implies that intron inclusion has not occurred as if the intron adjacent to the splicing mutation had not been spliced and remained in the final transcript, a PTC would have been introduced. As seen in the premature termination of translation mutants, introduction of a PTC into the transcript results in NMD and transcript levels are severely reduced. As transcript is present in all of the splicing mutants, it can be concluded that intron inclusion has not occurred in any of the splicing error mutants and exon deletion is much more likely to have occurred.

As the RNA samples were separated by gel electrophoresis before being transferred to the membrane, this provided the opportunity to examine overall transcript size. All transcripts appear to be of the same size and show no major size differences. The minor size differences that are seen, specifically in S75, can be attributable to RNA loading, with samples containing more RNA migrating slightly further than those with less RNA. If, as the Northern blot data suggests, the exon adjacent to each of the splicing mutants was deleted, this would result in size differences of 75-189 base pairs. These small size differences may not be distinguishable on an agarose gel therefore further analysis is needed to determine whether exon deletion had occurred in any of the splicing error mutants.

3.3.2.3 Further RT-PCR analysis of splicing error mutants

Previous Northern blot and RT-PCR analysis had shown that both the 5' and 3' ends of the *Sad1* transcript were intact in all of the splicing mutants, and therefore ruled out intron inclusion being caused by the splice site mutations. This left exon deletion as the most likely effect of the splice site mutations, however Northern blotting had failed to reveal any obvious transcript size

differences, possibly because the expected size differences were too small to be resolved. Therefore in order to determine whether transcript size had been reduced and hence exon deletion had occurred, RT-PCR was carried out to amplify the region surrounding each of the splicing error mutants to look for small size differences.

A region between 200 bp and 550 bp surrounding each of the splicing error mutations was amplified for each of the splicing mutants together with a wild type control. The RT-PCR reaction products were run on a 1% TAE agarose gel and visualised using ethidium bromide staining (Figure 3.7). The results clearly show that when compared to the wild-type, all of the splicing mutants have a smaller product in the region surrounding their corresponding splice site mutation. Figure 3.7A shows the amplification of the 315 bp region surrounding exon 16, the location of the splice site mutation in mutant #110. In the wild-type and mutants #225, #589 and #1001, a band just above the 300 bp marker

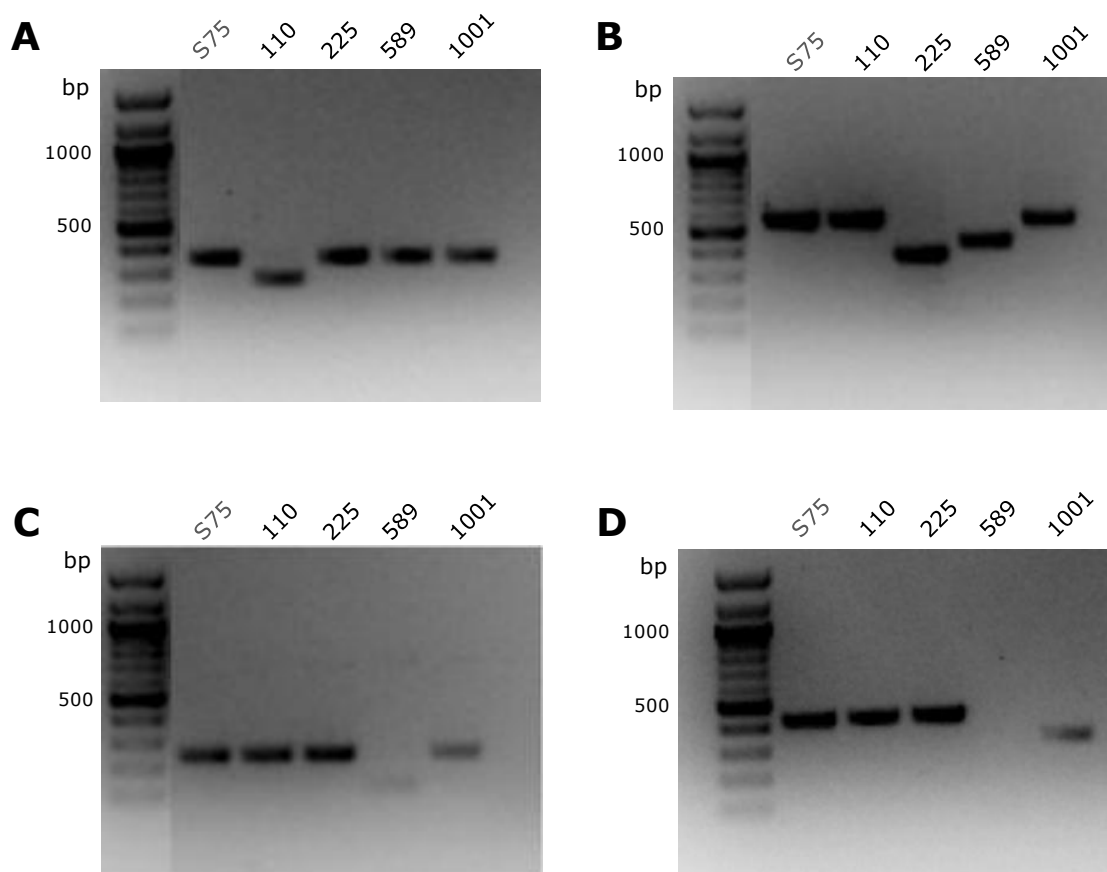


Figure 3.7: RT-PCR to amplify regions surrounding splice site mutations.

RT-PCR using various *Sad1* primer combinations to determine transcript size in regions surrounding each of the splice site mutations in wild-type and *Sad1* splicing error mutants. Fragment sizes for each of the *sad1* mutants were compared to the wild-type S75. Each of the agarose gel photo images corresponds to the amplification of the region surrounding each of the splice site mutations. **A.** G⁶⁶⁸⁹→A in mutant #110, 315bp. **B.** G³³⁰²→A and G³⁹¹⁴→A in mutants #225 and #589 respectively, 553bp. **C.** G³⁹¹⁴→A in mutant #589, 203bp. **D.** G⁴³⁶⁵→A in mutant #1001, 336bp.

corresponding to the expected size of the region is seen, however in mutant #110, the band seen is smaller, approximately 250 bp in size. Figure 3.7B shows the amplification of the 553 bp region surrounding exons 7-9, the locations of the splice site mutations in mutants #225 and #589. In the wild-type and mutants #110 and #1001, a band at approximately 550 bp is seen which corresponds to the expected size of the region, however in mutants #225 and #589, a smaller band is seen at 350 bp and 450 bp respectively. Figure 3.7C shows the amplification of the 203 bp region surrounding exon 9, the location of the splice site mutation in mutant #589. In the wild-type and mutants #110, #225 and #1001, a band just above the 200 bp marker is seen which corresponds to the expected size of the region, however in mutant #589 a faint smaller band at approximately 100 bp is seen. Figure 3.7D shows the amplification of the 435 bp region surrounding exon 11 which includes the splice site mutation in mutant #1001. In the wild-type and mutants #110 and #225 a band just above the 400bp marker is seen which corresponds to the expected size of the region, however in mutant #1001 a smaller band at approximately 350bp is seen. No band is seen in mutant #589 as the forward primer is located in the exon that has been deleted.

All of the splicing mutants had a reduced size transcript in the region where their point mutation occurred. This reduction in size corresponded to a deletion of the exon that each splicing mutation is associated with, therefore it can be concluded that errors at *Sad1* splice sites, whether they be at the start or end of the exon, result in deletion of the exon. This reduces the size of the transcript and may affect protein folding or protein function if important regions of the protein have been deleted. It was then of interest to determine whether any of the splicing mutants had protein present. These results will be discussed in the following chapter.

Chapter 4 - Protein analysis of *sad1* mutants

4.1 Introduction

4.1.1 Genetic analysis of *sad1* mutants

The previous chapter describes the genetic analysis of the 17 *sad1* mutants by sequencing the genomic DNA to determine the location of the mutations and by RT-PCR and Northern blotting to determine whether *sad1* transcript was present. At the gene level, genomic DNA sequencing showed that all mutants contained single point mutations in the *Sad1* gene which introduced predicted premature termination of translation codons, splicing errors or amino acid substitutions into the *Sad1* coding sequence. RT-PCR showed that all of the predicted premature termination of translation mutants had a severe reduction in *Sad1* transcript levels possibly due to degradation of the faulty transcripts by nonsense-mediated mRNA decay. Both the splicing error mutants and the amino acid substitution mutants had *Sad1* transcript, but Northern blotting and further RT-PCR experiments showed that the splicing error mutants contained a single exon deletion thereby shortening the transcript. The amino acid substitution mutants were the only mutants to produce full-length *sad1* transcript.

4.1.2 Aims

Following on from the genetic characterisation of the 17 *sad1* mutants in the previous chapter, the mutants will now be analysed at the protein level by Western blotting to determine whether full-length β -amyrin synthase protein (AsbAS1) is produced. Western blotting requires a primary antibody specific to AsbAS1, therefore a polyclonal AsbAS1 antibody will be generated using an *Escherichia coli* expression system. Once the Western blotting conditions have been optimised, oat root protein extracts from each of the mutants can be probed with the AsbAS1 antibody to determine which, if any, of the *sad1* mutants produce full-length protein. This will provide and insight into catalytic regions of the protein that play an important role in OSC function.

4.2 Materials and Methods

4.2.1 Cloning of *Sad1* into pET-14B vector

The β -amyrin synthase cDNA was amplified by PCR using a full-length cDNA clone, AsbAS1-pYES2 (49) as a template. The following primers were used

which contained *Nco*I and *Bam*HI restriction sites for cloning into the pET-14B vector (*Nco*I site is shown in red, *Bam*HI site is shown in blue): pET14-5' = GGTA CGCCATGGGCAGGCTAACAATAGGTGAGGGCGGCG, and pET14b-3' = CTATAT GGATCCTTAGCTCTTAATCGCAAGAAGTCGACG. PCR was carried out using Phusion™ High Fidelity DNA polymerase (Finnzymes) using the following program: 98°C for 1 minute followed by 30 cycles of 98°C for 10 seconds, 60°C for 30 seconds, 72°C for 3 minutes and a final extension at 72°C for 5 minutes. The presence of a product of the correct size was confirmed by 1% TBE agarose gel electrophoresis with ethidium bromide staining, and the resulting DNA fragments were purified using a Wizard PCR Preps Kit (Promega). Using the two new restriction sites, the amplified DNA fragments were cloned into the pET-14b vector (Invitrogen) and transformed by heat-shock into *Escherichia coli* DH5 α cells (Invitrogen) and spread onto LB agar plates (1% tryptone, 1% NaCl, 0.5% yeast extract, 1.5% agar) containing 50 μ g/ml ampicillin for growth overnight at 37°C. Plasmid DNA was purified by Wizard SV Minipreps Kit (Promega) and digested with *Eco*RI (Invitrogen) for 1 hour at 37°C to confirm the presence of the *Sad1* coding sequence. Those *E. coli* transformants containing the *Sad1* coding sequence were sequenced.

4.2.2 Expression of pET-14b-Sad1 in *Escherichia coli* BL21(DE3) cells

Purified pET-14b-Sad1 plasmid DNA was transformed by heat-shock into competent *E. coli* BL21(DE3) cells and spread onto LB plates containing 100 μ g/ml ampicillin for growth overnight at 37°C. For small scale expression trials, a single colony was grown in 10 ml of LB media (1% tryptone, 1% NaCl, 0.5% yeast extract) containing 50 μ g/ml ampicillin overnight at 37°C. Two ml of the culture was used to inoculate 50 ml of LB media containing 50 μ g/ml ampicillin and grown at 37°C with shaking until the culture reached log phase growth ($OD_{600} = 0.5$). Isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce protein expression, and the culture was grown at 37°C with shaking for 4 hours. After 4 hours, 1 ml of the culture was taken for expression analysis.

For large scale expression, a single colony was grown in 10 ml of LB media containing 100 μ g/ml ampicillin overnight at 37°C. Two 500 ml flasks of LB media were each inoculated with 5ml of the culture and grown at 37°C with shaking until the culture reached log phase growth ($OD_{600} = 0.5$). IPTG was added to a final concentration of 1 mM to induce protein expression and the culture was grown at 37°C with shaking for 3 hours. After 3 hours 1 ml of the

culture was taken for analysis of expressed protein and the remainder of the culture was centrifuged at 5,000 xg for 5 minutes and stored at 4°C.

Samples were prepared for SDS-PAGE analysis by standardising the OD₆₀₀ after 3 hours of induction to the OD₆₀₀ at 0 hours of induction in a 1 ml volume. Samples were centrifuged at 16,000 xg for 5 minutes and the pellet was resuspended in 20 µl of distilled water. Five µl NuPAGE® LDS Sample Buffer (Invitrogen) and 13 µl distilled water were added to 2 µl of the resuspended culture. The protein preparation was heated at 95°C for 30 minutes and then centrifuged at 16,000 xg for 5 minutes.

4.2.3 Purifying AsbAS1 from inclusion bodies

Cells harvested from the 1 L culture were resuspended in 50 ml lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 5% (v/v) glycerol, 1 mM DTT, 50 mM NaCl, 0.2% (w/v) sodium deoxycholate (NaDOC), 200 µg/ml lysosyme) supplemented with 1 tablet of Complete protease inhibitor (Roche) per 50 ml and incubated on ice for 30 minutes. The cell suspension was sonicated at 18 microns, using a Soniprep 150 (MSE), for 3 sets of 30 second pulses with a pause of 30 seconds on ice between each pulse. The cell lysates were centrifuged at 15,000 xg for 20 minutes and the supernatant (S/N1) was decanted and stored at 4°C for analysis. The cell pellet was resuspended in 50 ml Wash buffer (50 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 5% (v/v) glycerol, 50 mM NaCl, 2% (w/v) NaDOC, 0.1 mM DTT) and incubated at 4°C with slow stirring for 1 hour. The sonication, centrifugation and resuspension steps were repeated twice and the supernatants (S/N2 and S/N3) were stored at 4°C for analysis. The cell pellet was resuspended in 50 ml solubilisation buffer (50 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 5% (v/v) glycerol, 50 mM NaCl, 0.25% sarkosyl (N-lauroylsarcosine), 0.1 mM DTT) and incubated at 4°C with slow stirring for 1 hour. The cell suspension was dialysed against 2 L dialysis buffer (50 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 5% (v/v) glycerol, 50 mM NaCl, 0.1 mM DTT) at 4°C overnight, then centrifuged at 15,000 xg for 20 minutes. The supernatant (S/N4) was decanted and concentrated using an Amicon Ultra-15 Centrifugal Filter Unit 30,000 MWCO (Millipore) then stored at 4°C.

4.2.4 SDS-PAGE gel purification of AsbAS1 and antibody production

Concentrated S/N4 supernatant (1.5 ml) from inclusion body purification was added to 500 µl NuPAGE® LDS Sample Buffer (Invitrogen) and heated in a boiling water bath for 10 minutes then placed on ice for a further 5 minutes. Two

ml of the sample was loaded onto an 10% SDS-PAGE gel placed in a PROTEAN® II xi Cell (Bio-Rad). Both the upper and lower chambers were filled with Gel Running Buffer (0.25 M Tris, 1.92 M glycine, 1% (w/v) SDS) and the gel was electrophoresed at 50 mA through the stacking gel and 70 mA through the running the gel for 4½ hours. The gel was stained in Coomassie Brilliant Blue (Bio-Rad) for 1 hour and destained overnight.

For electroelution of AsbAS1 from the SDS-PAGE gel, the band corresponding to the molecular weight of AsbAS1 was excised from the gel and placed into dialysis tubing filled with 0.5x Gel Running Buffer. The tubing was placed horizontally in a large gel tank and protein was electroeluted overnight at 15 mA in 0.5x Gel Running Buffer. Electroeluted protein was dialysed against 5 L water at 4°C overnight and then freeze-dried before storage at -20°C.

Freeze-dried electroeluted AsbAS1 (1 mg) was used as an antigen for antibody production in two rats (BioGenes).

4.2.5 Protein analysis of *sad1* mutants

Total protein was extracted from the terminal 0.5 cm of 3 day old oat root tips by grinding the roots in protein extraction buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 10% (v/v) glycerol, 1% (w/v) PVPP, 1% (v/v) Triton X-100 (Boehringer Mannheim), 1x Complete protease inhibitor (Roche)) for 1 minute with a plastic pestle followed by incubation at 4°C for 2 hours. Protein extract was obtained by centrifugation at 16,000 xg for 20 minutes and stored at -20°C in aliquots of 10 µl with 5 µl distilled H₂O and 5 µl NuPAGE® LDS Sample Buffer (Invitrogen).

For sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), each 20 µl aliquot was heated at 95°C for 30 minutes then centrifuged at 16,000 xg for 5 minutes. Each sample (15 µl) was loaded onto a NuPAGE® Novex 4-12% Bis-Tris Gel (Invitrogen) placed in an XCell SureLock™ Mini-Cell (Invitrogen). Both the upper and lower chambers were filled with 1x MES SDS Running Buffer (Invitrogen) and 0.5 ml NuPAGE® Antioxidant (Invitrogen) was added to the upper buffer chamber. The gel was electrophoresed at 200 V for 40 minutes and stained by incubation with InstantBlue (Novexin) for 1 hour.

For Western blotting, samples were prepared and run on a NuPAGE® Novex 4-12% Bis-Tris Gel (Invitrogen) as described in section 4.2.4, then transferred to a nitrocellulose membrane using the XCell II™ Blot Module at 30V constant for 1 hour. The membrane was washed in Tris-Buffered Saline (TBS) (20 mM Tris-HCl pH 7.5, 7.5 mM NaCl) and incubated for 1 hour/overnight in blocking solution (3% Bovine Serum Albumin, 10 mM Tris-HCl pH 7.5, 7.5 mM

NaCl). The membrane was then washed with TBS with Tween-20 and Triton X-100 (TBSTT) (20 mM Tris-HCl pH 7.5, 0.5 mM NaCl, 0.05% v/v Tween[®]-20, 0.2% v/v Triton X-100) and TBS and incubated with *Sad1* polyclonal antibody diluted 1:10,000 in blocking solution for 1 hour. The membrane was washed again with TBSTT and TBS and incubated with Goat Anti-Rat IgG Horseradish peroxidase conjugate (Sigma) diluted 1:10,000 in blocking solution for 1 hour. The membrane was thoroughly washed in TBSTT and incubated Supersignal[®] West Pico Chemiluminescent Substrate (Thermo Scientific) for 7 minutes at room temperature. The membrane was then exposed to Super RX 18 x 24cm film (Fujifilm) and developed using a SRX-101A Film Processor (Konica Minolta).

4.3 Results and Discussion

4.3.1 Generation of AsbAS1 antibody

Transcript analysis of both the premature termination of translation mutants and the amino acid substitution mutants had the expected results, with the former having severely reduced transcript levels and the latter having wild-type levels of transcript present. Detailed analysis of the splicing error mutants revealed that transcript was present in all of the mutants, but at a reduced length due to deletion of the exon adjacent to the mutation in each mutant. It was clear that both the splicing error and amino acid substitution mutants were affected at a post-transcriptional stage, so it was of interest to investigate the levels of SAD1 (AsbAS1) protein present in each of the mutants. If AsbAS1 protein was found in any of the mutants, this would provide valuable information about important residues or regions that are required for function in pentacyclic oxidosqualene cyclases, a class of enzymes that has previously not been studied in great detail.

Western blotting was used in order to study levels of AsbAS1 protein in each of the *sad1* mutants, but this required a primary antibody specific to AsbAS1. Antibodies were available for other proteins encoded by different genes in the avenacin pathway (*Sad7*, *Sad9* and *Sad10*), but due to the membrane bound nature of AsbAS1, previous attempts to purify the protein for antibody generation had failed (62). However without a specific antibody, accurate protein detection of AsbAS1 protein in all of the *sad1* mutants would have been very difficult, therefore another attempt was made at generating an antibody for AsbAS1. If this was successful then the antibody could be used to probe root extracts of *sad1* mutants detect presence of AsbAS1 protein.

4.3.1.1 *Sad1* antibody generation

Attempts to purify AsbAS1 in *Pichia pastoris* (see Chapter 5) and *Nicotiana benthamiana* (86) using immobilised metal affinity chromatography had been unsuccessful as too many impurities remained in the fractions eluted from the column. Both the yeast and tobacco systems were originally selected as a means to generate functional protein for kinetic and structural studies, but as functional protein is not required for antibody generation, the *Escherichia coli* system could be used for protein purification. This system had been successfully used to purify the *Sad7* gene product, *Avena strigosa* serine carboxypeptidase-like acyltransferase (AsSCPL1) from inclusion bodies, which was then used to generate a highly specific antibody (62). Therefore this method was used to purify AsbAS1 protein for antibody generation.

4.3.1.1.1 Cloning of *Sad1* into the pET-14b vector

The full-length *Sad1* cDNA sequence was amplified by PCR using primers to add *NcoI* and *BamHI* restriction sites to the 5' and 3' ends of the coding sequence respectively. The products were analysed by 1% TAE gel electrophoresis to confirm the presence of a correctly sized product at ~2.3kb. The DNA fragments were then digested with *NcoI* and *BamHI* and ligated into the pET-14b vector to generate the construct BAS-E1, and transformed into *E. coli* DH5 α cells (Figure 4.1A). Plasmid DNA was isolated from the transformed cells

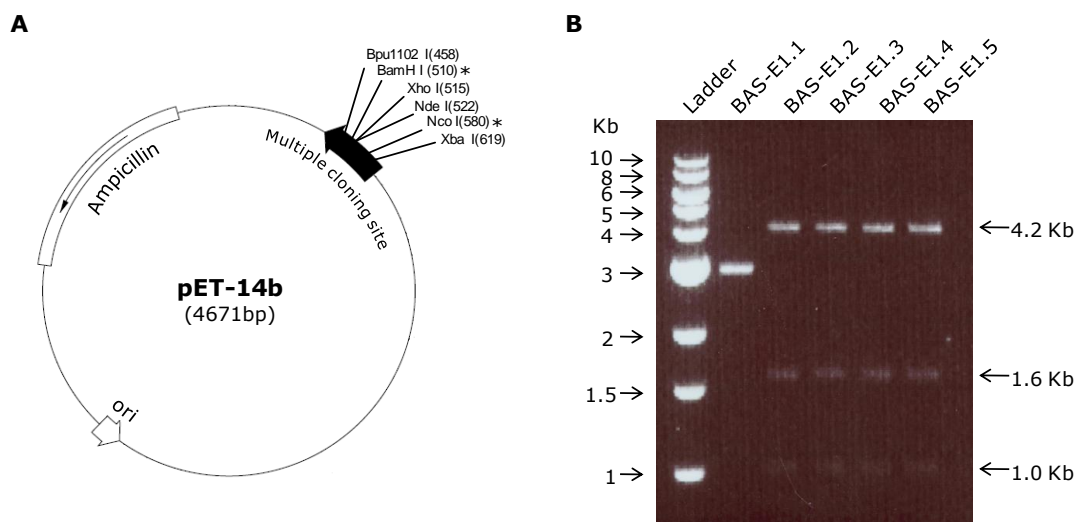


Figure 4.1: Cloning of *Sad1* into the pET-14b vector

A. Plasmid map of pET-14b (4.6kb). ori – origin of replication; Ampicillin – *bla* gene for ampicillin resistance. Multiple cloning site is also shown and restriction sites used for *Sad1* cloning are denoted by *. **B.** *EcoRI* digestion of plasmid DNA extracted from four BAS-E1 transformants to identify constructs containing the *Sad1* gene. Plasmids containing the *Sad1* gene have bands at 4.2kb, 1.6kb and 1kb whereas undigested plasmids that do not contain the *Sad1* gene have a single band at 3 kb. The digested plasmid DNA was run on a 1% TAE agarose gel and visualised with EtBr staining.

and digested with *EcoRI* to confirm presence of *Sad1* cDNA. The *Sad1* coding sequence has two *EcoRI* sites 1kb apart and pET-14b has one *EcoRI* site situated 512bp upstream of the *NcoI* restriction site, so successful integration of the *Sad1* cDNA into the pET-14b vector would generate fragments of 4.2kb, 1.6kb and 1kb (Figure 4.1B). The BAS-E1-pET14b construct had a number of *E. coli* transformants which produced the desired fragment sizes after digestion with *EcoRI*. Sequencing of the *Sad1* gene in a selection of these *E. coli* transformants confirmed that the gene had been successfully cloned into the pET-14b vector.

4.3.1.1.2 Expression of BAS-E1 in *Escherichia coli* BL21(DE3) cells

For small scale protein expression trials, the BAS-E1 construct was transformed into *E. coli* BL21(DE3) cells. BL21(DE3) cells were chosen as they produce high levels of expression for T7 promoter-driven vectors such as pET-14b (244). Expression trials in a 50ml culture volume were conducted on five *E. coli* BL21(DE3) transformants (BAS-E1.1:BL21(DE3) to BAS-E1.5:BL21(DE3)) to identify a highly expressing transformant. Protein expression was induced by addition of IPTG to a final concentration of 1mM and the culture was grown for four hours at 37°C with shaking. After four hours, 1ml of the culture was taken for expression analysis. Cell pellets were resuspended in distilled water and analysed by SDS-PAGE to detect protein expression.

SDS-PAGE analysis for each of the transformants was conducted by comparing a sample taken before protein induction to the sample taken after four hours of protein induction to look for the appearance of novel bands. After InstantBlue staining all transformants had a novel band between the 62 and 98kDa markers that was not seen in the uninduced sample. This band corresponded to the expected molecular weight of AsbAS1, 86.8kDa. There was no significant difference in the level of the novel protein produced by each of the transformants, so it did not matter which was chosen to carry forward to large scale expression.

4.3.1.1.3 Inclusion body purification of AsbAS1

BAS-E1.4:BL21(DE3) was selected from small scale expression trials as the transformant to be used for large scale expression and purification of AsbAS1. For large scale expression the culture volume was scaled up to two 500ml cultures to give a total culture volume of one litre. Protein expression was induced by addition of IPTG to a final concentration of 1mM and the cultures were grown for three hours at 37°C with shaking. After three hours 1ml of each

culture was taken for expression analysis and the remainder of the cells were harvested by centrifugation and stored at -80°C . The samples were analysed by SDS-PAGE to ensure that the novel band between the 62 and 98kDa markers that was seen in trial expression analysis was produced (Figure 4.2). The SDS-PAGE gel showed that both 500ml cultures (culture 1 and culture 2) produced a novel band between the 62 and 98kDa markers after three hours of protein induction (Lanes 6 and 9) which corresponded to the novel band seen in the small scale expression trial (Lane 3) and the expected molecular weight of AsbAS1.

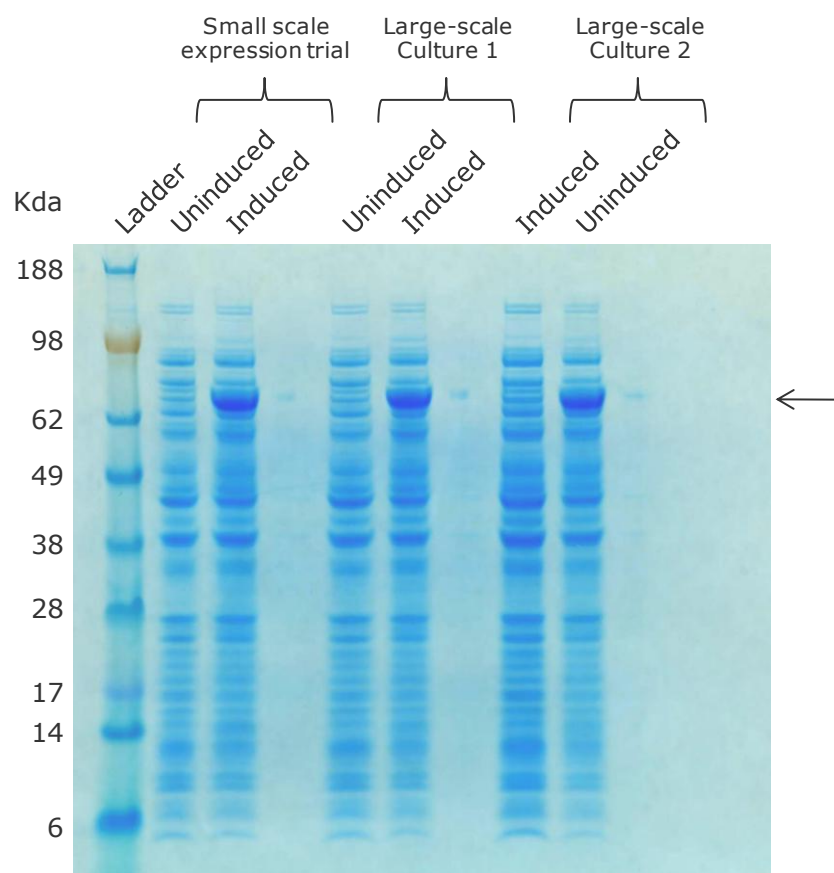


Figure 4.2: Instant Blue stained SDS-PAGE gel of samples taken following large scale expression of AsbAS1.

SDS-PAGE gel shows analysis of small and large scale BAS-E1.4:BL21(DE3) culture samples taken before and 3 hours after protein induction. Lane 1, SeeBlue® Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), myoglobin red (17kDa), lysozyme (14kDa), aprotinin (6kDa). Small scale expression of BAS-E1.4:BL21(DE3): Lane 2, uninduced culture (0 hours); Lane 3, induced culture (3 hours); Lane 4, blank. Large scale expression of BAS-E1.4:BL21(DE3): Lane 5, Culture 1 uninduced (0 hours); Lane 6, Culture 1 induced (3 hours); Lane 7, blank; Lane 8, Culture 2 uninduced (0 hours); Lane 9, Culture 2 induced (3 hours). Expected molecular weight of AsbAS1 (shown by the arrow) is predicted to be 86.8kDa, which falls between the 62 and 98kDa markers.

For inclusion body purification, the cells harvested from protein induction were lysed by sonication in a buffer containing 0.2% (w/v) sodium deoxycholate, a mild detergent used to lyse cells and solubilise cellular and membrane components (245). After centrifugation to remove soluble proteins, the pellet was then washed two times in a buffer containing 2% (w/v) sodium deoxycholate which involved sonication, incubation at 4°C with slow stirring for one hour followed by centrifugation. At this stage the pellet appeared brown-white and had a chalky appearance. The protein was solubilised from the inclusion bodies by resuspension and incubation at 4°C with slow stirring for one hour in a buffer containing sarkosyl, a detergent used to disrupt the interactions between inclusion bodies (246). The solution was dialysed overnight to remove the detergent and underwent a final centrifugation step to remove insoluble fragments. The final supernatant was then concentrated from a volume of 80 ml

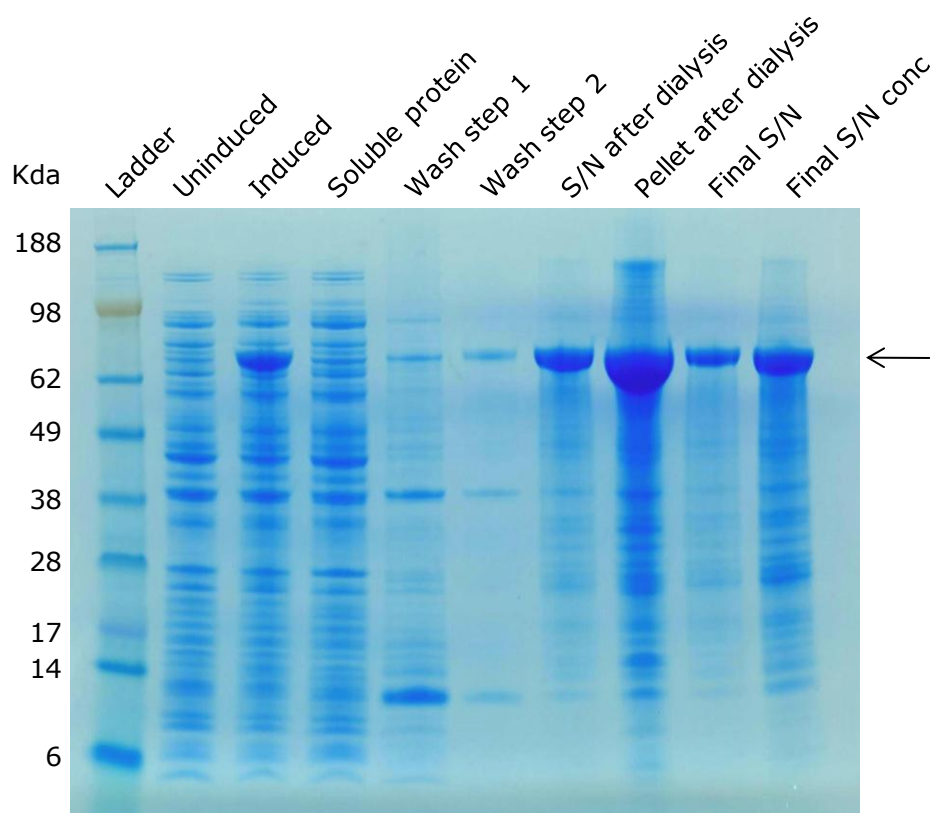


Figure 4.3: SDS-PAGE analysis of samples collected during inclusion body purification of AsbAS1.

SDS-PAGE gel of samples collected during inclusion body purification of AsbAS1. Lane 1, SeeBlue® Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), myoglobin red (17kDa), lysozyme (14kDa), aprotinin (6kDa); Lane 2, BAS-E1,4:BL21(DE3) uninduced (culture 1); Lane 3, BAS-E1,4:BL21(DE3) induced (culture 1); Lane 4, Soluble protein isolated following lysis step; Lanes 5 and 6, Supernatant following wash steps; Lane 7, Supernatant obtained following sarkosyl solubilisation and overnight dialysis; Lane 8, Pellet obtained following sarkosyl solubilisation and overnight dialysis; Lane 9, Supernatant obtained after final centrifugation step; Lane 10, Concentrated final supernatant. Expected molecular weight of AsbAS1 (shown by the arrow) is predicted to be 86.8kDa, which falls between the 62 and 98kDa markers.

to 7 ml and stored at 4°C. Samples taken at all stages of the purification process were then analysed by SDS-PAGE to determine whether the induced band from large scale expression had been successfully purified (Figure 4.3). Lanes 2 and 3 contain the samples from the trial expression (lanes 5 and 6 of Figure 4.2) which show a strong band produced at the expected molecular weight of AsbAS1 following protein induction. Lane 4 contains the soluble protein present after cell lysis and centrifugation steps to remove insoluble molecules and membrane fragments. The strong band between the 62 and 98kDa markers is not present here indicating that this protein was present in the insoluble pellet fraction. Both wash step fractions removed some minor impurities and after solubilisation and dialysis the large band was present in the supernatant which indicated that it had been successfully solubilised from the inclusion bodies. The pellet remaining after solubilisation and dialysis still contained a large amount of the protein of interest, so more protein could have been recovered with a longer solubilisation period, but the amount of protein recovered was sufficient for antibody production. In the final supernatant, the band of interest was clearly the most abundant protein in the sample and had been successfully solubilised from the inclusion bodies. Concentration of the protein showed that a lot of other lower molecular weight impurities were also present in the final sample, therefore another stage of purification was needed in order to obtain pure protein.

4.3.1.1.4 Gel extraction and electroelution of AsbAS1

The protein of interest had been partially purified from inclusion bodies, but still contained a number of impurities. As the protein was not expressed with any tags, column chromatography could not be used, therefore to obtain a pure protein sample gel extraction and electroelution were used. To separate the proteins based on size, a large SDS-PAGE gel was used to allow a large sample volume to be loaded. Two 10% SDS-PAGE gels were loaded with 1.5ml of the concentrated supernatant from inclusion body purification and 500µl of NuPAGE® loading dye. A tris-glycine buffer was used as it provides better resolution for large proteins. After electrophoresis and InstantBlue staining, a large thick band was seen approximately one third of the way down the gel which was likely to be the protein of interest. The top and bottom halves of the band, sample 1 and sample 2 respectively, were excised separately from the gel and the protein was electroeluted into a 0.5x tris-glycine buffer overnight. The electroeluted samples were dialysed overnight and then freeze-dried until the majority of the aqueous solution had been removed.

The two freeze dried samples were then analysed by SDS-PAGE to assess the purity of the sample and to quantify the amount of protein present. 10 μ l of each protein sample was analysed and a BSA dilution series was used to determine the concentration of the protein (Figure 4.4). The freeze dried electroeluted protein samples from the top and bottom halves of the gel extracted band are shown in Figure 4.4, lanes 3 and 4 respectively. Comparison with the partially purified sample in lane 2 shows that all of the lower molecular weight impurities have been removed and just a single band at the expected molecular weight of AsbAS1 is present. Quantification using the BSA dilution series in lanes 5-9 showed that the top and bottom halves of the excised band contained 1 μ g and 0.5 μ g protein respectively. This allowed the amount of protein in both of the

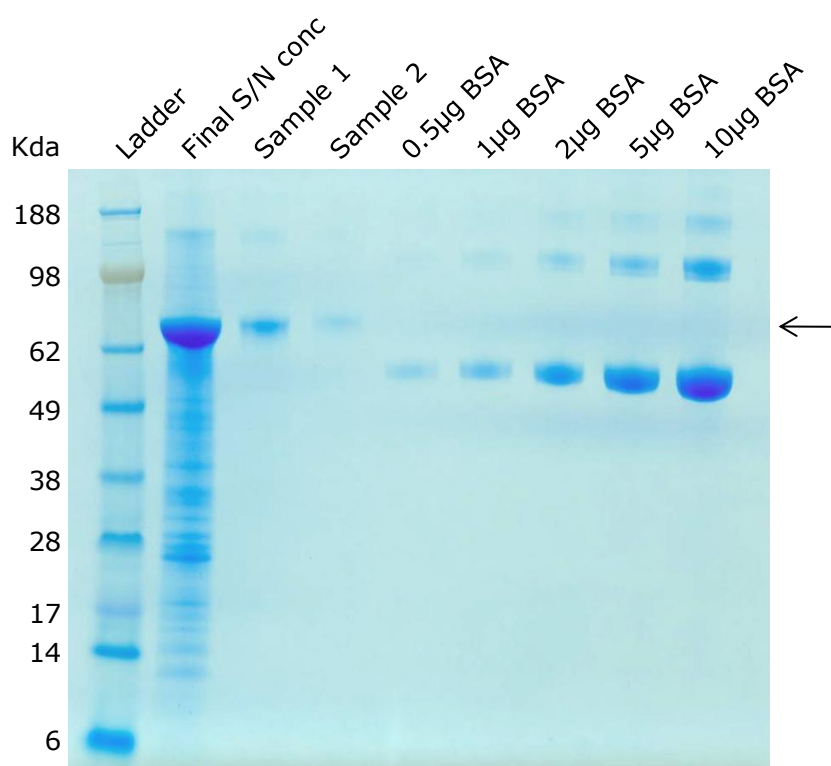


Figure 4.4: SDS-PAGE gel to quantify the concentration and purity of AsbAS1

SDS-PAGE gel to quantify the concentration of protein present in the freeze dried samples following gel extraction and electroelution of AsbAS1 from the large SDS-PAGE gel. Purity of samples was assessed by comparison to the partially purified sample following inclusion body purification. Concentration of samples was calculated by comparison of the purified samples to a BSA dilution series. Lane 1, SeeBlue® Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), myoglobin red (17kDa), lysozyme (14kDa), aprotinin (6kDa); Lane 2, Partially purified protein after inclusion body purification; Lane 3, Freeze-dried electroeluted protein excised from the top half of the gel extracted band (10 μ l); Lane 4; Freeze dried electroeluted protein excised from the bottom half of the gel extracted band (10 μ l); Lane 5, 0.5 μ g BSA standard; Lane 6, 1 μ g BSA standard; Lane 7, 2 μ g BSA standard; Lane 8, 5 μ g BSA standard; Lane 9, 10 μ g BSA standard. Expected molecular weight of AsbAS1 (shown by the arrow) is predicted to be 86.8kDa, which falls between the 62 and 98kDa markers.

samples to be calculated. The total volumes of samples 1 and 2 were 7 ml and 10 ml respectively therefore they contained 0.7 mg and 0.5 mg protein respectively. The total amount of pure protein obtained was 1.2mg and as both samples had equal purity they were combined and freeze dried.

To confirm the identity of the purified protein, the band in lane 3 of Figure 4.4 was excised and sent for Matrix Assisted Laser Desorption/Ionization – Time Of Flight – Mass Spectrometry (MALDI-TOF-MS fingerprinting). This technique involves digesting the proteins contained in the band of interest with trypsin and dissolving them in a matrix solution which protects the protein and facilitates ionisation. The peptide molecules are spotted onto a MALDI plate, then ionised using a laser which promotes the matrix molecules and peptides from a solid to a gaseous state. These molecules are then accelerated in the electric field of the mass spectrometer and their mass is calculated by measuring the time of flight (TOF) over a specified distance (247). The masses of the different peptide fragments are exported as peak lists which are then analysed by MASCOT, a search engine that uses the peak lists to search a number of databases containing protein sequence information to match the masses produced by tryptic digestion to known proteins. The results are exported as a list of possible matches each with a mascot score value. This score is calculated using the probability that the observed match between the MS data and the sequence database entry is a random event. Probability is converted to a mascot score using $-10\log_{10}(P)$, where P is the probability, therefore meaning that the best match has the highest score. A significance level is determined for each data set, but usually matches with a score above 70 are significant (248-249). Mascot analysis of the band confirmed the identity of the protein to be *Avena strigosa* β -amyrin synthase, with a mascot score of 235. This confirmed that the protein purified from *E. coli* BL21(DE3) cells was AsbAS1.

4.3.1.1.5 AsbAS1 antibody production and testing

The freeze dried protein was sent to BioGenes, Germany for polyclonal antibody production. Antibodies were raised in two rats which were immunised with 0.5mg of purified AsbAS1 protein solution over a three month period. Pre-immune serum was taken from each rat before immunisation and 3ml of final bleed serum was obtained from each rat at the end of the immunisation protocol (serum numbers 13082 and 13803). Each antibody was then tested against a number of different protein extracts to determine specificity of the antibody and to quantify the level of antibody dilution needed for Western blotting. Pre-

immune sera was also tested to determine whether any background signal was present in the rats before immunisation.

The protein extracts used to test the AsbAS1 antisera were *Avena strigosa* S75 root extract, *Avena strigosa* S75 leaf extract and *E. coli* expressed AsbAS1 (BAS-E1.4:BL21(DE3)). AsbAS1 is expressed in the roots of young oat seedlings but is not present in the leaves which provides a method of testing the specificity of the antibody (49). The closest related protein to AsbAS1 that is present in oats is the *Avena strigosa* cycloartenol synthase. This enzyme is expressed in all tissues and organs of oat, so if it is also detected by the AsbAS1 antibody then signal would be seen in both root and leaf extracts and the presence or absence of AsbAS1 protein would be masked (49). However, if a signal is seen in the oat root extract but not in the leaf extract this will show that the enzyme is highly specific to AsbAS1. The *E. coli* expressed AsbAS1 provides a positive control as it was used to raise the antisera, so should produce a strong signal. To optimise the protocol, membranes were probed with varying concentrations of the two rat antisera to find the concentration which produced the clearest signal with minimal background. Optimal dilution ratios for serum 13082 and 13083 were found to be 1:10,000 and 1:5,000 respectively.

Analysis of the three protein extracts produced the same results for both antisera. Western blots with the pre-immune sera were blank for both sets of sera at 30 seconds and 30 minutes of exposure. This confirmed that before immunisation neither of the rats contained antibodies that cross-reacted with any of the proteins present in the oat or *E. coli* extracts. Western blots with the final bleed sera had a clear signal at the expected molecular weight of AsbAS1 in the *A. strigosa* root extract and the *E. coli* expressed AsbAS1, but no signal was seen in the *A. strigosa* leaf extract (Figure 4.5). In the samples that produced a signal, only a single band was seen even after 30 minutes of exposure indicating that the AsbAS1 antibody is highly specific to AsbAS1, and lack of signal in the leaf extract showed that no cross-reaction with *A. strigosa* cycloartenol synthase is present. Therefore this highly specific, highly concentrated antibody could now be used to probe root extracts of the *sad1* mutants to look for presence of the AsbAS1 protein in the three different mutant groups.

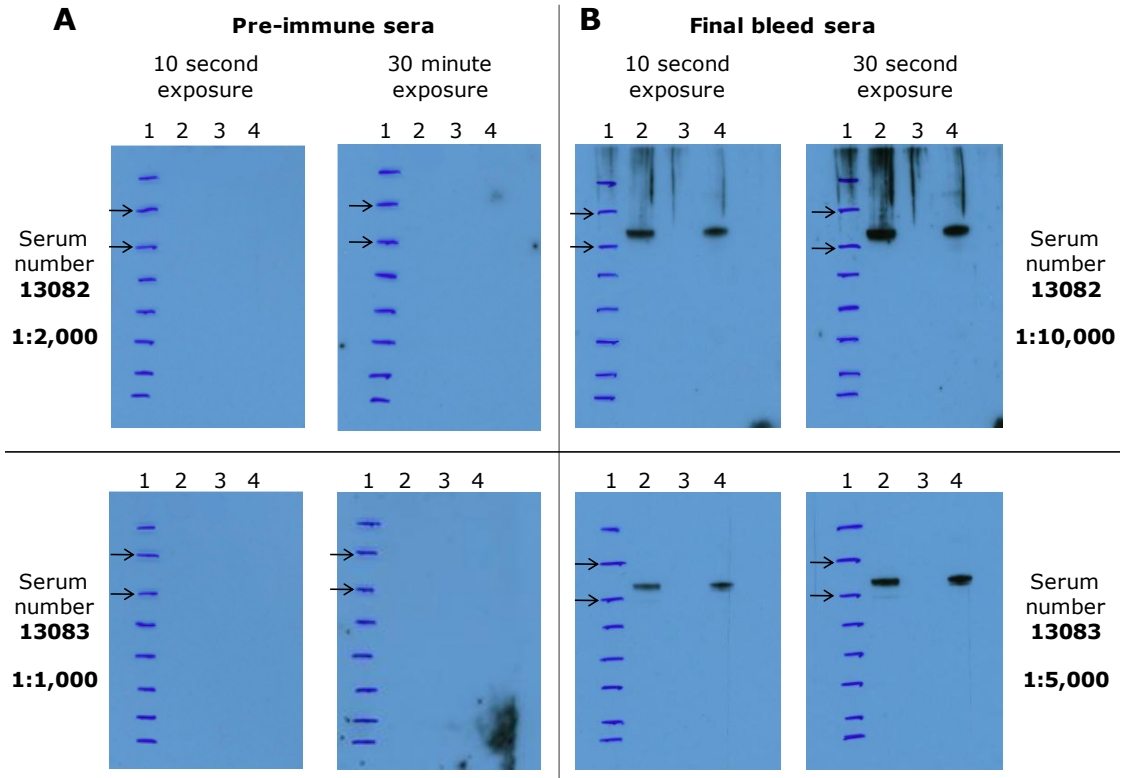


Figure 4.5: Testing of pre-immune and final bleed sera from AsbAS1 antibody production

Western blot analysis of pre-immune and final bleed sera from AsbAS1 antibody production in two rats (13082 and 13083). Lane 1, SeeBlue[®] Plus2 Pre-Stained Standard - phosphorylase (98kDa) and bovine serum albumin (62kDa) denoted by arrows; Lane 2, *A. strigosa* S75 root tip protein; Lane 3, *A. strigosa* S75 leaf protein; Lane 4, *E. coli* expressed AsbAS1 (BAS-E1.4:BL21(DE3) transformant). **A.** Pre-immune sera testing of serum numbers 13082 and 13083. Sera tested at 5x concentration of the final bleed sera. **B.** Final bleed sera testing of serum numbers 13082 and 13083.

4.3.2 Western blot analysis of *sad1* mutants

None of the *sad1* mutants had previously been analysed for presence of AsbAS1 protein, but it was expected that the premature termination of translation mutants would not contain AsbAS1 protein as the *Sad1* transcript had been severely degraded in each of these mutants. Transcript was present in the other the splicing error mutants but each mutant contains an exon deletion resulting in a shorter transcript so stable folded protein is unlikely to be produced. All of the amino acid substitution mutants have full length transcript so all have the potential to produce fully folded protein, but if the point mutation affects protein stability, then the protein will be degraded. However, those mutants that do contain full-length protein will provide valuable clues as to the amino acid residues and regions of the protein that are important for protein function.

AsbAS1 antibody generation had produced a highly specific, high titre antibody against AsbAS1 which could now be used to probe protein extracts of *sad1* mutants to look for presence or absence of AsbAS1 protein. Protein

extracted from the terminal 0.5cm of 3 day old root tips from each of the *sad1* mutants was run on an SDS-PAGE gel and transferred to a nitrocellulose membrane. The membrane was then probed with the AsbAS1 antibody from serum 13083 followed by a horseradish peroxidase conjugate for chemiluminescent detection using X-ray film.

4.3.2.1.1 Premature termination of translation mutants

As expected, when compared to the wild-type S75, none of the premature termination of translation mutants produced AsbAS1 protein (Appendix 8). These mutants had all had a predicted premature termination of translation codon within the coding sequence that resulted in a substantial reduction in *Sad1* transcript levels.

4.3.2.1.2 Splicing error mutants

Western blot analysis revealed that none of the splicing mutants had any AsbAS1 protein present (Figure 4.6). The wild type S75 root protein clearly shows a single band between the 62 and 98 kDa markers at the expected size of AsbAS1, but none of the splicing mutants have any evidence of a signal at any size. These mutants each had a point mutation at an intron-exon boundary which resulted in deletion of the exon adjacent to the mutation producing a shorter *Sad1* transcript. Although deletion of an exon does not alter the reading frame in any of the splicing mutants, relatively large regions of the coding sequence (110 – 25 amino acids, 225 – 63 amino acids, 589 – 41 amino acids and 1001 – 33 amino acids) would be deleted. The majority of the tertiary structure of AsbAS1 is formed of alpha helices, and all of the deleted regions would result in the removal of at least one helix which would seriously affect protein stability and folding. Accumulation of misfolded proteins can be toxic to cells, so eukaryotic cells have evolved a quality control mechanism to detect misfolded and unassembled proteins and either retain them in the endoplasmic reticulum for re-folding or transport them for degradation by a process known as ER-associated degradation (ERAD) (250-253). It is currently understood that misfolded proteins are recognised by detection of features such as, abnormal hydrophobic regions, unpaired cysteine residues or immature glycans. These misfolded proteins are then translocated to the cytoplasm where they are ubiquitinated and degraded by the 26S proteasome (254-255). Since no AsbAS1 protein is observed in any of these mutants, it seems likely that the shortened *sad1* transcripts produce

protein that is either unable to fold or is misfolded and is targeted for degradation.

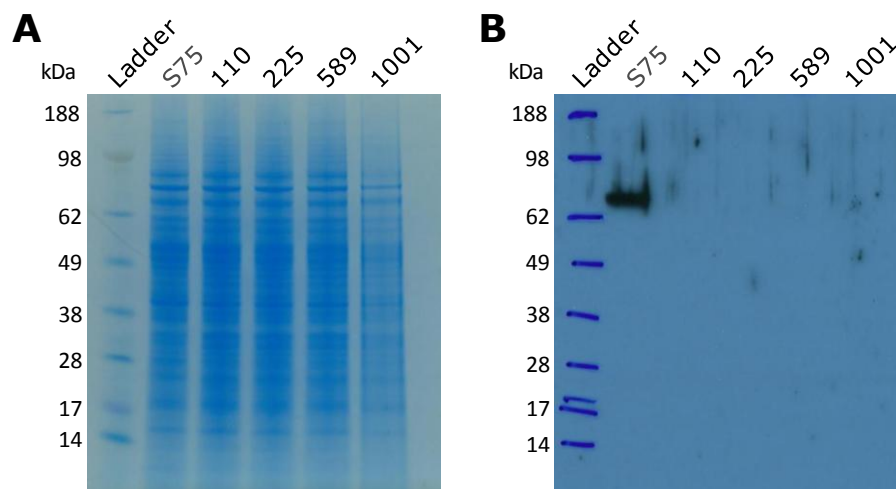


Figure 4.6: Western blot analysis of *sad1* splicing error mutants using the AsbAS1 antibody

Western blot analysis of the *sad1* splicing mutants to look for the presence of AsbAS1 protein. Membrane was probed with the AsbAS1 antibody at 1:10,000 dilution (13083 serum) and chemiluminescent signal detected by exposure to X-ray film for 30 seconds. Lane 1, SeeBlue® Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), myoglobin red (17kDa), lysozyme (14kDa), aprotinin (6kDa); Lane 2, Wild type S75 root tip protein; Lane 3-6, Root tip protein from mutant numbers 110, 225, 589 and 1001. **A.** SDS-PAGE gel to show even total protein loading. **B.** Western blot at 30 seconds exposure.

4.3.2.1.3 Amino acid substitution mutants

Western blot analysis of the seven amino acid substitution mutants revealed that three of the mutants had full length AsbAS1 protein present (Figure 4.7). In the wild-type, a single band is seen between the 62 and 98 kDa markers corresponding to the expected molecular weight of AsbAS1. This band is also seen in mutants #358, #384 and #1023 indicating that these mutants all produce full-length AsbAS1 protein, however the remaining four mutants, #297, #532, #599 and #1217 do not have any AsbAS1 protein as no signal is seen. As transcript is present in these four mutants, lack of protein is likely to be caused by ERAD in the same way as for the splicing mutants. Unlike the splicing mutants, these substitution mutants produce full-length protein, but the location of the mutations most likely destabilises the protein fold resulting in misfolded protein, which is then recognised by the cell and targeted for degradation. The mutation in #1217 is located near to one of the QW motifs which are postulated to be involved in protein stability, adjacent to the stacked glutamine and tryptophan residues. This mutation may therefore disrupt the stacking and destabilise the protein structure resulting in the protein being targeted for

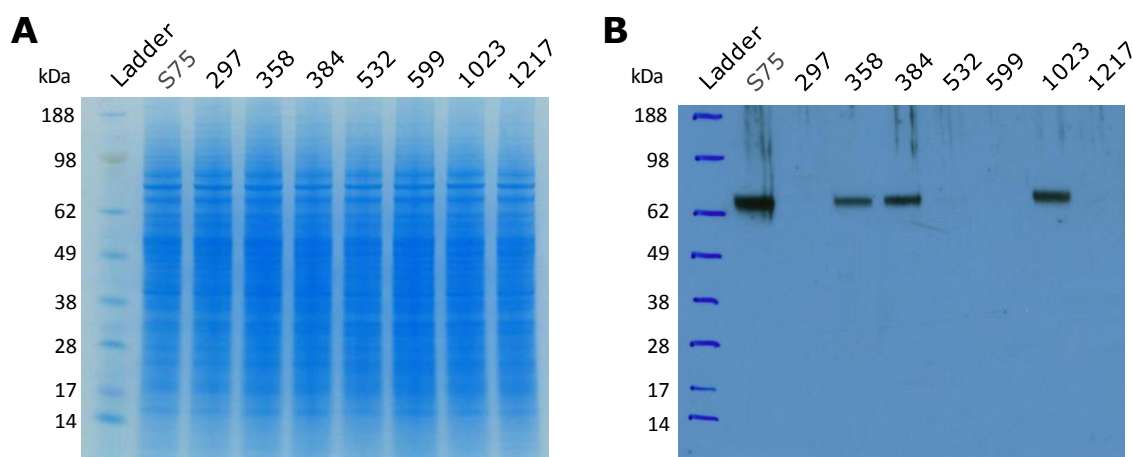


Figure 4.7: Western blot analysis of *sad1* amino acid substitution mutants using the AsbAS1 antibody

Western blot analysis of the *sad1* amino acid substitution mutants to look for the presence of AsbAS1 protein. Membrane was probed with the AsbAS1 antibody at 1:10,000 dilution (13083 serum) and chemiluminescent signal detected by exposure to X-ray film for 1 minute. Lane 1, SeeBlue® Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), myoglobin red (17kDa), lysozyme (14kDa), aprotinin (6kDa); Lane 2, Wild type S75 root tip protein; Lane 3-9, Root tip protein from mutant numbers 297, 358, 384, 532, 599, 1023, 1217. **A.** SDS-PAGE gel to show even total protein loading. **B.** Western blot at 1 minute exposure.

degradation. The mutation in #599 is found on the protein surface, and is the only substitution mutation not to be located near the active site or a region known to be important for stability. If this residue was involved in protein-protein interactions and mutation resulted in AsbAS1 being unable to interact with its binding partner, this could destabilise the protein or activate regulatory mechanisms leading to its degradation.

As for the three mutants which produce full-length AsbAS1 protein, the reason for their lack of function must come at a stage after protein folding as disruption to AsbAS1 protein folding would have resulted in protein degradation as seen for the other substitution mutants. As none of these mutants produce the end product of the avenacin pathway despite having fully folded protein, it can only be concluded that the AsbAS1 protein is non-functional. It was therefore of interest to examine the areas of the protein where the mutations were located in order to gain further insights into residues or regions of the protein that were important for protein function.

To ascertain what effect each of the three mutations would have on protein function, the location of the mutation and its proximity to residues predicted to be involved the enzyme mechanism were examined in detail on the homology model of AsbAS1. The point mutation in mutant #358 was found to be in the active site close to the catalytic aspartate which is essential for function

and is conserved throughout all OSCs (Figure 4.8A) (52, 124, 210-211). The mutation was found to be in one of the cysteine residues (Cys563) which coordinates the catalytic aspartate and increases its acidity to allow it to protonate the epoxide group of the substrate (144, 212). Both cysteines are required for activity, as site-directed mutagenesis in *Saccharomyces cerevisiae* lanosterol synthase has shown. The cysteine adjacent to the catalytic aspartate (not corresponding to the mutation in mutant #358) was mutated to an aspartate residue to form the catalytic motif DDTAE, as is seen in bacterial squalene hopene cyclases, and this resulted in a significant reduction in enzyme activity (256). The aspartate is of a similar size to the cysteine and is able to form a hydrogen bond with the catalytic aspartate which may explain why enzyme activity is not completely abolished. In mutant #358, the non-adjacent cysteine to the catalytic aspartate is mutated to a tyrosine residue. This residue is much bulkier than the original cysteine residue so the vital hydrogen bond required for reaction initiation is unlikely to be formed. With the reaction unable to commence, the enzyme is rendered inactive and the avenacin pathway is terminated at this stage. This demonstrates that two cysteine hydrogen bonds are required for activation of the catalytic aspartate.

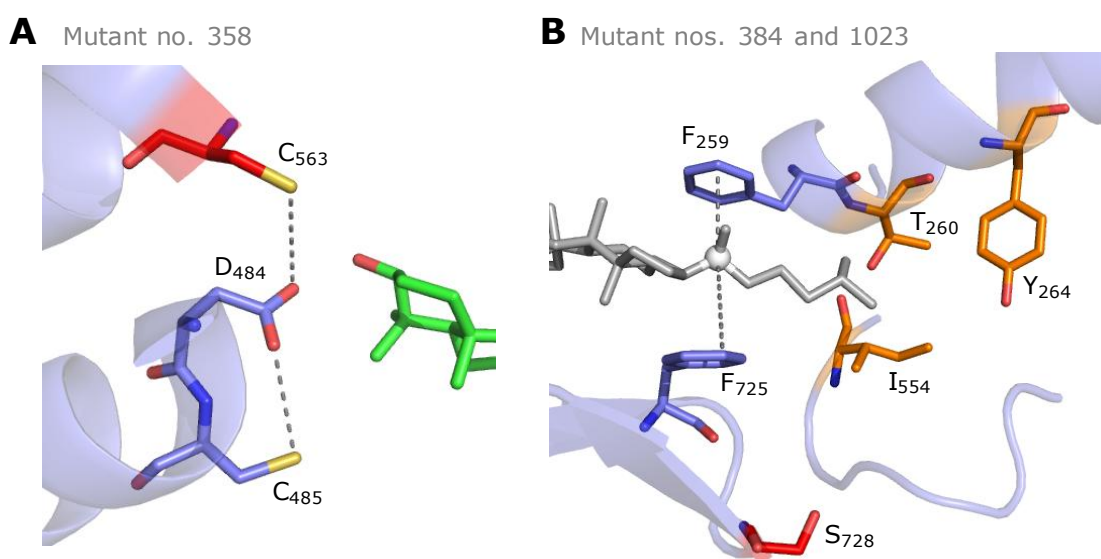


Figure 4.8: Effect of mutations in mutants #358, #384 and #1023 on enzyme function

Diagrams to illustrate the location of the residues affected by point mutations in mutants #358, #384 and #1023 and their proximity to conserved and important amino acid residues involved in the enzyme mechanism. The AsbAS1 tertiary structure is shown in purple. **A.** The residue affected by the point mutation in mutant #358. Mutation is found in a cysteine residue (shown in red) hydrogen bonded to the catalytic aspartate. Hydrogen bonds are shown by grey dashed lines. **B.** The residue affected by the point mutation in mutants #384 and #1023. Mutation is found in a serine residue (shown in red) which has close proximity to the three residues involved in allowing substrate access (shown in orange) and the Phe725 (shown in purple) which is involved in stabilisation of the D-ring during cyclisation. Cation- π interactions are shown by grey dashed lines.

Despite being independent from each other, mutants #384 and #1023 contain the same point mutation, which was also found to be in the active site close to two mechanistic features that are vital for correct enzyme function (Figure 4.8B). The mutation substitutes a serine residue (Ser728) for a phenylalanine. Ser728 is found to be in close proximity to the three residues (Tyr264, Cys260 and Ile554) postulated to block the substrate access channel which then undergo a conformational change to allow substrate entry into the active site (144). Substrate access is dependent on all of the residues being in the correct orientation so that when they change conformation the substrate can fit through the newly created opening. The loop on which Ser728 is located is adjacent to the loop containing isoleucine 554 (Ile554) so the position of the loop could easily be shifted by mutation of a small serine residue to a large bulky phenylalanine. If the orientation of Ile554 is altered then the substrate access may be permanently blocked resulting in a non-functional enzyme. Ser728 is also located three residues along from a phenylalanine residue (Phe725) which is involved in stabilising the positive charge formed upon D-ring closure during 2,3-oxidosqualene cyclisation to allow ring expansion to a 6-membered ring (144, 165). The positive charge is stabilised through cation- π interactions between the delocalised electrons of the phenylalanine and the positive charge formed on the newly cyclised ring of the intermediate (164). Phe725 works together with Phe259 to stabilise the positive charge, and stabilisation is dependent on the correct orientation of the delocalised electrons on the aromatic residues. AsbAS1 mechanism analysis has demonstrated that at least two aromatic residues are required for ring stabilisation, therefore substitution of Ser728 with a phenylalanine residue may affect the orientation of Phe725. If Phe725 is shifted to a position in which it is no longer able to stabilise the positive charge, Phe259 will not be able to provide enough negative potential to stabilise the charge so the cyclisation will be terminated prematurely resulting in a truncated tetracyclic product. This effect is seen in the sterol oxidosqualene cyclases which only have a single phenylalanine residue for D-ring stabilisation and as a result can only produce tetracyclic products as cyclisation is terminated at this stage (144). Also, site-directed mutagenesis experiments on *Panax ginseng* β -amyrin synthase where one of the D-ring stabilising phenylalanine residues is mutated have also found that tetracyclic products were produced (176). If this, rather than blocked substrate access, occurs in mutants #384 and #1023 then truncated tetracyclic products may be present in the root tips of these mutants, so this could be investigated in future experiments.

All, except one, of the amino acid substitution mutants, whether they produced folded protein or not, were found in the active site or in a region known to be important for stability (Figure 4.9). This pattern can be explained by the method used to select the mutants. Mutant selection was done using a loss-of-fluorescence screen, therefore only mutants which lacked the ability to express functional protein or produced folded proteins with mutations in critical regions would have been selected as they would not have produced a fluorescent product. There may have been mutants present in the mutagenised oat population that had single point mutations in the *Sad1* gene that resulted in no effect on enzyme function, but these would not have been selected for as they would have displayed a wild-type phenotype. This implies that OSC enzymes are very

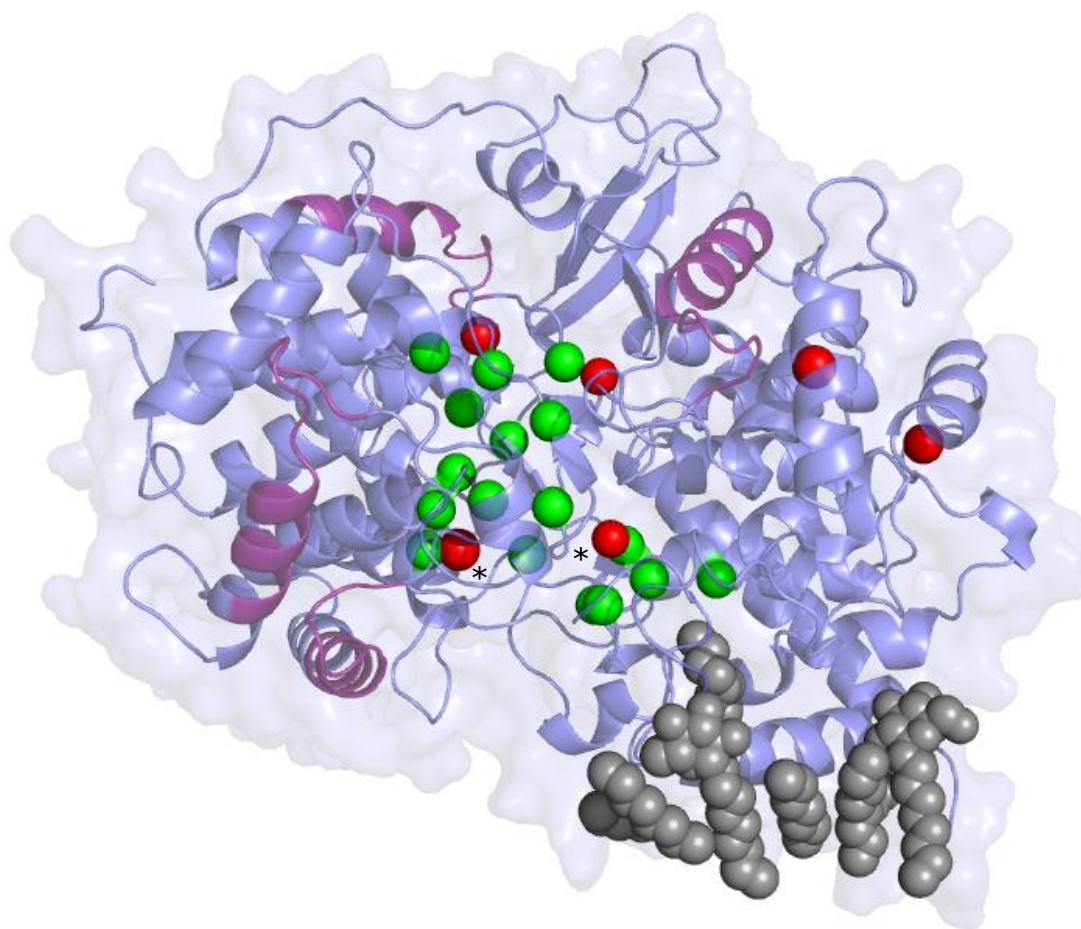


Figure 4.9: Location of amino acid substitution mutations on AsbAS1 homology model

Location of amino acid residues affected by amino acid substitutions on the AsbAS1 homology model. The tertiary structure of AsbAS1 is shown in lilac with the four QW motifs shown in purple. Detergent molecules are shown as grey spheres which represent the lipid bilayer and show how the enzyme is oriented in the membrane. Amino acid residues directly involved in the catalytic mechanism (initiation, stabilisation, deprotonation and substrate access) are shown as green spheres. Amino acid residues affected by the mutations are shown as red spheres and the mutations which produce full-length AsbAS1 protein are indicated by asterisks (*).

tolerant of mutations in areas of the enzyme that are not critical for function, and only those mutations that seriously affect enzyme expression or function that result in premature termination of the pathway.

Chapter 5 - Expression and purification of AsbAS1 in *Pichia pastoris*

5.1 Introduction

Plant terpene synthases synthesise a diverse range of secondary metabolites which have a wide range of functions in plants, for example, plant defence, pigmentation, scents, signalling and allelopathy. Many of these compounds have been exploited commercially as drugs and medicines (taxol – an anticancer drug from the yew tree), flavourings (glycyrrhizin – an artificial sweetener from licorice root), fragrances (eucalyptol – from the eucalyptus tree) and agrochemicals (rotenone – an insecticide from the vine plant). It is therefore of huge interest to both the scientific and commercial industries to characterise the enzymes involved in synthesising these compounds. Oxidosqualene cyclases (OSCs) are found at the initial stage of triterpenoid secondary metabolite synthesis and cyclise 2,3-oxidosqualene to a number of cyclic products, for example cycloartenol and β -amyrin, which are the cyclic backbones for all steroids and triterpene secondary metabolites, respectively.

5.1.1 Cloning and purification of oxidosqualene cyclases

As mentioned previously, a large number of sterol and triterpene OSCs have been cloned from mammalian, fungal and plant species. The majority of these OSCs have been heterologously expressed in yeast, and *in vitro* or *in vivo* assays coupled with GC-MS, LC-MS or HPLC have been used to determine their function and product specificity.

Despite their importance to the commercial industry, very few OSCs have been purified and only one, the *Homo sapiens* lanosterol synthase (hOSC), has been subjected to in depth biochemical and structural studies. Prior to the cloning of the OSCs, lanosterol synthase, cycloartenol synthase and β -amyrin synthase had been purified from their native organisms. OSCs had proved difficult to purify due to their membrane-bound nature, their instability and their insolubility in the absence of detergents. Lanosterol synthase from rat liver was the first OSC to be solubilised. This was achieved by isolating liver microsomes by centrifugation then solubilising the enzyme from the membrane using 0.08 M deoxycholate. The enzyme was partially purified by centrifugation and ether extraction and was found to be active, but it proved to be unstable and rapidly lost activity (257). Subsequent experiments found that solubilisation using 1%

Triton X-100 greatly improved the solubility and stability of the enzyme, which could then be stored at 0°C for two weeks without significant loss of activity (258). Lanosterol synthases have since been purified both partially and fully from *Saccharomyces cerevisiae* (211, 258-259), *Rattus norvegicus* (260) and *Bos taurus* (261). Cycloartenol synthase and β -amyrin synthase have also been partially purified from pea seedlings (262). All purification protocols involve solubilisation of the enzyme from microsomes using Triton X-100, followed by various purification steps that include ion-exchange chromatography, affinity chromatography, isoelectric focussing and gel filtration. Most of the enzymes were purified to homogeneity and retained activity but yields were relatively low, typically approximately 2 mg (2-5% final yield from the crude homogenate) (211, 260-261).

Heterologous overexpression of proteins in yeast or bacteria dramatically increases the amount of protein that can be produced in cultures and therefore increases yield. However attempts to overexpress OSCs in a bacterial system proved unsuccessful as the proteins mainly formed inclusion bodies possibly due to insolubility, improper folding or lack of post-translational modification (211). *Saccharomyces cerevisiae* lanosterol synthase has also been successfully expressed in insect cells. The majority of OSCs have been heterologously expressed in yeast, but despite the large number of OSCs cloned into a heterologous overexpression system, only hOSC has been successfully purified on a large scale and been subjected to kinetic analysis (151).

5.1.2 Purification of hOSC in *Pichia pastoris*

hOSC was heterologously expressed in the methylotrophic yeast *Pichia pastoris*. It was cloned into the pPICZB vector to give a construct with a C-terminal 6-histidine tag attached to the OSC protein and the construct was transformed into *Pichia pastoris* GS115 cells. Western blotting with an hOSC polyclonal antibody identified a single highly expressing clone that produced 3 mg hOSC per 1 g cell mass. For large-scale purification, 0.2% Triton X-100 was used to solubilise hOSC from the membrane after cell disruption and the C-terminal 6-histidine tag facilitated purification by metal-affinity chromatography. Detergent exchange from Triton X-100 to 0.8% β -D-octylglucopyranoside (β -OG) was carried out before elution of hOSC from the column and the fractions containing hOSC were further purified by gel filtration. The final yield of pure hOSC was 105 mg and subsequent analysis showed that hOSC was monomeric with a molecular mass of 80 kDa (151).

hOSC was successfully crystallised in a solution containing 2.4% β -OG and diffracted to a resolution of 2.1 Å. The enzyme was crystallised with both the product, lanosterol, and an inhibitor, Ro 48-8071, bound in the active site. The detergents used for solubilisation of the enzyme during the purification process were also visible in the final crystal structure and these detergent molecules identified a single amphipathic helix implicated in anchoring hOSC to the membrane (151).

5.1.3 Challenges in purifying membrane proteins

Currently there are just over 600 membrane protein crystal structures in the Protein Data Bank, which represent just 1% of the total number of crystal structures in the database. So why are structures of membrane proteins so scarce? The main problems lie in the stages of protein expression, protein purification and crystal growth. The first membrane protein crystals were grown using protein purified from natural sources rather than from heterologous expression. Compared to soluble proteins, after ribosomal synthesis many membrane proteins undergo complex processes involved in targeting and inserting proteins into the membrane. Over-expression of a membrane protein can overload the secretory pathways leading to formation of toxic intermediates or inclusion body formation (149, 263-265). Eukaryotic membrane proteins present an additional challenge as bacterial hosts have a very different translocation mechanisms and membrane composition and also lack post translational machinery (266). Other hosts, such as yeast, have been much more successful for eukaryotic membrane protein expression and purification. The methylotrophic yeast *Pichia pastoris* has been successfully used for heterologous expression of a number of membrane proteins, such as hOSC (151); multidrug resistance transporters (267-268) and G-protein coupled receptors (269-270).

After successful expression in a host, protein purification presents the next challenge and selection of the right detergent for solubilisation of the membrane protein is key. The detergent needs to stabilise the protein structure in an unaggregated state which may involve mixtures of detergents and non-detergent additives (271-272). Crystal growth is the final hurdle and is unpredictable. Crystals have been grown using detergent:protein complexes or by reintroducing the protein to a lipid bilayer environment (although the former method is more commonly used). Searching for the right conditions requires the correct choice of precipitant and must take into consideration the complex phase behaviour of the detergent as well as the protein (273). New methods for crystallisation of membrane proteins such as lipidic cubic phase and bicelle techniques are being

developed which may improve the success rate for membrane protein crystallisation (274-275).

5.1.4 Aims

To date no triterpene OSCs have been purified on a large scale from heterologous expression. The aim of this part of the project was to develop methods for purification of AsbAS1. As hOSC is currently the only OSC to be successfully purified from heterologous expression, the expression system and purification strategy were used as templates for AsbAS1 expression and purification. This involved cloning and heterologous functional expression of tagged AsbAS1 in *Pichia pastoris*, followed by expression and purification using metal-affinity chromatography. The ultimate aim was to generate sufficient pure protein to generate crystals and solve the structure of AsbAS1, which would provide important new insights into the mechanism of triterpene OSCs and enable the hypotheses made by the AsbAS1 homology model generated in Chapter 2 to be tested.

5.2 Materials and Methods

5.2.1 Cloning of *Sad1* into pPICZB vector

β -amyrin synthase was amplified by PCR using a full-length cDNA clone, AsbAS1-pYES2 (49) as a template. The following primers were used which contained *Pm*I and *Xba*I restriction sites for cloning into the pPICZB vector (*Pm*I site is shown in red, *Xba*I site is shown in blue, start and stop codons are shown in bold): pPICZB-5'-native = 5'GTACAT**CACGTG**ACC**ATGT**GAGGCTAACAATAGGT GAGG-3', pPICZB-3'-native = 5'-CATGTA**TCTAGA****TTAG**CTCTTAATCGCAAGAAGTCG ACGGC-3', pPICZB-3'-6His = 5'-CATGTA**TCTAGA****TTAG**TGATGGTGATGGTGATGGCT CTTAATCGCAAGAAGTCGACG-3', and pPICZB-5'-6His = 5'GTACAT**CACGTG**ACC **ATG**CATCACCATCACCATCACTGGAGGCTAACAATAGGTGAG-3'. Three different primer combinations were used to generate the constructs: pPICZB-BAS1 (native protein) using pPICZB-5'-native and pPICZB-3'-native; pPICZB-BAS2 (C-terminally 6-histidine tagged protein) using pPICZB-5'-native and pPICZB-3'-6His; and pPICZB-BAS4 (N-terminally 6-histidine tagged protein) using pPICZB-5'-6His and pPICZB-3'-native. PCR was carried out using *Pfu* Ultra DNA polymerase (Stratagene Europe) for 30 cycles (program: 95°C, 0.5min, 50°C, 1 min, 70°C, 3 min, and final extension at 72°C, 10 min). The presence of a correctly sized product was confirmed by 1% TBE agarose gel electrophoresis

with ethidium bromide staining, and the resulting DNA fragments were purified using a Wizard PCR Preps Kit (Promega). Using the two new restriction sites, the amplified DNA fragments were cloned into the pPICZB-vector (Invitrogen) and transformed by heat-shock into *Escherichia coli* DH5 α cells (Invitrogen) and spread onto Low Salt LB pH 7.5 plates (1% tryptone, 0.5% NaCl, 0.5% yeast extract, 1.5% agar) containing 25 $\mu\text{g}/\mu\text{l}$ zeocin for growth overnight at 37°C. Plasmid DNA was purified by Wizard SV Minipreps Kit (Promega) and digested with *EcoRI* (Invitrogen) for 1 hour at 37°C to confirm the presence of the *Sad1* coding sequence. Those *E. coli* transformants containing the *Sad1* coding sequence were sequenced.

5.2.2 Transformation into *Pichia pastoris*

Competent cells were prepared from a 500 ml culture of *Pichia pastoris* (*P. pastoris*) X-33 or KM71H strain in Yeast Peptone Dextrose (YPD) medium (1% yeast extract, 2% peptone, 2% dextrose) grown at 30°C with shaking to an $\text{OD}_{600} = 1.3\text{-}1.5$. Cells were harvested by centrifugation at 1,500 xg for 5 minutes and resuspended in 500 ml ice-cold sterile water. The resulting pellet was washed with 250 ml ice-cold sterile water, followed by 20 ml ice-cold 1 M sorbitol and finally 1 ml ice-cold 1 M sorbitol.

Plasmid DNA for each of the constructs was linearised with *SacI*, purified by phenol/chloroform extraction and ethanol precipitation and transformed into the competent *P. pastoris* cells by electroporation. Increasing volumes (10, 25, 50, 100, 200 μl) of the electroporated cells were grown on Yeast extract Peptone Dextrose+ (YPDS) medium (1% yeast extract, 2% peptone, 2% dextrose, 1 M sorbitol, 2% agar) plates containing 100 $\mu\text{g}/\text{ml}$ zeocin at 30°C. Up to ten colonies were purified on fresh YPDS plates containing 100 $\mu\text{g}/\text{ml}$ zeocin at 30°C.

The Mut phenotype of the *P. pastoris* transformants was determined by growing up to 10 purified colonies on Minimal Dextrose (MD) (1.34% YNB, $4 \times 10^{-5}\%$ biotin, 2% dextrose) and Minimal Methanol (MM) (1.34% YNB, $4 \times 10^{-5}\%$ biotin, 0.5% methanol) media agar plates at 30°C.

5.2.3 Small scale expression trials of AsbAS1 *P. pastoris* X-33 transformants

A single Mut⁺ colony was grown in 25 ml of Buffered Glycerol-complex (1% yeast extract, 2% peptone, 100 mM potassium phosphate pH 6.0, 1.34% YNB, $4 \times 10^{-5}\%$ biotin, 1% glycerol) media (BMGY) at 30°C with shaking at 250 rpm to an $\text{OD}_{600} = 2\text{-}6$. The cells were harvested by centrifugation at 5,000 xg and resuspended to an $\text{OD}_{600} = 1.0$ in 100 ml Buffered Methanol-complex (1%

yeast extract, 2% peptone, 100 mM potassium phosphate pH 6.0, 1.34% YNB, 4×10^{-5} % biotin, 0.5% methanol) media (BMMY) to induce protein expression. The culture was grown at 30°C with shaking at 250 rpm. Methanol was added to a final concentration of 0.5% every 24 hours to maintain protein induction. On day 3 of protein induction, 1 ml of the culture was taken for expression analysis. Cells were pelleted by centrifugation at 16,000 xg for 2 minutes, the supernatant was decanted to a fresh 1.5 ml eppendorf tube and 150 μ l 80% glycerol was added. Both supernatant and pellet were frozen in liquid nitrogen and stored at -80°C.

5.2.4 SDS-PAGE analysis of protein expression

Cell pellets were resuspended in 300 μ l Breaking buffer (50 mM sodium phosphate pH 7.4, 1 mM protease inhibitor, 1 mM EDTA, 5% glycerol) and lysed by sonication at 18 microns, using a Soniprep 150 (MSE), for 5 sets of 15 second pulses with a pause of 30 seconds on ice between each pulse. Triton X-100 (Boehringer Mannheim) was added to the cell lysates to a final concentration of 1% and incubated for 1 hour at 4°C with inversion. After centrifugation at 16,000 xg for 15 minutes, the supernatant was collected. For sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), 7 μ l of NuPAGE[®] LDS Sample Buffer (Invitrogen) was added to 20 μ l of supernatant, incubated at 70°C for 10 minutes and placed on ice for a further 5 minutes. 20 μ l of each sample was loaded onto a NuPAGE[®] Novex 4-12% Bis-Tris Gel (Invitrogen) placed in an XCell SureLock[™] Mini-Cell (Invitrogen). Both the upper and lower chambers were filled with 1x MES SDS Running Buffer (Invitrogen) and 0.5 ml NuPAGE[®] Antioxidant (Invitrogen) was added to the upper buffer chamber. The gel was electrophoresed at 200 V for 40 minutes and stained by incubation with InstantBlue (Novexin) for 1 hour.

5.2.5 Western blot analysis of protein expression

Samples were prepared and run on a NuPAGE[®] Novex 4-12% Bis-Tris Gel (Invitrogen) as described in section 4.2.4, then transferred to a nitrocellulose membrane using the XCell II[™] Blot Module at 30V constant for 1 hour. The membrane was washed in Tris-Buffered Saline (TBS) (20mM Tris-HCl pH 7.5, 7.5 mM NaCl) and incubated for 1 hour/overnight in blocking solution (3% Bovine Serum Albumin, 10 mM Tris-HCl pH 7.5, 7.5 mM NaCl). The membrane was washed with TBS with Tween-20 and Triton X-100 (TBSTT) (20 mM Tris-HCl pH 7.5, 0.5 mM NaCl, 0.05% v/v Tween[®]-20, 0.2% v/v Triton X-100) and TBS and

incubated with His•Tag[®] Monoclonal Antibody (Novagen) diluted 1:1000 in blocking solution for 1 hour. The membrane was washed again with TBSTT and TBS and incubated with Goat-Anti-Mouse IgG Alkaline Phosphatase conjugate (Novagen) diluted 1:5000 in blocking solution for 1 hour. The membrane was thoroughly washed in TBSTT and incubated with the colourimetric reagent BCIP/NBT solution (50 mM Tris-HCl pH 8.8, 14 mM MgCl₂, 0.1 mg/ml 4-nitro blue tetrazolium chloride (NBT), 0.05 mg/ml 5-bromo-4-chloro-3-indolyl-phosphate (BCIP)). The colourimetric reaction was stopped by washing the membrane thoroughly in deionised water.

5.2.6 GC-MS analysis of *in-vivo* protein activity

P. pastoris transformants were grown under inducing conditions as described in 4.2.3. On day 3 of protein induction the cells were harvested by centrifugation at 5,000 xg for 5 minutes. Cells were disrupted and saponified by resuspension in 20 ml Extraction Buffer (80% ethanol, 10% w/v potassium hydroxide, 0.5 mg/ml butylated hydroxytoluene (Sigma)) and incubated at 65°C for 2 hours. The mixture was diluted with 5 ml water and the products were extracted 3 times with 1 volume of hexane. The hexane fraction was washed three times each with 1 volume of water and 3 M NaCl solution. The hexane was then evaporated under reduced pressure and the resulting extract derivatised with Tri-Sil Z (Pierce) and analysed by gas-chromatography mass-spectrometry (GC-MS) on an Agilent 5973 MSD (Agilent, Stockport, UK) coupled to an Agilent 6890 Gas Chromatograph. The column used was a Phenomenex Zebron ZB-5 column (0.25 mm ID x 30 m, 0.25 µm film) (Macclesfield, UK) with a built-in 5 m guard column. The injector port, source and transfer line temperatures were set at 250°C and an oven temperature program from 170°C (2.0 min) to 340°C (3.0 min) at 6°C /min was used, with helium as the carrier gas at 0.8 ml/min constant flow. Splitless injections (1 µl) were made from the derivatised sample. EI+ mass spectral data were acquired for the duration of the GC program from m/z 50-800.

5.2.7 Large scale expression of AsbAS1

A single Mut⁺ colony was grown in 25 ml of Buffered Glycerol-complex (1% yeast extract, 2% peptone, 100 mM potassium phosphate pH 6.0, 1.34% YNB, 4x10⁻⁵% biotin, 1% glycerol) media (BMGY) at 30°C with shaking at 250 rpm to an OD₆₀₀ = 2-6. The 25 ml culture was used to inoculate 500 ml of BMGY media and grown at 30°C with shaking at 250 rpm until the culture reached log

phase growth ($OD_{600} = 2-6$). The cells were harvested by centrifugation at 5,000 xg and resuspended to an $OD_{600} = 1.0$ in 5 litres Buffered Methanol-complex (1% yeast extract, 2% peptone, 100 mM potassium phosphate pH 6.0, 1.34% YNB, 4×10^{-5} % biotin, 0.5% methanol) media (BMMY) to induce protein expression. The culture was grown at 30°C with shaking at 250 rpm. Methanol was added to a final concentration of 0.5% every 24 hours to maintain protein induction. On day 3 of protein induction, 1 ml of the culture was taken for expression analysis. Cells were pelleted by centrifugation at 5,000 xg for 10 minutes and stored on at 4°C or frozen at -80°C until required.

5.2.8 Extraction and solubilisation of AsbAS1 protein

The following protocol was adapted from Ruf *et al.* (151). Cells harvested from the 5 litre culture were resuspended in 250 ml Resuspension buffer (50 mM Tris-HCl pH 7.5, 5% glycerol (v/v), 2 mM DTT, 2 mM $MgCl_2$) supplemented with 1 tablet of Complete protease inhibitor (Roche) per 50 ml. The cell suspension was disrupted with $P_g = 1000$ psi in a French pressure cell press (SLM Aminco) and then sonicated for 10 minutes at 18 microns on a Soniprep 150 (MSE). Triton X-100 (Boehringer Mannheim) was added to the crude extract to a final concentration of 0.2% (v/v) and incubated for 1 hour at 4°C with stirring. Unbroken cells and cell fragments were removed by centrifugation at 30,000 xg and the supernatant was stored at 4°C until required.

5.2.9 Immobilised metal affinity chromatography

Purification was carried out using an AKTA Purifier (Amersham Biosciences). The supernatant was loaded, using a 150 ml Superloop (GE Healthcare), onto a HiTrap™ Chelating HP 5ml column (GE Healthcare) equilibrated with Loading buffer (50 mM Tris-HCl pH 7.5, 5% glycerol (v/v), 2 mM DTT, 2 mM $MgCl_2$, 0.2% Triton X-100 (v/v)). The column was washed with 1 column volume (CV) of loading buffer and 5 CVs of Wash buffer (50 mM Tris-HCl pH 7.5, 5% glycerol (v/v), 2 mM DTT, 2 mM $MgCl_2$, 0.8% β -D-octylglucopyranoside (β -OG) (w/v) (Sigma). Bound protein was eluted with a linear gradient of Elution buffer (50 mM Tris-HCl pH 7.5, 5% glycerol (v/v), 2 mM DTT, 2 mM $MgCl_2$, 0.5 M imidazole) and 5 ml fractions were collected on a Frac-920 (Amersham Biosciences). Fractions with an absorbance at 280 nm were analysed by SDS-PAGE and Western blotting. Fractions containing AsbAS1 were pooled and concentrated using an Amicon Ultra-15 Centrifugal Filter Unit 30,000 MWCO (Millipore) and stored at -80°C until required

5.2.10 Gel filtration

Gel filtration was carried out using an AKTA Purifier. The IMAC purified concentrated protein solution was loaded at 0.5 ml/minute onto a Superdex 200 16/60 120 ml column (GE Healthcare) equilibrated with Gel filtration buffer (50 mM Tris-HCl pH 7.5, 5% glycerol (v/v), 2 mM DTT, 2 mM MgCl₂, 0.15 M NaCl). The column was washed with 1 CV of Gel filtration buffer and 1 ml fractions were collected on a Frac-920. Fractions were analysed by SDS-PAGE.

5.2.11 Hydrophobic interaction chromatography

Hydrophobic interaction chromatography (HIC) was carried out using an AKTA purifier. The IMAC purified concentrated protein solution was diluted with 1 volume of Buffer C (0.1 M sodium phosphate buffer pH 7.6, 2 M (NH₄)₂SO₄). It was then loaded at 0.5 ml/minute onto a Phenyl Sepharose Fast Flow (low-sub) 25 ml column (GE Healthcare) equilibrated with 5 CV of Buffer B (50 mM sodium phosphate buffer pH 7.6) followed by 5CV of Buffer A (0.1 M sodium phosphate buffer pH 7.6, 1 M (NH₄)₂SO₄). The column was washed with Buffer A until unbound protein had been eluted and then bound protein was eluted with a linear gradient of Buffer B over 5 CVs and 0.25 ml fractions were collected on a Frac-920. Fractions were analysed by SDS-PAGE.

5.3 Results and Discussion

5.3.1 *Pichia pastoris* expression system

Both prokaryotic and eukaryotic systems have been used for overexpression of oxidosqualene cyclases (OSCs) (151, 211). Attempts to express *Saccharomyces cerevisiae* lanosterol synthase (ScLS) in *Escherichia coli* (*E. coli*) failed to produce active protein despite a large number of strain and vector combinations being used. Prokaryotes do not contain the post-translational modification machinery that is present in eukaryotes so this may explain why active enzyme was not produced. Subsequent attempts to overexpress ScLS in its native organism did result in active protein, however the expression levels were very low (211).

The expression system chosen for heterologous expression of AsbAS1 was the yeast *Pichia pastoris*. *Pichia* is an attractive system for heterologous protein expression as it maintains many of the properties seen in other higher eukaryotic expression systems, such as protein folding, protein processing and post-translational modification, which are not found in prokaryotic expression systems

such as *E. coli*. In addition to this, it has also been shown to produce higher expression levels than other yeast systems, with expression levels 10- to 100-fold higher than the *Saccharomyces cerevisiae* expression system. The *P. pastoris* expression system has been successfully utilised for the functional expression of other plant OSCs for activity assays (123, 151), purification (151) and crystallisation (144).

P. pastoris is a methylotroph which gives it the ability to use methanol as its sole carbon source. Methanol is oxidised to formaldehyde and hydrogen peroxide in the first step of the methanol utilisation pathway. Hydrogen peroxide is highly toxic to cells, so the reaction takes place in peroxisomes where it is detoxified by catalase. Methanol oxidation is catalysed by the enzyme alcohol oxidase (AOX) which has a very low affinity for oxygen, so AOX expression is upregulated to produce large quantities of the enzyme. AOX is encoded by two genes, *AOX1* and *AOX2*, although it is the gene product of the former that provides the majority (85%) of the AOX activity (276-278). Expression of AOX is tightly regulated at the transcriptional level and the promoter is highly inducible by methanol (279). This property has been exploited for expression of foreign proteins in *P. pastoris*, and a plasmid-borne version of the *AOX1* promoter has been developed to drive heterologous protein expression in *P. pastoris* (279-281).

pPICZ vectors were used for heterologous expression of AsbAS1. This set of vectors has three different sub-types (A, B and C) which contain different restriction sites to allow the target gene to be cloned in-frame with the C-terminal tags which are present in the vector. The vectors have a C-terminal *myc* epitope tag and a C-terminal polyhistidine tag to enable detection purification and detection of the expressed protein. For positive selection in both *E. coli* and *P. pastoris*, all vectors contain the *Sh ble* gene which confers resistance to the antibiotic Zeocin™. Zeocin™ is a water-soluble copper-chelated glycopeptide antibiotic with a broad spectrum against bacteria and eukaryotes. Copper-bound Zeocin™ is inactive, but upon cell entry it is activated by removal of the copper cation by sulfhydryl compounds. It then causes cell death by binding and cleaving cellular DNA. Zeocin™ is light sensitive and inactivated by high salt concentrations, so reduced salt media must be used in cultures and incubation and storage must be in dark conditions (282). The *Sh ble*, Zeocin™ resistance, gene was isolated from *Streptoalloteichus hindustanus* and inhibits Zeocin™ by binding to it in a stoichiometric manner which prevents its DNA cleavage activity (283).

5.3.2 Design of expression constructs

Three constructs were designed for heterologous expression of oat β -amyrin synthase (AsbAS1) in *P. pastoris* with the aim of producing large quantities of protein that could be easily purified, for use in kinetic and structural studies. Constructs consisted of the full-length *Sad1* cDNA sequence with no tag (BAS1), C-terminal 6-histidine tag (BAS2) or N-terminal 6-histidine tag (BAS4) (Figure 5.1). 6-histidine tags were chosen as a C-terminal 6-histidine tag had been successfully used to purify the human lanosterol synthase (hOSC) (151). None of the constructs utilised the C-terminal 6-histidine and *myc* epitope tags originally contained in the vector as the restriction sites chosen resulted in a 60bp gap between the C-terminus of the protein and the end of both tags. The resulting size of the tag may have hindered crystallisation and as only a histidine tag was required for purification, 6-histidine tags were added directly at the end or start of the appropriate constructs by PCR. pPICZB was chosen as the vector for heterologous expression of AsbAS1 as it contained the desired *Pm*I and *Xba*I restriction sites.

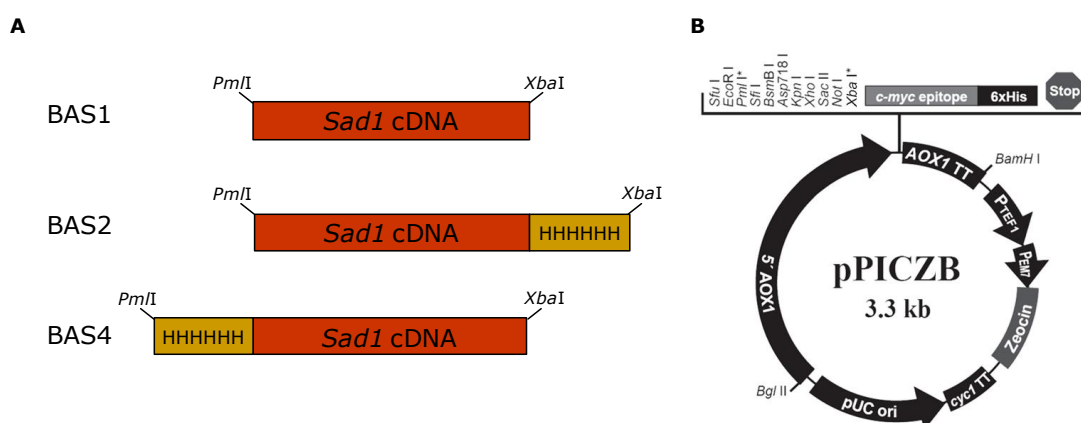


Figure 5.1: Design of constructs for β -amyrin synthase expression

A. Three constructs designed for expression of *Sad1* in *Pichia pastoris*. All constructs contain the full length *Sad1* cDNA sequence with and without 6-histidine tags. BAS1 does not have a tag, BAS2 and BAS4 have a 6-histidine tag at the C- and N-terminus respectively. The restriction sites used for cloning into the pPICZB vector are shown. **B.** Plasmid map of pPICZB (3.3 kb). 5'*AOX1* – *AOX1* promoter; *AOX1* TT – *AOX1* transcription termination and polyadenylation signal; P_{TEF1} – Transcription elongation factor 1 gene (for *Sh ble* expression); P_{EM7} – constitutive promoter which drives *Sh ble* expression; Zeocin – *Sh ble* gene for Zeocin™ resistance; *cyc1*TT – 3' end of *Saccharomyces cerevisiae* *CYC1* gene for efficient 3' processing of *Sh ble* gene; pUC ori – pUC origin. Multiple cloning site is shown between 5'*AOX1* and *AOX1* TT and restriction sites used for *Sad1* cloning are denoted by *.

5.3.3 Cloning of *Sad1* into the pPICZB vector

The full-length *Sad1* cDNA sequence was amplified by PCR using the primers required to add the desired restriction sites and tags for each of the 3 constructs. The products were analysed by 1% TBE gel electrophoresis to

confirm the presence of a correctly sized DNA fragment (Figure 5.2A). The gel showed a single band between the 2.0 and 3.0kb markers for all constructs which corresponded to the size of the *Sad1* coding sequence, ~2.3kb.

The *Sad1* DNA fragments were then digested with *Pml*I and *Xba*I and ligated into the pPICZB vector to generate the constructs pPICZB-BAS1, pPICZB-BAS2 and pPICZB-BAS4, and transformed into *E. coli* DH5 α cells. Plasmid DNA was isolated from the transformed cells and digested with *Eco*RI to confirm presence of *Sad1* cDNA. The *Sad1* coding sequence has two *Eco*RI sites 1kb apart and pPICZB has one *Eco*RI site situated ~200bp upstream of the *Xba*I restriction site, so successful integration of the *Sad1* gene into the pPICZB vector would generate fragments of 4.4kb, 1kb and 0.2kb. All three constructs had *E. coli* transformants which produced the desired fragment sizes after digestion with *Eco*RI (Figure 5.2B). Sequencing of the *Sad1* cDNA in each of these *E. coli* transformants confirmed that the gene had been successfully cloned into the pPICZB vector.

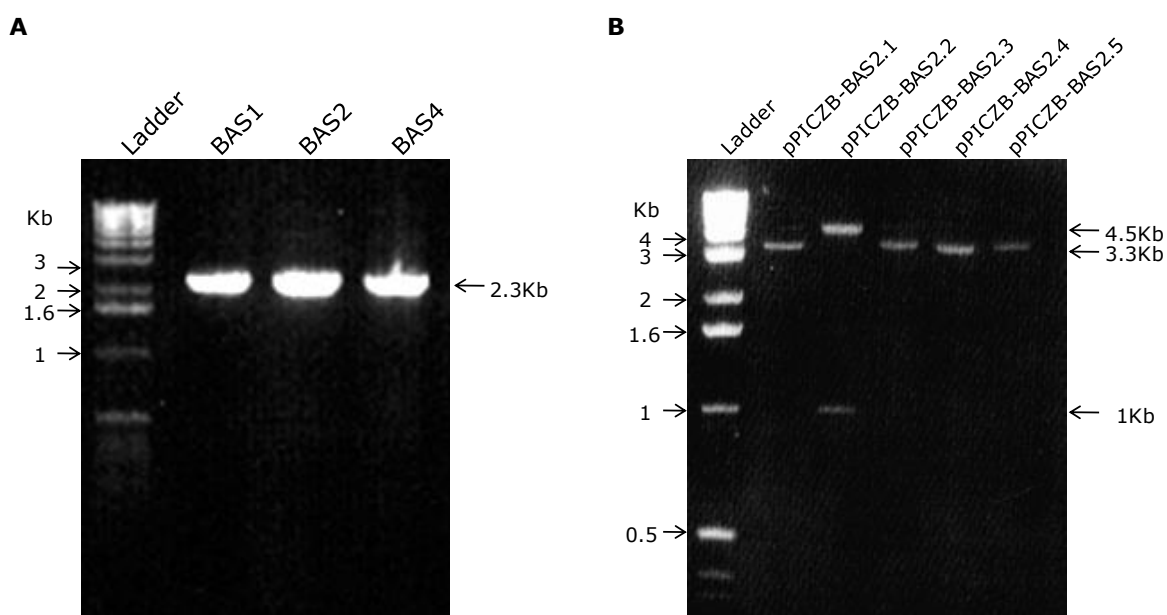


Figure 5.2: Cloning of *Sad1* into pPICZB vector

A. PCR amplification of *Sad1* from full-length cDNA clone using primers to add desired restriction sites and 6-histidine tags for the three constructs. The amplified PCR products were run on a 1% TBE agarose gel and visualised with EtBr staining. **B.** *Eco*RI digestion of plasmid DNA extracted from five pPICZB-BAS2 transformants to identify constructs containing the *Sad1* gene. Plasmids containing the *Sad1* gene have bands at 4.5 kb and 1 kb (N.B the 0.2 kb band is not visible) whereas plasmids that do not contain the *Sad1* gene have a single band at 3.3 kb. The digested plasmid DNA was run on a 1% TBE agarose gel and visualised with EtBr staining.

5.3.4 Expression of pPICZB-BAS1, BAS2 and BAS4 in *Pichia pastoris* X-33

For protein expression, the three pPICZB constructs were transformed into *Pichia pastoris* wild-type strain X-33 cells. X-33 was chosen as it has good growth levels in large-scale cultures and has been used for heterologous expression of other plant proteins (284-286).

Linearised pPICZB-BAS1, pPICZB-BAS2 and pPICZB-BAS4 plasmid DNA were transformed into competent *P. pastoris* X-33 cells via electroporation. Ten colonies were purified on fresh plates for each construct to select for single transformants. The methanol utilisation (Mut) phenotype of each colony was then determined. The linearised plasmid DNA is integrated into the *Pichia* genome via homologous recombination which can result in either gene insertion or gene replacement. Constructs generated in pPICZB vectors integrate into the *Pichia* genome at the *AOX1* locus by crossover events between the either of the two *AOX1* regions on the pPICZB vector, 5'*AOX1* or *AOX1* TT (Figure 5.1). Single crossover events result in insertion of the gene of interest upstream or downstream of the *AOX1* gene. These transformants maintain a functional copy of the *AOX1* gene and therefore have a wild type Mut phenotype, designated Mut⁺. Double crossover events result in replacement of the *AOX1* gene by the gene of interest. As the *AOX1* gene has been replaced, these transformants have to rely solely on the weaker *AOX2* gene for methanol metabolism and therefore consume methanol at a slower rate so have a Mut slow phenotype, designated Mut^S (277, 287-288). Mut phenotype is tested by growth of the transformant on MD and MM media agar plates. MD media agar plates have dextrose as the sole carbon source so both Mut⁺ and Mut^S transformants should grow well. MM media agar plates have methanol as the sole carbon source, therefore Mut⁺ transformants should grow well on these plates, whereas Mut^S transformants, which can only utilise the *AOX2* gene, will grow at a much slower rate (Figure 5.3). Each of the three constructs had nine independent *P. pastoris* X-33 transformants. Growth comparisons on MD and MM media agar showed all nine pPICZB-BAS1:X-33 transformants, five pPICZB-BAS2:X-33 transformants and eight pPICZB-BAS4:X-33 transformants had a Mut⁺ phenotype, whilst the remainder displayed a Mut^S phenotype. As Mut⁺ transformants are more desirable for protein expression, all Mut⁺ *P. pastoris* X-33 transformants were carried forward to small-scale expression trials.

Expression trials were conducted on all Mut⁺ *P. pastoris* X-33 transformants for each of the three constructs. Protein expression was induced by resuspension in 100 ml methanol growth media (BMMY) and a 1 ml sample

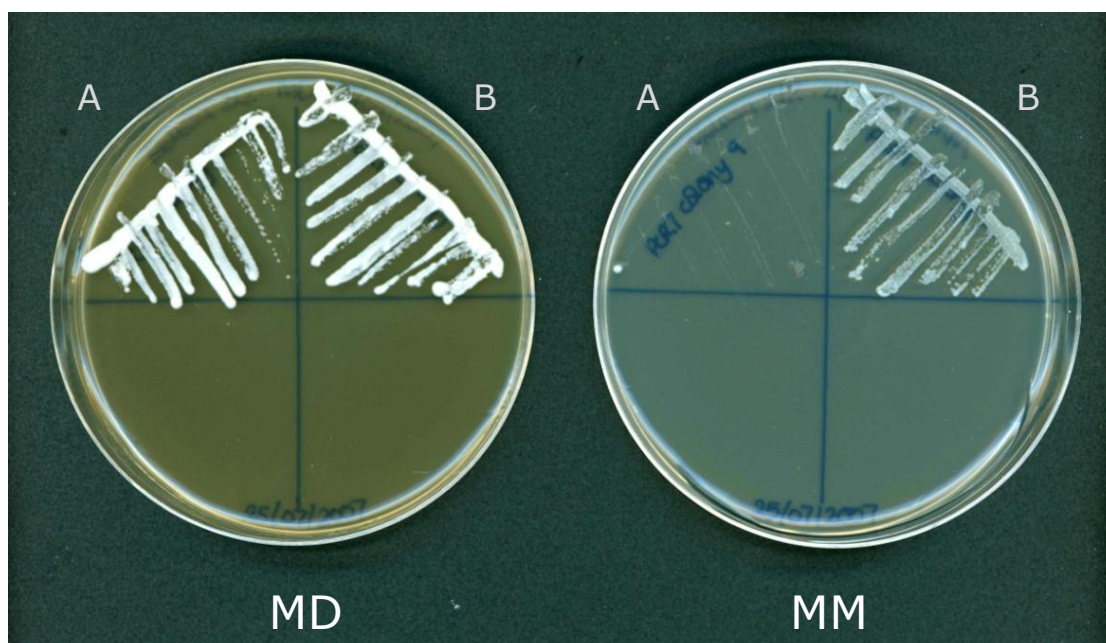


Figure 5.3: Determination of Mut phenotype of *P. pastoris* X-33 transformants

P. Pastoris X-33 cells transformed with each of the three pPICZB constructs were grown on minimal dextrose (MD) and minimal methanol (MM) media agar to determine the methanol utilisation (Mut) phenotype of each colony. Examples of the two phenotypes are shown. Colony A has a Mut^S phenotype as it grows well on MD media but grows poorly on MM media. Colony B has a Mut⁺ phenotype as it grown well on both MD and MM media.

was taken from each culture after three days of protein induction. Cell pellets were disrupted by sonication and the lysate was incubated with Triton X-100 (1%). Detergent addition was essential in order to release the membrane-bound AsbAS1 into the supernatant. Triton X-100 had been used successfully during the purification of *Homo sapiens* lanosterol synthase (hOSC) and in *in-vitro* assays for many other plant and fungal OSC, so was selected as the detergent used for AsbAS1 solubilisation in pPICZB-BAS1, pPICZB-BAS2 and pPICZB-BAS4 lysates (52, 57, 94, 101-102, 104). The cell-free extracts of pPICZB-BAS1, pPICZB-BAS2 and pPICZB-BAS4 *P. pastoris* X-33 transformants were then analysed using SDS-PAGE to look for expression of AsbAS1 protein.

5.3.4.1 pPICZB-BAS1:X-33

SDS-PAGE analysis showed no significant difference between the nine Mut⁺ pPICZB-BAS1:X-33 transformants and a pPICZB:X-33 vector only control. The predicted size of AsbAS1 is 86.8kDa so should be seen between the 62 and 98kDa markers on the gel if it has been induced in any of the pPICZB-BAS1:X-33 transformants. However, as all pPICZB-BAS1:X-33 transformants are Mut⁺ and under the control of the *AOX1* promoter, *Pichia pastoris* alcohol oxidase (PpAOX1)

would be co-expressed with AsbAS1. The molecular weight of PpAOX1 is approximately 75kDa which is approximately 12kDa smaller than AsbAS1. The theoretical calculations imply that AsbAS1 and PpAOX1 could be separated on an SDS-PAGE gel, however SDS-PAGE analysis of hOSC, which is predicted to have a molecular weight of 85.6 kDa, showed that it migrates at a distance of 80 kDa, 6 kDa lower than predicted (151). If AsbAS1 travels in a similar way to hOSC on an SDS-PAGE gel it may migrate a similar distance on an SDS-PAGE gel to PpAOX1 which may obscure the AsbAS1 band. At the time, the AsbAS1 antibody was not available, and as this construct did not have a tag, presence of AsbAS1 protein could not be detected by SDS-PAGE analysis alone.

The nine pPICZB-BAS1:X-33 transformants were analysed for production of AsbAS1 protein by function expression assays which are described in a later section.

5.3.4.2 pPICZB-BAS2:X-33 and pPICZB-BAS4:X-33

Analysis of the five pPICZB-BAS2:X-33 and eight pPICZB-BAS4:X-33 Mut⁺ transformants was more conclusive as both constructs contained 6-histidine tags at the C- and N-terminus respectively, so cell-free extracts could be analysed by Western blotting as well as SDS-PAGE.

One out of the five pPICZB-BAS2:X-33 transformants (pPICZB-BAS2.1:X-33) had a strong band present between the 62 and 98kDa markers that was not seen in the pPICZB:X-33 vector only control (Figure 5.4). As mentioned previously, due to the co-expression and similar molecular weight of AsbAS1 and PpAOX1, it was possible that the band observed was alcohol oxidase, therefore a Western blot using the His•Tag[®] monoclonal antibody was carried out to detect whether C-terminally tagged AsbAS1 was present. The Western blot detected a very faint signal which corresponded to the location of the band seen on the SDS-PAGE gel which confirms that AsbAS1 has been expressed from pPICZB-BAS2.1:X-33 (data not shown). The faint signal seen on the Western blot does not correspond to the intensity of the band seen on the SDS-PAGE gel, therefore it is likely that this band also contains the co-expressed PpAOX1 which appears to be expressed at much higher levels than AsbAS1.

Two out of the eight pPICZB-BAS4:X-33 transformants (pPICZB-BAS4.4:X-33 and pPICZB-BAS4.5:X-33) had a band present between the 62 and 98kDa markers that was not seen in the pPICZB:X-33 vector only control (Figure 5.5). A Western blot using the His•Tag[®] monoclonal antibody detected a faint signal

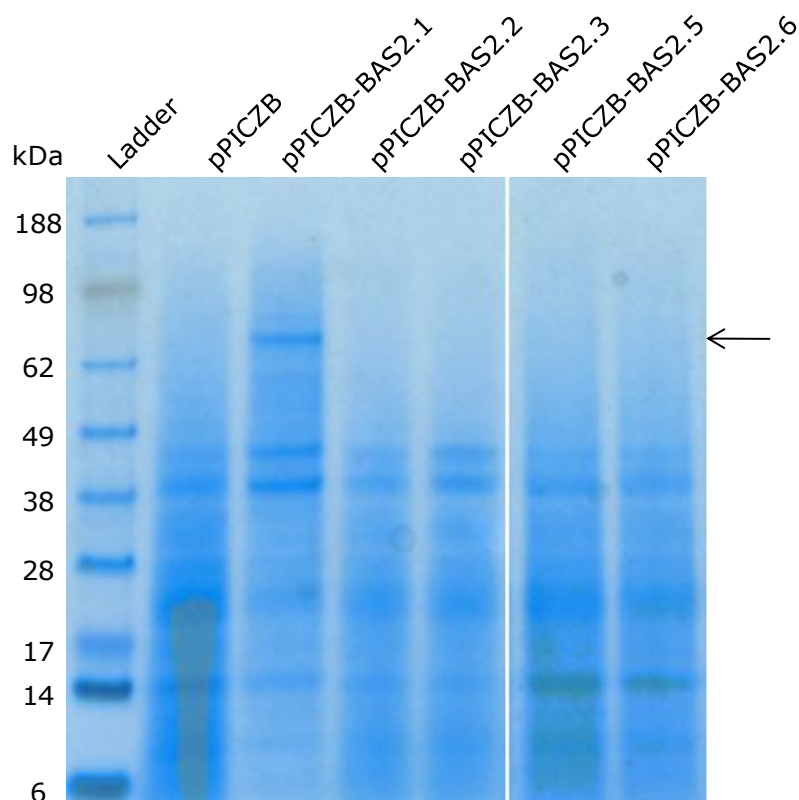


Figure 5.4: InstantBlue stained SDS-PAGE gel of cell-free extracts from trial expression of pPICZB-BAS2:X-33 Mut⁺ transformants.

SDS-PAGE gel of Triton X-100 treated cell-free extracts from Day 3 of the trial expression of pPICZB-BAS2:X-33 Mut⁺ transformants. Lane 1, SeeBlue[®] Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), myoglobin red (17kDa), lysozyme (14kDa), aprotinin (6kDa); Lane 2, pPICZB:X-33 vector only control cell-free extract; Lanes 3-7, pPICZB-BAS2:X-33 Mut⁺ transformants cell-free extracts. Expected molecular weight of AsbAS1 (shown by the arrow) is predicted to be 86.8kDa, which falls between the 62 and 98kDa markers.

corresponding to the location of the band seen on the SDS-PAGE gel for pPICZB-BAS4.4:X-33 but not for pPICZB-BAS4.5:X-33 (data not shown). As for pPICZB-BAS2.1:X-33, the signal from the Western blot did not correspond to the intensity of the band on the gel, so the band is also likely to contain the co-expressed PpAOX1. The lack of signal for pPICZB-BAS4.5:X-33 may be a result of insufficient AsbAS1 protein loaded onto the gel rather than lack of AsbAS1 expression as the band seen on the SDS-PAGE gel is much weaker than that of pPICZB-BAS4.4:X-33.

From these preliminary protein expression results, it seems that both pPICZB-BAS2:X-33 and pPICZB-BAS4:X-33 constructs have a single transformant which produced AsbAS1 protein to detectable levels on a Western blot. Both transformants were then analysed to determine whether the expressed protein in each of the transformants was active.

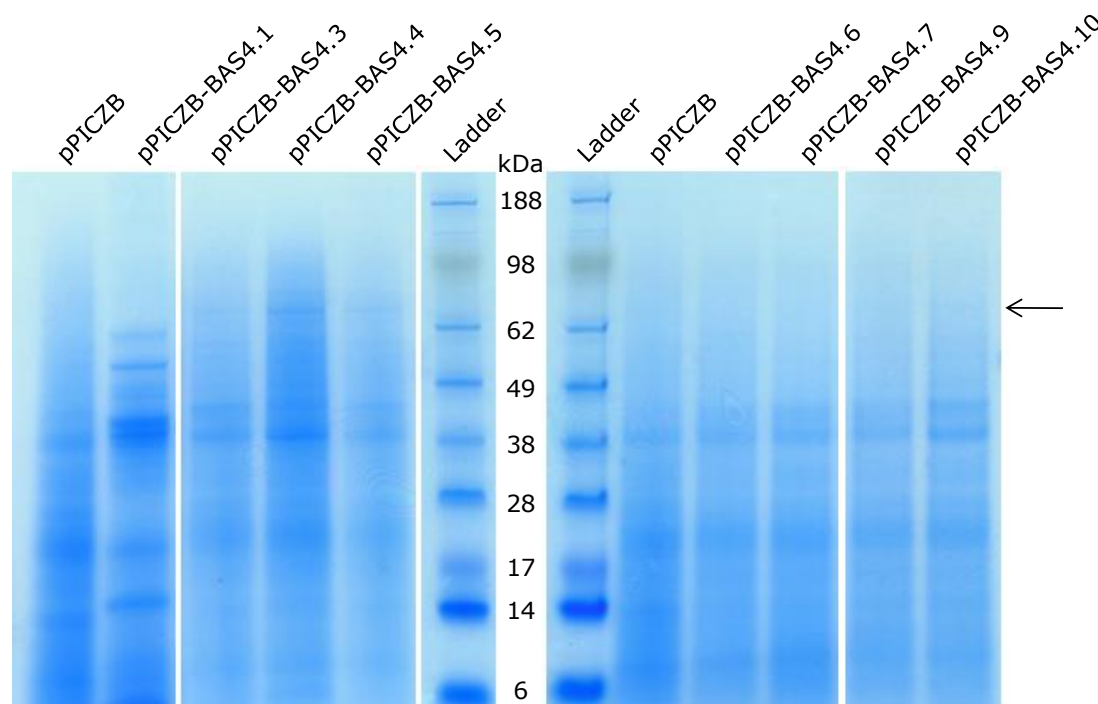


Figure 5.5: InstantBlue stained SDS-PAGE gel of cell-free extracts from trial expression of pPICZB-BAS4:X-33 Mut⁺ transformants.

SDS-PAGE gel of Triton X-100 treated cell-free extracts from Day 3 of the trial expression of pPICZB-BAS4:X-33 Mut⁺ transformants. Lane 1, pPICZB:X-33 vector only control cell-free extract; Lane 2-5, pPICZB-BAS2:X-33 Mut⁺ transformants cell-free extracts; Lane 6-7, SeeBlue[®] Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), myoglobin red (17kDa), lysozyme (14kDa), aprotinin (6kDa); Lane 8, pPICZB:X-33 vector only control cell-free extract; Lane 9-12, pPICZB-BAS4:X-33 Mut⁺ transformants cell-free extracts. Expected molecular weight of AsbAS1 (shown by the arrow) is predicted to be 86.8kDa, which falls between the 62 and 98kDa markers.

5.3.5 In-vivo functional assay of AsbAS1 activity in *Pichia* X-33 transformants

Protein expression analysis in *P. pastoris* X-33 transformants with 6-histidine tagged constructs had identified two transformants (pPICZB-BAS2.1:X-33 and pPICZB-BAS4.4:X-33) expressing AsbAS1 protein. These two transformants were then tested using an *in-vivo* assay to determine whether the expressed protein was active. All nine Mut⁺ pPICZB-BAS1:X-33 constructs were also analysed using the *in-vivo* assay as protein could not be detected via SDS-PAGE analysis.

Pichia pastoris cells produce the membrane sterol ergosterol as their major sterol product by cyclisation of 2,3-oxidosqualene to lanosterol, which is catalysed by lanosterol synthase. They do not, however, produce any secondary metabolite triterpenes such as β -amyrin, therefore any β -amyrin detected in the transformants would be a direct result of the heterologously expressed AsbAS1 activity. The enzyme can use the endogenous 2,3-oxidosqualene as a substrate

so the enzyme activity of AsbAS1 can be assayed *in-vivo* without feeding a substrate (123, 289).

Each transformant was grown using the same protocol as for the small-scale expression trials. After disruption and saponification of the cell pellets, triterpenes were extracted using hexane and analysed by GC-MS. Raw GC-MS data was analysed using the AMDIS software package which removes background noise from the raw data, extracts individual component spectra and compares them to a library of common triterpenes (290). The total ion chromatogram (TIC) of each clone was compared to the TIC of the pPICZB:X-33 vector only control to identify novel peaks. The spectra of any novel peaks were compared to the library of common triterpenes to identify the product(s) that had been produced.

The nine pPICZB-BAS1:X-33 Mut⁺ transformants together with the pPICZB-BAS2.1:X-33 and pPICZB-BAS4.4:X-33 transformants selected from small scale expression trials were then tested for enzyme function using the *in vivo* assay.

5.3.5.1 pPICZB-BAS1:X-33

Of the nine pPICZB-BAS1:X-33 Mut⁺ transformants analysed, all produced a novel peak which had a fragmentation pattern identical to that of β -amyrin. Figure 5.6A shows the AMDIS analysis of one of the transformants pPICZB-BAS1.6:X-33 compared to the pPICZB:X-33 vector only control. Both TICs have a peak at a retention time of 26.6 minutes and a fragmentation pattern corresponding to that of ergosterol, the primary sterol in *Pichia pastoris*. pPICZB-BAS1.6:X-33 contains a novel peak with a retention time of 27.7 minutes that is not observed in the pPICZB:X-33 vector only control. The extracted spectrum of this novel peak is shown in Figure 5.6B and displays the characteristic fragmentation pattern of β -amyrin, with the molecular ion at m/z 426, and ions at m/z 393, 218 and 203, thereby confirming that the compound in the novel peak to be β -amyrin. As stated previously, *Pichia pastoris* cells do not make secondary metabolite triterpenes, therefore the production of β -amyrin is a direct result of heterologously expressed AsbAS1 activity (123, 289). The same analysis was conducted for the other eight pPICZB-BAS1:X-33 transformants (data not shown) and all had a novel peak corresponding to β -amyrin, thereby confirming that all transformants contained active AsbAS1 protein.

In order to determine which transformant had the highest levels of expressed protein, and could therefore be carried forward to large-scale

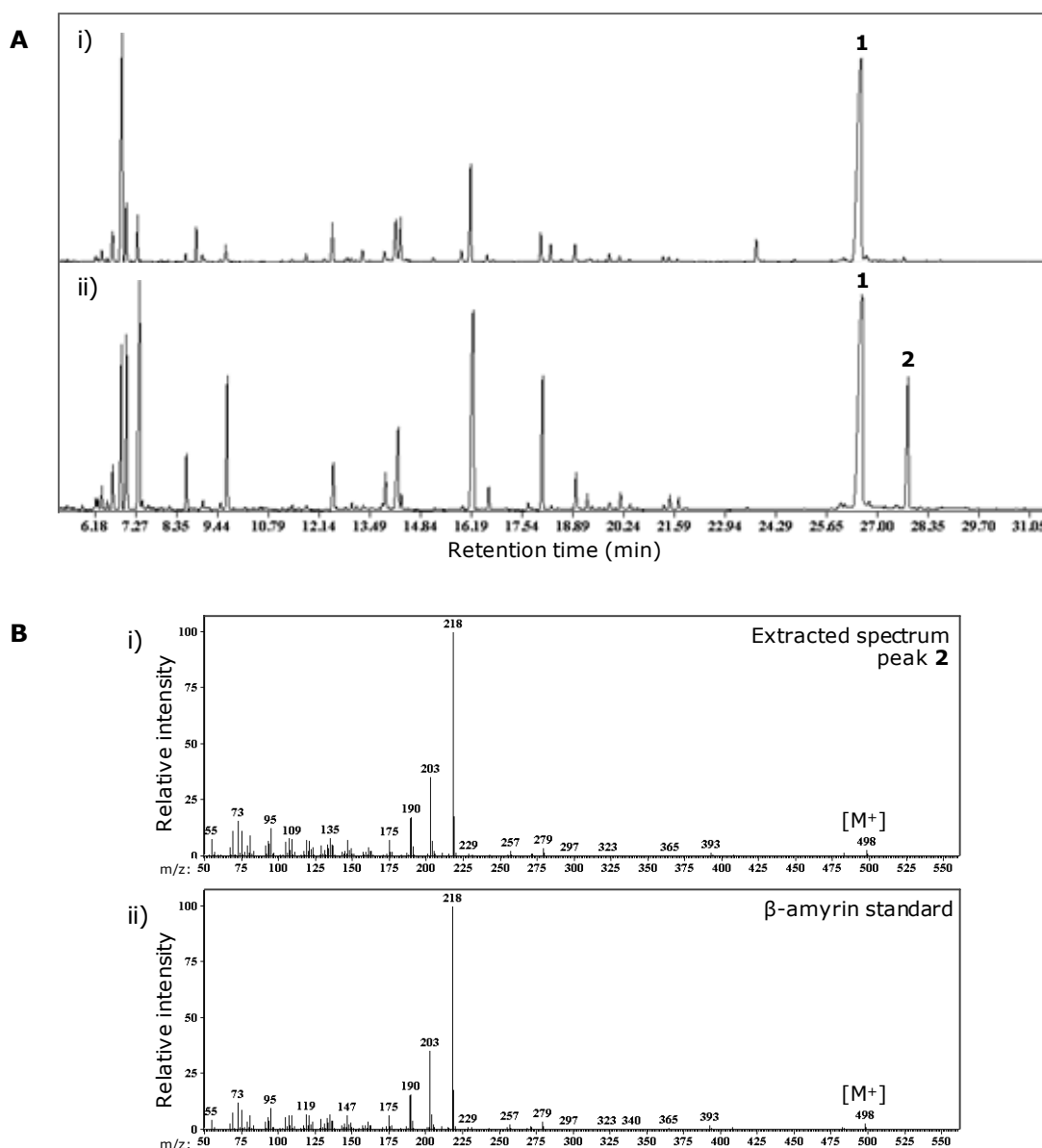


Figure 5.6: GC-MS analysis of pPICZB-BAS1:X-33

A. TIC chromatogram of hexane extracts from *Pichia* transformants with (i) pPICZB:X-33 as an empty vector control and (ii) pPICZB-BAS1.6. pPICZB-X-33 control shows only presence of the primary membrane sterol, ergosterol (1), whereas in pPICZB-BAS1.6:X-33, a novel peak at 27.7 min is detected. **B.** Mass spectra of the extracted spectrum of (i) the novel peak (2) from pPICZB-BAS1.6:X-33 and (ii) a β -amyrin standard.

expression trials, the levels of β -amyrin production were compared between all pPICZB-BAS1:X-33 transformants. Although this method would not give quantitative data on the amounts of AsbAS1 protein present, measuring the amounts of β -amyrin synthesised in each of the transformants would allow indirect quantification of protein levels on the assumption that the transformant with the highest product levels produced more functional AsbAS1 protein.

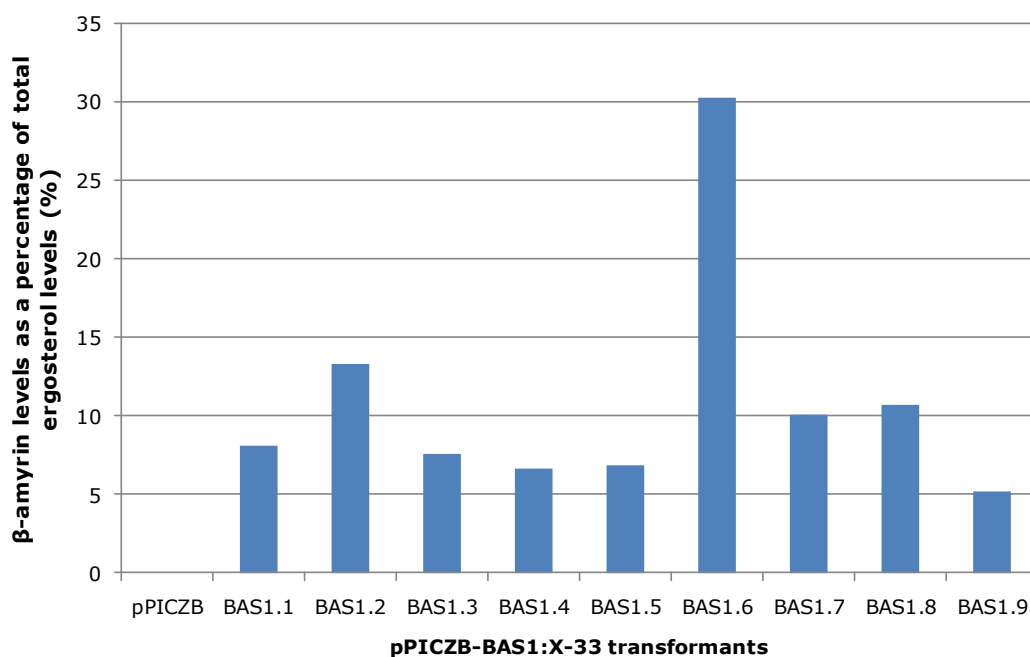


Figure 5.7: Comparison of β -amyryn levels in pPICZB-BAS1:X-33 transformants

Peak area of β -amyryn produced by each transformant is expressed as a percentage of total peak area of ergosterol. The vector only control pPICZB:X-33 produces no β -amyryn and all transformants show some degree of AsbAS1 activity.

Comparisons were made between the nine pPICZB-BAS1:X-33 transformants by expressing β -amyryn levels as a percentage of the total ergosterol level. Ergosterol levels remained at a similar level to the pPICZB:X-33 vector only control in all transformants so did not appear to be affected by pPICZB-BAS1 expression. The peak area of the β -amyryn peak was expressed as a percentage of the peak area of the ergosterol peak for each of the transformants and the data was displayed on a graph (Figure 5.7). The highest levels of β -amyryn production was observed in pPICZB-BAS1.6 which had 30% peak area compared to ergosterol levels, and twice the ratio of β -amyryn:ergosterol seen in any other transformant. This high ratio indicated that pPICZB-BAS1.6:X-33 had high levels of functional AsbAS1 protein, possibly as a result of multiple gene insertion events at the *AOX1* locus during transformation of the pPICZB-BAS1 construct into *Pichia pastoris* X-33. This data indicated that pPICZB-BAS1.6:X-33 is likely to be a highly expressing transformant.

5.3.5.2 pPICZB-BAS2.1:X-33 and pPICZB-BAS4.4:X-33

The two transformants selected from the small scale expression trials as producing the highest AsbAS1 protein levels, pPICZB-BAS2.1:X-33 and pPICZB-BAS4.4:X-33, both produced a novel peak which had a fragmentation pattern

identical to β -amyrin. Figure 5.8 shows the AMDIS analysis of both transformants compared to the pPICZB:X-33 vector only control. All TICs had a peak at 26 minutes which corresponded to the primary membrane sterol ergosterol. The TICs of the transformants also had a novel peak at 27.7 minutes that is not observed in the vector only control with a fragmentation pattern identical to that of β -amyrin. The extracted ion chromatogram at m/z 218, a characteristic fragmentation size of β -amyrin, showed a single peak thereby further confirming the identity of the novel peak to be β -amyrin.

Estimates for β -amyrin production were calculated in the same way as the pPICZB-BAS1:X-33 transformants by expressing the peak area of the β -amyrin peak as a percentage of ergosterol peak area. β -amyrin production was 8% and 2% of the total ergosterol levels for pPICZB-BAS2.1:X-33 and pPICZB-BAS4.4:X-33 respectively. If it is assumed that β -amyrin levels are a direct reflection of AsbAS1 protein levels then pPICZB-BAS1.6:X-33, which had 30% of total ergosterol levels, contains 4-fold and 15-fold more functional AsbAS1 protein than pPICZB-BAS2.1:X-33 and pPICZB-BAS4.4:X-33 respectively. This implied that the untagged *Sad1* construct had much higher functional protein levels than both the C- and N-terminally 6-histidine tagged constructs. This could be a result of higher expression levels for the untagged construct, or that the untagged construct is more active than both of the tagged constructs which implies that the

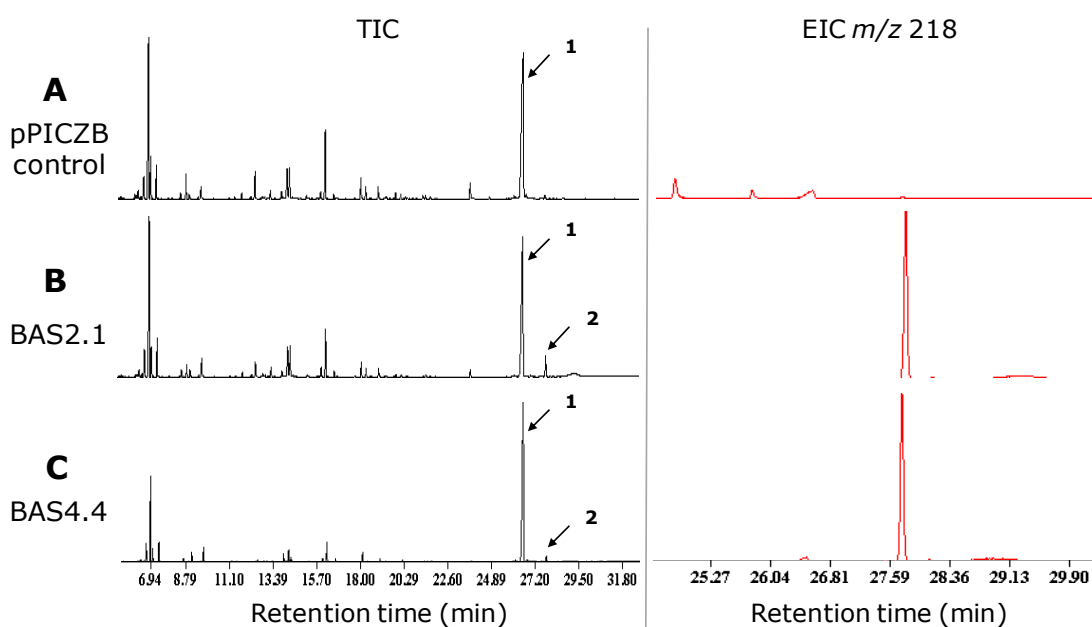


Figure 5.8: GC-MS analysis of pPICZB-BAS2.1:X-33 and pPICZB-BAS4.4:X-33

TIC and EIC (m/z 218) chromatograms of hexane extracts from *Pichia* transformants with **A.** pPICZB:X-33 as an empty vector control, **B.** pPICZB-BAS2.1 and **C.** pPICZB-BAS4.4. pPICZB-X-33 control shows only presence of the primary membrane sterol, ergosterol (1), whereas in pPICZB2.1 and pPICZB4.4 a novel peak (2) at 27.7 min is detected. EIC at m/z 218 shows both of these transformants have the characteristic fragmentation ion of β -amyrin.

6-histidine tag may interfere with enzyme function. However, the presence of a tag on both of these constructs enables the protein to be purified on a large scale by immobilised metal affinity chromatography (IMAC), so despite their reduced expression levels compared to the untagged transformant, pPICZB-BAS1.6:X-33, both tagged constructs would be much easier to purify. Therefore both tagged transformants, pPICZB-BAS2.1:X-33 and pPICZB-BAS4.4:X-33 were carried forward for large scale expression trials.

5.3.6 Large scale expression of pPICZB-BAS2.1:X-33

Small scale expression trials identified a single highly expressing transformant for each of the three different constructs designed for AsbAS1 expression in *Pichia pastoris* X-33. pPICZB-BAS1.6:X-33 was selected as the transformant for the untagged construct, pPICZB-BAS2.1:X-33 for the C-terminally 6-histidine tagged construct and pPICZB-BAS4.4:X-33 for the N-terminally 6-histidine tagged construct. All transformants produced functional protein and GC-MS analysis allowed indirect quantification of protein levels through comparison of β -amyrin product levels produced by each of the transformants (on the assumption that the transformant with the highest product levels produced more functional AsbAS1 protein). pPICZB-BAS1.6:X-33 was shown to produce the most functional protein followed by pPICZB-BAS2.1:X-33 (4-fold less) and finally pPICZB-BAS4.4:X-33 (15-fold less). pPICZB-BAS1.6:X-33 was therefore the first choice for large scale expression as it produced the most functional protein. However, as this construct lacked a tag and at the time, no antibody against the AsbAS1 protein was available, purification of this construct would have been very difficult. The other two transformants contained constructs with 6-histidine tags, so would allow for purification of the protein via IMAC followed by gel filtration, therefore the highest expressing of these two transformants, pPICZB-BAS2.1:X-33 was the first to be chosen for large scale expression of AsbAS1.

For large scale expression of pPICZB-BAS2.1:X-33, the culture volume was scaled up to five litres. Protein expression was induced by resuspension in five litres of methanol growth media (BMMY) and the cells were harvested by centrifugation after three days of protein induction and stored at -80°C.

5.3.7 Immobilised metal affinity chromatography of pPICZB-BAS2.1:X-33

Purification of *Homo sapiens* lanosterol synthase (hOSC) was carried out using IMAC, which utilised the 6-histidine tag on the expressed protein followed

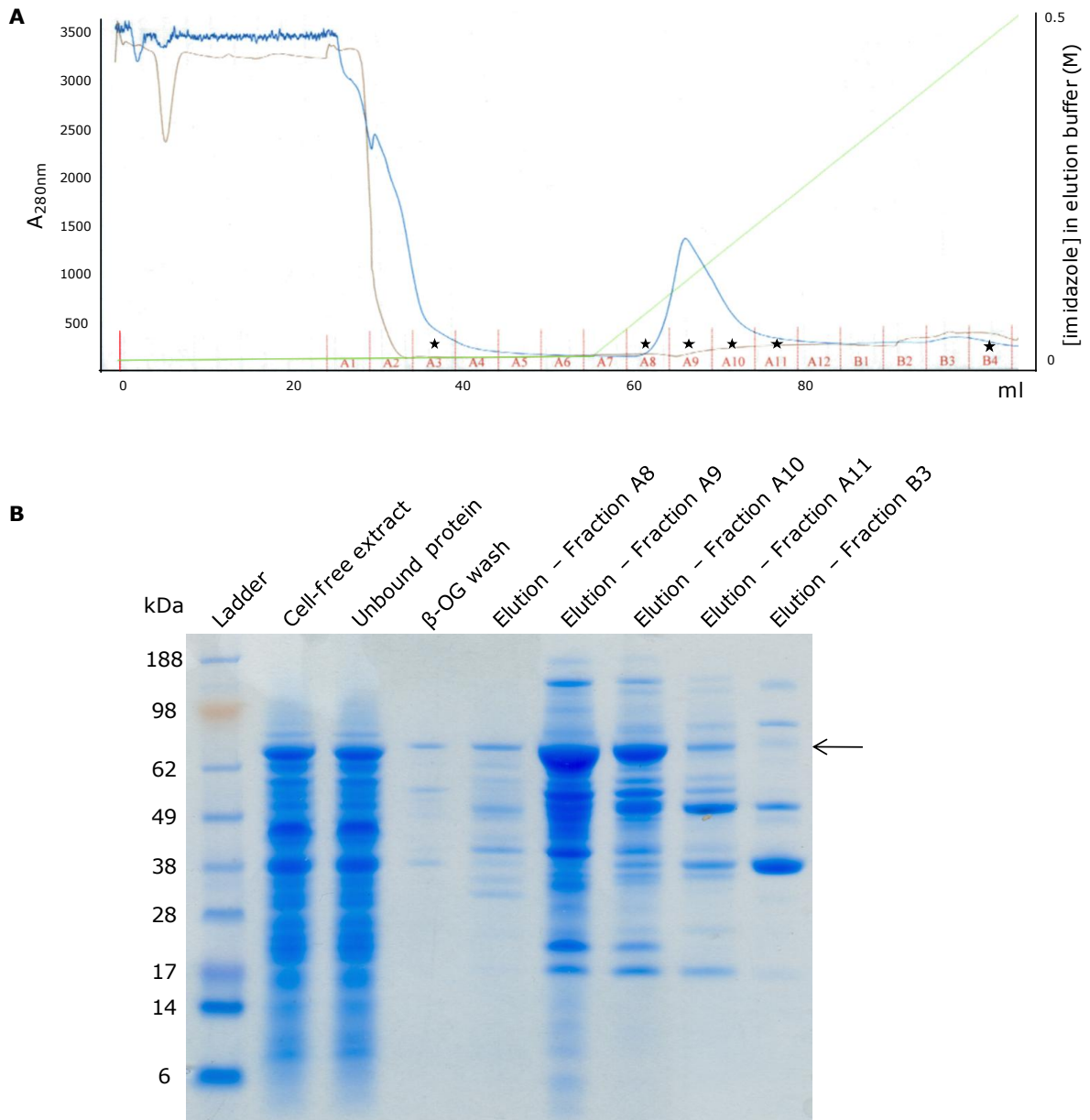


Figure 5.9: IMAC purification of pPICZB-BAS2.1:X-33 and SDS-PAGE analysis of eluted fractions

A. Linear elution profile of AsbAS1 from a nickel charged metal affinity column using a 0.5M imidazole gradient. Absorbance at 280nm of eluate from the column is represented by the blue line (left axis), conductivity is shown by the brown line and concentration of imidazole in the elution buffer is shown by the green line (right axis). Sample injection at the start of the purification is shown by the red line and fraction numbers are also shown in red. Fraction A1 was collected during the 1CV Triton X-100 wash, Fractions A2-A6 were collected during the 5CV β -OG wash and the remaining fractions A7-B4 were collected during elution with an imidazole gradient. The fractions labelled with a star denote those chosen for SDS-PAGE and Western blot analysis. **B.** SDS-PAGE gel of fractions eluted from the column. Lane 1, SeeBlue® Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), myoglobin red (17kDa), lysozyme (14kDa), aprotinin (6kDa); Lane 2, Cell free extract from Day3 pPICZB-BAS2.1:X33 culture; Lane 3, Unbound protein; Lane 4, Fraction A3 – β -OG detergent exchange wash; Lane 5-9, Fractions A8-A11 - 1st elution peak; Lane 10, Fraction B3 – 2nd elution peak. Expected molecular weight of AsbAS1 (shown by the arrow) is predicted to be 86.8kDa, which falls between the 62 and 98kDa markers.

by gel filtration (151). Homology modelling predicted hOSC and AsbAS1 to have highly similar structures (see Chapter 2) and both proteins have an almost identical molecular weight, therefore the conditions used for hOSC purification could be applied to AsbAS1 purification.

The cells harvested from protein induction were lysed and incubated with Triton X-100 to solubilise the protein from the membrane. After centrifugation the cell-free extract was loaded onto a nickel-charged metal affinity column and washed to remove unbound protein. Detergent exchange was carried out by first washing the column with Wash buffer containing Triton X-100 (0.2% v/v) followed by Wash buffer containing β -OG (0.8% w/v). Detergent exchange was performed to prepare the protein for further crystallisation trials, as the optimal conditions for hOSC crystallisation were achieved using β -OG (144, 151). Bound protein was eluted with a linear gradient of 0.5 M imidazole and as all fractions were eluted, their absorbance was measured at 280 nm and the data displayed in graphical form. The elution profile showed two peaks, one large peak at the start of the imidazole gradient and a much smaller peak towards the end (Figure 5.9A). The fractions containing these eluted peaks together with fractions collected from each of the wash steps were analysed by SDS-PAGE and Western blotting (Figure 5.9B and 5.10).

The InstantBlue stained SDS-PAGE gel showed presence of a large band which corresponded to the expected molecular weight of AsbAS1 in fractions containing the cell-free extract, unbound protein and elution fractions A8-A11 (Figure 5.9B). Upon Western blotting using an antibody against the 6-histidine tag, the band observed in fractions A8-A11 gave a strong signal, but no signal was observed for the band seen in the cell-free extract and unbound protein fractions (Figure 5.10). This indicated that the band observed in these fractions is likely to be *Pichia pastoris* AOX1 as it is co-expressed with AsbAS1 and has an almost identical molecular weight. As PpAOX1 does not have a 6-histidine tag, it would not bind to the column so would be present in the unbound fraction as is observed on the SDS-PAGE gel. The intensity of the signal seen on the Western blot corresponds with the size of the band on the SDS-PAGE gel which implied that AsbAS1 was a major constituent of this band.

To confirm the identity of this band, it was excised from the gel, from fractions A8 and A11 and sent for MALDI-TOF-MS fingerprinting. Mascot analysis of the bands from fraction A8 and A11 confirmed the identity of protein to be *Avena strigosa* β -amyrin synthase, with mascot scores of 137 and 250 respectively. This provided convincing evidence that the large band seen in the

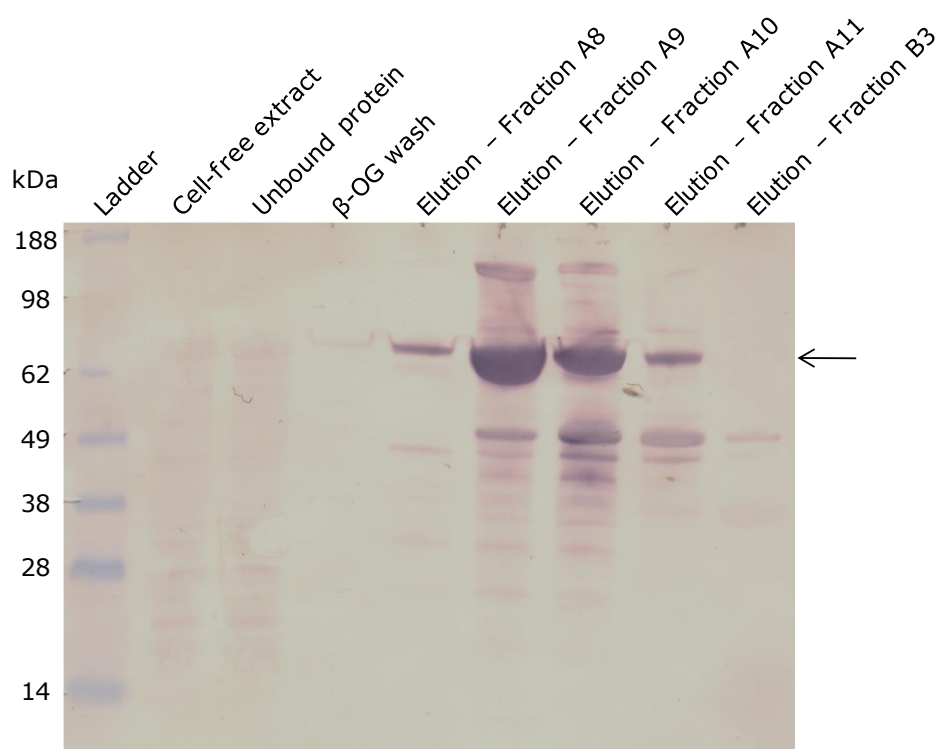


Figure 5.10: Western blot of fractions from IMAC purification of pPICZB-BAS2.1:X-33.

Colorimetric detection of fractions eluted from the metal affinity column with an antibody directed against the 6-histidine tag. Lane 1, SeeBlue® Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), lysozyme (14kDa); Lane 2, Cell free extract from Day3 pPICZB-BAS2.1:X33 culture; Lane 3, Unbound protein; Lane 4, Fraction A3 – β -OG detergent exchange wash; Lane 5-9, Fractions A8-A11 - 1st elution peak; Lane 10, Fraction B3 – 2nd elution peak. Expected molecular weight of AsbAS1 (shown by the arrow) is predicted to be 86.8kDa, which falls between the 62 and 98kDa markers.

fractions eluted from the IMAC column was AsbAS1. Therefore fractions A8-A11 were pooled and concentrated from 20 ml to 3 ml and stored in 1 ml aliquots at -80°C.

Comparisons between the crude extract and the fractions eluted from the column demonstrated that IMAC purification had been successful in partially purifying AsbAS1. Western blotting and MALDI-TOF-MS analysis showed that AsbAS1 was the major protein observed in the eluted fractions A8-A11, however there was still a large number of contaminating proteins present in these fractions. It is unlikely that all of the contaminating proteins were histidine-rich, so there must be another explanation for the presence of impurities. One cause could be protein aggregation as a result of detergent addition to the cell lysate. Although necessary for AsbAS1 solubilisation from the membrane, Triton X-100 may have bound to other proteins thereby causing aggregation with AsbAS1 and resulting in aggregated protein binding to the IMAC column via the AsbAS1 6-histidine tag. Another reason for the eluted impurities could be that the

detergent exchange wash step was too short, so weakly bound protein may have remained on the column and been eluted in the final elution step and contaminating the fractions. In an attempt to solve this problem, the IMAC purification was repeated using a fresh pPICZB-BAS2.1:X-33 culture and the purification protocol was modified to include longer wash steps with 5CV and 10CV of Triton X-100 wash buffer and β -OG wash buffer respectively. However, these elongated steps resulted in AsbAS1 being washed off the column indicating that it is only weakly bound. Weak binding could be as a result of protein aggregation or due to the 6-histidine tag not being fully exposed.

5.3.8 Secondary purification of pPICZB-BAS2.1:X-33

Despite the large number of impurities present in the pooled fractions from IMAC purified AsbAS1, two different secondary purification steps, gel filtration and hydrophobic interaction chromatography, were carried out. The aim was to remove some of the major impurities and to determine which method would be most successful for subsequent purifications.

Gel filtration separates proteins based on the relative size of the protein molecules. The gel filtration matrix contains porous beads which allow small molecules to diffuse into them, but large molecules cannot enter as they are too large. The column is pre-equilibrated with a buffer which fills both the area inside and outside the beads. Once the protein sample is applied to the column, it moves through the bead matrix. Large molecules cannot enter the beads, so pass through the column and are eluted first whereas small molecules are able to migrate into the matrix so spend longer in the column and are eluted later (291).

For secondary purification of AsbAS1, 1 ml of the concentrated pooled fractions from IMAC was loaded onto the gel filtration column. The column was washed with 1 CV of gel filtration buffer and 1 ml fractions were collected which were monitored at an absorbance of 280 nm as they were eluted from the column. AsbAS1 has a high molecular weight and was one of the largest proteins present in the IMAC purified sample, so was expected to be eluted in one of the early peaks. The elution profile showed five peaks, none of which were clearly defined indicating that complete protein separation was not achieved. SDS-PAGE analysis of fractions eluted at the maxima of each peak showed all fractions contained multiple proteins but there was some evidence of size separation as fractions eluted earlier contained the largest proteins and fractions eluted later contained smaller proteins (Figure 5.11). A band corresponding to the expected molecular weight of AsbAS1 was observed most strongly in the second eluted

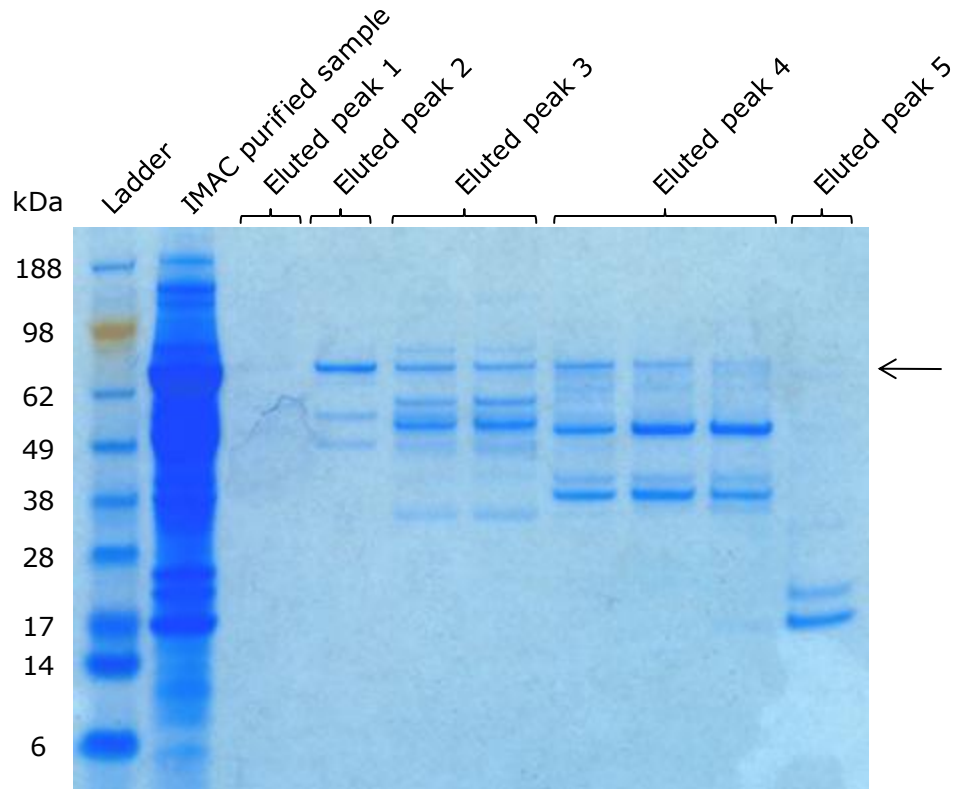


Figure 5.11: SDS-PAGE analysis of fractions eluted from gel filtration of IMAC purified AsbAS1

SDS-PAGE gel of fractions eluted from the column. Lane 1, SeeBlue® Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), myoglobin red (17kDa), lysozyme (14kDa), aprotinin (6kDa); Lane 2, Concentrated protein from pooled IMAC fractions A8-A11; Lane 3, Fraction 51 – 1st elution peak; Lane 4, Fraction 64 – 2nd elution peak; Lane 5, Fraction 68 – 3rd elution peak; Lane 6, Fraction 69 – 3rd elution peak; Lane 7, Fraction 74 – 4th elution peak; Lane 8, Fraction 76 – 4th elution peak; Lane 9, Fraction 78 – 4th elution peak; Lane 10, Fraction 89 – 5th elution peak. Expected molecular weight of AsbAS1 (shown by the arrow) is predicted to be 86.8kDa, which falls between the 62 and 98kDa markers.

peak and contained three lower molecular weight contaminating proteins. Evidence of the band was also seen at lower levels in peaks three and four however these fractions contained much higher levels of lower molecular weight contaminating proteins.

This result indicates that gel filtration would be a suitable method for secondary purification of AsbAS1 as clear size separation was seen for the IMAC purified sample which removed a number of the lower molecular weight contaminants. Had the purity of the IMAC purified sample been higher, gel filtration would have been much more successful, therefore once the IMAC purification step has been optimised, gel filtration can be utilised as the final polishing step to remove minor impurities.

Hydrophobic interaction chromatography (HIC) separates proteins based on the strength of hydrophobic interactions to hydrophobic ligands immobilised

on an uncharged matrix. The sample is loaded onto the column in high salt conditions as hydrophobic interactions are enhanced in these conditions enabling hydrophobic regions of the protein to bind to the column. Proteins are then eluted by a linear decrease in salt concentration, allowing weakly hydrophobic proteins to elute first followed by strongly hydrophobic proteins (292-294). AsbAS1 has hydrophobic helix on the protein surface that facilitates membrane attachment. It was thought that this slight increase in hydrophobicity might enable it to bind more strongly to the hydrophobic column than the soluble proteins present in the IMAC purified sample.

For secondary purification of AsbAS1, 1 ml of the concentrated pooled fractions from IMAC was loaded onto the HIC column. Non-bound proteins were eluted with a high salt wash and bound proteins were eluted with a linear gradient of low salt buffer and 0.25 ml fractions were collected which were monitored at an absorbance of 280 nm. The elution profile showed one major peak which eluted at the end of the low salt gradient. SDS-PAGE analysis of alternate fractions from the peak showed that sample purity had not been improved, as the majority of the major impurities were still present in the eluted fractions (Figure 5.12). This indicated that protein aggregation had occurred which may have been caused by high salt concentrations. However despite the possibility of protein aggregation, there was no evidence that the purity of the

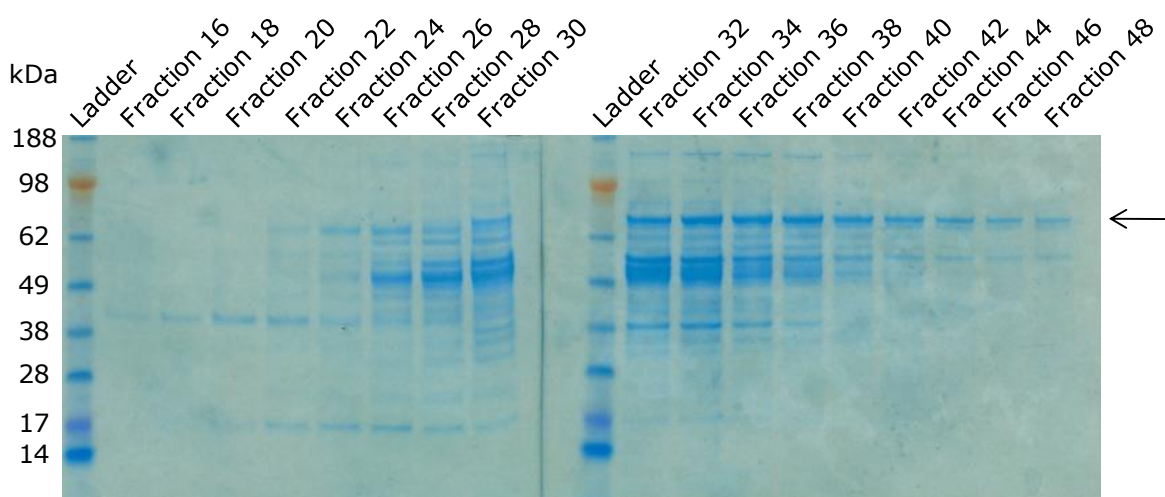


Figure 5.12: SDS-PAGE analysis of fractions eluted from hydrophobic interaction chromatography of IMAC purified AsbAS1

SDS-PAGE gel of fractions eluted from the column. Lane 1, SeeBlue® Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), myoglobin red (17kDa), lysozyme (14kDa), aprotinin (6kDa); Lanes 2-9, Even fractions 16-39; Lane 10, SeeBlue® Plus2 Pre-Stained Standard; Lanes 11-19, Even fractions 31-48. Expected molecular weight of AsbAS1 (shown by the arrow) is predicted to be 86.8kDa, which falls between the 62 and 98kDa markers.

sample had been improved, therefore HIC is unlikely to be a beneficial step should the IMAC purification step be optimised.

5.3.9 Repeat purification problems of pPICZB-BAS2.1:X-33

Subsequent attempts to purify AsbAS1 using IMAC were significantly less successful than the first attempt (see 4.3.7). After extending the wash steps to ensure all unbound protein was eluted proved to be unsuccessful as AsbAS1 was washed off the column, IMAC was repeated several times using the original protocol which had been so successful for hOSC. However the peak seen at the start of the imidazole elution gradient never appeared at the same magnitude and although SDS-PAGE analysis revealed a band was seen at the expected molecular weight of AsbAS1, MALDI analysis repeatedly found this to be PpAOX1. This implied that the yeast transformant was no longer making AsbAS1. However *in vivo* functional analysis carried out on each of the cultures used for large scale expression showed that all the cultures produced β -amyrin at similar levels to those seen in the original functional assay of pPICZB-BAS2.1:X-33 activity. This indicated that AsbAS1 was being produced at similar levels.

The subcellular localisation of OSCs in mammalian and yeast cells has been shown to be both in microsomes and in lipid particles. Partial purification of native OSCs in mammals and yeast has been achieved by detergent solubilisation of the microsomal fraction (258-260). Therefore microsomes were isolated from induced cultures of pPICZB-BAS2.1:X-33 in order to increase the purity of the crude extract before IMAC. The microsome fraction was obtained and a number of different detergents were used to solubilise pPICZB-BAS2.1:X-33 from the microsomes. SDS-PAGE analysis showed a single band at the expected molecular weight of AsbAS1 but this band was also found to be PpAOX1 by MALDI analysis (data not shown).

The problems encountered with AsbAS1 purification seemed to be as a result of PpAOX1 contamination which produced ambiguous results. As the AOX1 promoter is used to drive heterologous protein expression of AsbAS1 in *Pichia pastoris* and pPICZB-BAS2.1:X-33 has the Mut⁺ phenotype, PpAOX1 is co-expressed with AsbAS1. Unfortunately as PpAOX1 has a similar molecular weight to AsbAS1, they both migrate to the same point on an SDS-PAGE gel and cannot be separated by gel filtration. Presence of a histidine tag on the expressed AsbAS1 should have facilitated separation by IMAC as AsbAS1 is able to bind to the column whereas PpAOX1 is not. However, although this was seen on the first purification attempt, subsequent attempts indicated that PpAOX1 also bound to the column as it was seen in the eluted fractions.

If PpAOX1 was not expressed, this would solve the problem of PpAOX1 contamination leading to ambiguous results. Therefore the three pPICZB constructs, pPICZB-BAS1, pPICZB-BAS2 and pPICZB-BAS4, were transformed into the *Pichia pastoris* KM71H strain. This strain contains a non-functional AOX1 gene and relies on the secondary AOX gene in *Pichia pastoris*, AOX2, for production of the alcohol oxidase enzyme. AOX2 has the same activity as AOX1 but due to a weaker promoter has a much lower expression level (277, 279). pPICZB constructs expressed in KM71H will not be co-expressed with AOX1 so this should prevent the ambiguous results seen with the Mut⁺ transformants. Although AOX2 also has the same molecular weight as AOX1, due to its much lower expression levels, it should not interfere with AsbAS1 detection.

5.3.10 Expression of pPICZB-BAS1, BAS2 and BAS4 in *Pichia pastoris* KM71H

Linearised pPICZB-BAS1, pPICZB-BAS2 and pPICZB-BAS4 plasmid DNA were transformed into competent *P. pastoris* KM71H cells via electroporation. Colonies were purified on fresh plates for each construct to select for single transformants. As all *P. pastoris* KM71H transformants are Mut^S, there is no need for selection on MM and MD media. After purifying selection, pPICZB-BAS1:KM71H had seven transformants, pPICZB-BAS2:KM71H had six transformants, and pPICZB-BAS4:KM71H had ten transformants. Previously, *P. pastoris* X-33 transformants were tested for expression of AsbAS1 by SDS-PAGE analysis. However this technique was not suitable for the pPICZB-BAS1 constructs as they did not have a tag and no AsbAS1 antibody was available at the time. Tagged constructs were tested using Western blotting against the His•Tag[®] monoclonal antibody but these proved to be unclear due to the low amounts of AsbAS1 protein loaded onto the SDS-PAGE gel. Testing the transformants for functional expression proved a much more conclusive technique as product levels could be detected even if present at very low levels. If it was assumed that product levels mirrored protein levels, transformants which expressed the highest levels of functional protein could be easily identified and carried forward to large scale purification.

For *in-vivo* functional analysis of all *P. pastoris* KM71H transformants, each transformant was grown and analysed using the same protocol as for pPICZB:X-33 transformants (see 4.3.4 and 4.3.5).

5.3.10.1 pPICZB-BAS1:KM71H

Of the seven pPICZB-BAS1:KM71H Mut^S transformants analysed, all produced a novel peak which has a fragmentation pattern identical to that of β -amyirin which confirmed that all transformants produced active AsbAS1 protein. To identify a candidate for large scale expression, the β -amyirin level in each of the transformants was analysed as a percentage of ergosterol level for an indirect quantification of AsbAS1 protein levels. The peak area of the β -amyirin peak was expressed as a percentage of the peak area of the ergosterol peak for each of the transformants and the data was displayed on a graph (Figure 5.13). The highest level of β -amyirin production was observed in pPICZB-BAS1.4:KM71H which had 5.8% peak area compared to ergosterol levels. However this level of β -amyirin production was significantly lower than that seen for the highest expressing pPICZB-BAS1:KM71H transformant which produced β -amyirin to 30% of ergosterol levels. This low production of β -amyirin may be linked to the slow growth levels of the *Pichia* KM71H strain. The growth of the yeast depends on supply of carbon for use as an energy source, which *Pichia* KM71H can obtain

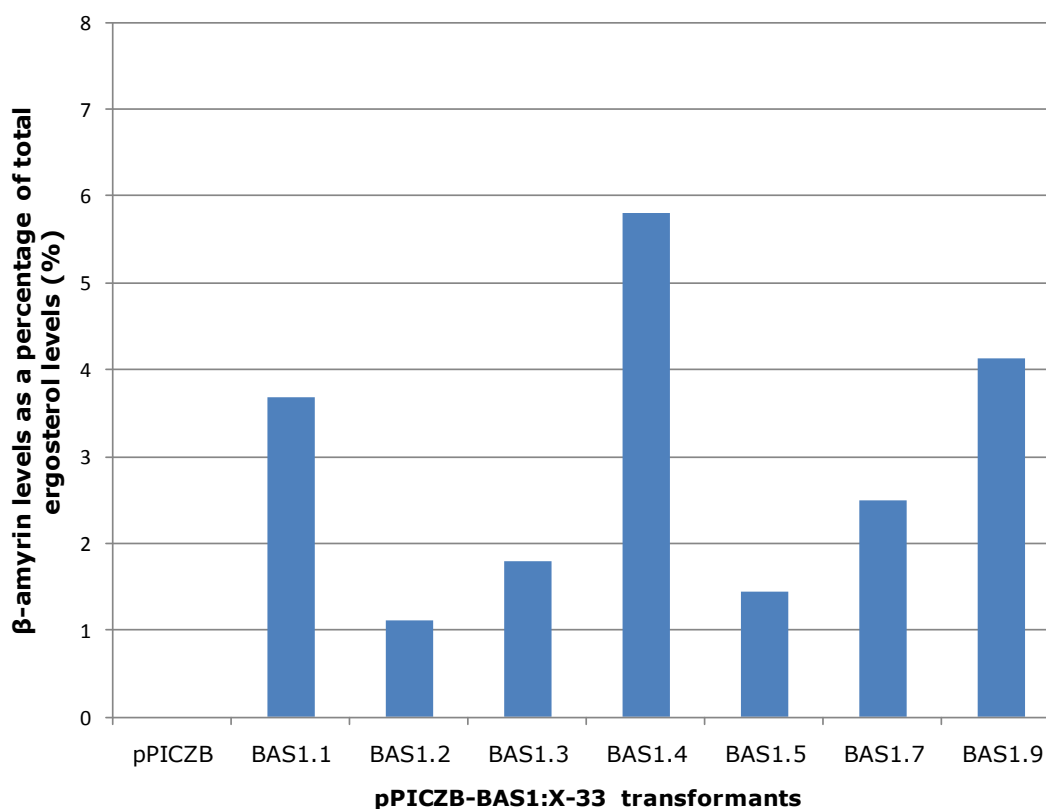


Figure 5.13: Comparison of β -amyirin levels in pPICZB-BAS1:KM71H transformants

Peak area of β -amyirin produced by each transformant is expressed as a percentage of total peak area of ergosterol. The vector only control pPICZB:KM71H produces no β -amyirin and all transformants show some degree of AsbAS1 activity.

from methanol should glycerol not be available. In wild-type *Pichia*, both the AOX1 and AOX2 genes can be used to metabolise methanol but *Pichia* KM71H lacks a functional AOX1 gene so can only use the much lower expressed AOX2 gene for methanol metabolism and as a result grows at a slower rate. Both the KM71H and the X-33 transformants were grown using the same conditions and protein was induced over the same number of days, so the slower growing KM71H would have had a lower final cell density therefore would produce lower amounts of AsbAS1 protein overall. However, as AsbAS1 levels were measured as a percentage of ergosterol levels rather than quantitative amounts, this implies that the KM71H cells were expressing AsbAS1 protein at lower levels than X-33 cells. Therefore, optimisation of the growth and induction conditions is needed to maximise the density of the cell culture and the amount of AsbAS1 protein expressed. This could be done via finding the optimum concentration of methanol required for induction, increasing length of protein induction, using a fermentation system and increasing shaking speed to maximise aeration of the culture. *Pichia* transformation could also be repeated to try to generate a high copy number transformant with multiple integration events.

Chapter 6 - Conclusion

This work describes the detailed characterisation of β -amyrin synthase from the diploid oat *Avena strigosa* (AsbAS1). AsbAS1 catalyses the first committed step in biosynthesis of avenacin A-1, an antimicrobial triterpene glycoside localised in oat roots that confers broad spectrum disease resistance.

Chapter 2 discusses the characterisation of AsbAS1 on a structural level to explore the mechanism of enzyme action and identify residues and regions of the protein that are likely to be catalytically important. In the absence of a crystal structure for the enzyme, a homology model was generated using the human lanosterol synthase crystal structure as a template. Validation of the AsbAS1 model showed that the indicators calculated from the model were indeed similar to those calculated from known protein structures of a similar size – a necessary condition for a reliable model. The AsbAS1 model, together with information gathered from a comprehensive sequence alignment and phylogenetic tree of all cloned plant OSCs, was used to explore AsbAS1 mechanism and the mechanistic differences between the sterol and triterpene OSC classes.

During the first stage of the OSC mechanism, when the substrate is prefolded in the active site, clear differences between sterol and triterpene OSCs, and monocot and dicot triterpene OSCs are observed. Triterpene OSCs fold the B-ring into the chair conformation, whereas sterol OSCs form the energetically unfavourable boat conformation. This is mediated by Tyr98 (hOSC numbering) which forces the C₁₀ carbon of 2,3-oxidosqualene below the molecular plane thereby enforcing the boat confirmation (144). Triterpene OSCs, have a smaller residue at the equivalent position thereby removing the steric pressure on C₁₀ and allowing the chair conformation to form. However, this was found to be restricted to dicot triterpene OSCs only. Monocot triterpene OSCs have retained a bulky aromatic residue at the equivalent Tyr98 position, so to allow the favourable chair conformation to be formed, have reversed the residues that position the tyrosine. This elevates the position of the tyrosine with respect to the C₁₀ of the substrate and no longer causes steric hindrance. Phylogenetic evidence indicates that monocot and dicot triterpene OSCs have evolved separately, with monocot triterpene OSCs being more closely related to sterol OSCs. Therefore it seems that monocot triterpene OSCs have retained many of the features of sterol

OSCs, but these have been modified to allow the enzyme to function as a triterpene OSC.

After the substrate has been positioned in the active site, very few differences were found between sterol and triterpene OSCs during cyclisation of the first three rings. A number of amino acid residues were compared which had previously been identified using site-directed mutagenesis or inhibitor studies. All amino acid residues involved in reaction initiation and stabilisation of the first three rings were highly conserved in AsbAS1 and hOSC as well as in all other plant OSCs indicating that despite the different evolutionary paths of monocot and dicot OSCs, they have maintained a very similar catalytic mechanism. The major difference between sterol and triterpene OSC mechanism occurs at cyclisation of the D-ring. It is here that the two mechanisms diverge, with D-ring formation signalling the end of sterol OSC cyclisation, whereas triterpene OSC cyclisation continues. In sterol OSCs, D-ring stabilisation is coordinated by His232 and Phe696, which do not provide enough negative potential to stabilise the positive charge, so deprotonation takes place and a tetracyclic product is formed (144). However in AsbAS1 and other triterpene OSCs, His232 is replaced by an aromatic residue (Phe or Tyr) therefore the positive charge is coordinated by two aromatic residues which provide sufficient negative potential to stabilise the positive charge and allow the formation of the 5th E-ring. This single residue change has been shown to differentiate between tetracyclic and pentacyclic products and the multiple sequence alignment clearly shows all sterol OSCs contain a His232 residue whereas all pentacyclic triterpene OSCs have an aromatic residue (Phe or Tyr) at the equivalent position (176).

E-ring stabilisation occurs only in pentacyclic triterpene OSCs and in AsbAS1 stabilisation appears to be mediated by Trp257 and Phe259 which allow the E-ring to be expanded to a 6-membered carbon ring. Both of these residues are conserved in one class of pentacyclic triterpene OSC, β -amyrin synthases, but not in lupeol synthases where the tryptophan is replaced by a leucine residue. This reduction in stabilising negative potential results in lupeol synthases being unable to expand the E-ring resulting in a product, lupeol, with a 5-membered E-ring. Deprotonation at the end of the cyclisation cascade still remains unclear as due to the non-polar nature of the active site cavity, there are currently no candidate basic residues in the vicinity of the product cation to facilitate deprotonation.

The predictions from the AsbAS1 model and the multiple sequence alignment indicate that a very small subset of active site residues are responsible for defining product identity for plant OSCs. Just a single residue change can

completely alter the product produced by the enzyme and presumably over time, it is these single base changes that have facilitated the diversification of the pentacyclic triterpene OSCs. It is also important to note that the two AsbAS1 residues Trp257 and Phe 259, appear only to be a requirement for defining β -amyrin/lupeol product identity and pentacyclic products in triterpene OSCs that produce single products (i.e. solely β -amyrin or lupeol). There are multifunctional triterpene OSCs from *Pisum sativum* (54), *Arabidopsis thaliana* (52, 132, 139), *Lotus japonicus* (134) and *Rhizospora stylosa* (56) that synthesise more than one product, often including both β -amyrin and lupeol, that contain different combinations of these apparent defining residues. It therefore seems that the active site environment as well as the conserved active site residues are also important for dictating product identity which adds another level of complexity to the OSC mechanism.

Chapters 3 and 4 discuss the characterisation of 17 AsbAS1 (*sad1*) mutants at the gene, transcript and protein level. The 17 *sad1* mutants were sequenced and found to contain single point mutations resulting in three types of predicted mutations: premature termination of translation, splicing errors and amino acid substitution. Two groups of mutants (#A1, #B1 and #610; and #384 and #1023) were found to have identical point mutations. Diversity array analysis (DART) showed that the majority of the *sad1* mutants were independent. Mutants #A1 and #B1 were genetically the most similar, but there was no conclusive proof that they were identical, so were treated as independent mutants in the analysis. Further DART analysis using more replicates for each mutant will be needed to confirm whether these mutants are truly independent.

RT-PCR analysis of the *sad1* mutants at the transcript level revealed that all of the predicted premature termination of translation mutants had severely reduced transcript levels compared to the wild type S75. Nonsense mediated mRNA decay is thought to be the cause for transcript reduction and occurs post-transcriptionally as a result of mRNA instability (236). All of the splicing error mutants had transcript present and further analysis by Northern blotting and RT-PCR showed that each of the splicing mutants had a reduced transcript size. This reduction in size corresponded to a deletion of the exon that each splicing mutation is associated with. Finally, all of the amino acid substitution mutants had full-length transcript present which was expected as nonsynonymous mutations were not anticipated to affect transcript stability.

To analyse the *sad1* mutants at the protein level, a highly specific AsbAS1 polyclonal antibody was generated from AsbAS1 protein purified from inclusion

bodies. Western blotting showed that both the premature termination of translation and splicing error mutants had no AsbAS1 protein present. As the splicing mutants produce shortened *sad1* transcripts, it seems likely that the protein produced is unable to fold or misfolded and is hence targeted for degradation. Three out of the seven amino acid substitution mutants produced full length AsbAS1 protein which indicates that the protein is able to fold but is non-functional. The location of these mutations were visualised on the AsbAS1 homology model to ascertain what effect the mutations might have on protein function. Mutant #358 contained a mutation in one of the cysteine residues that coordinates the catalytic aspartate. The cysteine is essential for activation of the aspartate and therefore reaction initiation is likely to be affected. Mutants #384 and #1023 both contained the same mutation in a serine residue close to the substrate access channel, and Phe725 which is involved in stabilising the D-ring. These mutants could therefore have blocked substrate access or product exit from the active site or be unable to stabilise the D-ring leading to deprotonation and formation of a tetracyclic product. If the latter occurs then truncated tetracyclic products may be present in the root tips of mutants #384 and #1023.

The location of all the amino acid substitution mutants on the AsbAS1 structure gives an insight into the stability of OSC enzymes. As the majority of the AsbAS1 secondary structure is helical, it is assumed that in the original mutagenised oat population there were other *sad1* mutants that contained amino acid substitutions in other areas of the protein. However, if these had no effect on enzyme function then they would not have been selected for in the loss of function screen. This therefore implies that OSC enzymes are very tolerant of mutations in areas of the protein that are not critical for function, and only those mutations that seriously affect enzyme expression or function result in premature termination of the pathway.

Chapter 5 describes the cloning and heterologous expression of AsbAS1 (*Sad1* cDNA) using the *Pichia pastoris* expression system with the aim of developing a purification strategy for AsbAS1. Three constructs were designed for heterologous expression and consisted of the full-length *Sad1* cDNA sequence with no tag (BAS1), C-terminal 6-histidine tag (BAS2) or N-terminal 6-histidine tag (BAS4). All three constructs were transformed into wild-type *Pichia pastoris* X-33 and Western blotting and an *in vitro* functional assay was used to identify the highest expressing functional transformant for each construct.

The construct with a C-terminal tag was used for large scale expression to develop a purification strategy for AsbAS1. The purification protocol was based

on the conditions used for purification of human lanosterol synthase (hOSC) and a combination of IMAC and gel filtration enabled partial purification of AsbAS1. The partially purified sample still contained a number of contaminants and subsequent purification attempts were hindered by the presence of the expression host enzyme PpAOX1, which is co-expressed with AsbAS1 and has a similar molecular weight. Therefore to prevent PpAOX1 contamination, the three AsbAS1 constructs were transformed into a *Pichia* strain lacking the AOX1 gene (KM71H) and a highly expressing functional transformant was identified for the un-tagged protein. In future, highly expressing functional tagged constructs in the KM71H strain can be identified and used to further develop the purification protocol to obtain pure AsbAS1 which can then be used for kinetic and structural studies.

This work has raised interesting questions about the evolution of the plant oxidosqualene cyclase family, in particular, how triterpene OSCs have evolved from ancestral sterol OSCs and how monocot and dicot triterpene OSCs have evolved separately. This study can be extended to look at the evolution of OSCs from their bacterial ancestors across all kingdoms of life. Evolutionary analysis can be combined with product profiles and known conserved residues of OSCs to generate a catalytic landscape of the OSC family. This landscape together with the continuing discovery of novel OSCs and structural information from homology modelling or crystal structures is expected to give clues as to the direction of OSC evolution. This will facilitate the design of novel OSCs which synthesise new secondary metabolites not previously seen in nature that may have applications in the medical or commercial industries.

Chapter 7 - Future work

The characterisation of β -amyrin synthase from *Avena strigosa* has provided an insight into the structure and function of the enzyme, but has also raised a number of questions and hypotheses that can be explored through further studies.

Chapter 2 described the structural characterisation of AsBAS1 via homology modelling. Docking of some of the reaction intermediates and the β -amyrin product into the active site identified residues and regions of the enzyme that were likely to be catalytically important. Homology modelling allows hypotheses to be made regarding enzyme structure and mechanism, but alone it cannot provide experimental evidence for the involvement of specific residues or regions of the enzyme in catalysis. Also, as the enzyme is modelled on a template structure there is a possibility that errors may be introduced into the homology model. Therefore, the involvement of these residues in catalysis and ring stabilisation should be examined experimentally, and this can be done using site directed mutagenesis. The effect of the mutations would be ascertained by heterologous expression in yeast, followed by GC-MS of metabolite extracts to look for novel cyclisation products. Experiments could be conducted in *Pichia pastoris* or *Saccharomyces cerevisiae* as neither produce secondary metabolites so accumulation of a novel product can be detected.

Residues that are targets for mutagenesis include Tyr119 which is implicated in prefolding of the B-ring of the substrate prior to catalysis. Tyr119 could be mutagenised to asparagine or isoleucine which are seen in the dicot β -amyrin synthases to see if β -amyrin is still produced. Alternatively the two residues (Phe731 and Leu473) that position Tyr119 could be reversed to determine whether a chair-boat-chair conformation of the A-B-C rings can be achieved. This residue reversal mirrors what is seen in sterol OSCs which prefold the B-ring into the boat conformation. A number of aromatic residues have been implicated in ring stabilisation (letters in brackets denote which ring(s) each residue is implicated in stabilising); Trp611 (A, B), Phe731 (A, B, C), Phe725 (C, D, E), Phe259 (C, D, E). Each of these candidate residues could be mutagenised to non-aromatic residues and GC-MS could be used to detect novel cyclisation products. The ring structure of any truncated cyclisation products will provide valuable information as to which stage of the cyclisation each residue is involved

in. Product specificity of the enzyme could also be investigated using site directed mutagenesis. Mutagenesis of Phe259, which is implicated in defining pentacyclic products in dicot triterpene OSCs, to a histidine as is seen sterol OSCs will determine whether Phe259 has a similar role in monocot triterpene OSCs. If this is the case, it would be expected that tetracyclic dammarenyl products would be produced. The role of Trp257 in defining β -amyrin or lupeol product identity could be investigated by mutagenising this residue to a leucine as is seen in lupeol synthases and see if lupenyl based products are produced.

In addition to targeting individual residues, mutagenesis could also be used to investigate the role of specific regions of the enzyme. Sequence analysis shows that all plant OSCs have an insertion at the N-terminus of approximately 30 amino acids when compared to animal and fungal OSCs. Homology modelling of AsbAS1 showed that this inserted region was modelled as an unordered loop that extended away from the protein surface. The role of this loop region in plants is unclear as its secondary structure cannot be assigned, but it may be involved in protein-protein interactions or (add other reason here). To ascertain whether the loop is essential for protein function, it could be removed by mutagenesis and the truncated protein expressed in the *Pichia pastoris* expression system. AsbAS1 protein could then be detected using the AsbAS1 antibody and protein function could be assayed by GC-MS.

Mutagenesis could also be used to make point mutations in the amphipathic membrane-binding helix in order to solubilise the protein. Due to the membrane bound nature of the enzyme, purification currently requires the application of detergents which hinder the subsequent column purification steps. If the protein was able to be solubilised, purification may be more successful which would enable crystallisation trials to be done.

Chapter 2 also discussed the phylogenetic analysis of all cloned plant OSCs. This study could be extended to include animal and fungal OSCs as well as bacterial squalene hopene cyclases (SHCs). This will provide insights into how each of the kingdoms evolved, whether each evolved separately from a common ancient ancestor or whether the eukaryotic cyclases evolved from bacterial SHCs. The current plant OSC multiple sequence alignment could also be examined in more detail to study the sequence homology of inserted N-terminal loop region that is seen in plant OSCs.

Chapters 3 and 4 described the characterisation of the 17 *sad1* mutants at the gene, transcript and protein level. Protein analysis using the AsbAS1

antibody showed that three of the amino acid substitution mutants, #384, #532 and #1023, had full length AsbAS1 protein present. Mapping the point mutations to the AsbAS1 homology model indicated that #532 was likely to be affected in reaction initiation, and #384 and #1023 (which have identical mutations) could be affected in substrate entry or ring stabilisation. Studying the metabolite profiles of these three mutants by GC-MS to look for accumulation of novel cyclisation products may provide further insights into the cyclisation mechanism.

Further applications of the AsbAS1 antibody could be to look for the presence of β -amyrin synthase in other oat species as well as other cereals and grasses. The AsbAS1 antibody could also be used to determine the subcellular localisation of AsbAS1. Currently studies on yeast have shown that OSCs could be associated with lipid particles, but the subcellular localisation of OSCs in plants is currently unclear. Immunolocalisation of oat root sections with the AsbAS1 antibody could determine whether AsbAS1 is localised to the endoplasmic reticulum or a subcellular organelle.

Chapter 5 described the cloning and expression of AsbAS1 in *Pichia pastoris*. AsbAS1 was successfully expressed and resulted in production of transformants that expressed functional AsbAS1. Partial purification of C-terminally 6-histidine tagged AsbAS1 protein has been achieved in wild-type *Pichia* cells but contamination with the *Pichia* AOX1 led the AsbAS1 constructs to be expressed in an AOX1 deficient strain. It is clear that AsbAS1 expression levels in both the wild-type and AOX1 deficient strains need to be optimised and following the identification of highly expressing functional tagged transformants (pPICZB-BAS2 and pPICZB-BAS4) in the AOX1 deficient strain, large scale purification can be repeated. The availability of a highly specific AsbAS1 antibody will allow for accurate detection of AsbAS1 in induced cultures. If expression of AsbAS1 in the AOX1 deficient strain removes the AOX1 protein contamination, then the purification protocol can be optimised to obtain pure AsbAS1 protein that can be used for kinetic and structural studies.

Appendices

Appendix 1: BLAST search of AsbAS1 against the UniProt database

Results table for BLAST search carried out against the UniProt Knowledge database using the AsbAS1 cDNA sequence as a query.

UNIPROT:ID	Source	Length (AA)	Score	Identity %	Positives %	E value
Q93WU1_9POAL	Beta-amyrin synthase <i>Avena strigosa</i>	757	4059	98	98	0.0
Q6IWA0_9POAL	Beta-amyrin synthase <i>Avena longiglumis</i>	757	4055	98	98	0.0
Q6IW99_9POAL	Beta-amyrin synthase <i>Avena prostrata</i>	757	4041	97	98	0.0
Q6IWA1_9POAL	Beta-amyrin synthase <i>Avena longiglumis</i>	757	4031	97	98	0.0
Q6IWA2_9POAL	Beta-amyrin synthase <i>Avena clauda</i>	757	3859	92	96	0.0
Q6IW97_9POAL	Beta-amyrin synthase <i>Avena ventricosa</i>	757	3852	92	95	0.0
C5YJE0_SORBI	Putative uncharacterized protein <i>Sorghum bicolor</i>	753	2845	66	80	7.9e-295
Q2R712_ORYSJ	Cycloartenol synthase, putative, expressed <i>Oryza sativa</i> subsp. <i>japonica</i>	760	2789	66	78	6.8e-289
B6SVP8_MAIZE	Cycloartenol synthase <i>Zea mays</i>	760	2680	63	78	2.4e-277
C0PGA6_MAIZE	Putative uncharacterized protein <i>Zea mays</i>	762	2677	62	78	5.0e-277
Q0IT98_ORYSJ	Os11g0286800 protein <i>Oryza sativa</i> subsp. <i>japonica</i>	756	2662	62	77	2.0e-275
B6U7M4_MAIZE	Cycloartenol synthase <i>Zea mays</i>	762	2661	62	78	2.5e-275
B9GB84_ORYSJ	Putative uncharacterized protein <i>Oryza sativa</i> subsp. <i>japonica</i>	762	2593	61	78	4.0e-268
Q0IS49_ORYSJ	Os11g0562100 protein <i>Oryza sativa</i> subsp. <i>japonica</i>	762	2593	61	78	4.0e-268
C5YRX4_SORBI	Putative uncharacterized protein Sb08g021260 <i>Sorghum bicolor</i>	733	2211	59	75	1.9e-251
C5Y5B0_SORBI	Putative uncharacterized protein Sb05g022745 (Fragment) <i>Sorghum bicolor</i>	627	2410	68	82	9.9e-249
A4VB63_9LILI	Cycloartenol synthase <i>Dioscorea zingiberensis</i>	759	2376	54	74	4.0e-245
Q8VWY4_COSSP	Cycloartenol synthase <i>Costus speciosus</i>	759	2376	55	73	4.0e-245
C7U323_9LILI	Cycloartenol synthase <i>Dioscorea zingiberensis</i>	759	2374	54	74	6.5e-245
Q6Z2X6_ORYSJ	Os02g0139700 protein <i>Oryza sativa</i> subsp. <i>japonica</i>	759	2365	55	73	5.8e-244
A7BJ36_9ROSI	Cycloartenol synthase <i>Kandelia candel</i>	758	2364	54	74	7.4e-244
Q2WGL6_LOTJA	Cycloartenol synthase <i>Lotus japonicus</i>	757	2360	54	73	2.0e-243
Q2XPU6_RICCO	Cycloartenol synthase <i>Ricinus communis</i>	759	2358	54	74	3.2e-243
B0FPA2_9APIA	Cycloartenol synthase <i>Panax notoginseng</i>	758	2357	54	73	4.1e-243
Q9SXV6_GLYGL	Cycloartenol synthase <i>Glycyrrhiza glabra</i>	757	2357	55	73	4.1e-243
O82139_PANGI	Cycloartenol Synthase <i>Panax ginseng</i>	758	2355	54	73	6.7e-243
Q9SPJ0_ORYSA	Putative cycloartenol synthase (Fragment)	757	2349	54	73	2.9e-242
A7BJ35_RHISY	Cycloartenol synthase <i>Rhizophora stylosa</i>	758	2347	54	73	4.7e-242

Q941S0_9POAL	Cycloartenol synthase <i>Avena strigosa</i>	759	2347	54	73	4.7e-242
Q6IWA6_9POAL	Cycloartenol synthase <i>Avena longiglumis</i>	759	2346	54	73	6.0e-242
Q6IWA5_9POAL	Cycloartenol synthase <i>Avena prostrata</i>	759	2343	54	73	1.2e-241
B9HBB8_POPTR	Predicted protein <i>Populus trichocarpa</i>	761	2340	54	72	2.6e-241
Q6IWA7_9POAL	Cycloartenol synthase <i>Avena longiglumis</i>	759	2340	54	72	2.6e-241
Q8W3Z3_9ROSI	Cycloartenol synthase <i>Betula platyphylla</i>	757	2337	53	73	5.4e-241
Q6IWA4_9POAL	Cycloartenol synthase <i>Avena strigosa</i>	759	2335	53	73	8.8e-241
C5Z987_SORBI	Putative uncharacterized protein Sb10g029175 (Fragment) <i>Sorghum bicolor</i>	649	1349	71	84	5.2e-240
C5XUJ6_SORBI	Putative uncharacterized protein Sb04g003200 <i>Sorghum bicolor</i>	757	2325	54	72	1.0e-239
Q6IWA3_9POAL	Cycloartenol synthase <i>Avena ventricosa</i>	759	2317	53	72	7.1e-239
Q6IWA8_9POAL	Cycloartenol synthase <i>Avena clauda</i>	759	2316	53	72	9.1e-239
CAS1_ARATH	Cycloartenol synthase <i>Arabidopsis thaliana</i>	759	2314	53	72	1.5e-238
Q6QZW7_9APIA	Cycloartenol synthase <i>Centella asiatica</i>	757	2312	53	72	2.4e-238
A9QW75_9FABA	Cycloartenol synthase 1 <i>Polygala tenuifolia</i>	761	2306	54	72	1.0e-237
Q8W3Z4_9ROSI	Cycloartenol synthase <i>Betula platyphylla</i>	767	2299	53	71	5.7e-237
Q2TU69_9APIA	Cycloartenol synthase <i>Bupleurum kaoi</i>	757	2296	52	72	1.2e-236
C0P7K9_MAIZE	Putative uncharacterized protein <i>Zea mays</i>	759	2295	53	72	1.5e-236
D3WI27_CIMRA	Putative 2,3 oxidosqualene cyclase <i>Cimicifuga racemosa</i>	757	2295	53	72	1.5e-236
O23909_PEA	Cycloartenol synthase <i>Pisum sativum</i>	756	2290	53	71	5.2e-236
Q6BE25_CUCPE	Cycloartenol synthase <i>Cucurbita pepo</i>	766	2287	52	72	1.1e-235
Q9SLP9_LUFCY	Cycloartenol synthase <i>Luffa cylindrica</i>	765	2282	52	72	3.6e-235
B1PLH3_SOLLC	Cycloartenol synthase <i>Solanum lycopersicum</i>	757	2280	51	73	5.9e-235

Appendix 2: BLAST search of AsbAS1 against the Protein Data Bank

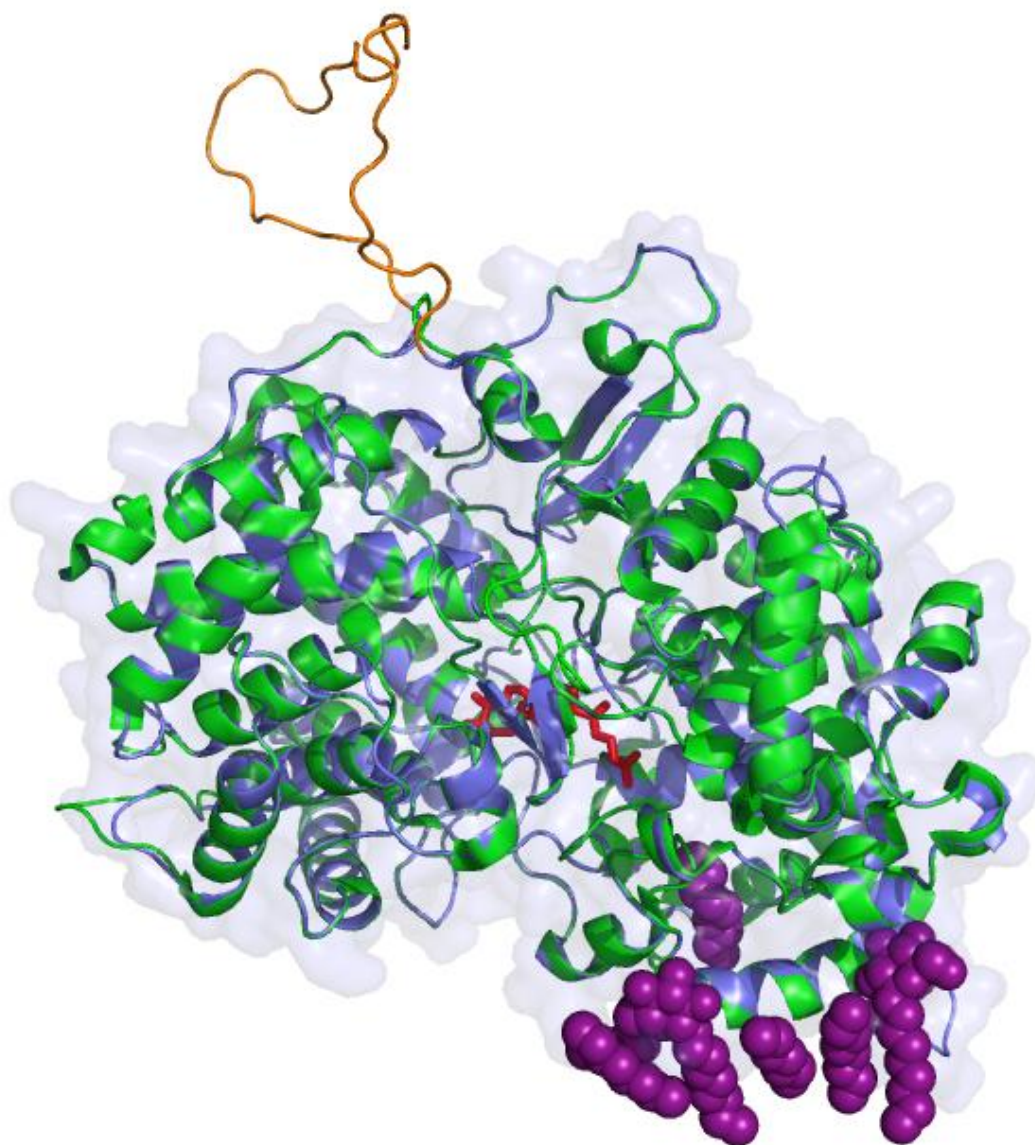
Results table for BLAST search carried out against the RCSB Protein Data Bank (PDB) using the AsbAS1 cDNA sequence as a query.

PDB ID	Source	Length	Score	Identity (%)	Positives (%)	E value
1W6K_A	LANOSTEROL SYNTHASE	732	1310	40	56	e-144
1W6J_A	LANOSTEROL SYNTHASE	732	1309	40	56	e-144
2SQC_B	SQUALENE-HOPENE CYCLASE	631	316	23	40	7e-29
2SQC_A	SQUALENE-HOPENE CYCLASE	631	316	23	40	7e-29
3SQC_C	SQUALENE--HOPENE CYCLASE	631	312	23	40	2e-28
3SQC_B	SQUALENE--HOPENE CYCLASE	631	312	23	40	2e-28
3SQC_A	SQUALENE--HOPENE CYCLASE	631	312	23	40	2e-28
1UMP_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1UMP_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1UMP_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1SQC_	SQUALENE-HOPENE CYCLASE	631	309	23	40	5e-28
1O79_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1O79_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1O79_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1O6R_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1O6R_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1O6R_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1O6Q_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1O6Q_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1O6Q_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1O6H_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1O6H_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1O6H_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H3C_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H3C_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H3C_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H3B_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H3B_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H3B_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H3A_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H3A_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H3A_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H39_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H39_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H39_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H37_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H37_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H37_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H36_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H36_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H36_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H35_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H35_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1H35_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1GSZ_C	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1GSZ_B	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28

1GSZ_A	SQUALENE--HOPENE CYCLASE	631	309	23	40	5e-28
1LTX_B	GERANYLGERANLTRANSFERASE	331	81	30	51	0.12
1DCE_D	GERANYLGERANYLTRANSFERASE	331	81	30	51	0.12
1DCE_B	GERANYLGERANYLTRANSFERASE	331	81	30	51	0.12

Appendix 3: Structure of original AsbAS1 model

Superimposed image of hOSC crystal structure (green) and the original AsbAS1 model (lilac) with the 33 amino acid unordered loop shown in orange. The product of hOSC catalysis – lanosterol (red sticks) is shown bound in the active site. Detergent molecules (purple spheres) used for solubilisation and crystallisation of hOSC show the orientation of the enzyme at the membrane. A detergent molecule is also bound in the substrate access channel. The unordered loop extends away from the protein surface and no secondary structure could be assigned to this region. It did not appear to interact with any of the structural or mechanistic features of the enzyme, hence was removed from the final model.



Appendix 4: Table of characterised plant oxidosqualene cyclases

Table of all currently characterised plant OSCs used to construct the phylogenetic tree in Figure 2.7. The OSCs are grouped according to their enzyme classification and are shaded using the colours of the corresponding clade on the phylogenetic tree. A total of 101 sequences were analysed and consisted of 81 sequences from cloned and functionally characterised plant OSCs and 20 sequences for predicted OSCs from a variety of plant species.

Enzyme name	Organism	Abbreviation in phylogenetic tree	Common abbreviation	Cloned	Functionally expressed	GenBank Accession number
Cycloartenol synthase	<i>Cucurbita pepo</i>	CP-CAS	CPX	Yes	Yes	BAD34644
Cycloartenol synthase	<i>Luffa cylindrica</i>	LC-CAS1	-	Yes	Yes	BAA85266
Cycloartenol synthase	<i>Betula platyphylla</i>	BP-CAS1	BPX	Yes	Yes	BAB83085
Cycloartenol synthase	<i>Solanum lycopersicum</i>	SL-CAS1	-	Yes	No	ACA28830
Cycloartenol synthase	<i>Olea europaea</i>	OE-CAS	OEX	Yes	No	BAA86931
Cycloartenol synthase	<i>Kandelia candel</i>	KC-CAS	KcCAS	Yes	Yes	BAF73930
Cycloartenol synthase	<i>Rhizophora stylosa</i>	RS-CAS	RsCAS	Yes	Yes	BAF73929
Cycloartenol synthase	<i>Ricinus communis</i>	RC-CAS	RcCAS	Yes	Yes	ABB76767
Cycloartenol synthase	<i>Arabidopsis thaliana</i>	AT-CAS1	-	Yes	Yes	AAC04931
Cycloartenol synthase	<i>Betula platyphylla</i>	BP-CAS2	BPX2	Yes	Yes	BAB83086
Cycloartenol synthase	<i>Polygala tenuifolia</i>	PT-CAS1	-	Yes	No	ABX75046
Cycloartenol synthase	<i>Bupleurum kaoi</i>	BK-CAS	-	Yes	No	AAS83469
Cycloartenol synthase	<i>Centella asiatica</i>	CA-CAS	-	Yes	No	AAS01524
Cycloartenol synthase	<i>Glycyrrhiza glabra</i>	GG-CAS	GgCAS1	Yes	Yes	BAA76902
Cycloartenol synthase	<i>Lotus japonicus</i>	LJ-CAS	OSC1	Yes	Yes	BAE53431
Cycloartenol synthase	<i>Panax ginseng</i>	PG-CAS	pMM31	Yes	Yes	BAA33460
Cycloartenol synthase	<i>Pisum sativum</i>	PS-CAS	PNX	Yes	Yes	BAA23533
Cycloartenol synthase	<i>Allium macrostemon</i>	AM-OSC	-	Yes	No	BAA84603
Cycloartenol synthase	<i>Avena strigosa</i>	AS-CAS	AsCS1	Yes	Yes	AAT38891
Cycloartenol synthase	<i>Costus speciosus</i>	CS-CAS	CSOCS1	Yes	Yes	BAB83253
Cycloartenol synthase	<i>Dioscorea zingiberensis</i>	DZ-CAS	-	Yes	?	CAM91422
Putative cycloartenol synthase	<i>Sorghum bicolor</i>	SB-OSC3	-	No	No	XP_002453283
Putative cycloartenol synthase	<i>Sorghum bicolor</i>	SB-OSC4	-	No	No	XP_002453282
Putative cycloartenol synthase	<i>Oryza sativa</i>	OS-OSC1	-	No	No	-
Putative cycloartenol synthase	<i>Oryza sativa</i>	OS-OSC2	-	No	No	-

Cycloartenol synthase	<i>Dictyostelium discoideum</i>	DD-CAS	-	Yes	Yes	AAF80384
Putative cycloartenol synthase	<i>Chlamydomonas reinhardtii</i>	CR-CAS1	-	No	?	XP_001689874
Predicted oxidosqualene cyclase	<i>Physcomitrella patens</i>	PP-OSC	-	No	No	XP_001758555
Cycloartenol synthase	<i>Abies magnifica</i>	AbM-CAS1	-	Yes	?	AAG44096
Cycloartenol synthase	<i>Adiantum capillus-veneris</i>	AC-CAS	ACX	Yes	Yes	BAF93208
Cycloartenol synthase	<i>Stigmatella aurantiaca</i>	SA-CAS	-	Yes	Yes	AJ494839
Cucurbitadienol synthase	<i>Cucurbita pepo</i>	CP-CuS	CPQ	Yes	Yes	BAD34645
Lanosterol synthase	<i>Arabidopsis thaliana</i>	AT-LAS	LSS	Yes	Yes	NP_190099
Putative lanosterol synthase	<i>Cucurbita pepo</i>	CP-OSC	CPR	Yes	No	BAD34646
Lanosterol synthase	<i>Lotus japonicus</i>	LJ-LAS	OSC7	Yes	Yes	BAE95410
Putative lanosterol synthase	<i>Luffa cylindrica</i>	LC-OSC	Lc-OSC2	Yes	No	BAA85267
Lanosterol synthase	<i>Panax ginseng</i>	PG-LAS	PNZ	Yes	Yes	BAA33462
Putative lanosterol synthase	<i>Taraxacum officinale</i>	TO-OSC	TRV	Yes	Non-functional	BAA86933
Multifunctional triterpene synthase	<i>Costus speciosus</i>	CS-MTS	CSOSC2	Yes	Yes	BAB83254
Putative triterpene synthase	<i>Oryza sativa</i>	OS-OSC5	-	No	No	-
Putative triterpene synthase	<i>Oryza sativa</i>	OS-OSC6	-	No	No	-
Putative pentacyclic triterpene synthase	<i>Triticum aestivum</i>	TA-OSC	-	No	No	AX382063
Triterpene synthase	<i>Sorghum bicolor</i>	SB-OSC1	-	No	No	XP_002445263
β -amyrin synthase	<i>Avena strigosa</i>	AS-bAS	AsbAS1	Yes	Yes	CAC84558
putative β -amyrin synthase	<i>Oryza sativa</i>	OS-OSC8	-	No	No	-
cycloartenol synthase	<i>Zea Mays</i>	ZM-CAS	-	No	No	ACG28931
Putative triterpene synthase	<i>Oryza sativa</i>	OS-OSC10	-	No	No	-
Putative triterpene synthase	<i>Sorghum bicolor</i>	SB-OSC2	-	No	No	XP_002442519
Putative triterpene synthase	<i>Oryza sativa</i>	OS-OSC11	-	No	No	-
Putative triterpene synthase	<i>Oryza sativa</i>	OS-OSC7	-	No	No	-
Putative triterpene synthase	<i>Oryza sativa</i>	OS-OSC12	-	No	No	-
Lupeol synthase	<i>Betula platyphylla</i>	BP-LUS	BPW	Yes	Yes	BAB83087
Lupeol synthase	<i>Glycyrrhiza glabra</i>	GG-LUS1	GgLUS1	Yes	Yes	BAD08587
Lupeol synthase	<i>Lotus japonicus</i>	LJ-LUS	OSC3	Yes	Yes	BAE53430
Lupeol synthase	<i>Olea europaea</i>	OE-LUS	OEW	Yes	Yes	BAA86930
Lupeol synthase	<i>Taraxacum officinale</i>	TO-LUS	TRW	Yes	Yes	BAA86932
putative β -amyrin synthase	<i>Centella asiatica</i>	CA-DAS	CaDDS	Yes	Yes	AAS01523










Multifunctional triterpene synthase	<i>Olea europaea</i>	OE-MTS	OEA	Yes	Yes	BAF63702
Dammarenediol-II-synthase	<i>Panax ginseng</i>	PG-DAS	PNA	Yes	Yes	BAF33291
Arabidiol synthase	<i>Arabidopsis thaliana</i>	AT-ARS	At4g15340	Yes	Yes	BAF33292
Baurool synthase	<i>Arabidopsis thaliana</i>	AT-BARS1	BARS1	Yes	Yes	NP_193272
Thalianol synthase	<i>Arabidopsis thaliana</i>	AT-THAS	At5g48010	Yes	Yes	BAB11065
Marneral synthase	<i>Arabidopsis thaliana</i>	AT-MARS	MRN1	Yes	Yes	NP_199074
Multifunctional triterpene synthase	<i>Arabidopsis thaliana</i>	AT-MTS	T30F21.16	Yes	Yes	NP_177971
Predicted lupeol synthase	<i>Arabidopsis thaliana</i>	AT-PEN3	-	No	No	NP_198464
Multifunctional triterpene synthase	<i>Rhizophora stylosa</i>	RS-MTS2	RsM2	Yes	Yes	BAF80442
Lupeol synthase	<i>Bruguiera gymnorrhiza</i>	BG-LUS	BgLUS	Yes	Yes	BAF80444
putative oxidosqualene cyclase	<i>Euphorbia tirucalli</i>	ET-OSC	EtOSC	Yes	No	BAE43643
Multifunctional triterpene synthase	<i>Kandelia candel</i>	KC-MTS	KcMS	Yes	Yes	BAF35580
Lupeol synthase	<i>Ricinus communis</i>	RC-LUS	RcLUS	Yes	Yes	ABB76766
Isomultiflorenol synthase	<i>Luffa cylindrica</i>	LC-IMS	LcIMS1	Yes	Yes	BAB68529
Lupeol synthase	<i>Arabidopsis thaliana</i>	AT-LUP1	LUP1	Yes	Yes	NP_178018
Multifunctional triterpene synthase	<i>Arabidopsis thaliana</i>	AT-LUP2	LUP2	Yes	Yes	NP_178017
Multifunctional triterpene synthase	<i>Arabidopsis thaliana</i>	AT-LUP3	F1019.4	Yes	Yes	NP_176868
Camelliol synthase	<i>Arabidopsis thaliana</i>	AT-CamS	CAMS1	Yes	Yes	NP_683508
β -amyirin synthase	<i>Arabidopsis thaliana</i>	AT-bAS	AtBAS	Yes	Yes	AB374428
Baccharis oxide synthase	<i>Stevia rebaudiana</i>	SR-BOS	StrBOS	Yes	Yes	BAH23676
β -amyirin synthase	<i>Gentiana straminea</i>	GS-bAS	GSAs1	Yes	Yes	ACO24697
β -amyirin synthase	<i>Nigella sativa</i>	NS-bAS	Ns-bAS1	Yes	Yes	ACH88048
β -amyirin synthase	<i>Bupleurum kanoi</i>	BK-bAS	-	Yes	No	AAS83468
β -amyirin synthase	<i>Panax ginseng</i>	PG-bAS1	PNY	Yes	Yes	BAA33461
β -amyirin synthase	<i>Saponaria vaccaria</i>	VH-bAS	SvBS	Yes	Yes	ABK76265
β -amyirin synthase	<i>Gypsophila paniculata</i>	GP-bAS	-	?	?	ABC33922
β -amyirin synthase	<i>Euphorbia tirucalli</i>	ET-bAS	EtAS	Yes	Yes	BAE43642
β -amyirin synthase	<i>Betula platyphylla</i>	BP-bAS	BPY	Yes	Yes	BAB83088
Multifunctional triterpene synthase	<i>Rhizophora stylosa</i>	RS-MTS1	RsM1	Yes	Yes	BAF80441
β -amyirin synthase	<i>Bruguiera gymnorrhiza</i>	BG-bAS	BgbAS	Yes	Yes	BAF80443
β -amyirin synthase (isogene)	<i>Panax ginseng</i>	PG-bAS2	-	Yes	?	BAA33722
β -amyirin synthase	<i>Artemisia annua</i>	AA-bAS	AaBAS	Yes	Yes	ACB87531

β -amyrin synthase	<i>Aster sedifolius</i>	AstS-bAS	AsOXA1	Yes	?	AAX14716
Mixed amyrin synthase	<i>Pisum sativum</i>	PS-MTS	PSM	Yes	Yes	BAA97559
Multifunctional triterpene synthase	<i>Lotus japonicus</i>	LJ-MTS	LjAMY2	Yes	Yes	AAO33580
β -amyrin synthase	<i>Polygala tenuifolia</i>	PT-bAS	-	Yes	Yes	ABL07607
β -amyrin synthase	<i>Medicago truncatula</i>	MT-bAS1	-	Yes	Yes	CAD23247
β -amyrin synthase	<i>Pisum sativum</i>	PS-bAS	PSY	Yes	Yes	BAA97558
β -amyrin synthase	<i>Glycine max</i>	GM-bAS	-	Yes	No	AAM23264
β -amyrin synthase	<i>Glycyrrhiza glabra</i>	GG-bAS1	GgbAS1	Yes	Yes	BAA89815
β -amyrin synthase	<i>Lotus japonicus</i>	LJ-bAS	LjAMY1	Yes	Yes	BAE53429







Appendix 5: Multiple sequence alignment of characterised plant OSCs

Multiple sequence alignment of all characterised plant OSCs shown in Appendix 4. The alignment was generated in MUSCLE and viewed and edited using Jalview. The OSCs are grouped in the alignment according to their enzyme classification and are shaded using the colours of the corresponding clade on the phylogenetic tree (Horizontal key). Important conserved structural features and amino acids involved in the enzyme mechanism are also annotated (Vertical key). Both the horizontal and vertical keys are shown below. The OSCs are named according to their abbreviations in the phylogenetic tree (Figure 2.7).

HORIZONTAL KEY

-  DICOT β -amyrin synthases
-  Arabidopsis triterpene synthases
-  DICOT multifunctional triterpene synthases
-  DICOT lupeol synthases
-  MONOCOT triterpene synthases
-  DICOT lanosterol synthases
-  MONOCOT cycloartenol synthases
-  DICOT cycloartenol synthases
-  AsbAS

VERTICAL KEY

-  Substrate access
-  QW motif
-  DCTAE catalytic motif
-  Pre-folding of the B-ring
-  Catalytic initiation
-  Ring stabilisation (letter(s) denote(s) which ring residue is involved in stabilising)
 - Except monocot triterpene OSCs
 - ◆ Only in monocot triterpene OSCs

Frame 1

RS-NTS1	-----MWR-LKIA-----EG--GNDP-YLYSTNNYVGRQ-IWEFD-P-D-AGTPEERA---KAE EARQNFYKNR--YQVKPSGDLLW-----RLQ--FLREKH--FKQ---TIPQ--V
BG-bAS	-----MWR-IKIA-----EG--GKDP-YLYSTNNYVGRQ-TWEFD-P-D-AGTPEERA---EVE EARQNFYKNR--YQVKPCGDLLW-----RLQ--FLGEKH--FEQ---TIPQ--V
ET-bAS	-----MWK-LKIA-----EG--G-NDE-YLYSTNNYVGRQ-TWVFD-P-Q-PPTPQELA---QQQARLNHFYNNR--YHVKPSDILLW-----RFQ--FLREKH--FKQ---TIPQ--A
BP-bAS	-----MWR-LKIA-----DG--GSDP-YIYSTNNFVGRQ-TWEFD-P-Q-AGSPQERA---EVE EARRRNFYDNR--YQVKPSGDLLW-----RMQ--FLKEKH--FKQ---TIPP--V
PG-BAS1	-----MWK-LKIA-----EG--NKNDP-YLYSTNNFVGRQ-TWEFD-P-DYVASPGELE---EVE QVRRQFWDNR--YQVKPSGDLLW-----RMQ--FLREKH--FRQ---TIPQ--V
BK-bAS	-----MWR-LKIA-----EG--SEDDK-YLYSTNNYVGRQ-TWEFD-P-N-YGSREEKA---QVDEARLNHFQNR--YKVKPSGDVLLW-----QMQ--FLREKH--FKQ---TIPQ--V
PG-BAS2	-----MWR-LMTA-----KG--GNDP-YLYSTNNFVGRQ-TWEFD-P-D-YGTPAERA---EVE EARLHFVNNR--YQVKPSDVLW-----RMQ--FLKEKH--FKQ---IIPQ--V
VH-bAS	-----MWR-LKIA-----EG--ANDP-YLYSTNNFVGRQ-TWEFD-T-D-YGTPEAIK---EVE EARQDFYKNR--FQVKPCGDLLW-----RFQ--FLREKH--FKQ---TIPQ--V
GP-bAS	-----MWR-LKIA-----EG--RNDP-YLYSTNNFVGRQ-IWEFD-S-N-YGTPEERA---EVE QARVDFWNNR--HEVKPSDVLW-----RMQ--FLREKH--FEQ---TIPQ--V
AstS-bAS1	-----MWR-MNIA-----KG--GNDP-YLYSTNNYVGRQ-IWEFD-P-N-YGTPEELA---EVE QARAEFVNNR--HKVKTSSDVLW-----RMQ--FLREKH--FKQ---TIPQ--V
PS-NTS	-----MWK-LKIG-----DG--G-KDR-NIFSTNNFVGRQ-TWEFD-P-D-AGTSQEKA---QVE AARQHFDNR--FEVKACSDLLW-----RFQ--FLREKH--FKQ---TIES--V
LJ-NTS	-----MWK-LKVA-----DG--GKHP-YIFSIHNFVGRQ-TWEYD-P-D-AGTPEERA---QVE EARQDFYNNR--YKVKTCGDRLLW-----RFQ--VMRENN--FKQ---TIPS--V
NT-BAS1	-----MWK-LKIG-----EG--KNEP-YLFSTNNFVGRQ-TWEYD-P-E-AGSEERA---QVE EARKNFYDNR--FQVKPCGDLLW-----RFQ--VLRENN--FMQ---TIDG--V
GG-bAS	-----MWR-LKIA-----EG--GKDP-YIYSTNNFVGRQ-TWEYD-P-D-GGTPEERA---QVE AARLHFYNNR--FQVKPCADLLW-----RFQ--VLRENN--FKQ---TIDG--V
GN-bAS	-----MWR-LKIA-----DG--GNDP-YIFSTNNFVGRQ-TWEFD-P-E-AGSPEERA---QVE AARQHFDNR--FQVKPCADLLW-----RFQ--VLRENN--FKQ---TIPR--V
LJ-bAS	-----MWK-LKVA-----DG--GKDP-YIFSTNNFVGRQ-TWEYD-P-D-AGTPEERA---QVE EARQDFYNNR--YKVKPCGDLLW-----RFQ--VLRENN--FKQ---TIPS--V
PT-bAS	-----MWR-LKVG-----EG--KNDP-YLFSTNDYTGRQ-TWEFD-P-D-AGTPEERA---EVE AARQAFYDNR--FQFKHCGDLLW-----RFQ--FLRDKH--FKQ---TIPK--V
PS-bAS	-----MWR-LKIA-----EG--GNDP-YLFSTNNFVGRQ-TWEYD-P-E-AGSEERA---QVE EARRRNFYNNR--FEVKPCGDLLW-----RFQ--VLRENN--FKQ---TIGG--V
NS-bAS	-----MWK-LKIA-----DATGPFSE-YLYSTNNYIGRQ-TWEFD-P-D-AGTPEERA---EVE KARQDFHKNR--FDIKPCGDVLL-----RLQ--MLKENKDRFDL---SIPP--V
GS-bAS	-----MWR-LKIG-----EG--GNDP-NLYSTNNYVGRQ-IWEYD-R-N-AGIAEERE---EVE EARLHFVNNR--RHVKPSDILLW-----RLQ--FLREKH--FKQ---SIPA--V
SR-BOS	-----MWR-LKIA-----DG--NNHP-YLYSTNNFVGRQ-TWEFD-P-N-YGTQEERD---EVE QARQHFWNNR--HQFKATGDVLLW-----RMQ--FIREKR--FKQ---TIPQ--V
AT-LUP1	-----MWK-LKIG-----KG--NGEDP-HLFSSNNFVGRQ-TWVFD-H-K-AGSPEERA---AVE EARRGFLDNR--FRVKGCSDLLW-----RMQ--FLREKH--FEQ---GIPP--L
AT-LUP3	-----MWR-LKVG-----EG--GKDP-YLFSSNNFVGRQ-TWEFD-P-K-AGTREERT---AVE EARRFDDNR--SRVKPSDILLW-----KMQ--FLKEAK--FEQ---VIPP--V
AT-LUP2	-----MWK-LKIG-----EG--NGEDP-YLFSSNNFVGRQ-TWEFD-P-K-AGTPEERA---AVEDARRHYLDNR--PRVKGCSDLLW-----RMQ--FLKEAK--FEQ---VIPP--V
AT-bAS	-----MWR-LKIG-----EG--NGDDP-YLFTTNNFVGRQ-TWEFD-P-D-GGSPEERH---SVVEARRIFYDNR--FHVKASDILLW-----RMQ--FLREKH--FEQ---RIAP--V
AT-CamS	-----MWK-LKIA-----NG--NKEEP-YLFSTNNFLGRQ-TWEFD-P-D-AGTVEELA---AVE EARRKIFYDNR--FRVKASDILLW-----RMQ--FLKEKK--FEQ---VIPP--A
LC-INS	-----MWR-LKVA-----DG--GNDP-YIYSMNNFVGRQ-IWEFD-P-N-AGTPEERA---EIERLRHHFTKNR--HKGFPASDILLW-----RVQ--LLREKH--FKQ---SIPA--V
RC-LUS	-----MWR-IKIA-----EG--GNHP-YIYSTNNFVGRQ-IWVFD-P-N-AGTPEEQA---EVE EARQNFVKNR--FQVKPNSDILLW-----QLQ--FLREKH--FKQ---KIPK--V
ET-OSC	-----MWR-LKIA-----EQG--NHP-NIYSTNNFVGRQ-IWEYD-A-N-AGTPQDIA---EVEDARLRFVNNR--FHIKPSDILLW-----QFQ--FLREKH--FKQ---RIPA--V
KC-NTS	-----MWR-LKIA-----EG--GDHP-YIYSTNNFVGRQ-TWEFE-P-E-AGTPEERA---QVE EARQNFVWRDR--FRIKPCSDILLW-----RFQ--FLREKH--FKQ---IIPQ--G
BG-LUS	-----MWR-LKIA-----EG--GNHP-YIYSTNNFVGRQ-TWEFD-P-E-AGTPEERA---QVE EARNFVWRDR--FLIKPSDILLW-----RFQ--FLSEKK--FKQ---RIPQ--V
RS-NTS2	-----MGVWR-LKIG-----EG--ANHP-YLTSTNNFVGRQ-TWVFE-P-D-GGTPEERD---QVE EARQNFYKNR--FRVPCSDILLW-----QMQ--FLREKH--FRQ---KIPQ--V
AT-NTS	-----MWR-LKIG-----AKG--GDETHLFTTNNYVGRQ-TWEFD-A-D-ACSPEELA---EVE EARQNFVNNR--SRFKISADLLW-----RMQ--FLREKH--FEQ---KIPR--V
AT-ARS	-----MWR-LRIG-----AKAG--NDT-HLFTTNNYVGRQ-IWEFD-A-N-AGSPQELA---EVE EARRRNFVNNR--SHYKASADLLW-----RMQ--FLREKH--FEQ---KIPR--V
AT-THAS	-----MWR-LRTG-----PK-----AG--EDT-HLFTTNNYVGRQ-IWEFD-A-N-AGSPQELA---EVEDARHKFSNNT--SRFKTTADLLW-----RMQ--FLREKH--FEQ---KIPR--V
AT-PEN3	-----MWR-LRIG-----AKAG--DDP-HLCTTNNFVGRQ-IWEFD-A-N-AGSPAELS---EVDQARQNFVNNR--SQYKACADLLW-----RMQ--FLREKH--FEQ---KIPR--V
AT-BARS1	-----MWR-LRIG-----AK--DHT-HLFTTNNYVGRQ-IWEFD-A-N-AGSPEELA---EVE EARRRNFVNNR--SRFKASADLLW-----RMQ--FLREKH--FEQ---KIPR--V
AT-NARS	-----MWR-LRIG-----AE-----AR--QDP-HLFTTNNFVGRQ-IWEFD-A-N-GGSPEELA---EVE EARLNFAANNK--SRFKASPDLLW-----RRQ--FLREKH--FEQ---KIPR--V
CA-bAS	-----MWK-LKIA-----EG--NGA-YLYSTNNFVGRQ-IWEYD-P-D-AGTPEERL---EVE KLRETYKYNNL INNGIHPCGDMLM-----RLQ--LIKESG--LDL--LSIPP--V
OE-NTS	-----MWK-LKIA-----EG--HGP-YLYSTNNFVGRQ-IWEYD-P-N-GGTPEERE---AYDKAREEFQRNRKLKGVHPCGDLLM-----RIQ--LIKESG--IDL--MSIPP--V
PG-DAS	-----MWK-QKGA-----QG--NDP-YLYSTNNFVGRQ-YWVFD-P-D-AGTPEERE---EVE KARQYVNNKHLHGHPGDMLM-----RRQ--LIKESG--IDL--LSIPP--L
TO-LUS	-----MWK-LKIA-----EG--GDDE-WLTTTNNHVGGRQ-HWQFD-P-D-AGTEEERA---EIEKIRLNFKLNR--FQFKQASDILLM-----RTQ--LRKENP--INK---IPDA--I
OE-LUS	-----MWK-LKIA-----DG--TGP-WLTTTNNHIGRQ-HWEFD-P-E-AGTPDERV---EVERLREEFKKNR--FRTKQASDILLM-----RMQ--LVKEHQ--RVQ---IPPA--I
BP-LUS	-----MWK-LKIA-----EG--GLVSGNDFVGRQ-HWEFD-P-D-AGTPQERA---EVE KVRREEFTKNR--FQMKQASDILLM-----RMQ--LRKENP--CQP---IPPA--V
LJ-LUS	-----MWK-LKVA-----EG--GK-GLVSVSNFVGRQ-HWVFD-P-N-AGTPQEHE---EIERMRQEFKNR--FSIKQASDILLM-----RMQ--LRKENP--CGP---IPPA--V
GG-LUS1	-----MWK-LKIG-----EG--GA-GLISVNNFVGRQ-HWEFD-P-N-AGTPQEHA---EIERLRREFTKNR--FSIKQASDILLM-----RMQ--LRKENHYGTHN---HIPAA--V

OS-OSC7	-----MWR-LKVS-----EG--GSP-WLRVNNLLGRQ-VWEFD-P-D-LGTPEERA---DVEKARREFAEHR--FERKHSDDLMM-----RMQ--FAKENC--QKL---DLLA--V
OS-OSC12	-----MWK-LKIA-----EG--GP-WLKSGNSHVGRE-TWEFD-P-N-FGTSEERE---AVEAARIEFQKNR--FRTRHTSDVLA-----RMQ--LAKANN--FSI---DLQK--E
OS-OSC3	-----MWR-LKVA-----EG--GGA-LLRSTNGFLGRA-VWELD-P-D-HGTPEERA---DVERVRRFTDDR--LRRRESADLLM-----RMQ--FAKQKK--LQRRRDSIPPA-V
OS-OSC4	-----MWR-LKVA-----EG--GGTP-LLRSTNGFLGRA-VWEFD-P-D-HGTPEERA---DVERVRRFTDHR--LHHRRESADLLM-----RMQ--QNKHQK--RRY---RIPPVNN
OS-OSC11	-----MWR-LKVA-----QG--GGA-LLRSTNGFAGRA-VWEFD-P-D-HGTPEERA---HVERVRRDFTDHR--LRRRESADLLM-----RMQ--FARENN--HQRRGDRIPPAVN
SB-OSC2	-----MWR-LKTS-----EG--GGS-WMQTVSGFHGRQGVWEFD-P-G-AGTQEERT---KVEQLRQEFTENR--FRRRESQDLLM-----RMQ--FTGQKH--LHA---DMPAA-T
OS-OSC9	-----MWR-LKIG-----EG--AGHP-LLRSPNGFLGRE-TWEFD-P-D-AGTPEERA---EVERLRRDFTRNR--FTRRESI IKLLLP LLTASLFQ--FAKQNN--LYT---NLRV--N
OS-OSC10	-----MWK-LKFG-----EG--ASHP-LLRSSNGFLGRE-TWEFD-P-N-GGSPEEHA---VVERLRRDFTRNR--FTQRECSDDLMM-----RMQ-----EQNQVYSK---HEV--S
ZN-CAS	-----MWR-LKIG-----ER--PSHP-LLRSPNGF-GRE-VWEFD-P-A-AGTPEERA---EVDRLRQEYTRNR--FTRRQCSDDLMM-----RMQ--YAKQNL--HHT---NLPA--I
OS-OSC8	-----MWR-LKIA-----AESGGGSG---SSP-LLHTGNGFLGRA-VWEFD-P-D-AGTPEERA---EVARLRRDFTRHR--FQRKESQDLLM-----RMQALYAKLGH--LQP---DLA--V
AS-bAS	-----MWR-LTIG-----EG--GGP-WLKSNHNGFLGRQ-VWEYD-A-D-AGTPEERA---EVERVRAEFTKNR--FQRKESQDLLM-----RLQ--YAKDNP--LPA---NIPTE-A
SB-OSC1	-----MWR-LKVA-----EG--GGP-WLKSNHFRGRV-VWEFD-P-E-LGTPEERA---EVERVRRFTERR--FEKRESSDLLL-----RMQ--YAKKHK--LQV---DPPP--T
TA-OSC	-----MWR-LKVS-----EG--GGP-WLRVNNFLGRA-VWEFD-P-D-YGTPEERA---EVKRVRRFTDRR--FEKKESQDLLM-----RMQ--YAKEKH--LQV---DLPA--I
OS-OSC5	-----MWR-LKVA-----EG--GAP-GLRSCNGFLGRA-VWEFD-P-N-AGTPEERA---EVERMRREFTLRR--FERREAQDLLM-----RMQ--YAKQHR--LQV---DVPA--S
OS-OSC6	-----MWR-LKVA-----EG--SGP-WLRSTNNHVGRE-VWEFD-P-S-GGTPEEIA---EVERARETFRDHW--VEHTNSADLIM-----RLQ--FEKENP--AEM---KYSV--I
CS-NIS	-----MWR-LKVA-----EG--SGP-WLRSTNNHVGRE-VWEFD-P-S-GGTPEEIA---EVERARETFRDHW--VEHTNSADLIM-----RLQ--FEKENP--AEM---KYSV--I
PG-LAS	-----MWK-LILS-----Q--GDP-GLKSVNNHIGRQ-FWEFD-P-N-LGTPEERA---HIDKLRQQFHNNR--FRVKHSDDLMM-----RYQ--FEREKS--RKL---GDDDE--V
TO-OSC	-----MWK-LQLS-----KG--DDDP-EVRVNNHHMGRQ-YWEFD-P-H-AGTPEEQS---QIESMREFFTRNR--LNVTHSDDLMM-----RFQ--LTSKIRGGIEK---SH--V
CP-OSC	-----MWT-LKFS-----KG--WETSDNAHLGRQ-FWEFD-P-NLQPSLEEQA---RVHNVRCDFYTHR--FQAKHSDDLMM-----RLQ--LKKANG--SEV---KLPTQ--V
LC-OSC	-----MWK-LKVS-----KG--WETSENDHVGRQ-YWEFD-P-NLEPSEEEA---EIEVNCHEFHKNR--FHVKQSSDLLM-----RLQ--LRKEHASNVKL---LTKQ--I
LJ-LAS	-----MWK-LRIS-----ES---KEDELIRSVNNHVGRQ-FWEFD-P-D-LGTEQERA---QEVARKEFNQNR--FKTKHSDDLMM-----RLQ--FEREHG--VHM---KVKN--V
AT-LAS	-----MWR-LKLS-----EG--DEE---SVNHQHVGRQ-FWEYD--NQFGTSEERH---HINHRLSNFTLNR--FSSKHSDDLLY-----RFQ--CWKEKGKGMER---LPQ--V
CP-Cus	-----MWR-LKVG---AE---SVGEE---DEK-WVKSVSNHHLGRQ-VWEFCADAA-ADTPHQLL---QIQNARHNFHNNR--FHRKQSSDFLF-----AIQ--YEKEIA--KGA---KGG--V
DD-CAS	FTRMTTTHWS-LKVD-----RGRQ-TWEYS-----QEKKEATDVDI-----HLL--RLKEPG-----T
SA-CAS	-----MGR-----FE-P-----
CR-CAS1	-----MWKFISAG-----TT--GGP-LLRSLNGNKGRQ-TWEYD-P-A-AGTPEERA---EAERLREFTANK--DKHHSGDDELL-----RLQ--SADRIR--AKK---HSPPS--G
PP-OSC	-----MWR-LKIA-----EGGGGGGL-PLRTLNNHVGRQ-VWEYD-EHGAAGDLAHRV---AVDKARVAEENR--FEQHSDDLIM-----RLQ--FARENP--LPE---LPTQ--V
AbN-CAS1	-----MWK-LKIA-----EG--GSP-WLQTLNNHVGRQ-VWEFD-P-K-AGTSEDHL---AVEKARVDFYNNR--FIQHHSADLLM-----RLQ--CGGENP--LSP---LPAQ--V
AB-CAS	-----MWT-LKIA-----DG--DSGT-WLHTLNEHTGRQ-TWHFD-P-D-AGSPSDDL---AVENAREFFENR--FTKKHSADLLM-----RMQ--YAKRNP--LPP---LPNP--V
SB-OSC3	-----MWR-LRIT-----EG--GGDP-WLRTKNAHVGRQ-VWEFD-A---AADPP-A---VDAARAFVDTR--RQLKHSADLLM-----RIQ--FAKENP--LEL---DLPA--I
SB-OSC4	-----MWR-LRVA-----EDGGGGGG-GLRTKNGHVGRQ-VWEFD-A---AADPP-PA---VDAARAFVDTR--HHLKHSADLLM-----RIQ--LAKEHP--LQL---DLLA--M
AS-CAS	-----MWR-LKIA-----EG--GGDP-WLRTKNAHVGRQ-VWEFD-P-E-AGDPDELA---AVEAARRDFAAGR--HRLKHSADRLM-----RIQ--FAKENP--LKL---DLA--I
OS-OSC2	-----MWR-LRVA-----EG--GGDP-WLRTKNAHVGRQ-VWEFD-P-A-AGDPDELA---AVEAARRDFAAGR--HELKHSADLLM-----RMQ--FAKANP--LKL---DIPA--I
OS-OSC1	-----MWR-LRVA---EE---AGAG--GEG-WLSSNAHVGRQ-VWEFD-A-A-AADDDDDAAAAAEEVEAARREFFTRRR--HQIKHSADLLL-----LLQ--FTRSNP--SKL---EIPG--I
DZ-CAS1	-----MWR-LKIA-----EG--GHP-WLRTTNNHVGRQ-VWEFD-P-N-LRTPEELA---EVERAREEAFHQHR--FEKKHSDDLMM-----RLQ--FAKENP--LEL---TLPQ--V
AN-OSC	-----MWKLLKMA-----DG--GHP-WLRTTNNHVGRQ-IWEFD-P-N-LGTPEEIE---AVENAREIYRKNR--FEKKHSADLLM-----RLQ--FAKENP--AEI---GLPQ--V
CS-CAS1	-----MWR-LKIA-----EG--GGP-WLRTKNNHVGRQ-VWEFD-P-S-LGTPEEIA---EVERVREAFRETR--FEKKHSADLLM-----RLQ--FAKENP--LEM---NYP--I
KC-CAS	-----MWR-LKIA-----EG--GDP-WLRTI NNHVGRQ-AWEFD-P-SLVGSPEDIA---DIENARRNFTINR--FRHRHSADLLM-----RLQ--FAKENR--LPE---VLPK--V
RS-CAS	-----MWR-LKIA-----EG--GDP-WLRTI NNHVGRQ-VWEFD-P-SLVGSPEDIA---DIENARRNFTINR--FRHRHSADLLM-----RLQ--FAKENR--LPP---VLPK--V
RC-CAS	-----MWK-LRIA-----EG--SGHP-WLRTTNDHIGRQ-VWEFD-S-SKIGSPEELS---QIENARQNFTKNR--FIHKHSADLLM-----RIQ--FSKENP--ICE---VLPQ--V
BP-CAS2	-----MWK-LKIA-----EG--GSP-WLRTLNNHVGRQ-VWEFD-P-K-LGSPEELA---EIERARETSLKVR--FEKKHSADLLM-----RIQ--FAKENP--RGA---VLPQ--V
AT-CAS1	-----MWR-LKIA-----EG--GSP-WLRTTNNHVGRQ-FWEFD-P-N-LGTPEDLA---AVEEARKSFSNDR--FVQKHSADLLM-----RLQ--FSRENL--ISP---VLPQ--V
PG-CAS	-----MWK-LKIA-----EG--GHP-WLRTLNDHVGRQ-IWEFD-P-N-IGSPEELA---EVEKVENFRNHR--FEKKHSADLLM-----RIQ--FANEHP--GSV---VLPQ--V
CA-CAS	-----MWK-LKVA-----EG--GHP-WLRTVNNHVGRQ-IWEFD-P-K-LGSPEELA---EIEKARENFNHR--FEKQHSADLLM-----RLQ--FANEHP--RHE---VLPQ--V
BK-CAS	-----MWK-LKIA-----EG--GHP-WLRTVNDHVGRQ-VWEFD-P-K-LGSPEELA---EVEKARDNFRHR--FEKKHSADLLM-----RLQ--FANEHP--SRE---VLPQ--V
PS-CAS	-----MWK-LKVA-----EG--GTP-WLRTLNNHVGRQ-VWEFD-P-H-SGSPQDLD---DIETARNFHDR--FTHKHSADLLM-----RLQ--FAKENP--MNE---VLPK--V
GG-CAS	-----MWK-LKIA-----EG--GSP-WLRTVNNHVGRQ-VWEFD-P-K-LGSPEDLL---EIEKARQNFHNR--FTHKHSADLLM-----RTH--FAKENP--MNE---VLPK--V
LJ-CAS	-----MWK-LKIA-----EG--GHP-WLRTSNHVGRQ-VWEFD-P-K-LGSPQDLA---EIEETARNFHDR--FTHKHSADLLM-----RIQ--FSKENP--IGE---VLPK--V
SL-CAS	-----MWK-LKVA-----EG--GSP-WLRTTNNHVGRE-VWEFE-A-D-FGSPEDRA---QIEKFRHEFTKHR--LEQKHSADLLM-----RYQ--LSKEHQ--GISI---LPQ--V
OE-CAS	-----MWR-LKIA-----EG--SGSP-WLRTGNDHIGRQ-VWEFDNELE-VGSTQCLE---EIEENARCLFYENR--FQNKHSADLLM-----RIQ--FAKEYP--GTK---VLPQ--V
PT-CAS1	-----MWK-LKIG---AETARGDGGGGGGSETWLRSLNHHLGRQ-IWEFH-P-E-LGTQEELQ---QIDDARRRFRERR--FERRHSADLLM-----RIQ--FAKENP--SSA---NIPQ--V
BP-CAS1	-----MWQ-LKIGADTVPSDP SHAGG---WLSTLNNHVGRQ-VWHFH-P-E-LGSPEDLQ---QIQARQHFSDHR--FEKKHSADLLM-----RMQ--FAKHS--SFV---NLPQ--V
CP-CAS	-----MWQ-LKIGADTVPSDP SHAGG---WLSTLNNHVGRQ-VWHFH-P-E-LGSPEDLQ---QIQARQHFSDHR--FEKKHSADLLM-----RMQ--FAKHS--SFV---NLPQ--V
LC-CAS1	-----MWQ-LKIGADTVPADPSHAGG---WLSLNNHVGRQ-VWHFH-P-E-LGTPEDLQ---QIQARQHFSDHR--FEKKHSADLLM-----RMQ--FAKHS--SFV---NLPQ--I

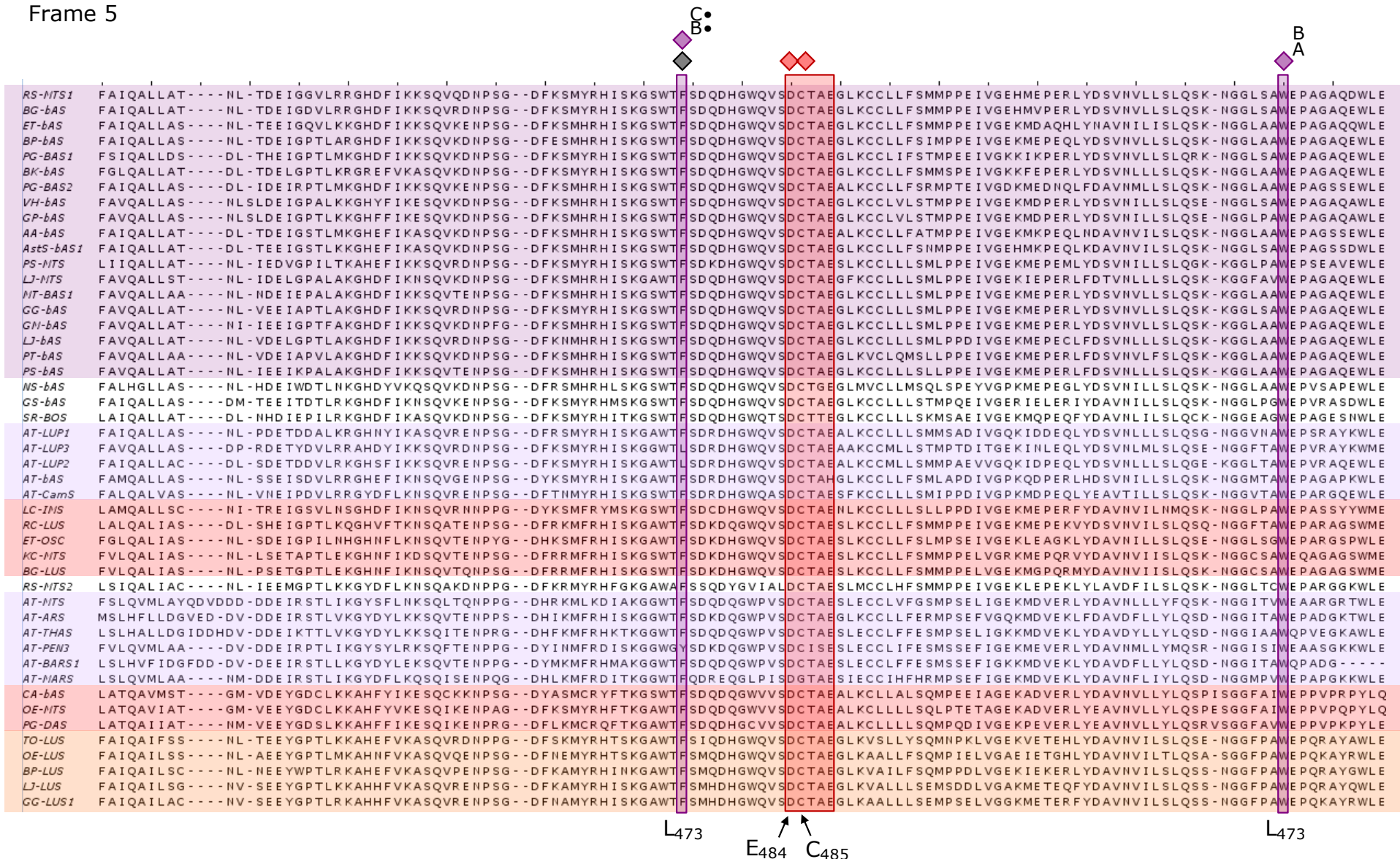
OS-OSC7	KRGEHED--VMG-----EAVWSSLK	RAISRVCNLQAHGDGHWPGDY	A-GLMFFLPGLIITLHVSGVLNTVLSSEH	QKEMRRYIYNHQ-----NE-DGGW	GLHIEGHSTMLGSSLNHYVALRLLGEG
OS-OSC12	KDGNPIN--IDT-----ATVSDILK	KALSYFSAIQAYDGHWPDDFP	-GPLFTTATMIIVLYVTESLTITLSSEH	HKEICRYLYNRQ-----NI-DGGW	GLHAEGESSMLSTALNYVALRLLGEN
OS-OSC3	KLGEKEQ--VTE-----EIAMASLR	RALDEFSSLQADDGHWPGDFF	-GVMFIMPGLIFALYVTGSLDTTISPEH	RQECRYIYHMQAKKQTKKNYIRADP-K---	---ISSSTMFGTCSNYITLRLGGEV
OS-OSC4	KLGEKEE--VTE-----EIAMASLR	RALDEFSSLQADDGHWPGDFF	-GVMFIMPGLIFALYVTGSLDTTISPEH	RREICRYIYNHQ-----AYFFPNE-DGGW	SSLSMSSTMFGTCSNYITLRLGEE
OS-OSC11	KLGEKEQ--VTE-----ETVMASLR	RALDEFSSLQADDGHWPGDLS	-GAIFIMPVLIFSLYATGSLDTVISSEH	RREICRYIYNHQ-----NE-DGGW	GMLILGSSTMFATCLNYVTLRLIGEE
SB-OSC2	KLEDGDE--VTE-----EIVQESLR	LALGWMSLQAEDGHWPGDFF	-GIMYILPFWFIFALHITGSDVVLSEH	KREICRHIYNHQ-----NE-DGGW	SFNILDDSAMFGTCLNYVTLRLGEEV
OS-OSC9	NHKDSSE--ITE-----EVLTLTKR	RVLQDQSSSLQAHGDGHWPGDFF	-GVLFILPLMIFALHVTGSLNDVLSSEH	IREICRYIYNIQ---ASHNLFIFNE-DGGW	STLTLGSPSTMFSGCVNYATLRLGEE
OS-OSC10	NHKDSSE--VTE-----EVLTLTKR	RVLQDQSSSLQAPDGYWPGGY	-GILFILPLMIFALHVTGSLNDVLSSEH	IREICRYIYNIQ-----AY--GGW	STHTLGPSSMFSGCVNYATLRLGEEV
ZN-CAS	KIDQSD--VTE-----ETILITLR	RALNQYSSLQAGPAGHWPGDY	-GISFLLPLMIFALHVTGSMDEVLSIEH	IREICRYIYNHQ-----NE-DGGW	STHTLGPSSMFGTCSNYATLRLGEEV
OS-OSC8	IVEDNQ--VTE-----ETILSSLR	RALNQYSSLQAHGDGHWPGDYS	-GILFIMPVLIFSLMVTGTLNDVLSLEH	KREICRYIYNHQ-----AY--GGW	STQVLGSPSTMFSGSCLNYATLKLGEA
AS-BAS	KLEKSTE--VTH-----ETIYESLR	RALHQYSSLQADDGHWPGDYS	-GILFIMPVLIFSLYVTRSLDTFLSPEH	RHEICRYIYNHQ-----NE-DGGW	GKMLVLPSTMFSGSCLNYATLMLGEEK
SB-OSC1	KLAENVE--VTE-----EILAAALR	RALDQHSLSLQAGDGCWPDADY	-GILFIMPVLIFALYTTGSLDTVLSKEH	RREICRYIYNHQ-----NE-DGGW	GKQVLGPSTMFSGSCLNYVALRLIGEK
TA-OSC	KLADSAQ--VTE-----ETLLTSLR	RCLSQHSALQAHGDGHWAGDFF	-GILFIMPILIFALHVTGSLNTVLSSTEH	RCEICRYIYNHQ-----NE-DGGW	STQVLGPSTMFSGSCLNYVTLRLGEEV
OS-OSC5	KLVDSTQ--VTE-----QIILASLR	RALTQHSLSLQAHGDGHWPGDFF	-GIMFIMPILIFALYITGSLDAVLSSEH	RREICRYIYNHQ-----AY--QHW	STQVLGPSTMFSGSCLNYVTLRLGEEV
CS-NTS	RIQAEEN--ISK-----EAVDICLR	KAVTRISTLQAHGDGHWPGDFF	-GAGPMFLPLPGLVIGLHITGALNTILTPEH	QKEMRRYIYNHQ-----NV-DGGW	GLSHNGPSTMFSGSCLNYVALRLLGEEG
PG-LAS	KSGSEGE--ITTSSSGVEGVKMALR	RALKFYSTIQADDGHWPGDY	-GPLFLLPGLVIGLYVMGVMDTILAKEH	QREMCRYIYNHQ-----NV-DGGW	GLHIEGCTMLCTALNYITLRLLRG
TO-OSC	KEEGDD--VTE-----EVVVKTLK	KALKFYSSLQAGDGSWPGDY	-GPLFLLPGLVIGLHVGMKDLVLSIEQ	QNEIRRYLYNHQ-----NV-DGGW	GLHIEGNTMLCTALNYVALRLLGEE
CP-OSC	KLRSSEE--MSE-----EAVETTLR	RAIRFYSTMQTDGDFWPGDY	-GPLFLLPGLVIGLVTGALNTVLSRHH	QKEMRRYIYNHQ-----NE-DGGW	GLHIEGNTMLCTALNYVALRLLGEE
LC-OSC	KVASEEE--ISE-----EAVETTLR	RAIRFYSTMQTDGDFWPGDY	-GPLFLLPGLVIGLVTGALNTVLSRHH	RQMCRYLYNHQ-----NE-DGGW	GLHIEGNTMLCTALSYVALRLLGEE
LJ-LAS	NIQKEED--ITE-----EVEDTLKR	RALRCYSTLQAGDGSWPGDY	-GAMFMLPGLVIGLVTGALNAALSPEH	QSEMCRYLYNHQ-----NE-DGGW	GLHIEGPTTFGTCLNYVALRLLGED
AT-LAS	KVKEGEERLINE-----EVLNVTLR	RSLRFYSILQSQDGFWPGDY	-GPLFLLPALVIGLYVTGALNTVLSRHH	QNEIRRYLYNHQ-----NK-DGGW	GLHVEGNTMFCTVLSVALRLLGEE
CP-CuS	KVKEGEE--VGK-----EAVKSTLR	RALGFYSAVQTRDGNWASDLG	-GPLFLLPGLVIALHVTGVLNSVLSKHH	RVEMCRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVALRLLGED
DD-CAS	HCPGEGCD--LNRA-----KTPQQAIR	KAFQYFSKVQTEDGHWAGDY	-GPMFLLPGLVITCYVTGYQ---LPEST	QREIRYLYNRQ-----NPVDGGW	GLHIEAHSDFGTTLQVVALRLLGV-
SA-CAS	KLVDSTQ--VTE-----QIILASLR	RALTQHSLSLQAHGDGHWPGDFF	-GIMFIMPILIFALYITGSLDAVLSSEH	RREICRYIYNHQ-----AY--QHW	STQVLGPSTMFSGSCLNYVTLRLGEEV
CR-CAS1	PVPDAPD--A-----ERVEEHLK	GAI SFYECLQEDGDFWPGDY	-GPMFLLPGLVITLYTCGALDQIFSPA	KKEALRYLYNHQ-----NE-DGGW	GLHIEGNTMFGTGLNYVMARLLGLA
PP-OSC	KIENLSD--LDE-----EAVTTTLR	RSMRFYSTLQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNTVLSPEH	QKEMRRYIYNHQ-----NK-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
AbN-CAS1	KLESAND--ITE-----EVIQTTLV	RAIRFYSTIQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNAVLSPEH	KKEICRYLYNHQ-----NE-DGGW	GLHIEGNTMFGTGLNYVTLRLGEE
AC-CAS	KVNDQSK--LPE-----QNVVDTLK	RAVLFYSTIQAHGDGHWAGDY	-GPMFLMPGLVIALYVTGSLNVVLSPEH	KKEMVRYLYNHQ-----NK-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
SB-OSC3	KLGEHED--VTE-----EAVSTTLK	RAISRFSLQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNTVLSPEH	QKEIRRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
SB-OSC4	KLDEHED--VTV-----EAVSTTLK	RAISRFSLQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNTVLSPEH	QKEIRRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
AS-CAS	KLEENED--VTE-----EAVSTSLK	RAISRFSLQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNAVLSPEH	QKEIRRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
OS-OSC2	KLEEEHA--VTG-----EAVLSSLK	RAIRYRSTLQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNTVLSSEH	QKEIRRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
OS-OSC1	KLGEDED--VTE-----EAVLSSLK	RAIRYRSTLQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNTVLSSEH	QKEIRRYLYNHQ-----NE-DGGW	GLHIEGNTMFGTCLNYVTLRLGEE
DZ-CAS1	KVRDDED--VTE-----EAVTTVTR	RAISRHSTLQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNTVLSPEH	QREICRYLYNHQ-----NK-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEEK
AN-OSC	KLAENEN--ITE-----EAAAITLR	RAMHRYSTLQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNTVLSSEH	QKEIRRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
CS-CAS1	KIEEHED--VTE-----ELVVTSLR	KAIRSVSTLQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNTVLSSEH	QKEIRRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
KC-CAS	TVKDDER--VTE-----QAVTVALR	RTLDYFSTIQAHGDGHWPGDY	-GPMFLMPGLVITL SVAGALNAILSREH	QGEICRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
RS-CAS	AVKDDER--VTE-----QAVTVTR	RALDYFSTIQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNAILSREH	QREICRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVTLRLIGEG
RC-CAS	KVKESEQ--VTE-----EKVKITLR	RALNYYSLQAHGDGHWPGDY	-GPMFLMPGLVIALYVTGALNAILSSEH	KREMCRYLYNHQ-----NR-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
BP-CAS2	KVNETED--VTE-----EMVTRMLR	RAISFHSSTLQAHGDGHWAGDY	-GPMFLMPGLVITL SITGALNTVLSSEH	KKEMCRYLYNHQ-----NK-DGGW	GLHIEGNTMFGTCLNYVTLRLGEE
AT-CAS1	KIEDTDD--VTE-----EMVETTLK	RGLDFYSTIQAHGDGHWPGDY	-GPMFLMPGLVITL SITGALNTVLSSEH	QKEMRRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
PG-CAS	KVNDGED--ISE-----DKVTVTLK	RAMSFYSTLQAHGDGHWPGDY	-GPMFLMPGLVITL SITGALNTVLSSEH	QKEICRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
CA-CAS	KVKDIGD--ISE-----DKVITLTK	RALSFYSTLQAHGDGHWAGDY	-GPMFLMPGLVITL SITEALNAILSKEH	KREICRYLYNHQ-----NR-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
BK-CAS	KVKSTED--ITE-----DVTVTTLK	RAVSFHASIQACDGHWPDY	-GPMFLMPGLVITL SITGALNAVLSKEH	KHEICRYLYNHQ-----NR-DGGW	GLHIEGNTMFGTCLNYVTLRLGEE
PS-CAS	KVKDVED--VTE-----EAVTTTLR	RGLNFYSTLQSHDGHWPDY	-GPMFLMPGLVITL SITGALNAVLTDEH	RKEMCRYLYNHQ-----NR-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
GG-CAS	RVKDIED--VTE-----ETVKTTLR	RAINFHSTLQSHDGHWPDY	-GPMFLMPGLVITL SITGALNAVLTDEH	RKEICRYLYNHQ-----NK-DGGW	GLHIEGNTMFSGSCLNYVALRLLGEE
LJ-CAS	KVKDVED--VTE-----EAVTTTLR	RAISFHSSTLQSHDGHWPDY	-GPMFLMPGLVITL SITGALNAVLTDEH	RKEMCRYLYNHQ-----NK-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
SL-CAS	KIINDNEV--ITE-----DINVATLR	RALSFYSTLQSHDGHWPAGDY	-GPMFLMPGMI IALSITGALNAVFSSEH	KREMCRYLYNHQ-----NK-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
OE-CAS	KVNDGED--ISE-----DKVTVTLK	RAMSFYSTLQAHGDGHWPGDY	-GPMFLMPGLVITL SITGAPNAVLSSEH	KKEMCRYLYNHQ-----NK-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
PT-CAS1	KVKDGHG--ITE-----EVVTTTLR	RAISFHSSTLQAHGDGHWPGDY	-GPLFLMPGLVITL SITGAPNAVLSSEH	QKEICRYLYNHQ-----NE-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
BP-CAS1	KIKDTEE--VRE-----EAVGMTLR	RAINFYSTIQAHGDGHWPGDY	-GPMFLIPGLVITL SITGALNAVLSKEH	QCEICRYLYNHQ-----NE-DGGW	GLHIEGNTMFGTALNYITLRLGEP
CP-CAS	KVKDKED--VTE-----EAVTRTLR	RAINFYSTIQAHGDGHWPGDY	-GPMFLIPGLVITL SITGALNAVLSSTEH	QREICRYLYNHQ-----NK-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE
LC-CAS1	KVKDKED--VTE-----EAVSRTLR	RAINFYSTIQAGDGHWPDY	-GPMFLIPGLVITL SITGALNAVLSSTEH	QREICRYLYNHQ-----NK-DGGW	GLHIEGNTMFSGSCLNYVTLRLGEE

Frame 4

<i>RS-NTS1</i>	IHWKKNCHQCAPEDLYYPHPFIQDLIWDCLYISMEPLLTRWPLNMIIRKKALETMKHIHYEDGSSRYITIGCVEK---VLCMLACWVEDPHGDYFKKHLARIPDYIWAEDGMKMQSF-GSQWDTG---
<i>BG-bas</i>	IHWKTRHQCAPEDLYYPHPFVQDLIWDCLYIFTEPLLTRWPLNEIRKKALETMKHIHYEDESSRYITIGCVEK---VLCMLACWVEDPHGDYFKKHLARIPDYIWAEDGMKMQSF-GSQWDTG---
<i>ET-bas</i>	IHWTKTRHLCAPEDVYYPHPLIQDLMWDSLYIFTEPLLTRWPFNKIRKKALETMKHIHYEDENSRYITIGCVEK---VLCMLACWAEDPHGVFFKKHLARIPDYMWVAEDGMKMQSF-GSQWDTG---
<i>BP-bas</i>	VHWKVRHLCAPEDLYYPHPLIQDLMWDSLYIFTEPLLTRWPFNKLVREKALQVTMKHIHYEDENSRYITIGCVEK---VLCMLACWVEDPHGDYFKKHLARIPDYIWAEDGKMQSF-GSQWDTG---
<i>PG-BAS1</i>	IHWKTRHVCAPEDLYYPHPLIQDLMWDSLYVLTTEPLLTRWPFNKLRKALQVTMKHIHYEDENSRYITIGCVEK---VLCMLVCWVEDPHGDYFRKHLARIPDYIWAEDGMKMQSF-GSQWDTG---
<i>BK-bas</i>	IHWPKVRHCCAPEDLYYPHPLVQDLIWGSLYICTEPLLTRWPFNKIRKALATTMEHIHYEDENSRYITIGCVEK---VLCMPACWAEDPDGDYFKKHLARIPDYIWAEDGMKMQSF-GSQWDTG---
<i>PG-BAS2</i>	IHWKVRHHCAPEDLYYPHPLIQDLMWDSLYIFTEPLLTRWPFNKLRKALQVTMKHIHYEDENSRYITIGCVEK---VLCMLACWVEDPHGDYFKHLARIPDYIWAEDGMKMQSF-GSQWDTG---
<i>VH-bas</i>	IWKVKVRHFCAPEDLYYPHPLIQDLMWDSLYLFTTEPLLTRWPFNGLIRKALQVTMDHIHYEDENSRYITIGCVEK---VLCMLACWVEDPHGVCKKHLARVPDYVWIAEDGLKMQSF-GSQWDCG---
<i>GP-bas</i>	IWKVKVRHLCAPEDLYYPHPLIQDLMWDSLYLFTTEPLLTRWPFNGLIRKALQVAMDHVHYEDENSRYITIGCVEK---VLCMLACWVEDPHGVCKKHLARVPDYIWAEDGLKMQSF-GSQWDCG---
<i>AA-bas</i>	IKWR SIRHLCAPEDLYYPHPLIQDLMWDSLYVFTTEPLNHWPFNKLRKALQVTMKHIHYEDENSRYITIGSVEK---ALCMLACWVEDPHGVCFKKHLARIPDYIWAEDGMKMQSF-GSQWDTG---
<i>AstS-bas1</i>	IHWK SIRHVAKEDLYYPHPLIQDLMWDSLYIFTEPLLRWPFNKLRKALRTTMHHIHYEDENSRYITIGSVEK---ALCMLSCWVEDPHGICFKKHLARVPDYLVWGEDGMKMQSF-GSQWDTG---
<i>PS-NTS</i>	VHWKTRHLCAPEDLYYPHPLIQDLIWDCLYIFTEPLLTRWPFNKLRKALQVTMKHIHYEDENSRYITIGCVEK---VLCILACWVEDPHGDAFKKHLARIPDYLVWSEDGMTLHSF-GSQWDTG---
<i>LJ-NTS</i>	VHWKTRHLCAPEDLYYPHPLIQDLIWDCLYIFTEPLLTRWPFNKLRKALQVTMKHIHYEDENSRYITIGAVEK---VLCMLACWVEDPHGIAFKRHLARIPDYIWLWSEDGLKMQSF-GSQWDTG---
<i>NT-BAS1</i>	IHWTKSRHLCAPEDLYYPHPLIQDLIWDCLYIFTEPLLTRWPFNKLRKALEVTMKHIHYEDENSRYITIGCVEK---VLCMLACWVEDPHGDYFKKHLARVQDYLVWSEDGMTMQSF-GSQWDTG---
<i>GG-bas</i>	VHWKTRHLCAPEDLYYPHPLIQDLIWDCLYIFTEPLLTRWPFNKLVREKALQVTMKHIHYEDENSRYITIGCVEK---VLCMLACWVEDPHGDAFKKHLARVPDYLVWSEDGMTMQSF-GSQWDTG---
<i>GN-bas</i>	VHWKTRHLCAPEDLYYPHPLIQDLIWDCLYIFTEPLLTRWPFNKLRKALEVTMKHIHYEDENSRYITIGAVEK---VLCMLACWVEDPHGDAFKKHLARVPDYLVWSEDGMTMQSF-GSQWDTG---
<i>LJ-bas</i>	VHWKTRHLCAPEDLYYPHPLIQDLIWDCLYIFTEPLLTRWPFNKLVREKALEVTMKHIHYEDENSRYITIGAVEK---VLCMLACWVEDPHGDAFKKHLARIPDYLVWSEDGMCMQSF-GSQWDTG---
<i>PT-bas</i>	IHWKTRHLCAPEDLYYPHPLIQDLIWDCLYIFTEPLLTRWPFNKLRKALQVTMKHIHYEDENSRYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARVPDYLVWSEDGMCMQSF-GSQWDTG---
<i>PS-bas</i>	IHWTKTRHLCAPEDLYYPHPLIQDLIWDCLYIFTEPLLTRWPFNKLVREKALQVTMKHIHYEDENSRYITIGAVEK---VLCMLACWVEDPHGDAFKKHLARVPDYLVWSEDGMTMQSF-GSQWDTG---
<i>NS-bas</i>	IRWFKSRNACAPEDLYYPHPLVQHLLWDTLNVFGETVLRWPFNPKLRKALQVTMKHIHYEDENSRYITIGAVEK---SLCMLACWVEDPDSDAFKKHLARVQDYLVWVAEDGMRMQSF-GSQWDTG---
<i>GS-bas</i>	IRWFKSRNACAPEDLYYPHPLVQHLLWDTLNVFGETVLRWPFNPKLRKALQVTMKHIHYEDENSRYITIGAVEK---VLCMLACWVEDPDSDAFKKHLARVQDYLVWVAEDGMRMQSF-GSQWDTG---
<i>SR-BOS</i>	IWKVSTRHLVVKEDLHYPHPLVQDLIWDCLYIFTEPLLTRWPFNPKLRKALQVTMKHIHYEDENSRYITIGAVEK---VLCMLACWVEDPDSDAFKKHLARVQDYLVWVAEDGMRMQSF-GSQWDTG---
<i>AT-LUP1</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGCVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>AT-LUP3</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>AT-LUP2</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>AT-bas</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>AT-CamS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>LC-INS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>RC-LUS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>ET-OSC</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>KC-NTS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>BG-LUS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>RS-NTS2</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>AT-NTS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>AT-ARS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>AT-THAS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>AT-PEN3</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>AT-BARS1</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>AT-MARS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>CA-bas</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>OE-NTS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>PG-DAS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>TO-LUS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>OE-LUS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>BP-LUS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>LJ-LUS</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---
<i>GG-LUS1</i>	IHWKSRRLYAKEDMYAHPLVQDLSDTLQNFVEPLLTRWPLNKLVRKALQVTMKHIHYEDENSHYITIGAVEK---VLCMLACWVEDPHGDYFKKHLARIPDYMWVAEDGMKMQSF-GCQLWDTG---

OS-OSC7 IDWIKARTQCAKEDMYPRSSKLDMFWSFLHKFIEPVLRLWRPGRKL-REKALATSMRNVHYEDECTRYICFGGVPK---ALNIIACWIEDPSEAFKCHIRVYDYLWIAEDGMKMQIYDGSQVWDAG---
 OS-OSC12 IDWKEARKLCAKEDAYNPHMWLQCECLSDCLYSFGEFPLTRWPISYM-RKRALYQIAEFLKYEDENSQYICIGAAQK---ALSMCCWIEHPNSDAFKRHLARVADFLVWGEDGMKVRVC-AGQLWDVA---
 OS-OSC3 IDWAQARSACAKEDLVCPRTLRQNAVWSWLYKWVPEVMSWSWAMNKL-RGRALDALMEHIHYEDEHTQYLCICSVNKK---ALNMVCCWAEDPNSDAFKRHLARVPDFLWLESDGMKAQVYDGCQSWETA---
 OS-OSC4 IDWAQARSACAKEDLVCPRTLRQNAVWSWLYKWVPEVMSWSWAMNKL-RGRALDRML-HIHYEDEHTQYLCICSVNKK---ALNMVCCWAEDPNSDAFKQHLARVPDFLWLESDGMKAQVYDGCQSWETA---
 OS-OSC11 IDWAQACHASCKEDLVCPRTLRQNAVWTSLSYKWVPEVLSGRPMNKL-RERALDRMEHIHYEDEHTQYLCICSVNKK---ALNMVCCWAEDPNSDAFKRHLARIPDFLWLESDGMKAQIYDGCQSWETA---
 SB-OSC2 IDWIKARNTCAKEDTRYQPSAIFSVIASCLNTFVEPFLHCWPLNKL-RKRALSHILEHIQYEDETTQYVGLSPVNN---ALNICRWVEHPNPDVLRHITRINDYLWIAEDGMKTKVCDGAQHWEIA---
 OS-OSC9 INWNHARSSCKDDIYPPSWLQIVAMASLHNFVEPLFNLWPMNVL-RQRALTNLMDHIHYEDENHYIIGLCPMNK---VLNMICCWIEHPNSYAFRQHLPRIHDFLWLESDGMKSKVYGGCQSWETA---
 OS-OSC10 INWNHARISCKKDDIYPPSWFQNIAMASLHNFVEPLFNLWPMNKL-RKRALTNLMDHIHYEDENHYIIGLCPMNK---VLNMICCWIEHPNSYAFRRHVPRIHDFLWLESDGMKSKVYGGCQSWETA---
 ZN-CAS IHNWIKACHSCAKEDLYEPSWLQSIAMAYLNKFIEPLSHLWPMNKL-RERAMSDLMEHIHYEDETSNYVGLCPINK---ALNMICCWIEKPNSEFRQHLPRIHDFLWLESDGMKSKVYVGGCHSWETA---
 OS-OSC8 IDWIKARDCCAKEDLYPCSWIQDQIVWTYLNKYVDPMFNVWPFNKL-REISLRNLMKHIHYEDEHTKYIGLCPINK---ALNMICCWIEDPNSDAFKRHLPRIDYFLWLESDGMKAQVYDGCQSWETA---
 AS-BAS IHNWDKARDYCAKEDLHYPRSRADQLISGCLTKIVEPILNHWPAANKL-RDRALTNLMEHIHYEDESTKYVGLCPINK---ALNMICCWVEHPNSPEFQQHLPRFHDYLWMAEDGMKAQVYDGCQSWETA---
 SB-OSC1 VDWEKARDTCAKVDLIYPRTMAQHLVWTCNLNKVIEPMLHCWVPVHKL-RDIALKNIMKHIHYEDETSKYICTCPINK---ALDMICCWAEHPNSDSFKKHLPRIDYDMLWIAEDGMKAQVYDGNPTWETS---
 TA-OSC IDWIKARDTCAKEDLRYPRSLLNQVITWTCNLNKVIEPVLHCWPIINKL-RDTALKNLMKHIHYEDESTKYIGVCPINK---ALDMICCWSEDPNSDALKLHLPRIDYDYLWLESDGMKAQVYDGCQSWETA---
 OS-OSC5 -----QYDGCQSWETS-----
 OS-OSC6 IHNWKNARHNCACAKEDLRYPRSFVQHVIVTGLNKKVVEPILSLWPFNTL-RHAALNHLKHIRYEDESTKYIGICPINK---ALDMICCWIDHPNSDAFKLHLPRIDYDYLWLESDGMKAQVYDGCQSWETA---
 CS-MTS IDWIKARTDCAKEDQYYPHPLIQDIWGSLSHNFVEPILMRWPGSKL-REKALSTVMQHIHYEDENRYICVGGVNN---ALNMLCCWIDDPNSEAFKLHLPRVYDYLWLESDGMKMKAYDGFQLWEAA---
 PG-LAS IHNWDLARINQCAKEDLYYPHPLIQDMLWLSLHKFVEPLIMQVPLSKI-RQRALTTVMQHIHYEDENRYICIGPVNN---VLNMVCCWVEDPNSMANILHLSRIDYDYLWVAEDGMKMKQYVNGSQLWDV---
 TO-OSC VHNWDLARKQCAKDDLYYPHPLIQDMLWLSLHKFVEPILMQVPAVAKL-RKEALKTVMQHIIMRMRTRTIFCIGPVNN---VLNMLCCWVEDPKTPIINLHLSRIDYDYLWVAEDGMKMKQYVNGSQLWDV---
 CP-OSC IDWHLRSRLCAKEDLYTPHSKIQDMLWDSIHKLGEPLLLKWPVLSKL-RQKALDFVIKHIHYEDENHYLCLGVPVSK---VNMVCCWVEDPNSDAFKRHLPRIDYDYLWLESDGMKMKQYVNGSQLWDV---
 LC-OSC VDWNHARINQCAKEDLYTPHSKIQDMLWDSIHKFVEPILMQVPAVAKL-RQKALDFVIKHIHYEDENHYLCLGVPVSK---VLNMVCCWVEDPNSDAFKRHLPRIDYDYLWLESDGMKMKQYVNGSQLWDV---
 LJ-LAS LDWNEARINLCAKEDLHSHRPGIQNHLWGLLHHVGEPLLTHKLFVSR-LRQKALHHVMEHIHNEDEASNYICIGPVNN---VLNMICCWLEDPNSQAFKYHISRIDYDYLWVAEDGMKMKQYVNGSQLWDV---
 AT-LAS IDWDTARINQCAKEDLYYPHPLIQDMLWLSLHKFVEPLLFRWPLNKL-RNHALQVVMQHIHYEDENSHYICIGPVNN---VLNMLCCWVESSNSEAFKSHLSRIDYDYLWVAEDGMKMKQYVNGSQLWDV---
 CP-CUS IDWIKSRINTCAKEDLYPHPKMQDILWGSYIHYVEPLFRWPGSKL-REKALQAAMKHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLQRVHDYLWVAEDGMRMQYVNGSQLWDTA---
 DD-CAS IHNWSEQRNHSKLDMYEHTSLLNVIINGSLHAY-EKVHSHKWL-----RDKAIDYTFDHIRYEDEQTKYIDIGPVNN---TVNMLCVWDRREGKSPAFYKHADRDKDYDYLWLSFDGMKMKQYVNGSQLWDTA---
 SA-CAS VDWWAARARVSPDVTYPTLWLVKVAHQAMYGY-----EWLAGKQLRARALDFALEQIRAEDEATHFCIGPVNN---VLNMVWHVFPNPEVEVRAHLERLDPYFFEAADGIMKHGSSSELDWTS---
 CR-CAS1 VDWNHARINQCAKEDLYYPHPLIQDMLWLSLHKFVEPLIMQVPLSKI-RQRALTTVMQHIHYEDENRYICIGPVNN---VLNMLACWLEEPGGKAFQKHLPPQGGFSW---KGCFAAGSSCGLQPTAPVL---
 PP-OSC VHNWNEARINLCAKEDLYYPHPIQDVLWGLTKVVEPILTRWPGSLL-RKKALARTMEHIHYEDENRYICIGPVNN---VMNMLSCWIEDPNSDAFKLHLARVYDYLWVAEDGMKMKQYVNGSQLWDTT---
 AbM-CAS1 IDWIKARINQCAKEDLYYPHPLIQDMLWGLTKVVEPALMHWPVGSML-RERALQVSMKHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKRHLARVYDYLWVAEDGMKMKQYVNGSQLWDTA---
 AC-CAS IDWIKARNECAKEDLYYPHPIQDVLWGLTKVVEPALMHWPVGSML-RERALQVSMKHIHYEDENRYICIGPVNN---VLNMLCCWIEDPNSDAFKCHLPRIDYDYLWVAEDGMKMKQYVNGSQLWDTA---
 SB-OSC3 IDWDKARINLCAKEDLYYPHPPVQDVLWATLHKFVEPVMHWPVGSKL-REKALETVMQHVHYEDENRYICIGPVNN---VLNMLACWIEDPNSDAFKLHIPRVDYLWLESDGMKMKQYVNGSQLWDTA---
 SB-OSC4 IDWDKVRNICAKEDLYRPHPPVQDVLWASLHKFVEPVMHWPVGSKL-REKALETVMQHVHYEDENRYICIGPVTK---VLNMLACWIEDPNSDAFKLHIPRVDYLWLESDGMKMKQYVNGSQLWDTA---
 AS-CAS IDWDKARINLCAKEDLYYPHPLIQDMLWGLTKVVEPVMHWPVGSKL-REKALTNLMDHIHYEDENRYICIGPVNN---VLNMLTCWIEDPNSDAFKLHIPRVDYLWVAEDGMKMKQYVNGSQLWDTA---
 OS-OSC2 VDWDKARINLCAKEDLYYPHPPVQDVLWATLHKFVEPAMLRWPGNKL-REKALDVTVMQHIHYEDENRYICIGPVNN---VLNMLACWIEDPNSDAFKLHIPRVDYLWIAEDGMKMKQYVNGSQLWDTA---
 OS-OSC1 IHNWDKARINQCAKEDLYRPHPLGQDILWATLYKVEPVLSHWPGSKL-REKALKNAMQHIHYEDENRYICSGAVQK---VLNMLSCWIEHPNSEAFRFRHIPRVDYLWVAEDGMKMKQYVNGSQLWDTA---
 DZ-CAS1 IHNWHLARINQCAKEDLYYPHPLIQDMLWASLHKFVEPILLRWPGSRL-RQRALHSTMQHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKHLPRIDYDYLWVAEDGMKMKQYVNGSQLWDTA---
 AN-OSC VDWNHARINQCAKEDLYYPHPLIQDMLWGLTKVVEPILMRWPGGKL-RGKALKTVMHVEHYEDENRYICLGHFSK---VLNMLCCWVEDQNSEAFKLHLPRINDYLWIAEDGMKIQYVNGSQLWDTA---
 CS-CAS1 LDWIKARNECAKEDLYYPHPPVQDVLWASLHKFVEPILMHWPVGSKL-REKAVNTAMQHVHYEDENRYICIGPVNN---VLNMLCCWIEDPNSDAFKLHLPRVMDYLWLESDGMKMRQYVNGSQLWDTA---
 KC-CAS IDWNEARINLCAKEDLYYPHPLVQDMLWAFLDKAIIEPILMHWPVGSKL-REKALQVSMHVEHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWIAEDGMKMKQYVNGSQLWDLA---
 RS-CAS IDWINDARINLCAKEDLYYPHPLVQDMLWAFLDKAIIEPILMRWPGKKL-RERAFQVSMHVEHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWIAEDGMKMKQYVNGSQLWDLA---
 RC-CAS IDWIKARINQCAKEDLYYPHPLVQDMLWATLHKFVEPILMHWPVGSKL-REKAIQTAIEHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWLESDGMKMKQYVNGSQLWDTA---
 BP-CAS2 IDWIKARINLCAKEDLYYPHPLVQDMLWASLHKFVEPILMRWPGKKL-REKALRTVLEHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRINDYLWIAEDGMKMKQYVNGSQLWDTA---
 AT-CAS1 VHNWNEARINLCAKEDLYYPHPLVQDMLWASLHKFVEPILMRWPGANL-REKAIQTAIEHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWLESDGMKMKQYVNGSQLWDTG---
 PG-CAS IDWIKARINLCAKEDLYYPHPLIQDMLWASLHKFVEPILMHWPVGSKL-REKSLRTVMEHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWLESDGMKMKQYVNGSQLWDTA---
 CA-CAS IDWIKARINLCAKEDLYYPHPLVQDMLWASLHKFVEPILMRWPGKKL-REKALRTVMEHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWLESDGMKMKQYVNGSQLWDTA---
 BK-CAS IDWIKARINQCAKEDLYYPHPLVQDMLWASLHKFVEPILMHWPVGSKL-REKSLRTVMEHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWVAEDGMKMKQYVNGSQLWDTA---
 PS-CAS IDWIKARINLCAKEDLYYPHPLVQDMLWATLHKFVEPILMHWPVGSKL-REKAIQTAIEHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWVAEDGMKMKQYVNGSQLWDTA---
 GG-CAS IDWIKARINLCAKEDLYYPHPLVQDMLWASLHKFVEPILMRWPGKKL-REMAIKTAIEHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWIAEDGMKMKQYVNGSQLWDTA---
 LJ-CAS IDWIKARINLCAKEDLYYPHPLVQDMLWASLHKFVEPILMQVPAVAKL-REKAIHNSVMEHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWIAEDGMKMKQYVNGSQLWDTA---
 SL-CAS VDWDKARNECAKEDLYYPHPLVQDMLWATLHKFVEPILMHWPVGSKL-REKALRTVMEHIHYEDENRYICIGPVNN---IHNMLCCWVEDPNSDAFKLHLPRIDYDYLWVAEDGMKMKQYVNGSQLWDTA---
 OE-CAS IDWIKARINQCAKEDLYYPHPLVQDMLWASLHKFVEPILMRWPGKKL-REKALRTVIEHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWLESDGMKMKQYVNGSQLWDTA---
 PT-CAS1 IHNWDHARSQCAKEDLYYPHPLVQDMLWGLTKVVEPILMRWPGKKL-RDKALRTVMEHIHYEDENRYICIGPVNN---ALNMLCCWVEDPNSDAFKLHIPRVDYLWIAEDGMKMKQYVNGSQLWDTA---
 BP-CAS1 IDWIKARINQCAKEDLYYPHPLVQDMLWASLYAYEPIFMHWPVGSKL-REKALDVTVMQHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHLPRIDYDYLWIAEDGMKMKQYVNGSQLWDTA---
 CP-CAS1 VDWNHARINQCAKEDLYYPHPLVQDMLWATLHHVVEPILMHWPVGSKL-REKALDVTVMQHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHIPRVDYLWIAEDGMKMKQYVNGSQLWDTA---
 LC-CAS1 VDWNHARNECAKEDLYYPHPLVQDMLWASLHHVVEPILMRWPAKRL-REKALQVSMQHIHYEDENRYICIGPVNN---VLNMLCCWVEDPNSDAFKLHIPRVDYLWIAEDGMKMKQYVNGSQLWDTA---

Frame 5



Frame 6

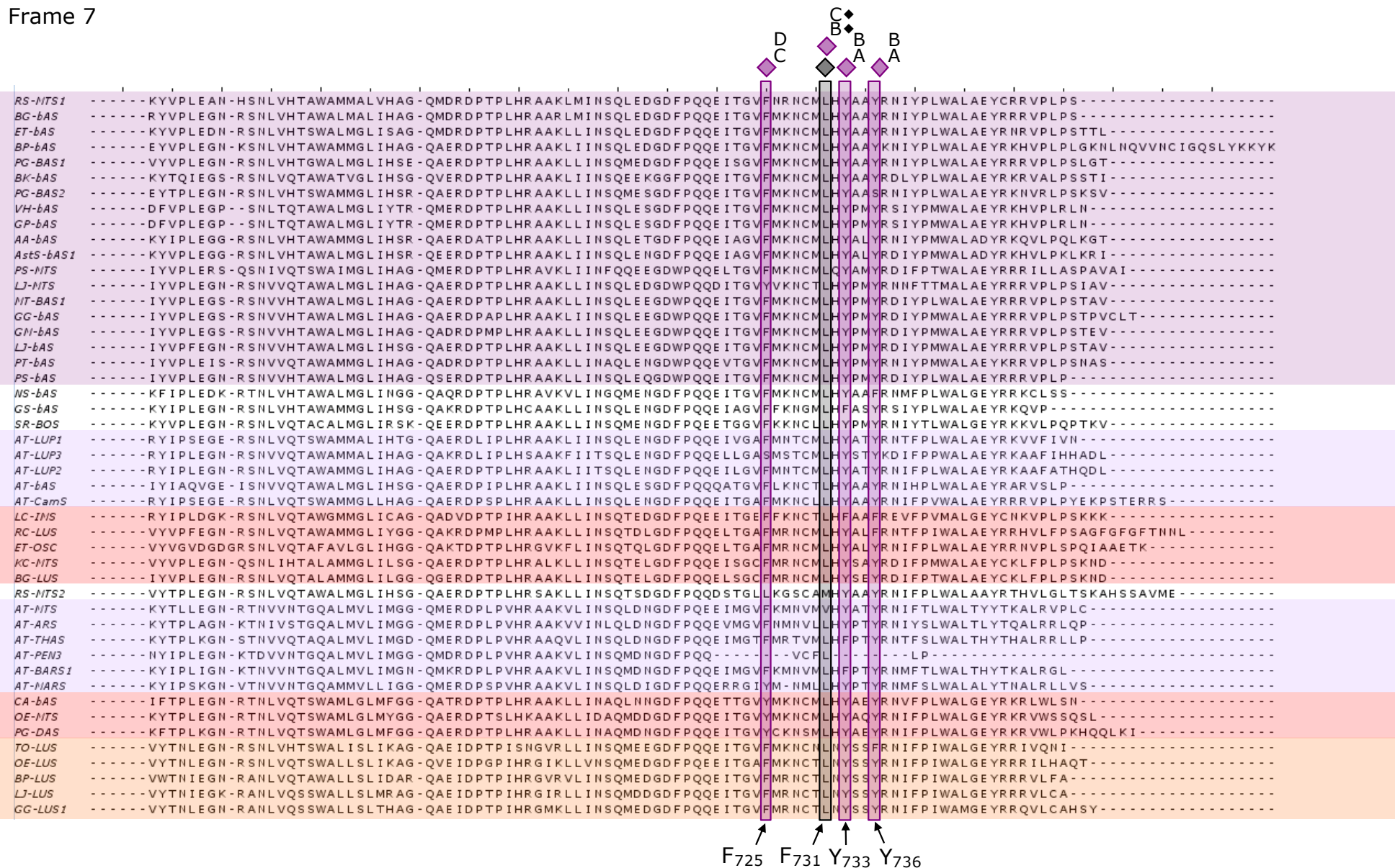
<i>RS-NTS1</i>	LLNPTEFFAD	IVIEHEYVER	TSSA IHALVLFKFLYPGH-RKKEIEDFIA	KSVRFLESIQTSDDGTW	YGNWGVCFYTGTFWALGGLAAAGKTYNSCLAMR	KAVDFLLRIQ-KDDGGW	GESYLSCPK	-----
<i>BG-bas</i>	LLNPTEFFAD	IVIEHEYVECT	TSSA IHALVLFKFLYPGH-RKKEIDHFI	NAVRYLESIQTSDDGGW	YGNWGVCFYTGTFWALGGLAAAGKTYNNCLAMR	KAVDFLLRIQ-RDNGGW	GESYLSCPK	-----
<i>ET-bas</i>	MLNPTEFFAD	IVIEHEYVECT	TASA IHALIMFKFLYPGH-RKKEIENFI	NAVKYLEDEVQTADGGW	YGNWGVCFYTGTFWAVGGAAAGKHYHNCAMR	KAVDFLLRTQ-KDQGGW	GESYLSCPHK	-----
<i>BP-bas</i>	LLNSTEFFAD	IVIEHEYIECT	TASAMQTLVLFKFLYPGH-RKKEIENFIK	NAAQFLQVIQMPDGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNNCLAVR	RAVDFLLRAQ-RDNGGW	GESYLSCPKK	-----
<i>PG-BAS1</i>	LLNPTEFFAD	IVIEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIDHFI	NAVRYLEDTQMPDGSW	YGNWGVCFYTGSWFALGGLAAAGKTYNCAAVR	KAVEFLLKSQ-MDDGGW	GESYLSCPKK	-----
<i>BK-bas</i>	LLNPTEFFQD	IVIEHEYVECT	TASA IQALVLFKFLYPGH-RKKEINAFI	NAAKYLEDIQMPDGSW	YGNWGVCFYTGSWFALGGLAAAGKTYNNCAAVR	KGVDFFLLKSQ-LDDGGW	GESYLSCPLE	-----
<i>PG-BAS2</i>	LLNPTEFFED	IVIEHEYVECT	TSSA IQAMVMFKFLYPGH-RKKEIEVSI	NAVQYLEDIQMPDGSW	YGNWGVCFYTGTFWAMGGLTAAGKTYNNCLAVR	KAVDFLLKRSQ-RSDGGW	GESYLSCPNK	-----
<i>VH-bas</i>	LLNPTEFFAD	IVIEHEYVECT	TGSA IQALVLFKFLYPGH-RKKEIENFIL	KASKYLEDTQYPNGSW	YGNWGVCFYTGTFWALGGLTAAGRFSNCAAIR	KGVEFLLKSQ-KEDGGW	GESYISCPKK	-----
<i>GP-bas</i>	LLNPTEFFAD	IVIEHEYVECT	TGSA IQALVLFKFLHYPGH-RKKEIENFIS	KASKYLEDTQYPNGSW	YGNWGVCFYTGTFWALGGLTAAGR TYSNCAAIR	KGVEFLLKSQ-KEDGGW	GESYISCPKK	-----
<i>AA-bas</i>	ILNPTEFFAD	IVIEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFLL	GSSGYLEKIQMEDGSW	YGNWGVCFYTGTFWALGGLSAVGKTYDNCAAIR	KAVKFLLETTQ-LEDGGW	GESYKSCPEK	-----
<i>AstS-bAS1</i>	YLNPTTEFFAD	IVIEHEYVECT	TSSA MQALVLFKFLYPGH-RKKEIENFLP	NACRYLENIQMPGGSW	YGNWGVCFYTGTFWALGGLTSGKTYENCPAIR	KGVKFLETTQ-LKDDGGW	GESYKSSPEK	-----
<i>PS-NTS</i>	LFNPIEFLEE	IVVEHEYVECT	TSSA IQALVLFKFLYPEH-RKKEIVENFI	NAVRFLEYKQTSDDGSW	YGNWGVCFYTGSWFALNGLVAAGKTYDNCAAIR	KGVEFLLTTQ-REDGGW	GESHLSSSKK	-----
<i>LJ-NTS</i>	LLNPTEFFED	IVIEHELVVECT	TGSA IGAALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAAIR	KGVKFLLTTQ-LEDGGW	GESYLSCPKK	-----
<i>NT-BAS1</i>	LLNPTEFFAD	IVVEHEYVECT	TGSA IQALVLFKFLYPGH-RKKEIENFIS	EAVRFIEDIQTADGGSW	YGNWGVCFYTGSWFALGGLAAAGKTYTNCAAIR	KAVKFLLETTQ-REDGGW	GESYLSPPKK	-----
<i>GG-bas</i>	LLNPTEFFAD	IVVEHEYVECT	TGSA IQALVLFKFLYPGH-RKKEIENFIA	NAVRFLEDTQTADGGSW	YGNWGVCFYTGSWFALGGLAAAGKTFANCAAIR	KAVKFLLETTQ-REDGGW	GESYLSPPKK	-----
<i>GM-bAS</i>	YSNPTTEFFAD	IVVEHEYVECT	TGSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRFLEDTQTADGGSW	YGNWGVCFYTGSWFALGGLAAAGKTYTNCAAIR	KAVKFLLETTQ-REDGGW	GESYLSPPKK	-----
<i>LJ-bas</i>	LLNPTEFFAD	IVVEHEYVECT	TGSA IGAALVLFKFLYPGH-RKKEIENFIS	EAVRFLEDTQTADGGSW	YGNWGVCFYTGSWFALGGLAAAGKTYANCAAIR	KAVKFLLETTQ-RDGGW	GESYLSPPKK	-----
<i>PT-bAS</i>	LLNPTEFFAD	IVVEHEYVECT	TGSA IQALVLFKFLYPGH-RKKEIDHFI	NAVRFLEDTQTADGGSW	YGNWGVCFYTGSWFALGGLAAAGKTFNCAAIR	KAVHFLLETTQ-KEDGGW	GESYLSPPKK	-----
<i>PS-bAS</i>	LLNPTEFFAD	IVVEHEYVECT	TGSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRFLEDTQTEDGSW	YGNWGVCFYTGSWFALGGLAAAGKTYTNCAAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>NS-bAS</i>	VINPTEFFQD	IVIEHEYVECT	TASG IAALVLFKFLYPGH-RKKEIESFIA	KAVHFLLEETQMPDGSW	YGNWGVCFYTGTFWALRGLAAGVHNCNHSPTVR	KACDFLLSTQ-LESGGW	GESYKSCPEK	-----
<i>GS-bAS</i>	LLNPTEFFAD	IVIEHEHVECT	TASA IDALVLFKFLYPGH-RKKEIESFIA	NASHYLEDVQMPDGSW	YGNWGVCFYTGTFWALRGLAAGKTYNNCLAVR	KAVHFLLETTQ-QKDDGGW	GESYRSCPEN	-----
<i>SR-BOS</i>	FLNPTEFFED	IVIEHELVVECT	TATGMQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAAIR	RAVHFLLETTQ-RDGGW	GESYRSPKK	-----
<i>AT-LUP1</i>	LLNPTEFFAD	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>AT-LUP2</i>	LLNPTEFFAD	IVVEHEYVECT	TGSA IGAALVLFKFLYPGH-RKKEIENFIS	EAVRFLEDTQTADGGSW	YGNWGVCFYTGSWFALGGLAAAGKTYANCAAIR	KAVKFLLETTQ-RDGGW	GESYLSPPKK	-----
<i>AT-LUP3</i>	LMNPTEFFAD	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>AT-LUP2</i>	LLNPTEFFAD	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>AT-bAS</i>	LLNPTEFFAD	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>AT-CamS</i>	LLNPTEFFAD	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>LC-INS</i>	WLNPTEFFED	IVIEHELVVECT	TSSA IGAALVLFKFLYPGH-RKKEIENFIS	EAVRFLEDTQTADGGSW	YGNWGVCFYTGSWFALGGLAAAGKTYANCAAIR	KAVKFLLETTQ-RDGGW	GESYLSPPKK	-----
<i>RC-LUS</i>	WLNPTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>ET-OSC</i>	WLNPTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>KC-NTS</i>	WLNPTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>BG-LUS</i>	WLNPTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>RS-NTS2</i>	VLNPLEFFED	IVVEHEYVECT	TASA INALVLMFKFLYPGH-RKKEIENFIS	KAVHYLIQTFPNDGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KAVDFLLKTQ-CQDGGW	GESYLSGTTK	-----
<i>AT-NTS</i>	WLSPTTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>AT-ARS</i>	WLSPTTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>AT-THAS</i>	WLSPTTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>AT-PEN3</i>	WLSPTTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>AT-BARS1</i>	WLSPTTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>AT-MARS</i>	WLSPTTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>CA-bas</i>	VLNPTEFFAD	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>OE-NTS</i>	MLNPTEFFAD	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>PG-DAS</i>	MLNPTEFFAD	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>TO-LUS</i>	KFNPTTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>OE-LUS</i>	KFNPTTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>BP-LUS</i>	KFNPTTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>LJ-LUS</i>	KFNPTTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----
<i>GG-LUS1</i>	KFNPTTEFFED	IVVEHEYVECT	TSSA IQALVLFKFLYPGH-RKKEIENFIS	NAVRYLEDTQADGGSW	YGNWGVCFYTGTFWALGGLAAAGKTYNCAPAIR	KGVKFLLTTQ-REDGGW	GESYLSPPKK	-----

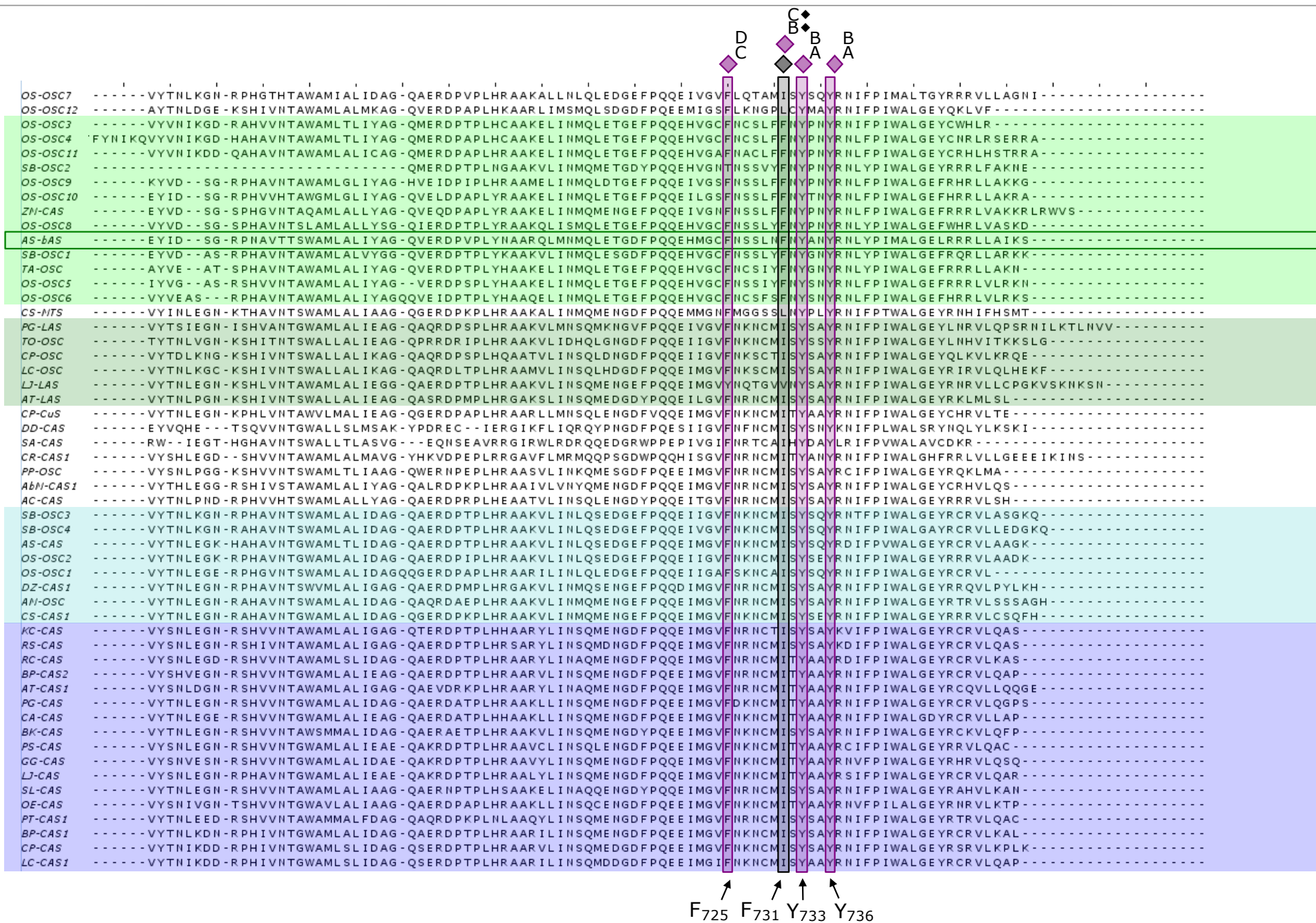
B
A

I₅₅₄ C₅₆₃

W₆₁₁

Frame 7





Appendix 6: Genomic DNA sequence of *Sad1*

Genomic DNA sequence of the *Sad1* gene showing the positions of the primers used for gDNA sequencing of the *sad1* mutants. Exons and introns are shown in uppercase and lowercase respectively. Locations of the forward and reverse primers identified in table 3.1 are shown in green and pink respectively. The start (ATG) and STOP codon (TGA) are shaded in grey.

```

ATGTGGAGGCTAACAATAGGTGAGGGCGGGTCCGTGGCTGAAGTCGAACAATGGCTTCCTTGGCCGCCAAGTGT
GGGAGTACGACGCCGATGCCGGCAGCCCGAAGAGCGTGCCGAGGTTGAGAGGGTGCCTGCGGAATTCACAAAGAA
CAGGTTCCAGAGGAAGGAGTCACAGGACCTTCTTCTACGCTTGCAGgtacatgCGTcttctttccctacttccat
atacaccagtagtatatggttgccactgCGgttagctctagcttttaggactgagaaaagggtctcagataatcca
tatctctcttttagatggagggtttgcttttatttattattacatattgttcaatccttgcgtgtatcatcaac
tgCagTACGCAAAGACAACCCTCTCCGGCGAATATTCCGACAGAAGCCAAGCTTGAAAAGAGTACAGAGGTAC
TCAGGAGACTATCTACGAATCATTGATGCGAGCTTTACATCAATATTCCTCTCTACAAGCAGACGATGGGCATTGG
26R ←
CCTGGTGATTCAGTGGGATTCTCTTCATTATGCCTATCATTgtaagtattttactatttttatgatacagcaa
tttgcaattaatatatgcatacagagtttcttattttcgtaaatactcaagacaatatagcatgtggaatcttata
atcttataatgaatatgtaccgtcttgtgtgCGcaatacgtatactatattattccgctatgcatatagtattac
ataccaatattgatagatgttcaaccaattatgaagagtttaactacaagatttaatatagtagtttctgttatt
ctagcagcaagttacctccattaggttccggaagttctactcttaccacctatatatatgtattattgcttatact
accttCGTctcaaagtttaagacttttttttaaagtcaatttatggaaagtttgaactaacttttataaaactatc
aagaactatgatattatatttgccatgtgaaaatagttttattatgatcaaagggtatggatttcgtaccgtaa
25R ←
atataaatattgttgcTaaaatcttggttgaactttacttagtttgacttttggaagatataagccttaaactt
taaaatagacgtagtaattcagatgcacttcaactgatatcccgacaaaagtacaaaatacatttatggaatgcaa
atattttgaaaacaacacatttggttagcttcaatatttcggaaaagaaaatagaggagtgatttaataaagt
tcttaaggttttcatgaaaacaaatctgttatggggactttatgcaaagagaacaagattggctcttagaaattt
ctttagatatgattaaatataaacagtgtttgcaactaaaaccacatttggtttgattgaaatatttgaaagaga
tagaaaatcttgaacatttatttttaggaatataggctttattactaccatcctatgtatcatcagatgtggctc
18F
atcacattgatcacaactctgaaaactaagaagtctccaacatttagacaatgatattggtttttcaaatttcagt
aacacttacaagaattccggttgattttattctccatccgagaactcatttctcctctcctaataatgatgcacata
tatgatgggatcttttctttatggttgCagATATTCCTTTATATGTTACTAGATCACCTTGACACCTTTTTATCTCC
GGAACATCGTCATGAGATATGTCGCTACATTTACAATCAACAGGcagggattaaacctaacacatatttccatatt
ttgtttctatatgtttgatgattttgtgacaaaataaaaacagtaacttaatgcaacatatattgagcagAATGAA
GATGGTGGTTGGGGAAAAATGGTTCTTGGCCCAAGTACCATGTTTGGATCGTGTATGAATTATGCAACCTTAATGA
TTCTTGGCGAGAAGCGAAATGGTGATCATAAGGATGCATTGGAAAAAGGGCGTTCTTGGATTTTATCTCATGGAAC
610R ←
TGCAACTGCAATACCACAGTGGGGAAAAATATGGTTGTCGgtatgttaaataacacaagatatcaatgctcatata
tgttctcttctgaactaacgttaaatcaacctactatttgataacatcatagATAATTGGCGTTTACGAATGGTCA

```

GGAAACAATCCTATTATACCTGAATTGTGGTTGGTTCCACATTTTCTCCGATTCACCCAGgtatttctatctagc
 ttgcatatataacaaaattgttgtagaacgcatgcttagaccatcattctgtggaattattctgtgcaatttggtg
 cttgtggaagcaatttaaccatataatcaacaaggaatattgaggcatggtacctgaaatagttttttgaaaaata
 catgccgaaaaggaaatcaatgtttcaattagcatgtttgacgtagattccacaagattctcttgtatatgttt
 tgatcttgagatacatgtatataatttatgtatctttcatattatctcaaaaaataacatgttactaccccctct
 atccataataagtgtcggtcacttagtacaactttatactagcttagtacaanaatggacgactcttattatggat
 tgcagggagtactaaatattatgaagttgaaccttatcattcacaagtaatttattgaaaataatccttcatatg
 tagGTCGTTTTTGGTGTTTTACCCGGTTGATATACATGTCAATGGCATACTCTATGGTAAGAAATTTGTTGGGCC
 TATTAGTCTACAATATTAGCT **CTGCGACAAGACCTCTATA** GTATAC **CTTACTGCAACATTAATTGGGAC** AAGGCG
 CGTGATTATTGTGCAAAAGgttagtttagttaatcaatcactatataatgtattcagtttgttagaataatattaatt
 tagcccagtgctactacataatattttcatggattcaagattaagaacatcacgtagaataatgaagtacatcattt
 cagtacttggatctcagaaaaatatagactaagaaagctagtgttcttcaaaaattttatgttgtttcagGAGG
 ACCTTCATTACCCACGCTCACGGGCACAAGATCTTATATCTGGTTGCCTAACGAAAATTTGTTGGAGCCAATTTTGAA
 TTGGTGGCCAGCAACAAGCTAAGAGATAGAGCTTTAACTAACCTCATGGAGCA **TATCCATTATGACGACGAATCA**
ACCAAATATGTGGGCATTTGCCCTATTAACAAGgtgaaattattttcaaatgatttgcaccttttacttttaataa
 tgacggatgttattccattctaattgttttaacatgtttattgtaattagGCATTGAACATGATTTGTTGTTGGGTA
 GAAACCCAAATTCGCCTGAATTCCAACAACATCTTCCACGATTCCATGACTATTTGTGGATGGCGGAGGAT **GGAA**
109R **TGAAGGCACAGgttggatagag** gctcttgtcagatattttgccaatttaactacgtgccaattcttcacaaccatt
 aacctttttcatgaatataatatttctcaacaaaatgtgagaatcttttgggttacaaggattttttattttcat
 ctatatctaggttgcatcaataagcatgtttgtgcatgtccgagttctcctgaaccaaactaaaatgcatattct
 ctttagctgcacatagtgatatgaaataaaattatggtaataatatttttacttttagttaattctaagacgaaa
 tagttgatatgcctatatcgtttcgaatatataaatcagaggtagttagaaaaattattggacttacatcaaatgc
 aaactgtgaatgtataagtaaatatgtatacaatcgcagGTATATGA **TGGATGTCATAGCTGGGA** ACTAGCGTTCAT
 AATTCATGCCTATTGTTCCACGGATCTTACTAGCGAGTTTATCCCAGCTCTAAAAAAGGCGCACGAGTTCATGAAG
 AACTCACAGgtttgttgttctccatattatattattgtcacaattctgaaaagatctaacattaattgtctaccct
 tgaagGTTCTTTTCAACCACCAAATCATGAAAGCTAT **TATCGCCACAGATCAAAAAGGCT** CATGGACCCTTCAAG
 TGTAGATAATGGTTGGTCTGTATCTGATGTACTGCGGAAGCTGTTAAGgttaacataagaaccatgtcttccaat
 tgtacatatataagtacatatgtgaatacatgacgggttaccctgtataagttgaaatgaactattcatgaatata
 ttgaatctacattaatattcattattttttcagGCATTGCTACTATTATCAAAGA **TATCCGCTGACCTTGTG** GCG
 ATCCAATAAAAACAAGACAGGTTGTATGATGCCATTGATTGCATCTTTCATGgtatgagaatctaaattgga
 tcaattaacaaacgtacattactaacaaggaaactatgcagaccattactaaagaaatgtgagcccacctagc
 tagataattttatctaaaagtattaaattatattttgcacaacatacaaaaagttaaatttgttgataatgcatat
 tattttctaaaaaaatgcaaaaataatttggagaaattttataaggtagtcacggtaatttaattcatttctat
 aatgcaaatagagctcactaatagcaggtctccttttctcgttttgaagAATACAGATGGAACATTTTCTACCTA

CGAATGCAAACGGACATTCGCTTGGTTAGAGgtttagtgatattcctttaagttttataacatggtacaattaaga
 tgaatatcatttttgtattgtatgacttgtccatgagaacaaggtattgggattgaataagaagtcaaaagaaaa
 ccaatacaacaatgatataattaattgtaattccttatggtcattttgcatctctcttcataccaagaaattttt
 tctcctgaacaataagtttgataaccctatcccctttaacaaaatatctctctacgagctagGTTCTCAACCTT
 TCTGAGAGTTTTCGGAACATTGTCGTGGACTATCCgtaagacaaaaaacctacttcataaattatctttacttc
 23R ←
 tatattcaaatattcattttcggaactgacttgatatacataataatggtcagATCTGTTGAATGCACATCATCT
 GTGGTTGATGCTCTCATATTATTTAAAGAGACGAATCCACGATATCGAAGAGCAGAGATAGATAAATGCATTGAAG
 AAGCTGTTGTATTTATGAGAACAGTCAAATAAAGGATGGTTCATGgtaagtgacatgatataaattatgcgttac
 aataacttttacttttgattaaatttgaaaatttattacttcttgtatctcatagGTATGGCTCATGGGGTATATG
 TTTTCGCATATGGATGCATGTTTGCAGTAAGGGCGTTGGTTGCTACAGGAAAAACCTACGACAATTGTGCTTCTATC
 AGGAAATCATGCAAATTTGTCTTATCAAAGCAACAAACAACAGGTGGATGGGGTGAAGACTATCTTTC TAGTGACA
 08F →
 ATGGGgtaatatacaaaactactttaccctataacattttactaatggtaaatcaatccatcatgattatcat
 agatttaagtcatacatgatataagtaaacatagaaattgatactagttgagagttttgttgttctataaataact
 agttgagagttcagtagttctataagccatgcatggaattacgaaacattataaactcaacgcaagatgatatagt
 tgacaaattttaaaatcatatgtcttgtaaaaaaaagatcctagttatattttgagcatgaatatcttcaaa
 tattgtgtatgtgagaaaggtgatgtttaattgcacaaggtacactaaataaaatggttaatttgtgtatgcc
 aaaaaagagagatatagaaagagctaaagtaaacttttaattgacacatcttgttcttaacatttattttttatga
 aagtcctcagtacattcacgatacctagcaaatatgaaaattcattgattagacagaagataaaattgccattccaca
 attattaaatcatatatgttaatttttgcctttttgcttatttttgtcatgataatggatgcataacaccatgttt
 tccaatggtctcatctactaatctcagatatcttaatacagcttggtctacaactgttacaccagttttgatga
 22R ←
 ccgtgtcca caaacattataaaactactactaatgacctaacacaaaaaatgacgacaaagacaaatccaat
 agaaggatcttttgtactgaacaaatgaagaaaattgtacatatatattgtgtaatatattaatttgttttccttta
 gtgtactccatccatgaaaatgatattaaatcaatatttgcacatcatgaggtcaaatctgttcttctagatgcgta
 agctaaagtcattatatgtatatatatattttcaagaacacataggtcattgtgttttctaatacagttttgtac
 agGAATATATTGATAGCGGTAGGCCAATGCTGTGACCACCTCA TGGGCAATGTTGGCTTTAATTTATGCTGGACA
 19F →
 Ggtttgtcaaatatttttcttgtttgtctagaatatgaattttttatataaaaggaaaagttctcactattcttg
 aatagtcgagttatctaacagaataatttatattttgttttttaataagGTTGAACGTGACCCAGTACCACTGTA
 TAATGCTGCAAGACAGCTAATGAATATGCAGCTAGAAACAGGTGACTTCCCCAACAGgtaaatatgtttccgtcct
 acatgttttcaacaaaaatgcaaagtaacttaagtttaattgaaactgatcttttgtaaatgaaactcaatgta
 gaccttaagggaaacaaccagtagaataaaaacttgtgaattgataactctggaaagtgtatgcattaatgttgg
 tgtgaatgtggtaaatgttggcattgcgtcataatttttgcatcgggtacttacaagtttaattaacactaatctc
 ttgtcagattcaatgatcattaaaaattaagatataacctccatctagcttcttacttacagtttcaccttgtaa
 tagGAACACATGGGTGCTTCAACTCCTCCTTGAACTTCAACTACGCCAACTACCGCAATCTATACCGATTATGG
 CTCTTGGGAACTTCGCCGTCGACTTCTTGCGATTAAGAGCTGA

Appendix 7: Original data from DArT analysis of *sad1* mutants

Table A: DArT binary data. The binary data for the 35 genetic markers found to be polymorphic between all of the 17 *sad1* mutants. Presence of absence of the marker was scored using “1” and “0” respectively and missing data is shown by a dash (-). Mutants that show high levels of variability between replicate samples are highlighted in red.

CloneID	CloneName	MarkerName	P	Q	Reproducibility	Call Rate	PIC	Discordance	S75-1	S75-2	A1-1	A1-2	B1-1	B1-2	109-1	109-2	610-1	610-2	1146-1	1146-2	1293-1	1293-2	110-1	110-2	225-1	225-2	589-1	
415701	802106272005_N_24	oPt-8320	88.438	87.571	100	96.078	0.48	0	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	
468690	802107109015_N_15	oPt-17991	83.785	82.964	100	94.118	0.5	0	0	1	1	1	1	1	-	1	0	0	1	1	1	1	1	1	0	0	0	
415431	802106272005_C_18	oPt-3946	82.194	81.388	100	88.235	0.495	0	0	1	1	1	1	1	1	1	0	0	1	1	-	1	1	1	0	0	0	
468690	802107109015_N_15	oPt-17991	81.915	81.112	100	90.196	0.499	0	0	1	1	1	1	1	1	1	0	0	1	1	-	1	1	1	0	0	0	
466873	802107109011_B_22	oPt-16174	81.404	80.605	100	100	0.109	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	
464253	802107109004_E_18	oPt-13554	79.327	78.549	100	86.275	0.498	0	0	1	1	1	1	1	-	1	0	0	1	1	-	1	-	1	0	0	0	
467153	802107109011_N_14	oPt-16454	78.949	78.175	100	92.157	0.219	0	-	0	1	1	1	1	0	0	-	-	0	0	0	0	0	0	0	0	0	
465310	802107109007_A_19	oPt-14611	77.514	76.671	100	90.196	0.065	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
455550	800907094014_K_3	oPt-11379	76.888	76.134	100	93.137	0.245	0	1	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	1	
451487	800907094004_A_20	oPt-2489	73.515	73.625	100	89.216	0.252	0	-	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1	-	1	
450979	800907094002_L_16	oPt-7377	73.474	72.754	100	91.176	0.292	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	
468422	802107109015_C_11	oPt-17723	72.307	71.598	100	93.137	0.285	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	
467089	802107109011_K_22	oPt-16390	72.225	71.551	100	87.255	0.266	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	
451823	800907094004_O_20	oPt-7303	70.929	71.036	100	92.157	0.225	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	1	1	0	
464996	802107109006_D_17	oPt-14297	67.635	67.164	100	88.235	0.252	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	-	0	
466143	802107109009_D_12	oPt-15444	66.243	65.594	100	93.137	0.175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	
397225	802106198003_M_4	oPt-9687	65.35	64.703	100	89.216	0.228	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	-	-	0	0	0	
794668	804909173005_A_1	oPt-794668	64.406	63.774	100	95.098	0.117	0	0	0	0	0	0	0	0	1	1	0	-	0	0	0	0	0	0	0	0	
393089	802106192010_P_20	oPt-2374	63.354	62.733	100	94.118	0.192	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	
793423	804909173001_M_4	oPt-793423	60.718	60.152	100	97.059	0.058	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
412054	802106271006_G_1	oPt-6715	64.346	63.786	96.97	82.353	0.215	0.022	-	0	0	0	0	0	0	0	1	-	0	0	0	0	-	-	0	0	0	
467187	802107109011_O_24	oPt-16488	66.992	66.673	96.97	80.392	0.337	0.022	1	-	1	1	1	-	1	1	1	1	-	1	1	-	-	0	0	0	0	
454314	800907094011_G_15	oPt-10143	71.103	70.986	96.97	83.333	0.27	0.022	1	0	-	0	0	-	1	1	0	-	-	0	0	0	0	0	0	0	0	
794493	804909173004_I_18	oPt-794493	70.015	69.682	96.97	87.255	0.299	0.019	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	-	0	0	
390960	802106192005_H_3	oPt-1586	69.409	68.729	96.97	84.314	0.483	0.021	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	-	1	1	0	
454390	800907094011_J_19	oPt-10219	67.033	68.164	96.97	87.255	0.18	0.021	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	
465377	802107109007_D_14	oPt-14678	73.056	72.597	96.97	85.294	0.452	0.022	0	0	0	0	1	1	0	0	-	1	0	0	-	1	0	0	0	0	0	
415431	802106272005_C_18	oPt-3946	83.566	83.02	96.97	91.176	0.5	0.018	0	1	1	1	1	1	1	1	0	0	1	1	-	1	1	1	0	0	0	
467633	802107109013_B_14	oPt-16934	74.16	73.588	96.97	84.314	0.456	0.022	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1	
467672	802107109013_D_5	oPt-16973	74.233	73.506	96.97	84.314	0.481	0.022	-	1	0	0	0	-	1	1	1	1	0	0	0	0	1	1	0	0	1	
467633	802107109013_B_14	oPt-16934	74.122	73.395	96.97	82.353	0.492	0.021	0	0	-	-	1	1	0	0	1	1	0	0	0	0	-	0	0	0	-	
453764	800907094009_P_17	oPt-9106	73.829	73.105	96.97	83.333	0.492	0.024	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	-	0	0	
464962	802107109006_C_7	oPt-14263	75.059	74.323	96.97	94.118	0.242	0.017	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
794975	804909173005_M_20	oPt-794975	74.558	73.827	96.97	86.275	0.488	0.021	0	58	0	-	0	-	1	0	0	1	1	0	0	1	1	0	0	0	0	
392831	802106192010_F_2	oPt-9413	83.599	82.779	96.97	95.098	0.488	0.018	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	
			74.029	73.393	100	92.0098	0.273		0	0.278	0.55	0.42	0.4	0.4	0.4	0.278	0.35	0.211	0.211	0.4444	0.4	0.4375	0.45	0.353	0.421	0.2	0.188	0.25

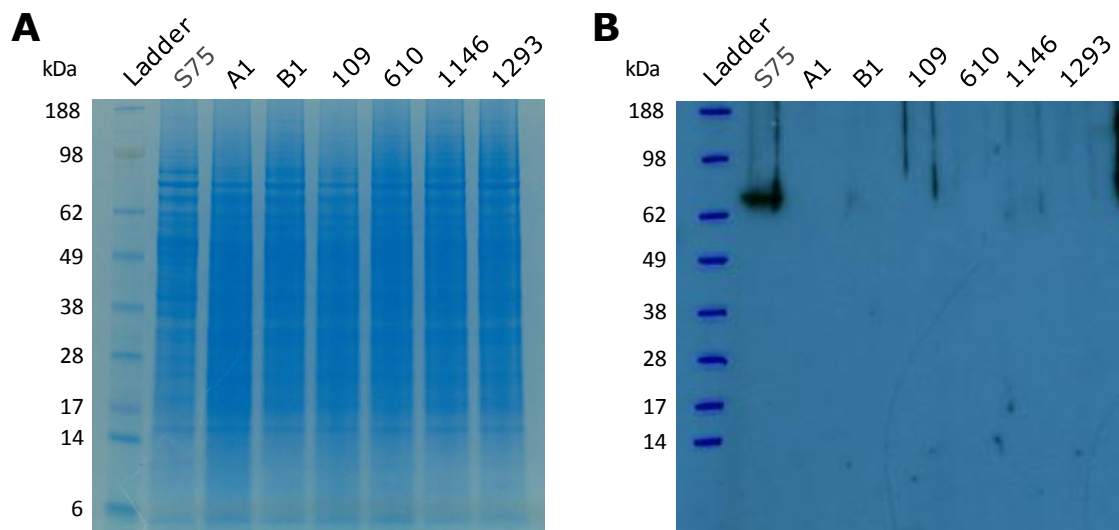
CloneID	CloneName	MarkerName	P	O	Reproducibility	Call Rate	PIC	Discordance	1001-1	1001-2	297-1	297-2	358-1	358-2	384-1	384-2	532-1	532-2	599-1	599-2	1023-1	1023-2	1217-1	1217-2
415701	802106272005_N_24	oPt-8320	88.438	87.571	100	96.078	0.48	0	1	1	0	0	1	1	0	0	0	0	0	0	1	-	0	0
468690	802107109015_N_15	oPt-17991	83.785	82.964	100	94.118	0.5	0	0	0	0	0	1	0	0	0	0	0	1	1	1	1	0	0
415431	802106272005_C_18	oPt-3946	82.194	81.388	100	88.235	0.495	0	0	0	0	0	1	0	-	0	0	0	1	1	1	1	0	0
468690	802107109015_N_15	oPt-17991	81.915	81.112	100	90.196	0.499	0	0	0	0	0	1	0	-	0	0	0	1	1	1	1	0	0
466873	802107109011_B_22	oPt-16174	81.404	80.605	100	100	0.109	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
464253	802107109004_E_18	oPt-13554	79.327	78.549	100	86.275	0.498	0	0	0	0	0	-	0	0	0	0	0	1	1	1	1	0	0
467153	802107109011_N_14	oPt-16454	78.949	78.175	100	92.157	0.219	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
465310	802107109007_A_19	oPt-14611	77.514	76.671	100	90.196	0.065	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
455550	800907094014_K_3	oPt-11379	76.888	76.134	100	93.137	0.245	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
451487	800907094004_A_20	oPt-2489	73.515	73.625	100	89.216	0.252	0	-	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1
450979	800907094002_L_16	oPt-7377	73.474	72.754	100	91.176	0.292	0	-	1	0	0	0	0	0	0	0	0	1	1	0	0	1	1
468422	802107109015_C_11	oPt-17723	72.307	71.598	100	93.137	0.285	0	-	1	0	0	0	0	0	0	0	0	1	1	0	0	1	1
467089	802107109011_K_22	oPt-16390	72.225	71.551	100	87.255	0.266	0	1	1	0	0	0	0	0	0	0	0	0	0	-	0	0	0
451823	800907094004_O_20	oPt-7303	70.929	71.036	100	92.157	0.225	0	-	1	0	0	0	0	1	0	0	0	0	0	0	0	-	0
464996	802107109006_D_17	oPt-14297	67.635	67.164	100	88.235	0.252	0	-	-	0	0	0	0	0	0	0	0	1	1	0	0	1	1
466143	802107109009_D_12	oPt-15444	66.243	65.594	100	93.137	0.175	0	1	1	0	0	0	0	0	0	0	0	0	0	0	-	0	0
397225	802106198003_M_4	oPt-9687	65.35	64.703	100	89.216	0.228	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	-	1
794668	804909173005_A_1	oPt-794668	64.406	63.774	100	95.098	0.117	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0
393089	802106192010_P_20	oPt-2374	63.354	62.733	100	94.118	0.192	0	1	-	1	1	1	1	1	1	1	1	1	1	-	1	0	0
793423	804909173001_M_4	oPt-793423	60.718	60.152	100	97.059	0.058	0	1	1	1	1	1	1	1	1	1	1	0	-	1	1	1	1
412054	802106271006_G_1	oPt-6715	64.346	63.786	96.97	82.353	0.215	0.022	0	0	-	0	0	0	-	1	0	0	0	0	0	0	1	1
467187	802107109011_O_24	oPt-16488	66.992	66.673	96.97	80.392	0.337	0.022	1	1	-	0	1	1	1	1	1	1	1	-	1	1	1	1
454314	800907094011_G_15	oPt-10143	71.103	70.986	96.97	83.333	0.27	0.022	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
794493	804909173004_I_18	oPt-794493	70.015	69.682	96.97	87.255	0.299	0.019	0	0	1	0	0	-	1	-	0	0	-	0	0	0	1	1
390960	802106192005_H_3	oPt-1586	69.409	68.729	96.97	84.314	0.483	0.021	-	-	1	-	0	0	0	0	1	0	1	1	-	1	1	-
454390	800907094011_J_19	oPt-10219	67.033	68.164	96.97	87.255	0.18	0.021	1	-	1	-	-	1	0	0	1	1	1	1	1	1	1	1
465377	802107109007_D_14	oPt-14678	73.056	72.597	96.97	85.294	0.452	0.022	1	1	-	0	0	0	0	1	1	1	0	0	-	1	1	1
415431	802106272005_C_18	oPt-3946	83.566	83.02	96.97	91.176	0.5	0.018	0	0	0	0	-	0	0	0	0	1	1	1	1	0	0	0
467633	802107109013_B_14	oPt-16934	74.16	73.588	96.97	84.314	0.456	0.022	1	1	0	0	0	-	1	-	1	1	0	0	-	0	1	1
467672	802107109013_D_5	oPt-16973	74.233	73.506	96.97	84.314	0.481	0.022	1	1	1	1	0	0	-	0	1	1	1	1	1	1	-	0
467633	802107109013_B_14	oPt-16934	74.122	73.395	96.97	82.353	0.492	0.021	1	1	0	0	0	0	1	1	1	1	0	0	-	-	-	1
453764	800907094009_P_17	oPt-9106	73.829	73.105	96.97	83.333	0.492	0.024	1	1	-	0	0	-	0	1	1	1	-	0	-	1	1	1
464962	802107109006_C_7	oPt-14263	75.059	74.323	96.97	94.118	0.242	0.017	-	0	0	0	0	0	0	0	0	0	1	1	0	0	-	1
794975	804909173005_M_20	oPt-794975	74.558	73.827	96.97	86.275	0.488	0.021	1	1	-	0	0	0	-	1	1	1	0	0	1	1	1	1
392831	802106192010_F_2	oPt-9413	83.599	82.779	96.97	95.098	0.488	0.018	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	0
			74.029	73.393	100	92.0098	0.273	0	0.4	0.5	0.1	0.2	0.421	0.2	0.222	0.1	0.15	0.15	0.5	0.526	0.3889	0.3889	0.2941	0.3

Table B: Humming matrix analysis of binary data. The binary data was analysed using a Humming matrix which gives a numerical score of the similarity between each of the mutant samples. The samples are scored at distances between 0 and 100, with identical samples being at a distance of 0 from each other. The two groups of mutants with identical point mutations are highlighted in red and blue.

Clone	A1-1	A1-2	1001-1	1001-2	1023-1	1023-2	109-1	109-2	110-1	110-2	1146-1	1146-2	1217-1	1217-2	1293-1	1293-2	225-1	225-2	297-1	297-2	358-1	358-2	384-1	384-2	532-1	532-2	589-1	589-2	599-1	599-2	610-1	610-2	B1-1	B1-2	S75-1	S75-2
A1-1	0	0	63	67	16	21	14	16	22	14	11	10	59	53	24	20	41	44	37	31	14	29	30	38	36	32	36	23	28	24	45	45	10	10	35	27
A1-2	0	0	59	63	18	22	16	18	21	13	10	9	59	55	25	21	38	40	35	28	16	29	28	38	35	32	32	21	28	25	44	45	9	13	35	24
1001-1	63	59	0	0	32	32	60	61	48	52	56	52	33	36	27	44	46	46	35	44	56	32	39	31	18	18	26	32	62	62	19	16	44	44	29	48
1001-2	67	63	0	0	39	39	68	68	58	57	59	57	28	36	36	50	41	39	46	48	59	39	39	38	26	26	33	40	62	62	28	25	50	50	37	50
1023-1	16	18	32	39	0	0	24	21	9	4	8	7	52	54	9	7	48	54	36	37	16	30	35	41	25	25	36	25	33	31	26	24	15	12	35	19
1023-2	21	22	32	39	0	0	24	22	15	11	18	16	48	52	15	10	43	50	30	37	24	35	41	37	22	25	44	33	30	30	27	24	16	14	44	26
109-1	14	16	60	68	24	24	0	0	15	18	14	16	56	61	37	32	37	36	22	27	20	28	36	40	34	38	36	26	30	27	47	45	29	30	32	32
109-2	16	18	61	68	21	22	0	0	14	16	16	18	59	65	38	32	42	42	27	30	22	34	40	46	37	40	38	28	27	24	49	47	29	28	37	29
110-1	22	21	48	58	9	15	15	14	0	4	12	11	60	61	29	24	44	42	27	30	19	31	40	44	38	41	38	28	26	21	33	37	32	33	32	17
110-2	14	13	52	57	4	11	18	16	4	0	4	3	62	61	20	16	41	44	32	27	11	25	37	48	39	39	33	20	24	20	35	38	27	28	30	13
1146-1	11	10	56	59	8	18	14	16	12	4	0	0	67	67	16	13	38	41	41	31	4	18	35	48	45	42	37	24	35	30	45	45	23	27	27	16
1146-2	10	9	52	57	7	16	16	18	11	3	0	0	64	61	14	12	41	43	38	31	3	16	35	44	41	38	36	25	31	28	41	41	21	25	24	15
1217-1	59	59	33	28	52	48	56	59	60	62	67	64	0	0	57	57	37	36	25	44	69	42	32	26	21	24	45	54	41	44	22	23	46	48	60	54
1217-2	53	55	36	36	54	52	61	65	61	61	67	61	0	0	54	55	44	43	35	49	65	42	35	25	27	27	46	54	41	44	28	29	42	45	59	52
1293-1	24	25	27	36	9	15	37	38	29	20	16	14	57	54	0	4	44	52	36	37	21	19	42	33	31	28	32	30	48	48	32	31	25	27	21	32
1293-2	20	21	44	50	7	10	32	32	24	16	13	12	57	55	4	0	47	53	40	41	19	29	45	38	35	32	46	36	41	39	38	36	18	19	41	27
225-1	41	38	46	41	48	43	37	42	44	41	38	41	37	44	44	47	0	0	21	16	40	28	21	31	30	33	28	30	52	48	39	40	47	50	36	53
225-2	44	40	46	39	54	50	36	42	42	44	41	43	36	43	52	53	0	0	23	17	43	32	27	35	36	39	30	32	48	41	48	50	53	57	37	47
297-1	37	35	35	46	36	30	22	27	27	32	41	38	25	35	36	40	21	23	0	11	41	21	22	14	13	17	21	21	41	41	17	18	40	39	27	47
297-2	31	28	44	48	37	37	27	30	30	27	31	31	44	49	37	41	16	17	11	0	26	20	21	29	21	21	13	11	42	36	32	33	41	40	24	38
358-1	14	16	56	59	16	24	20	22	19	11	4	3	69	65	21	19	40	43	41	26	0	14	33	43	44	41	39	32	40	37	43	41	29	31	21	23
358-2	29	29	32	39	30	35	28	34	31	25	18	16	42	42	19	29	28	32	21	20	14	0	22	28	25	22	23	26	50	50	20	17	36	38	7	36
384-1	30	28	39	39	35	41	36	40	40	37	35	35	32	35	42	45	21	27	22	21	33	22	0	14	20	17	21	23	50	50	21	25	30	32	33	48
384-2	38	38	31	38	41	37	40	46	44	48	48	44	26	25	33	38	31	35	14	29	43	28	14	0	15	12	31	36	59	61	19	20	28	30	32	59
532-1	36	35	18	26	25	22	34	37	38	39	45	41	21	27	31	35	30	36	13	21	44	25	20	15	0	3	21	24	46	46	12	9	24	22	37	50
532-2	32	32	18	26	25	25	38	40	41	39	42	38	24	27	28	32	33	39	17	21	41	22	17	12	3	0	18	21	49	49	15	13	21	19	33	47
589-1	36	32	26	33	36	44	36	38	38	33	37	36	45	46	32	46	28	30	21	13	39	23	21	31	21	18	0	0	50	44	31	32	36	36	17	42
589-2	23	21	32	40	25	33	26	28	28	20	24	25	54	54	30	36	30	32	21	11	32	26	23	36	24	21	0	0	41	33	37	39	29	27	17	32
599-1	28	28	62	62	33	30	30	27	26	24	35	31	41	41	48	41	52	48	41	42	40	50	50	59	46	49	50	41	0	0	55	57	38	40	61	16
599-2	24	25	62	62	31	30	27	24	21	20	30	28	44	44	48	39	48	41	41	36	37	50	50	61	46	49	44	33	0	0	58	60	38	39	57	12
610-1	45	44	19	28	26	27	47	49	33	35	45	41	22	28	32	38	39	48	17	32	43	20	21	19	12	15	31	37	55	58	0	0	34	33	31	50
610-2	45	45	16	25	24	24	45	47	37	38	45	41	23	29	31	36	40	50	18	33	41	17	25	20	9	13	32	39	57	60	0	0	32	30	31	52
B1-1	10	9	44	50	15	16	29	29	32	27	23	21	46	42	25	18	47	53	40	41	29	36	30	28	24	21	36	29	38	38	34	32	0	0	48	36
B1-2	10	13	44	50	12	14	30	28	33	28	27	25	48	45	27	19	50	57	39	40	31	38	32	30	22	19	36	27	40	39	33	30	0	0	50	38
S75-1	35	35	29	37	35	44	32	37	32	30	27	24	60	59	21	41	36	37	27	24	21	7	33	32	37	33	17	17	61	57	31	31	48	50	0	41
S75-2	27	24	48	50	19	26	32	29	17	13	16	15	54	52	32	27	53	47	47	38	23	36	48	59	50	47	42	32	16	12	50	52	36	38	41	0

Appendix 8: Western blot analysis of *sad1* PTC mutants

Western blot analysis of the *sad1* premature termination of translation mutants to look for the presence of AsbAS1 protein. Membrane was probed with the AsbAS1 antibody at 1:10,000 dilution (13083 serum) and chemiluminescent signal detected by exposure to X-ray film for 30 seconds. Lane 1, SeeBlue® Plus2 Pre-Stained Standard: myosin (188kDa), phosphorylase (98kDa), bovine serum albumin (62kDa), glutamic dehydrogenase (49kDa), alcohol dehydrogenase (38kDa), carbonic anhydrase (28kDa), myoglobin red (17kDa), lysozyme (14kDa), aprotinin (6kDa); Lane 2, Wild type S75 root tip protein; Lane 3-8, Root tip protein from mutants #A1, #B1, #109, #610, #1146, #1293. **A.** SDS-PAGE gel to show even total protein loading. **B.** Western blot at 30 seconds exposure.



References

1. Ingrouille, M. J., and Eddie, B. (2006) 6.7.1 Grasses, In *Plants: Evolution and Diversity* 1st ed., pp 295-298, Cambridge University Press, New York.
2. Ingrouille, M. J., and Eddie, B. (2006) 7.2 Plants for food, In *Plants: Evolution and Diversity* 1st ed., pp 295-298, Cambridge University Press, New York.
3. (1999) FAOSTAT [electronic resource], Food Agriculture Organization of the United Nations, Rome, Italy.
4. Chapman, G. P. (1996) *The Biology of Grasses*, 1st ed., CABI Publishing, Wallingford, UK.
5. Colhoun, J. (1971) Cereals, In *Diseases of Crop Plants*, p 181.185, Macmillan, UK.
6. Bateman, G., Gutteridge, R., Jenkyn, J., Spink, J., and McVittie, J. (2006) Take-all in winter wheat – management guidelines, Home Grown Cereals Authority, London, UK.
7. Jellis, G., Kelly, C., Clark, B., Bryson, R., and Tonguç, L. (2008) Take-all, In *The Encyclopaedia of Cereal Diseases*, p 81, Agriculture and Horticulture Development Board, Warwickshire, UK.
8. Hemming, B. C., and Houghton, J. M. (1993) Chapter 2: Influence of Biotechnology on Biocontrol of Take-all Disease of Wheat, In *Biotechnology in Plant Disease Control*, pp 15-38, Wiley-Liss Inc., Wilmington, DE, USA.
9. Morrissey, J. P., and Osbourn, A. E. (1999) Fungal resistance to plant antibiotics as a mechanism of pathogenesis, *Microbiol Mol Biol R* 63, 708-724.
10. Osbourn, A. E. (1999) Antimicrobial phytoprotectants and fungal pathogens: a commentary, *Fungal Genet Biol* 26, 163-168.
11. Dickinson, M. (2003) *Molecular Plant Pathology*, BIOS Scientific Publishers, New York, USA.
12. Bowyer, P. (1999) Chapter 10: Plant disease caused by fungi, In *Molecular Fungal Biology*, pp 294-321, Cambridge University Press, Cambridge, UK.
13. Osbourn, A. E. (1996) Preformed antimicrobial compounds and plant defense against fungal attack, *The Plant cell* 8, 1821-1831.
14. Hammond-Kosack, K. E., and Jones, J. D. (1996) Resistance gene-dependent plant defense responses, *The Plant cell* 8, 1773-1791.
15. Dangl, J. L., and Jones, J. D. (2001) Plant pathogens and integrated defence responses to infection, *Nature* 411, 826-833.
16. Martin, G. B. (1999) Functional analysis of plant disease resistance genes and their downstream effectors, *Curr Opin Plant Biol* 2, 273-279.
17. VanEtten, H. D., Mansfield, J. W., Bailey, J. A., and Farmer, E. E. (1994) Two Classes of Plant Antibiotics: Phytoalexins versus "Phytoanticipins", *The Plant cell* 6, 1191-1192.
18. Papadopoulou, K., Melton, R. E., Leggett, M., Daniels, M. J., and Osbourn, A. E. (1999) Compromised disease resistance in saponin-deficient plants, *Proceedings of the National Academy of Sciences of the United States of America* 96, 12923-12928.
19. Dixon, R. A. (2001) Natural products and plant disease resistance, *Nature* 411, 843-847.
20. Osbourn, A. E., Clarke, B. R., Lunness, P., Scott, P. R., and Daniels, M. J. (1994) An Oat Species Lacking Avenacin Is Susceptible to Infection by *Gaeumannomyces-Graminis* Var *Tritici*, *Physiological and Molecular Plant Pathology* 45, 457-467.
21. Behal, V. (2001) Nontraditional microbial bioactive metabolites, *Folia Microbiol* 46, 363-370.

22. Saleem, M., Nazir, M., Ali, M. S., Hussain, H., Lee, Y. S., Riaz, N., and Jabbar, A. (2010) Antimicrobial natural products: an update on future antibiotic drug candidates, *Nat Prod Rep* 27, 238-254.
23. Haralampidis, K., Trojanowska, M., and Osbourn, A. E. (2002) Biosynthesis of triterpenoid saponins in plants, *Advances in biochemical engineering/biotechnology* 75, 31-49.
24. Osbourn, A. E. (2003) Saponins in cereals, *Phytochemistry* 62, 1-4.
25. Hostettman, K. A., and Marston, A. (1995) *Saponins*, Cambridge University Press, Cambridge, UK.
26. Trojanowska, M. R., Osbourn, A. E., Daniels, M. J., and Threlfall, D. R. (2000) Biosynthesis of avenacins and phytosterols in roots of *Avena sativa* cv. Image, *Phytochemistry* 54, 153-164.
27. Price, K. R., Johnson, I. T., and Fenwick, G. R. (1987) The chemistry and biological significance of saponins in foods and feedingstuffs, *Crit Rev Food Sci Nutr* 26, 27-135.
28. Pasi, S., Aligiannis, N., Pratsinis, H., Skaltsounis, A. L., and Chinou, I. B. (2009) Biologically active triterpenoids from *Cephalaria ambrosioides*, *Planta Med* 75, 163-167.
29. Li, J., Huang, X., Du, X., Sun, W., and Zhang, Y. (2009) Study of chemical composition and antimicrobial activity of leaves and roots of *Scrophularia ningpoensis*, *Nat Prod Res* 23, 775-780.
30. Zhang, J. D., Xu, Z., Cao, Y. B., Chen, H. S., Yan, L., An, M. M., Gao, P. H., Wang, Y., Jia, X. M., and Jiang, Y. Y. (2006) Antifungal activities and action mechanisms of compounds from *Tribulus terrestris* L, *J Ethnopharmacol* 103, 76-84.
31. Bowyer, P., Clarke, B. R., Lunness, P., Daniels, M. J., and Osbourn, A. E. (1995) Host-Range of a Plant-Pathogenic Fungus Determined by a Saponin Detoxifying Enzyme, *Science* 267, 371-374.
32. Crombie, W. M. L., and Crombie, L. (1986) Distribution of Avenacins a-1, a-2, B-1 and B-2 in Oat Roots - Their Fungicidal Activity Towards Take-All Fungus, *Phytochemistry* 25, 2069-2073.
33. Crombie, L., Crombie, W. M. L., and Whiting, D. A. (1984) Isolation of avenacins A-1, A-2, B-1, and B-2 from oat roots: Structures of their 'aglycones', the avenestergenins, *Journal of the Chemical Society, Chemical Communications*, 244-246.
34. Armah, C. N., Mackie, A. R., Roy, C., Price, K., Osbourn, A. E., Bowyer, P., and Ladha, S. (1999) The membrane-permeabilizing effect of avenacin A-1 involves the reorganization of bilayer cholesterol, *Biophysical Journal* 76, 281-290.
35. Mylona, P., Owatworakit, A., Papadopoulou, K., Jenner, H., Qin, B., Findlay, K., Hill, L., Qi, X., Bakht, S., Melton, R., and Osbourn, A. (2008) Sad3 and sad4 are required for saponin biosynthesis and root development in oat, *The Plant cell* 20, 201-212.
36. Wink, M., R.A. Leigh, D. S., and Callow, J. A. (1997) Compartmentation of Secondary Metabolites and Xenobiotics in Plant Vacuoles, In *Advances in Botanical Research*, pp 141-169, Academic Press.
37. Luckner, M., Diettruch, B., and Lerbs, W. (1980) Cellular compartmentation and channeling of secondary metabolism in microorganisms and higher plants, *Progress in Phytochemistry* 6, 103-142.
38. Nelson, D. L., and Cox, M. M. (2000) Part III: Bioenergetics and Metabolism, In *Lehninger: Principles of Biochemistry*, pp 485-903, Worth Publishers, New York, USA.
39. Sacchettini, J. C., and Poulter, C. D. (1997) Creating isoprenoid diversity, *Science* 277, 1788-1789.
40. Dewick, P. M. (2002) The biosynthesis of C5-C25 terpenoid compounds, *Nat Prod Rep* 19, 181-222.

41. Liu, Y., Wang, H., Ye, H. C., and Li, G. F. (2005) Advances in the plant isoprenoid biosynthesis pathway and its metabolic engineering, *J Integr Plant Biol* 47, 769-782.
42. McGarvey, D. J., and Croteau, R. (1995) Terpenoid metabolism, *The Plant cell* 7, 1015-1026.
43. Dubey, V. S., Bhalla, R., and Luthra, R. (2003) An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants, *Journal of biosciences* 28, 637-646.
44. Rohmer, M. (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants, *Nat Prod Rep* 16, 565-574.
45. Kirby, J., and Keasling, J. D. (2009) Biosynthesis of Plant Isoprenoids: Perspectives for Microbial Engineering, *Annu Rev Plant Biol* 60, 335-355.
46. Laden, B. P., Tang, Y., and Porter, T. D. (2000) Cloning, heterologous expression, and enzymological characterization of human squalene monooxygenase, *Archives of biochemistry and biophysics* 374, 381-388.
47. Kolesnikova, M. D., Xiong, Q., Lodeiro, S., Hua, L., and Matsuda, S. P. (2006) Lanosterol biosynthesis in plants, *Archives of biochemistry and biophysics* 447, 87-95.
48. Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L., and Gruissem, W. (2004) GENEVESTIGATOR. Arabidopsis Microarray Database and Analysis Toolbox, *Plant Physiol.* 136, 2621-2632.
49. Haralampidis, K., Bryan, G., Qi, X., Papadopoulou, K., Bakht, S., Melton, R., and Osbourn, A. (2001) A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots, *Proceedings of the National Academy of Sciences of the United States of America* 98, 13431-13436.
50. Woldemichael, G. M., and Wink, M. (2002) Triterpene glycosides of *Lupinus angustifolius*, *Phytochemistry* 60, 323-327.
51. Xiang, T., Tezuka, Y., Wu, L.-J., Banskota, A. H., and Kadota, S. (2000) Saponins from *Lonicera bournei*, *Phytochemistry* 54, 795-799.
52. Herrera, J. B., Bartel, B., Wilson, W. K., and Matsuda, S. P. (1998) Cloning and characterization of the *Arabidopsis thaliana* lupeol synthase gene, *Phytochemistry* 49, 1905-1911.
53. Kawano, N., Ichinose, K., and Ebizuka, Y. (2002) Molecular cloning and functional expression of cDNAs encoding oxidosqualene cyclases from *Costus speciosus*, *Biological & pharmaceutical bulletin* 25, 477-482.
54. Morita, M., Shibuya, M., Kushiro, T., Masuda, K., and Ebizuka, Y. (2000) Molecular cloning and functional expression of triterpene synthases from pea (*Pisum sativum*) new alpha-amyrin-producing enzyme is a multifunctional triterpene synthase, *European journal of biochemistry / FEBS* 267, 3453-3460.
55. Basyuni, M., Oku, H., Inafuku, M., Baba, S., Iwasaki, H., Oshiro, K., Okabe, T., Shibuya, M., and Ebizuka, Y. (2006) Molecular cloning and functional expression of a multifunctional triterpene synthase cDNA from a mangrove species *Kandelia candel* (L.) Druce, *Phytochemistry* 67, 2517-2524.
56. Basyuni, M., Oku, H., Tsujimoto, E., Kinjo, K., Baba, S., and Takara, K. (2007) Triterpene synthases from the Okinawan mangrove tribe, Rhizophoraceae, *The FEBS journal* 274, 5028-5042.
57. Fazio, G. C., Xu, R., and Matsuda, S. P. (2004) Genome mining to identify new plant triterpenoids, *J Am Chem Soc* 126, 5678-5679.
58. Lodeiro, S., Xiong, Q., Wilson, W. K., Kolesnikova, M. D., Onak, C. S., and Matsuda, S. P. (2007) An oxidosqualene cyclase makes numerous products by diverse mechanisms: a challenge to prevailing concepts of triterpene biosynthesis, *J Am Chem Soc* 129, 11213-11222.

59. Shibuya, M., Xiang, T., Katsube, Y., Otsuka, M., Zhang, H., and Ebizuka, Y. (2007) Origin of structural diversity in natural triterpenes: direct synthesis of seco-triterpene skeletons by oxidosqualene cyclase, *J Am Chem Soc* 129, 1450-1455.
60. Xiang, T., Shibuya, M., Katsube, Y., Tsutsumi, T., Otsuka, M., Zhang, H., Masuda, K., and Ebizuka, Y. (2006) A new triterpene synthase from *Arabidopsis thaliana* produces a tricyclic triterpene with two hydroxyl groups, *Organic letters* 8, 2835-2838.
61. Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., Eagles, J., Werck-Reichhart, D., Schaller, H., Lesot, A., Melton, R., and Osbourn, A. (2006) A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defense, *Proceedings of the National Academy of Sciences of the United States of America* 103, 18848-18853.
62. Mugford, S. T., Qi, X., Bakht, S., Hill, L., Wegel, E., Hughes, R. K., Papadopoulou, K., Melton, R., Philo, M., Sainsbury, F., Lomonosoff, G. P., Roy, A. D., Goss, R. J., and Osbourn, A. (2009) A Serine Carboxypeptidase-Like Acyltransferase Is Required for Synthesis of Antimicrobial Compounds and Disease Resistance in Oats, *The Plant cell*.
63. Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R., and Osbourn, A. (2004) A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants, *Proceedings of the National Academy of Sciences of the United States of America* 101, 8233-8238.
64. Osbourn, A. E., and Field, B. (2009) Operons, *Cell Mol Life Sci*.
65. Field, B., and Osbourn, A. E. (2008) Metabolic diversification--independent assembly of operon-like gene clusters in different plants, *Science* 320, 543-547.
66. Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grun, S., Winklmeier, A., Eisenreich, W., Bacher, A., Meeley, R. B., Briggs, S. P., Simcox, K., and Gierl, A. (1997) Analysis of a chemical plant defense mechanism in grasses, *Science* 277, 696-699.
67. Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K. W., Toyomasu, T., Sassa, T., Hasegawa, M., Kodama, O., Shibuya, N., Koga, J., Nojiri, H., and Yamane, H. (2007) Identification of a biosynthetic gene cluster in rice for momilactones, *J Biol Chem* 282, 34013-34018.
68. Stuiver, M. H., and Custers, J. H. H. V. (2001) Engineering disease resistance in plants, *Nature* 411, 865-868.
69. Jolidon, S., Polak, A. M., Guerry, P., and Hartman, P. G. (1990) Inhibitors of 2,3-oxidosqualene lanosterol-cyclase as potential antifungal agents, *Biochem Soc Trans* 18, 47-48.
70. Rommens, C. M., Salmeron, J. M., Oldroyd, G. E., and Staskawicz, B. J. (1995) Intergeneric transfer and functional expression of the tomato disease resistance gene *Pto*, *The Plant cell* 7, 1537-1544.
71. Hammond-Kosack, K. E., Tang, S., Harrison, K., and Jones, J. D. (1998) The tomato Cf-9 disease resistance gene functions in tobacco and potato to confer responsiveness to the fungal avirulence gene product *avr 9*, *The Plant cell* 10, 1251-1266.
72. Vaeck, M., Reynaerts, A., Hofte, H., Jansens, S., De Beuckeleer, M., Dean, C., Zabeau, M., Montagu, M. V., and Leemans, J. (1987) Transgenic plants protected from insect attack, *Nature* 328, 33-37.
73. Whitham, S., McCormick, S., and Baker, B. (1996) The N gene of tobacco confers resistance to tobacco mosaic virus in transgenic tomato, *Proceedings of the National Academy of Sciences of the United States of America* 93, 8776-8781.
74. Naoumkina, M. A., Modolo, L. V., Huhman, D. V., Urbanczyk-Wochniak, E., Tang, Y., Sumner, L. W., and Dixon, R. A. (2010) Genomic and

- coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*, *The Plant cell* **22**, 850-866.
75. Suzuki, H., Achnine, L., Xu, R., Matsuda, S. P., and Dixon, R. A. (2002) A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*, *Plant J* **32**, 1033-1048.
 76. Shibuya, M., Hoshino, M., Katsube, Y., Hayashi, H., Kushiro, T., and Ebizuka, Y. (2006) Identification of beta-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay, *The FEBS journal* **273**, 948-959.
 77. Meesapyodsuk, D., Balsevich, J., Reed, D. W., and Covello, P. S. (2007) Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding beta-amyrin synthase and a triterpene carboxylic acid glucosyltransferase, *Plant Physiol* **143**, 959-969.
 78. Frey, M., Kliem, R., Saedler, H., and Gierl, A. (1995) Expression of a cytochrome P450 gene family in maize, *Mol Gen Genet* **246**, 100-109.
 79. Gierl, A., and Frey, M. (2001) Evolution of benzoxazinone biosynthesis and indole production in maize, *Planta* **213**, 493-498.
 80. Wilderman, P. R., Xu, M., Jin, Y., Coates, R. M., and Peters, R. J. (2004) Identification of Syn-Pimara-7,15-Diene Synthase Reveals Functional Clustering of Terpene Synthases Involved in Rice Phytoalexin/Allelochemical Biosynthesis, *Plant Physiol.* **135**, 2098-2105.
 81. Swaminathan, S., Morrone, D., Wang, Q., Fulton, D. B., and Peters, R. J. (2009) CYP76M7 Is an ent-Cassadiene C11{alpha}-Hydroxylase Defining a Second Multifunctional Diterpenoid Biosynthetic Gene Cluster in Rice, *The Plant cell* **21**, 3315-3325.
 82. Walton, J. D. (2000) Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: An hypothesis, *Fungal Genetics and Biology* **30**, 167-171.
 83. Trojanowska, M. R., Osbourn, A. E., Daniels, M. J., and Threlfall, D. R. (2001) Investigation of avenacin-deficient mutants of *Avena strigosa*, *Phytochemistry* **56**, 121-129.
 84. Gollub, E. G., Liu, K. P., Dayan, J., Adlersberg, M., and Sprinson, D. B. (1977) Yeast mutants deficient in heme biosynthesis and a heme mutant additionally blocked in cyclization of 2,3-oxidosqualene, *J Biol Chem* **252**, 2846-2854.
 85. Kremer, C. A., Lee, M., and Holland, J. B. (2001) A restriction fragment length polymorphism based linkage map of a diploid *Avena* recombinant inbred line population, *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada* **44**, 192-204.
 86. Giesler, K. (2010) Personal communication, John Innes Centre, Norwich, UK.
 87. Owatworakit, A. (2009) The role of glucosylation in plant defence in oats, In *Department of Metabolic Biology*, John Innes Centre, Norwich.
 88. Osbourn, A. E., Clarke, B. R., Dow, J. M., and Daniels, M. J. (1991) Partial characterization of avenacinase from *Gaeumannomyces graminis* var. *avenae*, *Physiological and Molecular Plant Pathology* **38**, 301-312.
 89. Vogt, T., and Jones, P. (2000) Glycosyltransferases in plant natural product synthesis: characterization of a supergene family, *Trends in plant science* **5**, 380-386.
 90. Townsend, B., Jenner, H., and Osbourn, A. (2006) Saponin glycosylation in cereals, *Phytochemistry Reviews* **5**, 109-114.
 91. Kelly, R., Miller, S. M., Lai, M. H., and Kirsch, D. R. (1990) Cloning and characterization of the 2,3-oxidosqualene cyclase-coding gene of *Candida albicans*, *Gene* **87**, 177-183.
 92. Karst, F., and Lacroute, F. (1977) Ertosterol biosynthesis in *Saccharomyces cerevisiae*: mutants deficient in the early steps of the pathway, *Mol Gen Genet* **154**, 269-277.

93. Corey, E. J., Matsuda, S. P., Baker, C. H., Ting, A. Y., and Cheng, H. (1996) Molecular cloning of a *Schizosaccharomyces pombe* cDNA encoding lanosterol synthase and investigation of conserved tryptophan residues, *Biochem Biophys Res Commun* 219, 327-331.
94. Corey, E. J., Matsuda, S. P., and Bartel, B. (1994) Molecular cloning, characterization, and overexpression of ERG7, the *Saccharomyces cerevisiae* gene encoding lanosterol synthase, *Proceedings of the National Academy of Sciences of the United States of America* 91, 2211-2215.
95. Milla, P., Viola, F., Oliaro Bosso, S., Rocco, F., Cattell, L., Joubert, B. M., LeClair, R. J., Matsuda, S. P., and Balliano, G. (2002) Subcellular localization of oxidosqualene cyclases from *Arabidopsis thaliana*, *Trypanosoma cruzi*, and *Pneumocystis carinii* expressed in yeast, *Lipids* 37, 1171-1176.
96. Abe, I., Naito, K., Takagi, Y., and Noguchi, H. (2001) Molecular cloning, expression, and site-directed mutations of oxidosqualene cyclase from *Cephalosporium caerulens*, *Biochimica et biophysica acta* 1522, 67-73.
97. Sung, C. K., Shibuya, M., Sankawa, U., and Ebizuka, Y. (1995) Molecular cloning of cDNA encoding human lanosterol synthase, *Biological & pharmaceutical bulletin* 18, 1459-1461.
98. Baker, C. H., Matsuda, S. P., Liu, D. R., and Corey, E. J. (1995) Molecular cloning of the human gene encoding lanosterol synthase from a liver cDNA library, *Biochem Biophys Res Commun* 213, 154-160.
99. Abe, I., and Prestwich, G. D. (1995) Molecular cloning, characterization, and functional expression of rat oxidosqualene cyclase cDNA, *Proceedings of the National Academy of Sciences of the United States of America* 92, 9274-9278.
100. Nakano, C., Motegi, A., Sato, T., Onodera, M., and Hoshino, T. (2007) Sterol biosynthesis by a prokaryote: first in vitro identification of the genes encoding squalene epoxidase and lanosterol synthase from *Methylococcus capsulatus*, *Bioscience, biotechnology, and biochemistry* 71, 2543-2550.
101. Corey, E. J., Matsuda, S. P., and Bartel, B. (1993) Isolation of an *Arabidopsis thaliana* gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen, *Proceedings of the National Academy of Sciences of the United States of America* 90, 11628-11632.
102. Shibuya, M., Adachi, S., and Ebizuka, Y. (2004) Cucurbitadienol synthase, the first committed enzyme for cucurbitacin biosynthesis, is a distinct enzyme from cycloartenol synthase for phytosterol biosynthesis, *Tetrahedron* 60, 6995-7003.
103. Hayashi, H., Hiraoka, N., Ikeshiro, Y., Yazaki, K., Tanaka, S., Kushiro, T., Shibuya, M., and Ebizuka, Y. (1999) Molecular Cloning of a cDNA Encoding Cycloartenol Synthase from *Luffa cylindrica* (Accession No. AB033334) (PGR 99-183), *Plant Physiol.* 121, 1383-1385.
104. Zhang, H., Shibuya, M., Yokota, S., and Ebizuka, Y. (2003) Oxidosqualene cyclases from cell suspension cultures of *Betula platyphylla* var. *japonica*: molecular evolution of oxidosqualene cyclases in higher plants, *Biological & pharmaceutical bulletin* 26, 642-650.
105. Shibuya, M., Zhang, H., Endo, A., Shishikura, K., Kushiro, T., and Ebizuka, Y. (1999) Two branches of the lupeol synthase gene in the molecular evolution of plant oxidosqualene cyclases, *European journal of biochemistry / FEBS* 266, 302-307.
106. Basyuni, M., Oku, H., Tsujimoto, E., and Baba, S. (2007) Cloning and functional expression of cycloartenol synthases from mangrove species *Rhizophora stylosa* Griff. and *Kandelia candel* (L.) Druce, *Bioscience, biotechnology, and biochemistry* 71, 1788-1792.
107. Guhling, O., Hobl, B., Yeats, T., and Jetter, R. (2006) Cloning and characterization of a lupeol synthase involved in the synthesis of

- epicuticular wax crystals on stem and hypocotyl surfaces of *Ricinus communis*, *Archives of biochemistry and biophysics* 448, 60-72.
108. Hayashi, H., Hiraoka, N., Ikeshiro, Y., Kushiro, T., Morita, M., Shibuya, M., and Ebizuka, Y. (2000) Molecular cloning and characterization of a cDNA for *Glycyrrhiza glabra* cycloartenol synthase, *Biological & pharmaceutical bulletin* 23, 231-234.
 109. Sawai, S., Shindo, T., Sato, S., Kaneko, T., Tabata, S., Ayabe, S., and Aoki, T. (2006) Functional and structural analysis of genes encoding oxidosqualene cyclases of *Lotus japonicus*, *Plant Science* 170, 247-257.
 110. Morita, M., Shibuya, M., Lee, M. S., Sankawa, U., and Ebizuka, Y. (1997) Molecular cloning of pea cDNA encoding cycloartenol synthase and its functional expression in yeast, *Biological & pharmaceutical bulletin* 20, 770-775.
 111. Kushiro, T., Shibuya, M., and Ebizuka, Y. (1998) Beta-amyrin synthase--cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants, *European journal of biochemistry / FEBS* 256, 238-244.
 112. You, S., Kawano, N., Ichinose, K., Yao, X.-S., and Ebizuka, Y. (1999) Molecular Cloning and Sequencing of an *Allium macrostemon* cDNA Probably Encoding Oxidosqualene Cyclase, *Plant Biotechnology* 16, 311-314.
 113. Shinozaki, J., Shibuya, M., Masuda, K., and Ebizuka, Y. (2008) Squalene cyclase and oxidosqualene cyclase from a fern, *FEBS letters* 582, 310-318.
 114. Godzina, S. M., Lovato, M. A., Meyer, M. M., Foster, K. A., Wilson, W. K., Gu, W., de Hostos, E. L., and Matsuda, S. P. (2000) Cloning and characterization of the *Dictyostelium discoideum* cycloartenol synthase cDNA, *Lipids* 35, 249-255.
 115. Bode, H. B., Zeggel, B., Silakowski, B., Wenzel, S. C., Reichenbach, H., and Muller, R. (2003) Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2,3(S)-oxidosqualene cyclase from the myxobacterium *Stigmatella aurantiaca*, *Molecular microbiology* 47, 471-481.
 116. Phillips, D. R., Rasbery, J. M., Bartel, B., and Matsuda, S. P. T. (2006) Biosynthetic diversity in plant triterpene cyclization, *Curr Opin Plant Biol* 9, 305-314.
 117. Gibbons, G. F., Goad, L. J., Goodwin, T. W., and Nes, W. R. (1971) Concerning Role of Lanosterol and Cycloartenol in Steroid Biosynthesis, *J Biol Chem* 246, 3967-&.
 118. Giner, J. L., and Djerassi, C. (1995) A Reinvestigation of the Biosynthesis of Lanosterol in *Euphorbia-Lathyris*, *Phytochemistry* 39, 333-335.
 119. Giner, J. L., Berkowitz, J. D., and Andersson, T. (2000) Nonpolar components of the latex of *Euphorbia peplus*, *J Nat Prod* 63, 267-269.
 120. Suzuki, M., Xiang, T., Ohyama, K., Seki, H., Saito, K., Muranaka, T., Hayashi, H., Katsube, Y., Kushiro, T., Shibuya, M., and Ebizuka, Y. (2006) Lanosterol synthase in dicotyledonous plants, *Plant Cell Physiol* 47, 565-571.
 121. Sawai, S., Akashi, T., Sakurai, N., Suzuki, H., Shibata, D., Ayabe, S. I., and Aoki, T. (2006) Plant lanosterol synthase: Divergence of the sterol and triterpene biosynthetic pathways in eukaryotes, *Plant Cell Physiol* 47, 673-677.
 122. Kalinin, V., Silchenko, A., Avilov, S., Stonik, V., and Smirnov, A. (2005) Sea Cucumbers Triterpene Glycosides, the Recent Progress in Structural Elucidation and Chemotaxonomy, *Phytochemistry Reviews* 4, 221-236.
 123. Kajikawa, M., Yamato, K. T., Fukuzawa, H., Sakai, Y., Uchida, H., and Ohyama, K. (2005) Cloning and characterization of a cDNA encoding beta-amyrin synthase from petroleum plant *Euphorbia tirucalli* L, *Phytochemistry* 66, 1759-1766.

124. Kirby, J., Romanini, D. W., Paradise, E. M., and Keasling, J. D. (2008) Engineering triterpene production in *Saccharomyces cerevisiae*-beta-amyrin synthase from *Artemisia annua*, *The FEBS journal* **275**, 1852-1859.
125. Hayashi, H., Huang, P., Kirakosyan, A., Inoue, K., Hiraoka, N., Ikeshiro, Y., Kushiro, T., Shibuya, M., and Ebizuka, Y. (2001) Cloning and characterization of a cDNA encoding beta-amyrin synthase involved in glycyrrhizin and soyasaponin biosyntheses in licorice, *Biological & pharmaceutical bulletin* **24**, 912-916.
126. Scholz, M., Lipinski, M., Leupold, M., Luftmann, H., Harig, L., Ofir, R., Fischer, R., Prufer, D., and Muller, K. J. (2009) Methyl jasmonate induced accumulation of kalopanaxsaponin I in *Nigella sativa*, *Phytochemistry* **70**, 517-522.
127. Liu, Y., Cai, Y., Zhao, Z., Wang, J., Li, J., Xin, W., Xia, G., and Xiang, F. (2009) Cloning and Functional Analysis of a beta-amyrin synthase gene associated with oleanolic acid biosynthesis in *Gentiana straminea* MAXIM, *Biological & pharmaceutical bulletin* **32**, 818-824.
128. Shibuya, M., Katsube, Y., Otsuka, M., Zhang, H., Tansakul, P., Xiang, T., and Ebizuka, Y. (2009) Identification of a product specific beta-amyrin synthase from *Arabidopsis thaliana*, *Plant Physiol Biochem* **47**, 26-30.
129. Hayashi, H., Huang, P., Takada, S., Obinata, M., Inoue, K., Shibuya, M., and Ebizuka, Y. (2004) Differential expression of three oxidosqualene cyclase mRNAs in *Glycyrrhiza glabra*, *Biological & pharmaceutical bulletin* **27**, 1086-1092.
130. Segura, M. J., Meyer, M. M., and Matsuda, S. P. (2000) *Arabidopsis thaliana* LUP1 converts oxidosqualene to multiple triterpene alcohols and a triterpene diol, *Organic letters* **2**, 2257-2259.
131. Saimaru, H., Orihara, Y., Tansakul, P., Kang, Y. H., Shibuya, M., and Ebizuka, Y. (2007) Production of triterpene acids by cell suspension cultures of *Olea europaea*, *Chemical & pharmaceutical bulletin* **55**, 784-788.
132. Husselstein-Muller, T., Schaller, H., and Benveniste, P. (2001) Molecular cloning and expression in yeast of 2,3-oxidosqualene-triterpenoid cyclases from *Arabidopsis thaliana*, *Plant Mol Biol* **45**, 75-92.
133. Shibuya, M., Sagara, A., Saitoh, A., Kushiro, T., and Ebizuka, Y. (2008) Biosynthesis of baccharis oxide, a triterpene with a 3,10-oxide bridge in the A-ring, *Organic letters* **10**, 5071-5074.
134. Iturbe-Ormaetxe, I., Haralampidis, K., Papadopoulou, K., and Osbourn, A. E. (2003) Molecular cloning and characterization of triterpene synthases from *Medicago truncatula* and *Lotus japonicus*, *Plant Mol Biol* **51**, 731-743.
135. Godio, R. P., and Martin, J. F. (2009) Modified oxidosqualene cyclases in the formation of bioactive secondary metabolites: biosynthesis of the antitumor clavarinic acid, *Fungal Genet Biol* **46**, 232-242.
136. Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M., and Yuan, Y. (1998) Predicting function: from genes to genomes and back, *J Mol Biol* **283**, 707-725.
137. Rouzé, P., Pavy, N., and Rombauts, S. (1999) Genome annotation: which tools do we have for it?, *Curr Opin Plant Biol* **2**, 90-95.
138. Schilling, C. H., Schuster, S., Palsson, B. O., and Heinrich, R. (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era, *Biotechnol Prog* **15**, 296-303.
139. Ebizuka, Y., Katsube, Y., Tsutsumi, T., Kushiro, T., and Shibuya, M. (2003) Functional genomics approach to the study of triterpene biosynthesis, *Pure and Applied Chemistry* **75**, 369-374.
140. Xiong, Q., Wilson, W. K., and Matsuda, S. P. (2006) An *Arabidopsis* oxidosqualene cyclase catalyzes iridal skeleton formation by Grob fragmentation, *Angewandte Chemie (International ed)* **45**, 1285-1288.

141. Kolesnikova, M. D., Wilson, W. K., Lynch, D. A., Obermeyer, A. C., and Matsuda, S. P. (2007) Arabidopsis camelliol C synthase evolved from enzymes that make pentacycles, *Organic letters* 9, 5223-5226.
142. Buntel, C. J., and Griffin, J. H. (1994) Evolution of Sterol and Triterpene Cyclases, *Acs Sym Ser* 562, 44-54.
143. Wendt, K. U., Poralla, K., and Schulz, G. E. (1997) Structure and function of a squalene cyclase, *Science* 277, 1811-1815.
144. Thoma, R., Schulz-Gasch, T., D'Arcy, B., Benz, J., Aebi, J., Dehmlow, H., Hennig, M., Stihle, M., and Ruf, A. (2004) Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase, *Nature* 432, 118-122.
145. Abe, I., and Prestwich, G. D. (1994) Active-Site Mapping of Affinity-Labeled Rat Oxidosqualene Cyclase, *J Biol Chem* 269, 802-804.
146. Abe, I., and Prestwich, G. D. (1995) Identification of the active site of vertebrate oxidosqualene cyclase, *Lipids* 30, 231-234.
147. Feil, C., Sussmuth, R., Jung, G., and Poralla, K. (1996) Site-directed mutagenesis of putative active-site residues in squalene-hopene cyclase, *European journal of biochemistry / FEBS* 242, 51-55.
148. Poralla, K. (1994) The Possible Role of a Repetitive Amino-Acid Motif in Evolution of Triterpenoid Cyclases, *Bioorganic & Medicinal Chemistry Letters* 4, 285-290.
149. Loll, P. J. (2003) Membrane protein structural biology: the high throughput challenge, *Journal of structural biology* 142, 144-153.
150. Wendt, K. U., Feil, C., Lenhart, A., Poralla, K., and Schulz, G. E. (1997) Crystallization and preliminary X-ray crystallographic analysis of squalene-hopene cyclase from Alicyclobacillus acidocaldarius, *Protein Sci* 6, 722-724.
151. Ruf, A., Muller, F., D'Arcy, B., Stihle, M., Kuszniir, E., Handschin, C., Morand, O. H., and Thoma, R. (2004) The monotopic membrane protein human oxidosqualene cyclase is active as monomer, *Biochem Biophys Res Commun* 315, 247-254.
152. Eschenmoser, A., Ruzicka, L., Jeger, O., and Arigoni, D. (1955) Zur Kenntnis Der Triterpene .190. Eine Stereochemische Interpretation Der Biogenetischen Isoprenregel Bei Den Triterpenen, *Helv Chim Acta* 38, 1890-1904.
153. van Tamelen, E. E. (1982) Bioorganic Characterization and Mechanism of the 2,3-Oxidosqualene --> Lanosterol Conversion, *J Am Chem Soc* 104, 6480-6481.
154. Boar, R. B., Couchman, L. A., Jaques, A. J., and Perkins, M. J. (1984) Isolation from Pistacia Resins of a Bicyclic Triterpenoid Representing an Apparent Trapped Intermediate of Squalene 2,3-Epoxyde Cyclization, *J Am Chem Soc* 106, 2476-2477.
155. Corey, E. J., Russey, W. E., and de Montellano, P. R. O. (1966) 2,3-Oxidosqualene, an Intermediate in the Biological Synthesis of Sterols from Squalene, *J Am Chem Soc* 88, 4750-4751.
156. Wendt, K. U. (2005) Enzyme mechanisms for triterpene cyclization: new pieces of the puzzle, *Angewandte Chemie (International ed)* 44, 3966-3971.
157. Woodward, R. B., and Bloch, K. (1953) The Cyclization of Squalene in Cholesterol Synthesis, *J Am Chem Soc* 75, 2023-2024.
158. Stork, G., and Burgstahler, A. W. (1955) The Stereochemistry of Polyene Cyclization, *J Am Chem Soc* 77, 5068-5077.
159. Maudgal, R. K., Tchen, T. T., and Bloch, K. (1958) 1,2-Methyl Shifts in the Cyclization of Squalene to Lanosterol, *J Am Chem Soc* 80, 2589-2590.
160. Cornforth, J. W., Cornforth, R. H., Donniger, C., Popják, G., Shimizu, Y., Ichii, S., Forchielli, E., and Caspi, E. (1965) The Migration and Elimination

- of Hydrogen during Biosynthesis of Cholesterol from Squalene, *J Am Chem Soc* 87, 3224-3228.
161. Corey, E. J., and Russey, W. E. (1966) Metabolic Fate of 10,11-Dihydrosqualene in Sterol-Producing Rat Liver Homogenate, *J Am Chem Soc* 88, 4751-4752.
 162. van Tamelen, E. E., Willett, J. D., Clayton, R. B., and Lord, K. E. (1966) Enzymic Conversion of Squalene 2,3-Oxide to Lanosterol and Cholesterol, *J Am Chem Soc* 88, 4752-4754.
 163. Abe, I., Rohmer, M., and Prestwich, G. D. (1993) Enzymatic Cyclization of Squalene and Oxidosqualene to Sterols and Triterpenes, *Chemical Reviews* 93, 2189-2206.
 164. Wendt, K. U., Schulz, G. E., Corey, E. J., and Liu, D. R. (2000) Enzyme Mechanisms for Polycyclic Triterpene Formation, *Angewandte Chemie (International ed)* 39, 2812-2833.
 165. Schulz-Gasch, T., and Stahl, M. (2003) Mechanistic insights into oxidosqualene cyclizations through homology modeling, *Journal of Computational Chemistry* 24, 741-753.
 166. Lenhart, A., Reinert, D. J., Aebi, J. D., Dehmlow, H., Morand, O. H., and Schulz, G. E. (2003) Binding Structures and Potencies of Oxidosqualene Cyclase Inhibitors with the Homologous Squalene-Hopene Cyclase, *Journal of Medicinal Chemistry* 46, 2083-2092.
 167. Wu, T. K., Liu, Y. T., Chang, C. H., Yu, M. T., and Wang, H. J. (2006) Site-saturated mutagenesis of histidine 234 of *Saccharomyces cerevisiae* oxidosqualene-lanosterol cyclase demonstrates dual functions in cyclization and rearrangement reactions, *J Am Chem Soc* 128, 6414-6419.
 168. Wu, T. K., Liu, Y. T., Chiu, F. H., and Chang, C. H. (2006) Phenylalanine 445 within oxidosqualene-lanosterol cyclase from *Saccharomyces cerevisiae* influences C-ring cyclization and deprotonation reactions, *Organic letters* 8, 4691-4694.
 169. Wu, T. K., Yu, M. T., Liu, Y. T., Chang, C. H., Wang, H. J., and Diau, E. W. G. (2006) Tryptophan 232 within oxidosqualene-lanosterol cyclase from *Saccharomyces cerevisiae* influences rearrangement and deprotonation but not cyclization reactions, *Organic letters* 8, 1319-1322.
 170. Wu, T. K., Wang, T. T., Chang, C. H., Liu, Y. T., and Shie, W. S. (2008) Importance of *Saccharomyces cerevisiae* oxidosqualene-lanosterol cyclase tyrosine 707 residue for chair-boat bicyclic ring formation and deprotonation reactions, *Organic letters* 10, 4959-4962.
 171. Wu, T. K., Wen, H. Y., Chang, C. H., and Liu, Y. T. (2008) Protein plasticity: a single amino acid substitution in the *Saccharomyces cerevisiae* oxidosqualene-lanosterol cyclase generates protosta-13(17),24-dien-3beta-ol, a rearrangement product, *Organic letters* 10, 2529-2532.
 172. Lodeiro, S., Wilson, W. K., Shan, H., and Matsuda, S. P. (2006) A putative precursor of isomalabaricane triterpenoids from lanosterol synthase mutants, *Organic letters* 8, 439-442.
 173. Wu, T. K., Chang, C. H., Liu, Y. T., and Wang, T. T. (2008) *Saccharomyces cerevisiae* oxidosqualene-lanosterol cyclase: a chemistry-biology interdisciplinary study of the protein's structure-function-reaction mechanism relationships, *Chem Rec* 8, 302-325.
 174. Herrera, J. B. R., Wilson, W. K., and Matsuda, S. P. T. (2000) A tyrosine-to-threonine mutation converts cycloartenol synthase to an oxidosqualene cyclase that forms lanosterol as its major product, *J Am Chem Soc* 122, 6765-6766.
 175. Wu, T.-K., and Griffin, J. H. (2002) Conversion of a Plant Oxidosqualene-Cycloartenol Synthase to an Oxidosqualene-Lanosterol Cyclase by Random Mutagenesis, *Biochemistry* 41, 8238-8244.

176. Kushiro, T., Shibuya, M., Masuda, K., and Ebizuka, Y. (2000) Mutational studies on triterpene syntheses: Engineering lupeol synthase into beta-amyrin synthase, *J Am Chem Soc* 122, 6816-6824.
177. Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O., and Shaw, D. E. (2009) Long-timescale molecular dynamics simulations of protein structure and function, *Current Opinion in Structural Biology* 19, 120-127.
178. Khalili-Araghi, F., Gumbart, J., Wen, P.-C., Sotomayor, M., Tajkhorshid, E., and Schulten, K. (2009) Molecular dynamics simulations of membrane channels and transporters, *Current Opinion in Structural Biology* 19, 128-137.
179. Shan, Y., Seeliger, M. A., Eastwood, M. P., Frank, F., Xu, H., Jensen, M. Ø., Dror, R. O., Kuriyan, J., and Shaw, D. E. (2009) A conserved protonation-dependent switch controls drug binding in the Abl kinase, *Proceedings of the National Academy of Sciences of the United States of America* 106, 139-144.
180. Martínez-Mayorga, K., Pitman, M. C., Grossfield, A., Feller, S. E., and Brown, M. F. (2006) Retinal Counterion Switch Mechanism in Vision Evaluated by Molecular Simulations, *J Am Chem Soc* 128, 16502-16503.
181. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic acids research* 25, 3389-3402.
182. Berman, H., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank, *Nat Struct Biol* 10, 980.
183. Kleywegt, G. J., Harris, M. R., Zou, J. Y., Taylor, T. C., Wahlby, A., and Jones, T. A. (2004) The Uppsala Electron-Density Server, *Acta Crystallogr D Biol Crystallogr* 60, 2240-2249.
184. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation, *Proteins* 50, 437-450.
185. Henrick, K., and Thornton, J. M. (1998) PQS: a protein quaternary structure file server, *Trends in biochemical sciences* 23, 358-361.
186. Shi, J., Blundell, T. L., and Mizuguchi, K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties, *J Mol Biol* 310, 243-257.
187. Poirot, O., O'Toole, E., and Notredame, C. (2003) Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments, *Nucleic acids research* 31, 3503-3506.
188. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, CA, USA.
189. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2006) Comparative protein structure modeling using Modeller, *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al Chapter 5, Unit 5 6.*
190. Wiederstein, M., and Sippl, M. J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic acids research* 35, W407-410.
191. Pontius, J., Richelle, J., and Wodak, S. J. (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures, *Journal of Molecular Biology* 264, 121-136.
192. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *Journal of Applied Crystallography* 26, 283-291.
193. Edgar, R. C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity, *BMC Bioinformatics* 5, 113.

194. Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004) The Jalview Java alignment editor, *Bioinformatics* 20, 426-427.
195. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, *Mol Biol Evol* 24, 1596-1599.
196. Allen, F. (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising, *Acta Crystallographica Section B* 58, 380-388.
197. Liu, R., Wang, X. B., and Kong, L. Y. (2006) Dammaradienyl acetate, *Acta Crystallographica Section E-Structure Reports Online* 62, O3544-O3546.
198. Madureira, A. M., Duarte, M. T., Piedade, M. F. M., Ascenso, J. R., and Ferreira, M. J. U. (2004) Isoprenoid compounds from *Euphorbia portlandica*. X-ray structure of lupeportlandol, a new lupane triterpene, *Journal of the Brazilian Chemical Society* 15, 742-747.
199. Watson, W. H., Ting, H. Y., and Domingue, X. A. (1972) Molecular Structure of 3beta-Acetoxy-20-Hydroxylupane, C₃₂H₅₄O₃, *Acta Crystallographica Section B-Structural Crystallography and Crystal Chemistry B* 28, 8-8.
200. Flekhter, O. B., Medvedeva, N. I., and Suponitsky, K. Y. (2007) (6aR)-1-Hydroxy-2,2,6a,6b,9,9,12a-heptamethyl-10-(methylcarbonyloxy)perhydro-4-picenylmethyl acetate, *Acta Crystallographica Section E-Structure Reports Online* 63, O2289-O2290.
201. Silva, G. D. F., Duarte, L. P., Vieira, S. A., Doriguetto, A. C., Mascarenhas, Y. P., Ellena, J., Castellano, E. E., and Cota, A. B. (2002) Epikatonic acid from *Austroplenckia populnea*: structure elucidation by 2D NMR spectroscopy and X-ray crystallography, *Magnetic Resonance in Chemistry* 40, 366-370.
202. Schuttelkopf, A. W., and van Aalten, D. M. F. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes, *Acta Crystallographica Section D* 60, 1355-1363.
203. Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis, *Journal of Molecular Modeling* 7, 306-317.
204. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr D Biol Crystallogr* 50, 760-763.
205. Rodriguez, R., Chinea, G., Lopez, N., Pons, T., and Vriend, G. (1998) Homology modeling, model and software evaluation: three related resources, *Bioinformatics* 14, 523-528.
206. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics* 23, 2947-2948.
207. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* 22, 2577-2637.
208. Sippl, M. J. (1993) Recognition of Errors in 3-Dimensional Structures of Proteins, *Proteins-Structure Function and Genetics* 17, 355-362.
209. Gill, G., Pawar, D. M., and Noe, E. A. (2005) Conformational Study of cis-1,4-Di-tert-butylcyclohexane by Dynamic NMR Spectroscopy and Computational Methods. Observation of Chair and Twist-Boat Conformations, *The Journal of Organic Chemistry* 70, 10726-10731.
210. Kushiro, T., Shibuya, M., Masuda, K., and Ebizuka, Y. (2000) A novel multifunctional triterpene synthase from *Arabidopsis thaliana*, *Tetrahedron Letters* 41, 7705-7710.
211. Corey, E. J., Cheng, H. M., Baker, C. H., Matsuda, S. P. T., Li, D., and Song, X. L. (1997) Methodology for the preparation of pure recombinant *S-cerevisiae* lanosterol synthase using a Baculovirus expression system. Evidence that oxirane cleavage and A-ring formation are concerted in the

- biosynthesis of lanosterol from 2,3-oxidosqualene, *J Am Chem Soc* 119, 1277-1288.
212. Gandour, R. D. (1981) On the importance of orientation in general base catalysis by carboxylate, *Bioorganic Chemistry* 10, 169-176.
213. Fischer, W. W., and Pearson, A. (2007) Hypotheses for the origin and early evolution of triterpenoid cyclases, *Geobiology* 5, 19-34.
214. Kushiro, T., Shibuya, M., and Ebizuka, Y. (1999) Chimeric triterpene synthase. A possible model for multifunctional triterpene synthase, *J Am Chem Soc* 121, 1208-1216.
215. Qin, B., Eagles, J., Mellon, F. A., Mylona, P., Pena-Rodriguez, L., and Osbourn, A. E. (2010) High throughput screening of mutants of oat that are defective in triterpene synthesis, *Phytochemistry* 17, 1245-1252.
216. Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucleic Acids Symposium Series* 41, 95-98.
217. Rines, H. W. (1985) Sodium-Azide Mutagenesis in Diploid and Hexaploid Oats and Comparison with Ethyl Methanesulfonate Treatments, *Environmental and Experimental Botany* 25, 7-16.
218. Kleinhofs, A., Owais, W. M., and Nilan, R. A. (1978) Azide, *Mutat Res* 55, 165-195.
219. Kleinhofs, A., Sander, C., Nilan, R. A., and Konzack, S. (1974) Azide mutagenicity—mechanism and nature of mutants produced, pp 196-199, International Atomic Energy Agency, Vienna.
220. Kleinhofs, A., Warner, R. L., Murhlbauer, F. J., and Nilan, R. A. (1978) Induction and selection of specific gene mutations in hordeum and pisum, *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 51, 29-35.
221. Sander, C., Nilan, R. A., Kleinhofs, A., and Vig, B. K. (1978) Mutagenic and chromosome-breaking effects of azide in barley and human leukocytes, *Mutat Res* 50, 67-75.
222. Fasken, M. B., and Corbett, A. H. (2005) Process or perish: quality control in mRNA biogenesis, *Nat Struct Mol Biol* 12, 482-488.
223. Chang, Y. F., Imam, J. S., and Wilkinson, M. F. (2007) The nonsense-mediated decay RNA surveillance pathway, *Annu Rev Biochem* 76, 51-74.
224. Silva, A. L., and Romao, L. (2009) The mammalian nonsense-mediated mRNA decay pathway: to decay or not to decay! Which players make the decision?, *FEBS letters* 583, 499-505.
225. Stalder, L., and Muhlemann, O. (2008) The meaning of nonsense, *Trends Cell Biol* 18, 315-321.
226. Reddy, A. S. (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era, *Annu Rev Plant Biol* 58, 267-294.
227. Lorkovic, Z. J., Wieczorek Kirk, D. A., Lambermon, M. H., and Filipowicz, W. (2000) Pre-mRNA splicing in higher plants, *Trends in plant science* 5, 160-167.
228. Lewin, B. (2004) The interrupted gene, In *Genes VIII* (Carlson, G., Ed.), pp 33-49, Pearson Education, Inc., Upper Saddle River, New Jersey.
229. Jaccoud, D., Peng, K., Feinstein, D., and Kilian, A. (2001) Diversity arrays: a solid state technology for sequence information independent genotyping, *Nucleic acids research* 29, E25.
230. Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Kleinhofs, A., and Kilian, A. (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley, *Proceedings of the National Academy of Sciences of the United States of America* 101, 9915-9920.
231. Roberts, A., and Withers, P. (2007) statistiXL, v 1.8 ed., Broadway-Nedlands.

232. Petracek, M. E., Nuygen, T., Thompson, W. F., and Dickey, L. F. (2000) Premature termination codons destabilize ferredoxin-1 mRNA when ferredoxin-1 is translated, *Plant J* 21, 563-569.
233. Saito, M., and Nakamura, T. (2005) Two point mutations identified in emmer wheat generate null Wx-A1 alleles, *Theor Appl Genet* 110, 276-282.
234. Gadjeva, R., Axelsson, E., Olsson, U., Vallon-Christersson, J., and Hansson, M. (2004) Nonsense-mediated mRNA decay in barley mutants allows the cloning of mutated genes by a microarray approach, *Plant Physiol Biochem* 42, 681-685.
235. Isshiki, M., Yamamoto, Y., Satoh, H., and Shimamoto, K. (2001) Nonsense-mediated decay of mutant waxy mRNA in rice, *Plant Physiol* 125, 1388-1395.
236. Jofuku, K. D., Schipper, R. D., and Goldberg, R. B. (1989) A frameshift mutation prevents Kunitz trypsin inhibitor mRNA accumulation in soybean embryos, *The Plant cell* 1, 567.
237. Voelker, T. A., Moreno, J., and Chrispeels, M. J. (1990) Expression analysis of a pseudogene in transgenic tobacco: a frameshift mutation prevents mRNA accumulation, *The Plant cell* 2, 255-261.
238. Voelker, T. A., Staswick, P., and Chrispeels, M. J. (1986) Molecular analysis of two phytohemagglutinin genes and their expression in *Phaseolus vulgaris* cv. Pinto, a lectin-deficient cultivar of the bean, *EMBO J* 5, 3075-3082.
239. Que, Q., Wang, H. Y., English, J. J., and Jorgensen, R. A. (1997) The Frequency and Degree of Cosuppression by Sense Chalcone Synthase Transgenes Are Dependent on Transgene Promoter Strength and Are Reduced by Premature Nonsense Codons in the Transgene Coding Sequence, *The Plant cell* 9, 1357-1368.
240. Kusumi, J., and Iba, K. (1998) Characterization of a nonsense mutation in FAD7, the gene which encodes Δ³ desaturase in *Arabidopsis thaliana*, *Journal of Plant Research* 111, 87-91.
241. Marchant, A., and Bennett, M. J. (1998) The *Arabidopsis* AUX1 gene: a model system to study mRNA processing in plants, *Plant Mol Biol* 36, 463-471.
242. Hilleren, P., and Parker, R. (1999) mRNA surveillance in eukaryotes: kinetic proofreading of proper translation termination as assessed by mRNP domain organization?, *RNA* 5, 711-719.
243. Chiba, Y., and Green, P. J. (2009) mRNA degradation machinery in plants, *Journal of Plant Biology* 52, 114-124.
244. Weiner, M. P., Anderson, C., Jerpseth, B., Wells, S., Johnson-Browne, B., and Vaillancourt, P. (1994) *Strategies* 7, 41-43.
245. Neugebauer, J. M. (1990) Detergents: an overview, *Methods Enzymol* 182, 239-253.
246. Frankel, S., Sohn, R., and Leinwand, L. (1991) The use of sarkosyl in generating soluble protein after bacterial expression, *Proceedings of the National Academy of Sciences of the United States of America* 88, 1192-1196.
247. Karas, M., and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, *Analytical Chemistry* 60, 2299-2301.
248. Pappin, D. J. C., Hojrup, P., and Bleasby, A. J. (1993) Rapid Identification of Proteins by Peptide-Mass Fingerprinting, *Current Biology* 3, 327-332.
249. Perkins, D. N., Pappin, D. J. C., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data, *Electrophoresis* 20, 3551-3567.
250. Ellgaard, L., and Helenius, A. (2003) Quality control in the endoplasmic reticulum, *Nat Rev Mol Cell Biol* 4, 181-191.

251. Goldberg, A. L., and Dice, J. F. (1974) Intracellular protein degradation in mammalian and bacterial cells, *Annu Rev Biochem* 43, 835-869.
252. Sherman, M. Y., and Goldberg, A. L. (2001) Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases, *Neuron* 29, 15-32.
253. Goldberg, A. L. (1972) Degradation of abnormal proteins in *Escherichia coli* (protein breakdown-protein structure-mistranslation-amino acid analogs-puromycin), *Proceedings of the National Academy of Sciences of the United States of America* 69, 422-426.
254. Malhotra, J. D., and Kaufman, R. J. (2007) The endoplasmic reticulum and the unfolded protein response, *Semin Cell Dev Biol* 18, 716-731.
255. Hoseki, J., Ushioda, R., and Nagata, K. (2010) Mechanism and components of endoplasmic reticulum-associated degradation, *J Biochem* 147, 19-25.
256. Oliaro-Bosso, S., Schulz-Gasch, T., Balliano, G., and Viola, F. (2005) Access of the substrate to the active site of yeast oxidosqualene cyclase: an inhibition and site-directed mutagenesis approach, *Chem Bio Chem* 6, 2221-2228.
257. Dean, P. D. G., Demontel, P., Bloch, K., and Corey, E. J. (1967) A Soluble 2,3-Oxidosqualene Sterol Cyclase, *J Biol Chem* 242, 3014-&.
258. Corey, E. J., and Matsuda, S. P. T. (1991) Purification of the 2,3-Oxidosqualene-Lanosterol Cyclase from *Saccharomyces-Cerevisiae*, *J Am Chem Soc* 113, 8172-8174.
259. Balliano, G., Viola, F., Ceruti, M., and Cattell, L. (1992) Characterization and partial purification of squalene-2,3-oxide cyclase from *Saccharomyces cerevisiae*, *Archives of biochemistry and biophysics* 293, 122-129.
260. Moore, W. R., and Schatzman, G. L. (1992) Purification of 2,3-Oxidosqualene Cyclase from Rat-Liver, *J Biol Chem* 267, 22003-22006.
261. Wu, T. K., Huang, C. Y., Ko, C. Y., Chang, C. H., Chen, Y. J., and Liao, H. K. (2004) Purification, tandem mass characterization, and inhibition studies of oxidosqualene-lanosterol cyclase enzyme from bovine liver, *Archives of biochemistry and biophysics* 421, 42-53.
262. Abe, I., Sankawa, U., and Ebizuka, Y. (1992) Purification and Properties of Squalene-2,3-Epoxyde Cyclases from Pea-Seedlings, *Chemical & pharmaceutical bulletin* 40, 1755-1760.
263. de Gier, J. W., and Luirink, J. (2001) Biogenesis of inner membrane proteins in *Escherichia coli*, *Molecular microbiology* 40, 314-322.
264. Johnson, A. E., and van Waes, M. A. (1999) The translocon: a dynamic gateway at the ER membrane, *Annu Rev Cell Dev Biol* 15, 799-842.
265. Muller, M., Koch, H. G., Beck, K., and Schafer, U. (2001) Protein traffic in bacteria: multiple routes from the ribosome to and across the membrane, *Prog Nucleic Acid Res Mol Biol* 66, 107-157.
266. Tate, C. G. (2001) Overexpression of mammalian integral membrane proteins for structural studies, *FEBS letters* 504, 94-98.
267. Cai, J., Daoud, R., Alqawi, O., Georges, E., Pelletier, J., and Gros, P. (2002) Nucleotide binding and nucleotide hydrolysis properties of the ABC transporter MRP6 (ABCC6), *Biochemistry* 41, 8058-8067.
268. Lerner-Marmarosh, N., Gimi, K., Urbatsch, I. L., Gros, P., and Senior, A. E. (1999) Large scale purification of detergent-soluble P-glycoprotein from *Pichia pastoris* cells and characterization of nucleotide binding properties of wild-type, Walker A, and Walker B mutant proteins, *J Biol Chem* 274, 34711-34718.
269. Sarramegna, V., Demange, P., Milon, A., and Talmont, F. (2002) Optimizing functional versus total expression of the human mu-opioid receptor in *Pichia pastoris*, *Protein expression and purification* 24, 212-220.

270. Schiller, H., Molsberger, E., Janssen, P., Michel, H., and Reilander, H. (2001) Solubilization and purification of the human ETB endothelin receptor produced by high-level fermentation in *Pichia pastoris*, *Receptors Channels* 7, 453-469.
271. Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., and Hughes, C. (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export, *Nature* 405, 914-919.
272. Weiss, H. M., and Grisshammer, R. (2002) Purification and characterization of the human adenosine A(2a) receptor functionally expressed in *Escherichia coli*, *European journal of biochemistry / FEBS* 269, 82-92.
273. Garavito, R. M., and Ferguson-Miller, S. (2001) Detergents as Tools in Membrane Biochemistry, *J Biol Chem* 276, 32403-32406.
274. Landau, E. M., and Rosenbusch, J. P. (1996) Lipidic cubic phases: a novel concept for the crystallization of membrane proteins, *Proceedings of the National Academy of Sciences of the United States of America* 93, 14532-14535.
275. Faham, S., and Bowie, J. U. (2002) Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure, *J Mol Biol* 316, 1-6.
276. Cregg, J. M., Madden, K. R., Barringer, K. J., Thill, G. P., and Stillman, C. A. (1989) Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*, *Mol Cell Biol* 9, 1316-1323.
277. Daly, R., and Hearn, M. T. (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production, *J Mol Recognit* 18, 119-138.
278. Cereghino, J. L., and Cregg, J. M. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*, *Fems Microbiology Reviews* 24, 45-66.
279. Tschopp, J. F., Brust, P. F., Cregg, J. M., Stillman, C. A., and Gingeras, T. R. (1987) Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*, *Nucleic acids research* 15, 3859-3876.
280. Ellis, S. B., Brust, P. F., Koutz, P. J., Waters, A. F., Harpold, M. M., and Gingeras, T. R. (1985) Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*, *Mol Cell Biol* 5, 1111-1121.
281. Koutz, P., Davis, G. R., Stillman, C., Barringer, K., Cregg, J., and Thill, G. (1989) Structural comparison of the *Pichia pastoris* alcohol oxidase genes, *Yeast* 5, 167-177.
282. Berdy, J. (1980) *Handbook of Antibiotic Compounds*, CRC Press, Boca Raton, Fl.
283. Drocourt, D., Calmels, T., Reynes, J. P., Baron, M., and Tiraby, G. (1990) Cassettes of the *Streptoalloteichus hindustanus* ble gene for transformation of lower and higher eukaryotes to phleomycin resistance, *Nucleic acids research* 18, 4009.
284. Hasslacher, M., Schall, M., Hayn, M., Bona, R., Rumbold, K., Lückl, J., Griengl, H., Kohlwein, S. D., and Schwab, H. (1997) High-Level Intracellular Expression of Hydroxynitrile Lyase from the Tropical Rubber Tree *Hevea brasiliensis* in Microbial Hosts, *Protein expression and purification* 11, 61-71.
285. Klein, C., de Lamotte-Guéry, F., Gautier, F., Moulin, G., Boze, H., Joudrier, P., and Gautier, M.-F. (1998) High-Level Secretion of a Wheat Lipid Transfer Protein in *Pichia pastoris*, *Protein expression and purification* 13, 73-82.
286. Sonia, H., Mayte, V., Eva, G. I., Antonio, M.-R., and Rosalía, R. (1999) Production and detailed characterization of biologically active olive pollen

- allergen Ole e 1 secreted by the yeast *Pichia pastoris*, *European Journal of Biochemistry* 261, 539-546.
287. Inan, M., and Meagher, M. M. (2001) Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of *Pichia pastoris*, *Journal of Bioscience and Bioengineering* 92, 585-589.
288. Cregg, J. M., Vedvick, T. S., and Raschke, W. C. (1993) Recent advances in the expression of foreign genes in *Pichia pastoris*, *Biotechnology* 11, 905-910.
289. Sakaki, T., Zahringer, U., Warnecke, D. C., Fahl, A., Knogge, W., and Heinz, E. (2001) Sterol glycosides and cerebrosides accumulate in *Pichia pastoris*, *Rhynchosporium secalis* and other fungi under normal conditions or under heat shock and ethanol stress, *Yeast* 18, 679-695.
290. Stein, S. E. (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data, *Journal of the American Society for Mass Spectrometry* 10, 770-781.
291. Stellwagen, E. (1990) Gel filtration, *Methods in Enzymology* 182, 317-328.
292. Hofstee, B. H. J. (1973) Hydrophobic affinity chromatography of proteins, *Analytical Biochemistry* 52, 430-448.
293. Porath, J., Sundberg, L., Fornstedt, N., and Olsson, I. (1973) Salting-out in Amphiphilic Gels as a New Approach to Hydrophobia Adsorption, *Nature* 245, 465-466.
294. Tiselius, A. (1948) Adsorption separation by salting out, *Arkiv för Kemi, Mineralogi och Geologi* 26B, 1-5.