FACULTE DE BIOLOGIE ET MEDECINE

Département de Biologie Moléculaire Végétale

CHARACTERIZATION AND FUNCTIONAL IDENTIFICATION OF A NOVEL PLANT EXTRADIOL 4,5-DIOXYGENASE INVOLVED IN BETALAIN PIGMENT BIOSYNTHESIS IN *PORTULACA GRANDIFLORA*

Thèse de doctorat

présentée à la

Faculté de Biologie et Médecine de
l'Université de Lausanne

par

LAURENT CHRISTINET

Diplômé en biotechnologie de
L'Ecole Supérieure de Biotechnologie de Strasbourg

Jury

Prof. Dieter Haas, Président
Prof. Jean-Pierre Zrÿd, Directeur de thèse
Prof. Nikolaus Amrhein, Expert
Prof. Dieter Strack, Expert

LAUSANNE

2004

To my wife Anouk, and my children Marie and Nathan

AKNOWLEDGMENTS

I would like to thank Professor Jean-Pierre Zrÿd for giving me the opportunity to undertake this fascinating project in his laboratory and for its critical contribution to the "in-silico" identification of the DODA gene. I strongly appreciated the freedom and the confidence he granted me during the last five years.

My particular thanks go to Dr. Didier Schaefer for his support and his expertise concerning *Physcomitrella patens* transformation and molecular biology in general and Andrija Finka for his advices in microscopy and the construction of the DsRed2 vector.

I would like to thanks Dr. Ursula Hinz for introducing me to the world of *Portulaca grandiflora,* betalains and HPLC analysis during the first two years of the project.

I particularly appreciated the contribution of Silvia Marchesini to the GC-MS analysis and her technical support concerning PHA analysis.

I am grateful to Dr. Pagni and to Alexander Diemand from the Swiss Institute of Bioinformatics (Epalinges) for their advice concerning three-dimensional modeling and to Professor Yves Poirier of the DBMV (Lausanne University) for his fruitful discussions on fatty acid degradation.

My sincere thanks go to Professor Dieter Strack and Nikolaus Amrhein for accepting to be experts on the thesis committee and their important suggestions concerning the writing of this manuscript.

My special thanks go to Adrian Kessler for the correction of the english language of this thesis.

I would like to thank all members of the IE-PC and IE-BPV for their technical help, suggestions and the pleasant atmosphere.

Finally, my thoughts go to my family for their essential logistical help, and especially to my wife Anouk for her interest in my work and her love.

TABLE OF CONTENT

RESUME

Les bétalaïnes sont des pigments chromo-alcaloïdes violets et jaunes présents dans les plantes appartenant à l'ordre des Caryophyllales et dans les champignons des genres *Amanita* et *Hygrocybe*. Leur courte voie de biosynthèse est élucidée chimiquement depuis de nombreuses années, mais les enzymes impliquées dans cette biosynthèse chez les plantes ne sont toujours pas caractérisées. L'enzyme de la DOPA-dioxygénase d'*Amanita muscaria* a été identifiée (Girod et Zrÿd, 1991a), mais de nombreuses tentatives d'isolation d'un homologue chez les plantes ont échoué.

Afin d'isoler les gènes spécifiques des bétalaïnes chez les plantes, nous avons construit des banques soustraites d'ADNc à partir d'ARN total de pétales immatures de *Portulaca grandiflora* (*Pg*) de génotypes jaunes et blancs, respectivement violets et blancs. Les clones couleur-spécifiques ont été détectés en premier par analyse Northern du RNA de pétales blancs et colorés. Les candidats positifs ont alors été soumis à une analyse de transcription au niveau des tiges colorées, vertes et des feuilles, afin d'établir leur expression spécifique. Deux ARNs messagers complets ont une expression corrélée avec l'accumulation des bétalaïnes dans les tissus. Le premier de ces clones, *A.16,* code pour une oxydase de l'acyl-Coenzyme A (ACX) putative, mais le domaine de liaison du FAD essentiel pour l'activité d'ACX est absent. Toutes nos tentatives pour démontrer sa fonction ont échoué. Le rôle de cette protéine dans la voie de synthèse des bétalaïnes reste inconnu.

Le deuxième de ces clones spécifique aux bétalaïnes, *L.6* (isolé par Zaiko, 2000), a été renommé *DODA* en raison de son homologie avec le domaine LigB (pfam02900) d'une 4,5-dioxygénase extradiol bactérienne. *DODA* a été identifié in silico comme une dioxygénase extradiol en raison de la conservation stricte, au niveau de sa séquence peptidique, des résidus catalytiques de LigB et de ceux liant le cofacteur fer. Une analyse de transfert Southern a montré que ce gène est unique dans *Pg*. L'expression transitoire de *DODA* par transformation biolistique dans des pétales blancs de *Pg* a produit des taches violettes ou jaunes dans des cellules transformées. Une analyse HPLC de ces taches a démontré leur identité avec les bétalaïnes présentes naturellement dans les pétales violets et jaunes de *Pg*, confirmant ainsi la complémentation par le gène *Pg DODA* de l'allèle récessif *cc* présent dans les pétales blancs de *Pg*.

Des homologues de *DODA* (DOPA-dioxygénase) ont été identifiés dans de nombreuses espèces de plantes, y compris dans celles sans bétalaïne. L'alignement de ces homologues a permis l'identification d'un motif spécifique aux bétalaïnes à côté d'une histidine catalytique conservée. Ce motif [H-P-(S,A)-(N,D)-x-T-P] remplace le motif [H-N-L-R] conservé dans les plantes sans bétalaïne et le motif [H-N-L-x] présent dans tous les homologues bactériens et archaebactériens. Une modélisation tridimensionnelle préliminaire du site actif de *Pg* DODA et de son homologue dans la mousse *Physcomitrella patens* a montré l'importance de ce motif spécifique aux bétalaïnes pour l'accessibilité du substrat au site actif.

L'analyse phylogénétique de DODA a confirmé l'évolution séparée de cette protéine chez les plantes à bétalaïnes par comparaison avec celle des plantes sans bétalaïne. Nous avons donc conclu que les bétalaïnes sont apparues par modification de l'affinité pour un substrat d'enzymes similaires à DODA, chez un ancêtre unique des Caryophyllales qui a perdu toute capacité de biosynthèse des anthocyanes. Finalement, *Pg* DODA n'a aucune similarité avec la protéine DODA d'*Amanita muscaria*, bien que celle-ci complémente aussi la pigmentation des pétales blancs de *Pg*. La biosynthèse des bétalaïnes est un exemple remarquable de convergence évolutive biochimique indépendante entre espèces de règnes différents.

ABSTRACT

Betalains are violet and yellow chromo-alkaloid pigments present in plants belonging to the order Caryophyllales and also in the fungal genera *Amanita* and *Hygrocybe*. Their short biosynthetic pathway is chemically well understood since many years, but enzymes involved in the plant pathway are still uncharacterized. The DOPA-dioxygenase from *Amanita muscaria* was identified (Girod and Zrÿd, 1991a), but numerous attempts to identify a plant homologue to the corresponding gene, failed.

In order to isolate betalain-specific genes in plants, subtractive cDNA libraries were built with total RNA from white and yellow and respectively, violet immature petals from *Portulaca grandiflora* (*Pg*) genotypes. Colour-specific clones were first detected by Northern blot analysis using RNA from white and coloured petals. Positive candidates were submitted to further transcription analysis in coloured, green stems and leaves in order to assess their specific expression. Two full-length mRNAs showed a correlated expression with betalain accumulation in tissues. One of them, *A.16*, encodes a putative acyl-Coenzyme A oxidase (ACX), but missing the FAD binding domain essential for the ACX activity. Thus, all attempts to demonstrate its function failed. The role of this protein in the betalain biosynthesis pathway, if any, is still unknown.

The second betalain-specific mRNA, *L.6* (isolated by Zaiko, 2000) shows a homology with a LigB domain (pfam02900) from a bacterial extradiol 4,5-dioxygenase. It was then renamed *DODA* (DOPA-dioxygenase). *DODA* was identified in silico as a highly conserved extradiol dioxygenase due to the strict conservation of its peptidic sequence with LigB catalytic residues and iron-binding cofactor residues. Southern blot analysis showed that this gene is a single copy-gene in *Pg*. Transient expression of DODA protein through biolistic transformation of *Pg* white petals produced violet or yellow spots in individual cells. HPLC analysis of these spots showed an identity with betalain pigments present naturally in yellow and violet *Pg* petals, thus confirming the complementation of the recessive *cc* allele present in *Pg* white petals by *Pg DODA* gene.

DODA homologues were identified in numerous plant species including those without betalain. Alignment of these homologues allowed the identification of a betalain-specific pattern beside a highly conserved catalytic histidine. This [H-P-(S,A)-(N,D)-x-T-P] pattern replaces a [H-N-L-R] pattern strictly conserved in non-betalain plants and a [H-N-L-x] pattern present in all bacterial and archaebacterial homologues. Preliminary three-dimensional modeling of the active site of *Pg* DODA and its *Physcomitrella patens* moss homologue revealed the importance of this betalain-specific pattern for the substrate accessibility to the DODA active site.

DODA phylogenetic analysis confirmed the separate evolution of this protein in betalain-producing plants. We conclude that betalain pigments appeared in a unique ancestor of the Caryophyllales order in which anthocyanin biosynthetic pathway was impaired, by a modification of enzymes of the DODA family for substrate affinity. The *Pg* DODA protein has no sequence similarity with *Amanita muscaria* DODA, despite the fact that they both complement *Pg* white petals for their pigmentation. Betalain biosynthesis is an interesting example of independent biochemical evolutionary convergence between species from different kingdoms.

ABBREVIATIONS

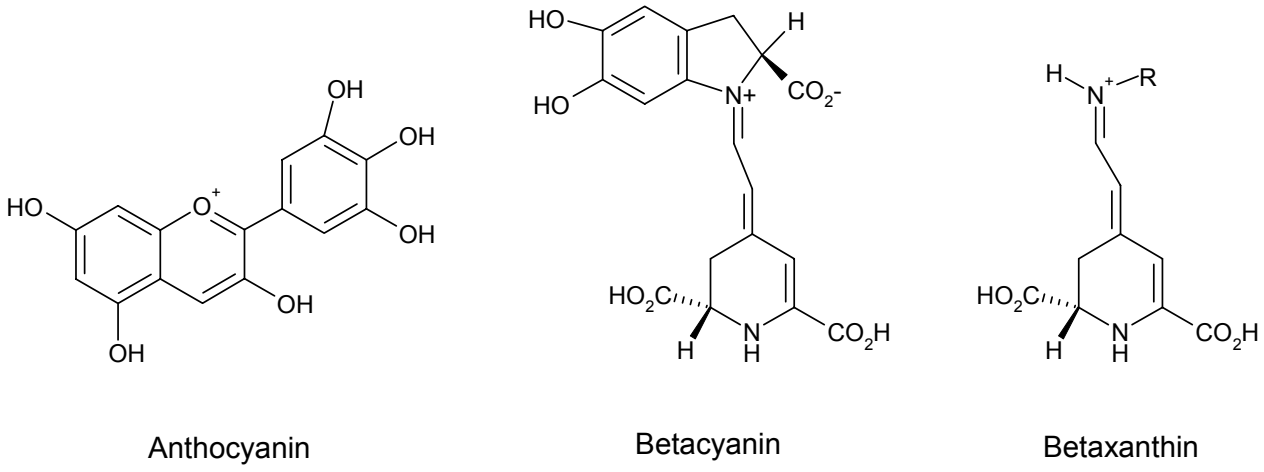
2,4-D	2,4-dichlorophenoxyacetic acid
2-ODD	2-oxoacid dependent dioxygenase
6-BAP	6-benzylaminopurine
ACX	acyl-Coenzyme A oxidase
BLAST	Basic Local Alignment Research Tool
bp	base pair
CaMV	cauliflower mosaic virus
cDNA	complementary deoxyribonucleic acid
Da	dalton
dCTP	deoxycytidine triphosphate
DFR	dihydroflavonol 4-reductase
DODA	DOPA-dioxygenase
DOPA	dihydroxyphenylalanine
EMBL	European Molecular Biology Laboratory
EST	expressed sequence tag
FAD	flavin adenin dinucleotide
GA	gibberellic acid
GC-MS	gas chromatography mass spectrometer
gDNA	genomic DNA
GMO	genetically modified organism
GPD	glyceraldehyde 3-phosphate dehydrogenase
GSP	gene specific primer
GT	glucosyl transferase
HPLC	high-performance liquid chromatography
Kan	kanamycin
kb	kilobase
λmax	maximum wavelength

LDL	low density lipoprotein
mRNA	messenger ribonucleic acid
o/n	overnight
2-ODD	2-oxoglutarate-dependent dioxygenase
ORF	open reading frame
PCR	polymerase chain reaction
PEG	polyethylene glycol
Pg	*Portulaca grandiflora*
PHA	polyhydroxyalcanoates
PI	isoelectric point
PPO	polyphenol oxidase
PTS	peroxisomal targeting signal
RACE	rapid amplification of cDNA ends
ROS	reactive oxygen species
RT	reverse transcriptase
SDS	sodium dodecyl sulfate
SSC	sodium chloride/sodium citrate buffer
SSH	suppression subtractive hybridization
TE	tris-ethylenediamine-tetra-acetate buffer
UDP-glucose	uridine 5'-diphosphate glucose
UTR	untranslated region
UV	ultraviolet
YEP	yeast extract peptone

I. GENERAL INTRODUCTION

1. The betalains a particular class of pigments

Betalains are a class of yellow (betaxanthins) or violet (betacyanins) vacuolar water-soluble chromo-alkaloid pigments accumulating in flowers, fruit and sometimes in vegetative tissues of plants from the Caryophyllales order. These secondary metabolites are also produced in vegetative tissues after insect, pathogen or virus attacks against plants, as observed in the case of the beet necrotic yellow vein virus (Steddom *et al.*, 2003). Wohlpart and Mabry (1968) introduced the term "betalain" to describe these pigments as derivatives of betalamic acid, identified in the red beet plant (*Beta vulgaris*). Previously, they were usually called "nitrogenous anthocyanins", referring incorrectly to a direct structural relationship with the flavonoid anthocyanins (Fig. 1)



Anthocyanin Betacyanin Betaxanthin

Figure 1. Comparison of the structure of anthocyanins (example: delphinidin) with betacyanins (ex: betanidin) and betaxanthins (R = amino acid lateral chain or an amine).

In fact, these violet betacyanins and yellow betaxanthins replace anthocyanins in most families of the order Caryophyllales, except in the Caryophyllaceae and the Molluginaceae families where anthocyanins are still present (see Table 1, (Clement and Mabry, 1996)). Mutual exclusion of betalain and anthocyanin pigments (Wyler and Dreiding, 1961) provides the main chemotaxonomic criterion to differentiate the Chenopodiineae from the Caryophyllineae in the Caryophyllales order.

Betalains and anthocyanins have a similar localisation in plants and operate with the same spectrum of colours to attract insects and animals for pollination and seed dispersal (Clement and Mabry, 1996). The origin of betalains is a fascinating question still unresolved; therefore several hypotheses coexist. Surprisingly, these pigments have also been identified in the pilei from a very restricted number of basidiomycete species belonging to the genera *Amanita* and *Hygrocybe* (Döpp and Musso, 1973b; von Ardenne *et al.*, 1974). Their role is completely unknown in fungi. Phylogenetic significance of betalains will be further discussed in the light of new results obtained in the understanding of betalain biosynthesis genetics. Due to their colourant properties, betalains can be used in various applications in the food industry, the horticultural sector or the agroscience industry. People are more often in contact with betalains than they think, while eating red beet or strawberry yoghourt coloured with beet juice, or looking at cactus or amaranth flowers.

Suborder	Family	Examples of genus
Chenopodiineae		
Betalain-producing anthocyanin-free taxa	Achatocarpaceae	*Achatocarpus*
	Aizoaceae	*Dorotheanthus*, *Mesembryanthemum*
	Amaranthaceae	*Amaranthus*, *Iresine*, *Gomphrena*
	Basellaceae	*Basella*
	Cactaceae	*Mammillaria*, *Opuntia*, *Pereskia*
	Chenopodiaceae	*Beta*, *Chenopodium*, *Spinacia*
	Didiereaceae	*Didierea*
	Halophytaceae	*Halophytum*
	Hectorellaceae	*Hectorella*
	Nyctaginaceae	*Bougainvillea*, *Mirabilis*
	Phytolaccaceae	*Phytolacca*, *Gisekia*
	Portulacaceae	*Portulaca*, *Claytonia*
	Stegnospermataceae	*Stegnosperma*
Caryophyllineae		
Betalain-free anthocyanin-producing taxa	Caryophyllaceae	*Dianthus*, *Silene*
	Molluginaceae	*Mollugo*, *Limeum*

Table 1. The most recent classification of Caryophyllales (Clement and Mabry, 1996).

2. The structure of betalains

The main structural characteristic of all betalains is the presence of the betalamic acid chromophore, a dihydropyridine moiety attached via a vinyl group to another nitrogenous group (Miller *et al.*, 1968). Its lemon yellow colour (λmax 424 nm) results from the resonance system induced by the presence of three conjugated double bonds (Fig. 2). Betaxanthins are formed by the condensation of an amino acid or an amine with the aldehyde group of the betalamic acid, resulting in a Schiff-base (Fig. 2). This basic structure gives the stronger yellow or yellow-orange colours to the betaxanthins with a maximum of absorbance at 470-486 nm. In spite of their colour differences, betacyanins are also made of a betalamic acid unit linked to a molecule of cyclo-DOPA (Wyler and Dreiding, 1961) (Fig. 2). The violet colour of the latter is due to an aromatic structure inducing a strong shift of 60 nm in the absorbance maximum (λmax 534-554 nm). Around the conserved betalamic acid moiety, betacyanin and betaxanthin structures can vary in numerous ways by conjugation reactions like glycosylation or acylation. A short classification will be given for both pigments. A more detailed view on new structure elucidations and the methods used for their characterization has been reviewed recently (Strack *et al.*, 2003).

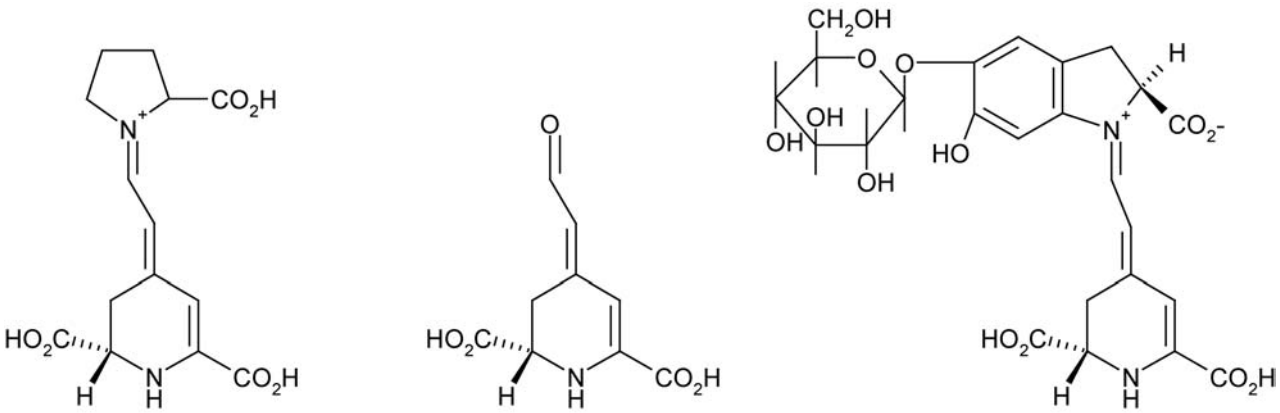
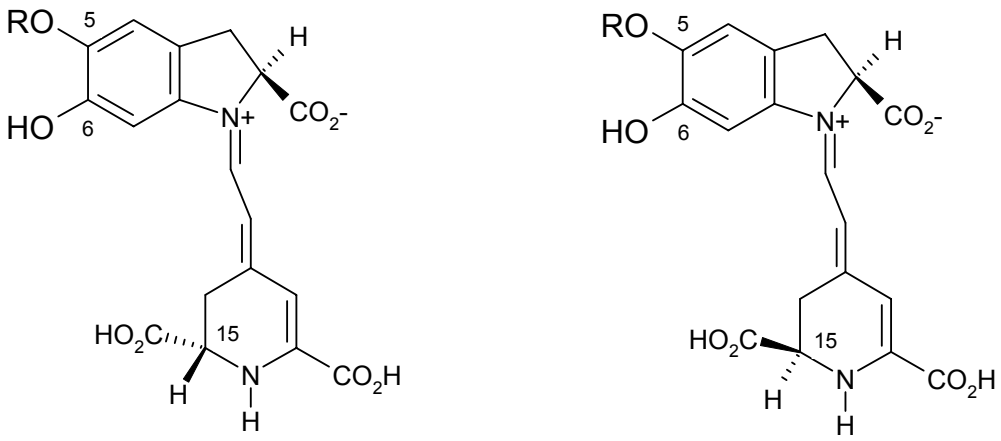


Figure 2. Chemical structure of betalamic acid, the main chromophore of betalains. This molecule is present in all betaxanthins linked with an amino acid or an imino compound (ex: indicaxanthin from *Opuntia ficus-indica*), and in all betacyanins linked with the cyclo-DOPA molecule (ex: betanin from red beet).

2.1 The betacyanins

Natural betacyanins are mainly composed of the non-glycosidated betanidin or isobetanidin chromophores obtained by the condensation of cyclo-DOPA with betalamic acid (Wyler *et al.*, 1963). Both molecules differ only by the absolute configuration of their C_{15} chiral center (Wilcox *et al.*, 1965) (Fig. 3).



Betanidin Isobetanidin

Figure 3. Non-glycosidated betacyanin chromophores differ only by the C_{15} chiral center (R=H).

Most betacyanins are derived from these two compounds by *O*-glucosylation on one of the two free hydroxyl groups of cyclo-DOPA (see Table 2). Glucosylation on position 5 is the most frequent one, as observed in the red beet pigment betanin (Fig. 4). Some rare pigments, like gomphrenins have been identified as 6-*O*-glucosides (Minale *et al.*, 1967). Until recently, the sole non-glycosidated betacyanin was the 2-descarboxy-betanidin (Fig. 4), a minor pigment from *Carpobrotus acinaciformis* flowers (Aizoaceae) derived from dopamine rather than cyclo-DOPA (Piattelli and Impellizzeri, 1970). Lately, small amounts of this pigment were also identified in yellow beet (Schliemann *et al.*, 1999) accompanied by two new similar pigments, the 2-descarboxy-betanin and the 6'-*O*-malonyl-2-descarboxy-betanin (Kobayashi *et al.*, 2001).

The majority of betacyanins are acylated on their glucoside part via an ester linkage mostly with ferulic or *para*-coumaric acid. Malonylation also occurs in the betacyanin structure and is known for generally stabilizing pigments in flowers, preventing for example anthocyanins from beta-glucosidases attack (Suzuki *et al.*, 2002). A new type of acylated betacyanin containing both an aliphatic and a hydroxycynnamoyl aromatic acyl residue has been detected in *Phytolacca americana* (Schliemann *et al.*, 1996). This kind of acylation is also observed in complex anthocyanins (Strack and Wray, 1989).

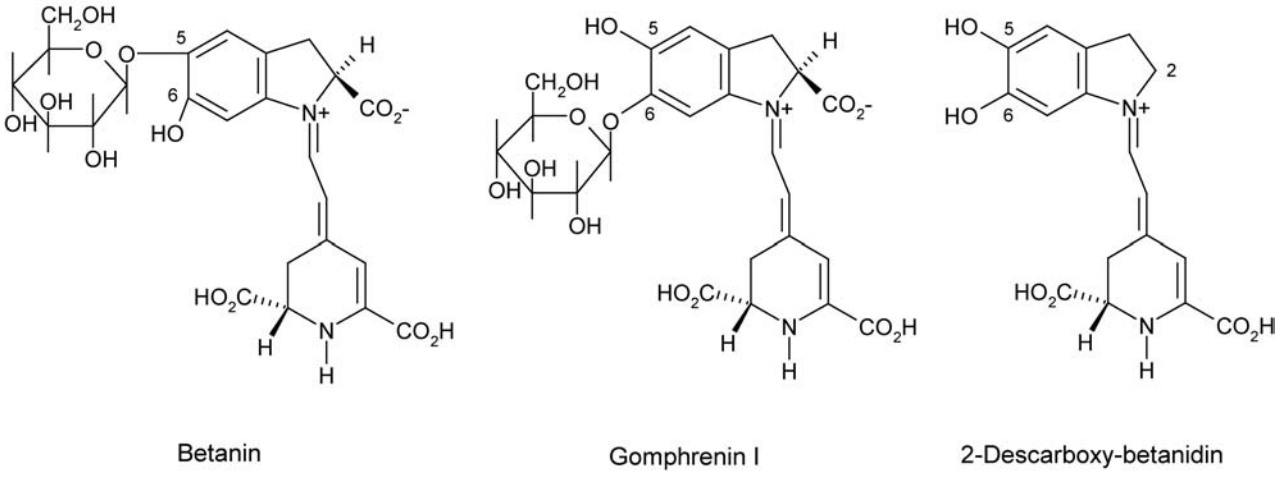


Figure 4. Comparison of different types of glycosylation. Betanidin 5-*O*-glucoside (betanin), betanidin 6-*O*-glucoside (gomphrenin I) and the non-glycosidated 2-descarboxy-betanin.

Aglycones	**References**
Betanidin	(Wyler *et al.*, 1963)
Isobetanidin	(Wilcox *et al.*, 1965)
2-descarboxy-betanidin group	
2-descarboxy-betanidin	(Piattelli and Impellizzeri, 1970)
2-descarboxy-betanin	(Kobayashi *et al.*, 2001)
6'-*O*-malonyl-2-descarboxy-betanin	(Kobayashi *et al.*, 2001)
Glycosides	
5-*O*-glucosylation	
Betanin group	
Betanin: betanidin 5-*O*-glucoside	(Wyler and Dreiding, 1961)
Phyllocactin: 6'-*O*-malonyl-betanin	(Piattelli and Imperato, 1969)
2'-apiosyl-phyllocactin	(Kobayashi *et al.*, 2000)
2'-(5''-*O*-*E*-feruloylapiosyl)-betanin	(Schliemann *et al.*, 1996)
2'-(5''-*O*-*E*-feruloylapiosyl)-phyllocactin	(Schliemann *et al.*, 1996)
Hylocerenin: 6'-*O*-(3''-hydroxy-3''-methyl)-betanin	(Wybraniec *et al.*, 2001)
Amaranthin group	
Betanidin 5-*O*-(glucuronide)glucoside-amaranthin	(Sciuto *et al.*, 1974)
Iresinin I : hydroxymethylglutaryl-amaranthin	(Cai *et al.*, 2001a)
Celosianin I, II : coumaroyl and feruloyl-amaranthin	(Cai *et al.*, 2001a)
6-*O*-glucosylation	
Gomphrenin group	
Gomphrenin I: betanidin 6-*O*-glucoside	(Piattelli and Minale, 1964)
Gomphrenin II: coumaroyl derivative of Gomphrenin I	(Heuer *et al.*, 1992)
Gomphrenin III: feruloyl derivative of Gomphrenin I	(Heuer *et al.*, 1992)
Betanidin 6-*O*-(6-*O*-hydroxycinnamoyl)-β-sophoroside derivatives	(Heuer *et al.*, 1994)

Table 2. A selection of betacyanin representative and recent structures.

2.2 The betaxanthins

The yellow betaxanthins are immonium conjugates of betalamic acid with an amine or an amino acid. All proteinogenic amino acids and any of the 220 known non-protein amino acids found in plants can contribute to the betaxanthins constitution (Trezzini and Zryd, 1991a). Therefore numerous betaxanthins can be found in plant petals, but only a few have been characterized because most of them are present only in trace amounts in plant tissues. The classification of betaxanthins distinguishes the amino acid-derived compounds from the amine-derived conjugates (see Table 3). The first isolated structure was indicaxanthin (Fig. 2) from the fruit of Cactus pear (*Opuntia ficus-indica*) (Piattelli *et al.*, 1964). Numerous new structures have been identified during the last fifteen years. In *Portulaca grandiflora*, two new pigments, Portulacaxanthin II and III were characterized (Trezzini and Zryd, 1991b). Whereas several pigments have been identified in *P. grandiflora* flowers (see Table 3), little was known in the betaxanthin pattern of the Amaranthaceae. Recently, this lack of knowledge was filled by the characterization of a tryptophan-derived betaxanthin and of the first methylated betaxanthin, the 3-methoxytyramine-betaxanthin, in *Celosia argentea* species (Schliemann *et al.*, 2001). The latter compound seems to be methylated already at the catecholic stage rather than at the betaxanthin stage.

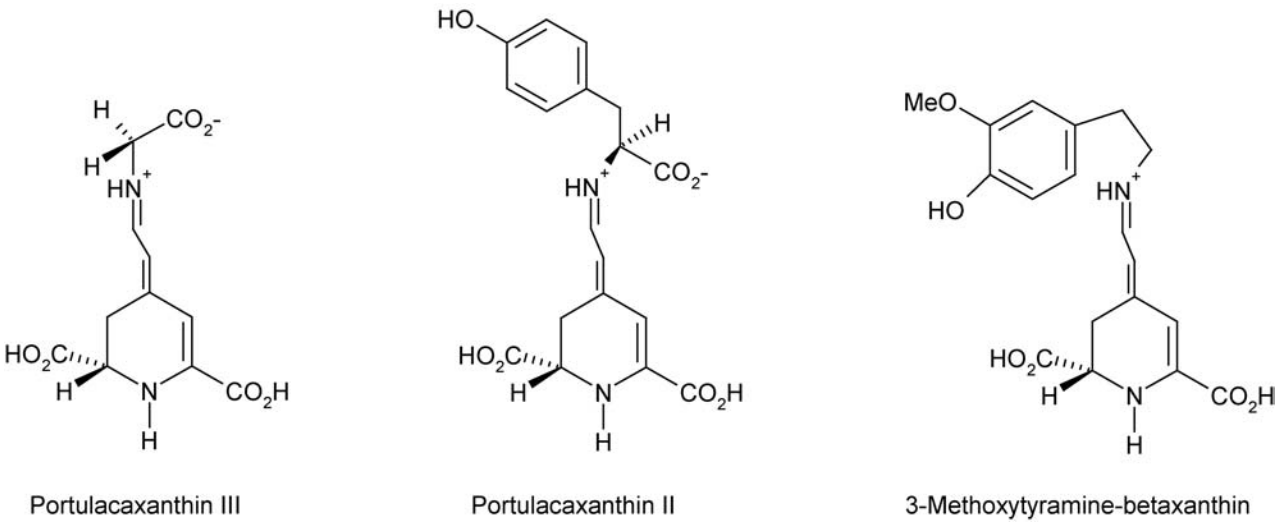


Figure 5. Chemical structure of some betaxanthins. Portulacaxanthin III (glycin-betaxanthin), Portulacaxanthin II [(*S*)-tyrosine-betaxanthin] and 3-methoxytyramine-betaxanthin (derived from dopamine-betaxanthin).

Recently, seven new pigments from cactus pear were characterized (Stintzing *et al.*, 2002b), six of which were new in plants. That brings the question of the choice of the methodology to identify betaxanthins often present in trace amount. According to Stintzing and co-workers, enrichment of the plant extract in betaxanthins, after separating them from sugars and pectins, gives a more detailed profile close to the "in vivo" chemical diversity.

In *Amanita* fungi, several betaxanthin compounds have been identified, including seven that are orange musca-aurins (λmax 480 nm) (Döpp and Musso, 1973b, a; Trezzini and Zryd, 1991a). No betacyanin pigment is known, but other red and yellow betalamic acid derived compounds, called, respectively, muscapurpurin (λmax 540 nm) and muscaflavin (λmax 420 nm), are present (Terradas and Wyler, 1991a; Mueller *et al.*, 1997b).

Amino acid-derived conjugates	Amino acid / amine	References
Dopaxanthin	DOPA	(Impellizzeri *et al.*, 1973)
Indicaxanthin	proline	(Piattelli *et al.*, 1964)
Miraxanthin I	methionine sulfoxide	(Piattelli *et al.*, 1965a)
Miraxanthin II	aspartic acid	(Piattelli *et al.*, 1965a)
Portulacaxanthin I	hydroxyproline	(Piattelli *et al.*, 1965b)
Portulacaxanthin II	tyrosine	(Trezzini and Zryd, 1991b)
Portulacaxanthin III	glycine	(Trezzini and Zryd, 1991b)
Vulgaxanthin I	glutamine	(Piattelli *et al.*, 1965c)
Vulgaxanthin II	glutamic acid	(Piattelli *et al.*, 1965c)
Tryptophan-betaxanthin	tryptophan	(Schliemann *et al.*, 2001)
Amine-derived conjugates		
Miraxanthin III	tyramine	(Piattelli *et al.*, 1965a)
Miraxanthin V	dopamine	(Piattelli *et al.*, 1965a)
3-Methoxytyramine-betaxanthin	methoxytyramine	(Schliemann *et al.*, 2001)
Humilixanthin	hydroxynorvaline	(Strack *et al.*, 1987)
Miraxanthin I	methionine sulfoxide	(Schliemann *et al.*, 2001)

Table 3. A selection of betaxanthin representative structures distributed in two groups according to the type of molecule conjugated with betalamic acid (amines or amino acids).

3. Betalain biosynthesis pathway and its regulation

In comparison with the anthocyanin biosynthetic pathway, the betalain pathway is rather simple, requiring only a few steps for either betaxanthin or betacyanin synthesis (see Fig. 6). Betalains are derivatives of tyrosine, which usually distinguishes them from phenylalanine-derived anthocyanins (Miller *et al.*, 1968). Anthocyanins can be derived from tyrosine in maize, because the phenylalanine ammonia-lyase has also a tyrosine ammonia-lyase activity (Rösler *et al*., 1997). Betacyanins synthesis requires a minimum of two tyrosines as precursor, whereas one is sufficient for betaxanthins. A feeding experiment with ^{14}C-radiolabeled tyrosine demonstrated that the entire C_6C_3-skeleton of this amino acid is incorporated, respectively, into betalamic acid and cyclo-DOPA molecules (Liebisch *et al.*, 1969).

Tyrosine is hydroxylated by a tyrosinase to give 3,4-dihydroxyphenylalanine (DOPA). Betalamic acid, the common chromophore of betalains, is formed by the successive enzymatic cleavage of the DOPA aromatic ring at the position 4,5-, giving the unstable *seco*-DOPA molecule (Fischer and Dreiding, 1972; Terradas and Wyler, 1991a), and the spontaneous recyclization of this intermediate (Schliemann *et al.*, 1998). A 4,5-DOPA-dioxygenase catalyses this extradiol cleavage (Girod and Zryd, 1991a; Terradas and Wyler, 1991a).

Spontaneous condensation of the chromophore with an amino acid or an amine forms the betaxanthins (Schliemann *et al.*, 1999), whereas betacyanin biosynthesis requires the presence of a cyclo-DOPA molecule. The formation of cyclo-DOPA results from the oxidation of DOPA into a dopaquinone intermediate by a polyphenol-oxidase, followed by a spontaneous cyclization of this intermediate. The following enzymatic step leading to the formation of betacyanins is still an open question. It is not clear if the betalamic acid molecule condenses with a cyclo-DOPA molecule or with its glucosylated form, depending on the different species studied. Heuer and Strack (1992) synthesised betanin from betanidin and UDP-glucose by a protein preparation obtained from cell cultures of *Dorotheanthus bellidiformis*. According to these results, the final step in betacyanin synthesis is the 5- or 6-*O*-glucosylation of betanidin or isobetanidin, preceded by a condensation step between cyclo-DOPA and betalamic acid. Kobayashi and co-workers obtained similar results with a short-term dopamine administration experiment in fodder beet seedlings (Kobayashi *et al.*, 2001). Thus, both groups concluded that the condensation step occurs between a non-glucosylated cyclo-DOPA and a betalamic acid molecule, followed by glucosylation and acylation steps.

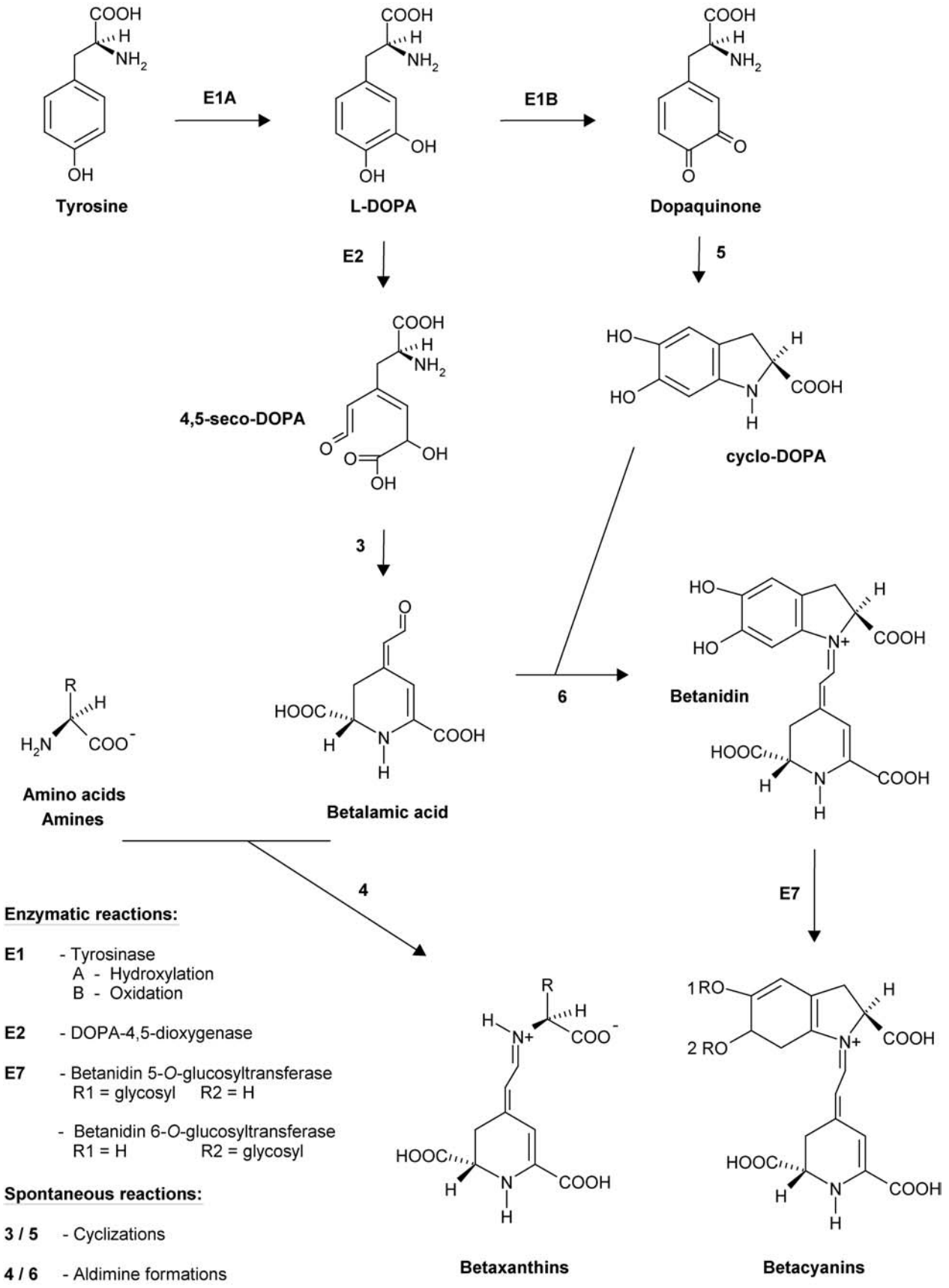


Figure 6. Current betacyanin and betaxanthin biosynthesis pathway (Strack *et al.*, 2003)

On the other hand, Bauer (2001) detected the presence of glucosylated cyclo-DOPA in beet cell cultures after feeding these cultures with radiolabeled DOPA molecule, showing that glucosylation was operative before the condensation step with betalamic acid. Mathematical modelization of the kinetics of incorporation of DOPA confirms this hypothesis. Nevertheless, Strack and colleagues (2003) concluded that the glucosylated form of cyclo-DOPA, also observed in their experiments at trace levels, results from the hydrolysis of betanin. On the other hand, the non-glycosylated form of cyclo-DOPA is not very stable and could be transformed into dopaquinone. The question is still open and will need further experimental evidence to be solved.

Betacyanins are frequently acylated via an ester linkage to the sugar moiety. Glycosylation and acylation of the betanidin nucleus allow for the diversity in betacyanin structures. Dopamine-derived betalains like the miraxanthine V or the 2-descarboxy-betanin require a supplementary decarboxylation step of the DOPA into dopamine (Dunkelbl *et al.*, 1972).

Numerous environmental and internal factors affect the betalain pathway. Synthesis of betacyanins or betaxanthins does not require light in most of the species examined till now, although light is a major factor controlling the quantities of pigments produced (Wohlpart and Mabry, 1968). Appearance of betacyanins on green calli from *Beta vulgaris* was observed when they were transferred from dim-light to high-light intensity (Girod and Zryd, 1987). A light induction experiment with amaranth seedlings (Giudici de Nicola *et al.*, 1975) showed that the photocontrol of amaranthin synthesis occurs at the level of the formation of the dihydropyridine portion of the molecule. In a further similar experiment, it was demonstrated that amaranthin photocontrol is operated by a simultaneous action of phytochrome (red) and cryptochrome (blue/UV) photoreceptors (Kochhar *et al.*, 1981). Recently, it was proven that UV-A light was sufficient to induce the production of betacyanins and flavonol glycosides in bladder cells of *Mesembryanthemum crystallinum* (Vogt *et al.*, 1999b). The same group refined this result by showing that the induction of betacyanins requires only a UV-light with a wavelength between 305-320 nm (Ibdah *et al.*, 2002). Light is known to affect the competition between the dopamine pathway leading to catecholamines (adrenaline) and the betalain biosynthesis pathway. In *P. grandiflora* callus, cathecholamines are only synthesised from dopamine in the dark (Endress *et al*., 1984).

In *Amaranthus caudatus*, betacyanins are synthesised in the dark in the presence of adenine-derived plant hormones (Bigot, 1968). Thus, a bioassay for the determination of cytokinins in

plant extracts was developed, based on the cytokinin-induced formation of betacyanins in the dark-grown *A. caudatus* seedlings in the presence of tyrosine (Biddington and Thomas, 1973). In fact, cytokinins can mimic light treatment and control the utilisation of tyrosine (Stobart and Kinsman, 1977). Recently, the molecular mode of cytokinin action was investigated with the help of the *Amaranthus* betacyanin assay (Romanov *et al.*, 2000). Studies on the inhibitory effects of actinomycin D, alpha-amanitin and cycloheximide indicated that rapid transcript induction and their translation were needed for amaranthin formation. The mean intervals between transcription, translation and pigment formation were estimated to be approximately two hours. In cultured cells of *B. vulgaris*, the ratio of auxin (2,4-D) to cytokinin (6-BAP) is an important factor in the regulation of betalain biosynthesis (Girod and Zryd, 1991b).

Regulatory mechanisms of betacyanin biosynthesis in suspension cultures of *Phytolacca americana* were investigated in relation to cell division activity (Sakuta *et al.*, 1994). By inhibiting the cell division, a reduced betacyanin accumulation and an inhibition of the incorporation of radioactivity from labelled tyrosine into betacyanin were observed. Thus, the conversion of tyrosine to DOPA is coupled with cell division activity. By using the DNA-methylation inhibitor 5-azacytidine, it was demonstrated that methylation plays a key role in the repression of genes encoding enzymes involved in betacyanin biosynthesis (Girod and Zryd, 1991b).

Betalain synthesis is frequently induced in injured tissues after a pathogen or an insect attack. Thus, sugar beets infested with Beet necrotic Yellow vein virus contain higher levels of betacyanins than healthy beets (Steddom *et al.*, 2003). The ratio of betacyanins to chlorophyll in the leaves was clearly increased in symptomatic beets, thus allowing the detection of the development of the rhizomania disease associated with this virus attack. Wounding and infiltration with *Pseudomonas syringae* or *Agrobacterium tumefaciens* also induce the synthesis of betacyanins in red beet leaves (Sepúlveda-Jiménez *G.*, personal communication). The pigment accumulation is restricted to wounded or infiltrated sites. A gene encoding a glucosyltransferase from *B. vulgaris* (GTBv) was identified and its expression was induced by infiltration with *A. tumefaciens, P. syringae* and wounding. These results were confirmed by a transient expression of an antisense construct of the GTBv gene leading to the reduction of betanin and transcript accumulation, suggesting a role of this gene in the synthesis of betanin.

4. Recent advances in betalain genetics and enzymology

In the absence of the isolation of the main betalain genes in higher plants, several enzyme activities have been described from enzyme preparation:

A decarboxylase activity, transforming DOPA into dopamine, was identified in protein extracts from red beet, thus supporting the synthesis of dopamine-derived betacyanins as observed previously (Terradas, 1989).

A tyrosinase activity was highlighted and partially purified in callus cultures from *P. grandiflora* and red beet (Steiner *et al.*, 1996; Steiner *et al.*, 1999). This bifunctional enzyme manages both the hydroxylation of tyrosine into DOPA and the oxidation of DOPA into dopaquinone. Recently, a tyrosine-hydroxylase activity from betacyanin producing callus cultures from *P. grandiflora* was separated from a polyphenol-oxidase activity (Yamamoto *et al.*, 2001). The tyrosine-hydroxylase fraction catalysed the formation of DOPA from tyrosine and was activated by Fe^{2+} and Mn^{2+}, and inhibited by metal chelating agents. The isolation of a tyrosine hydroxylase activity is in contradiction to the existence of a bifunctional tyrosinase / PPO (Steiner *et al.*, 1999). However, Yamamoto and colleagues (2001) neither purified this enzyme nor studied its kinetic activity and its substrate specificity. Therefore, it is difficult to take into account these partial results. Moreover, a bifunctional PPO was recently isolated from snapdragon flowers (*Antirrhinum majus*), catalysing the hydroxylation and oxidation of chalcones leading to aurones (Nakayama *et al.*, 2000, 2001; Sato *et al.*, 2001). Considering these biochemical studies, isolation of the gene coding for the tyrosinase / PPO enzyme is now necessary to further study this enzymatic step.

Enzymes involved in the final modification of betacyanins have been well studied. An hydroxycinnamic acid transferase was purified from *Chenopodium rubrum* cell cultures (Bokern *et al.*, 1992) and the glucosylation of betanidin by a betanidin 5-*O*-glucosyltransferase and a betanidin 6-*O*-glucosyltransferase was studied in *Dorotheanthus bellidiformis* (Heuer *et al.*, 1996). Recently cDNAs have been obtained for both enzymes in this species (Vogt *et al.*, 1999a; Vogt, 2002). Sequence comparison of these two enzymes shows that only 19% of the amino acids are conserved, suggesting a paraphyletic origin of these two glucosyltansferases (GT). Their kinetic properties are similar to other GTs involved in flavonoid biosynthesis (Vogt *et al.*, 1997).

No other betalain genes have been identified in higher plants. Till now, the 4,5-DOPA-dioxygenase activity has not been observed in plants. However 4,5-DOPA-dioxygenase (DODA) was purified from the mushroom *Amanita muscaria* and its kinetic activity evaluated (Girod and Zryd, 1991a). This dioxygenase was composed of four to seven identical subunits. This protein has some similarities with bacterial extradiol dioxygenases, not only due to the reaction they catalyse, but also to their homomultimeric structure and the implication of Fe(II) in its catalytic center (Girod, 1989; Girod and Zryd, 1991a). Nevertheless, no sequence homology at all was detected with existing dioxygenases. Such extradiol dioxygenases are not present in eukaryotes. *Amanita* DODA cDNA and genomic clones were isolated for the DOPA-dioxygenase and its functional activity characterized by recombinant production of the protein in *Escherichia coli* (Hinz *et al.*, 1997; Mueller *et al.*, 1997b). Mueller and co-workers observed for this fungal enzyme a unique double 4,5- and 2,3- aromatic ring-cleaving activity, allowing the simultaneous biosynthesis of, respectively, betalains and muscaflavin, a derivative of betalamic acid only present in fungal species (Barth *et al.*, 1979). Particle bombardment transformation of white *P. grandiflora* petals (unable to produce betalains) with *Amanita DODA* cDNA complemented betalain biosynthetic pathway in plants (Mueller *et al.*, 1997a). Analysis of the synthesised pigments demonstrated that the fungal DOPA-dioxygenase conserved its particularity of producing muscaflavin in the plant cell context, confirming the difference in the specificity of the fungal and the plant enzyme.

5. Portulaca grandiflora as a model for studying betalain genetics

Red beet is the most common betalain-containing crop, producing betanin in large amount in its hypocotyl. Nevertheless, this crop is not a good system for genetic analysis of betalain biosynthesis due to crossing and cultivation difficulties. At the beginning of the last century, betalains were already studied at the phenotypic level on the ornamental purslane plant *Portulaca grandiflora* Hook (*Pg*) which presents a large spectrum of petal colourations ranging from yellow through orange and red to violet (Yashi, 1920; Ikeno, 1921; Adachi, 1972). More recently, an elaborated genetic model of betalain biosynthesis based on the crossing of pure *P. grandiflora* lines exhibiting different petal colours (see Fig. 7), was developed (Trezzini, 1990a). The resultant segregation patterns indicated that a minimum of three loci *C*, *R* and *I* were involved in petal pigmentation (see Table 4).

The first locus, called *C* for colouration, directs the extradiol cleavage of the DOPA precursor into *seco*-DOPA. According to the "three-gene model", this is the last key enzymatic step necessary for betaxanthin biosynthesis, whereas the cyclization of the *seco*-DOPA to betalamic acid should be a spontaneous reaction. Expression of this dominant locus in either the homozygous or the heterozygous state gives always a coloured phenotype. In the recessive state, no pigmentation was observed. The plant petals and the stamens filet are colourless, and the stem and the pistil are green-coloured. Locus *C* in plant should correspond to a DOPA-4,5-dioxygenase with a similar activity in the mushroom *Amanita muscaria* (Mueller *et al.*, 1996).

Locus *R*, responsible for the synthesis of violet betacyanins, is involved either in the transformation of DOPA to cyclo-DOPA or to its glucosylated form. Orange and red phenotypes are also observed in the presence of *R* expression, due to its incomplete dominance, resulting in lower amounts of synthesised betacyanins. The conjugation of *R* and *C* gene products is not under genetic control because no flower containing only cyclo-DOPA and betalamic acid has been obtained by crossing (Trezzini, 1990a).

The third locus, called *I* for inhibitor, is involved in the modulation of the betaxanthin colouration. *I* could prevent the conjugation of amino acid or imino residues (others than cyclo-DOPA) with

betalamic acid, thus decreasing the synthesis of betaxanthins. The chromophore could be formed in the cytoplasm, while the conjugation to give the final betaxanthins would take place spontaneously in the vacuole where betaxanthins are mainly stored (Trezzini, 1990b). This hypothesis postulates the existence of a specific carrier for betalamic acid, which could be inhibited by a protein, encoded by locus *I*. Nevertheless, in the presence of the expression of locus *C* yellow pigment was always produced, despite the expression of the inhibitor *I*. This could be explained by residual betaxanthin synthesis in the cytoplasm due to the presence of tiny amount of free amino acids in the cytoplasm. This hypothesis is confirmed by the analysis of the pigment content from pale-yellow *CCrrII P. grandiflora* plants. A high amount of betalamic acid is present, whereas traces of dopaxanthin and no miraxanthin V are present (Trezzini and Zryd, 1991a). More betacyanins are synthetised when *I* is expressed, as observed by comparison of betacyanin amounts in violet and red *Pg* petals. *R* and *I* genes are linked together (Trezzini, 1990b).

These results obtained in betalain genetics from the selection of pure *P. grandiflora* lines exhibiting different petal colours allow defining a strategy based on the difference in pigment expression in order to isolate genes *C*, *R* and *I* (Table 4). Finally this metabolic pathway seems to be strictly compartmented.

	White	***Clear Yellow***	***Deep Yellow***	***Red***	***Violet***
C	-	+	+	+	+
R	-	-	-	+	+
I	-	+	-	-	+

Table 4. “Three genes model”. Expression (+) or absence (-) of genes *C*, *R* and *I* explain the variety of colours observed in *Portulaca grandiflora* flowers.

Figure 7. *Portulaca grandiflora* white, clear yellow, deep yellow, red and violet phenotypes.

6. Aims of the present work

The purpose of this work is to fill the lack of knowledge concerning the genes responsible for the biosynthesis of betalains in plants, by isolating such genes in the model plant *Portulaca grandiflora*. Numerous homology and antibody screens to identify the plant DOPA-dioxygenase via the *Amanita* gene or enzyme failed. These results suggested a different phylogenetic origin and a completely independent evolution of the plant enzyme, but resulting in a similar 4,5-dioxygenase activity (Hinz, personal communication). The *Amanita* enzyme is still today a unique case without clear homologue either in fungi or plant kingdom. Moreover, the formation of muscaflavin due to a 2,3-ring cleavage, indicated that the catalytic activity is partially different, which should correspond to differences in their three-dimensional structure. The number of DOPA-dioxygenase molecules in a plant cell, estimated from the kinetic parameter of the fungal enzyme, is low (24000) (Girod, 1989). The direct isolation of a protein is impossible at this concentration, suggesting rather a gene-based approach. Therefore, in order to directly isolate the plant genes involved in betalain biosynthesis, a subtractive library strategy will be developed in the present work from violet and white petals of *P. grandiflora* pure lines obtained by (Trezzini, 1990a).

By isolating major betalain genes and their homologues in non-betalain crops, we will be in a better position to study the problem of the phylogenetic origin(s) of betalains. We will try to answer the question of the appearance of betalains in plants and if they have coexisted in parallel with anthocyanins. Finally, we hope to suggest new potential commercial applications based on betalain pigment properties.

II. RESULTS AND DISCUSSION

1. Isolation of betalain specific cDNAs by subtractive hybridisation in P. grandiflora

1.1 Introduction

The aim of this thesis project is to isolate genes responsible for betalain biosynthesis in *P. grandiflora* using a subtractive strategy. Numerous methods exist to isolate genes having an expression pattern related to a particular developmental or physiological stage or a specific tissue or cell type. Among them, subtractive hybridization is a powerful technique that enables the comparison of two populations of mRNA and the isolation of genes that are expressed in one population but not in the other. Although traditional subtractive hybridization methods have been successful in some cases, they require several rounds of hybridization and are not well suited for the identification of rare transcripts. The major drawback of subtractive cDNA libraries is that the original disproportion in concentrations of different types of transcripts is preserved, thus making the isolation of rare transcripts very difficult.

A new method for the generation of subtracted cDNA libraries that overcomes these problems was developed (Diatchenko *et al.*, 1996; Gurskaya *et al.*, 1996). This suppression subtractive hybridization (SSH) method is based on selective PCR amplification of differentially expressed sequences and combines normalization and subtraction in a single procedure. The normalization step equalizes the abundance of cDNAs within the target population and the subtraction step excludes the common sequences between the two populations of mRNA. This method is sold as a kit by Clontech under the trademark PCR-Select cDNA subtraction, and has already been used to isolate betalain-specific transcripts of yellow petals from *P. grandiflora* (Zaiko, 2000).

By subtractive hybridization of the mRNA populations of violet and white petals from *Pg* isogenic lines, we hoped to isolate genes specifically expressed in coloured petals, but absent from white petals. This strategy is based on the assumption that most betalain genes are transcriptionally regulated, as observed for the *Amanita muscaria DODA* gene (Hinz *et al.*, 1997). Apart from

genes directly involved in the betalain biosynthesis pathway, we expected to find a number of genes indirectly related to betalain accumulation such as transport proteins for the export of tyrosine or DOPA from plastids, or vacuolar transporters for the import of betalains in the vacuole. This method does not exclude the isolation of genes presenting the same expression pattern, but unrelated with betalain biosynthesis. However, the genetic background of these two differentially coloured sets of plants should be fairly homogeneous due to the numerous crossing performed to obtain pure lines (Trezzini, 1990a). Obtaining an overview of the number and type of genes co-regulated with betalain pigments will be highly interesting by itself.

1.2 Construction of the subtractive library

We encountered difficulties in obtaining sufficient quantities of high-quality total RNA from *Pg* petals due to the presence of huge amount of polysaccharides. Therefore, we decided to preamplify total RNA samples with the SMART PCR cDNA Synthesis kit (Clontech). This kit utilizes SMART technology from Clontech (Borson *et al.*, 1992) and long-distance PCR (Barnes, 1994) to provide high yields of full-length double-stranded cDNA. The specificity of the SMART technology allows a selective amplification of full-length cDNAs, leaving partially reverse transcripted cDNAs non-amplified. cDNAs generated with this method can be directly used for the PCR-Select subtraction that we will use thereafter.

Thus, total RNA populations from violet and white *Pg* bud-extracted immature petals were converted by reverse transcription and SMART PCR amplification technology into double-stranded cDNA. In the subtraction experiment, we refer to the violet cDNA that contains betalain differentially expressed transcripts as “tester”, and the reference white cDNA as “driver”. Amplified cDNAs correspond to a smear from 0.3 to 5 kb on the gel. After purification on a CHROMA-spin column (Pharmacia), cDNA samples were digested with *Rsa*I to obtain shorter blunt-end cDNA molecules. Two different subtraction experiments were then performed. One direct subtraction, where violet cDNA was used as tester and white cDNA as driver, and one reverse subtraction with inversion of the roles of the two cDNAs. The overview of the procedure is shown in Figure 8.

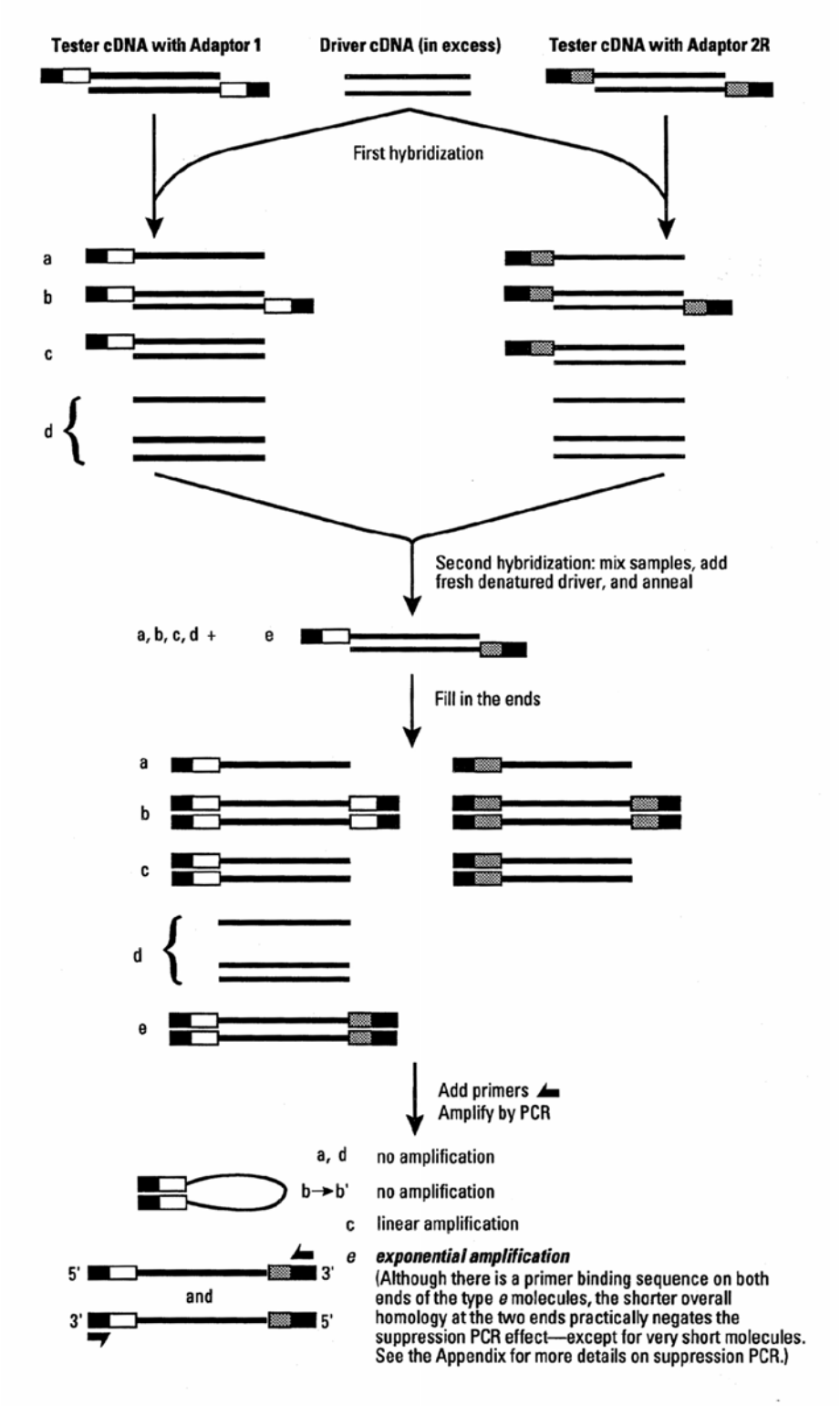


Figure 8. Overview of the different steps of the subtraction (figure copied from the PCR-Select subtraction guide, Clontech). White boxes represent the outer part of the Adaptator 1 and 2R longer strands, and corresponding PCR primer 1 sequence. Black boxes indicate the inner part of the Adaptator 1 and the corresponding nested primer 1 sequence. Grey boxes represent the inner part of the Adaptator 2R and the corresponding Nested primer 2 sequence.

Two tester populations were created by adding different adaptators; the driver had to remain with naked ends. Two hybridizations were then performed. In the first one, an excess of driver was added to each sample of tester and the kinetics lead to equalization and enrichment of differentially expressed sequences. A second hybridization was made by mixing the two primary hybridization samples together without denaturing and adding fresh denatured driver, finally followed by annealing. Using suppression PCR, only differentially expressed sequences were amplified exponentially.

During the second nested PCR amplification, the background is reduced and violet-specific differentially expressed sequences further enriched. Finally, a comparison of the amplification results of subtracted and non-subtracted cDNA samples was performed on an agarose gel to verify the quality of the subtraction (see Fig. 9). We observed several individual bands in the violet and white subtracted cDNA samples, whereas a strong smear was present in the non-subtracted samples confirming that the subtraction procedure was correct.

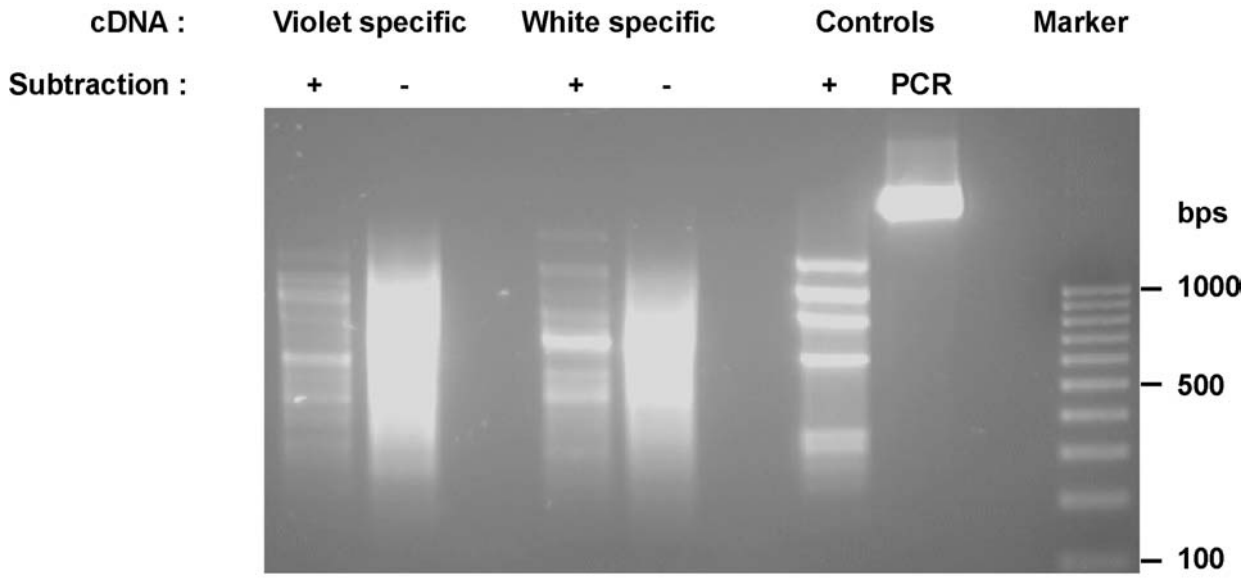


Figure 9. Evaluation of the cDNA subtraction efficiency after the second PCR amplification.

Previous isolation of three betalain-specific cDNAs called *L.6*, *L.13* and *V.33* (Zaiko, 2000) enabled us to monitor their presence from the very start of the violet-specific cDNA library construction, in order to evaluate its quality. Beyond the two-strand cDNA synthesis step, we detected only *V.33* cDNA. The two other cDNAs were missing, perhaps lost during the reverse-transcription step. Quantification of *V.33* cDNA before and after the subtraction shows a 20-fold

enrichment corresponding to the minimal expectation. Finally, the A-tailed amplified cDNA enriched in violet-specific sequences was directly ligated in the T-tailed pGemT Easy cloning vector (Promega), and XL10-Gold ultracompetent *E. coli* (Stratagene, La Jolla) were transformed to construct a violet-specific cDNA subtractive library.

1.3 Identification of betalain specific clones by subtractive hybridization

Due to the great number (946) of independent colonies present in the violet subtracted library, the screening was made by successive stages, starting initially with a coarse sorting to limit the number of candidates subsisting for the final Northern analysis in different tissues from *Portulaca grandiflora*. First, each clone was amplified by PCR with T7-Sp6 primers present on the pGemT Easy cloning vector (Promega). The average insert size was 550 bp (Fig. 10) and around 10% of the clones were eliminated because of the repetitive unspecificity of their amplification.

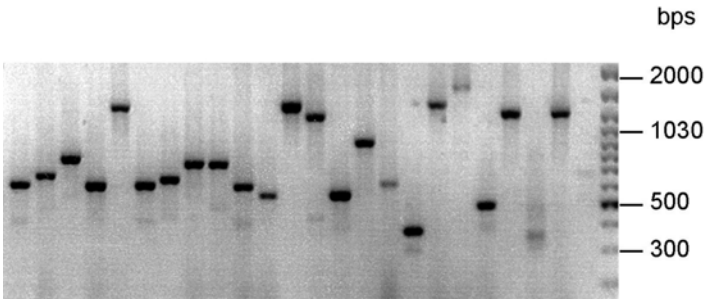


Figure 10. Inserts amplified from violet subtracted cDNA library. Molecular weight marker: 100 bp DNA ladder.

Four clones of different size were picked randomly for sequencing in order to verify the diversity and the quality of the library. All four cDNA fragments were different from the usual abundant genes such as those encoding the ribosomal proteins or the Rubisco. The first, *A.3*, to which no function has been assigned yet, is a 550-bps fragment, identical to *V.33* betalain specific cDNA (Zaiko, 2000). The second, *A.13*, has 70% of homology with an EST coding for a hypothetical protein from *Arabidopsis thaliana* (Genbank entry AAD18143) and 35% of similarity with the mouse and human *DMR-N9* genes containing a WD repeat domain. Thus, it has no clear function. Similarly, we observed for *A.8* and *A15* cDNAs (see Annex. 5) clear homologies with plant genes encoding different hypothetical proteins, but without known function.

1.3.1 Preliminary screening of the violet subtracted library by slot blot hybridization

Slot blot analysis facilitates the simultaneous evaluation of several hundreds clones by differential hybridization. This technique consists of a double blotting on a nitrocellulose membrane of a small amount (10 pg) of the denatured PCR-product obtained for each independent clone. The blotting was made under vacuum in the grid of the slot-blotter and the DNA covalently fixed to the membrane under UV light exposition. Thereafter, the membrane is cut in its middle to give two identical half-membranes. Differential expression of the PCR products in *Pg* white and violet immature petals was performed by hybridization of a half-membrane with the ^{32}P-radiolabeled violet-specific probe corresponding to the violet subtracted cDNA before cloning, whereas the second half-membrane is hybridized in parallel with the ^{32}P-radiolabeled white specific probe. After an autoradiography exposition, the intensity of the respective signals on each half-membrane are compared together. A strong difference in intensity is considered as a positive results (see Fig. 11). *V.33* positive control, plus a negative control consisting in the loading buffer, the violet-specific probe and the white-specific probe are also deposited twice, on each half-membrane, to ensure the quality of the analysis. In the violet immature petals, 226 clones (26% of the total colonies) were more strongly expressed. They will be called positives candidates. Only a few clones were repressed. Hybridization of the violet positives candidates with *V.33* full-length cDNA and some other previously identified abundant clones (Zaiko, 2000), left for sequencing analysis 105 candidates overexpressed in violet petals.

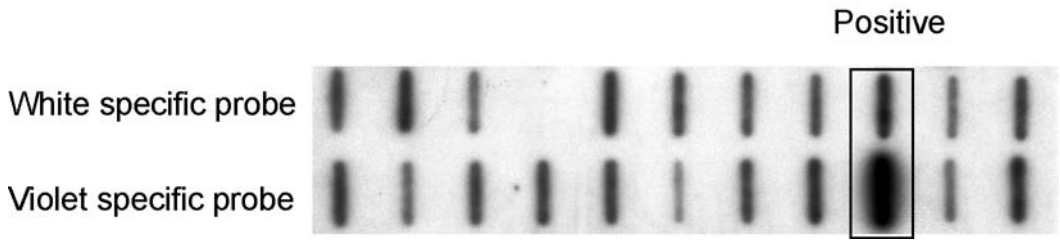


Figure 11. Slot blot of subtractive cDNA from *Pg* petals hybridized with violet specific and white specific probes. Selected positive clones are indicated in boxes.

1.3.2 Sequencing and expression analysis of candidates identified by differential screening

Sequences comparison of the 105 positives cDNA fragments among each other, and with EST and protein databases, revealed the existence of only 26 different cDNAs due to redundancy and *Rsa*I cutting of the cDNAs in different small pieces. Most of them have no homologue in the databases or are only hypothetical proteins without function. Nevertheless, defined ORFs were identified for most of these cDNAs. We identified cDNA fragments having a clear similarity with the VIP3 protein from *Arabidopsis thaliana* (*N.22*), others with acyl-Coenzyme A oxidases (*A.16*) or with hydrolase-like protein (*H.13*) from *A. thaliana* (see Annex. 5 for sequences). We still found some sequences homologous to the *V.33* clone. Thus, this still non-identified betalain-specific cDNA corresponds to 5% of the total colonies (20% of the positives) present in the violet subtracted library, as also observed in the yellow specific subtracted library (Zaiko, personal communication). About 14% of the positive candidates correspond to the *A.13* clone, encoding a protein homologous to human and mouse DMR-N9 proteins. As no candidate encoding a dioxygenase (*C* gene), a tyrosine hydroxylase or a vacuolar transporter (*I* gene) was identified in the gene or protein databases, we decided to compare the expression levels of the transcripts corresponding to the remaining hypothetical violet-specific cDNAs in coloured and white petals from *Pg*. Northern blot analysis showed that only *A.16* and *P.34* transcripts present a strong expression in betalain-coloured petals and are absent from white petals, as expected from a betalain-specific gene (Fig. 12). *A.16* cDNA corresponds to the 3' end of an ORF, showing 55% homology with the acyl-Coenzyme A oxidase 3 (ACX3) from *A. thaliana* (*At*). This protein is mainly involved in the degradation of fatty acids. The *P.34* partial cDNA has no homology in the databases and no clear ORF.

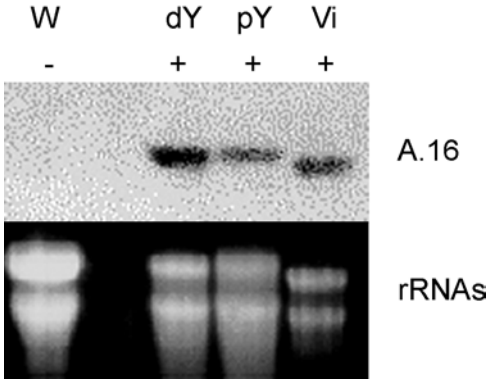


Figure 12. Northern blot analysis of *A.16* partial cDNA in white and betalain-coloured petals from *Pg*. Total RNA from white (W), deep-yellow (dY), pale-yellow (pY) and violet (Vi) petals.

Further Northern blot analysis of different pigmented and green tissues from *P. grandiflora* were performed to assess the betalain-specific expression of *A.16* and *P.34* genes. Betalains are strongly present in petals and stamens and slightly in the stem epidermis. It was demonstrated by HPLC analysis that betalain pigments and their precursors accumulate during petal development (Trezzini, 1990a). Nevertheless, the level of transcripts directly involved in the betalain biosynthesis pathway decreases beyond the synthesis of betalains in petals. Thus, the comparison of the expression level of hypothetical betalain-specific genes in pigmented immature and mature petals extracted from buds can provide evidence of their implication in this pathway. Extensive Northern blot analysis as described above confirmed *A.16* and *P.34* clones as betalain-specific genes (see Fig. 13). Both genes were expressed in petals and stems, with a decrease in mature petals. As expected, the level of expression of *A.16* and *P.34* in violet coloured stems is less pronounced than in petals. No signal was detected in white petals, green stems or leave samples.

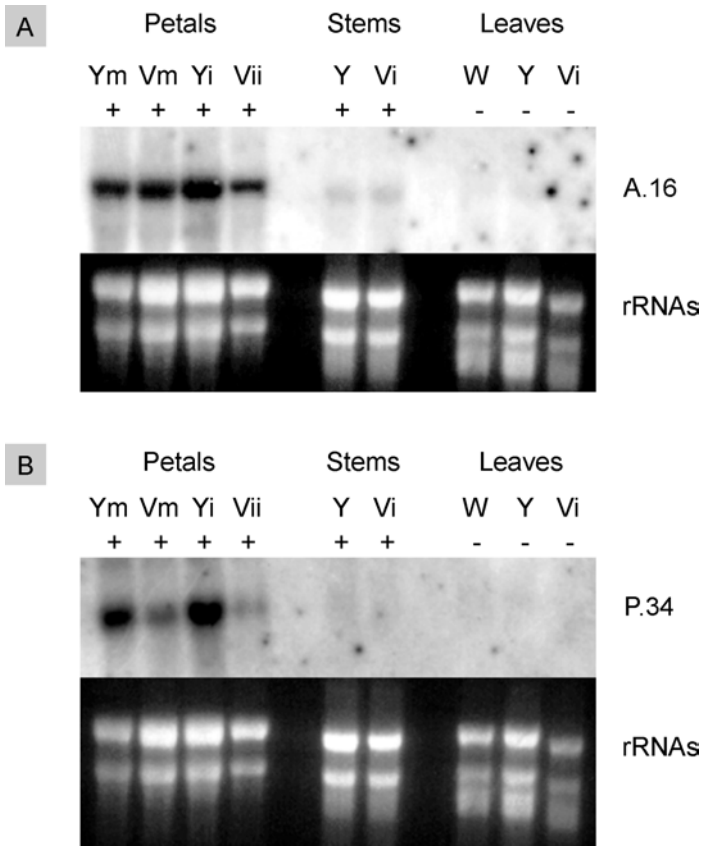


Figure 13. (A) Northern blot analysis of *A.16* partial cDNA in *Portulaca grandiflora*. Presence or absence of betalain pigments in plant tissue is indicated by (+) or (-) signs. Expression at different stages of bud development (with Yi yellow immature, Ym yellow mature, Vii violet immature and Vim violet mature petals), and in stems and leaves from different white (W), yellow (Y) and violet (Vi) *Pg* genotypes. (B) The same for *P.34* partial cDNA clone.

1.4 Isolation of full-length cDNAs by a RACE amplification strategy

Full-length *A.16* and *P.34* cDNAs had to be isolated in order to perform further functional analyses of the proteins corresponding to these two betalain-specific cDNAs. *A.16* corresponds to the 3'-end fragment (515 bps) of an acyl-CoA oxidase and *P.34* has no clear ORF, suggesting its location in an untranslated region. Missing parts of *A.16* and *P.34* partial cDNAs were elongated by a modified "rapid amplification of cDNA ends" method (RACE) (Chenchik *et al.*, 1996). In this approach a specially designed double-strand (ds) adaptator is ligated to both ends of a library of ds cDNA by T4 DNA ligase (see Fig. 14). This adaptator-ligated ds cDNA is then used to selectively amplify 5'- or 3'- cDNA fragments by PCR with a combination of 26 to 30 nucleotide-long gene-specific primers (GSP1, GSP2) and adaptator-specific primers (AP1, AP2). Full-length cDNA will be further obtained by a new amplification with a specific primer designed on the newly acquired sequence information from the previous missing end (GSPStart), combined with another specific primer chosen on the opposite ends (GSPStop) as shown in Fig. 14.

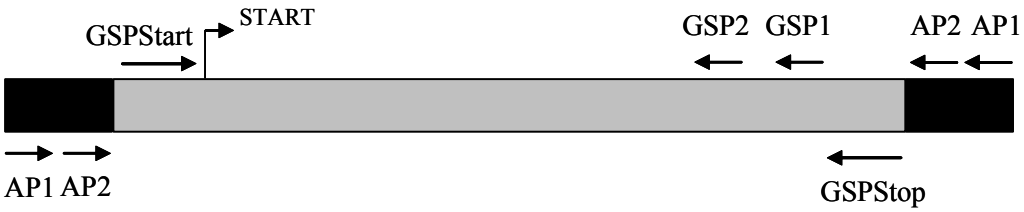


Figure 14. Schematic representation of the amplification of partial cDNA by the RACE method. Black boxes represent adaptators ligated to a full-length cDNA (grey). Adaptators contain the adaptator-specific primers (AP1, AP2). GSP indicated the pair of gene-specific primers.

In this context, *A.16* was elongated from its 3' end to the direction of the start codon by nested PCR with primers A161 and A162 combined with AP1/AP2 primers present on the ligated adaptators. We obtained a 2.3 Kb amplification product, the sequence of which is homologous to the beginning of the *At ACX3* gene. A16Start and A16Stop primers were designed respectively on the 5' UTR and 3' UTR regions to amplify *A.16* full-length cDNA modified on its non-coding extremities for cloning facilities. We tried to amplify *P.34* cDNA with the same procedure, but no specific PCR product was obtained. An equivalent strategy was undertaken on genomic DNA from violet *Pg* plants without success. Due to the absence of a clear ORF, *P.34* cDNA was left dropped from further investigations.

1.5 Functional analysis of *A.16* betalain-specific cDNA

1.5.1 Characterization of *A.16* cDNA

The in silico translation of *A.16* full-length cDNA gives an ORF encoding a 689 amino acid-long polypeptide with a theoretical molecular weight of 76 kDa and an isoelectric point (pI) of 6 (see Fig. 16) The A.16 translated protein is homologous to *At* acyl-Coenzyme A Oxidase 3 (ACX3). The hydropathy plots of both proteins are rather similar as shown in Fig. 15, but *At* ACX3 (NP_172120) is shorter in size (675 amino-acids) and has a quite different pI value of 7.6. The A.16 amino acid sequence shows 50% of identity and 68% of similarity with *At* ACX3 (see Fig. 17). Alignment is globally good with highly conserved area.

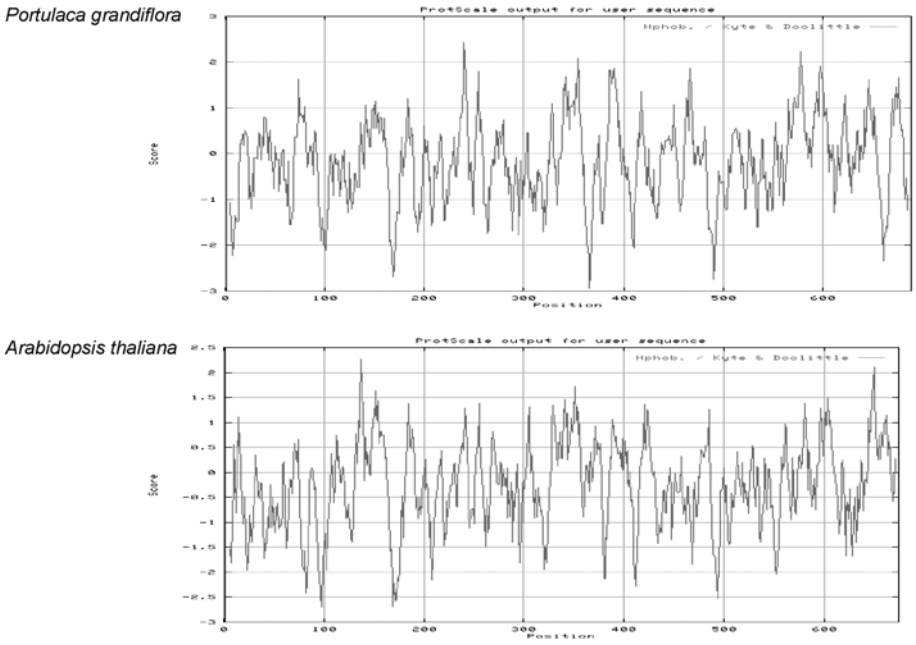


Figure 15. Hydropathy plots of *At* ACX3 and A.16 (Kyte and Doolittle, 1982). Note that scales are different.

The A.16 protein is also recognized as an acyl-CoA oxidase by the CDD domain search tool (Fig. 18). By comparison with the *At* ACX3 conserved domains, we can see a partial conservation of the ACOX domain (pfam01756) in *Pg* A.16 sequence, with its C-terminal end missing. The conserved acyl-CoA dehydrogenase middle domain (pfam02770), containing a β-barrel fold, is entirely present in the A.16 sequence, whereas the acyl-CoA C-terminal domain, usually accompanying the central domain, is present in *At* ACX3 sequence, but absent from *Pg* A.16 sequence.

```
-170  gaagacgacagaagggtacggctgcgagaagacgacagaagggtacggctgcgagaagacgacagaaggggactatcaatcagt   -85
 -85  attcagtagtcactctctgattgtctctcattctctctcttgagtactgattcctcgatctctggctcattcgtttcatcgatca    -1

   1  ATGGCGTCCTCTGATGATGGATCTGATCATCAAGCCTCCGCCAGAAGTGGAATTCTAAACCTACATCTCGACCCCTCACATTCAT    85
      M  A  S  S  D  D  G  S  D  H  Q  A  S  A  R  S  G  I  L  N  L  H  L  D  P  S  H  S

  86  CGTCATTGTCAGTACCTTCTGACATCTCCCCGTCACTATGTCTGTCCAAAAGCGAGGCAGCTTCCTCCTTCAACATACAACAAAT   170
      S  S  L  S  V  P  S  D  I  S  P  S  L  C  L  S  K  S  E  A  A  S  S  F  N  I  Q  Q  M

 171  GCGAGCTCTGTTGGATCCACACAACAATATTCAAGATCTCTACTCTCTGTTAGAGATCATGTGTGAAAGCCCTCTTTTTGGCCCT   255
        R  A  L  L  D  P  H  N  N  I  Q  D  L  Y  S  L  L  E  I  M  C  E  S  P  L  F  G  P

 256  AAATTGCACGCTCCAGCATCCCTATTTAGAGAGGAGGAAAGGGATCTCACTATGAGAAGAATCCAGTATCTGGTTGATCAAGGTG   340
        K  L  H  A  P  A  S  L  F  R  E  E  E  R  D  L  T  M  R  R  I  Q  Y  L  V  D  Q  G

 341  GTTTCAAGGGTTGCCCCACTGCTCCCGCTGATGAACAATCTCTGCTCAGGAGGTTGACTTACAATGAGCTTCTCGGTGTTTCTGA   425
      G  F  K  G  C  P  T  A  P  A  D  E  Q  S  L  L  R  R  L  T  Y  N  E  L  L  G  V  S  D

 426  TCACGCCCTTCCTATTCAACTCTCCCTTCATTTCTCCCTTTGGGGTGGTGCAATCAAGTACTTGGGGACCAAGCGTCATCATGAC   510
        H  A  L  P  I  Q  L  S  L  H  F  S  L  W  G  G  A  I  K  Y  L  G  T  K  R  H  H  D

 511  AAATGGCTGAAAGTTACTGAAGACTTTGGGATAAAAGGTTGCTTCTCTTTGACAGAGATGGGGCATCAAAACAATGTAAAAGGAA   595
        K  W  L  K  V  T  E  D  F  G  I  K  G  C  F  S  L  T  E  M  G  H  Q  N  N  V  K  G

 596  TTGAAACAGTTGCCACATATGACGCTCAAACTAGAGAATTTGTTATCAACACTCCTTGTGAATCGGCCCAAAAGCACTGGATTGG   680
      I  E  T  V  A  T  Y  D  A  Q  T  R  E  F  V  I  N  T  P  C  E  S  A  Q  K  H  W  I  G

 681  TGGGGCCCTTAAACATGCAACGCATACCATTGTTTTCGCGCAACTCATCATTGAAGGGCAAAATCAAGGGGTTCATGCATTTGTA   765
        G  A  L  K  H  A  T  H  T  I  V  F  A  Q  L  I  I  E  G  Q  N  Q  G  V  H  A  F  V

 766  GTTCCAGTCAGAGATTCTAAGGGAAATCTCTATCCAAATATCCACATGGCTGACTGTGGGATCAAAACGGGGATAACCGGGGTGG   850
        V  P  V  R  D  S  K  G  N  L  Y  P  N  I  H  M  A  D  C  G  I  K  T  G  I  T  G  V

 851  ATGATGGGCGAGCTTGGTTTGATAACGTCCGCATACCTAGGGAAAACCTATTGAATTCTGCTGCTGACGTTAATCCTGATGGTCA   935
      D  D  G  R  A  W  F  D  N  V  R  I  P  R  E  N  L  L  N  S  A  A  D  V  N  P  D  G  Q

 936  ATATGTCAGTTCCGTAGAATACCCTAACCAGAGGTTTGCAACGCTTGTTGACCCATTGACTTCAGAACGAGTCAATATAACAGTT  1020
        Y  V  S  S  V  E  Y  P  N  Q  R  F  A  T  L  V  D  P  L  T  S  E  R  V  N  I  T  V

1021  AGTGCTATCTACTTATCTCAGATTGATTTAGCTATTGGCGTAAAATTTGCTGTATCAAACTGCGATAAACCAAAATATTCAGACC  1105
        S  A  I  Y  L  S  Q  I  D  L  A  I  G  V  K  F  A  V  S  N  C  D  K  P  K  Y  S  D

1106  AAAGAGGGTTGTTGATCAACAGCCCAAGCTATCGGCGCTACATAATGCCCCTGCTTGCAACAGTATATGCAGTGAGTTTTGCTGG  1190
      Q  R  G  L  L  I  N  S  P  S  Y  R  R  Y  I  M  P  L  L  A  T  V  Y  A  V  S  F  A  G

1191  GAACCACTTGAAATCAATATATGTTAAGAGAACACCTGAAACAAATAAGCTTATCCAAATTCTTTCCAATGGGTTAAACGCTAAG  1275
        N  H  L  K  S  I  Y  V  K  R  T  P  E  T  N  K  L  I  Q  I  L  S  N  G  L  N  A  K

1276  TTGAATTGGCTTAACTACCAGACATTTGAGGACAATGTCATGAGAATCTTGGCTAGAAAGACGGTTGCAAATTTGCTAGGTCATC  1360
        L  N  W  L  N  Y  Q  T  F  E  D  N  V  M  R  I  L  A  R  K  T  V  A  N  L  L  G  H

1361  TCAAAGCCCAGGAAGATGTGCAATCAGTTTTCTTGGGGGTCAGATATGTCCTCTGGCAACAGGTGAGCAAAGCATTGGTTAGGGA  1445
      L  K  A  Q  E  D  V  Q  S  V  F  L  G  V  R  Y  V  L  W  Q  Q  V  S  K  A  L  V  R  E

1446  GTATCTGACTGCTCAAAATCAAAACAGGCCATTTAAAGAAGGATTGTGGTTAGAGCACATGAACGAGTCAACTCCTGTAATACCA  1530
        Y  L  T  A  Q  N  Q  N  R  P  F  K  E  G  L  W  L  E  H  M  N  E  S  T  P  V  I  P

1531  TCTCAGGTGACAAGCTCTATTTTAAGGAGCACCAAGTTTCAGACTGATGTATTCTTTTTGAGGGAGAGAGATATGCTGAATCGTC  1615
        S  Q  V  T  S  S  I  L  R  S  T  K  F  Q  T  D  V  F  F  L  R  E  R  D  M  L  N  R

1616  TTGTAAACGAAGTAGCTAAGCATCAAAATCTGGGAAGTTTAGAGCATGCATTTGCTCAGTGCCAGGAGCTTGGCAAGGAATTAGC  1700
      L  V  N  E  V  A  K  H  Q  N  L  G  S  L  E  H  A  F  A  Q  C  Q  E  L  G  K  E  L  A

1701  AATAGCTTTCACGGACAAGGCAATTTTGTGTGCTTTTATTGAAGCTGAAGCCAAGGTTCCTGTTGGTCCTCTAAAGGATGTGTTA  1785
        I  A  F  T  D  K  A  I  L  C  A  F  I  E  A  E  A  K  V  P  V  G  P  L  K  D  V  L

1786  GCTTTGTTGCGGGCATTGTACGCGCTAAAGATGATAGAAGAGGATGTTGGATTCACTCAGAATGGCTACATTTCTGCTGGTATTG  1870
        A  L  L  R  A  L  Y  A  L  K  M  I  E  E  D  V  G  F  T  Q  N  G  Y  I  S  A  G  I

1871  CTGAAGCTGCTGAAGAAGAAGTGACATCGCTCAGTGCTGAGCTCCATCCCCATGTGGTGGAACTGGTAGACTGTTTTGGGATCCC  1955
      A  E  A  A  E  E  E  V  T  S  L  S  A  E  L  H  P  H  V  V  E  L  V  D  C  F  G  I  P

1956  GGACTCATTTTTCAGCCCGGATGGCTACCGTAGATGGGATCTCCTCTACTCTCTTCATGTCGTGGCTCCTGGTTGTGCCGGCTTC  2040
        D  S  F  F  S  P  D  G  Y  R  R  W  D  L  L  Y  S  L  H  V  V  A  P  G  C  A  G  F

2041  CTTCATTCTTATGACCACCATCTTGACTGATGCTCCGAAAATGTCATTTGGAGTTGGATGATTTTCTATTTTTCAGATGCTAGTT  2125
        L  H  S  Y  D  H  H  L  D  *

2126  gagtagtagtctgttctgtttagtcagtgggacatgactctttgcagcaaagtttgtttcctgtctgtaagacttggctgattga  2210
```

Figure 16. Full-length *A.16* cDNA sequence. The coding region is in capital letters. The start codon is in bold. Primers A16Start and A16Stop are highlighted in grey, primer A161is indicated in bold and A162 is underlined.

```
AtACX3   2   SDNRALRRAHVLANHI--LQSNPPSSNPSLSRELCLQYSPPELNESYGFDVKEMRKLLDG 59
             SD++A  R+ +L  H+    S   S    +S  LCL  S      +  F++++MR LLD
PgA.16   8   SDHQASARSGILNLHLDPSHSSSLSVPSDISPSLCLSKS----EAASSFNIQQMRALLDP 63

AtACX3   60  HN-VVDRDWIYGLMMQSNLFNRKERGGKIFVSPDYNQTMEQQREITMKRIWYLLENGVFK 118
             HN + D   +  +M +S LF  K       +    +   E++R++TM+RI YL++ G FK
PgA.16   64  HNNIQDLYSLLEIMCESPLFGPK-------LHAPASLFREEERDLTMRRIQYLVDQGGFK 116

AtACX3   119 GWLTETGPEAEL-RKLALLEVCGIYDHSVSIKVGVHFFLWGNAVKFFGTKRHHEKWLKNT 177
             G  T    E  L R+L   E+ G+ DH++ I++ +HF LWG A+K+ GTKRHH+KWLK T
PgA.16   117 GCPTAPADEQSLLRRLTYNELLGVSDHALPIQLSLHFSLWGGAIKYLGTKRHHDKWLKVT 176

AtACX3   178 EDYVVKGCFAMTELGHGSNVRGIETVTTYDPKTEEFVINTPCESAQKYWIGGAANHATHT 237
             ED+ +KGCF++TE+GH +NV+GIETV TYD +T EFVINTPCESAQK+WIGGA  HATHT
PgA.16   177 EDFGIKGCFSLTEMGHQNNVKGIETVATYDAQTREFVINTPCESAQKHWIGGALKHATHT 236

AtACX3   238 IVFSQLHINGTNQGVHAFIAQIRDQDGSICPNIRIADCGHKIGLNGVDNGRIWFDNLRIP 297
             IVF+QL I G NQGVHAF+  +RD  G++ PNI +ADCG K G+ GVD+GR WFDN+RIP
PgA.16   237 IVFAQLIIEGQNQGVHAFVVPVRDSKGNLYPNIHMADCGIKTGITGVDDGRAWFDNVRIP 296

AtACX3   298 RENLLNAVADVSSDGKYVSSIKDPDQRFGAFMAPLTSGRVTIASSAIYSAKVGLSIAIRY 357
             RENLLN+ ADV+ DG+YVSS++ P+QRF   + PLTS RV I  SAIY +++ L+I +++
PgA.16   297 RENLLNSAADVNPDGQYVSSVEYPNQRFATLVDPLTSERVNITVSAIYLSQIDLAIGVKF 356

AtACX3   358 SLSRRAFSVTANGPEVLLLDYPSHQRRLLPLLAKTYAMSFAANELKMIYVKRTPETNKAI 417
             ++S        +    LL++ PS++R ++PLLA  YA+SFA N LK IYVKRTPETNK I
PgA.16   357 AVS-NCDKPKYSDQRGLLINSPSYRRYIMPLLATVYAVSFAGNHLKSIYVKRTPETNKLI 415

AtACX3   418 HVVSSGFKAVLTWHNMHTLQECREAVGGQGVKTE-NLVGQLKGEFDVQTTFEGDNNVLMQ 476
              ++S+G  A L W N  T ++    +  +  KT  NL+G LK + DVQ+ F G   VL Q
PgA.16   416 QILSNGLNAKLNWLNYQTFEDNVMRILAR--KTVANLLGHLKAQEDVQSVFLGVRYVLWQ 473

AtACX3   477 QVSKALFAEYVSCKKRNKPFK-GLGLEHMNSPRPVLPTQLTSSTLRCSQFQTNVFCLRER 535
             QVSKAL  EY++ + +N+PFK GL LEHMN   PV+P+Q+TSS LR ++FQT+VF LRER
PgA.16   474 QVSKALVREYLTAQNQNRPFKEGLWLEHMNESTPVIPSQVTSSILRSTKFQTDVFFLRER 533

AtACX3   536 DLLEQFTSEVAQLQGRGESREFSFLLSHQLAEDLGKAFTEKAILQTILDAEAKLPTGSVK 595
             D+L +  +EVA+ Q  G S E +F    +L ++L  AFT+KAIL   ++AEAK+P G +K
PgA.16   534 DMLNRLVNEVAKHQNLG-SLEHAFAQCQELGKELAIAFTDKAILCAFIEAEAKVPVGPLK 592

AtACX3   596 DVLGLVRSMYALISLEEDPSLLRYGYLSQDNVGDVRREVSKLCGELRPHALALVTSFGIP 655
             DVL L+R++YAL  +EED    + GY+S         EV+ L  EL PH + LV  FGIP
PgA.16   593 DVLALLRALYALKMIEEDVGFTQNGYISAGIAEAAEEEVTSLSAELHPHVVELVDCFGIP 652

AtACX3   656 DSFLSP 661
             DSF SP
PgA.16   653 DSFFSP 658
```

Figure 17. Partial alignment of *At* ACX3 and A.16 translated protein using BLAST 2 sequences (Tatusova and Madden, 1999). Conserved amino acids (aas) are indicated by capital letter and equivalent aas by + sign. *At* ACX3 FAD-binding domain is highlighted in black. In the N-terminal part, peroxisomal PTS2 targeting signal is underlined with conserved or equivalent aas highlighted in grey. The partially conserved acyl-CoA dehydrogenase protein signature 1 (aas 187-199) is indicated in grey (Prosite PS00072) (Bairoch *et al.*, 1997).

Portulaca grandiflora

Arabidopsis thaliana

Significant alignments:		E value
COG1960	CaiA, acyl-CoA dehydrogenases [Lipid metabolism]	1e-12
pfam01756	ACOX, acyl-CoA oxidase	1e-10
pfam02770	acyl-CoA_dh_M, acyl-CoA dehydrogenase, middle domain	3e-04

Figure 18. Conserved domain search for *Pg* A.16 and *At* ACX3 with NCBI CDD search tool (cddv1.62) using RPS-BLAST algorithm (v2.2.6). Scores are indicated for A.16 protein.

Proteins of the acyl-CoA oxidases family (EC:1.3.3.6) participate in fatty acid degradation. These enzymes catalyse the first step of the beta-oxidation cycle, i.e. the conversion of acyl-CoA into *trans*-2-enoyl-CoA in the presence of oxygen and flavin adenine dinucleotide (FAD) as a cofactor.

$$\text{acyl-CoA} + \text{FAD} \rightarrow \textit{trans}\text{-2-enoyl-CoA} + \text{FADH}_2 \quad ; \quad \text{FADH}_2 + \text{O}_2 \rightarrow \text{FAD} + \text{H}_2\text{O}_2$$

Four genes have been identified in *A. thaliana* till now (Eastmond *et al.*, 2000a). Each of the corresponding proteins are specialized in fatty acids of different chain-length. The *At* ACX3 single gene degrades only medium-chain length fatty acids (C8:0 to C14:0-CoA) and is localized to the peroxisome (Eastmond *et al.*, 2000b; Froman *et al.*, 2000). By comparison with the *At* ACX3 sequence, we identified a putative type-II peroxisomal targeting signal (PTS2) (Subramani, 1996) of 34 residues at the amino terminus of *Pg* A.16 (Fig. 17). The consensus sequence of the plant PTS2 [R-(I/Q/L)-X5-HL-X15-22-C] is present in most peroxisomal proteins such as ACXs, thiolases or the malate dehydrogenase (Froman *et al.*, 2000). The cysteine present at the end of this pattern is the signal for the maturation of the protein entering the peroxisome.

By compiling all these results, we conclude that A.16 is a peroxisomal acyl-CoA oxidase-like protein, which diverges locally. Nevertheless, all eukaryotic ACXs are flavoproteins containing a FAD-binding motif (Dubourdieu and Fox, 1977), which is not the case with *Pg* A.16. The oxidation of acyl-CoA requires FAD as a cofactor; the resulting $FADH_2$ is subsequently re-oxidized by O_2 to form H_2O_2. Until now, this FAD-binding motif has been found to be present in all types of plant ACXs (Do and Huang, 1997; Hayashi *et al.*, 1998, 1999). By aligning A.16 with *At* ACX3, we observed the absence of a homologous sequence of 15 residues at the place of the *At* ACX3 FAD-binding motif (Fig. 17). Its first-half has no homology with A.16 sequence and 2 gaps are necessary to align both sequences correctly. Alignment of the *Pg* A.16 and *At* ACX3 sequences with Caryophyllales homologues shows that *Pg* A.16 is the sole sequence diverging at the site of the FAD-binding motif (see Fig. 19). This could have been due to errors in the cDNA sequence. However after isolating the corresponding genomic DNA by proofreading PCR, the *A.16* cDNA sequence was confirmed. In order to verify a possible duplication of *A.16* in the *Pg* genome, we performed a Southern blot analysis. Results showed that *A.16* is a low-copy number single gene (Fig. 20). In conclusion, we have no functional evidence that *Pg* A.16 is an ACX, therefore we still had to test it.

```
At ACX3          ETNKAIHVVSSGFKAVLTWHNMHTLQECREAVGGQGVKT-ENLVGQLKGE
Mc               ETNKIIHVLSSGLKATLTWHNMRTLRXCREXCGGQGLKT ENRVGHLKAR
Bv               ETNKIIHVASIALKSTLTWHNMHTLQDCRETCGGHRLNT-----------
Pg A.16          ETNKLIQILSNGLNAKLNWLNYQTFEDNVMRILA R KTVANLLGHLKAQ
                 **** *:: * .::: *.* * :*:.:      .:  .*  *  * **.
```

Figure 19. Partial alignment of *At* ACX3 with Caryophyllales homologues. Grey boxes correspond to *At* ACX3 positions conserved in other sequences. *At* ACX3 FAD-binding motif is underlined. Mc, *Mesembryanthemum crystallinum* (BE035467); Bv, *Beta vulgaris* (BI096127).

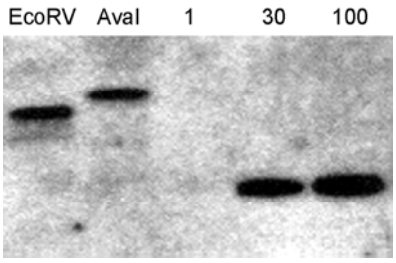


Figure 20. High-stringency Southern blot analysis of *Pg A.16*. *Pg* genomic DNA digested with *EcoR*V and *Ava*I was loaded, as well as 1, 30 and 100 equivalent-copies of *A.16* gene in *Pg* genomic DNA.

1.5.2 Complementation analysis of *Saccharomyces cerevisiae fox-1* mutant by *Pg A.16* cDNA

A.16 cDNA is homologous to ACXs but does not possess the FAD-binding site which is essential for the protein function. In order to verify the ACX function of this protein, we decided to test its capacity to complement a *Saccharomyces cerevisiae* ACX mutant. This species is ideal for ACX complementation analysis because it has only one ACX protein called FOX-1 (Dmochowska *et al.*, 1990; Kunau and Hartig, 1992), which is rather an exception in eukaryotic organisms. Mutagenesis of *S. cerevisiae* cells gave rise to mutants unable to use oleic acid as sole carbon source. All these *fox* mutants are defective in beta-oxidation cycle, but possess morphologically normal peroxisomes.

In order to test A.16 functionality, its cDNA was introduced into the p415GPD yeast expression vector (Mumberg *et al.*, 1995) at *Xba*I/*Pst*I sites. The p415GPD plasmid contains a glyceraldehydes-3-phosphate dehydrogenase (GPD) constitutive promoter and a cytochrome-C-oxidase *CYC1*-terminator. GPD415-A16 construct was transformed into DH5α *E. coli* and verified by restriction analysis. Finally, GPD415-A16 was transformed into wild-type (Eurofan number BY4742) and *fox-1 S. cerevisiae* strains (Eurofan number Y14571). We decided to employ three different methods in parallel to test functional complementation of the *fox-1* mutant by *A.16* full-length cDNA.

Monitoring the formation of clear zones on oleic acid/Tween 80 solid medium

First, we decided to study the consumption of oleic acid as a single carbon and energy source on a solid medium by monitoring the formation of clear zones around yeast cell growth on turbid media containing fatty acids dispersed with Tween 80 (Gurvitz *et al.*, 1997). Tween 80 does not serve as an alternative carbon source for *S. cerevisiae* when used as a dispersing agent in fatty-acid medium. The *FOX-1* steady state mRNA level is fully induced by oleic acid as a carbon source (Stanway *et al.*, 1995). The results showed that on the opaque medium consisting of oleic acid/Tween 80, the wild-type strain produced the characteristic clear zone indicative of the utilization of oleic acid as a sole carbon and energy source, whereas the *fox-1* mutant failed to do so. After trying different concentrations of oleic acid and Tween 80, we observed no clear-zone formation around the strain *fox-1* complemented with *A.16* cDNA, thus suggesting the absence of complementation of *fox-1* by *A.16* cDNA.

Comparative growth study of wt and *fox-1*/*A.16* *S. cerevisiae* strains in oleic acid medium

We performed a comparative growth study of *fox-1*, *fox-1*/*A.16* and wt strains in liquid Yeast Extract Peptone medium (YEP) completed by oleate as a sole source of carbon. A preliminary overnight preculture was done in rich YPD medium (YEP + Dextrose) or selective Leu- medium for the wt, because *fox-1* mutant does not grow at all on YEP/oleic acid medium, but just survives. In order to solubilise the oleic acid, pluronic acid was added in low concentration in the medium. Before inoculating the YEP/oleic acid test medium, cells were quantified in order to inoculate an equivalent amount of each strain. Many cultures were grown in parallel to avoid growth disturbances due to sampling. Growth was followed from the time of the inoculation (time 0) for 72 hours by measurement of the optical density at 600 nm (OD 600) as shown in Fig. 21.

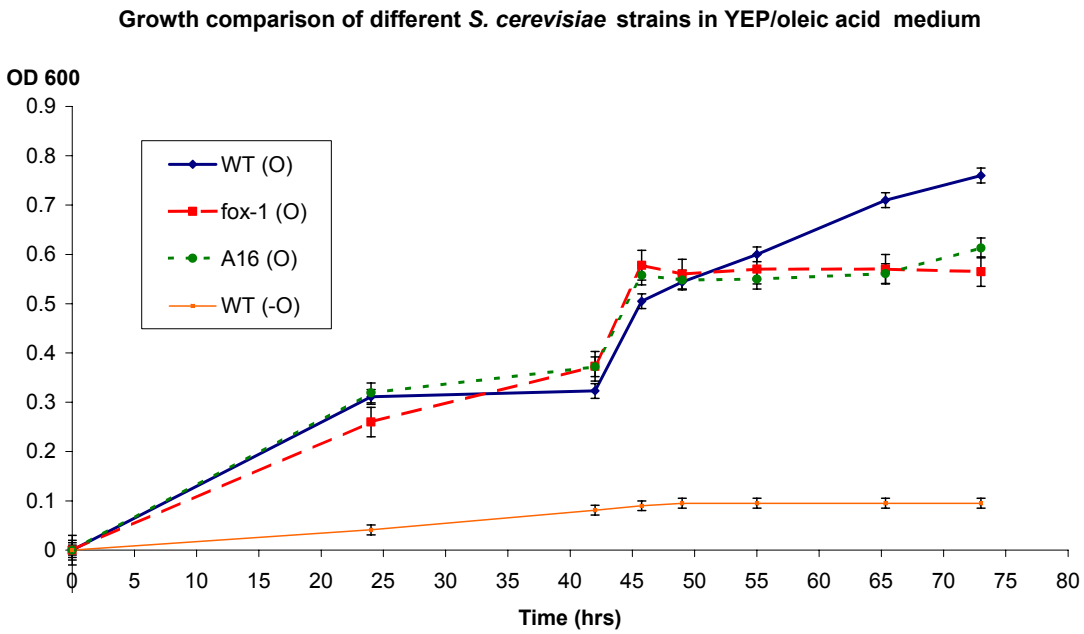


Figure 21. Comparison of the capacity of different yeast strains to use oleate as unique source of carbon. Mediums are YEP (-O) or YEP completed by oleate (+O). Strains are the wild-type (WT), *fox-1*, *fox-1* complemented with *A.16* cDNA (A16).

The wt strain grows at the same speed as *fox-1* and *fox-1*/*A.16* strains during the first two days and accelerates its growth during the third day. This phenomenon can be explained by the consumption during the first two days of the carbon reserves accumulated during the overnight pre-culture in a dextrose containing medium. Thus, wt and *fox-1* growth differ only during the third day. No difference was observed between the growth curves of *fox-1* mutant and *fox-1* complemented with *A.16*. These results show that *fox-1* complementation failed. This strain seems still unable to metabolise oleic acid because of its defective acyl-CoA-oxidase.

Monitoring the formation of PHA in the presence of oleic acid and the PHA synthase

Polyhydroxyalkanoates (PHA) are polyesters of hydroxyacids naturally synthesised in bacteria as a carbon reserve. PHA synthesis can be used as a tool to study the quality and relative quantity of the carbon flow through beta-oxidation in eukaryotic organisms as well as to analyse the degradation pathway of unusual fatty acids (Mittendorf *et al.*, 1998, 1999). A peroxisomal PHA pathway was engineered in *S. cerevisiae* by introducing the *Pseudomonas aeruginosa* PHAC1 synthase modified for peroxisome targeting by the addition of the carboxy-terminal 34 amino acids from the *Brassica napus* isocitrate lyase (Poirier *et al.*, 2001). PHA synthase expression and PHA accumulation were found in recombinant yeast growing in media containing fatty acids (see Fig. 22).

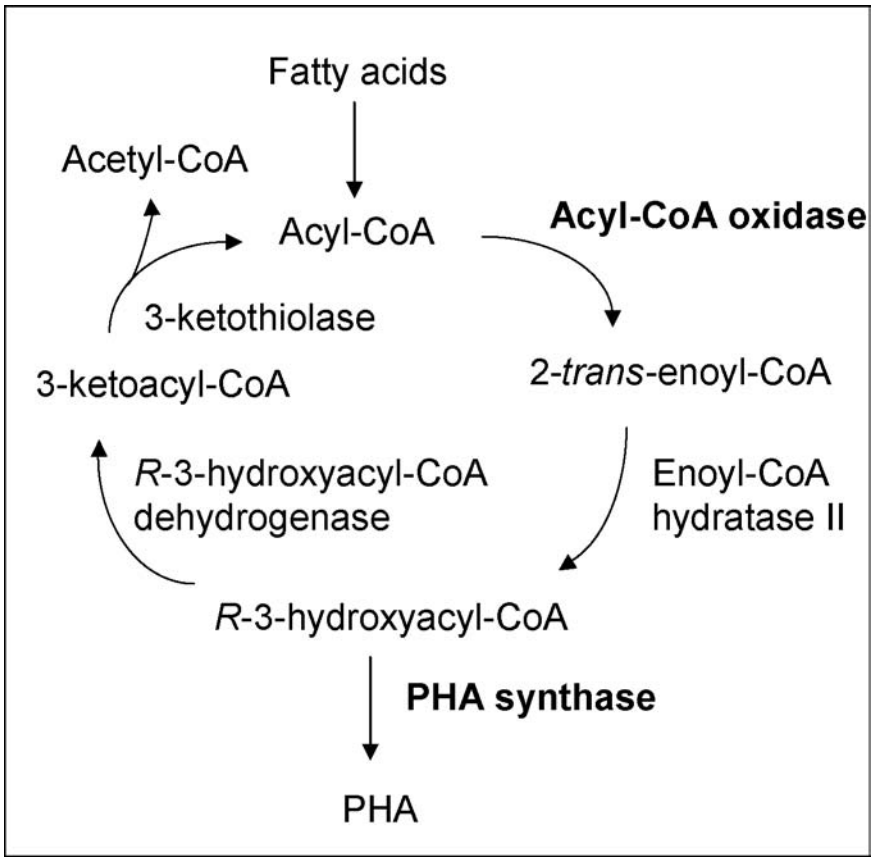


Figure 22. Beta-oxidation cycle and PHA synthesis in *S. cerevisiae*.

Consequently, we decided to transform *fox-1* and wt strains with pGPD415-A16 and the modified *Pseudomonas aeruginosa* PHAC1 synthase cloned into a yeast expression vector under the control of the catalase promoter (Poirier *et al.*, 2001), in order to monitor the fatty acid degradation through the newly introduced PHA biosynthesis. Transformed cells were grown 48 hrs on rich medium, inoculated in YEP/oleate medium (C17) and grown for 4 days as described by Poirier and colleagues. PHA were analysed on GC-MS after lyophylization of the cells, extraction of the PHA and transesterification of the extracted products. GC-MS analysis revealed that no PHA was synthesised in *fox-1* transformed with *A.16* cDNA and PHA synthase, whereas the whole PHA range was present in the wt control transformed with PHA and *A.16* as expected (see Figs. 24-25), confirming the inability of *A.16* cDNA to complement *fox-1* mutant deficient in ACX.

Three different experiments showed that A.16 is unable to replace the unique *S. cerevisiae* ACX. That raises the questions of the expression and the functionality of *Pg* A.16 protein in yeast. We can ask if A.16 is non-functional in yeast due to a too high level of dissimilarity with the yeast FOX1 (50% of indentity only) or due to the absence of the FAD-binding site absolutely necessary for the course of the chemical reaction. The problem of A.16 functionality could be solved by testing the capacity of A.16 to metabolise acyl-CoA in vitro according to the enzymatic assay described in Fig. 23 (Hooks *et al.*, 1996). Overexpression of A.16 in *E. coli* could be a solution, nevertheless an alternative is its transitory expression in the bryophyte *Physcomitrella patens*. None of these solutions will be investigated here due to the simultaneous identification of the more promising *P. grandiflora* DOPA-dioxygenase gene.

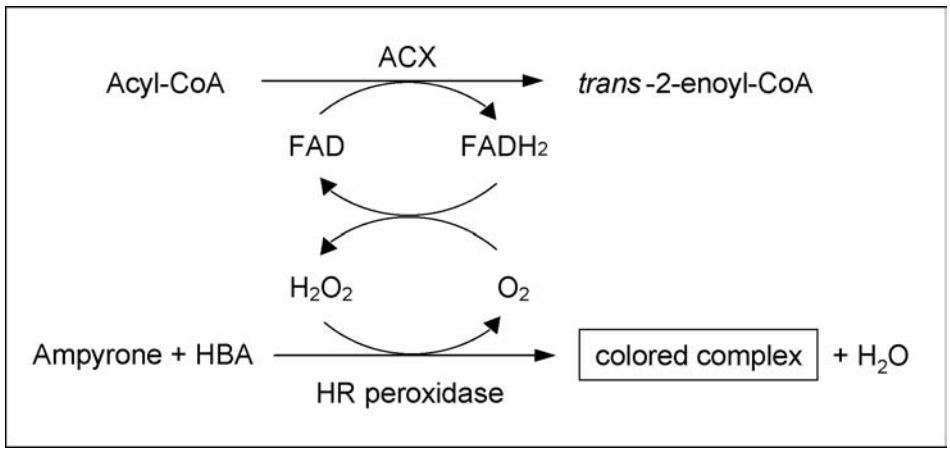


Figure 23. Acyl-CoA oxidase enzymatic assay. The acyl-CoA-dependant H_2O_2 production is indirectly measured in a coupled reaction. Production of a coloured complex is quantified by spectrophotometry at 500 nm. HBA, *para*-hydroxybenzoic acid; HR, horse radish peroxidase.

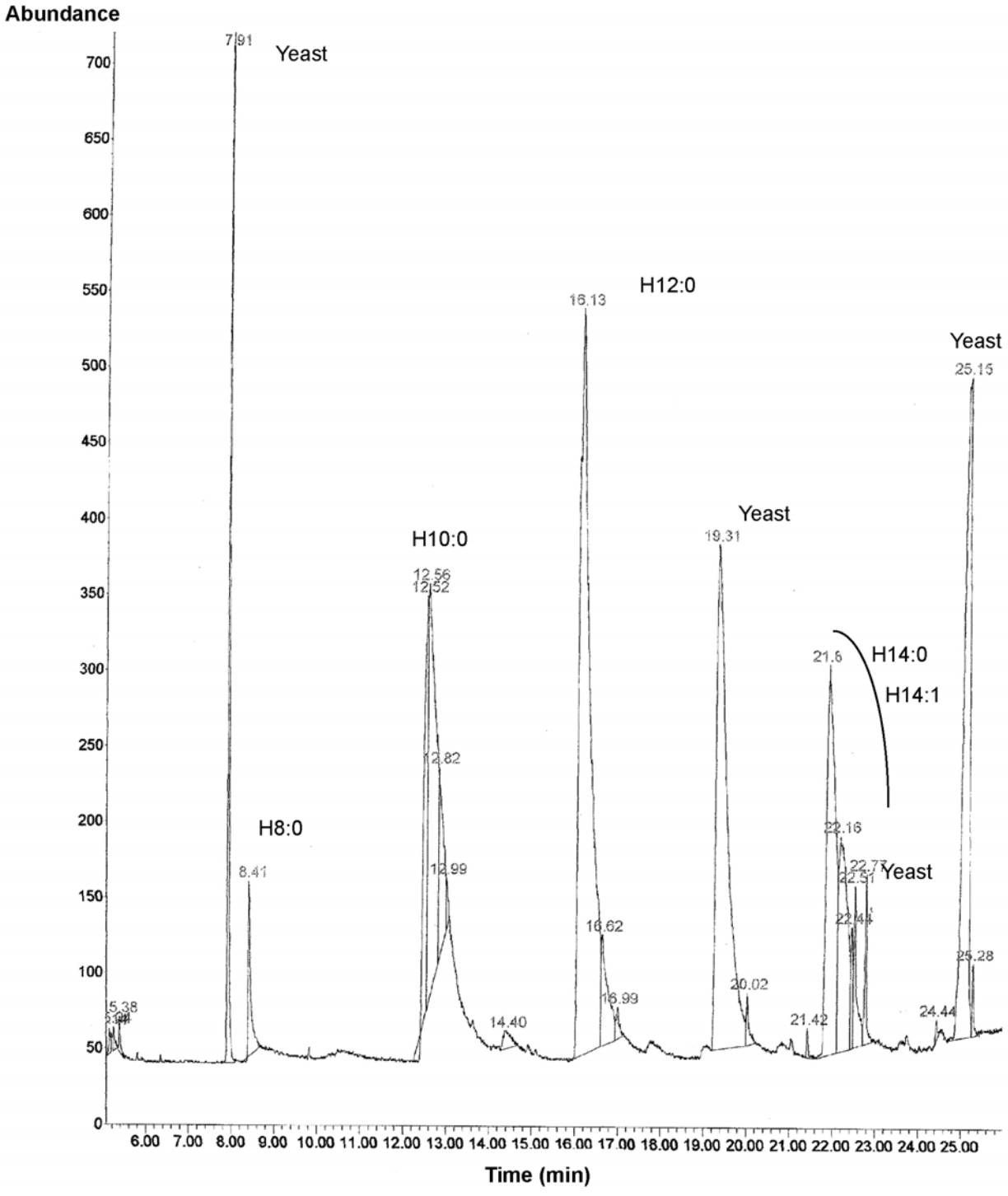


Figure 24. GCMS analysis of PHA synthesis in wt yeast transformed with PHA synthase and A.16. PHA peaks are labelled according to the PHA carbon chain length (H14:0, H14:1, H12:0, H10:0, H8:0). Peaks corresponding to yeast background noise are labelled by "Yeast". Time is indicated in minutes.

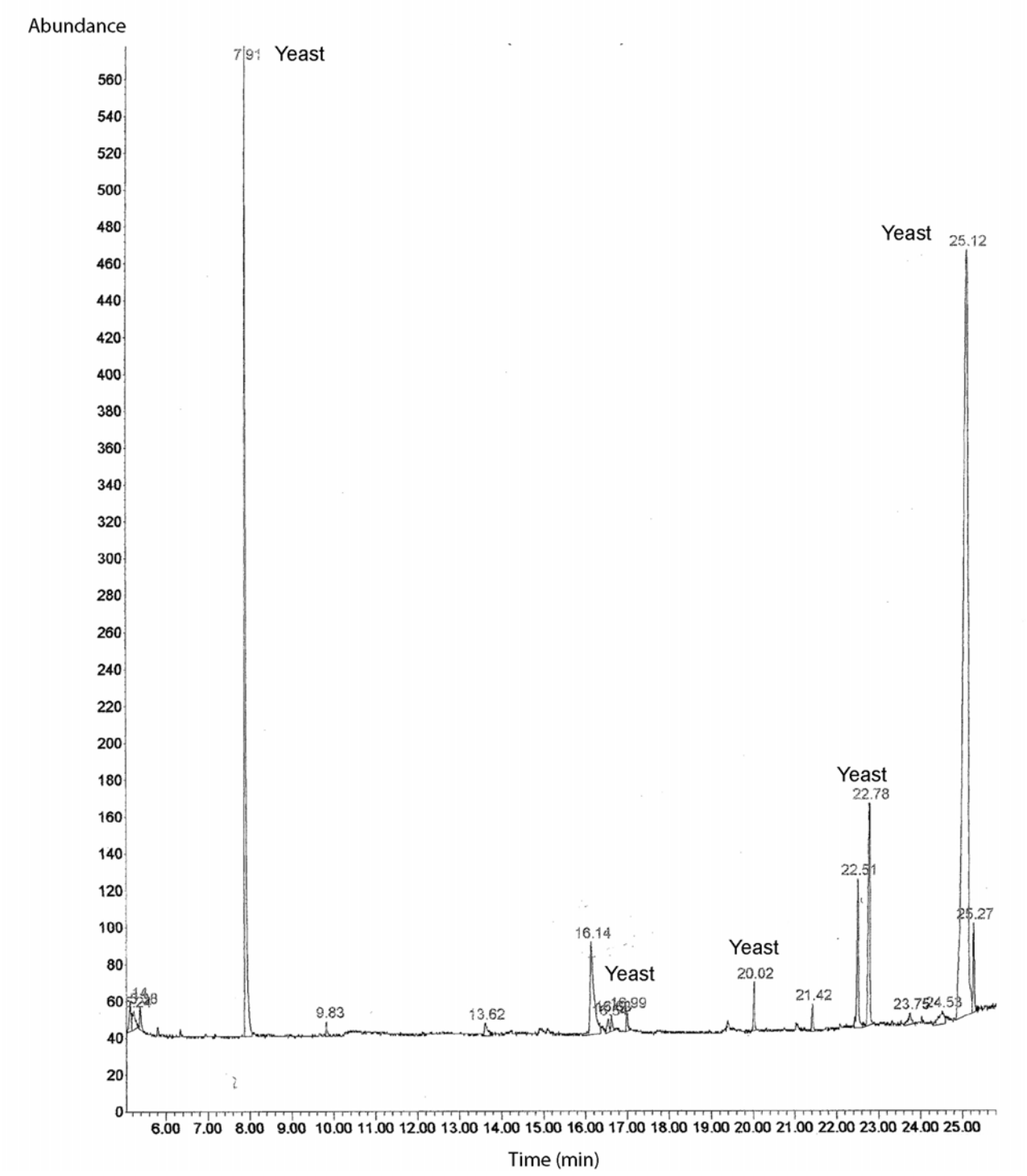


Figure 25. GC-MS analysis of PHA synthesis in *fox-1* yeast transformed with PHA synthase and A.16. PHA peaks are labelled according to the PHA carbon chain length (H14:0, H14:1, H12:0, H10:0, H8:0). Peaks corresponding to yeast background noise are labelled by "Yeast". Time is indicated in minutes.

In conclusion, the A.16 translated protein is homologous to eukaryotic ACX3 proteins, but we have for the moment no functional evidence of its participation in the beta-oxidation cycle. Moreover, the correlation observed between *A.16* mRNA expression and betalain biosynthesis in *Pg* pigmented tissues is not obvious if we consider A.16 as an ACX. It would be interesting to verify if other enzymes of the beta-oxidation cycle are also overexpressed in betalain-containing tissues. In the case of a positive answer, this raises the question of a up-regulation of the beta-oxidation cycle correlated with the presence of betalain pigments. For the moment, the sole functional indication that we have for A.16, is the presence of the conserved acyl-CoA dehydrogenase middle domain (pfam02770), containing a β-barrel fold. Such enzymatic activity is present in numerous pathways. Therefore, the role of this protein in relation with betalain biosynthesis remains open.

1.6 Discussion

In the first part of our work, we described the use of a PCR subtractive hybridization strategy to isolate genes expressed specifically during betalain biosynthesis in *P. grandiflora* violet petals. After a first screening step by slot blot hybridization, we obtained 26% of betalain-specific clones. Northern blot analysis reduced this number to 10% violet-specific clones, but corresponding to only three different cDNAs called *A.16*, *P.34* and *V.33*. Two cDNAs (*L.6* and *L.13*), previously isolated in a similar yellow-specific subtracted library from *Pg* petals (Zaiko, 2000), were absent from our subtraction library despite their demonstrated betalain specific expression in *Pg* petals (Zaiko, 2000). All these genes failed to complement *Pg* white petals.

We can conclude that the construction of such a cDNA subtracted library combined with several hybridization screens allows the identification of some genes taking part in a biosynthesis pathway of interest, but does not give at all an exhaustive image of all genes specifically involved in this metabolic pathway. Many candidates are lost during the numerous steps necessary to the construction of the library. This can be the consequence of the low expression level of an mRNA despite the amplification technique that we use. However, this strategy is powerful if used in replicate experiments and/or similar experiments with different tissues leading all to the same result. Thus, a comparison of the different results obtained in several experiments will give a better view. It could be also interesting to undertake in parallel a second strategy. RNA fingerprinting and mRNA differential display have already been used to isolate betalain-specific genes without success (Zaiko, 2000). Therefore, we would suggest other strategies that could avoid the loss of candidates and give a more exhaustive view of all genes involved directly or indirectly in betalain biosynthesis. The comparison and the micro-sequencing of the spots of 2-D protein gels from betalain-containing tissues versus green tissues is a possibility. Nevertheless, the number of DOPA-dioxygenase molecules in a plant cell could be smaller than 24000 according to kinetic studies of DODA in *Amanita muscaria* (Girod, 1989). At this concentration, the direct observation of a protein is critical, suggesting rather a gene-based approach. For the future, we propose the construction of a microarray containing all genes expressed in *Pg* petals and its screening by mRNA probes from the whole range of *Pg* coloured petals versus white probes. This strategy will offer the advantage of detecting not only single specific genes, but also whole pathways, thus allowing a better understanding of the regulation of betalain biosynthesis and its influence on the plant physiology.

2. A new type of plant dioxygenase involved in betalain biosynthesis

2.1 Dioxygenases and their participation in betalain biosynthesis

2.1.1 Introduction

Enzymes catalysing the incorporation of two atoms of molecular oxygen into their substrate(s) are called dioxygenases. A classification of the main families of dioxygenases is given in Table 5. They are set apart in two groups depending if they catalyse reactions acting on a single oxygen acceptor (EC 1.13.11.xx.) or on paired acceptors (EC 1.14.11.xx and 1.14.12.xx). Dioxygenase activity generally requires the presence of a cofactor to obtain oxygen in its activated form. The cofactor is usually a transition metal, iron, or more rarely copper or manganese. In the mononuclear non-heme iron enzymes, the reaction is catalysed either by a high-spin ferrous site involved in dioxygen activation or by a high-spin ferric site which activates substrates. For some of the ferrous enzymes, an additional organic cofactor, α-ketoglutarate, flavin or pterin participates in a coupled reaction with dioxygen, where both the substrate and the cosubstrate are oxygenated.

Dioxygenase family	Functional characteristics	Product / reaction type
Lipoxygenases	Fe III fatty acids hydroperoxidation	Jasmonic acid Flavour / aroma
2-oxoacid dependent dioxygenases	Fe II + 2 O2 acceptors hydroxylation / desaturation epoxidation	Flavonoids / ethylene Gibberellins / alkaloids
Intradiol dioxygenases mainly in bacteria	Fe III *ortho*-cleavage	aromatic compound catabolism
Extradiol dioxygenases mainly in bacteria	Fe II *meta*-cleavage	aromatic compound catabolism

Table 5. Classification of non-heme iron dioxygenases.

Until now, plants dioxygenases are distributed in two main classes, lipoxygenases (LOX, EC 1.13.11.12) and 2-oxoacid-dependent dioxygenases (2-ODD, EC 1.14.11.xx) as described in Table 5 (Prescott and John, 1996). The former are non-heme ferric monomeric dioxygenases catalysing the hydroperoxidation of (*Z*,*Z*)-1,4-pentadiene polyunsaturated fatty acids to produce unsatured fatty acid hydroxyperoxide. The predominant substrates of LOXs in plants are generally linoleic and linolenic acids. LOX activity is at the origin of the biosynthesis of aromas, flavours and oxylipins, the precursors of jasmonic acid in plants (Siedow, 1991). Progress in this field has been reviewed recently (Porta and Rocha-Sosa, 2002). Ring-cleavage of aromatic compounds by dioxygenases are poorly studied in plants.

2-ODDs are also called 2-oxoglutarate-dependent dioxygenases due to the obligatory requirement of 2-oxoglutarate as cosubstrate for most of them. These Fe^{2+}-containing enzymes catalyse hydroxylation, epoxidation or desaturation reactions, thus participating in substrate conversion in numerous metabolic pathways and notably in gibberellin, ethylene, flavonoid or alkaloid biosynthesis (Prescott and Lloyd, 2000). They also play catabolic or anabolic roles in the degradation of essential metabolites like amino acids, lipids, sugars and nucleic acids.

Dioxygenases from plants and animals have been much less studied than bacterial dioxygenases, in part because they are often present in low titres in the producing tissues. This problem has been partially overcome by the use of recombinant DNA technology. However many eukaryotic dioxygenases still cause significant difficulties for study. Among them, we notice the multienzymatic structure of dioxygenases, the presence of polymeric substrates and/or products and sometimes the exhibition of features such as auto-inactivation/auto-degradation (IturriagagoitiaBueno *et al.*, 1996; Kosugi *et al.*, 1997). It seems that many dioxygenases have multiple activities, making it difficult to identify these enzymes and their relevant physiological reaction(s).

2.1.2 Extradiol dioxygenases in betalain biosynthesis

In the plant betalain biosynthesis pathway, the existence of a dioxygenase catalysing the cleavage of DOPA between C_4-C_5 was predicted from the elucidation of the indicaxanthin structure (Impellizzeri and Piattelli, 1972). Similar results were confirmed with the isolation and the characterization of the *Amanita muscaria* DOPA-dioxygenase (*Am* DODA) involved in the synthesis of betalamic acid (Girod and Zryd, 1991a; Hinz *et al.*, 1997).

Am DODA differs completely from any known plant dioxygenases. *Am* DODA is a homomultimeric dioxygenase (Girod, 1989; Girod and Zryd, 1991a), whereas plant LOXs are monomeric enzymes, and eukaryotic ODDs are generally heterotetramers ($\alpha 2\beta 2$) (Clifton *et al.*, 2001). *Am* DODA structural and functional characteristics are nearer to those of bacterial extradiol dioxygenases than to any existing fungal or plant dioxygenases due to their similar extradiol ring-opening reaction, their usual homomultimeric structure and the presence of a ferrous cofactor in their active sites (Tatsuno *et al.*, 1980). Dihydroxylation of the aromatic ring with two adjacent hydroxyl groups is the initial step required for the degradation of organic compounds like benzene by bacteria. This reaction gives a catechol molecule possessing a diphenolic structure close to the DOPA structure (dihydroxyphenylalanine). In a second step, the catechol is degraded by an aromatic ring-opening dioxygenase accompanied by a non-heme iron cofactor. On the basis of the position of the ring-cleavage, the bacterial catecholic ring-opening dioxygenases can be classified into two types, the intradiol and the extradiol (Lipscomb and Orville, 1992). The intradiol type cleaves the C-C bound between the two hydroxyl groups (*ortho*-cleavage) in the presence of non-heme ferric iron cofactor, whereas the extradiol type cut the ring at the outside C-C bond adjacent to the two hydroxyl groups (*meta*-cleavage) in the presence of non-heme ferrous iron in their active site (see Table 5). The latter type exerts the same activity as *Am* DODA and the predicted plant DOPA-dioxygenase.

Am DODA is not specific to DOPA. It can also catalyse the 4,5-extradiol ring-cleavage of catechol, dopamine and caffeic acid (Girod and Zryd, 1991a), and also the 2,3-extradiol-cleavage of DOPA, resulting in the synthesis of muscaflavin (Mueller *et al.*, 1997b). This could explain why no sequence similarity has been detected at the amino acid level between bacterial extradiol dioxygenases and *Am* DODA (Hinz *et al.*, 1997).

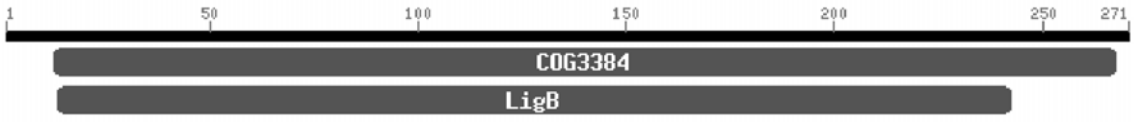
To date three classes of extradiol-type (catecholic) dioxygenases have been identified in bacteria (Spence *et al.*, 1996). Alignment of the primary structures of the class I and II enzymes revealed that these two sorts of enzymes have a common ancestor. The class II enzymes come from a gene-duplication of class I enzymes (Eltis and Bolin, 1996; Spence *et al.*, 1996). Numerous crystal structures confirm these results by revealing the presence of the same folding pattern in the two domains composing the class II enzymes (Han *et al.*, 1995). On the contrary, the class III extradiol dioxygenases do not show similarity at the amino acid level with the class I and II enzymes (Eltis and Bolin, 1996). However, spectroscopic studies suggested several common active sites between class II and III (Arciero *et al.*, 1985; Arciero and Lipscomb, 1986). In an identical manner, *Am* DODA and bacterial extradiol dioxygenase could originate from different ancestors, but possess a similar functional activity. Thus, *Am* DODA appears to constitute a new class of extradiol dioxygenases specifically present in fungi. Different phylogenetic origins were also predicted for the plant and the fungal DOPA-dioxygenase because of the absence of muscaflavin compounds in betalain plants (Mueller *et al.*, 1997a). Moreover, no homologue was found for *Am* DODA in betalain plants despite many homologous screening attempts (Hinz, personal communication).

In plants, DOPA ring-cleavage activity seems also to occur in the biosynthesis of stizolobic or stizolobinic acid (Saito *et al.*, 1976). These products have been identified in leguminous (Fabaceae) species, where DOPA is present at especially high concentration, notably around 5% in the seeds from the genera *Mucuna* and *Stizolobium* (Yang *et al.*, 2001). In these species, DOPA has been identified as the precursor for the synthesis of stizolobic or stizolobinic acid. DOPA ring-cleavage is one of the steps leading to the formation of these compounds, but is not sufficient by itself. Their formation requires a supplementary reduction step that the DOPA-dioxygenase cannot catalyse. Analysis of the products of the recombinant *Am* DODA did not detect stizolobic or stizolobinic acid molecules (Mueller, 1997), despite the fact that both compounds are naturally synthesised in the genus *Amanita* (Saito *et al.*, 1978). Ring-opening of tyrosine was also detected in the Droseraceae plant family. Feeding *Drosera* tissue cultures, as well as cell suspension cultures, with radiolabeled β-^{14}C-tyrosine, showed that tyrosine is broken down to acetate via the homogentisate pathway by an oxygenase activity (Durand and Zenk, 1974). Since then, the enzyme was identified as a 1,2-homogentisate-dioxygenase (EC: 1.13.11.5) (Fernandezcanon and Penalva, 1995).

2.2 Identification of a betalain specific 4,5-dioxygenase in *P. grandiflora*

2.2.1 In silico functional identification of a putative extradiol 4,5-dioxygenase

A cDNA subtractive hybridization strategy between pigmented (*C*-) and white (*cc*) *P. grandiflora* genotypes allowed the isolation of several betalain-specific cDNAs in our laboratory. Among them, *L.6* and *V.33* were without function and not able to complement *Pg* white petals by biolistic transformation (Zaiko, 2000). Nevertheless, both candidates had transcripts present in all pigmented tissues, except coloured stems for *L.6*, and absent from the non-pigmented tissues. Therefore, we were still interested in identifying a putative function for these genes. The NCBI CDD domain search tool (Marchler-Bauer *et al.*, 2002) revealed a homology of the *L.6* translated protein (34% identity, 48% similarity) with the LigB domain (pfam02900) present in several bacterial extradiol 4,5-dioxygenases (Fig. 26). This domain was first identified on the β-subunit of a bacterial heterotetrameric class III extradiol-type 4,5-dioxygenase (Genbank entry AAA17728) of *Sphingomonas paucimobilis* (Noda *et al.*, 1990). This extradiol dioxygenase function corresponded exactly to that expected for the plant DOPA-dioxygenase gene *C*. We have tentatively assigned *L.6* as *DODA*.



Significant alignment:		E value
COGG3384	Uncharacterized conserved protein [function unknown]	4e-47
pfam02900	Catalytic LigB subunit of aromatic ring-opening dioxygenase	1e-30

```
L.6    MLADESFIARNFL--LG------------WKKNVF----PVK-----PKSI-LVVSAHWETDVPCVSAGQYPNVIYDFTEV
LigB   MARVTTGITSSHIPALGAAIQTGTSDNDYWGP-VFKGYQPIRDWIKQPGNMPDVVILVYNDHASAFDMNIIPTFAIGCAET
       *    : *: ..:  **            *   **    *::     * .:  **   :: ...... .  *..  . :*.

L.6    --PASMFQMKYPAP---GCPKLAKRVQELLIAGGFKSAKLDEERGFDHSSWVPLSMM--------CPEADIPVCQLSVQPG
LigB   FKPADEGWGPRPVPDVKGHPDLAWHIAQSLILDEFDMTIMN-QMDVDHGCTVPLSMIFGEPEEWPCKVIPFPVNVVTYPP-
         **.      *.*   * *.** :: : ** . *. : :: : ..**.. *****:        *    :**  ::  *

L.6    LDATHHFNVG---RALAPLKGE--GVLFIGSGGAVHPSDDTPHWFDGVAPWAAEF-DQWLEDALLEGRYEDVNNYQTKAPE
LigB   PSGKRCFALGDSIRAAVESFPEDLNVHVWGTGGMSHQLQG-PRAGLINKEFDLNFIDKLISDPEELSKMPHIQYLRESGSE
        ...: * :*   ** .    *  .* . *:**  *  :. *:       :  :* *: :.*.   .:  .::  : ...*

L.6    G-----WKLAH-PIPEHFLPLHVAMGAGGEKSK-AELIYRTWDHGTLGYASYKFTSI------
LigB   GVELVMWLIMRGALPEKVRDLYTFYHIPASNTALGAMILQPEETAGTPLEPRKVMSGHSLAQA
       *     * : : .:**:.  *:.     ..::  . :* :. : .     . *. *
```

Figure 26. Identification of a LigB domain (pfam02900) in L.6 translated protein.

As a consequence of this result, we concluded that the methionine identified as the start codon (Zaiko, 2000), was in fact not the first methionine. Twenty-one amino acid residues were missing before this methionine, thus explaining the fact that the size of *DODA* mRNA on Northern blot was slightly longer than the size of *DODA* cDNA. Therefore we decided to clone again the cDNA by RACE strategy with primers PgDODAP2, PgDODAP3. Due to difficulties in obtaining the expected product by proofreading PCR amplification on *Pg* cDNA, we tried the same strategy on *Pg* genomic DNA. Sequencing revealed that this genomic PCR product contained no intron and thus was equivalent to the cDNA. A new first methionine was identified 21 amino acids upstream the previous one as predicted. The *DODA* cDNA sequence (Genbank entry AJ580598) is 1249 bp long with an ORF predicted to encode a 271 amino acid protein with a theoretical molecular mass of 29.9 kDa and an acidic pI of 5.6. PSORTII tool (Nakai and Horton, 1999) predicts a clear cytoplasmic localization of the protein (see Fig. 28).

A Southern blot analysis performed on genomic DNA from *P. grandiflora*, using a radiolabeled *Nco*I-digested 789 bp fragment from *L.6* cDNA as a probe, confirmed the presence of a single gene without intron. *Pg* Genomic DNAs were individually digested with *Eco*RI and *Nhe*I. They were then separated on agarose gel and transferred to a nylon membrane. Two bands were observed for *Nhe*I and one band for *Eco*RI as expected in the case of a single gene (Fig. 27). The intensity of the obtained bands corresponded to a one-copy gene equivalent from the plasmid. *Pg DODA* cDNA is therefore a single copy gene.

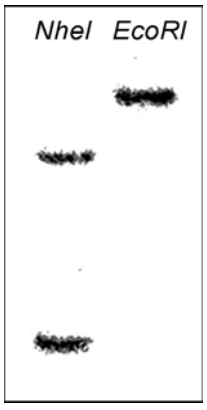


Figure 27. Southern blot analysis of the *Pg* DODA gene. Washes in SSC buffer were done to exclude the hybridization of sequences with less than 95% of homology.

```
-85                                                                cgctctagaaccat   -71
-70   gggcagtattatcatttatcagttatccagtatccagtatctggtaagtgagagtataacaaacaaagaa    -1

  1   ATGGGTGTTGGGAAGGAAGTGTCGTTTAAGGAGAGTTTCTTCTTGTCTCATGGGAATCCAGCCATGTTGG    70
        M  G  V  G  K  E  V  S  F  K  E  S  F  F  L  S  H  G  N  P  A  M  L

 71   CGGATGAATCCTTTATAGCGAGGAACTTCCTTTTGGGATGGAAAAAGAATGTGTTTCCTGTGAAACCCAA   140
        A  D  E  S  F  I  A  R  N  F  L  L  G  W  K  K  N  V  F  P  V  K  P  K

141   GTCCATTTTAGTAGTCTCTGCTCACTGGGAGACTGATGTCCCTTGTGTATCTGCTGGTCAGTACCCTAAT   210
        S  I  L  V  V  S  A  H  W  E  T  D  V  P  C  V  S  A  G  Q  Y  P  N

211   GTTATCTACGATTTCACTGAAGTCCCTGCTTCCATGTTTCAGATGAAGTACCCAGCTCCAGGGTGCCCAA   280
        V  I  Y  D  F  T  E  V  P  A  S  M  F  Q  M  K  Y  P  A  P  G  C  P

281   AGCTGGCAAAAAGAGTGCAGGAACTGCTGATAGCGGGAGGTTTCAAGAGCGCGAAGCTAGATGAGGAGCG   350
        K  L  A  K  R  V  Q  E  L  L  I  A  G  G  F  K  S  A  K  L  D  E  E  R

351   GGGTTTCGACCACAGCTCATGGGTGCCACTGAGCATGATGTGCCCGGAGGCGGACATCCCGGTGTGCCAG   420
        G  F  D  H  S  S  W  V  P  L  S  M  M  C  P  E  A  D  I  P  V  C  Q

421   CTTTCGGTGCAGCCTGGGCTTGACGCGACCCACCACTTCAACGTGGGGCGAGCGTTGGCCCCACTCAAAG   490
        L  S  V  Q  P  G  L  D  A  T  H  H  F  N  V  G  R  A  L  A  P  L  K

491   GGGAAGGTGTCCTCTTCATTGGCTCCGGTGGGGCTGTCCACCCTTCTGATGACACCCCTCATTGGTTCGA   560
        G  E  G  V  L  F  I  G  S  G  G  A  V  H  P  S  D  D  T  P  H  W  F  D

561   TGGTGTTGCTCCCTGGGCTGCTGAGTTTGATCAATGGCTTGAGGATGCTCTCCTAGAAGGAAGGTACGAA   630
        G  V  A  P  W  A  A  E  F  D  Q  W  L  E  D  A  L  L  E  G  R  Y  E

631   GATGTGAATAACTATCAAACAAAAGCACCAGAAGGGTGGAAGCTAGCACATCCAATTCCAGAACATTTTC   700
        D  V  N  N  Y  Q  T  K  A  P  E  G  W  K  L  A  H  P  I  P  E  H  F

701   TACCATTGCATGTAGCCATGGGTGCAGGTGGTGAGAAATCAAAGGCAGAGCTTATTTATCGTACTTGGGA   770
        L  P  L  H  V  A  M  G  A  G  G  E  K  S  K  A  E  L  I  Y  R  T  W  D

771   TCATGGTACTCTTGGCTACGCCTCTTACAAGTTCACTTCCATATGAtcatgctgaatttctctcactgtc   840
        H  G  T  L  G  Y  A  S  Y  K  F  T  S  I  *

841   tctcagctctcggggttgcctctcgtcgaccggctctatctctcaccatctgtctcaaaaaaa          903
```

Figure 28. *Pg DODA* cDNA and derived peptide sequences. Start codon is in bold and poly(A)-tail is cut before its end. Grey boxes indicate the 5 out of 6 conserved catalytic amino acids from LigB domains (pfam02900).

In-depth analysis of the *Pg* DODA amino acid sequence alignment with LigB-containing proteins and other plant homologues, enabled us to suppose that *Pg* DODA is an extradiol 4,5-dioxygenase (see Annex. 4). The crystal structure 1B4U of the LigAB protein (Sugimoto *et al.*, 1999) revealed the three amino acids coordinating the non-heme Fe^{2+} cofactor and the catalytic residues interacting with the protocatechuate (PCA) substrate (Figs. 30,31). The conservation of the iron-binding amino acid and of three out of five of the catalytic site amino acids is the main feature of interest (see Annex. 4). Referring to *P. grandiflora* 4,5-dioxygenase, the Ser16 (position A), the His121 (D) and the His177 (E), that orient the aromatic substrate, are conserved. Referring to PCA 4,5-dioxygenase structure, the Ser269 and Asn270, linked by hydrogen bonds to the two oxygen of the *PCA* lateral carboxylate group (Fig. 30), are conserved only in monocotyledons and *Lycopersicon* 1 (see Annex. 4). Substrates used by most plant homologues of the bacterial enzyme may not have a lateral carboxyl group or a modified one. The *Pg* DODA substrate DOPA possesses a longer (one carbon) lateral chain containing a $–NH_3$ group (Fig. 29). Thus, these two amino acids probably play a major role in substrate recognition.

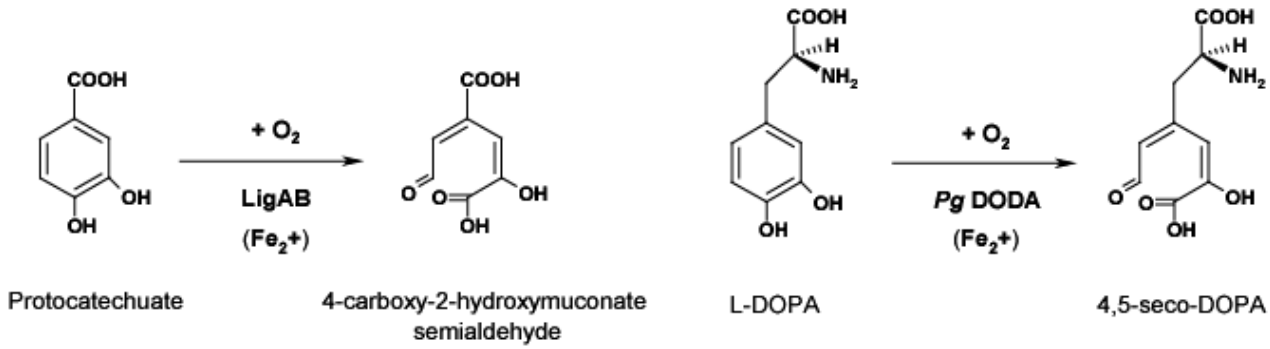


Figure 29. The L-DOPA substrate of *Pg* 4,5-DOPA extradiol dioxygenase DODA and the protocatechuate substrate for the bacterial LigAB differ in the complexity of the chain containing the carboxyl group (position 1).

The His17 (B), His55 (C) and Glu231 (F) coordinating the ferrous cofactor are conserved (Fig. 31), except that the latter is shifted one position and a conserved His232 (G) is present in its place (see Annex. 4), allowing as an alternative a possible triple His coordination of the Fe^{2+} ion as already observed in a soybean lipoxygenase (Boyington *et al.*, 1993). Secondary structure predictions realized with PROF tool (Rost and Sander, 1993) shows a high degree of conservation between the *Pg* DODA and the *PCA* LigAB dioxygenase secondary structure confirming our hypothesis (Annex. 4). There is no structural analogy of *Pg* DODA with the *Amanita muscaria* protein DODA.

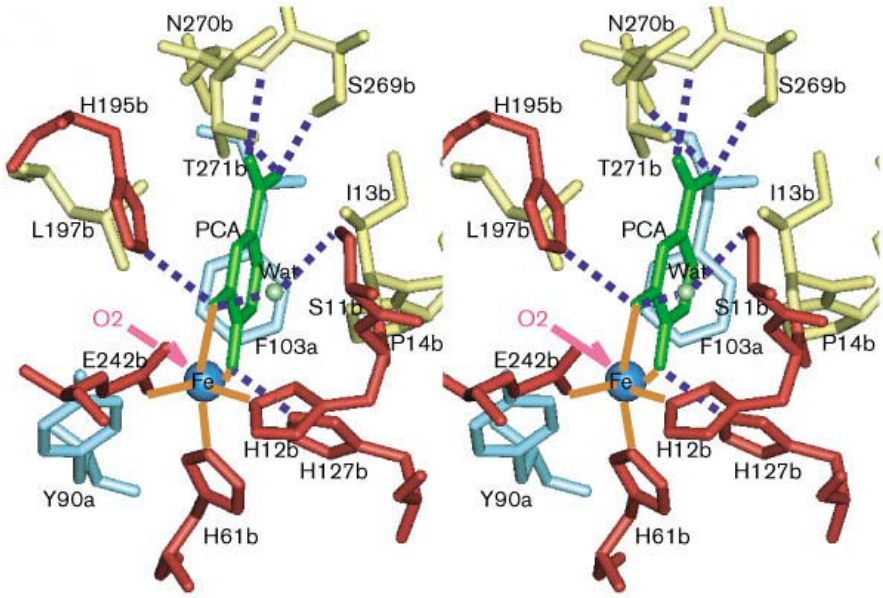


Figure 30. The active site of the LigAB-PCA complex shown in stereo (Sugimoto *et al.*, 1999). Amino acid residues completely conserved among the related class III extradiol dioxygenases are shown in red. α subunit residues are indicated in cyan. The PCA substrate (shown in green) makes six hydrogen bonds (dotted lines) with the surrounding residues. Coordination bonds to the Fe ion are coloured in orange. The putative O_2-binding site (shown in pink arrow) is presumed to be the fourth equatorial ligand in the tetragonal bipyramidal coordination sphere.

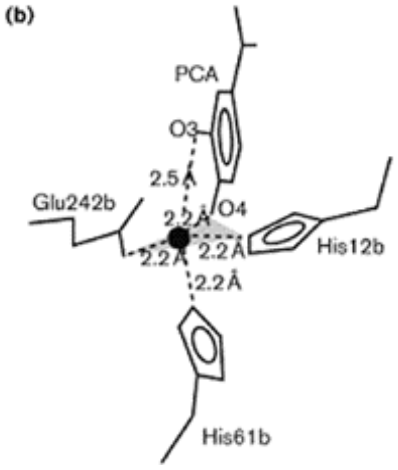


Figure 31. Representation of the iron coordination sphere in the presence of the PCA substrate (Sugimoto *et al.*, 1999).

2.2.2 Reevaluation of *DODA* transcriptional expression

As the putative extradiol dioxygenase function of *DODA* corresponded exactly to that expected for DOPA-dioxygenase gene *C* (Trezzini, 1990a), and all catalytic amino acids linked with the aromatic ring and iron-binding residues were conserved, we decided to examine again the expression profile of *L.6* mRNA.

According to Zaiko (2000), this promising candidate had transcripts present in all pigmented tissues, except the violet stems. We know that the level of betalain-specific transcripts reaches its maximum in immature pigmented petals and decreases during petal development. However, Zaiko detected no difference between the level expression of *DODA* in immature coloured petals and mature coloured petals, thus concluding that this gene was rather a coloured petal-specific gene than a betalain-specific gene.

From our previous work with *A.16*, we knew that the detection of transcripts in violet stems was difficult due to the restricted localization of pigment in the thin epidermal layer. So we had to use a higher amount of total RNA than usually to detect a signal. Thus, a novel expression analysis was performed with a lot of attention to the quantity of mRNA deposited in each line to verify the expected lower amount of transcript in mature pigmented tissues.

DODA has transcripts present in all pigmented tissues and absent from white petals, green stems and leaves (see Fig. 32). Its level of expression is very high in immature coloured petals and decreases in mature coloured petals. As expected, the level of expression in coloured stems is lower due to the localization of betalains restricted to the epidermal cell layer. Pooling these results together raised the expectation that this cDNA could well originate from the colour gene *C* necessary for the biosynthesis of the chromophore betalamic acid.

Similar results were obtained by Northern blot analysis of homologue in *Beta vulgaris* (red beet) callus. Green callus do not show any expression, but positive results were obtained in yellow, orange and red calli (Fig. 33).

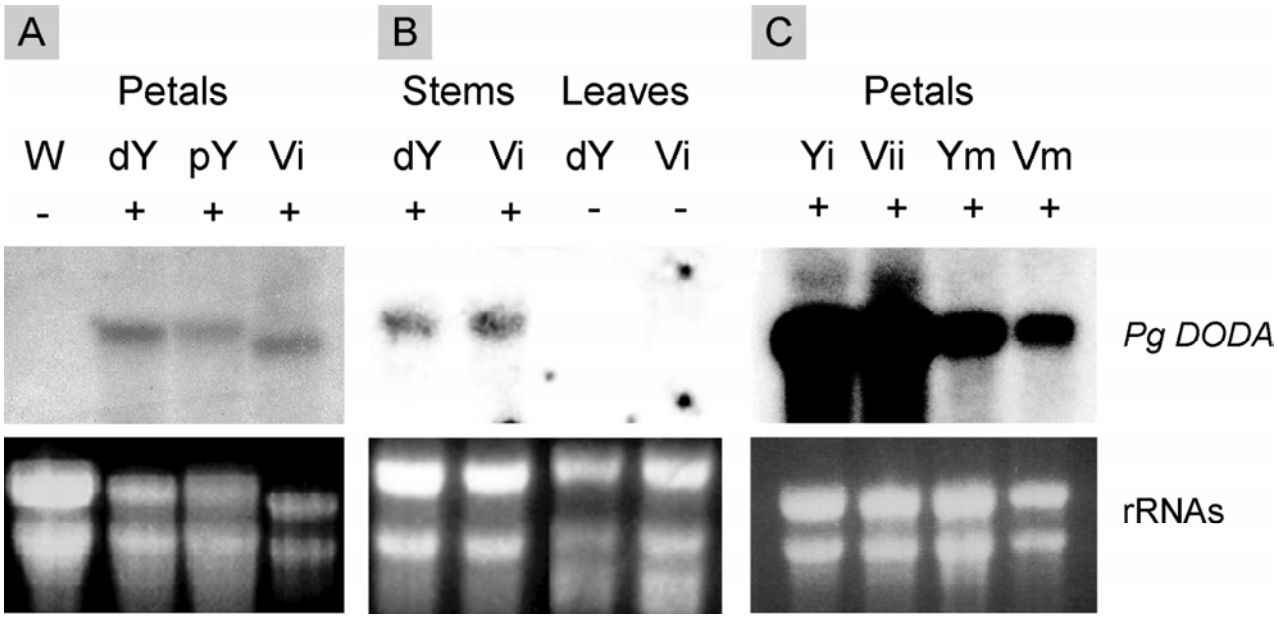


Figure 32. RNA gel blot analysis of DODA gene expression in *P. grandiflora* flowers. Presence or absence of betalain pigments (no *Pg* DODA) in plant tissue is indicated by (+) or (-) signs. **(A)** ***Pg*** petals of different genotypes (**W**, white cc, --, --; **dY**, deep yellow C-, rr, ii; **pY**, pale yellow C-, rr, I-; **Vi**, violet C-, R-,ii). **(B)** ***Pg*** stems and leaves. **(C)** Expression at different stages of bud development (with **Yi** yellow immature, **Ym** yellow mature, **Vii** violet immature and **Vim** violet mature).

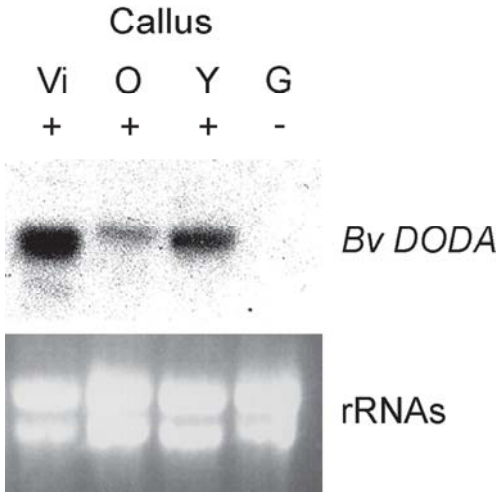


Figure 33. RNA gel blot analysis of expression of the putative *Beta vulgaris* (*Bv*) *DODA* in *Bv* coloured calli containing betalains and green calli lacking betalain (**Vi** violet, **O** orange, **Y** yellow and **G** green).

2.2.3 Immunodetection of DODA dioxygenase in *Portulaca grandiflora*

In order to verify if the expression of DODA protein is correlated with the synthesis of betalains in *Portulaca grandiflora*, Rabbit *Pg* DODA polyclonal antibodies were produced by injection of a synthetic peptide selected from *Pg* DODA sequence for its antigenic properties and specificity (Eurogentec®, Seraing, Belgium). The obtained serum was purified on a protein A affinity-column to decrease the background noise during immunodetection.

Western blot analysis was performed with crude plant protein extracts from *Pg* white, yellow and violet petals. DOPA-dioxygenase gene *C* is a cytoplasmic protein (Trezzini, 1990a), and on our side, we did not find any peptide signal sequence in DODA translated protein. Thus, we expected in SDS denaturing conditions a band of 30 kDa for DODA corresponding to the molecular weight predicted for DODA translated protein. 20 µg of each protein sample were loaded on a SDS-Page gel and blotted on nitrocellulose membrane. Immunodetection was done with the Immun-Star HRP kit (Bio-Rad®), *Pg* DODA specific antibody as primary antibody and an anti-rabbit IgG HRP-conjugate antibody as secondary antibody. Detection of the HRP-conjugate antibody was performed with the chemiluminescent Immune-Star HRP reagents mix.

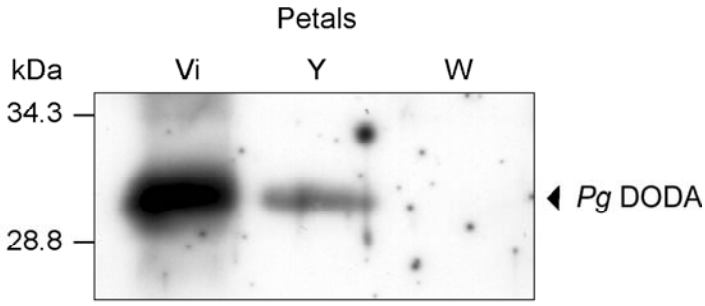


Figure 34. Western blot analysis of DODA in *P. grandiflora* petals of different colours (**Vi** violet; **Y** yellow; **W** white). *Pg* DODA (29.9 kDa) is present in high amount in violet petals, less in yellow petals and absent from non-pigmented white petals.

Immunodetection analysis confirmed the expression of DODA protein in violet and yellow *Pg* petals at the expected size and its absence from the non-pigmented white petals (Fig. 34). Since the size of the protein is equivalent to that predicted, we conclude that there is no major size-altering post-translational modification; this supports the cytoplasmic localization of DODA protein as predicted by PROF tool and the "three-gene model" (Trezzini, 1990a).

2.3 Functional complementation of the betalain biosynthetic pathway by DODA in *P. grandiflora* white petals

2.3.1 Introduction

Among several betalain-specific cDNAs isolated by subtractive hybridization, we have identified by sequence analysis *DODA* gene possessing an open reading frame homologous to the LigB extradiol dioxygenase from *Sphingomonas paucimobilis*. We showed by in depth analysis of its amino acid sequence that most of the LigB catalytic residues were conserved, thus confirming the preliminary result obtained by alignment. Finally, we demonstrated that the expression of *DODA* mRNA in different *P. grandiflora* tissues corresponds well to the profile envisaged for a candidate to the DOPA-dioxygenase gene *C* position, as predicted by the "three-gene model" for betalain biosynthesis in *P. grandiflora* (Trezzini, 1990a).

In order to confirm *Pg* DODA protein as an extradiol dioxygenase responsible for the conversion of DOPA into betalamic acid, we had to demonstrate that this enzyme is able to synthesise betalain pigments. This can be tested in vitro with an enzymatic assay or in planta by transient or stable transformation approach. As we had no information about the conditions of the dioxygenation reaction and the existence of other molecules participating to the reaction, it was safer to try an in planta experimental approach, corresponding more to the physiological environment of the enzyme. We know that *Pg* white flower phenotype is deficient in the gene *C* corresponding to the DOPA-dioxygenase (Trezzini, 1990b). Thus, successful biochemical complementation through transformation by a construct containing a putative gene candidate would constitute a proof that this gene encodes a specific dioxygenase. It was already demonstrated that shot gun (biolistic) transformation of white petals with a construct containing the *Amanita DODA* gene complemented pigment production (Mueller *et al.*, 1997a). Depending on the plant genetic background, (*cc*, *rr*, *ii*) or (*cc*, *R*-, *I*-), transformed cells produced yellow betaxanthin or violet betacyanin pigments (Trezzini, 1990b). Stable transformation of *Pg* plant would have been a possible approach (Rossi-Hassani, 1991), nevertheless due to difficulties to regenerate a sufficient number of transformants and the necessity to test several betalain-specific candidates together, the choice of the biolistic transient transformation method was done.

2.3.2 Complementation by biolistic strategy

We decided to follow the successful biolistic approach already developed in our lab. Therefore we subcloned *Pg DODA* full-length cDNA into CaMV35S pNco plant expression vector previously used for biolistic transformation with *Amanita DODA* cDNA (Mueller *et al.*, 1997a) and obtained the new pNco PgDOD construct (Fig. 35). Untranslated extremities (UTR) of *DODA* cDNA have been modified by proofreading PCR before the subcloning step. 5'-UTR residues adjacent to the start codon were modified to correspond to the Kozak translation initiation recognition sequence to guarantee a correct in frame expression of DODA protein (Kozak, 1991). *Xba*I and *Pst*I endonuclease restriction sites were also added at both extremities for cloning facilities. Finally, pNco PgDOD construct was double-sequenced to control its sequence integrity and the presence of a correct Kozak sequence.

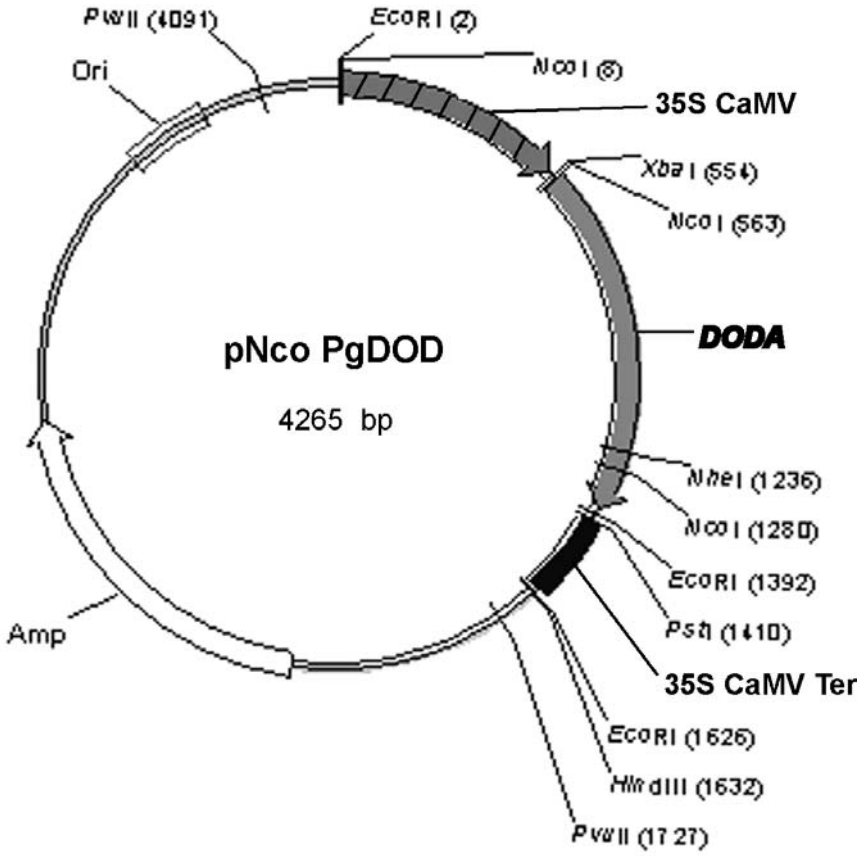


Figure 35. Map of pNco PgDOD vector expressing *P. grandiflora* DOPA-dioxygenase DODA under the control of 35S CaMV promoter and terminator (Ter). *Pg DODA* was cloned in place of *Amanita DODA* into pNco DOD (Mueller *et al.*, 1997a).

In order to determine the optimum parameters for the biolistic transformation method, the GUS gene reporter (Jefferson *et al.*, 1987) was introduced into a similar pNco expression vector (Mueller *et al.*, 1997a). Thereafter, this construct was used as a positive control to verify that all parameters were correct during the transformation experiment. Thus, pNco GUS construct was used simultaneously with betalain specific tested candidates. Nevertheless, as the GUS revelation system destroys the physiology of the transformed petals, we had to wait during one or two days for the appearance of betalain biosynthesis before doing the GUS test. Therefore we decided to replace GUS gene by a fluorescent gene reporter to avoid this disadvantage. We chose the *DsRed2* gene reporter (Clontech) because it emits a strong red fluorescent signal, which can be observed easily in plant tissue and for a longer time than that from the Green Fluorescent Protein. p35S DsRed2 NOS plant expression vector was constructed in our lab by Andrija Finka (unpublished) (Fig. 36). Usually 50 to 100 red fluorescent spots were observed per transformed *P. grandiflora* petal.

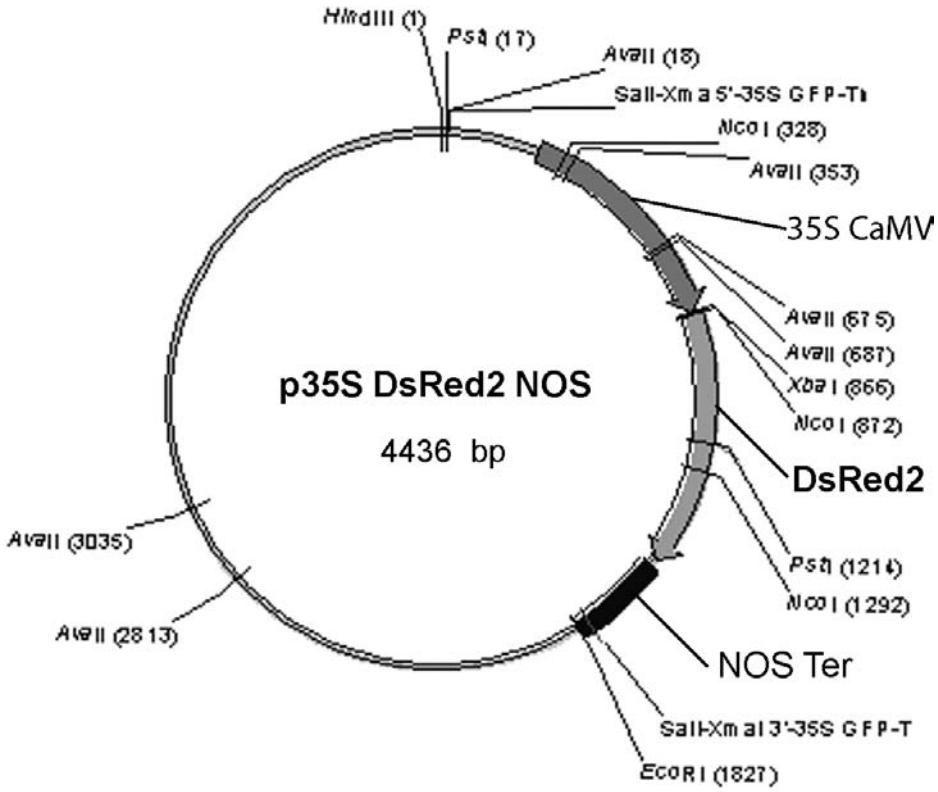


Figure 36. The plant expression vector p35S DsRed2 NOS expressing DsRED2 from *Discosoma sp*ecies. DsRed2 is under the control of 35S CaMV promoter and NOS terminator (NOS Ter) cloned into a pUC18 vector. This vector was constructed by cloning *DsRed2* gene in place of *GFP-talin* gene in pYSC14 (Kost *et al.*, 1998).

Other important points for the success of the experiment were the quality of the *Pg* white petals, the medium osmolarity and the shooting parameters. We know that *P. grandiflora* flowers have a short one-day life, and white petals have a maximum of DOPA concentration just before the opening of the buds (Trezzini, 1990a). In order to have optimal plant material, petals were delicately extracted from bud one day before their normal opening and transferred in a Petri dish containing a plant agar medium.

The physical parameters which have to be adjusted for the shooting are the helium pressure, the medium osmolarity, the distance between the opening of the particle gun and the petal sample and finally the distance between the opening of the particle gun and a filter necessary for a correct dispersion of the gold particles on the whole petal sample. We decided to keep the optimum parameters determined by Mueller and colleagues (1997a) as described below:

Medium osmolarity:	250 mOS
Helium pressure:	6.5 bar
Sample distance:	11.5 cm
Filter distance:	6.5 cm

2.3.3 Results of the transformation *of Portulaca grandiflora* white petals by *Pg* DODA

Transient overexpression of *Pg DODA* gene by shot gun transformation of white *Pg* petals produced, after 12 to 18 hours of incubation, bright yellow betaxanthin spots (Fig. 37A), or deep violet betacyanin spots (Fig. 37D), according to the plant genetic background (Trezzini, 1990b). Each spot is due to one individual cell (Fig. 37B, E). We observed no diffusion of the pigment during the next two days, what differs from the results obtained with *Amanita* DODA (Mueller *et al.*, 1997a). We counted 5 to 80 spots per petal, depending on the shot, petal size and maturity. More spots were visualized on fully developed petals. We used a p35S DsRed2 NOS expression construct as a positive control in the same transformation. All violet and yellow transformed cells also produced the DsRed2 red fluorescent protein, whereas some cells only emitted the red fluorescence. White petals transformed by DsRed2 alone, as a negative control, did not show any yellow or violet coloured spots.

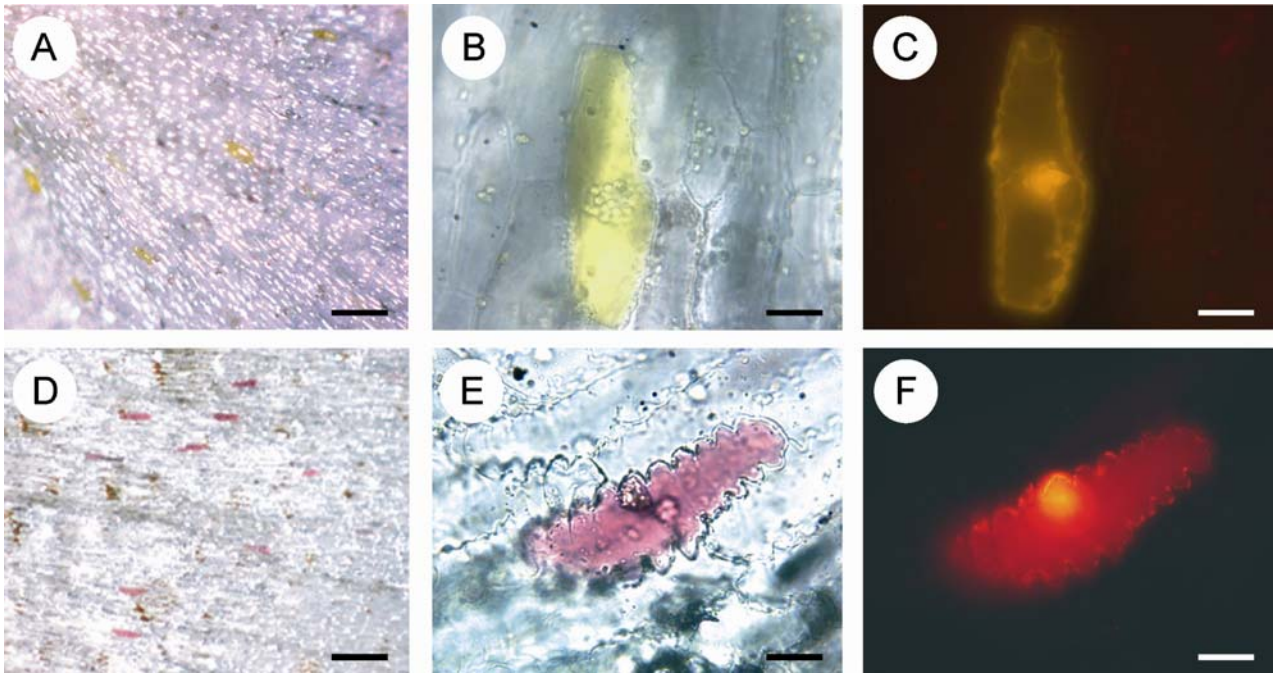


Figure 37. The specificity of the *Pg DODA* gene is demonstrated by biolistic complementation of the betalain pathway in the white petals of *Portulaca grandiflora* plants deficient in *DODA*. A pNco PgDOD expression vector containing full-length *Pg DODA* driven by CaMV35S promoter has been used with p35S DsRed2 NOS vector as a positive control. (A) Yellow spots revealed after biolistic transformation of a white petal from a plant with yellow genetic background. (B) Close up of a cell accumulating betaxanthins in its vacuole. (C) The same cell displaying the DsRed2 fluorescent protein modified to an orange one by the fluorescence of the betaxanthins. (D) Violet spots revealed in a white petal from a plant with a violet genetic background. (E) Close up of a cell accumulating betacyanins. (F) The same cell displaying the DsRed2 fluorescent protein. The red fluorescence hue from DsRed2 was slightly modified by filtration through yellow betaxanthin or violet betacyanin pigments (allowing for colour differences between C and F). Bars = 200 μm (A, D), 20 μm (B, C, E, F).

To check the identity of the pigments responsible for cell colouration, we collected about 500 yellow and 500 violet transformed cells. Pigments were extracted in acidified methanol and compared by HPLC analysis (Trezzini and Zryd, 1991a) for their spectra and retention times with naturally occurring betalain pigments from deep yellow and violet *P. grandiflora* petals as controls. These pigments showed a strict identity with the betalain pigments present naturally in violet and deep yellow *Pg* petals (Fig. 38A, B). We identified the violet pigment as betanin (λmax 536 nm) and the yellow pigment as dopaxanthin (λmax 476 nm) by comparison with standards. Both pigments have either DOPA or dopamine as amino conjugate, which are strongly present in *Pg* white petals.The quantity of dopaxanthin measured in the yellow-transformed cells was lower per cell than that present in untransformed cells of the yellow phenotype. This could be due to the synthesis of other betaxanthins in minor quantities, which were not measurable. We did not detect the presence of muscaflavin (λmax 424 nm), which would indicate a 2,3-dioxygenase activity.

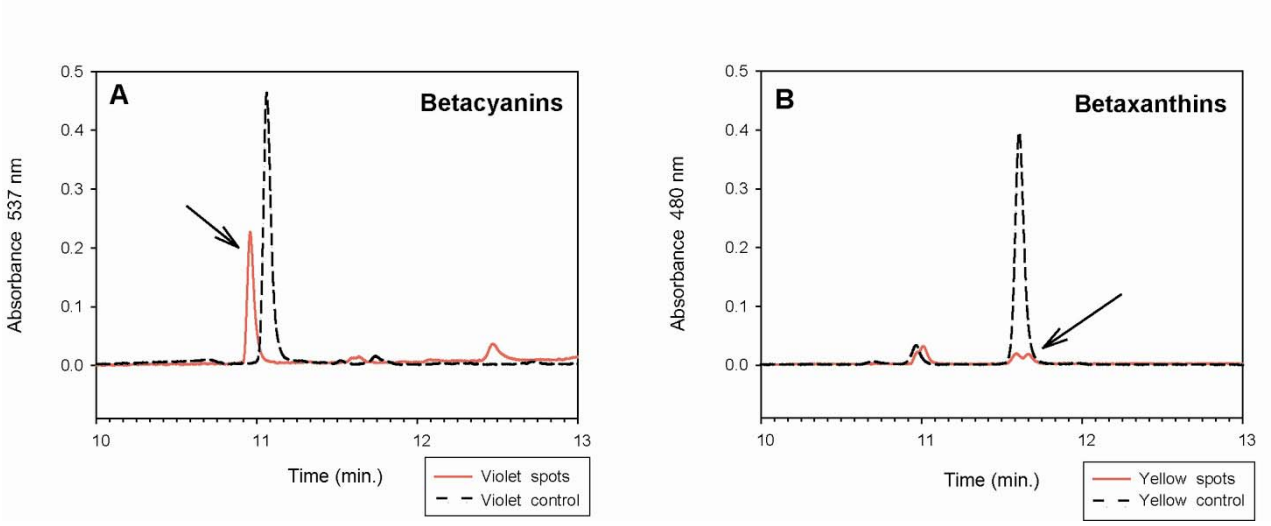


Figure 38. HPLC analysis of betalain pigments extracted from violet- (A) and yellow-transformed (B) cells in the white *P. grandiflora* background. These pigments were identified by comparing their elution profile with elution profile of the pigments extracted from violet and deep yellow *P. grandiflora* petals, respectively. Arrows indicate the major peaks of the violet betanin (A) and of the yellow dopaxanthin (B). the minor 3-s shift observed between the two betanin peaks is due to a slight inaccuracy of the injection process.

These results demonstrate the biochemical complementation of the missing C gene product in *Pg* white petals by the *Pg DODA* gene product. Our identification of *Pg* DODA *in silico* as an extradiol 4,5-dioxygenase is thus confirmed and we showed that, in contrary to *Amanita* DODA, the plant dioxygenase has no significant 2,3- ring cleaving activity. DODA represents a new and previously unidentified family of plant protein and the first example of a non-heme ring-opening dioxygenase from plant.

3. Phylogenetic analysis of P. grandiflora DODA and origins of betalains

3.1 Introduction

Betalains are only present in a small number of species from two different kingdoms, plants and fungi, whereas they are completely absent from animals and bacteria. This raises the general question of the evolutionary significance of betalain biosynthesis. Do the presence of these pigments constitutes a real evolutionary advantage for some plant and fungal species? Have betalains played similar role in the biology of plant and fungi? Before answering these two questions, we have to solve the question of the origin of betalain biosynthesis pathway. Does it result from a single event or from several independent events. Therefore, we will first compare the sequence of the plant DOPA-dioxygenase DODA with that of the fungal DOPA-dioxygenase.

Betalains play a major role in the flower colouration of the Caryophyllales plant order, where they replace the widespread anthocyanin pigments with the exception of the Caryophyllaceae and Molluginaceae families. Why did most species of this order develop the production of betalain pigments? Is it the consequence of the loss of the anthocyanin biosynthesis pathway in these species or only the development of another type of pigment in parallel with anthocyanins, prior to flower pigmentation?

Genes homologous to *Pg DODA* are expressed in most plant species. All these sequences show also a high degree of conservation when they are translated. The existence of these homologous proteins raises the query of their specific function in plant in absence of the betalain biosynthesis pathway. Is it possible to identify some conserved patterns in the primary sequence of DODA from betalain-producing plants or from non-betalain-containing plants? Therefore, we will try to widen our knowledge of the variability of the DOPA-dioxygenase peptidic sequence by isolating new homologous sequences from representative plant species and by making a phylogenetic analysis.

3.2 Biochemical evolutionary convergence of betalains in plants and fungi

3.2.1 Introduction

Comparative biochemical analysis of the betalain pathway in plants and fungi revealed the existence of a common reaction, the transformation of DOPA into betalamic acid. Consequently to the results we obtained with *P. grandiflora* DODA, we know that this reaction is catalysed in both cases by an extradiol 4,5-dioxygenase (Mueller *et al.*, 1997a; Christinet *et al.*, 2004). As we already suggested, it is surprising that such distant organisms synthesise the same pigments exceptionally found in both kingdoms. With the isolation of the plant DOPA-dioxygenase, we can now reduce this question to the problem of the origin of the DOPA-dioxygenases participating to the last key step of betalain chromophore synthesis. Thus two choices are available:

First, in the case of a plant and a fungal DODA protein with distinct origins, both proteins would have followed a convergent evolution, conferring them the same function. Such convergent evolution of similar enzymatic function on different protein folds was observed among the sugar kinases by comparison of the hexokinase, ribokinase and galactokinase protein families (Bork *et al.*, 1993). Each of these three families of sugar kinases appears to have a distinct three-dimensional fold, since conserved sequence patterns are strikingly different for the three families. However, each catalyses chemically equivalent reactions on similar or identical substrates. Thus, the enzymatic function of sugar phosphorylation is a good example of an independent evolution based on three distinct structural frameworks.

The second hypothesis supposes a minimum of conservation between the amino acid sequences of both enzymes. If plant and fungal DOPA-dioxygenases have had a common origin consequently to a horizontal gene transfer or the existence of a common ancestor, they would have conserved the same function. This conservation could be global or localized on a particular structural domain or only on catalytic amino acids. Nevertheless, many labs did not succeed in isolating a plant homologue from the *Amanita muscaria* DOPA-dioxygenase by either antibody screening or PCR amplification strategies. Thus, they concluded that fungal and plant DOPA-dioxygenase sequences were rather different. Moreover the observed bifunctional 3,4 and 4,5 ring-opening activity of the fungal enzyme and the absence of an homologue in the databases suggested that this enzyme was specific to fungi (Mueller *et al.*, 1997b). The muscaflavin compounds derived from the 3,4

aromatic cleavage were never identified in *P. grandiflora* or other betalain plants. Thus, we propose to compare the peptidic sequence of *Am* DODA with that of *Pg* DODA and its homologues in non-betalain-containing plants. This will surely reveal if they have a distinct or common origin.

3.2.2 Comparison of *P. grandiflora* and *Amanita* 4,5-DOPA-dioxygenases and discussion

Pg DODA and *Am* DODA 4,5-DOPA-dioxygenases have different amino acid composition, molecular weight, and isoelectric point and do not share a similar hydrophobic profile (Fig. 39). Finally, it was not possible to align together their nucleotidic or peptidic sequences with the different algorithms existing, neither to find the catalytic histidines conserved in the LigB domain (pfam02900) present in *Pg* DODA. Thus, *Amanita* DOPA-dioxygenase does not correspond to a type III extradiol dioxygenase like *Pg* DODA does. In conclusion, *Am* DODA is a new type of extradiol dioxygenase with a distinct origin from the plant DOPA-dioxygenase. This fact is supported by our recent observation that both enzymes have a different homologue in *Pseudomonas fluorescens*, definitively confirming the convergent evolution of betalain pathway in plants and fungi (see Figs. 40-41).

***Pg* DODA**	
MW	29.9
pI	5.6

***Am* DODA**	
MW	26.2
pI	6.6

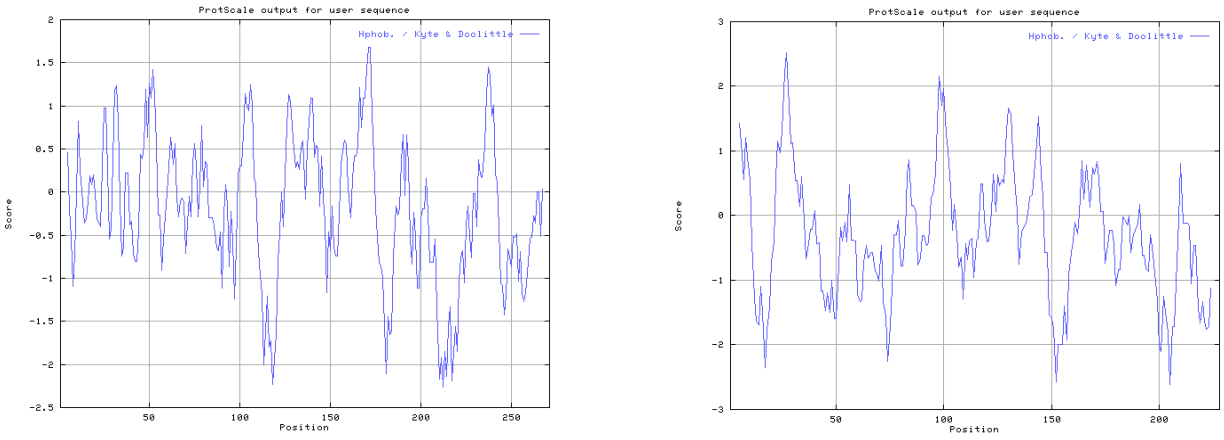


Figure 39. Comparison of *Pg* DODA and *Am* DODA physical parameters. Hydropathy plots according to Kyte and Doolittle (1982). Notes that scales are different.

```
Pg DODA            MGVGKEVSFKESFFLSHGNPAMLADESFIARNFLLGWKKNVFPVKPKSILVVSAHWETDV
P. fluorescens     -------MFP-SLFISHGSP-MLALEP-GASGPALARLAAELPR-PKAIIIVSAHWESHE
                          *  *:*:***.* *** *.  * .  *.     :*  **:*::******:.

Pg DODA            PCVSAGQYPNVIYDFTEVPASMFQMKYPAPGCPKLAKRVQELLIAGGFKSAKLDEERGFD
P. fluorescens     LLVSSHPQPETWHDFGGFPRALFEVQYPAPGNPQLAREVADLLTANNLP-ARLDPQRPFD
                     **:   *:. :**  .* ::*::::***** *:**:.* :** *..:  *:** :* **

Pg DODA            HSSWVPLSMMCPEADIPVCQLSVQPGLDATHHFN-VGRALAPLKGEGVLFIGSGGAVHPS
P. fluorescens     HGVWVPLSLMYPQADIPVVQVSL-PSRGGPALQNRVGQALARLREQGILLIGSGSITHNL
                   *. *****:* *:***** *:*: *. ...   * **:*** *: :*:*:****. .*

Pg DODA            DDTPHWFDG---VAPWAAEFDQWLEDALLEGRYEDVNNYQTKAPEGWKLAHPIPEHFLPL
P. fluorescens     RELD-WHAGPESVEPWARDFRDWIVDKLAANDEAALHDYRQQAPNAVR-SHPSDEHLLPL
                    :   *. *   * *** :* :*: * *  .    :::*: :**:. : :**  **:***

Pg DODA            HVAMGAGGEKSKAELIYRTWDHGTLGYASYKFTSI
P. fluorescens     YFARGAGGDFS---VAHQGFTMGALGMDIYRFD--
                   :.* ****: *   : :: :  *:**   *:*
```

Figure 40. Clustal W alignment of *Pg* DODA with its homologue in *Pseudomonas fluorescens* (ZP_00083959). Both proteins share 40% of identity and 57% of similarity at the amino acid level. LigB catalytic amino acids conserved in *Pg* DODA (grey boxes) are also present in ZP_00083959 protein.

```
Am DODA            KEWHFHIYFHQNNAAEHQAALELRDAVLRLRQDGAFVAVPLFRVNMDPMGPHPVGSYEIW
P. fluorescens     KGYHAHVYYDASTIKQARALCE---------QAAALFPLKMGRVHERPVGPHPDWSCQLA
                   * :* *:*:. ..  : :*  *         * .*:..: : **:  *:****  * ::

Am DODA            VPSETFASVFSYLCMNRGRLSILVHPLTREELRDHEIRNAWIGPSFPLNLANLPIKSDEI
P. fluorescens     FGPELIGDVLPWLALNRKGLVVFLHPDTGDDLLDHTEHAIWMGAMRPLDLSVF-------
                   . .* :..*:.:*.:**  * :::** * ::* **  :  *:*.  **:*: :
```

Figure 41. Clustal W alignment of *Am* DODA with a similar *Pseudomonas fluorescens* protein (ZP_00083152). Both proteins share 30% of identity and 48 % of similarity at the amino acid level.

The reasons for the synthesis of betalains in fungi are still unsolved. Till now, no serious study has been published concerning this question. Some analogies can be made with the general role of pigments in plants and animal. The colour of an organism is a structural adaptation that helps it to survive in a particular environment. Plants and animals use colour and patterns in different ways. Colour is frequently used in order to be dissimulated in a particular environment, which is certainly not the case for the red *A. muscaria*. Another aim of the colouration is to draw the attention of insect or animal species, either to attract them, like bees for pollination, or to warn them of possible danger, constituting a passive defence strategy. There is no known evidence of such attraction or repulsion strategy in *Amanita* species, but other fungi mimic bloody wounds to attract dipters during their sporulation phase. Thus, these insects take parts to the dissemination of fungi spores.

In this context, it is interesting to note that the different species of the genera *Amanita* and *Hygrocybe* do not contain the same DOPA derivatives. *A*manita *cesarea* contains the same pigments as *A. muscaria* with an exception of muscaurin 2. *A. pantherina* and *A. strobiliformis* possess the stizolobic and stizolobinic acids (Saito *et al.*, 1978), but no muscaflavin or betalamic acid. In *A. citrina*, *A. rubescens*, *A. spissa* and *A. fulva*, no pigment derived from the DOPA have been found. Finally, in *Hygrocybe* pigments derived from the muscaflavin are present, but not those derived from the betalamic acid (Musso, 1979; Terradas and Wyler, 1991a, b).

Thus, DOPA is not used in the same way in all *Amanita* or *Hygrocybe* species. The spectrum of affinity of *Am* DODA is very wide. This enzyme catalyses dioxygenation of catechol, as well as caffeic acid, but does not catalyse the formation of betalamic acid from DOPA in an optimal manner (Girod and Zryd, 1991a; Terradas and Wyler, 1991a). This is frequently the case when the enzyme has no definitive well-established structure or function. In fact, betaxanthin biosynthesis does not appear as a highly stable pathway in fungi, because these pigments are only present in two species. It is possible that betaxanthins have still not found a specific function and therefore could be considered as an available factor of variability in fungi.

Another similar case of convergent evolution between plants and fungi has been observed for the gibberellin (GAs) biosynthesis (Hedden *et al.*, 2001). GAs present in both kingdoms were first discovered in the fungus *Gibberella fujikuroi*, from which gibberellic acid (GA(3)) and other GAs are produced commercially. Although higher plants and fungi produce structurally identical GAs, there are important differences in the pathways and enzymes involved. Even where cytochromes

P450 have equivalent functions in plants and *Gibberella species*, they are unrelated in terms of amino acid sequence. This fact becomes particularly obvious with the identification of almost all the genes for GA-biosynthesis in *Arabidopsis thaliana* and *G. fujikuroi*, following the sequencing of the *Arabidopsis* genome and the detection of a GA-biosynthesis gene cluster in the fungus *G. fujikuroi* (Tudzynski and Holter, 1998). For example, 3-beta-hydroxylation occurs early in the pathway in G. fujikuroi and is catalysed by a cytochrome P450 monooxygenase, whereas it is usually the final step in plants and is catalysed by 2-oxoglutarate-dependent dioxygenases. Similarly, 20-oxidation is catalysed by dioxygenases in plants and a cytochrome P450 in fungi. Even where cytochrome P450s have equivalent functions in plants and *Gibberella* fungi , they are unrelated in terms of amino acid sequence. These profound differences indicate that higher plants and fungi have evolved their complex biosynthetic pathways to GAs independently and not by a horizontal gene transfer.

In conclusion, betalain biosynthesis pathway is a clear example of convergent evolution observed between two kingdoms. It would be interesting to have a three-dimensional structure for *Amanita* DOPA-dioxygenase to verify our hypothesis that this functional convergence is correlated with a similarity of the catalytic site geometry of both plant and fungal 4,5-DOPA-dioxygenases. Unfortunately, *Amanita* DOPA-dioxygenase has for the moment no characterized homologue. *Am* DODA is a new type of extradiol 4,5-dioxygenase with a distinct origin from the plant DOPA-dioxygenase.

3.3 Appearance of betalains in plants

3.3.1 Introduction

The question of the appearance of betalains in plants is more disputed than their presence within fungi. In plants, betalains are synthesised only in one order, the Caryophyllales, whereas anthocyanins are largely widespread. These two pigments are supposed to exclude themselves mutually because no species containing both pigments was found till now. In the Caryophyllales, only two families, the Caryophyllaceae and the Molluginaceae accumulate anthocyanins. Therefore, it has been suggested that the betalain- and anthocyanin-producing lineages diverged prior to the origin of red-violet floral pigmentation (Mabry, 1973). However, the flavonoid pathway, leading to the synthesis of anthocyanins, was interrupted in all betalain plants (Bate-Smith, 1962; Bittrich and Amaral, 1991). The transformation of leucoanthocyanidin in anthocyanidin is not possible due to the absence of the leucoantocyanidin synthase/dioxygenase enzyme. This was clearly demonstrated in betalain plants by the observation of an accumulation of leucoanthocyanidin, the anthocyanin precursor, and proanthocyanidin, a derived compound. Thus, Ehrendorfer (1976) postulated that the capacity to produce betalains evolved subsequently to the loss of anthocyanidin production in the betalain taxa. Anthocyanin synthesis was useless in arid or semi-arid environmental conditions with few pollinators and good conditions for anemophily, with low selective pressure for attracting insects. Later in less arid climate, pollinators were again present. In these climatic conditions, an ancestor of the betalain plants could have redeveloped new substitutive pigments, the betalains. However, recent work on the classification of the Caryophyllales does not support the basal position of anthocyanin-containing Caryophyllaceae and Molluginaceae plants (Cuenoud *et al.*, 2002). This being the case, the clear circumstances regarding the gain and loss of both independent pathways remains unsolved.

In this context, it is clear that an answer can only come from a phylogenetic analysis of a gene directly involved in betalain biosynthesis. Thus, after the identification and functional characterization of *P. grandiflora* DODA DOPA-dioxygenase, we decided to identify plant homologues in EST databases and to isolate them directly from plants of interest which genome have not been widely sequenced till now.

3.3.2 Isolation of *Pg* DODA homologues from betalain producing plants

Alignment of *Pg* DODA protein with betalain plant EST databases translated for the six frames allow to identify only one partial EST from *Beta vulgaris* (*Bv*) and two different from *Mesembryanthemum crystallinum* (*Mc*) as *Pg* DODA homologues. In non-Caryophyllales plants, many species had their genome completely or widely sequenced and therefore lots of ESTs were available for about twenty different species. In order to carry out a representative phylogenetic analysis of *Pg* DODA in the plant kingdom, it was necessary to isolate more homologues from betalain-containing plants chosen to give a complete representation of the Caryophyllales phylogeny. For this purpose, we decided to use a PCR amplification strategy of *Pg* DODA homologous genomic DNA sequences (gDNA) with nested degenerate primers designed on conserved parts of the *DODA* cDNA from betalain-producing plants. In order to determine a consensus sequence for the *DODA* cDNA, it was initially necessary to complete the *Beta vulgaris* homologous sequence.

The isolation of the *Beta vulgaris* full-length *DODA* cDNA was made from the sequence information of the partial EST BI095902. Total RNA was isolated from betalain producing *Bv* callus, and cDNA was obtained by reverse transcription of this total RNA. A rapid amplification of the cDNA ends (RACE) was used to elongate the 5' part of the partial *Bv* EST. For this purpose, Genome walker adaptators (Clontech) were added to the ends of the cDNAs. A nested PCR strategy was used with the two long gene-specific primers (*Bv*p1, *Bv*p3) and the two primers present on the adaptators (AP1, AP2) in order to allow a specific amplification of the cDNA as previously explained for *A.16* cDNA in chapter II.1.4 (Fig. 14).

The resulting sequence is shown in Fig. 42. The 1.2 kb-long *Bv DODA* cDNA contains an 810-bps length ORF and 3' and 5'-untranslated regions. It encodes a 269 amino acid-long protein with a predicted molecular weight of 29.6 kDa and an isolelectric point of 5.98, both parameters are similar to those of *Pg* DODA. Amino-acid sequence comparison with *Pg* DODA gives 61% of identity and 72% of similarity. All LigB catalytic amino acids conserved in *Pg* DODA were also present in *Bv* DODA translated protein as shown in Fig. 42. Both proteins have the same hydropathy plots (see Fig. 43). However Functional complementation of *Pg* white petals by a pNco *Bv* DODA vector failed.

Due to *Bv* DODA isolation, it was possible to determine a consensus sequence from the alignment of *Pg* DODA, *Bv* DODA and the *Mesembryanthemum crystallinum* ESTs BF480453 and BE131205. We have chosen to define one sense and two antisense "universal" primers designed on the two highest conserved sequence areas from these four cDNAs and included a few degenerated positions (L6Univsense, L6UnivAs1, L6UnivAs2). One disadvantage of this method is the production of a truncated genomic fragment because the extremities of *DODA* cDNA were too divergent to find a consensus sequence. Nevertheless, the extremities were not necessary for this phylogenetic study.

The following plants were chosen for the analysis:

- *Phytolacca americana*
- *Mammilaria* sp.
- *Iresine* sp.
- *Bougainvillea*
- *Dianthus caryophyllus* (non-betalain Caryophyllales)
- *Silene vulgaris* (non-betalain Caryophyllales)

and for the controls :

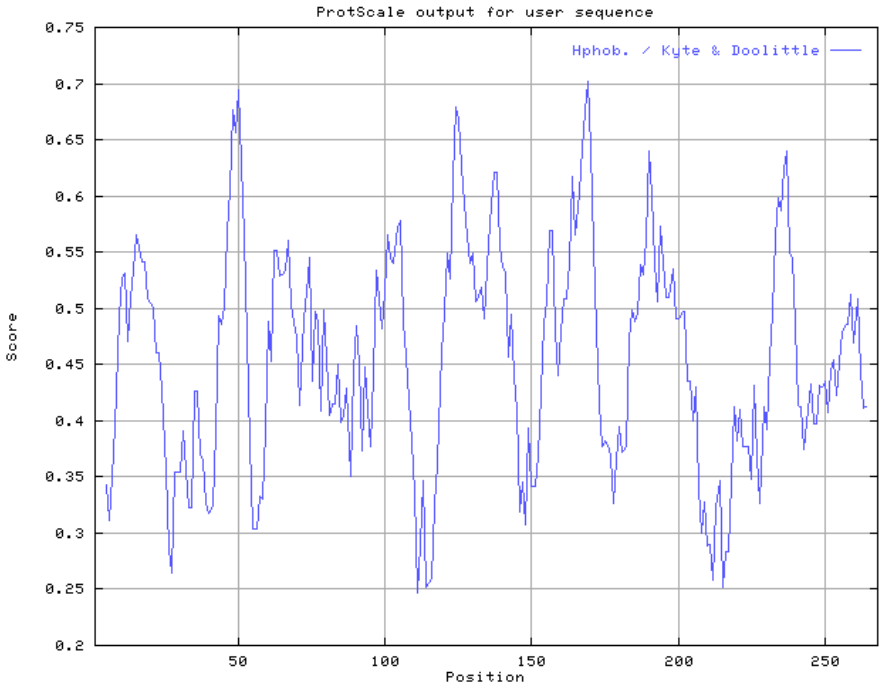
- *Arabidopsis thaliana* (non-Caryophyllales, negative control)
- *Portulaca grandiflora* (+ control)
- *Beta vulgaris* (+ control)
- *Mesembryanthemum crystallinum* (+ control)

They correspond to controls and to a representative choice of the missing main groups of the Caryophyllales classification according to Clement and Mabry (1996).

```
-136                                    CTATGCATCCAACGCGTTGGGAGCTC  -111
-110  TCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCACTAGTGATTCAATCTCAAC  -56
 -55  CCCAGAAATCAACAGAAGAACATTACTCAAATTCCTAATTAGCTCAAATCTGAAA  -1

   1  ATGGGTAGTGAAGATAACATCAAAGAAACCTTCTTTATTTCTCATGGAACTCCAA   55
        M  G  S  E  D  N  I  K  E  T  F  F  I  S  H  G  T  P

  56  TGATGGCAATTGATGATTCAAAGCCTTCTAAAAAATTCCTTGAAAGTTGGAGAGA  110
      M  M  A  I  D  D  S  K  P  S  K  K  F  L  E  S  W  R  E

 111  GAAAATCTTCTCTAAAAAACCCAAGGCTATTCTCGTCATTTCTGCTCACTGGGAA  165
        K  I  F  S  K  K  P  K  A  I  L  V  I  S  A  H  W  E

 166  ACTGATCAACCTTCTGTTAATGTTGTAGACATCAATGATACCATCTATGATTTCA  220
        T  D  Q  P  S  V  N  V  V  D  I  N  D  T  I  Y  D  F

 221  GAGGTTTTCCTGCTCGTTTGTACCAGTTTAAGTACTCGGCTCCTGGATCTCCAGA  275
        R  G  F  P  A  R  L  Y  Q  F  K  Y  S  A  P  G  S  P  E

 276  ACTGGCAAACAGGATACAAGATCTTCTCGCAGGATCTGGCTTCAAATCTGTAAAC  330
        L  A  N  R  I  Q  D  L  L  A  G  S  G  F  K  S  V  N

 331  ACTGACAAGAAACGAGGGCTTGATCACGGTGCATGGGTGCCTCTAATGCTGATGT  385
        T  D  K  K  R  G  L  D  H  G  A  W  V  P  L  M  L  M

 386  ATCCTGAAGCTGACATCCCTGTTTGTCAGCTCTCTGTCCAATCGCATTTGGATGG  440
      Y  P  E  A  D  I  P  V  C  Q  L  S  V  Q  S  H  L  D  G

 441  AACACACCACTATAAGTTGGGACAAGCGTTGGCTCCTCTAAAGGATGAAGGTGTC  495
        T  H  H  Y  K  L  G  Q  A  L  A  P  L  K  D  E  G  V

 496  CTCATCATTGGTTCCGGAAGTGCAACACACCCTTCAAATGGTACCCCTCCCTGTT  550
        L  I  I  G  S  G  S  A  T  H  P  S  N  G  T  P  P  C

 551  CTGATGGAGTTGCTCCGTGGGCTGCAGCTTTTGATAGCTGGCTTGAAACAGCTCT  605
      S  D  G  V  A  P  W  A  A  A  F  D  S  W  L  E  T  A  L

 606  GACAAATGGAAGCTATGAAGAAGTGAACAAATATGAAACTAAAGCACCAAACTGG  660
        T  N  G  S  Y  E  E  V  N  K  Y  E  T  K  A  P  N  W

 661  AAATTGGCTCATCCTTGGCCTGAGCACTTTTATCCACTGCATGTAGCCATGGGTG  715
        K  L  A  H  P  W  P  E  H  F  Y  P  L  H  V  A  M  G

 716  CCGCTGGTGAAAATTCGAAGGCAGAACTTATTCACAACAGTTGGGATGGTGGTAT  770
      A  A  G  E  N  S  K  A  E  L  I  H  N  S  W  D  G  G  I

 771  CATGTCGTATGGCTCCTACAAGTTCACTTCCACCTAATTTCAACATTGTAATACT  825
        M  S  Y  G  S  Y  K  F  T  S  T  *

 826  TGTTTTGTTTGTATCATCTCTTCAGCCCATTGGTATTCTCTTCAAGAATGAATTA  880
 881  ATCAGATGGAGTATATATATATCTATATGCTTATATATCTGAACTCATGTAGTTA  935
 936  TTATGCTAATTATTACATTTTCAAGCCTTTTGTCTTATAGTAGTTTCTGGTCTGG  990
 991  TCTGTGAATTCTTGCAGAAATTATGGGCTAACCATTTGGATATTGATTGACTGGT  1045
1046  GAAATTAATCAGGTGTGATTCCCGAAAAAAAGAAAAAAAAAAAAAAAAAAAA    1098
```

Figure 42. *Bv DODA* cDNA and peptide sequences. Start codon is in bold and poly(A)-tail is cut before its end. Grey boxes indicate the 5 out of 6 conserved catalytic amino acids from LigB domains (pfam02900).

Pg DODA protein



Bv DODA protein

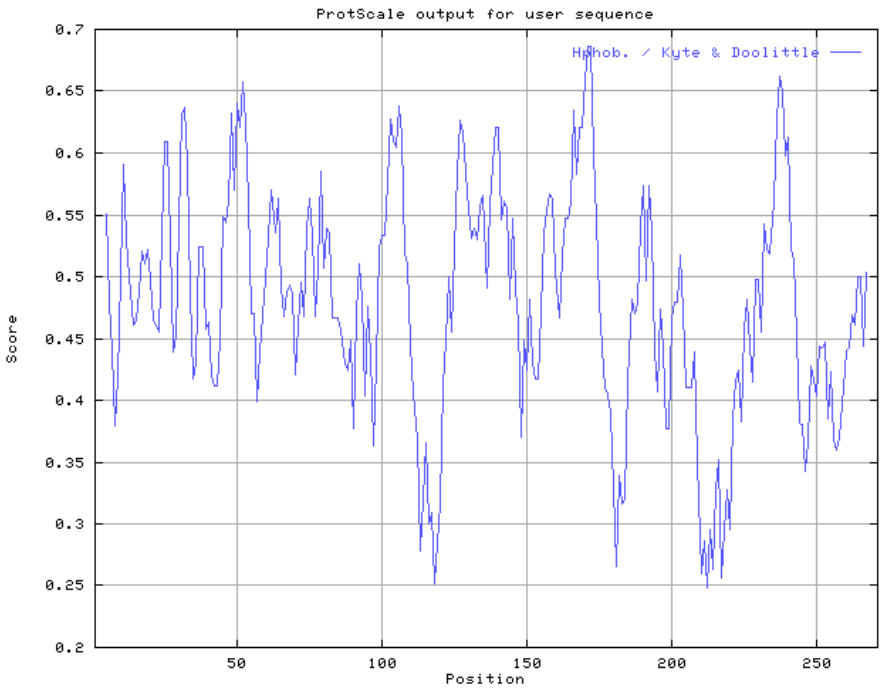


Figure 43. Comparison of the hydropathy plots of *Bv* DODA and *Pg* DODA according to Kyte and Doolittle (1982). Both proteins show similar hydropathy curves confirming their strong homology.

DODA PCR amplification on *Mammilaria* gDNA and positive controls gave one single band, whereas multiple bands were obtained for other species. As both primer combinations were tested for each sample, it was possible to identify the correct amplified product separated by around 100 bps length between the two PCR reactions. No band was obtained for *A. thaliana* as we expected, and all the amplification products obtained for *Silene* and *Dianthus* species were not specific. Universal primers are mainly based on the *P. grandiflora* DODA sequence for the positions where no consensus was found and no degeneration included. These primers seem to be "phylogenetic indicators" because they amplified effectively only the sequence closely related with *P. grandiflora* and worse and worse the more distant sequences. Thus non-Caryophyllales plants are not amplified. Finally, we obtained the central part of *DODA* genomic sequence for *Mammilaria*, *Phytolacca Americana*, *Iresine* sp. and the three positives controls (see Annex. 1). *DODA* gene from all these species except *P. grandiflora* possesses two small introns in the amplified part. These introns are located at the same place, but are variable in length. The translated proteins were aligned together to verify the conservation of the LigB catalytic amino acids (Fig. 44).

```
Pg DODA           MLADESFIARNFLLGWKKNVFPVKPKSILVVSAHWETDVPCVSAGQYPNVIYDFTEVPAS
Mc 2              MSIDETIPARHFLEEWQEKVYSKRPRSILVITAHWETVVPTVNAINHSDLIYDFGGFPAR
Bv DODA           MAIDDSKPSKKFLESWREKIFSKKPKAILVISAHWETDQPSVNVVDINDTIYDFRGFPAR
Mammilaria        MSIDDSIPARKFFQEWKEKVYSKRPKAILVISAHWETNVPAVNAVNHSDLIYDFRGFPAI
Phytolacca        MAIDKSVQARPFLEGWREKVLSKKPKSILMISAHWETDVPTVNAVHHSDLVYDFYGFPAP
Iresine           MSIDESIPARRFLEEWKDKVYSKRPSSILVITAHWLTVVPTVSAIDHSDLIYDFGGFPAC
                  *  *.:  :: *:  *:.:: . :* :**:::*** *  * *.. .  : :***  .**

Pg DODA           MFQMKYPAPGCPKLAKRVQELLIAGGFKSAKLDEERGFDHSSWVPLSMMCPEADIPVCQL
Mc 2              MYQLKYLAPGAPDLAKRLQELLAASGFECA-VDRKRGLDHGSWVPLMLMYPEANIPVCQL
Bv DODA           LYQFKYSAPGSPELANRIQDLLAGSGFKSVNTDKKRGLDHGAWVPLMLMYPEADIPVCQL
Mammilaria        MYQLKYPVPGAPNLARRVEELLTASGFSCV-VDKNRGLDHGSWVPLMLMYPEADIPVCQL
Phytolacca        MYQLKYPAPGAPHLARRIQEVLTASGLKCA-VDKKRGLDHGSWVPLRLMYPEASIPVCQL
Iresine           MYQIKYTAPGSPDLAKRVQELVTGSGFECA-VDTKRGLDHGSWVPLMLMYPEASIPVCQL
                  ::*:** .**.*.**.*::::: ..*:...  * :**:**.:**** :* ***.******

Pg DODA           SVQPGLDATHHFNVGRALAPLK
Mc 2              SVQPRLDGIHHYNLGKALAPLK
Bv DODA           SVQSHLDGTHHYKLGQALAPLK
Mammilaria        SVQSHLDGTHHYNLGKALAPLK
Phytolacca        SVQSNLDATHHYNLGRALAPLK
Iresine           SVQPHLDGKHHYDLGRALAPLK
                  ***. **. **:.:*:******
```

Figure 44. Alignment of DODA partial translated ORF from different betalain-containing Caryophyllales species representative of the classification of this order. The LigB conserved catalytic amino acids are indicated in grey boxes. For sequence reference see Annex. 2 (Mc 2 = *Mesembryanthemum* 2).

We observed a high degree of conservation between all these DODA fragments from different Caryophylalles species. Due to the strong conservation of this part of the protein, phylogenetic analysis will be distorted from the reality. Nevertheless, analysis of the multiple alignment revealed that *P. grandiflora* DODA is slightly different from the others clustering together. *Pg* DODA has between 56 to 62 % of identity with the other sequences, whereas 70 to 84% of identity has been found between them, aligned together by pairs.

The sequence Mc 1 (BF480453) from *Mesembryanthemum crystallinum* (*Mc*) is nearer to *Pg* DODA (68% of identity). Unfortunately, its N-terminal part is incomplete, which prevents aligning it with the other Caryophyllales fragments missing the DODA C-terminal part. As the DODA C-terminal end is more divergent than the N-terminal, we can conclude that Mc 1 is for the moment the nearest sequence to *Pg* DODA that we know. As both *Mc* DODA homologous sequences are as similar between them as with the others, we suppose that there is a gene duplication of *DODA* in *Mc*. It is possible that one of these two genes gives a non-functional DOPA-dioxygenase or is simply not expressed at the mRNA level. Both hypotheses can be tested by inducing the synthesis of betalains in *Mesembryanthemum* bladder cells in the presence of a specific wavelength of UV light (Vogt *et al.*, 1999b; Ibdah *et al.*, 2002). This would be followed by an observation of Mc 1 and Mc 2 mRNAs expression in this tissue or by the detection of their respective protein in the presence of two specific antibodies produced in rabbit by injection of short synthesised peptides chosen to differentiate both proteins.

3.3.3 Isolation of *Pg* DODA homologues from non-betalain producing plants

By aligning *Pg* DODA sequence against plant nucleotide translated databases, we identified one DODA homologous *Arabidopsis thaliana* expressed mRNA (NM_117597) and one or two *DODA* homologous ESTs per non-betalain plant species. Alignment of the different contigs obtained per species and the consecutive correction of the sequencing errors allowed us to reconstitute the complete ORF of DODA homologues from twelve different species (see Annex. 3). With these twelve species, we have a good representation of the plant kingdom, except for the ancestral plants like mosses. It would be interesting to isolate the DODA sequence from one of these species to study the evolution of DODA from bacteria to plant species.

As we found a partially sequenced EST from the moss *Physcomitrella patens* homologous to *Pg DODA* and we already constructed a cDNA library from this moss in our laboratory, we decided to isolate a full-length copy of this cDNA. The *Physcomitrella patens* homologue for *Pg DODA* (AJ583016) was amplified by nested PCR strategy on Lambda-UniZAP cDNA library (Girod *et al.*, 1999) with long primers (*PpL.6*p5, *PpL.6*p4) designed from the partially sequenced EST BJ195116 corresponding to the 5' region of the gene (Nishiyama *et al.*, 2003). Thus, we obtained a 961 bp-long cDNA encoding a 264 amino acid long ORF with a 29.5 kDa molecular weight and a pI of 6.2 (Fig. 45). LigB catalytic amino acid were also strictly conserved in the corresponding translated protein, confirming it as an extradiol 4,5-dioxygenase. We decided to call this protein DIOXA (for dioxygenase) because of the absence of betalains in this plant. There is no evidence that DOPA is the main substrate of this protein. Its hydrophobicity plot was the same as those obtained for *Pg* DODA and *Bv* DODA, confirming DIOXA in its function of extradiol 4,5-DOPA-dioxygenase.

We tried isolating DODA homologous sequences from non-betalain producing Caryophyllales plants such as *Silene vulgaris* and *Dianthus Caryophyllus*. Nevertheless specific PCR amplifications by using degenerated primers, either based on the betalain-producing plant sequences or on the non-betalain plant ones, failed. We obtained PCR fragment unrelated with DODA protein except the sequences corresponding to the amplification primers.

```
-49       AACTCAGTCCGATTCGGTTACAGCTAGCAAATCACTGGAGGCGGGAAGT -1

  1 ATGGCAACTTCAGCTGGTTTGAGTACTTTTTATGTATCGCATGGGTCGCCTATGA 55
    M  A  T  S  A  G  L  S  T  F  Y  V  S  H  G  S  P  M  M

 56 TGCCACTTGAGGATACTCCCATACGAGAATTCTTCTCCACTTGGACGGAGCGCTA 110
      P  L  E  D  T  P  I  R  E  F  F  S  T  W  T  E  R  Y

111 TCCCACAAGGCCAAAGGCCATTCTCGCAATTTCAGCTCACTGGGATACTCGAGAG 165
     P  T  R  P  K  A  I  L  A  I  S  A  H  W  D  T  R  E

166 CCTGCCGTTAATGCTGTCAGCCAGAACAGCACTATTCACGACTTTTATGGCTTCC 220
    P  A  V  N  A  V  S  Q  N  S  T  I  H  D  F  Y  G  F  P

221 CTCGCGAGCTATATCAATTGCAATACACGCCTCCAGGGGCTCCAGACGTTGCAAA 275
      R  E  L  Y  Q  L  Q  Y  T  P  P  G  A  P  D  V  A  K

276 GAGGGTGACATCTCTCCTCAAAGACGCTGGCTTCAAGACCGTCCTCGAGGACAAC 330
     R  V  T  S  L  L  K  D  A  G  F  K  T  V  L  E  D  N

331 AAGAGAGGGCTCGACCACGGAGCATGGACGCCCCTGATGCTAATGTACCCCAACG 385
    K  R  G  L  D  H  G  A  W  T  P  L  M  L  M  Y  P  N  A

386 CTGACATTCCTGTCCTCCAGGTCTCAATCCAAAGCAACAAAGATGGGCTTCACCA 440
      D  I  P  V  L  Q  V  S  I  Q  S  N  K  D  G  L  H  H

441 TTACCAGCTCGGCCGGGCTCTTGCACCGCTCAAGGACGAAGGAGTGTTAATTTTT 495
     Y  Q  L  G  R  A  L  A  P  L  K  D  E  G  V  L  I  F

496 GCCTCTGGAACTACAGTCCACAATTTGAGAGAGATAGATTTTTCTGCCAAGAAGC 550
    A  S  G  T  T  V  H  N  L  R  E  I  D  F  S  A  K  K  P

551 CAACTGTGTGGGCCAAGGCATTCGACGGATGGTTGACTGATGTGCTACTCAACAG 605
      T  V  W  A  K  A  F  D  G  W  L  T  D  V  L  L  N  S

606 CAAGCACAAGGAAGCAATGGAATGGGAGAAAGCACCATATGCTTCCAAAGCTCAC 660
     K  H  K  E  A  M  E  W  E  K  A  P  Y  A  S  K  A  H

661 CCTCATCCAGATCATTTCCTGCCTGTGTTGGTTGGGTTGGGAGCTGCGGGAGAGC 715
    P  H  P  D  H  F  L  P  V  L  V  G  L  G  A  A  G  E  Q

716 AATGCCAAGCTGAGAAGATATATGAGGAATTCGCATATGGTTTGGCGCTTTCATG 770
      C  Q  A  E  K  I  Y  E  E  F  A  Y  G  L  A  L  S  C

771 TTTTGCTTTCCATCCTCAAAACTAGTATGGTAGTTGTTGGTTAAAGTTTTCCACA 825
     F  A  F  H  P  Q  N  *

826 TGTGCTGCTTGAGTGGATGCTCACCACACAGCTGGTTAAACAATTTCCAATGTGT 880
881 GGATATTGATTTGTATTTTTCATTCTATTTTGAATAAATATGTTAGCTCCCGTAT 935
936 CTTGTAA                                                 942
```

Figure 45. *Physcomitrella patens DIOXA* cDNA and peptide sequences. Start codon is in bold and 3' and 5'UTR regions are incomplete. Grey boxes indicate the 5 out of 6 conserved catalytic amino acids from LigB domains (pfam02900).

3.3.4 Phylogenetic analysis of DODA and its homologues

DODA homologous protein sequences (Met22-Lys163) from betalain and non-betalain plants have been used for sequence alignments and construction of a phylogenetic tree (Fig. 46). In this tree, there are four different groups, one composed of betalain plants and three others composed of non-betalain-producing plants. Among the last, we can distinguish monocotyledons from dicotyledons and bryophytes (*Physcomitrella patens*). Caryophyllales sequences are clearly clustered together with 60 to 62% of identity with *Pg* DODA, whereas homology with other plant genes was lower with 52% of identity with *Arabidopsis thaliana*, 50% with major monocotyledons and 45% with the bryophyte *Physcomitrella patens* corresponding here to the root.

Paralogs were identified in several plant families like the Solanaceae (*Solanum tuberosum*, *Lycopersicon esculentum*) or the Fabaceae (*Glycine max*, *Populus sp.*). The same phenomenon exists also in the Caryophyllales order in the *Mesembryanthemum crystallinum* species, but it is not represented on this tree due to the incomplete sequence of *Mesembryanthemum crystallinum* 1 DODA homologue.

All these results correspond to the admitted plant phylogeny, except that betalain plants clearly form an out-group, whereas usually they are included in the dicotyledon group. This confirms that betalain synthesis constitutes a phylogeny criterion by itself, sufficient to place betalain-producing plants as a separated order in the dicotyledons. Thus, we suppose that DODA protein from betalain plants has diverged once from other plant DODA homologues (called DIOXA) to participate to the synthesis of betalains.

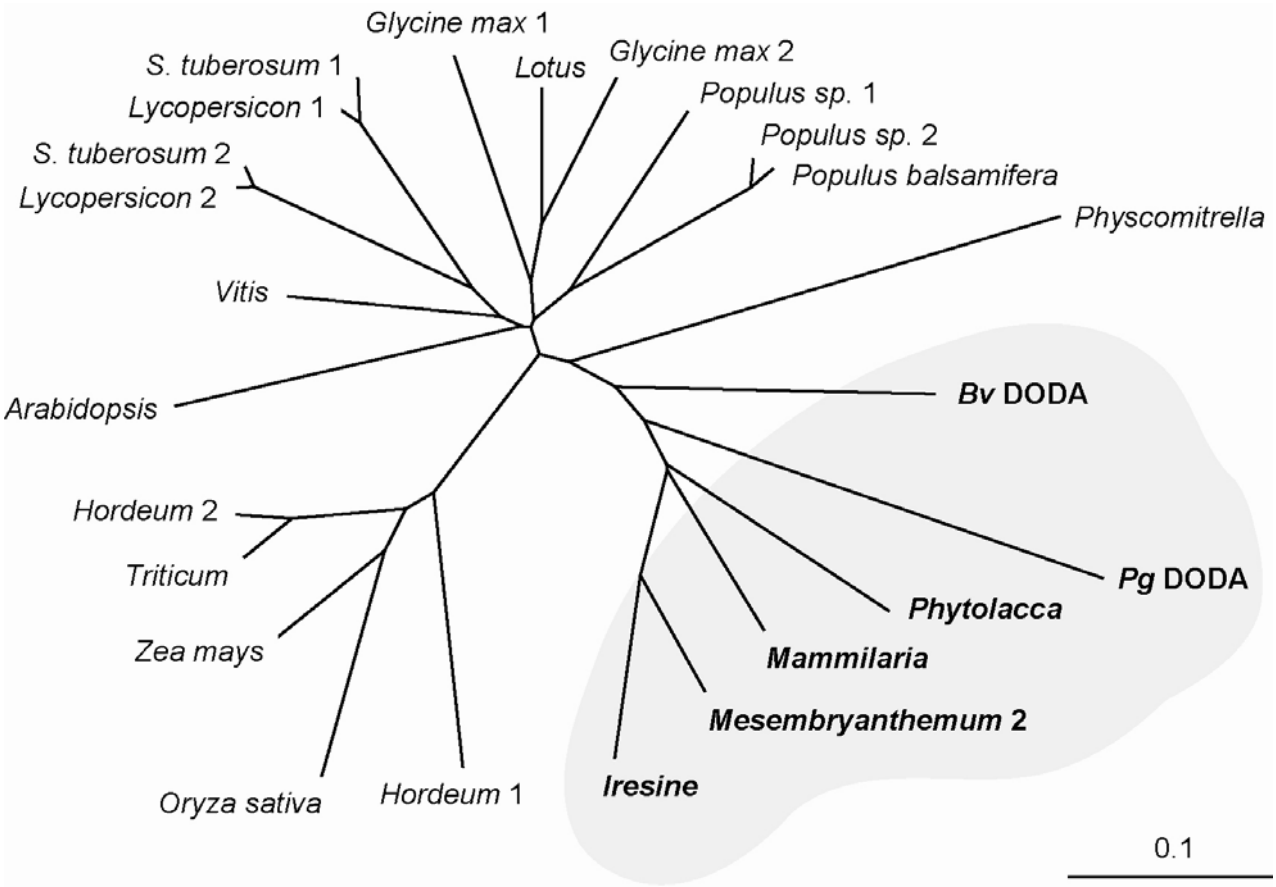


Figure 46. Phylogenetic analysis of *Pg* DODA homologues in plants. Multiple alignments from *Pg* DODA homologous fragment (Met22-Lys163) were done with Clustal W and the tree created with PHYLIP. The moss (bryophyte) *Physcomitrella patens* corresponds to the root. Betalain producing species in the grey surface clearly form a cluster distinct from other plants. For Genbank accession numbers see Annex 2.

New putative proteins having a higher level of homology with *Pg* DODA than *Sphingomonas* LigAB dioxygenase have been also identified in numerous bacteria and archaea species. These proteins also possess all LigAB catalytic amino acids. A global phylogenetic analysis has been done with manual corrections of the sequence alignment. We have chosen complete sequences from the plant and the bacteria kingdoms, and one archaea specie. Clear DODA homologous sequences were absent from the fungal and the animal kingdoms. The best consensus tree was constructed from this alignment and a bootstrap analysis was made to evaluate the consistence of the branches of the tree (see Fig. 47).

First, we can see that plant species cluster together with a high bootstrap support (99/100 replicates). On the bacteria side, Proteobacteria are divided in two main groups containing both γ-Proteobacteria species. Among the latter, the enterobacters clustering with α-Proteobacteria (99/100) are more distant from the plant cluster than the pseudomonales clustering with β-Proteobacteria (52/100). The β-Proteobacteria, represented here by *Burkholderia fungorum* specie, are the closest bacteria to plant DODA with around 40% of identity and 60% of similarity with *Physcomitrella patens*. *Pseudomonas* and *Xanthomonas* species are the nearest γ-Proteobacteria species with 38% of identity and 55% of similarity. Finally other bacteria corresponding all to Gram-positive bacteria (Firmicutes, Actinobacteria) are forming an out-group, which is not well supported by bootstrap analysis. We can suppose that the *DODA* gene from plant has a proteobacterial origin, thus explaining the distance observed for the other bacteria types. The archaea *Methanosarcina acetivorans* is included in the pseudomonal bacteria cluster, which is not surprising due to the incorporation by this methanogen archaea of an important part of bacterial genome in its genome by lateral gene transfer. Complete sequencing of the *Methanosarcina mazei* genome revealed that 544 ORFs (16%) only reach significant similarity values in the bacterial domain (Deppenmeier *et al.*, 2002). This finding might indicate that lateral gene transfer has played an important evolutionary role in forging the physiology of this metabolically versatile methanogen.

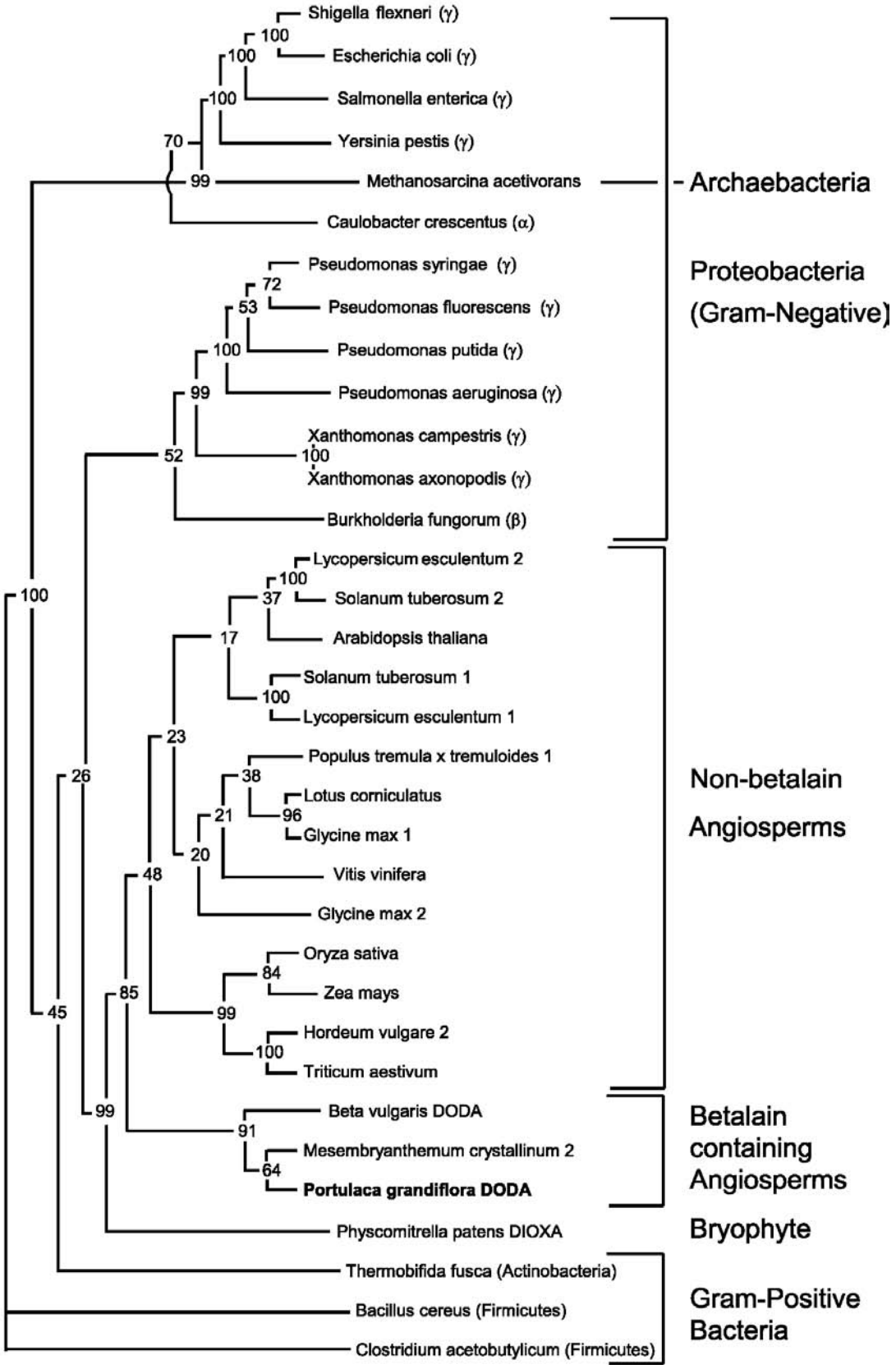


Figure 47. Phylogeny analysis of *Pg* DODA homologues (Majority-rule and Strict Consensus Tree program, version 3.572c). Numbers at the forks indicate the number of times the group consisting of the species which are to the right of that fork occurred among the trees, out of 100 replicate trees.

3.3.5 Identification of a pattern specific to betalain plants and analysis of its influence on the protein functionality by preliminary modeling of DODA catalytic site

Clustal W alignment of *P. grandiflora* DODA homologues from different kingdoms allowed the identification of a conserved motif (HNL-R/G) present in all organisms except for betalain-synthesizing plants. In these plants, the completely different conserved motif P-(S,A)-(N,D)-x-T-P is present (see Fig. 48). Both motifs begin with the strictly conserved His177, which is essential for the catalytic activity of extradiol dioxygenase class III enzyme (Sugimoto *et al.*, 1999) like *P. grandiflora* DODA. A single point mutation of His177 suppresses the enzyme activity.

The identification of one conserved pattern specific to DODA proteins from betalain-producing plants at the place of a pattern conserved in all other organisms suggested strongly, that DODA from all betalain plants resulted from modification events having taken place in a single species. Consequently, we conclude that all betalain-containing plants belong to the same phylogenetic cluster, differing from all other plants by their ability to synthesise betalain pigments.

We propose that modifications of the residues next to the conserved catalytic His177 could influence the capacity of DODA extradiol dioxygenase to open certain substrates. Thus in the past, the future betalain-producing plants would have modified the structure of some existing protein, like DODA, to use a new substrate like dopamine in order to develop the synthesis of a new pigment, called betalain.

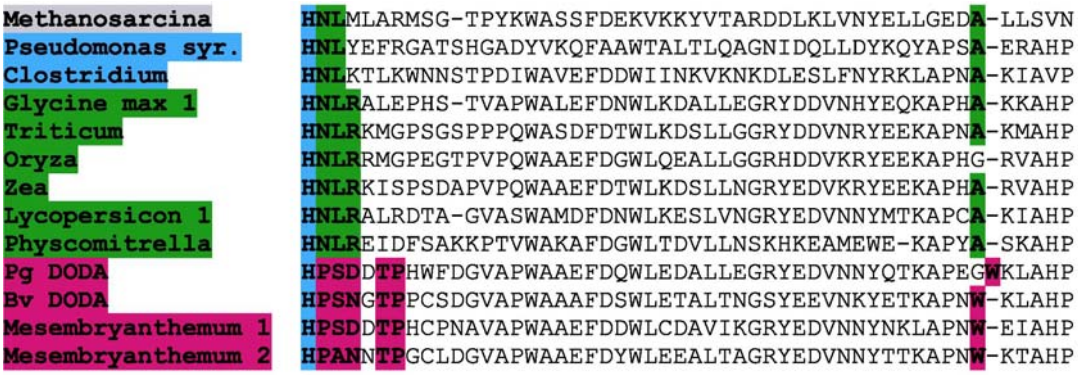


Figure 48. Identification of a conserved pattern specific to DODA proteins from betalain producing plants by alignment of the *Pg* DODA homologous sequences from different kingdoms. The conserved catalytic amino acid His177 is followed by the pattern P-(S,A)-(N,D)-x-T-P in all homologues of betalain producing plants, whereas at the same place a H-N-L pattern is conserved in all archaea and bacteria, and a H-N-L-R pattern is conserved in all plant homologues not producing betalains. For Genbank accession numbers see Annex. 2.

It will be very interesting to verify the importance that the pattern P-(S,A)-(N,D)-x-T-P, conserved in betalain-containing plants, has on the structure of the DODA catalytic site, in comparison with the H-N-L-R pattern present at this place in all other plants. For this purpose, we decided to make a preliminary modeling of the DODA catalytic site from a betalain-producing plant and a non-betalain-containing plant.

Three-dimensional models of *Pg* DODA and *Physcomitrella patens* DIOXA (Pp DIOXA) were created by using as a template the crystal structure from the *Sphingomonas* LigAB protein (1B4U) and the SWISS-MODEL and Swiss-Pdb Viewer softwares (Guex and Peitsch, 1997). *Pg* DODA and Pp DIOXA were aligned on the LigAB sequence according to the manual alignment presented in Annex. 4. Due to several insertions in the LigAB sequence by comparison with the two other sequences, a few errors were generated. In the presence of an insertion in the bacterial LigAB template, the modeling software has no other choice than to artificially lengthen the concerned peptidic bond in the model. For example, between the residues 180 and 190 from LigAB, an alpha-helix was suppressed from the model and replaced by a carbon chain of the same size. However, we observed that this helix also existed in the two modeled proteins, but just before. These errors were not corrected when they were sufficiently distant from the active site, even if they were important.

Fig. 49 shows a global view of the carbon-chain of the *Pg* DODA 3D model coloured according to the problems existing on the protein. Indeed there are several problems. The three histidines linking the iron cofactor are coloured in blue to better localize the active site. A large line is in particular present at the bottom-right corner of the figure, corresponding to the missing helix mentioned previously. As this missing structure is sufficiently distant from the catalytic site, we suppose that it has a negligible importance for the observations that we want to obtain.

Our model is based on the fact that the alignment of the important sites is sufficiently good that residues in the neighborhoods of the active site are correctly localized. However, a global homology of 40% does not allow to study the catalytic mechanism in details, but just to have a preliminary idea of the general structure of the active site. This will allow us to choose the right residues for a mutagenesis analysis for example, or to see the influence of some conserved pattern on the global structure.

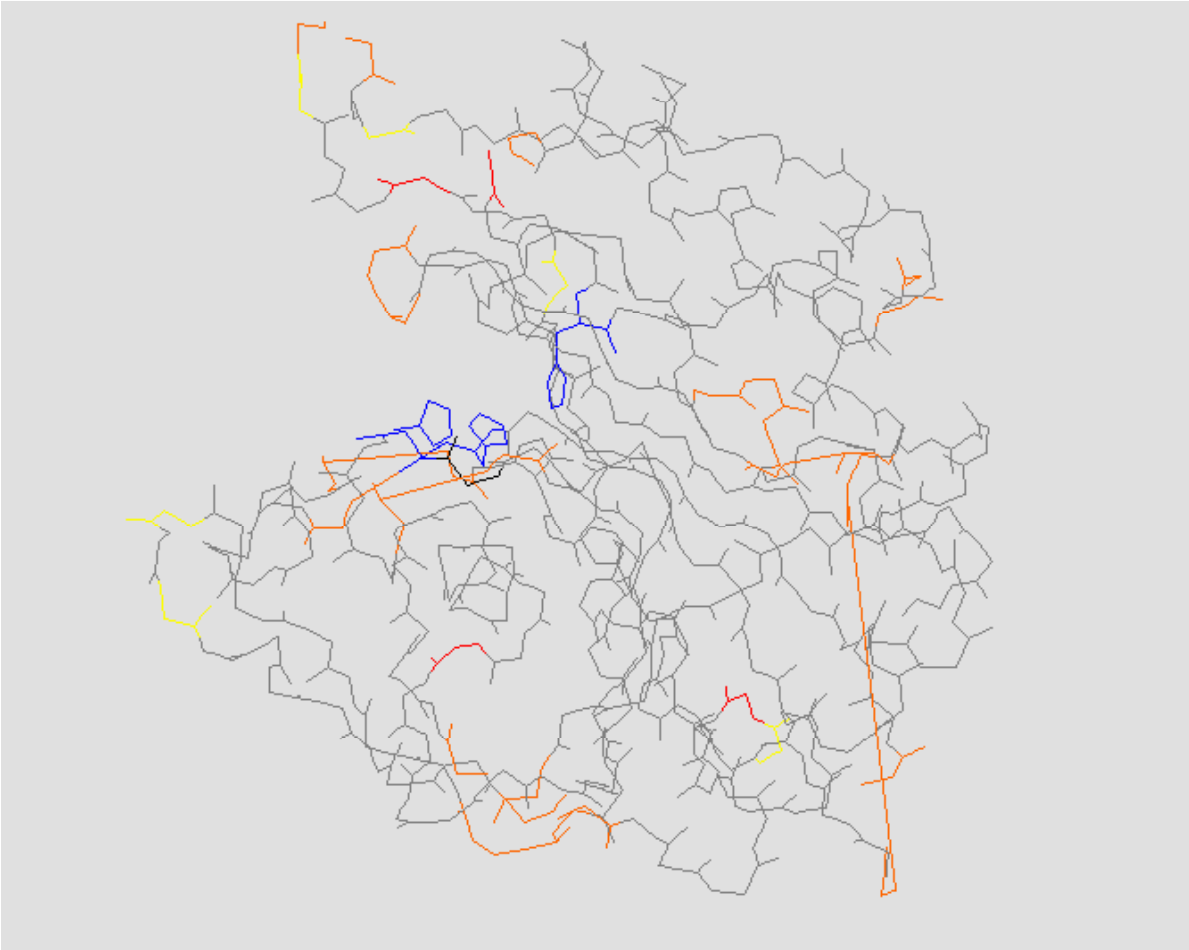


Figure 49. Global view of the three-dimensional model of *P. grandiflora* DOPA-dioxygenase DODA, coloured according to the problems detected on the protein model: in red, the prolines with an incorrect angle of torsion; in yellow, incorrect angle of torsion for other residues than proline; in orange, the internal residues without hydrogen bond; in pink, lateral chains making "collisions"; in blue the three conserved histidines coordinating the iron cofactor.

As it is relatively difficult to distinguish particular residues on the 3D model presented previously, we decided to focus our analysis on the residues localized in the direct proximity of the catalytic site, or those interacting with the substrate or the iron cofactor. The resulting stereo picture of the LigAB catalytic site is presented in Fig. 50. The three conserved residues linking the iron cofactor, the residues making hydrogen bonds with the PCA substrate and those making hydrophobic bonds with the aromatic ring of PCA are represented.

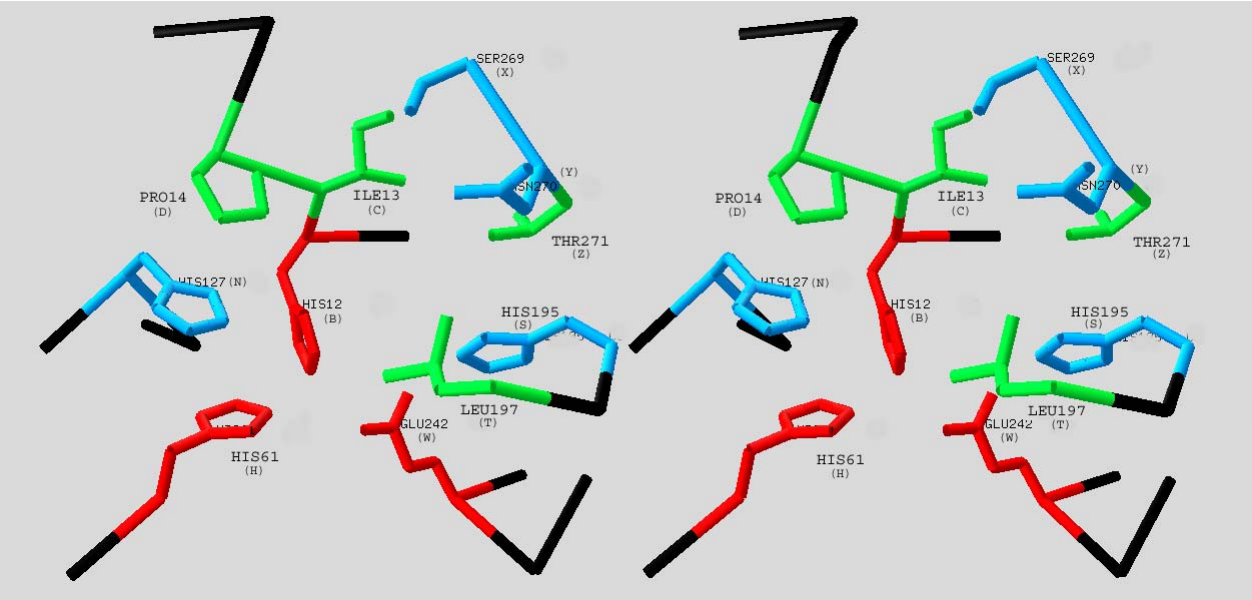


Figure 50. View of the LigAB active site constructed from the PDB file 1B4U. In red, the histidines linking the iron cofactor; in blue the residues making hydrogen bonds with the substrate; in green amino acids interacting by hydrophobic bonds with the substrate. Numbers are corresponding to the 1B4U sequence.

The same work was made with the *Pg* DODA and *Pp* DIOXA model with a conservation of the colour codes. Amino acids which are supposed to be involved in the extradiol dioxygenase catalytic mechanism are highlighted in pink, and the His262 was coloured in yellow for visibility reasons. This histidine is conserved in all DODA homologous sequences from plants and replaced by a glycine in the bacterial sequences. According to the orientation of its lateral chain, His 262 could interact with the substrate. Due to lateral chain interferences, we chose a better rotamer for His 262.

The angle of sight of the active site of *Pg* DODA and DODA homologues from plants are different from the angle of the bacterial structure for visibility reasons. On the *Physcomitrella patens* model (see Fig. 51), the three histidines linking the iron are localized in the same equatorial plane as described for LigAB (Sugimoto *et al.*, 1999), with the iron ion in the middle (not shown).As we know that the substrate makes a tetragonal bi-pyramidal coordination sphere with the three Histidines constituting its base, we supposed that the two conserved histidines (blue) could also make hydrogen bonds with the unknown aromatic substrate in the same way that His195b and His127b from LigAB protein do (Sugimoto *et al.*, 1999). The green coloured leucine making hydrophobic bonds with the substrate is conserved in all bacterial and plant DODA homologues, except those belonging to betalain species. The two adjacent amino acids, the asparagines 210 and the arginine 212 are also conserved similarly. Both residues could interact with the substrate and could be involved in the substrate recognition, thus responsible for the specificity of the enzyme. The same observations could be made for the *Pg* DODA active site model (see Fig. 52), excepted that Asparagine 210, leucine 211 and Arginine 212 (N-L-R motif) are replaced by other conserved amino acids, which could also interact with the substrate (P-(S,A)-(N,D)-x-T-P). It is difficult to know the real orientation of these amino acids due to the presence in the top part of this model of a loop of 5 amino acids absent from the bacteria structure (black). For the modeling, we used a function of Swiss Pdb Viewer allowing to look from the sequence in a database of existing loops. Therefore this loop is surely badly orientated in both models, but influencing them in the same way. It is interesting to notice that the glycine 203 residue (in pink, just behind histidine B) conserved in all species, is very close to the iron orientating site. If we had found another residue with a lateral chain at this place, it would have surely disturbed the iron coordination. Thus, the fact that this glycine is present at this place validates the modeling made from the alignment.

In conclusion, preliminary 3D-modeling of the *Pg* DODA and *Pp* DIOXA enzyme, based on the protocatechuate 4,5-dioxygenase LigAB crystal structure, revealed that the two different conserved amino acid patterns following His177 (according to *Pg* DODA sequence) could be involved in the substrate specificity of 4,5-DOPA-dioxygenase by controlling access to the catalytic site (Fig. 53). Thus, the membership of *DODA* genes from betalain plants to a cluster, separated from other *DODA* homologue genes from non-betalain-containing plants, is also well supported by the three-dimensional modeling of the DODA and DIOXA active site.

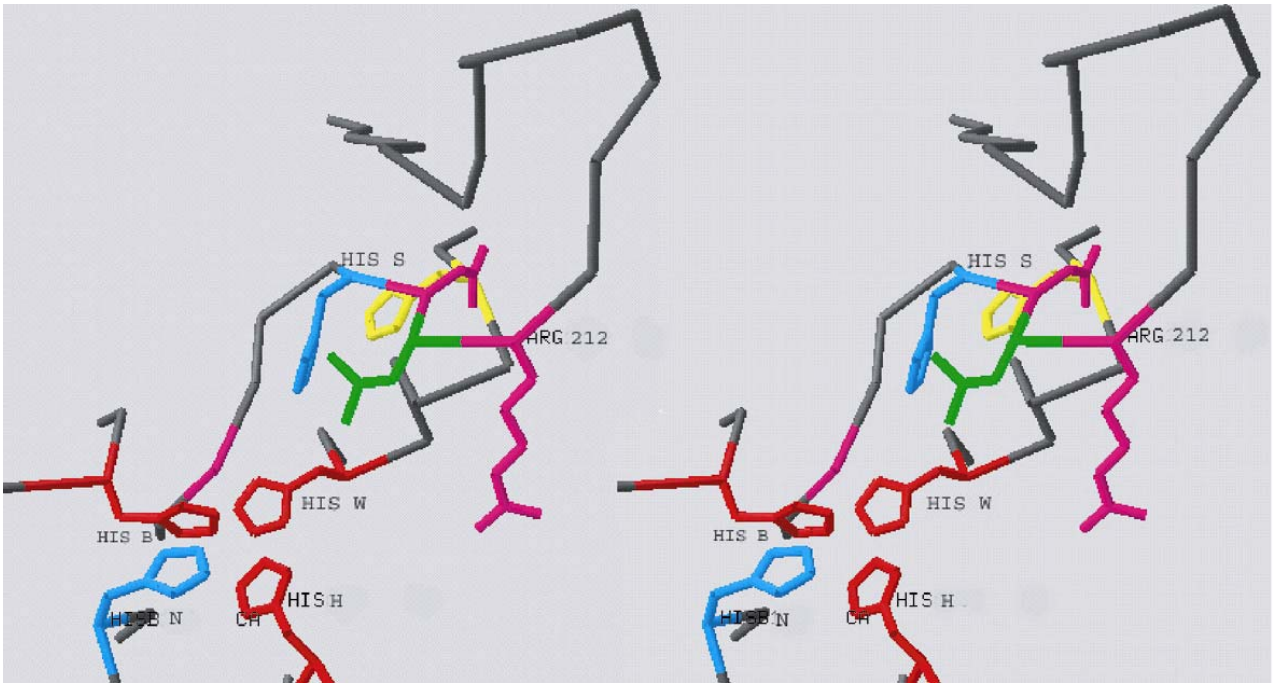


Figure 51. Stereo modeling of *Physcomitrella patens* DIOXA active site. Shown in red, the three histidines (B,H,W) potentially linking the iron cofactor; in blue, the two histidines (N,S) supposed to make hydrogen bonds with an unknown aromatic substrate; Following His S, shown in green a leucine, and in pink an arginine and an asparagine, which could anchor the substrate; Coloured in yellow, an histidine (position 262) could eventually interact with the substrate.

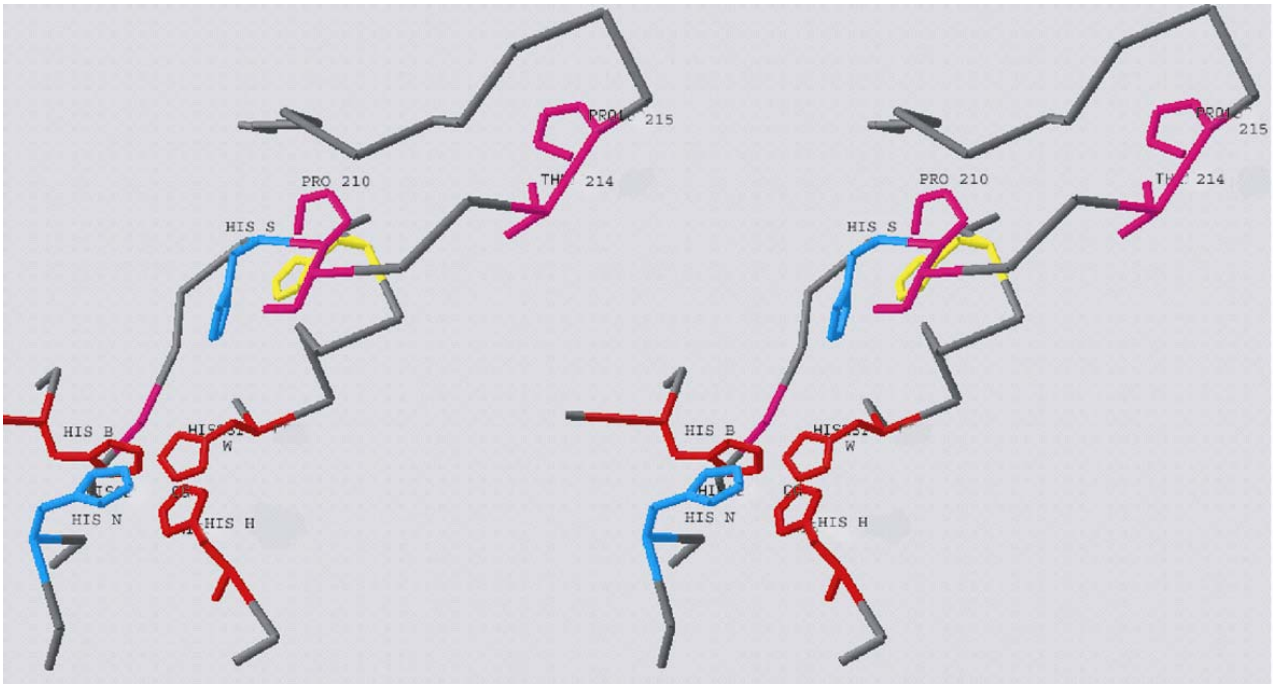


Figure 52. Stereo modeling of *P. grandilflora* DODA active site. Shown in red, the three histidines (B,H,W) potentially linking the iron cofactor; in blue, the two histidines (N,S) supposed to make hydrogen bonds with an unknown aromatic substrate; Following His S, in the large loop, coloured in pink, proline, asparagine or serine, and then several residues further, threonine and proline, which could anchor the substrate; Shown in yellow, an histidine (position 262) could eventually interact with the substrate.

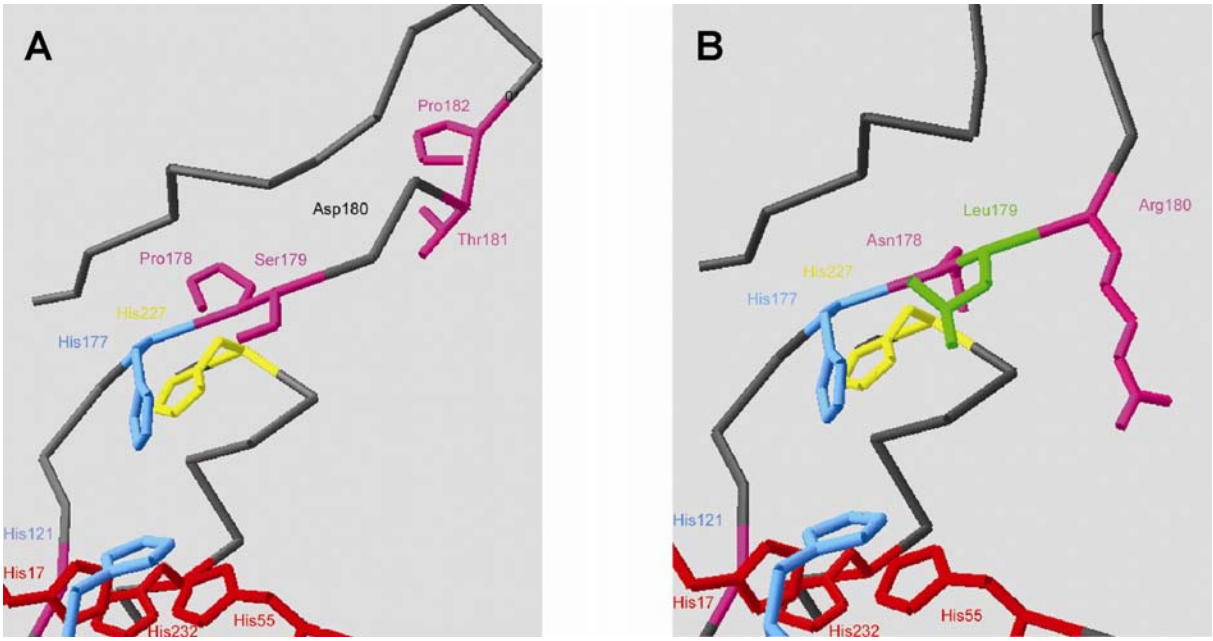


Figure 53. Preliminary modeling with SwissPdbViewer of the catalytic site of *Pg* DODA (A) and the moss *Physcomitrella patens* DODA homologue (B) based on the 3D structure 1B4U of *Sphingomonas* LigAB. Referring to *Pg* DODA sequence, catalytic His177 is conserved (shown in blue), whereas the three following residues conserved in betalain plants Pro178-Ser179-Asp180 are replaced in all non-betalain plants by Asn178-Leu179-Arg180, which reduce clearly the access to the catalytic site represented by the three conserved His linking the iron cofactor (red) and the two conserved His making hydrogen bonds with the substrate (blue). Pro178-Ser179-Asp180 and Thr182-Pro183 are conserved only in betalain plants potentially participating to the substrate recognition.

4. Functional analysis of P. grandiflora DODA homologue in non-Caryophylalles plants

4.1 Functional analysis of *Pg* DODA homologue from *Arabidopsis thaliana*

The first question that we raised, was to know if *Pg* DODA homologues from non-betalain plants were able in adequate reaction conditions to catalyse the conversion of DOPA into betalamic acid. For this purpose, we have at our disposal the biochemical complementation test of the betalain biosynthetic pathway in *Pg* white petals developed by Mueller and colleagues (Mueller *et al.*, 1997a). Successful complementation experiments have been already obtained with the novel *Portulaca grandiflora* DODA (Christinet *et al.*, 2004) and the non-homologous *Amanita muscaria* DODA (Mueller *et al.*, 1997a). *Pg* white petals deficient in the DOPA-dioxygenase gene *C* are containing all precursors necessary for the synthesis of betalamic acid, in particular the DOPA substrate.

We decided to exactly replicate the biolistic experiment made with *Pg DODA*, replacing this gene by its *Arabidopsis* homologue (NM_117597) to give pNco AtL6 construct (Annex. 6). This construct was double sequenced to verify the ORF integrity. The biolistic transformation experiment was conducted according to Mueller and colleagues (1997a) and repeated three times, but no cell pigmentation was observed in *P. grandiflora* white petals, despite the expression of the DsRed2 fluorescent control protein, testifying the correct experimental conditions.

We suppose therefore from this experiment that the *DODA* homologous gene from *Arabidopsis thaliana* is not able to take part to the betalain synthesis even placed in an ideal environment such as *P. grandiflora* white petals. It is possible that other *Pg* DODA homologues from non-betalain plants are in the same case. In fact, the function of this *DIOXA* gene correspond clearly to a 4,5-extradiol dioxygenase due to the presence of the conserved LigB catalytic amino acids and the conservation of the general structure of the active site as demonstrated in our three-dimensional model of *Pp* DIOXA. However, the substrate specificity of DIOXA dioxygenase could be different due to small modifications of the active site geometry and its concrete role in the plant physiology is still unknown.

4.2 Functional analysis of DIOXA dioxygenase from *Physcomitrella patens*

In order to determine the function of *Pg DODA* homologue in plants without betalain, we decided to construct a knock-out of this gene in the bryophyte *Physcomitrella patens*. To date this moss is the only plant species able to make DNA homologous recombination at sufficiently high frequency to use it concretely for specific gene targeting (Schaefer and Zryd, 1997).

For this purpose, it was necessary to isolate a genomic DNA fragment including intron sequence of the targeted gene. *DIOXA* genomic fragment was amplified on *Physcomitrella* genomic DNA by nested PCR with primers L6PpP1/ L6PpP4 and L6PpP2/ L6PpP5. pKO ori1 and pKO ori2 insertion vectors containing *Pp DIOXA* amplified genomic DNA fragment in opposite directions and the *NPT2* kanamycin resistance gene, were constructed from the vector pKO ori (Schaefer, unpublished) in order to practice *DIOXA* gene targeting in *Physcomitrella patens* protoplasts (see Annex. 6). Thanks to this system, the genomic DNA fragment accompanied by the kanamycin resistance gene will be integrated by homologous recombination in the homologous genomic DNA fragment from *Physcomitrella* patens genome, and thus will inactivate *DIOXA* gene. Frequently the plasmid will be integrated at several copies. Fig. 54 shows the different possible insertions. This scheme corresponds to a double integration of the pKO ori1 plasmid inside *DIOXA* genomic copy.

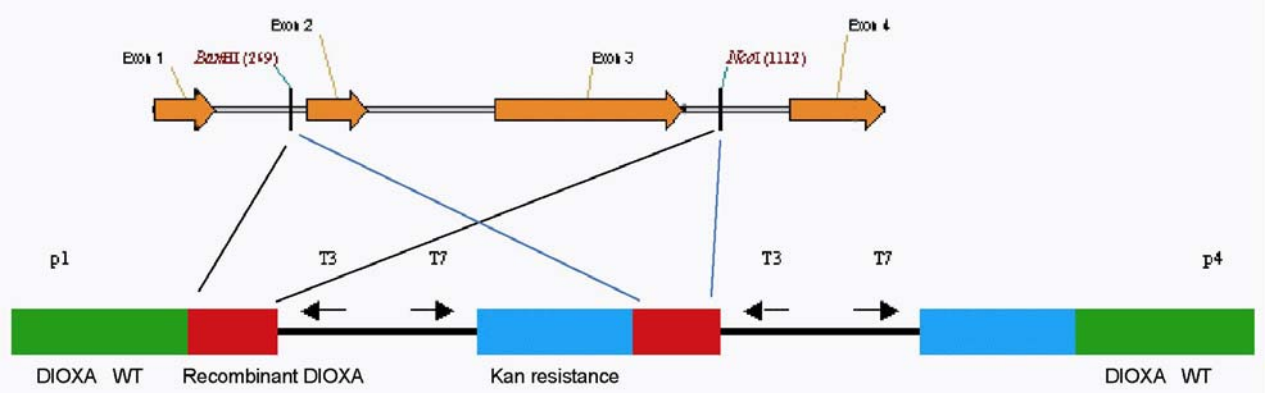


Figure 54. *DIOXA* homologous recombination in *Physcomitrella patens*. In green, the wild-type DIOXA; in red, DIOXA recombinant homologous fragment integrated in the insertion vector; in blue, the kanamycin resistance gene; in black, the insertion vector pKO ori1. As it is frequent to find multiple insertions of the plasmid, here we presented a double insertion event. On the top part, a representation of *DIOXA* genomic DNA with the two restriction sites *BamH*I(269) and *Nco*I(112) used for generating the fragment integrated in the insertion vector pKO ori1.

After transformation of *Physcomitrella patens* protoplasts (Schaefer and Zryd, 1997) and two successive kanamycin selection rounds separated by a non-selective step, several resistant clones were obtained. At this point, some false positive transformants frequently subsist. Non-integrated transformation can confer the kanamycin resistance to the moss. Therefore, it is vital to make several selection steps in alternation with non-selective steps. The non-selective steps allow the dilution of the number of copies of the non-integrated vector. DNA can also be integrated in a non-homologous way at low frequency, thus giving also false positives. Finally, as we do not know how many copies of *DIOXA* are present in *Physcomitrella patens* genome, it subsists, in the presence of several copies, the possibility that some copies are not inactivated, thus also giving false positives with a wrong phenotype.

All these uncertainties can be solved by a parallel Southern blot analysis of *DIOXA* in the wt and the supposed positive clones. For this purpose clones were propagated several weeks on non-selective medium and genomic DNA was extracted. Genomic DNA was digested by *Nhe*I and *Bsp*TI blotted on a nylon membrane. Hybridization was conducted with the complete *DIOXA* cDNA (AJ583016). Unfortunately, we detected only a high molecular band in two transfomants corresponding to the integration of multiple copies of the plasmid. The signal has a rather low intensity. This confirms that both clones were transformed, but does not answer the question if it is a *DIOXA* targeted transformation as expected, because no signal was detected for the wt type control. Therefore this hybridization has to be made a second time in order to have a clear answer.

It would be also interesting to overexpress *DIOXA*, but because of little time, we did not end the construct of the plasmids necessary for this strategy.

5. Biotechnological applications resulting from the isolation of DODA DOPA-dioxygenase

The isolation of the plant *DODA* gene performing the last enzymatic step for betalamic acid biosynthesis opens a new era in betalain research: the engineering of the betalain biosynthesis pathway in Caryophyllales plants. This could allow the development of commercial applications based at first on the colourant properties of these pigments, but also on their anti-oxidative properties demonstrated more recently.

5.1 Betalains as a food colourant

Using colour enhancement in food makes certain sorts of food more aesthetically appealing and appetizing. It is also a cheap manner to reconstitute the natural colour of food often lost after transformation processes or to disguise low quality foods (Delgado-Vargas *et al.*, 2000). Natural ingredients such as the Red Cochineal (*Dactylopius coccus*), saffron, carotenoids or red betalains have been used for a very long time for food colouration, but in the late 1800s, the developed food industry had available a vast array of synthetic colours, thus partially decreasing the interest for natural colourants.

The red colourant from betalains is obtained from garden beetroot as a juice, which can be concentrated or transformed in powder or paste. The main pigment is the betanin, constituting around 95% of the total content in betacyanins. Betaxanthins are also present, but their yellow colour is masked by violet betacyanins. The characteristic smell of geosmin from beet juice, as well as the high nitrate and nitrite concentrations, are minimised after ultrafiltration by fermentation of the juice concentrate in the presence of *Candida utilis* under partial anaerobic conditions (Adams *et al.*, 1976).

Betalains are used as a cost-effective natural colourant known under the designation E162 in numerous dairy products like strawberry ice-cream, yoghourt and dairy drinks (Pasch *et al.*, 1975). Enhancement of the colouration of meat substitutes and sausage is also possible in accordance with betalain physicochemical properties (Elbe and Maing, 1973; Dhillon and Maurer, 1974; Elbe *et al.*, 1974). The stability of their colour is mainly due to the pH, but light, temperature, air conditions and water availability are also important factors (Patkai and Barta, 1996; Reynoso *et al.*, 1997). No change in the red-blue betanin colour is detected between pH 3.5 and 7.0 corresponding to the pH of most foods (λmax 537 nm). Below pH 3.5, λmax shifts toward a lower wavelength (535 nm at pH 2.0) with a slight colour change to blue-violet, while above pH 7.0 λmax shifts toward a longer wavelength (544 nm at pH 9.0), the colour becoming more blue-violet. The degradation reaction of betanin in solution has been shown to be partially reversible. Betanin regeneration involved a Schiff-base condensation of the hydrolysis reactions products, the cyclo-DOPA-5-*O*-glucoside with the aldehyde group of betalamic acid. Both the degradation and regeneration reactions are pH dependent (Elbe *et al.*, 1981; Bilyk and Howard, 1982). Ascorbic, isoascorbic, metaphosphoric and gluconic acids improved the regeneration of red beet juice pigments after heating, retaining the initial concentration of pigments even after 5 cycles of heating (3 min at 100°C) and regeneration (Han *et al.*, 1998).

Nevertheless, after beetroot processing, the pigment concentration in beet juice concentrate or dry powder represents 1% or less. Thus, there is a need for a quantitative increase in betalain concentrations in beet to improve the commercial applications of these pigments as food colourants. In contrast to red beets, cactus pears offers a great palette of colour hues and therefore may be also used as a food colourant free from certification (Stintzing *et al.*, 2001). In Central and South America, the purple garambullo fruit from *Myrtillocactus geometrizans* growing in desert regions has also been confirmed as a new potential source of betacyanins (Reynoso *et al.*, 1997), as well as the fruit from red-purple pitaya (*Hylocereus polyrhizus*) (Stintzing *et al.*, 2002a).

The chemical stability and colourant properties of three betaxanthins recently identified from *Celosia argentea* varieties were evaluated (Cai *et al.*, 2001b). Bright yellow betaxanthins possess similar colourant properties, with colour stability between pH 2.2 and pH 7.0 in aqueous solution. Lyophilized betaxanthin powders from yellow inflorescences of *Celosia* exhibited bright yellow colour and high colour purity with strong hygroscopicity and a much better storage stability (mean 95.0% pigment retention) than corresponding aqueous solutions (14.8%) at 22°C after 20 weeks.

Refrigeration (4°C) significantly increased pigment retention of aqueous betaxanthins to 75%. Unfortunately these pigments are not used for the moment due to the absence of an agricultural crop, producing the pigment in sufficient amounts, so that it could be economically extracted.
However at present there is no safe, cheap and highly water-soluble bright yellow colourant of natural origin. Curcumin from turmeric rhizome is the main natural yellow pigment used as food additive at this time. Crocin, a water-soluble carotenoid corresponding to the saffron aroma from *Crocus sativus* is also used, but its colour is rather yellow-orange. Crocin has a strong taste and is terribly expensive due to the small quantities produced, whereas curcumin is badly hydrosoluble and light sensitive when solubilised. Many synthetic red or yellow dyes have been either banned by the Food and Drug Administration, or are in danger of losing their permitted status due to suspect carcinogenic effects. In addition consumer preference for natural over synthetic products has increased dramatically in recent years, suggesting the widespread level of acceptance for natural food additives derived from plant tissues (Downham and Collins, 2000).

Thus, the accumulation of betaxanthins was studied in *Beta vulgaris* callus cells cultivated in a bioreactor (Schwitzguébel *et al.*, 1991; Leathers *et al.*, 1992; De Jesus, 1995). Nevertheless, due to the weak production and excretion levels and the low economical value of food pigments in comparison with the strong cost of the cell culture process, their work was stopped. More recently other groups developed the production of betalains in hairy root culture from *Beta vulgaris* (Mukundan *et al.*, 1998). Despite numerous improvements made in betalain synthesis in hairy root culture and in the down-stream processing, such as a better pigment excretion level in the presence of an acidic solution (Mukundan *et al.*, 2001) or the development of a new medium for high betacyanin production in cell suspension culture (Akita *et al.*, 2002), the costs seem still too high for commercial applications.

Consequently, the development of new varieties of high pigment yellow beet by overexpression of *DODA* gene in the presence of a strong promoter could be an alternative far cheaper strategy, once the variety will be developed. This could be completed by an introgression step of the interesting character in fodder beet varieties already widely used in agriculture for their growth properties and their robustness to pathogens and to environmental conditions. It is important to notice that the modification of a pathway is not an easy process due to internal regulations and interactions with others pathways. In our case, DODA protein catalyses the last enzymatic step of betaxanthin biosynthesis and therefore needs to have a sufficient amount of its DOPA substrate. Therefore, the simultaneous overexpression of the tyrosinase would also have to be considered, as well as the

vacuolar transporter. Amino acids necessary for the final spontaneous conjugation in the vacuole should be in sufficient amount because betaxanthins can be synthesised from many of them.

The availability of such yellow beet varieties will allow for example the replacement of the artificial food colourant tartrazine termed FD&C Yellow N°5 or E102. This dye is still authorized in the U.S.A. and many other countries despite numerous cases of severe allergy like asthma observed following ingestion of food containing this chemical. This colourant can also generate an anaphylactic shock by cross-reaction with the frequently used aspirin medicine (Wuthrich, 1993). A recent toxicity study supported the use of betalain pigments as a food additive by demonstrating clearly their safety (Reynoso *et al.*, 1999).

Due to the general rejection of genetically modified organisms (GMO) in Europe, an alternative strategy could be a step-by-step marker-assisted selection of new table beet varieties with an enhanced level of betalain pigment concentration based on the evaluation of *DODA* mRNA expression by quantitative PCR and selection of the seedlings expressing it at most. A RAPD marker-assisted strategy had been used to select novel high pigment red beet variety (Eagen and Goldman, 1996). An introgression step will also be necessary in this case. Marker-assisted selection combined with an introgression step could take more time to obtain new beet varieties with a sufficient amount of pigment than the direct transformation, but it has the advantage to be far quicker at the time of its introduction to the market, needing few authorizations in comparison with GMO plants, the cultivation of which is for the moment forbidden in most European countries.

5.2 Betalains and their importance in the modification of flower colours

The floricultural industry has focused its attention on the development of longer living and novel coloured cut flowers. The basis for the latter point was laid down some years ago through the isolation of 'blue' genes (de Vetten *et al.*, 1999) and ethylene biosynthesis genes. The addition of a new branch to the phenylpropanoid pathway of white petunia varieties by introduction of a chalcone reductase cDNA from *Medicago sativa* under the control of a 35S promoter, allow changing the flower colour from white to yellow (Davies *et al.*, 1998). A similar strategy was followed successfully to obtain petunia coloured in orange due to the novel synthesis of pelargonidin (Oud *et al.*, 1995). Genetically commercially modified flowers are now available in many parts of the world. The first market introduction occurred in 1996 in Australia, when the firm Florigene launched the mauve Moondust^{TM} carnation, which is now produced and commercialized in the U.S.A and Japan. Since then many others modified flowers have been produced.

In this context, we suggest the use of *DODA* gene for engineering colour modified flowers or fruits in Caryophyllales plants. This order contains a great number of horticultural flowers like tricolour amaranth, *Celosia*, *Bougainvillaea*, *Gomphrena* or *Cactaceae*. It would be possible to intensify or inhibit their colour. The latter is providing white-flower varieties, which do not exist at this time in certain species of cactus for example. In a similar manner, the production of white *Chrysanthemum* has been developed by antisense gene construct or sense-suppression strategy (Courtneygutterson *et al.*, 1994). Others alternative methods like RNA interference or direct mutagenesis of the known catalytic amino acids of DODA protein can be used to produce unpigmented varieties. The latter strategy was applied recently in petunia to produce the orange pelargonidin pigment (Johnson *et al.*, 2001). Petunia does not produce orange flowers because dihydroflavonol 4-reductase (DFR) from this species, an enzyme involved in anthocyanin biosynthesis, inefficiently reduces dihydrokaempferol, the precursor to orange pelargonidin-type anthocyanins. The substrate specificity of DFR, however, has not been investigated at the molecular level. By analysing chimeric DFRs of Petunia and Gerbera, Johnson and colleagues identified a region that determines the substrate specificity of DFR. Furthermore, by changing a single amino acid in this presumed substrate-binding region, they developed a DFR enzyme that preferentially reduces dihydrokaempferol, demonstrating that the substrate specificity of DFR can be altered by minor changes in DFR.

Horticultural market for new coloured varieties seems to have an important potential (Mol *et al.*, 1999). Fruit colouration is also a determinant factor in the buying process. Consumer concern over the introduction of GMO remains an issue in the successful commercialisation of all transgenic plants, although recent surveys showed that people are more comfortable with the application of gene technology to flower crops than to food crops. Nevertheless, we can wonder whether similar applications are necessary, because they do not meet any need other than human curiosity.

5.3 Betalain as a tool for facilitating transgenesis

A method for visually selecting transgenic plant cells by betalain pigmentation

In the presence of transgenic overexpressed *Pg* DODA, DOPA is directly transformed into yellow betaxanthin pigment and accumulates in the vacuole. The development of this pigmentation can be used to select visually transgenic plant cells from non-transgenic plant cells in white sugar beet, thus reducing the tedious regeneration work usually necessary. At the plant level, the pigment accumulation could be used to visually identify proprietary transgenic Caryophyllales agricultural crops like the South-American *Chenopodium quinoa* or amaranth widely cultivated there for their high nutritive grains. A similar approach has already been developed by Seminis vegetable with the carotenoid pigmentation (Patent WO9714807).

.

5.4 Betalain as an antioxidative nutraceutical

Another important aspect of betalains is based on their antioxidative properties. These pigments contain a phenolic and a cyclic amine group which are both shown to be very good electron donors, acting as antioxidant (Kapadia *et al.*, 1996; Escribano *et al.*, 1998). Recently a patent (US20003036565) was deposited on extraction methods of betalain pigments from high pigment beet in order to use it as cancer chemopreventive agents. Inventors observed an induction of quinone reductases in murine hepatoma cells in the presence of betalains. These enzymes are considered as protectors against carcinogens (Fernandes *et al.*, 1996). A few months ago, the antioxidative protective effects were also tested by measuring the resistance of low density lipoprotein (LDL) to copper-induced oxidation in the presence of different betalain pigments (Tesoriere *et al.*, 2003). These results show that indicaxanthin and betanin can bind to LDL, thus preventing copper-induced lipid oxidation. Interaction with vitamin E appears to add a remarkable potential to indicaxanthin in the protection of LDL. Although molecular mechanisms remain incompletely understood, various aspects of the action of betanin and indicaxanthin in preventing LDL lipid oxidation are discussed in this paper. Betaxanthins were shown to have the highest antioxidative activity. The relationship between the chemical structure and antioxidant activity of betalains was examined by Zakharova and Petrova (Zakharova and Petrova, 1998).

Natural antioxidants are now thought to possibly prevent the increasing incidence of many western diseases such as cancers and heart disease which are directly linked with a lack of fruit and vegetables in the diet. In this context, it was important to study the effects of the pH and the temperature on betalains antiradical activity. Recently, it was demonstrated that the antiradical activity was greater at acidic pH and lower at higher temperatures (Pedreno and Escribano, 2001).

More concretely, the preventive administration of the prickly pear juice (*Opuntia ficus-indica*) inhibits the ulcerogenic activity of ethanol in rat (Galati *et al.*, 2003). Light microscopy observations showed an increase in mucus production and the restoration of the normal mucosal architecture. The juice is nutritionally interesting, and its dietary intake could provide protection against oxidative damage. The cancer chemopreventive potential of beetroot extract was investigated in three different chemical carcinogen initiation-promotion experimental tumour models in mice (Kapadia *et al.*, 2003). Oral administration of 0.0025% betanin inhibits significantly the mouse skin tumour promotion. The most interesting observation is that the cancer

chemopreventive effect was exhibited at a very low dose, thus indicating that beetroot warrants more attention for possible human applications in the control of malignancy.

All these results support the development of betalains based nutraceutical products in order to improve people's health. Red beet products used regularly in the diet may provide protection against certain oxidative stress-related disorders in humans (Kanner *et al.*, 2001).

5.5 Conclusion

At the beginning of the 21st century it is predicted that many colours will be used for both their colouring effect as well as nutraceutical value in the growing health food market (Downham and Collins, 2000). Due to these new commercial applications, required amounts of betalain pigments will increase. Thus, we suggest a higher level of expression of *DODA* gene in betalain crops as one answer to this problem, either by transgenesis or by marker-assisted selection.

Further analysis of the enzymatic properties of DODA protein could allow other developments such as a colourimetric detection test of DOPA in tissues. Considering the existence of homologues in bacteria and plant without betalain, it could be also possible to isolate these new enzymes catalysing the opening of an aromatic ring, leading perhaps to the detoxification of aromatic compounds.

III. GENERAL CONCLUSIONS

Isolation of genes responsible for betalain pigment biosynthesis in plants was the main objective of this PhD thesis project. For this purpose, we employed a PCR subtractive hybridization method to select betalain specific cDNAs in violet petals from *P. grandiflora* plant. Three cDNAs are expressed in correlation with betalain biosynthesis. *A.16* and *V.33* cDNAs correspond to an ORF, whereas *P.34* does not. As *V.33* cDNA had no homologue and was already studied in a previous PhD thesis work (Zaiko, 2000), and *P.34* cDNA elongation did not allow identification of an ORF. Therefore, we left them and focused our work on *A.16* cDNA. A.16 translated protein is strongly homologous to *Arabidopsis thaliana* acyl-CoA oxidase 3, but does not contain a FAD binding domain necessary for the putative ACX activity. This could explain the incapacity of A.16 protein to complement the *Saccharomyces cerevisiae fox-1* mutant deficient in ACX activity, whereas *Arabidopsis* ACX protein does. A.16 protein could be involved in the beta-oxidation cycle as an ACX, but its activity would not require the fixation of FAD as cofactor, or would imply another cofactor or another protein carrying the FAD cofactor. In order to verify if the beta-oxidation cycle is stimulated in the presence of betalain biosynthesis, it will be necessary to control the level of expression of other enzymes of the cycle such as the 3-Ketoacyl-CoA thiolase. In the absence of a proven ACX function, we can notice that A.16 contains a partial dehrydrogenase domain. Finally, we conclude that the link between A.16 protein and the betalain pathway is not obvious.

A similar subtractive hybridization experiment was previously made between yellow and white petals from *P. grandiflora* (Zaiko, 2000). Three betalain-specific cDNAs corresponding to an ORF were identified. However, two of them remained without function. By aligning regularly these proteins with protein databases, we identified *L.6* cDNA as putative type III extradiol 4,5-dioxygenase with all catalytic amino acids from LigB protocatechuate 4,5-dioxygenase conserved. This function corresponds to the DOPA-dioxygenase gene *C* proposed in the “three-gene model” for betalain biosynthesis in *Portulaca grandiflora* (Trezzini, 1990a). Consequently, we were interested again more closely in this gene and called it *DODA* for DOPA-dioxygenase. Northern blot analysis confirms the correlated expression with betalain biosynthesis of this gene in *P. grandiflora*. Similar results were obtained at the protein level. DODA expression was rather low in pigmented tissues. Finally, we confirmed DODA as a DOPA-dioxygenase by biolistic

complementation of *Pg* white petals deficient in DOPA-dioxygenase gene *C* (Christinet *et al.*, 2004). HPLC analysis showed that yellow and violet produced pigments were identical to pigments synthesised in natural *P. grandiflora* yellow and violet flowers.

Numerous *Pg* DODA homologous sequences were deduced from plants and bacteria ESTs, thus allowing us to make a phylogenetic study and to try answering the question of the betalain pigment origins. *DODA* genes from betalain-containing plant species are clustering together apart from *DODA* genes from non-producing betalain plants and from homologous bacterial genes. This result was confirmed by the identification of a conserved pattern in DODA from betalain-containing plants. This [H-P-(S-A)-(N,D)-x-T-P] betalain-specific pattern replaces a [H-N-L-R] pattern present in all other *Pg* DODA plant homologues and a [H-N-L-x] pattern conserved in all bacterial and archaeabacterial homologues (Christinet *et al.*, 2004).

The position of these two different conserved patterns next to the catalytic His177 essential for the dioxygenase activity, suggested us their implication on the DOPA-dioxygenase activity. For this purpose, we constructed a three-dimensional model of *Pg* DODA and of its *Physcomitrella patens* homologue DIOXA, based on the LigAB crystal structure 1B4U. Comparison of the two modeled active sites shows that their global geometry was rather similar and suggests an implication of these different conserved patterns in the access of the aromatic substrate to the active site. Therefore, it would be interesting to further analyse the functional importance of each residue of the betalain-specific pattern by a *Pg* DODA point-mutation experiment coupled with a functional test of this mutated protein by biolistic transformation of *Pg* white petals. In the same manner, it will be interesting to modify a DODA homologue from a non-Caryophyllales plant in such way that it can use DOPA as a substrate.

Only a few mutations are needed to switch from non-betalain producing plants DIOXA dioxygenase to betalain producing DODA DOPA-dioxygenase. The natural evolution of enzymes can eventually lead to new catalytic function or modified specificity. This can happen in two ways: either by formation of a new active site on the protein framework or by the transformation of the old active site for a new function or new specificity. Therefore, we hypothesized from the results we obtained, that betalain biosynthesis pathway takes its origins in the recruitment of an already existing plant metabolic pathway, thus replacing the anthocyanins deficiency in plants from the order Caryophyllales. Presence of a single conserved pattern in *DODA* from all betalain-producing species analysed, suggested the existence of a unique event of mutations in a species,

corresponding to a primitive ancestor of the betalain plants. This hypothesis is in agreement with the monophyletic origin of the Caryophylalles. It is possible that several events occurred in the past, but only one was successful and thus was subsisting in all betalain-producing plants. We tried to isolate *DODA* gene in *Silene vulgaris* and *Dianthus caryophyllus* species because they are phylogenetically related with betalain-producing plants considering several botanical criteria, except they synthesise anthocyanins. Nevertheless, we always obtained non-specific amplifications. Therefore, we did not find which of the two conserved patterns was present in their DODA protein. In the case of the existence of a pattern identical to the betalain one, this would have supported the coexistence of both pigment types in a Caryophyllales ancestor, as proposed by Clement and Mabry (1996). In the presence of the plant conserved pattern, this would confirm the independent development of the betalain biosynthesis pathway. Floral colour change has clearly evolved independently many times, most likely in response to selection by visually oriented pollinators, and reflects a widespread functional convergence within the angiosperms (Weiss, 1995).

By comparing the plant DOPA-dioxygenase DODA with the *Amanita muscaria* DODA, we noticed their complete amino acid sequence dissimilarity. Even the conserved LigB catalytic residues were not present in *Am* DODA. As we know that both *A. muscaria* and *P. grandiflora* DODA are able to complement *Pg* white petals by biolistic transformation, we consider them as functional equivalent. These observations lead us to conclude that the appearance of betalains in fungi was independent from that in plants. Betalain biosynthesis occurred on two different occasions and thus constitutes a good example of biochemical functional convergence.

The capacity to produce thousands of structurally diverse natural products due to secondary metabolism is considered a typical feature of plants and microbes. Although the huge number of secondary metabolites known today, there is still a clear trend to link each product with specific groups of organisms. Thus, a variety of structures, biosynthetic pathways, protein and genes are believed to be specific to plants, raising the question of the origin of biosynthetic pathways. Are these pathways all plant inventions without the use and the adaptation of prokaryotic genetic information?

In contrast to this generally accepted consideration, DODA phylogenetic analysis shows that genetic information related to plant specific pathways is in fact already present in several bacteria or archaea species and is used in a similar way. Plant genome sequencing programmes demonstrate that many non-heme iron(II)-dependant dioxygenase remain to be characterized.

Around one-hundred dioxygenases were predicted for *Arabidopsis thaliana*, but to date only 20% are known. Further three-dimensional structures are likely to assist with identification of dioxygenases from primary sequences.

We still need to better characterize the biochemistry of *Pg* DODA. Up to now, we have met some difficulties in detecting *Pg* DODA in vitro activity from a crude extract protein from *E. coli Pg* DODA overexpressing strain. The protein might be not stabile or non-functional in the different condition that we tested. Overexpression of a plant protein in bacteria could sometimes be problematic due to different patterns of glycosylation. Therefore we are currently transitorily expressing *Pg* DODA protein in the *Physcomitrella patens* moss protoplasts, thus avoiding such problems.

The identification of the plant DOPA-dioxygenase DODA, catalysing the last enzymatic reaction of the betalamic acid biosynthesis, opens the way to study the architecture of the betaxanthins and betacyanins pathways. Those pathways could be engineered to further produce high amount of natural water-soluble yellow or violet pigments in beetroot and could provide visual markers in genetic transformation. Nevertheless, it is important to be aware that pigmentation is generally regulated by trans-nuclear factors.

IV. MATERIAL AND METHODS

1. Plant cell and culture cultivation

We grew *Portulaca grandiflora* (*Pg*), *Mammilaria sp*. and *Iresine sp.* in a greenhouse under normal daylight conditions. *Pg* buds (6-8 mm) were collected from vegetatively propagated clones obtained from quasi-isogenic lines established in our lab (Trezzini, 1990b), their basal part cut-off, and petals extracted and frozen in liquid N_2. Beetroot calli cell cultures displaying different colours were selected from *Beta vulgaris* (*Bv*) Bikores monogerm (Girod and Zryd, 1991b) and cultivated on a maintenance medium at 26°C with a 16hr photoperiod (light from GRO-LUX WS SYLVANIA, 6mW/m^2). We collected samples for the RNA extraction by vacuum filtration of suspension cultures; these samples were stored frozen at –80°C and grinded in liquid N_2 just before extraction. The moss *Physcomitrella patens* (*Pp*) was grown in the same conditions as beet cell cultures on solid minimal medium (Ashton *et al.*, 1979) supplemented with 2.7 mM ammonium tartrate and 50 mM glucose.

2. RNA isolation and construction of a cDNA subtractive library

Total RNA from *Pg* was isolated from immature petals (1 g) using the hot phenol method (Rochester *et al.*, 1986), except that we used 5 x larger volumes throughout to reduce the high viscosity of the extracts. Messenger RNA was purified using the PolyATract ® mRNA isolation system (Promega ®) according to the suppliers' protocol. The Smart PCR cDNA Library Construction Kit (CLONTECH ®) was used to amplify a sufficient amount of cDNA from *Pg* violet, yellow and white immature petals for the construction of a subtractive cDNA library. cDNAs were purified on a chroma spin-1000 column (Clontech ®), digested by *RsaI* enzyme and purified on a PCR purification column (Quiagen ®). Four subtraction experiments were performed using PCR-Select cDNA Subtraction kit (Clontech ®) according to manufacturer's instructions: two direct subtractions where yellow and violet cDNA were used as testers and white cDNA as a driver, and two reverse subtractions with white cDNA as a tester. After PCR-based subtractive

hybridization, cDNA fragments were cloned into pGEM-T-easy vector (Promega ®) and transformed in *E. coli* DH5α strain.

For comparative studies, we isolated *Bv DODA* cDNA by RT-PCR from yellow *Beta vulgaris* callus total RNA with nested primers designed from the 5'UTR region of a partial EST (gi14522284) and nested primers annealing with the added adaptators from Genome walker kit (Clontech ®).

3. Northern blot analysis

Northern blot analysis and other molecular biology techniques were described in Molecular Cloning (Sambrook, 1989). We blotted 5-20 μg total RNA per lane onto a Zeta-Probe GT membrane (Bio-Rad ®). For hybridization, we used [^{32}P]dCTP-labelled-DNA probes in 0.5 M phosphate buffer (pH7.2), 7% SDS, 1 mM EDTA at 60°C. We washed the blots under stringent conditions and exposed them using X-OMAT AR film (Kodak ®).

4. Full-length cDNA isolation

Undigested *Pg* yellow-specific total cDNA amplified using capFinder cDNA Construction Kit (Clontech ®) was ligated to DNA adaptators from Universal Genome Walking kit (Clontech ®) containing annealing sites for two long primers (AP1, AP2) that follow one another. Two gene-specific primers were designed in the same manner from the partial *Pg DODA* cDNA (*PgDODA*p2, *PgDODA*p3), thus allowing a two-steps amplification of the missing 5' regions of *Pg DODA* cDNA with annealing and extension performed at 68°C in the presence of Expand High Fidelity Taq polymerase mix (Roche). We reconstituted the full-length cDNA from the extremities by high fidelity PCR amplifications and sequenced on both strands.

5. Southern blot analyis

We extracted *Pg* genomic DNA from young leaves and stems ground in liquid N_2 with the DNeasy Plant Maxi kit (Quiagen ®). We digested 4 µg of DNA with the restriction enzymes *Eco*RI and *Nhe*I (Invitrogen ®) and separated them by electrophoresis on a 0.7% agarose gel. A ^{32}PdCTP-probe was made from a *Nco*I-digested 789 bp-fragment of *Pg DODA* cDNA. Blotting and hybridization conditions were the same as described for the Northern blot analysis. We washed the filters progressively in SSC buffer from 70% to 95% of homology and exposed using X-OMAT AR film (Kodak ®).

6. Production of *Pg* DODA specific antibody and Western blot analysis

Rabbit *Pg* DODA polyclonal antibodies were produced by injection of a 16 amino acids peptide (H_2N-CRYEDVNNYQTKAPEG-$CONH_2$) selected from *Pg* DODA sequence for its antigenic properties and specificity, and followed by an affinity purification of the obtained serum on a protein A column (Eurogentec®, Seraing, Belgium). Crude plant proteins extracts were obtained by grinding tissues in the CCLR lysis buffer (Promega®). Proteins were quantified by the Bradford method (Bradford, 1976). SDS-PAGE was performed on a Miniprotean II (Bio-Rad®, Munich, Germany) according to Laemmli (1970) with the Low Range molecular weight marker (Bio-Rad®). Proteins were blotted on a Trans-Blot® nitrocellulose membrane by electro-transfer according to Bio-Rad® manual instruction. Transferred proteins were revealed by incubating the membrane 1 min. in Red Ponceau solution followed by a wash with water. Immunodetection was done with the Immun-Star HRP kit (Bio-Rad®), *Pg* DODA specific antibody as primary antibody (diluted 3500x) and an anti-rabbit IgG HRP-conjugate antibody (Promega®) as secondary antibody (diluted 2500x). Detection of the HRP-conjugate antibody was performed by 5 min. incubation of the membrane in the chemiluminescence immune-Star HRP reagents mix (Bio-Rad®) and exposed using Hyperfilm ECL film (Amersham®).

7. Expression study by particle bombardment and pigment analysis

Pg DODA full-length cDNA was amplified with high fidelity Taq polymerase and modified primers. Sense primers contained *XbaI* and antisense *PstI* restriction sites to facilitate cloning. The sequence before the start codon was also corrected according to Kozak sequence to facilitate the initiation of the translation (Kozak, 1991). Amplified cDNA was digested with *XbaI* and *PstI* and purified. The plant expression vector pNco *PgDOD* was constructed by placing the *XbaI /PstI* digested *Pg DODA* cDNA in pNco vector (Mueller *et al.*, 1997a). The cloned cDNA was constitutively expressed under the control of a CaMV 35S promoter and terminated by CaMV polyadenylation signal. We submitted the double-stranded DNA of this construct for sequencing at Microsynth Gmbh facilities.

We performed ballistic transformations of young white petals from flower buds with the pNco construct on gold particles (Mueller *et al.*, 1997a). Petals from plants homozygous for the *cc* locus (white phenotype) with different genetic background for gene *R* and *I* (hidden violet or yellow phenotypes), were tested (see (Trezzini, 1990a) for details). We counted the number of coloured spots per petals after 24 hours of incubation at 26°C under light. A positive control of transformation was done in parallel and simultaneously with a construct containing the cDNA encoding the red fluorescent protein DsRed2 (Clontech ®) driven by 35S promoter and terminated by Nos polyadenylation signal (p35S *DsRed2*, Finka A. in our laboratory unpublished).

Violet and yellow spots synthesised in pNco *PgDOD* transformed petals were extracted with acidified methanol and analysed by HPLC (Waters) using a Hypersil® column (Bischoff, Leonberg, Germany) as described previously (Trezzini and Zryd, 1991b). We compared for identification the spectra and retention time from the pigments produced by ballistic transformation with the naturally occurring betalain pigments from yellow and violet *P. grandiflora* petals.

8. Homologous genes and partial genomic sequence isolation

The *Arabidopsis thaliana* homologue for *Pg DODA (AtL.6*) was isolated by PCR amplification with primers *AtL.6*p1rev and *AtL.6*p2rev designed from the complete mRNA AY050948 on the 5' sequenced homologue EST N65678 (Newman *et al.*, 1994) from the ABRC Lambda-PRL2 EST library (Davis and Ware, 1998). The *Physcomitrella patens* homologue for *Pg DODA,* called *DIOXA* (AJ583016), was amplified by nested PCR strategy on Lambda-UniZAP cDNA library (Girod *et al.*, 1999) with long primers *PpDIOXA*p5 and *PpDIOXA*p4 designed from the partially sequenced EST BJ195116 corresponding to the 5' region of the gene (Nishiyama *et al.*, 2003). The *Beta vulgaris* homologue for *Pg DODA* (AJ583017) was isolated from yellow callus total RNA by RT-PCR amplification strategy. DNA adaptators from Universal Genome Walking kit (Clontech®) containing annealing sites for two long primers (AP1, AP2) were ligated to the amplified cDNA according to the manufacturer prescriptions. Gene specific nested primers *BvDODA*p1 and *BvDODA*p4 were designed on the 5'UTR-region from EST sequence BI095902 (los Reyes, unpublished) and used in combination with nested primer AP1 and AP2 to amplify the full-length cDNA.

9. Yeast LiAc Transformation

1. Precultures

20 ml YPD / Selective Media + 20 µl Kan G418 (! If mutants!))
1 colony of Yeast from a selective agar plate during 2 – 3 days at 28°C and ~230 rpm

2. Cultures

100 ml YPD / Selective Media + 100 µl Kan G418 (! If mutants!))
100, 50, 20 µl Yeast pre culture (≠ dilution), O/N at 28°C at 230 rpm until OD600 ≈ 0.4 – 0.9

3. Preparation of the cells (!Use cut tips!)

Pour the cells into two 50 ml Falcon tubes (T° < 20°C) and spin 2 min at 3'000 rpm
Resuspend the totality of the cells in 15 ml Li-TE buffer and spin 2 min at 3'000 rpm
Resuspend the cells in 1 ml Li-TE buffer and leave at room T° < 30 min to prepare the DNA

4. Transformation (!Use cut tips!)

Mix in a 1.5 ml sterile Eppendorf tube:
1 – 5 µg plasmid (max. 5 µg)
50 – 100 µg denaturised salmon sperm DNA (5 – 10 µl of 10 mg/ml)
100 µl cells
125 µl 63% PEG 4000 in Li-TE buffer

Mix gently by pipetting up and down and incubate at 30°C during 30 min with occasional mixing
Heat shock at 42°C during 15 min and then centrifuge the cells 1 min at 2'000 rpm
Resuspend the pellet in 200 µl selective media and plate 20 – 50 µl of cells on selective agar plates
Put at 28°C for 2 – 4 days

Yeast Li-Acetate Transformation Reagents:

<u>1x Li-TE buffer</u>
0.1M LiOAc
0.01M Tris HCl pH 7.5
0.001M EDTA

Check pH at 7.5; adjust with dilute acetic acid if necessary and autoclave this solution

<u>63% PEG 400 in Li-TE buffer</u>
PEG 4000 63g
1x Li-TE buffer 100ml
Dissolve the PEG 4000 with a magnetic bar and leave it in the bottle and autoclave this solution
Right after the autoclaving, stir again the solution

10. Crude Yeast PHA Extraction

1. Pre-culture Work in sterile conditions!

In a 50 ml Falcon tube put:
x ml Selection media
1 Yeast colony and incubate 2 days at 28°C and ~230 rpm
Spin the cells 2 min at 3'000 rpm
Resuspend the cells in x/2 ml autoclaved water

2. Feeding Work in sterile conditions!

Generally used FA feeding conditions
0.1% Glucose
0.1% Oleic acid / 17:1 Δ10 or (0.1% Lauric acid)
2% Pluronic
! X is calculated with the number of 20 ml cultures to be done!

In a 15 ml Falcon tube mix by vortexing:
x/1000 ml Oleic acid / 17:1 Δ10 ! Oleic acid at 4°C is solid → incubator for solubilization!
5 -10 ml 10% pluronic

In a 100 ml sterile Erlenmeyer mix and heat (w/microwave or on heating plate):
x/1000 g Lauric acid
10 ml 10% pluronic

In x ml of Yeast media /No dextrose/2% pluronic/a.a. solution add:
x/100 ml Dextrose 10%
Totality Oleic acid or 17:1 Δ10 / 10% pluronic solution previously prepared
(Totality Lauric acid / 10% pluronic solution previously prepared)

Put 20ml in sterile Erlenmeyer
Add 2 ml (dilution 1:10) of yeast pre-culture
Incubate 3 – 5 days at 28°C; ~230 rpm

3. Lyophilization No use to work in sterile conditions from this point on!

Transfer the 20 ml cultures in 50 ml Falcon tubes
Spin 2 min at 3'000 rpm
Resuspend the totality of the cells in 5 ml autoclaved water
Spin 2 min at 3'000 rpm
Repeat last two steps one more time
Resuspend the cells in 1 ml autoclaved water
Transfer in small glass tube with black screw caps
Put HORIZONTALLY (!) at -80°C or -20°C for 2 – 3 hours
Put the tubes w /caps a bit opened in the lyophilization machine when at -50°C O/N (or even more)

4. Transesterification Use glass tubes and pipets!

Weigh the lyophilized cells by transferring into new glass tubes
Wash the cells w /5 ml methanol (MeOH)
Boil the cells for 15 min at 65°C
Spin 5 min at maximum speed
Take out the supernatant w /a vacuum system
Repeat last four steps two more times

During the last wash, prepare:

1/ 10 ml MeOH
10 µl A4
2/ 10 ml $CHCl_3$
300 µl H_2SO_4

Pour GENTLY the two mixes in one Erlenmeyer (or aliquot separately in each sample glass tube).
Add 1 – 2 ml of prepared mix in each tube containing a magnetic stirrer
Prepare also one blank tube containing only the mix
Put on the heating plate for 4 hours at 94°C! Be sure that the caps are well tightened!
Take the tubes off the plate and wait until they cool down (2 – 5 min)
Add 1 Vol of NaCl 0.9% Mix vigorously by vortexing
Spin 5 min at maximum speed
Take off the interphase and the aqueous phase w /a vacuum system
Add ½ Vol of NaCl 0.9% Mix vigorously by vortexing
Transfer the chloroform phase (bottom phase) into prepared vials containing some $MgSO_4$
Transfer the chloroform into an insert when the $MgSO_4$ is not floating any more

Samples are now ready to be loaded on the GC-MS or kept in the -20°C freezer for an ulterior use.

11. Three-dimensional preliminary modeling

3D preliminary models of *Pg* DODA and *Pp* DIOXA proteins were created with Swiss-PdbViewer (Guex and Peitsch, 1997) from the *Sphingomonas paucimobilis* crystallized structure PDB 1B4U as a template. *Pg* DODA and *Pp* DIOXA sequences were aligned manually with the structure 1B4U based on the same alignment as described previously. The longer length of the bacterial sequence generated gaps in the modeling program; we introduced a carbon chain of the same size as the bacterial amino acid-sequence length to replace those gaps. We did no attempt to correct the errors that were located away from the active site amino acids.

12. Computational anlaysis

Similarity search of amino-acid and nucleotide sequences of *Pg* DODA was performed using the different options of the 'BLAST' algorithm (Altschul *et al.*, 1990; Gish and States, 1993). The NCBI CDD domain search tool (Marchler-Bauer *et al.*, 2002) was used to identify the LigB domain (pfam02900). Sequence analysis was performed with the Vector NTI® software package (Invitrogen). Sequence alignments were made using Clustal W software (Thompson *et al.*, 1994) and edited manually with Jalview tool (www.jalview.org). PHYLIP was used for phylogenetic analyses (Felsenstein, 1989).

ANNEX 1

Isolated genomic DNA and its corresponding ORF

Exons are indicated by underlined characters.

>*Portulaca grandiflora DODA* (AJ580598)

ATGGGTGTTGGGAAGGAAGTGTCGTTTAAGGAGAGTTTCTTCTTGTCTCATGGGAATCCAGCCATGTTGGCGGATGAAT
CCTTTATAGCGAGGAACTTCCTTTTGGGATGGAAAAAGAATGTGTTTCCTGTGAAACCCAAGTCCATTTTAGTAGTCTC
TGCTCACTGGGAGACTGATGTCCCTTGTGTATCTGCTGGTCAGTACCCTAATGTTATCTACGATTTCACTGAAGTCCCT
GCTTCCATGTTTCAGATGAAGTACCCAGCTCCAGGGTGCCCAAAGCTGGCAAAAAGAGTGCAGGAACTGCTGATAGCGG
GAGGTTTCAAGAGCGCGAAGCTAGATGAGGAGCGGGGTTTCGACCACAGCTCATGGGTGCCACTGAGCATGATGTGCCC
GGAGGCGGACATCCCGGTGTGCCAGCTTTCGGTGCAGCCTGGGCTTGACGCGACCCACCACTTCAACGTGGGGCGAGCG
TTGGCCCCACTCAAAGGGGAAGGTGTCCTCTTCATTGGCTCCGGTGGGGCTGTCCACCCTTCTGATGACACCCCTCATT
GGTTCGATGGTGTTGCTCCCTGGGCTGCTGAGTTTGATCAATGGCTTGAGGATGCTCTCCTAGAAGGAAGGTACGAAGA
TGTGAATAACTATCAAACAAAAGCACCAGAAGGGTGGAAGCTAGCACATCCAATTCCAGAACATTTTCTACCATTGCAT
GTAGCCATGGGTGCAGGTGGTGAGAAATCAAAGGCAGAGCTTATTTATCGTACTTGGGATCATGGTACTCTTGGCTACG
CCTCTTACAAGTTCACTTCCATATGA

>*Portulaca grandiflora* DODA (CAE45178)

MGVGKEVSFKESFFLSHGNPAMLADESFIARNFLLGWKKNVFPVKPKSILVVSAHWETDVPCVSAGQYPNVIYDFTEVP
ASMFQMKYPAPGCPKLAKRVQELLIAGGFKSAKLDEERGFDHSSWVPLSMMCPEADIPVCQLSVQPGLDATHHFNVGRA
LAPLKGEGVLFIGSGGAVHPSDDTPHWFDGVAPWAAEFDQWLEDALLEGRYEDVNNYQTKAPEGWKLAHPIPEHFLPLH
VAMGAGGEKSKAELIYRTWDHGTLGYASYKFTSI

>*Beta vulgaris DODA* cDNA (AJ583017)

GAATTCACTAGTGATTCAATCTCAACCCCAGAAATCAACAGAAGAACATTACTCAAATTCCTAATTAGCTCAAATCTGA
AAATGGGTAGTGAAGATAACATCAAAGAAACCTTCTTTATTTCTCATGGAACTCCAATGATGGCAATTGATGATTCAAA
GCCTTCTAAAAAATTCCTTGAAAGTTGGAGAGAGAAAATCTTCTCTAAAAAACCCAAGGCTATTCTCGTCATTTCTGCT
CACTGGGAAACTGATCAACCTTCTGTTAATGTTGTAGACATCAATGATACCATCTATGATTTCAGAGGTTTTCCTGCTC
GTTTGTACCAGTTTAAGTACTCGGCTCCTGGATCTCCAGAACTGGCAAACAGGATACAAGATCTTCTCGCAGGATCTGG
CTTCAAATCTGTAAACACTGACAAGAAACGAGGGCTTGATCACGGTGCATGGGTGCCTCTAATGCTGATGTATCCTGAA
GCTGACATCCCTGTTTGTCAGCTCTCTGTCCAATCGCATTTGGATGGAACACACCACTATAAGTTGGGACAAGCGTTGG
CTCCTCTAAAGGATGAAGGTGTCCTCATCATTGGTTCCGGAAGTGCAACACACCCTTCAAATGGTACCCCTCCCTGTTC
TGATGGAGTTGCTCCGTGGGCTGCAGCTTTTGATAGCTGGCTTGAAACAGCTCTGACAAATGGAAGCTATGAAGAAGTG
AACAAATATGAAACTAAAGCACCAAACTGGAAATTGGCTCATCCTTGGCCTGAGCACTTTTATCCACTGCATGTAGCCA
TGGGTGCCGCTGGTGAAAATTCGAAGGCAGAACTTATTCACAACAGTTGGGATGGTGGTATCATGTCGTATGGCTCCTA
CAAGTTCACTTCCACCTAATTTCAACATTGTAATACTTGTTTTGTTTGTATCATCTCTTCAGCCCATTGGTATTCTCTT
CAAGAATGAATTAATCAGATGGAGTATATATATATCTATATGCTTATATATCTGAACTCATGTAGTTATTATGCTAATT
ATTACATTTTCAAGCCTTTTGTCTTATAGTAGTTTCTGGTCTGGTCTGTGAATTCTTGCAGAAATTATGGGCTAACCAT
TTGGATATTGATTGACTGGTGAAATTAATCAGGTGTGATTCCCGAAAAAAAGAAAAAAAAAAAAAAAAAA

>*Beta vulgaris* DODA (CAE7100)

MGSEDNIKETFFISHGTPMMAIDDSKPSKKFLESWREKIFSKKPKAILVISAHWETDQPSVNVVDINDTIYDFRGFPAR
LYQFKYSAPGSPELANRIQDLLAGSGFKSVNTDKKRGLDHGAWVPLMLMYPEADIPVCQLSVQSHLDGTHHYKLGQALA
PLKDEGVLIIGSGSATHPSNGTPPCSDGVAPWAAAFDSWLETALTNGSYEEVNKYETKAPNWKLAHPWPEHFYPLHVAM
GAAGENSKAELIHNSWDGGIMSYGSYKFTST

>*Iresine sp. DODA*

ATGTCGATCGATGAGTCGATTCCTGCAAGACGCTTTCTTGAGGAATGGAAAGACAAAGTCTACTCCAAGAGGCCCAGCT
CCATTCTTGTGATCACTGCTCACTGGCTGACCGTTGTGCCCACTGTCAGTGCCATCGACCACTCCGATCTCATCTACGA
TTTTGGAGGCTTCCCTGCTTGCATGTACCAGGTCTGTATCTCCCTCTCTCCTTCCTTCCTCCCACTCCAAATTCAAAAT
CGATTTTCATATAGAATAACCACAATATAATCTTTCCTTTTCTTTTGGATTCTTGTGTTGGGGTTGGTAAAAATGAGAG
GGAGTGAATGGTTAAATCAACATTTGTGGAACTTGCAAAATTTGTTACTAAATCTATTTACTAAGTAGATTCTGAGATT
CAAGCACTGCTTAGAAGGATAATGCCAAAGATGCATATGCAAGGCATTTGACAGCAAGATATAGCATTTCTGCTTTATC
AGGAAATCCAAAAGAAAAAACAAAAAGACTGGAGGTTTCTATTTGTAACAGCCAATCACACAAGGTGTACTGTACTCAA
TCACAAAAGATCTAACATTTCTGCTTCATGGGGAAAAACATACAACGATACAAG

- intron sequence stopped -

GGAGTGAATGGTTAAATCAACATTTGTGGAACTTGCAAAATTTGTTACTAAATCTATTTACTAAGTAGATTCTGAGATT
CAAGCACTGCTTAGAAGGATAATGCCAAAGATGCATATGCAAGGCATTTGACAGCAAGATATAGCATTTCTGCTTTATC
AGGAAATCCAAAAGAAAAAACAAAAAGACTGGAGGTTTCTATTTGTAACAGCCAATCACACAAGGTGTACTGTACTCAA
TCACAAAAGATCTAACATTTCTGCTTCATGGGGAAAAACATACAACGATACAAGATTTTCCCTAGGTATCATGTTTTGA
AATTACCATTCTCATGCTGTTACTTGTTAGTACTACAATTAGAGTGAAGATGTACCAAGCTAGGAAACATGACTTTGTT
TTCTTGGCACCGGTTACCAGATCAAATATACGGCTCCTGGGTCTCCAGATTTGGCAAAAAGGGTACAAGAACTTGTCAC
TGGATCAGGGTTTGAATGTGCAGTAGATACGAAGCGTGGACTTGACCATGGTTCGTGGGTTCCTCTCATGCTTATGTAC
CCAGAAGCCAGTATCCCTGTTTGTCAGCTCTCTGTTCAACCGCACTTGGATGGTAAACACCATTATGACTTGGGAAGAG
CATTGGCCCCGCTCAAG

>*Iresine sp.* DODA

MSIDESIPARRFLEEWKDKVYSKRPSSILVITAHWLTVVPTVSAIDHSDLIYDFGGFPACMYQIKYTAPGSPDLAKRVQ
ELVTGSGFECAVDTKRGLDHGSWVPLMLMYPEASIPVCQLSVQPHLDGKHHYDLGRALAPLK

>*Mammilaria sp. DODA*

ATGGCGATCGACAAGTCGGTGCAAGCGAGGCCCTTTCTGGAGGGCTGGAGAGAGAAAGTGTTGTCTAAGAAGCCAAAGT
CGATTCTAATGATTTCAGCTCACTGGGAGACAGATGTGCCCACTGTTAATGCTGTTCATCATTCTGATCTCGTCTACGA
TTTCTATGGCTTCCCTGCTCCCATGTACCAGGTGAGTCTCTCTCTCTGCTTTTCAAGTTTGCGGTGTTTTTGTTTTACA
GAAACACATATGCCTTGTTTGTACATAACAGTTAGCAGTTAACAGTTAGCAGTTTGTATAAACATGTTTGATTAACAGT
TAGCAATTGGTATCAACAGTTAGCAGTTTCAAGTGTACATATGACATTTTAGGGCCAAAATACTCAATACCAAACATCA
CAATTAACATTTTGCTTGGTCAAACATGTTCTCCCAAGATGTTAAATACTCTTCCCAAATAATGTTTACCAAACAAGGC
TAATGAAAAATAGTAAGATAAAAAGTCATCCATAAATTTTACATTTTTTTTTTTAAAACCATCCAATAAAAAGATGCAT
TTTTTTAGTCACCTGAAAACAGTGATATTTTTTTTTAGATTTGGAGAACTGGGTTCCGAATCCTT

- intron sequence stopped -

AAATCTGTGCTCTCATAAGAGTAGTATTAGAAGCTAAACTGTGAGTGTACTTATAAGTTTCCTTCTAAAAAAGGTACTG
CAATGATGTATGCAGCTTTACACCCTGTTATCAAAAAGTTGTTTCCGTTAGACCCTTGATCCAAATCACAACAACCGTT
TGCGATTTCACAATAACTAAGAAGTGCAGATGTTTAATGAGTCATTTGATTTAGTCCATTATAATCCAAAGCCAAAGAA
TAGTGTTTCGTTTGGCTCCATTAAAAATGGAAATCATTCCTGGAACTACTGAAATATACAATTTACTCTGTTTACTGAG
TGGATTGTTTGTTTAAACACCGGCTTCAAGGATACCAAAGAATGTAGTATTAACATTGTTGTGTTGGGGTTACCACTAC
CAGCTCAAATACCCAGCACCTGGGGCTCCACATTTGGCAAGAAGGATACAAGAAGTGTTAACCGCATCAGGGTTAAAAT
GCGCGGTAGATAAGAAGCGTGGGCTTGATCATGGTTCGTGGGTTCCTCTGAGGCTCATGTACCCAGAGGCCAGTATCCC
TGTTTGTCAGCTCTCAGTTCAGTCCAACTTGGATGCTACACACCACTATAATTTGGGAAGGGCATTGGCCCCTCTTAAG

>*Mammilaria sp.* DODA

MAIDKSVQARPFLEGWREKVLSKKPKSILMISAHWETDVPTVNAVHHSDLVYDFYGFPAPMYQLKYPAPGAPHLARRIQ
EVLTASGLKCAVDKKRGLDHGSWVPLRLMYPEASIPVCQLSVQSNLDATHHYNLGRALAPLK

>*Phytolacca Americana DODA*

ATGGCGATCGACAAGTCGGTGCAAGCGAGGCCCTTTCTGGAGGGCTGGAGAGAGAAAGTGTTGTCTAAGAAGCCAAAGT
CGATTCTAATGATTTCAGCTCACTGGGAGACAGATGTGCCCACTGTTAATGCTGTTCATCATTCTGATCTCGTCTACGA
TTTCTATGGCTTCCCTGCTCCCATGTACCAGGTGAGTCTCTCTCTCTGCTTTTCAAGTTTGCGGTGTTTTTGTTTTACA
GAAACACATATGCCTTGTTTGTACATAACAGTTAGCAGTTAACAGTTAGCAGTTTGTATAAACATGTTTGATTAACAGT
TAGCAATTGGTATCAACAGTTAGCAGTTTCAAGTGTACATATGACATTTTAGGGCCAAAATACTCAATACCAAACATCA
CAATTAACATTTTGCTTGGTCAAACATGTTCTCCCAAGATGTTAAATACTCTTCCCAAATAATGTTTACCAAACAAGGC
TAATGAAAAATAGTAAGATAAAAAGTCATCCATAAATTTTACATTTTTTTTTTTAAAACCATCCAATAAAAAGATGCAT
TTTTTTAGTCACCTGAAAACAGTGATATTTTTTTTTAGATTTGGAGAACTGGGTTCCGAATCCTT-intron
stopped
AAATCTGTGCTCTCATAAGAGTAGTATTAGAAGCTAAACTGTGAGTGTACTTATAAGTTTCCTTCTAAAAAAGGTACTG
CAATGATGTATGCAGCTTTACACCCTGTTATCAAAAAGTTGTTTCCGTTAGACCCTTGATCCAAATCACAACAACCGTT
TGCGATTTCACAATAACTAAGAAGTGCAGATGTTTAATGAGTCATTTGATTTAGTCCATTATAATCCAAAGCCAAAGAA
TAGTGTTTCGTTTGGCTCCATTAAAAATGGAAATCATTCCTGGAACTACTGAAATATACAATTTACTCTGTTTACTGAG
TGGATTGTTTGTTTAAACACCGGCTTCAAGGATACCAAAGAATGTAGTATTAACATTGTTGTGTTGGGGTTACCACTAC
CAGCTCAAATACCCAGCACCTGGGGCTCCACATTTGGCAAGAAGGATACAAGAAGTGTTAACCGCATCAGGGTTAAAAT
GCGCGGTAGATAAGAAGCGTGGGCTTGATCATGGTTCGTGGGTTCCTCTGAGGCTCATGTACCCAGAGGCCAGTATCCC
TGTTTGTCAGCTCTCAGTTCAGTCCAACTTGGATGCTACACACCACTATAATTTGGGAAGGGCATTGGCCCCTCTTAAG

>*Phytolacca americana* DODA

MAIDKSVQARPFLEGWREKVLSKKPKSILMISAHWETDVPTVNAVHHSDLVYDFYGFPAPMYQLKYPAPGAPHLARRIQ
EVLTASGLKCAVDKKRGLDHGSWVPLRLMYPEASIPVCQLSVQSNLDATHHYNLGRALAPLK

>*Physcomitrella patens DIOXA* (AJ583016)

ATGGCAACTTCAGCTGGTTTGAGTACTTTTTATGTATCGCATGGGTCGCCTATGATGCCACTTGAGGATACTCCCATAC
GAGAATTCTTCTCCACTTGGACGGAGCGCTATCCCACAAGGTACTGCGTCAATCTTCTTTCTGATTCACATCGATTGAA
GCAAATAAACTCAGGTCAAGGACGCGAAATCACATCAATTCAATACAGCCGCGATAAGTTTTGAAAATTGCTCATCACA
GAAGCAACCTACTTCTATCAAAATAATATTGGATCCTATTGCATGTCCGCGGCTTTGATCAGGCCAAAGGCCATTCTCG
CAATTTCAGCTCACTGGGATACTCGAGAGCCTGCCGTTAATGCTGTCAGCCAGAACAGCACTATTCACGACTTTTATGG
CTTCCCTCGCGAGCTATATCAAGTAATTCCTGCACTAACACACCACTACTTTCACCAGATTCTGCGCTACACAAGTTTT
TGTTCACTGATTGGTACCTTGATTCATTTGTTTTCCTCCTTAGTATCTGAATCTTTTTTTCATTCTGGTCACGATTACG
TGACATATTTCATAGATTCTCTGACAAATTCAAGGTCTACTTATATAACACCAGTCTTTTCCTACGTCGTTTGTGCTCA
TGAGATAATTGTTTTCTCCTGTGTCCTCGTCAGTTGCAATACACGCCTCCAGGGGCTCCAGACGTTGCAAAGAGGGTGA
CATCTCTCCTCAAAGACGCTGGCTTCAAGACCGTCCTCGAGGACAACAAGAGAGGGCTCGACCACGGAGCATGGACGCC
CCTGATGCTAATGTACCCCAACGCTGACATTCCTGTCCTCCAGGTCTCAATCCAAAGCAACAAAGATGGGCTTCACCAT
TACCAGCTCGGCCGGGCTCTTGCACCGCTCAAGGACGAAGGAGTGTTAATTTTTGCCTCTGGAACTACAGTCCACAATT
TGAGAGAGATAGATTTTTCTGCCAAGAAGCCAACTGTGTGGGCCAAGGCATTCGACGGATGGTTGACTGATGTGCTACT
CAACAGCAAGTAAGAATTCTTCCATTCCTCCGTTTCTTATCCATTTGAAGAGCCTGTCCACAAAGATCAGAGTTCATAA
ATTCCCATGGATCATTTTACTGCATAATCTCTGTGGTTTTGATTTGTGACGGTTTTTAGGGTTGTTTGAGCTCTTTTCT
TTCTTTCTTTCTTTTTGTCTTTGTGAAGCTGACGTAGTGTTAAACGACATCGATGCAGGCACAAGGAAGCAATGGAATG
GGAGAAAGCACCATATGCTTCCAAAGCTCACCCTCATCCAGATCATTTCCTGCCTGTGTTGGTTGGGTTGGGAGCTGCG
GGAGAGCAATGCCAAGCTGAGAAGATATATGAGGAATTCGCATATGGTTTGGCGCTTTCATGTTTTGCTTTCCATCCTC
AAAACTAG

>*Physcomitrella patens* DIOXA (CAE47099)

MATSAGLSTFYVSHGSPMMPLEDTPIREFFSTWTERYPTRPKAILAISAHWDTREPAVNAVSQNSTIHDFYGFPRELYQ
LQYTPPGAPDVAKRVTSLLKDAGFKTVLEDNKRGLDHGAWTPLMLMYPNADIPVLQVSIQSNKDGLHHYQLGRALAPLK
DEGVLIFASGTTVHNLREIDFSAKKPTVWAKAFDGWLTDVLLNSKHKEAMEWEKAPYASKAHPHPDHFLPVLVGLGAAG
EQCQAEKIYEEFAYGLALSCFAFHPQN

ANNEX 2

DODA homologous Genes Table with Genbank Accession Number

DODA HOMOLOGUES	GENBANK ACCESSION NUMBER
Arabidopsis thaliana	NM_117597
Beta vulgaris DODA	AJ583017
Bacillus anthracis str. Ames	AAP25762.1
Burkholderia fungorum BCEP_397	AAAJ02000075
Escherichia coli str. K12	E65091
Glycine max 1	BE608113, BG651843, BE191341, AW733761
Glycine max 2	AW348985, BM523589
Hordeum vulgare 1	BI956621
Hordeum vulgare 2	BE558845, BI780256
Iresine sp.	-
Lotus corniculatus var. *japonicus*	AP004916
Lycopersicon esculentum 1	AI896661, BG128676
Lycopersicon esculentum 2	AI898647, AW222441, BF050515
Mammilaria sp.	-
Medicago truncatula	BF645123
Mesembryanthemum crystallinum 1	BF480453
Mesembryanthemum crystallinum 2	BE131205
Methanosarcina acetivorans str. C2A	NC_003552
Oryza sativa	AP003227
Physcomitrella patens DIOXA	AJ583016
Phytolacca americana	-

Populus balsamifera subsp. trichocarpa	BU879976, BU875430
Populus tremula x Populus tremuloides 1	BI127978, BI122325
Populus tremula x Populus tremuloides 2	BU812037, BU887517
Portulaca grandiflora DODA	AJ580598
Pseudomonas aeruginosa	NP_251529.1
Pseudomonas fluorescens PFO-1	ZP_00083959
Pseudomonas putida str. KT2440	NP_744024
Pseudomonas syringae pv. tomato str. DC3000	NC_004578
Salmonella enterica strain CT18	AE0890
Shigella flexneri 2a str. 2457T	AAP18370.1
Solanum tuberosum 1	BG098818, BG886949
Solanum tuberosum 2	BG590982
Sphingomonas paucimobilis str. SYK6 LigAB	AAA17728
Thermobifida fusca	ZP_00059271
Vitis vinifera	CF214153
Xanthomonas axonopodis pv. citri str. 306	AAM38549.1
Xanthomonas campestris str. ATCC 33913	NP_639012.1
Yersinia pestis CO92	CAC89513.1
Zea mays	AY106860.1

ANNEX 3

DODA homologous proteins predicted by translation:

Sequences obtained by cloning (bold character) or by conceptual translation of several ESTs or genomic DNA.

Caryophyllales plants

>***Beta vulgaris* DODA**

MGSEDNIKETFFISHGTPMMAIDDSKPSKKFLESWREKIFSKKPKAILVISAHWETDQPSVNVVDINDTIYDFRGFPAR
LYQFKYSAPGSPELANRIQDLLAGSGFKSVNTDKKRGLDHGAWVPLMLMYPEADIPVCQLSVQSHLDGTHHYKLGQALA
PLKDEGVLIIGSGSATHPSNGTPPCSDGVAPWAAAFDSWLETALTNGSYEEVNKYETKAPNWKLAHPWPEHFYPLHVAM
GAAGENSKAELIHNSWDGGIMSYGSYKFTST*

>***Iresine sp*.**

MSIDESIPARRFLEEWKDKVYSKRPSSILVITAHWLTVVPTVSAIDHSDLIYDFGGFPACMYQIKYTAPGSPDLAKRVQ
ELVTGSGFECAVDTKRGLDHGSWVPLMLMYPEASIPVCQLSVQPHLDGKHHYDLGRALAPLK

>***Mammilaria sp*.**

KMSIDDSIPARKFFQEWKEKVYSKRPKAILVISAHWETNVPAVNAVNHSDLIYDFRGFPAIMYQLKYPVPGAPNLARRV
EELLTASGFSCVVDKKRGLDHGSWVPLMLMYPEADIPVCQLSVQSHLDGTHHYNLGKALAPLK

>*Mesembryanthemum crystallinum* 1

GFDHGVWFPLQFMYPEADIPVCQLSVQPSMDGAHHFNMGKALAPLMDEGILIIGSGGAVHPSDDTPHCPNAVAPWAAEF
DDWLCDAVIKGRYEDVNNYNKLAPNWEIAHPGPEHLYPLHVALGAAGEKSIAETIHHSWARNGVFGYASFKFTSTSSTL

>*Mesembryanthemum crystallinum* 2

MGGVEKITETFFISHGTPRMSIDETIPARHFLEEWQEKVYSKRPRSILVITAHWETVVPTVNAINHSDLIYDFGGFPAR
MYQLKYLAPGAPDLAKRLQELLAASGFECAVDRKRGLDHGSWVPLMLMYPEANIPVCQLSVQPRLDGIHHYNLGKALAP
LKAEGVLIIGSGSAVHPANNTPGCLDGVAPWAAEFDYWLEEALTAGRYEDVNNYTTKAPNWKTAHPWAGHFYPLHV

>***Phytolacca americana***

MAIDKSVQARPFLEGWREKVLSKKPKSILMISAHWETDVPTVNAVHHSDLVYDFYGFPAPMYQLKYPAPGAPHLARRIQ
EVLTASGLKCAVDKKRGLDHGSWVPLRLMYPEASIPVCQLSVQSNLDATHHYNLGRALAPLK

>***Portulaca grandiflora* DODA**

MGVGKEVSFKESFFLSHGNPAMLADESFIARNFLLGWKKNVFPVKPKSILVVSAHWETDVPCVSAGQYPNVIYDFTEVP
ASMFQMKYPAPGCPKLAKRVQELLIAGGFKSAKLDEERGFDHSSWVPLSMMCPEADIPVCQLSVQPGLDATHHFNVGRA
LAPLKGEGVLFIGSGGAVHPSDDTPHWFDGVAPWAAEFDQWLEDALLEGRYEDVNNYQTKAPEGWKLAHPIPEHFLPLH
VAMGAGGEKSKAELIYRTWDHGTLGYASYKFTSI*

Non-Caryophyllales plants

>***Arabidopsis thaliana***

MEKVNQTFFLSHGSPTLSIDDSLEARQFFKSWTQKVLPQKPKSILVISAHWDTKFPSVNTVLRNNTIHDFSGFPDPMYK
LKYEAPGAIELGKRVKELLMKEGGMKRVDEDTKRGLDHGAWVPLMLMYPEADIPICQLSVQSNQNGSYHYNMGKALASL
KDEGVLIIGSGSATHNLRKLDFNITDGSPVPWALEFDHWLRDSLLQGRYGDVNEWEEKAPNAKMAHPWPEHLYPLHVVM
GAAGGDAKAEQIHTSWQLGTLSYSSYSFTSSL*

>*Glycine max* 1

MALKDTFYISHGSPTLSIDESIQARKFLQSWKKDVFPQRPSSILVISGHWETAVPTVNVVDSINDTIYDFYGFPKQMYQ
LKYPAPGAPQLARRVKELLKKSGFSHVDEDTKRGLDHGAWVPLFLMYPEADIPVCQISIQSQQDGTYHYNLGKALAPLK
DEGVLIMGSGSAVHNLRALEPHSTVAPWALEFDNWLKDALLEGRYDDVNHYEQKAPHAKKAHPWPDHFYPLHVAIGAAG
EEAKAKLIHSSIELGSLSYASYQFTSAAS*

>*Glycine max* 2

MGLKLKETFYLSHGAPSLVIDDSIPAWHFFNSWKEQFPTKPSSILVISAHWDTHVPTVNVVDQNDTIYDFSGFPKSMYK
LKYPAPGAPQLAKRVKELLLGSGFSHVDEDKKRGLDHGAWVPLFLMYPEADIPVCQLSISSNKGGTYHYNMGKALAPLK
DEGVLIIGSGSATHNLRAIAPRGTPPAPWASAFMSWLKTALLDGRYEEVNEYEEKAPYAKMAHPWPDHFFPLHVAMGAA
GENSKAKVVHDSWDGGSMSYASFGFTTADS*

>*Hordeum vulgare* 1

MALMETFYLSHGAPTLAIDETVPARKFFQSWQQSAYKEKPSSILVISAHWETAQPTVNVVDRNDTIYDFYGFPKPLYQI
KYTPPGAPELAKRVKELLLASGIEHVDEDKKRGLDHGAWVPLMFMYPEADIPVCQLSVQSDRDGTHHYNMGKALAPLRE
EGVLILGSGSAVHNLGSRLPDGSPVPSWALEFDNWLK

>*Hordeum vulgare* 2

MDTFFLSHGSPTLSIDETIPARSFFESWLPAAVAGPERPRAILIVSAHWETATPAVNVIRGSNDTIHDFGGFPKSMYQL
KYPAPGAPDLAKRTKELLEQGGFGPVKEDRSRGLDHGAWVPLMLMYPDADVPVCQLSVQTDRDATYHYNLGKALAPLRE
EAVLVLGSGSATHNLRKMGPTSSPPPQWASDFDTWLKDSLLGGRYDDVNRYEDKAPNAKMAHPRPEHFYPLHVALGAAG
DEAKAELVHHSWTNASLSYASYRFTT*

>*Lotus corniculatus*

MALKDTFYISHGSPTLSIDESLVARKFLQSWKKEVFPPRPTSILVISGHWDTAVPTVNVVDSTNDTIYDFYGFPKPMYQ
LKYPAPGAPHLAKRVKELLKEGGFSRVDEDKKRGLDHGAWVPLLLMYPEADIPVCQLSVQSNLDGTHHYNIGKALAPLK
DEGVLIVGSGSAVHNLRALERHATVAAPWAVEFDNWLKEALLEGRYEDVNHYEQKAPHAKKAHPWPDHFYPLHVAIGAA
GENSKAKLIHSSIDLGSLSYASYQFTSDVI*

>*Lycopersicon esculentum* 1

MGSQMAVPVNETFFISHGSPMLSIDDSLPARHFLKCFNQKVFTQKPNSILVISGHWETSEPTVNCITGLNDTIYDFYGF
PEQMYQLKYQAPGAPKLAKRVKELLKSSGFNRVHEDNNRGLDHGAWVPLMLMYPGADIPVCQLSVQTKKDGTHHFNIGK
ALAPLKEEGVLIVGSGSATHNLRALRDTAGVASWAMDFDNWLKESLVNGRYEDVNNYMTKAPCAKIAHPWPDHLYPLHV
AMGAAGENAKAELIHHSWSNHALSYASYKFESQLK*

>*Lycopersicon esculentum* 2

MSCPAISPVKETFFISHGSPTLSIDESLPARNFLKSFKQKFLMNQKPNSILVISAHWETSEPTVNSIRGRNDTIHDFYG
FPKSMYQLKYPAPGSPELAKRVKDVLMASGFPIVHEDKKRGLDHGAWVPLMLMYPEADIPVCQLSVQPNRDGTYHYNLG
KALASLKDEGVLIIGSGSATHNLRALGPSKNVSSWALEFDNWLKDALLSGRHQDVNNYDMKAPHAKVAHPWPEHIYPLH
VALGAAGEGVNGELIHHSWDLGALSYASYRFPSLNRSS*

>Medicago truncatula

PRAESLRNENTREIVNSISKNIQKMALKETFYISHGSPTLAIDETIPAWKFLTSWKEVFPERPSAILVISGHWDTSVPT
VNVVNHNETIHDFGGFPRSMYKLKYPAPGAPKLAKRVKELIEASGLSRVDEDKKRGLDHGTWVPLMLMYPEADIPVCQL
SVSSNRNGTYHYNLGKAIAPLKDEGVLIIGSGSATHNMRAIGPRESPPPPWALAFDSWLKESLVEGRYEDI

>*Oryza sativa*

MDTFFLSHGAPTLSIDDTIAAQGFFKSWLPAAVAGAELPRAILVVSGHWEAAAPTVNVIRGNNDTIHDFYGFPKAMYKL
KYPAPGAPDLAMKTKELLEQAGFGPVKENHSRGLDHGAWVPLMFMYPEANVPVCQLSLQSGRDGAYHYELGRALAPLRD
DGVLVLGSGSATHNLRRMGPEGTPVPQWAAEFDGWLQEALLGGRHDDVKRYEEKAPHGRVAHPSPDHFLPLHVALGAAG
EGAKAELIHRSWSNASLSYASYRFTTAKN*

>***Physcomitrella patens DIOXA***

MATSAGLSTFYVSHGSPMMPLEDTPIREFFSTWTERYPTRPKAILAISAHWDTREPAVNAVSQNSTIHDFYGFPRELYQ
LQYTPPGAPDVAKRVTSLLKDAGFKTVLEDNKRGLDHGAWTPLMLMYPNADIPVLQVSIQSNKDGLHHYQLGRALAPLK
DEGVLIFASGTTVHNLREIDFSAKKPTVWAKAFDGWLTDVLLNSKHKEAMEWEKAPYASKAHPHPDHFLPVLVGLGAAG
EQCQAEKIYEEFAYGLALSCFAFHPQN*

>*Populus balsamifera*

MALMETFYLSHGAPTLAIDETVPARKFFQSWQQSAYKEKPSSILVISAHWETAQPTVNVVDRNDTIYDFYGFPKPLYQI
KYTPPGAPELAKRVKELLLASGIEHVDEDKKRGLDHGAWVPLMFMYPEADIPVCQLSVQSDRDGTHHYNMGKALAPLRE
EGVLILGSGSAVHNLGSRLPDGSPVPSWALEFDNWLK

>*Populus tremula* x *Populus tremuloides* 1

MALMETFYLSHGAPTLAIDESIPARQFLKSWQQTVLRERPKAILVISGHWDTKEPTVNVVNINDTIYDFYGFPKPMYQL
KYTPPGAPQLAKRVKELLMENGFKHVHEDKKRGVDHGTWVPLMFMYPEADIPVCQLSVQSDRDGTYHYNMGKALAPLKE
EGILVMGSGATVHNLGSRLPDGSPVPPWAVEFDNWLKDALIEGRYEDVNHYESKAPHPKTAHPRPDHFYPLHVAMGAAG
EDSKAKLVHSSWTDGTLSYASYQFTAPK*

>*Populus tremula* x *Populus* tremuloides 2

MALMETFYLSHGAPTLAIDETVPARKFFQSWKQGVYKEKPSSILVISAHWETAQPTVNVVDRNDTICDFYGFPKPLYQI
KYTPPGAPELAKRVKELLLASGIEHVDEDKKRGLDHGAWVPLMFMYPEADIPVCQLSVQSDRDGTHHYNMGKALAPLRE
EGVLILGSGSAVHNLGSRLPDGSPVPSWAVEFDNWLKDALIDGRYEDVNHYESKAP

>*Solanum tuberosum* 1

MGSQIAVPVKETFFISHGSPMLSIDDSLPARHFLKCFKQKVFTQKPNSILVISGHWETSEPTVNCITGLNDTIYDFYGF
PEQMYQLKYQAPGAPKLAKRVKELLKSSGFNRVHEDNNRGLDHGAWVPLMLMYPESDIPVCQLSVQTKKDGTHHFNIGK
ALAPLKEEGVLIVGSGSATHNLRALRDTAGVASWAMDFDNWLKESLVNGRYEDVNNYMTKAPCAKIAHPWPDHLYPLHV
AMGAAGENAKAELIHHSWSSHALSYASYKFKSQPK*

>*Solanum tuberosum* 2

MSCPAISPVKETFFISHGSPTLSIDESLPARTFLKSFKERFLINQKPNSILVISAHWETSEPTVNSIRGRQDTIHDFYG
FPKSMYQLKYPAPGSPELAKRVKDVLMASGFPIVHEDKKRGLDHGAWVPLMLMYPEADIPVCQLSVQPNRDGTYHYNLG
KALASLKDEGVLIIGSGSATHNLRALGPSKNVSSWALEFDNWLKDALLSGRHQDVNNYDMKAPHAKVAHPWPEHIYPLH
VALGAAGEGVNGELIHHSWDLGALSYASYRFTTSS*

>*Triticum aestivum*

MDTFFLSHGSPTLSIDEAIPARSFFQSWLPAAVAGPERPRSILIVSAHWETATPAVNVIRGANDTIHDFYGFSKSMYQL
KYPAPGAPDLARRTKELLEQGGFGPVKEDRSRGLDHGAWVPLMLMYPDADIPVCQLSVQTDRDATYHYNLGRALAPLRE
EGVLVLGSGSATHNLRKMGPSGSPPPQWASDFDTWLKDSLLGGRYDDVNRYEEKAPNAKMAHPRPEHLYPLHVALGAAG
DESKAELIHSSWTNASLSYASYRFTTKN*

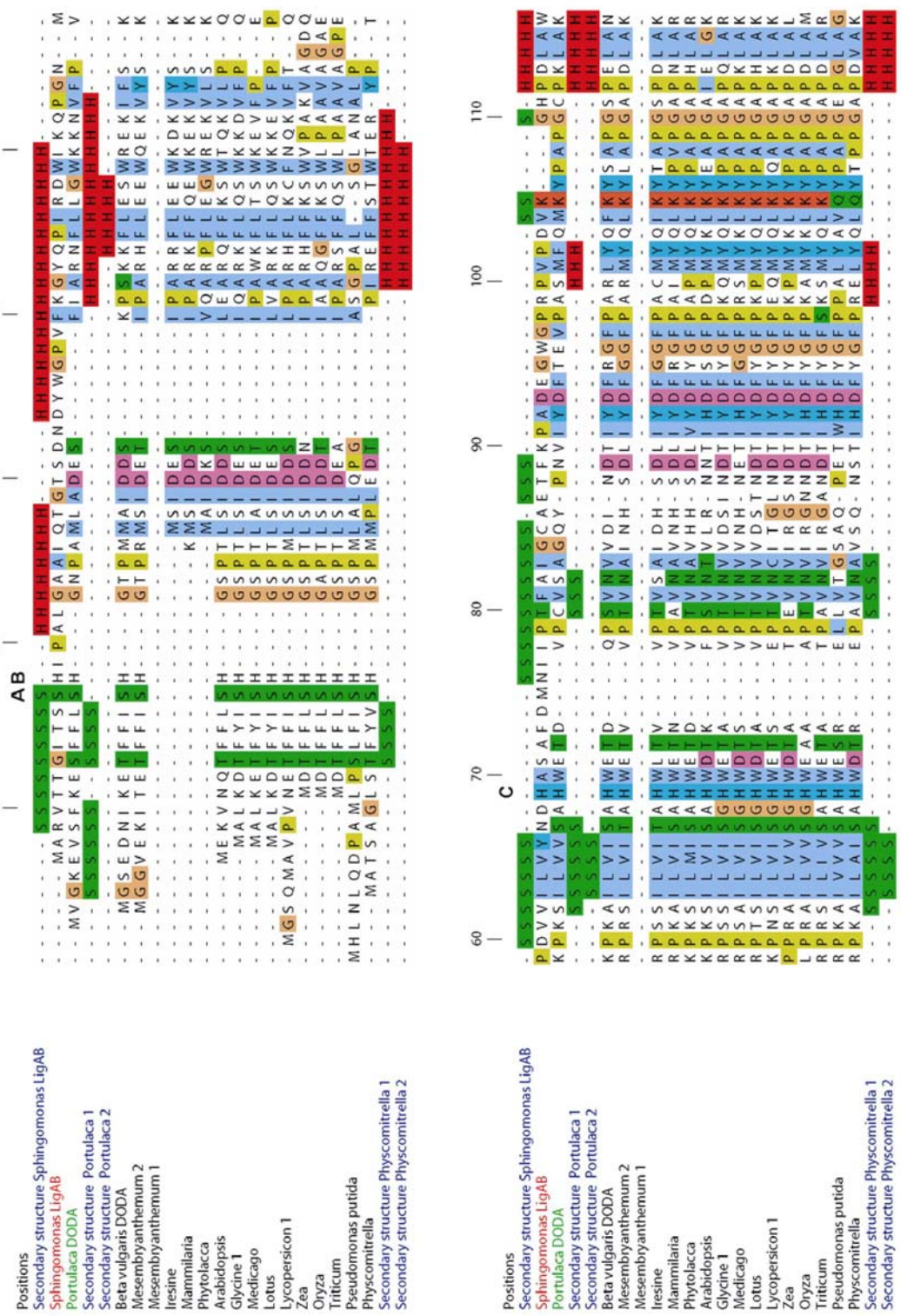
>*Vitis vinifera*

MDTFYISHGSPTLSIDESLPARHFLKSWREKVMGQPPTSILVISGHWETHEPTVNVVPRNDTIYDFYGFPKPMYEFKYP
APGAPELAKRVKELLMASDFKCVNEDKERGLDHGAWVPLMLMYPEADIPVCQLSVQTNRDGSYHYNMGKALAPLREEGV
LIFGSGSATHNLRTINRNSDSKAAVVPWAYEFDTWLKEALLDGRYEDVNCYEEKAPHAREAHPWPDHFYPLHVAMGAAG
KDAKAKLIHHSWSFGTLSYASYQFTTAAS*

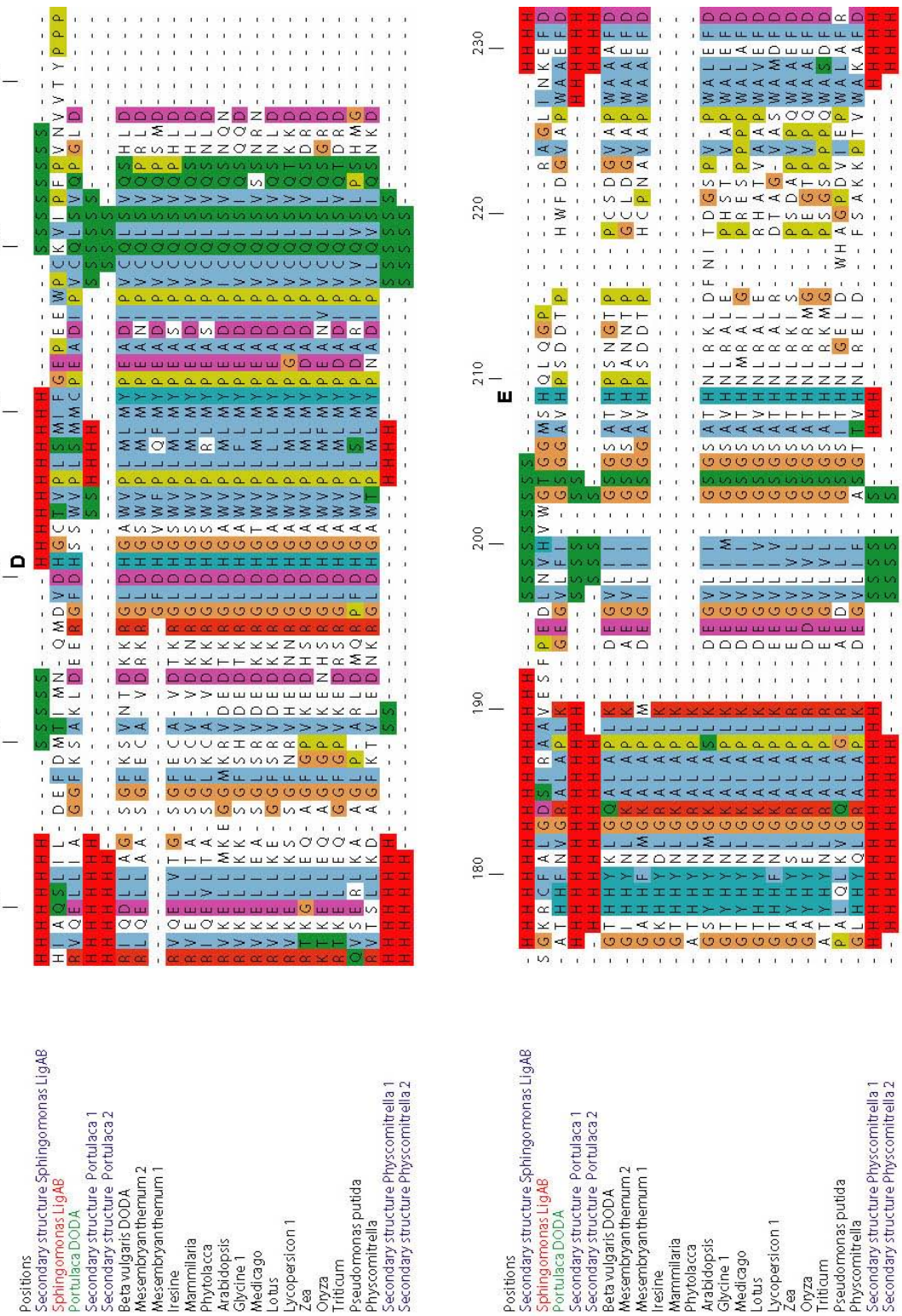
>Zea mays

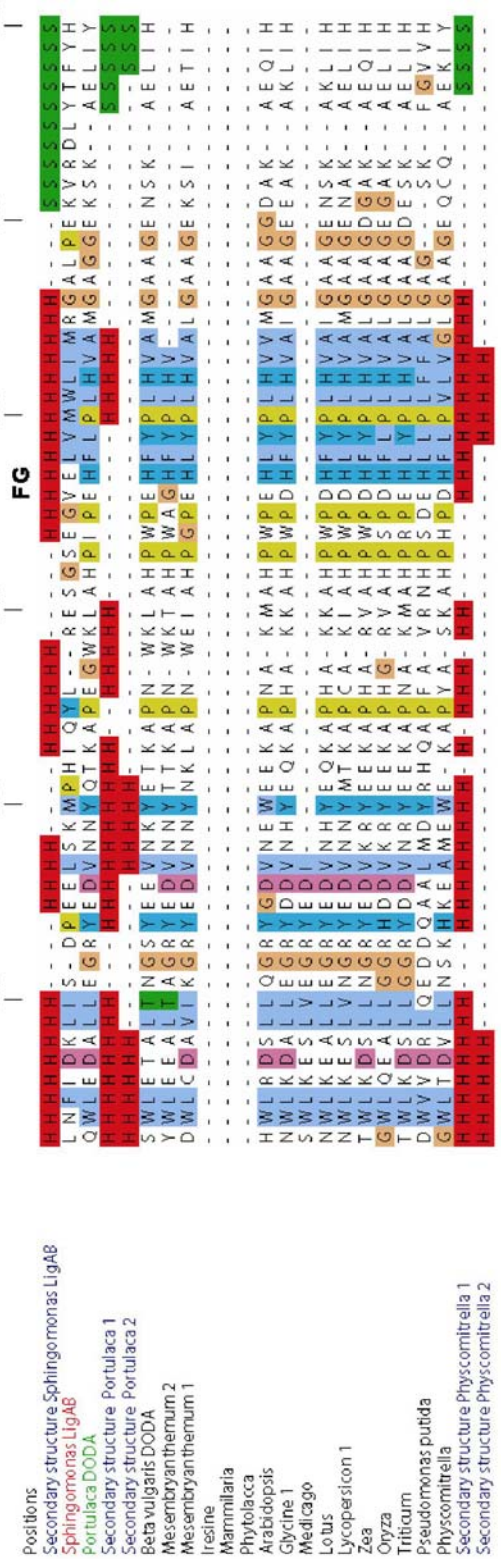
MDTFFLSHGSPTLSIDDNIPARHFFKSWVPAKVAGDQPPRAILVVSGHWDTATPEVNVIRGSNDTIYDFYGFPKPMYKL
KYPAPGAPDLALRTKGLLEQAGFGPVKEDHSRGLDHGAWVPLMLMYPDADIPVCQLSVQSDRDGAYHYSLGRALAPLRE
EGVLVLGSGSATHNLRKISPSDAPVPQWAAEFDTWLKDSLLNGRYEDVKRYEEKAPHARVAHPWPDHFYPLHVALGAAG
DGAKAEQIHQCWSNGTLSYASYRFTTNT*

ANNEX 4



Multiple alignment of *Pg* DODA and its homologues with *Sphingomonas* PCA 4,5-dioxygenase LigAB shows the conservation of the LigB domain (pfam02900). Based on the LigAB crystal structure 1B4U, we indicated on the first line in bold characters the position of the amino acids responsible for the orientation of the aromatic substrate (A, D, E) and those responsible for the orientation of the iron cofactor (B, F, G). Highlighted in blue characters, we show a comparison of *Sphingomonas* LigAB secondary structure with those predicted for *Portulaca* DODA and for its *Physcomitrella patens* homologue realized with PROF tool (1) and a selection with an accuracy >82% in PROF predictions in (2). Alpha-helixes are highlighted in red and beta-sheets in green. For Genbank accession numbers see Annex 2.


D
E
Positions
Secondary structure Sphingomonas LigAB
Sphingomonas LigAB
Portulaca DODA
Secondary structure Portulaca 1
Secondary structure Portulaca 2
Beta vulgaris DODA
Mesembryanthemum 2
Mesembryanthemum 1
Iresine
Mammillaria
Phytolacca
Arabidopsis
Glycine 1
Medicago
Lotus
Lycopersicon 1
Zea
Oryza
Triticum
Pseudomonas putida
Physcomitrella
Secondary structure Physcomitrella 1
Secondary structure Physcomitrella 2


Positions
Secondary structure Sphingomonas LigAB
Sphingomonas LigAB
Portulaca DODA
Secondary structure Portulaca 1
Secondary structure Portulaca 2
Beta vulgaris DODA
Mesembryanthemum 2
Mesembryanthemum 1
Iresine
Mammillaria
Phytolacca
Arabidopsis
Glycine 1
Medicago
Lotus
Lycopersicon 1
Zea
Oryza
Triticum
Pseudomonas putida
Physcomitrella
Secondary structure Physcomitrella 1
Secondary structure Physcomitrella 2
240
250
260
FG
270
280
290


Positions
Secondary structure Sphingomonas LigAB
Sphingomonas LigAB
Portulaca DODA
Secondary structure Portulaca 1
Secondary structure Portulaca 2
Beta vulgaris DODA
Mesembryanthemum 2
Mesembryanthemum 1
Iresine
Mammillaria
Phytolacca
Arabidopsis
Glycine 1
Medicago
Lotus
Lycopersicon 1
Zea
Oryza
Triticum
Pseudomonas putida
Physcomitrella
Secondary structure Physcomitrella 1
Secondary structure Physcomitrella 2
300
310
320

ANNEX 5

Partial cDNA from *Portulaca grandiflora*

>A.3 (=V.33)
ACAGCGTTGCATTCAAGTCAGCCGGTCATATGTTTCATGCAAATTTCACAGCCAGGCCAAGAAAGGATCCTTCCGCTGA
GCCAGAGCTCTTCTTTGCTGAGTTGAAAGGCGATATCTATAGGAAGGGGTATCATACTGTCCCTGCGGTCTGTTCACTG
GGACCAGCGAGTTCAGTGACTGGTGGTTTAAATGGGTGTAGGGCTTGTGGAAGTGTTTGCAGGTATGATCCGCTTTGTC
ACCCCGAAGGAGGATTCTGTCTTGGTTCTCCTGTTTACCTACTCTCAGATCGTAGTGTGCTCAAAAAATATACTGAATG
AGATTCTGCCTCGGCACAGGAAGACTGAAAAGAACTAGTTATGATATCATTATCAGAGGTGCATGTAAATAAGTGATGC
TAATGACATCTACGAGCAACTAAGACCCTGTTTTATCCTATTAGAATGTTTGTGTATGAAATTCTATCTATCTCAGAAT
GTGCATGTATGTATGGAGCTCTAACTAATGAAACGTATGCTAGTGATTGAGACAATGCTTTTATCGCGTCAAAAAA

>*A.8*
ACCTGGAAACTACACGGATCCCAATGCTTCAAAGGCATTAGCAAATGTCAGAAAGCTTGTGGATGATGGCAATTATTCT
GAAGCTACTGCTATGGCTGATAAGGACCTTTCTGGCACTTCTGCTAATGTTTACCAAGTCCTGGGTGATCTCAAGCTGG
ATTTTGATGAGCGACATGCTAACTATACAGCAGGAACTTACTACAGAGAGCTAGACCTCGACACTGCGACTGTTACTGT
AAAGTATACAGCGGGAGATGTTGAATTCACCAGAGAACACTTTGCGTCGAATCCTGATCAGGTCATTGTCACCAAGATC
TCTGCCAGTAAACCGGCTTCTGTTTCATTCAACATCACATTAAGCAGCCAACTGCAATACCGTTCCTCCATAAATGGCA
CAAACATGATCATTATGGAAGGAAGTTGTCCTGGGAAGAGAAACCAGGCCGATGATCATCAAGGGATCAAATTCTCAGC
AGTCCTAAACTTAGATATTACTGGAAAAAACGGGGCCATACATAATGTAGAAAGCCAGAAGTTGAGAGTTGAAAATGCA
GATTCTGCAGTGATTCTTGTGGCAGCTGCCTCTTCATTTTCCGGGCCATTCACTAACCCTTCAGAATCCAGCAAGAACC

>*A.13*
ACCACAATCTCATTTGCTTCCAAGTCCCACAGCAGTAATTGCGTGTCCTGACCAACAGAACCAAAACGATATACAACAT
TTTCTCCTGAATCATCAGAATTTGGCGGAGACCAAAATGGGTCAAATGCTACTCCACTAACCCAGGAGTTGTGCCCTTC
TCCCCAAGCAACAACCTTCCGGTCTTCCATGCTCCAAACTTGAACCAAGTCATCTTCACCTCCAGTCAAAATATATTTC
CCATCACCACTCCAAGCACAACATAGAAGTGCCCCATAGTAACTTTTTCCACCGCATATAAGTTGTTCTTTCAAGTAGT
CAAAAACCCGCAAATATCCATCTCTTCCAACAGCAGCCATATGGACACCATCATTCGAAAATGCAATGCTATTGATCGA
TCCTTGGCATATATGCCATCTCGCAATTGGATTACTCTTACTCGTGCGAGCATGGGCTACAGAAAACTGGGGTTGGTCT
TTGCAGCAGGAAAGGACGGATCAGCAGCACCATCTTTGTTCTTCTCATACACATACAAATTTCCATCTGCATGAGCAAC

>*A.15*
ACTGTATTTTATATACCAGAACAAATAAGCTCCGAAAATTAACACTGTTGTAATTAGATGTGATTTGCAAACATAATCA
AGCCCGAATGACTGAATTTTACATAGACAGATATTTGCTTGTAGAGATGATAGACCCGCGAACTTATAAAATTGATTAC
CGGCCTCCGGGATTCCTGTTCCATGACAAATAGGAAGGATCCCCTAGTTGCCTTGGCTTGACAAGATAGATCAAGGTGA
AGAGCATGGCGAGAAAGAAGCAGAAGCCCCCGATGGTGAGATAGGCAATTCCTAGAAAGTCATTTTTGCCACCAATCCA
GGTGGTGGTGGAGAGCACTAACATCTTTTTCCCATTGAAGCTGTATGTGTTGTAGTTGTTCTGCATTCTCACCGTTATG
AGATCGTGTGCTTGCAAATCCTGTTCTATCCTTCCATACAATTTCCTAAAAGTAGGAAGAGCTGCAGTGCGCATCCACA
CAATAAGACCCTCTTGCTCACTCAGAGGAATGGATTCATTCAAATGAGCACCACCG

>*H.13*
ACTGTGGCTGCCTCTGGTAGAAAGCAGTCTGAGATCATATTCATGGGAACAGGCACGAGTGAAGGCATACCTCGGCTGA
GCTGCCTTACTGATCCCGCCAAGAAATGCTCGGTGTGCAGGAAGGCTGTAGAGCCAGGTAACAAGAATAGAAGACTCAA
CACAGGCATCCTAATACGCTTTCCTGGATCGGTTGGGATGCACAATGTTTTAATTGATGCTGGAAAGTTTTTCTACCAT
AGTGCTCTTCGTTGGTTTCCTATTTATGGGCTAAGAACCATTGACGCGGTTATTATTACTCATTCTCATGCTGATGCTA
TTGGAGGCCTGGATGATCTACGGGATTGGACGAACAATGTTCAGCCTTCCATTCCAATATATGTTGCAGAGCGTGATTT
TGAGGTGATGAAGAAGACCCACTACTACCTAGTTGATACTAGTGTGGTTCTACCTGGTGCTGCTGTCTCCTCATTGCAA
TTCAATGTCATTAATGAAGANCCCTTTGTCGTGCACGGTCTAAAGATCACTCCATTACCCGTGTGGCATGGTGCANGCT
ATCGATCTCTTGGCTTTCGGTTTGGCAATGTTTGTTACATAAGTGATGTCAGCGACATCCCAGATGAAACTTACCTCTT
TTAAAGGATTGTGA

>N.22

CTTTTTTTTTTTTCTTTTTTTTCTGAGAAAAACCTCCCTTCGCTCCCTTAAAACTCAAAGGTGCTCTCTCTTCCTGCAA
CTCAAACTACCGCAAAAAGCTCCCACCTTTGTTCATCTGAATCATGCAGCAGGGCACAAGATCAGAGACCCAAGTCNCT
TGGGATGACCAGCAGAACATCNACNAGTTCAGCAAACTCAACAATNGGTTCCATGAGCTTGAGGATGAGATTAAATTCN
CTAAGGAAAAATGTGAGAATCTANAANATGCAGGCNACGANCTAATCCTTACAGATGAAGANATAATNCGGTTTCAAAT
AGGGGAGGTGTTCGCTCATGTTCCTAAGGAGGAAGTNNAAACTCGGCTANAGGAGATGAAAGAGGAGGCAACGAAAAAC
CTGTANAAACTGGAAGAACAGAAGGAATCCNTNGTGACCCNAATGACNCANTTGAAGAAGATTTTGTATGGAAAATTTG
GGGACTCCNTCCACTTGGAAGANGAATNCGAGATGGAGCTTGATCTTCCTCTTAATCTCAGTATCTCNCTATTCTTATT
ATTGTAGTGCAGAGTGCNCCCCNTNCCCTTGTATCTTCCTTTTATTTTTCTTTCTGAATTTTGTTGCCCTTGTTTNCAA
CTTGGGAAAGGTGATCTCATTCTGTTGCTTACTGTTGCNTTTGATCTGANGANAAATGGTGAAATGAANATTNCCCCCC
CCTCNNTAAAAAAAACC

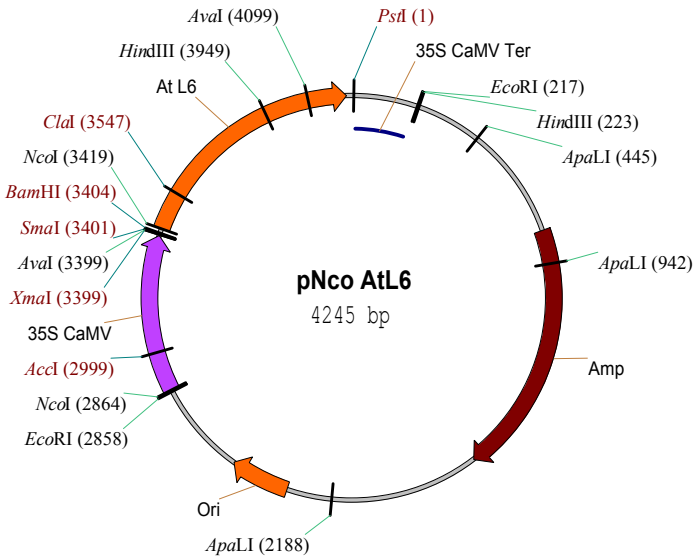
>P.34

ACATTAATACACAATCTGGCAATTAATCATCAATGACAAAAAAAAACTCACAATTCACAAATCCAGCTAGTCAAACCAT
TAATCTACCAAACACCCAGCACCAAAATTCAAATACAAATTCAAAAAATTATTAAAAACTATGAAACCCAGATTTGAAG
TAGTGGCAATGTGGCATCTATGACTGACCCAAACACTCATCAGAACCTGACCGATTCAAACCCAATTCAAATCCAAAAA
TGCAAAAAATAGATCCAAAATGAATTCGAACCCAACCTGACCCAGTTTGATCCAAACTTATCTTTCAAAAGATAAAGTG
ACCCAAATTAACTCGATCCAAACTCAACACAACCTAATCATTGTTCGCTCAAATTGAGTCAACCCAAACCCGACCCAAG
TGACTCAACCGCGATGATTTTTTGAATTTTATTTAGCAAGAAAATGACTAAAGTAAAATGTCTCGTTAAATCTCATCTG
CTTCTTTTCCAATAAAGATGCTTCTTCTTTTCCAATAACGTTTTCAAAATCTTGGGTGAATTAAAACCGGTCTCTCTAA
ATTGAAAAATTTAAGAGTAAGATTGTTACGTTCCACCTCTCCATAACCGCTTGGGCGAGATTTTGGAATATTGGTTTTT
GGTTTTTTTTTTTTATTTAGCAAGAAACGGAAAAGGAAAAACAAAAAAAAAAAAANTACTGCCNGGGCGGGCGCTCGAA
ATCACTAATT

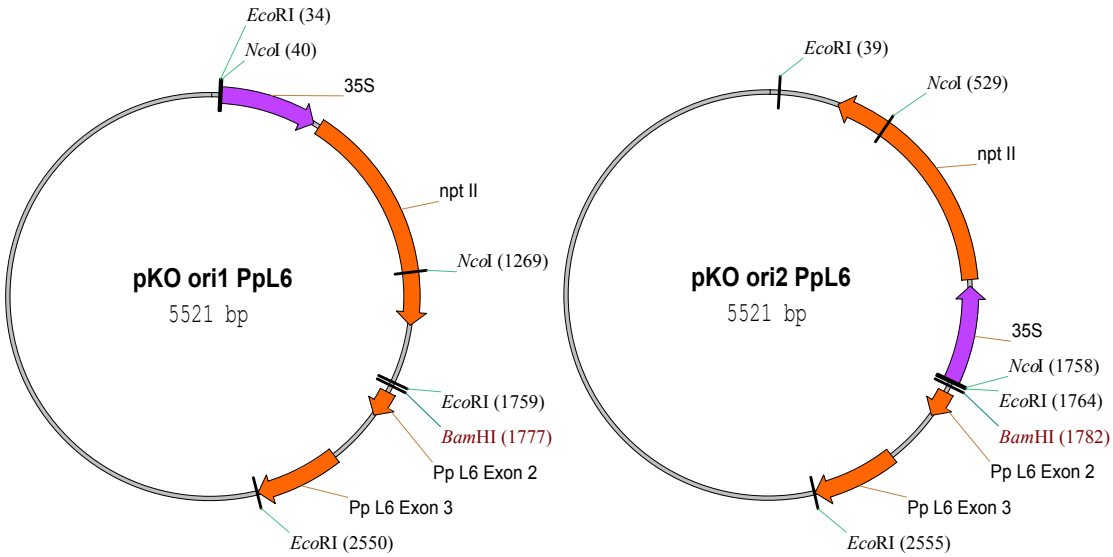
ANNEX 6

Plasmids restriction map

Biolistic transformation vector:



Knock-out transformation vectors:



ANNEX 7

List of PCR amplification primers

***A.16* specific primers:**

*A.16*p1	5’-CACTGAGCGATGTCACTTCTTCTTCAGCA-3’
*A.16*p2	5’-CAGAAATGTAGCCATTCTGAGTGAATCCAA-3’
*A.16*start	5’-TGCTCTAGACCATGGCGTCCTCTGA-3’
*A.16*stop	5’-CTTGACTGATGCTCCGAAAATG-3’

***A.t L6* specific primers:**

*AtL.6*p1rev	5’-TCTTGAAATTCACATGACATTCA-3’
*AtL.6*p2rev	5’-TGTATTCGTATACAAGACAGGAT-3’

***Bv DODA* specific primers:**

*BvDODA*start	5’-ATTCTAGACCACCATGGGTAGTGAAGATAACAT-3’
*BvDODA*stop	5’-ATACCAATGGGCTGAAGAGATGATAC-3’
*BvDODA*p1	5’-TACTTAATATGATACTTTCGTGCCA-3’
*BvDODA*p4	5’-GCTCAAATCTGAAAATGGGTAGTGAAGA-3’

***Pg DODA* specific primers:**

*PgDODA*p2	5’-CTTCCTTCTAGGAGAGCATCCTC-3’
*PgDODA*p3	5’-TTCACATCTTCGTACCTTCCTTCTAGGAGAG-3’
*PgDODA*start	5’-CGTCTAGACCACCATGGGTGTTGGGAAGGAAGTG-3’
*PgDODA*stop	5’-GTCGACCGGCTCTATCTCTCACCATCT-3’

***Pp DIOXA* specific primers:**

*PpDIOXA*p4	5’-ACCAACAACTACCATACTAGTTTTGAGGAT-3’
*PpDIOXA*p5	5’-CTTCCTTCTAGGAGAGCATCCTC-3’

***DODA* Universal primers:**

*L.6*univsense	5’-AGTTTCTTCTTGTCTCATGGGAATCCA-3’
*L.6*univas1	5’-CCGGAGGCGGACATCCCGGTGTGC-3’
*L.6*univas2	5’-GAAGGTGTCCTCTTCATTGGCTCCGG-3’
*L.6*univp4	5’-TGAGCATGATGTGCCCGG-3’
*L.6*univp5	5’-TGATCAATGGCTTGAGGA-3’

Genome Walker primers:

AP1	5’- GTAATACGACTCACTATAGGGC-3’
AP2	5’- ACTATAGGGCACGCGTGGT-3’

REFERENCES

Adachi T (1972) Chemogenetic studies on flower colour in the genus *Portulaca* in relation to breeding. Bull. Lab. plant breed. Miyazaki University, Japan **4:** 1-98

Adams JP, Elbe JH, Amundson CH (1976) Production of a betacyanine concentrate by fermentation of red beet juice with *Candida utilis*. Journal of Food Science **41:** 78-81

Akita T, Hina Y, Nishi T (2002) New medium composition for high betacyanin production by a cell suspension culture of table beet (*Beta vulgaris* L.). Bioscience Biotechnology and Biochemistry **66:** 902-905

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology **215:** 403-410

Arciero DM, Lipscomb JD (1986) Binding of O_{17}-labeled substrate and inhibitors to protocatechuate 4,5-dioxygenase-nitrosyl complex - evidence for direct substrate binding to the active-site Fe^{2+} of extradiol dioxygenases. Journal of Biological Chemistry **261:** 2170-2178

Arciero DM, Orville AM, Lipscomb JD (1985) O_{17}-water and nitric-oxide binding by protocatechuate 4,5-dioxygenase and catechol 2,3-dioxygenase - evidence for binding of exogenous ligands to the active-site Fe^{2+} of extradiol dioxygenases. Journal of Biological Chemistry **260:** 4035-4044

Ashton NW, Cove DJ, Featherstone DR (1979) Isolation and physiological analysis of mutants of the moss, *Physcomitrella patens*, which over-produce gametophores. Planta **144:** 437-442

Bairoch A, Bucher P, Hofmann K (1997) The PROSITE database, its status in 1997. Nucleic Acids Research **25:** 217-221

Barnes WM (1994) Pcr amplification of up to 35-Kb DNA with high-fidelity and high-yield from lambda-bacteriophage templates. Proceedings of the National Academy of Sciences of the United States of America **91:** 2216-2220

Barth H, Kobayashi M, Musso H (1979) Pigments of *Amanita muscaria* - 5. Synthesis of muscaflavin. Helvetica Chimica Acta **62:** 1231-1235

Bate-Smith EC (1962) The phenolic constituents of plants and their taxonomic signifcance. J Linn Soc Bot **58:** 95-173

Bauer J (2001) Etude des voies métaboliques conduisant à la bétanine dans des cultures de *Beta vulgaris* L. par incorporation de tyrosine et de dihydroxyphenylalanine. PhD Thesis. University of Lausanne, Switzerland, Lausanne

Biddington NL, Thomas TH (1973) Modified *Amaranthus* betacyanin bioassay for rapid determination of cytokinins in plant extracts. Planta **111:** 183-186

Bigot C (1968) Action d'adénine substituées sur la synthèse des bétacyanines dans la plantule d'*Amaranthus caudatus* L. Comptes Rendus Hébdomadaires de l'Académie des Sciences Série D (Paris) **266:** 349-352

Bilyk A, Howard M (1982) Reversibility of thermal-degradation of betacyanines under the influence of isoascorbic acid. Journal of Agricultural and Food Chemistry **30:** 906-908

Bittrich V, Amaral MDE (1991) Proanthocyanidins in the testa of centrospermous seeds. Biochemical Systematics and Ecology **19:** 319-321

Bokern M, Heuer S, Strack D (1992) Hydroxycinnamic acid transferases in the biosynthesis of acylated betacyanins - purification and characterization from cell-cultures of *Chenopodium rubrum* and occurrence in some other members of the Caryophyllales. Botanica Acta **105:** 146-151

Bork P, Sander C, Valencia A (1993) Convergent evolution of similar enzymatic function on different protein folds - The hexokinase, ribokinase, and galactokinase families of sugar kinases. Protein Science **2:** 31-40

Borson ND, Sato WL, Drewes LR (1992) A lock-docking oligo(dT) primer for 5' and 3' RACE PCR. PCR Methods & Appl. **2:** 144-148

Boyington JC, Gaffney BJ, Amzel LM (1993) The 3-dimensional structure of an arachidonic-acid 15-lipoxygenase. Science **260:** 1482-1486

Bradford MM (1976) Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. Analytical Biochemistry **72:** 248-254

Cai YZ, Sun M, Corke H (2001a) Identification and distribution of simple and acylated betacyanins in the Amaranthaceae. Journal of Agricultural and Food Chemistry **49:** 1971-1978

Cai YZ, Sun M, Schliemann W, Corke H (2001b) Chemical stability and colorant properties of betaxanthin pigments from *Celosia argentea*. Journal of Agricultural and Food Chemistry **49:** 4429-4435

Chenchik A, Diachenko L, Moqadam F, Tarabykin V, Lukyanov S, Siebert PD (1996) Full-length cDNA cloning and determination of mRNA 5' and 3' ends by amplification of adaptor-ligated cDNA. Biotechniques **21:** 526-534

Christinet L, Burdet FX, Zaiko M, Hinz U, Zryd JP (2004) Characterization and functional identification of a novel plant 4,5-extradiol dioxygenase involved in betalain pigment biosynthesis in *Portulaca grandiflora*. Plant Physiol. **134:** 265-274

Clement JS, Mabry TJ (1996) Pigment evolution in the Caryophyllales: A systematic overview. Botanica Acta **109:** 360-367

Clifton IJ, Hsueh LC, Baldwin JE, Harlos K, Schofield CJ (2001) Structure of proline 3-hydroxylase - Evolution of the family of 2-oxoglutarate dependent oxygenases. European Journal of Biochemistry **268:** 6625-6636

Courtneygutterson N, Napoli C, Lemieux C, Morgan A, Firoozabady E, Robinson KEP (1994) Modification of flower color in florists chrysanthemum - production of a white-flowering variety through molecular-genetics. Bio-Technology **12:** 268-271

Cuenoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. American Journal of Botany **89:** 132-144

Davies KM, Bloor SJ, Spiller GB, Deroles SC (1998) Production of yellow colour in flowers: redirection of flavonoid biosynthesis in Petunia. Plant Journal **13:** 259-266

Davis KR, Ware D (1998) Lambda-PRL2 EST library from the Arabidopsis Biological Resource Center (ABRC).

De Jesus M (1995) Etude de l'accumulation de bétaxanthines par des cellules de *Beta vulgaris* L. cultivées en bioréacteur. PhD Thesis n° 1453. EPFL, Switzerland, Lausanne

de Vetten N, ter Horst J, van Schaik HP, de Boer A, Mol J, Koes R (1999) A cytochrome b(5) is required for full activity of flavonoid 3 ',5 '-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors. Proceedings of the National Academy of Sciences of the United States of America **96:** 778-783

Delgado-Vargas F, Jimenez AR, Paredes-Lopez O (2000) Natural pigments: Carotenoids, anthocyanins, and betalains - Characteristics, biosynthesis, processing, and stability. Critical Reviews in Food Science and Nutrition **40:** 173-289

Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R, Henne A, Wiezer A, Baumer S, Jacobi C, Bruggemann H, Lienard T, Christmann A, Bomeke M, Steckel S, Bhattacharyya A, Lykidis A, Overbeek R, Klenk HP, Gunsalus RP, Fritz HJ, Gottschalk G (2002) The genome of *Methanosarcina mazei*: Evidence for lateral gene transfer between bacteria and archaea. Journal of Molecular Microbiology and Biotechnology **4:** 453-461

Dhillon AS, Maurer AJ (1974) Evaluation of betalain pigments as colorants in chicken and turkey meat sausages. Poultry Science **53:** 1918-1918

Diatchenko L, Lau YFC, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996) Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proceedings of the National Academy of Sciences of the United States of America **93:** 6025-6030

Dmochowska A, Dignard D, Maleszka R, Thomas DY (1990) Structure and transcriptional control of the *Saccharomyces cerevisiae POX1* gene encoding acyl-Coenzyme A oxidase. Gene **88:** 247-252

Do YY, Huang PL (1997) Characterization of a pollination-related cDNA from *Phalaenopsis* encoding a protein which is homologous to human peroxisomal acyl-CoA oxidase. Archives of Biochemistry and Biophysics **344:** 295-300

Döpp H, Musso H (1973a) Die Konstitution dess Muscaflavins aus *Amanita muscaria* und über Betalaminsäure. Naturwissenschaften **60:** 477-478

Döpp H, Musso H (1973b) Isolierung und Chromophore der Farbstoffe aus *Amanita muscaria*. Chemische Berichte **106:** 3473-3482

Downham A, Collins P (2000) Colouring our foods in the last and next millennium. International Journal of Food Science and Technology **35:** 5-22

Dubourdieu M, Fox JL (1977) Amino acid sequence of *Desulfovibrio vulgaris* flavodoxin. Journal of Biological Chemistry **252:** 1453-1463

Dunkelbl E, Dreiding AS, Miller HE (1972) Mechanism of decarboxylation of betanidine - Contribution to interpretation of biosynthesis of betalaines. Helvetica Chimica Acta **55:** 642-645

Durand R, Zenk MH (1974) Homogentisate Ring-Cleavage Pathway in Biosynthesis of Acetate-Derived Naphthoquinones of Droseraceae. Phytochemistry **13:** 1483-1492

Eagen KA, Goldman IL (1996) Assessment of RAPD marker frequencies over cycles of recurrent selection for pigment concentration and percent solids in red beet (*Beta vulgaris* L.). Molecular Breeding **2:** 107-115

Eastmond PJ, Hooks M, Graham IA (2000a) The *Arabidopsis* acyl-CoA oxidase gene family. Biochemical Society Transactions **28:** 755-757

Eastmond PJ, Hooks MA, Williams D, Lange P, Bechtold N, Sarrobert C, Nussaume L, Graham IA (2000b) Promoter trapping of a novel medium-chain acyl-CoA oxidase, which is induced transcriptionally during *Arabidopsis* seed germination. Journal of Biological Chemistry **275:** 34375-34381

Ehrendorfer F (1976) Closing remarks - Systematics and evolution of centrospermous families. Plant Systematics and Evolution **126:** 99-105

Elbe JH, Klement JT, Amundson CH, Cassens RG, Lindsay RC (1974) Evaluation of betalain pigments as sausage colorants. Journal of Food Science **39:** 128-132

Elbe JH, Maing IY (1973) Betalains as possible food colorants of meat substitutes. Cereal Science Today **18:** 263-265

Elbe JH, Schwartz SJ, Hildenbrand BE (1981) Loss and regeneration of betacyanin pigments during processing of red beets. Journal of Food Science **46:** 1713-1715

Eltis LD, Bolin JT (1996) Evolutionary relationships among extradiol dioxygenases. Journal of Bacteriology **178:** 5930-5937

Endress R, Jager A, Kreis W (1984) Catecholamine biosynthesis dependent on the dark in betacyanin-forming *Portulaca* callus. Journal of Plant Physiology **115:** 291-295.

Escribano J, Pedreno MA, Garcia-Carmona F, Munoz R (1998) Characterization of the antiradical activity of betalains from *Beta vulgaris* L. roots. Phytochemical Analysis **9:** 124-127

Felsenstein J (1989) Phylogeny Inference Package (Version 3.2). Cladistics **5:** 164-166

Fernandes CL, Dong JH, Roebuck BD, Chisari FV, Montali JA, Schmidt DE, Prochaska HJ (1996) Elevations of hepatic quinone reductase, glutathione, and alpha- and mu-class glutathione S-transferase isoforms in mice with chronic hepatitis: A compensatory response to injury. Archives of Biochemistry and Biophysics **331:** 104-116

Fernandezcanon JM, Penalva MA (1995) Molecular characterization of a gene encoding a homogentisate dioxygenase from *Aspergillus nidulans* and identification of its human and plant homologs. Journal of Biological Chemistry **270:** 21199-21205

Fischer N, Dreiding AS (1972) Biosynthesis of betalains - On the cleavage of aromatic ring during the enzymatic transformation of DOPA into betalamic acid. Helvetica Chimica Acta **55:** 649-658

Froman BE, Edwards PC, Bursch AG, Dehesh K (2000) ACX3, a novel medium-chain acyl-coenzyme A oxidase from *Arabidopsis*. Plant Physiology **123:** 733-741

Galati EM, Mondello MR, Giuffrida D, Dugo G, Miceli N, Pergolizzi S, Taviano MF (2003) Chemical characterization and biological effects of Sicilian *Opuntia ficus-indica* (L.) Mill. fruit juice: Antioxidant and antiulcerogenic activity. Journal of Agricultural and Food Chemistry **51:** 4903-4908

Girod PA (1989) Régulation de la biosynthèse des bétalaïnes chez les cellules de *Beta vulgaris* L. en culture. PhD Thesis. University of Lausanne, Switzerland, Lausanne

Girod PA, Fu HY, Zryd JP, Vierstra RD (1999) Multiubiquitin chain binding subunit MCB1 (RPN10) of the 26S proteasome is essential for developmental progression in *Physcomitrella patens*. Plant Cell **11:** 1457-1471

Girod PA, Zryd JP (1987) Clonal variability and light induction of betalain synthesis in red beet cell-cultures. Plant Cell Reports **6:** 27-30

Girod PA, Zryd JP (1991a) Biogenesis of betalains - purification and partial characterization of DOPA 4,5-dioxygenase from *Amanita muscaria*. Phytochemistry **30:** 169-174

Girod PA, Zryd JP (1991b) Secondary metabolism in cultured red beet (*Beta vulgaris* L.) cells - Differential regulation of betaxanthin and betacyanin biosynthesis. Plant Cell Tissue and Organ Culture **25:** 1-12

Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. Nature Genetics **3:** 266-272

Giudici de Nicola M, Amico V, Sciuto S, Piattelli M (1975) Light control of amaranthin synthesis in isolated *Amaranthus* cotyledons. Phytochemistry **14:** 479-481

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis **18:** 2714-2723

Gurskaya NG, Diatchenko L, Chenchik A, Siebert PD, Khaspekov GL, Lukyanov KA, Vagner LL, Ermolaeva OD, Lukyanov SA, Sverdlov ED (1996) Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: Cloning of Jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-myristate 13-acetate. Analytical Biochemistry **240:** 90-97

Gurvitz A, Rottensteiner H, Kilpelainen SH, Hartig A, Hiltunen JK, Binder M, Dawes IW, Hamilton B (1997) The *Saccharomyces cerevisiae* peroxisomal 2,4-dienoyl-CoA reductase is encoded by the oleate-inducible gene *SPS19*. Journal of Biological Chemistry **272:** 22140-22147

Han D, Kim SJ, Kim SH, Kim DM (1998) Repeated regeneration of degraded red beet juice pigments in the presence of antioxidants. Journal of Food Science **63:** 69-72

Han S, Eltis LD, Timmis KN, Muchmore SW, Bolin JT (1995) Crystal-structure of the biphenyl-cleaving extradiol dioxygenase from a pcb-degrading Pseudomonad. Science **270:** 976-980

Hayashi H, De Bellis L, Ciurli A, Kondo M, Hayashi M, Nishimura R (1999) A novel acyl-CoA oxidase that can oxidize short-chain acyl-CoA in plant peroxisomes. Journal of Biological Chemistry **274:** 12715-12721

Hayashi H, De Bellis L, Yamaguchi K, Kato A, Hayashi M, Nishimura M (1998) Molecular characterization of a glyoxysomal long chain acyl-CoA oxidase that is synthesized as a precursor of higher molecular mass in pumpkin. Journal of Biological Chemistry **273:** 8301-8307

Hedden P, Phillips AL, Rojas MC, Carrera E, Tudzynski B (2001) Gibberellin biosynthesis in plants and fungi: A case of convergent evolution? Journal of Plant Growth Regulation **20:** 319-331

Heuer S, Richter S, Metzger JW, Wray V, Nimtz M, Strack D (1994) Betacyanins from bracts of *Bougainvillea glabra*. Phytochemistry **37:** 761-767

Heuer S, Strack D (1992) Synthesis of betanin from betanidin and UDP-glucose by a protein preparation from cell-suspension cultures of *Dorotheanthus bellidiformis* (Burm F) NEBr. Planta **186:** 626-628

Heuer S, Vogt T, Bohm H, Strack D (1996) Partial purification and characterization of UDP-glucose: Betanidin 5-*O*- and 6-*O*-glucosyltransferases from cell suspension cultures of *Dorotheanthus bellidiformis* (Burm f) NEBr. Planta **199:** 244-250

Heuer S, Wray V, Metzger JW, Strack D (1992) Betacyanins from flowers of *Gomphrena globosa*. Phytochemistry **31:** 1801-1807

Hinz U, Fivaz J, Girod PA, Zyrd JP (1997) The gene coding for the DOPA-dioxygenase involved in betalain biosynthesis in *Amanita muscaria* and its regulation. Molecular & General Genetics **256:** 1-6

Hooks MA, Bode K, Couee I (1996) Higher-plant medium- and short-chain acyl-CoA oxidases: Identification, purification and characterization of two novel enzymes of eukaryotic peroxisomal beta-oxidation. Biochemical Journal **320:** 607-614

Ibdah M, Krins A, Seidlitz HK, Heller W, Strack D, Vogt T (2002) Spectral dependence of flavonol and betacyanin accumulation in *Mesembryanthemum crystallinum* under enhanced ultraviolet radiation. Plant Cell and Environment **25:** 1145-1154

Ikeno S (1921) Studies on the genetics of flower-colours in *Portulaca grandiflora*. Journal of the College of Agriculture, Imperial University Tokyo **8:** 93-133

Impellizzeri G, Piattelli M (1972) Pigments of Centrospermae-16. - Biosynthesis of indicaxanthine in *Opuntia ficus-indica* L. fruits. Phytochemistry **11:** 2499-2502

Impellizzeri G, Piattelli M, Sciuto S (1973) Pigments of Centrospermae-17. - A new betaxanthin from *Glottiphyllum longum*. Phytochemistry **12:** 2293-2294

IturriagagoitiaBueno T, Gibson EJ, Schofield CJ, John P (1996) Inhibition of 1-aminocyclopropane-1-carboxylate oxidase by 2-oxoacids. Phytochemistry **43:** 343-349

Jefferson RA, Kavanagh TA, Bevan MW (1987) Gus fusions - Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher-plants. EMBO Journal **6:** 3901-3907

Johnson ET, Ryu S, Yi HK, Shin B, Cheong H, Choi G (2001) Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. Plant Journal **25:** 325-333

Kanner J, Harel S, Granit R (2001) Betalains - A new class of dietary cationized antioxidants. Journal of Agricultural and Food Chemistry **49:** 5178-5185

Kapadia GJ, Azuine MA, Sridhar R, Okuda Y, Tsuruta A, Ichiishi E, Mukainake T, Takasaki M, Konoshima T, Nishino H, Tokuda H (2003) Chemoprevention of DMBA-induced UV-B promoted, NOR-1-induced TPA promoted skin carcinogenesis, and DEN-induced phenobarbital promoted liver tumors in mice by extract of beetroot. Pharmacological Research **47:** 141-148

Kapadia GJ, Tokuda H, Konoshima T, Nishino H (1996) Chemoprevention of lung and skin cancer by *Beta vulgaris* (beet) root extract. Cancer Letters **100:** 211-214

Kobayashi N, Schmidt J, Nimtz M, Wray V, Schliemann W (2000) Betalains from Christmas cactus. Phytochemistry **54:** 419-426

Kobayashi N, Schmidt J, Wray V, Schliemann W (2001) Formation and occurrence of dopamine-derived betacyanins. Phytochemistry **56:** 429-436

Kochhar VK, Kochhar S, Mohr H (1981) An analysis of the action of light on betalain synthesis in the seedling of *Amaranthus caudatus*, var *Viridis*. Planta **151:** 81-87

Kost B, Spielhofer P, Chua NH (1998) A GFP-mouse talin fusion protein labels plant actin filaments in vivo and visualizes the actin cytoskeleton in growing pollen tubes. Plant Journal **16:** 393-401

Kosugi Y, Oyamada N, Satoh S, Yoshioka T, Onodera E, Yamada Y (1997) Inhibition by 1-aminocyclobutane-1-carboxylate of the activity of 1-aminocyclopropane-1-carboxylate oxidase obtained from senescing petals of carnation (*Dianthus caryophyllus* L.) flowers. Plant and Cell Physiology **38:** 312-318

Kozak M (1991) Structural features in eukaryotic messenger-RNAs that modulate the initiation of translation. Journal of Biological Chemistry **266:** 19867-19870

Kunau WH, Hartig A (1992) Peroxisome biogenesis in *Saccharomyces cerevisiae*. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology **62:** 63-78

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology **157:** 105-132

Laemmli UK (1970) Cleavage of structural proteins during assembly of head of bacteriophage-T4. Nature **227:** 680-685

Leathers RR, Davin C, Zryd JP (1992) Betalain producing cell-cultures of *Beta vulgaris* L. var. Bikores Monogerm (red beet). In Vitro Cellular & Developmental Biology-Plant **28P:** 39-45

Liebisch HW, Matschin.B, Schutte HR (1969) Studies on physiology and biosynthesis of betanin. Zeitschrift Fur Pflanzenphysiologie **61:** 269-271

Lipscomb JD, Orville AM (1992) Mechanistic aspects of dihydroxybenzoate dioxygenases. Metal Ions in Biological Systems **28:** 243-298

Mabry TJ (1973) Is the order Centrospermae monophyletic? In: Chemistry in botanical classification, Bend G and Santesson J (eds), Academic press, London

Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Research **30:** 281-283

Miller HE, Rosler H, Wohlpart A, Wyler H, Wilcox ME, Frohofer H, Mabry TJ, Dreiding AS (1968) Biogenesis of Betalaine. Biotransformation of DOPA and tyrosine into betalamic acid part of betanin. Helvetica Chimica Acta **51:** 1470-&

Minale L, Piattelli M, De Stefano S (1967) Pigments of Centrospermae-7. - Betacyanins from *Gomphrena globosa* L. Phytochemistry **6:** 703-709

Mittendorf V, Bongcam V, Allenbach L, Coullerez G, Martini N, Poirier Y (1999) Polyhydroxyalkanoate synthesis in transgenic plants as a new tool to study carbon flow through beta-oxidation. Plant Journal **20:** 45-55

Mittendorf V, Robertson EJ, Leech RM, Kruger N, Steinbuchel A, Poirier Y (1998) Synthesis of medium-chain-length polyhydroxyalkanoates in *Arabidopsis thaliana* using intermediates of peroxisomal fatty acid beta-oxidation. Proceedings of the National Academy of Sciences of the United States of America **95:** 13397-13402

Mol J, Cornish E, Mason J, Koes R (1999) Novel coloured flowers. Current Opinion in Biotechnology **10:** 198-201

Mueller LA (1997) Enzymologie de la biosynthèse des bétalaines chez *Amanita Muscaria*. PhD Thesis. Lausanne university

Mueller LA, Hinz U, Uze M, Sautter C, Zryd JP (1997a) Biochemical complementation of the betalain biosynthetic pathway in *Portulaca grandiflora* by a fungal 3,4-dihydroxyphenylalanine dioxygenase. Planta **203:** 260-263

Mueller LA, Hinz U, Zryd JP (1997b) The formation of betalamic acid and muscaflavin by recombinant dopa-dioxygenase from *Amanita*. Phytochemistry **44:** 567-569

Mueller LA, Hinz U, Zrÿd JP (1996) Characterization of a tyrosinase from *Amanita muscaria* involved in betalain biosynthesis. Phytochemistry **42:** 1511-1515

Mukundan U, Bhagwat V, Singh G, Curtis W (2001) Integrated recovery of pigments released from red beet hairy roots exposed to acidic medium. Journal of Plant Biochemistry and Biotechnology **10:** 67-69

Mukundan U, Carvalho EB, Curtis WR (1998) Growth and pigment production by hairy root cultures of *Beta vulgaris* L. in a bubble column reactor. Biotechnology Letters **20:** 469-474

Mumberg D, Muller R, Funk M (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene **156:** 119-122

Musso H (1979) The pigments of fly agaric, *Amanita muscaria*. Tetrahedron **35:** 2843-2853

Nakai K, Horton P (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends in Biochemical Sciences **24:** 34-35

Nakayama T, Sato T, Fukui Y, Yonekura-Sakakibara K, Hayashi H, Tanaka Y, Kusumi T, Nishino T (2001) Specificity analysis and mechanism of aurone synthesis catalyzed by aureusidin synthase, a polyphenol oxidase homolog responsible for flower coloration. FEBS Letters **499:** 107-111

Nakayama T, Yonekura-Sakakibara K, Sato T, Kikuchi S, Fukui Y, Fukuchi-Mizutani M, Ueda T, Nakao M, Tanaka Y, Kusumi T, Nishino T (2000) Aureusidin synthase: A polyphenol oxidase homolog responsible for flower coloration. Science **290:** 1163-1166

Newman T, Debruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M, Retzel E, Somerville C (1994) Genes galore - a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. Plant Physiology **106:** 1241-1255

Nishiyama T, Fujita T, Shin-I T, Seki M, Nishide H, Uchiyama I, Kamiya A, Carninci P, Hayashizaki Y, Shinozaki K, Kohara Y, Hasebe M (2003) Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: Implication for land plant evolution. Proceedings of the National Academy of Sciences of the United States of America **100:** 8007-8012

Noda Y, Nishikawa S, Shiozuka KI, Kadokura H, Nakajima H, Yoda K, Katayama Y, Morohoshi N, Haraguchi T, Yamasaki M (1990) Molecular-cloning of the protocatechuate 4,5-dioxygenase genes of *Pseudomonas paucimobilis*. Journal of Bacteriology **172:** 2704-2709

Oud JSN, Schneiders H, Kool AJ, Vangrinsven MQJM (1995) Breeding of transgenic orange *Petunia hybrida* varieties. Euphytica **84:** 175-181

Pasch JH, Vonelbe JH, Sell RJ (1975) Betalaines as colorants in dairy-products. Journal of Milk and Food Technology **38:** 25-28

Patkai G, Barta J (1996) Decomposition of betacyanins and betaxanthins by heat and pH changes. Nahrung-Food **40:** 267-270

Pedreno MA, Escribano J (2001) Correlation between antiradical activity and stability of betanine from *Beta vulgaris* L roots under different pH, temperature and light conditions. Journal of the Science of Food and Agriculture **81:** 627-631

Piattelli M, Impellizzeri G (1970) Pigments of Centrospermae-12. - 2-descarboxybetanidin, a minor betacyanin from *Carpobrotus acinaciformis*. Phytochemistry **9:** 2553-2556

Piattelli M, Imperato F (1969) Betacyanins of family Cactaceae. Phytochemistry **8:** 1503-1507

Piattelli M, Minale L (1964) Pigments of Centrospermae-2. - Distribution of betacyanins. Phytochemistry **3:** 547-557

Piattelli M, Minale L, Nicolaus RA (1965a) Pigments of Centrospermae-V. - Betaxanthins from *Mirabilis jalapa* L. Phytochemistry **4:** 817-823

Piattelli M, Minale L, Nicolaus RA (1965b) Ulteriori ricerche sulle betaxanthine. Rend. Accad. Sci. Fis. Mat. **32:** 55-56

Piattelli M, Minale L, Prota G (1964) Isolation structure and absolute configuration of indicaxanthin. Tetrahedron **20:** 2325-2329

Piattelli M, Minale L, Prota G (1965c) Pigments of Centrospermae-III. - Betaxanthins from *Beta vulgaris* L. Phytochemistry **4:** 121-125

Poirier Y, Erard N, Petetot JMC (2001) Synthesis of polyhydroxyalkanoate in the peroxisome of *Saccharomyces cerevisiae* by using intermediates of fatty acid beta-oxidation. Applied and Environmental Microbiology **67:** 5254-5260

Porta H, Rocha-Sosa M (2002) Plant lipoxygenases. Physiological and molecular features. Plant Physiology **130:** 15-21

Prescott AG, John P (1996) Dioxygenases: Molecular structure and role in plant metabolism. Annual Review of Plant Physiology and Plant Molecular Biology **47:** 245-271

Prescott AG, Lloyd MD (2000) The iron(II) and 2-oxoacid-dependent dioxygenases and their role in metabolism. Natural Product Reports **17:** 367-383

Reynoso R, Garcia FA, Morales D, deMejia EG (1997) Stability of betalain pigments from a Cactaceae fruit. Journal of Agricultural and Food Chemistry **45:** 2884-2889

Reynoso RC, Giner TV, De Mejia EG (1999) Safety of a filtrate of fermented garambullo fruit: Biotransformation and toxicity studies. Food and Chemical Toxicology **37:** 825-830

Rochester DE, Winer JA, Shah DM (1986) The Structure and Expression of Maize Genes Encoding the Major Heat-Shock Protein, Hsp70. EMBO Journal **5:** 451-458

Romanov GA, Getman IA, Schmulling T (2000) Investigation of early cytokinin effects in a rapid *Amaranthus* seedling test. Plant Growth Regulation **32:** 337-344

Rösler J, Krekel F, Amrhein N, Schmid J (1997) Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase activity. Plant Physiology **113:** 175-179

Rossi-Hassani BD (1991) Mutabilité et biosynthèse des bétalaïnes chez *Portulaca grandiflora* Hook. PhD Thesis. University of Lausanne, Switzerland, Lausanne

Rost B, Sander C (1993) Prediction of protein secondary structure at better than 70-percent accuracy. Journal of Molecular Biology **232:** 584-599

Saito K, Komamine A, Hatanaka SI (1978) Biochemical studies on nitrogen-compounds of fungi-17. Biosynthesis of stizolobic and stizolobinic Acids in *Amanita pantherina*. Zeitschrift Fur Naturforschung C-a Journal of Biosciences **33:** 793-795

Saito K, Komamine A, Senoh S (1976) Further studies on biosynthesis of stizolobinic acid and stizolobic acid in etiolated seedlings of *Stizolobium hassjoo*. Zeitschrift Fur Naturforschung C-a Journal of Biosciences **31:** 15-17

Sakuta M, Hirano H, Kakegawa K, Suda J, Hirose M, Joy RW, Sugiyama M, Komamine A (1994) Regulatory mechanisms of biosynthesis of betacyanin and anthocyanin in relation to cell-division activity in suspension-cultures. Plant Cell Tissue and Organ Culture **38:** 167-169

Sambrook J, Fritsch, E.F., Maniatis, T: (1989) Molecular cloning: a laboratory manual. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York

Sato T, Nakayama T, Kikuchi S, Fukui Y, Yonekura-Sakakibara K, Ueda T, Nishino T, Tanaka Y, Kusumi T (2001) Enzymatic formation of aurones in the extracts of yellow snapdragon flowers. Plant Science **160:** 229-236

Schaefer DG, Zryd JP (1997) Efficient gene targeting in the moss *Physcomitrella patens*. Plant Journal **11:** 1195-1206

Schliemann W, Cai YZ, Degenkolb T, Schmidt J, Corke H (2001) Betalains of *Celosia argentea*. Phytochemistry **58:** 159-165

Schliemann W, Joy RW, Komamine A, Metzger JW, Nimtz M, Wray V, Strack D (1996) Betacyanins from plants and cell cultures of *Phytolacca americana*. Phytochemistry **42:** 1039-1046

Schliemann W, Kobayashi N, Strack D (1999) The decisive step in betaxanthin biosynthesis is a spontaneous reaction. Plant Physiology **119:** 1217-1232

Schliemann W, Steiner U, Strack D (1998) Betanidin formation from dihydroxyphenylalanine in a model assay system. Phytochemistry **49:** 1593-1598

Schwitzguébel JP, Zryd JP, Leathers RR (1991) From plant cells to biotechnology. Swiss Biotechnology**:** 17-24

Sciuto S, Oriente G, Piattell.M, Impelliz.G, Amico V (1974) Pigments of Centrospermae-19. - Biosynthesis of amaranthin in *Celosia plumosa*. Phytochemistry **13:** 947-951

Siedow JN (1991) Plant Lipoxygenase - Structure and Function. Annual Review of Plant Physiology and Plant Molecular Biology **42:** 145-188

Spence EL, Kawamukai M, Sanvoisin J, Braven H, Bugg TDH (1996) Catechol dioxygenases from *Escherichia coli* (MhpB) and *Alcaligenes eutrophus* (MpcI): Sequence analysis and biochemical properties of a third family of extradiol dioxygenases. Journal of Bacteriology **178:** 5249-5256

Stanway CA, Gibbs JM, Berardi E (1995) Expression of the *Fox1* gene of *Saccharomyces cerevisiae* is regulated by carbon source, but not by the known glucose repression genes. Current Genetics **27:** 404-408

Steddom K, Heidel G, Jones D, Rush CM (2003) Remote detection of rhizomania in sugar beets. Phytopathology **93:** 720-726

Steiner U, Schliemann W, Bohm H, Strack D (1999) Tyrosinase involved in betalain biosynthesis of higher plants. Planta **208:** 114-124

Steiner U, Schliemann W, Strack D (1996) Assay for tyrosine hydroxylation activity of tyrosinase from betalain-forming plants and cell cultures. Analytical Biochemistry **238:** 72-75

Stintzing FC, Schieber A, Carle R (2001) Phytochemical and nutritional significance of cactus pear. European Food Research and Technology **212:** 396-407

Stintzing FC, Schieber A, Carle R (2002a) Betacyanins in fruits from red-purple pitaya, *Hylocereus polyrhizus* (Weber) Britton & Rose. Food Chemistry **77:** 101-106

Stintzing FC, Schieber A, Carle R (2002b) Identification of betalains from yellow beet (*Beta vulgaris* L.) and cactus pear [*Opuntia ficus-indica* (L.) Mill.] by high-performance liquid chromatography-electrospray ionization mass spectrometry. Journal of Agricultural and Food Chemistry **50:** 2302-2307

Stobart AK, Kinsman LT (1977) Hormonal-Control of Betacyanin Synthesis in *Amaranthus caudatus*. Phytochemistry **16:** 1137-1142

Strack D, Schmitt D, Reznik H, Boland W, Grotjahn L, Wray V (1987) Humilixanthin, a new betaxanthin from *Rivina humilis*. Phytochemistry **26:** 2285-2287

Strack D, Vogt T, Schliemann W (2003) Recent advances in betalain research. Phytochemistry **62:** 247-269

Strack D, Wray V (1989) Anthocyanins, Vol Plant Phenolics. In: Method in plant biochemistry, Dey PM and Harborne JB (eds), Academic Press, London

Subramani S (1996) Protein translocation into peroxisomes. Journal of Biological Chemistry **271:** 32483-32486

Sugimoto K, Senda T, Aoshima H, Masai E, Fukuda M, Mitsui Y (1999) Crystal structure of an aromatic ring opening dioxygenase LigAB, a protocatechuate 4,5-dioxygenase, under aerobic conditions. Structure with Folding & Design **7:** 953-965

Suzuki H, Nakayama T, Yonekura-Sakakibara K, Fukui Y, Nakamura N, Yamaguchi M, Tanaka Y, Kusumi T, Nishino T (2002) cDNA cloning, heterologous expressions, and functional characterization of malonyl-Coenzyme A : anthocyanidin 3-*O*-glucoside-6 "-*O*-malonyltransferase from dahlia flowers. Plant Physiology **130:** 2142-2151

Tatsuno Y, Saeki Y, Nozaki M, Otsuka S, Maeda Y (1980) Mossbauer-spectra of metapyrocatechase. FEBS Letters **112:** 83-85

Tatusova TA, Madden TL (1999) BLAST 2 SEQUENCES, a new tool for comparing protein and nucleotide sequences. FEMS Microbiology Letters **174:** 247-250

Terradas F (1989) Etude par voie enzymatique des métabolites des bétalaïnes dans *Beta vulgaris* et *Amanita muscaria*. PhD Thesis. University of Lausanne, Switzerland, Lausanne

Terradas F, Wyler H (1991a) 2,3-*seco*-DOPA and 4,5-*seco*-DOPA, the biosynthetic intermediates generated from L-DOPA by an enzyme-system extracted from the fly agaric, *Amanita muscaria* L., and their spontaneous conversion to muscaflavin and betalamic acid, respectively, and betalains. Helvetica Chimica Acta **74:** 124-140

Terradas F, Wyler H (1991b) The *seco*-DOPAs, natural pigments in *Hygrocybe conica* and *Amanita muscaria*. Phytochemistry **30:** 3251-3253

Tesoriere L, Butera D, D'Arpa D, Di Gaudio F, Allegra M, Gentile C, Livrea MA (2003) Increased resistance to oxidation of betalain-enriched human low density lipoproteins. Free Radical Research **37:** 689-696

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through tequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22:** 4673-4680

Trezzini GF (1990a) Génétique des bétalaïnes chez *Portulaca grandiflora* Hook. PhD Thesis. University of Lausanne, Switzerland, Lausanne

Trezzini GF, Zryd JP (1991a) Characterization of some natural and semisynthetic betaxanthins. Phytochemistry **30:** 1901-1903

Trezzini GF, Zryd JP (1991b) Two betalains from *Portulaca grandiflora*. Phytochemistry **30:** 1897-1899

Trezzini GF, Zryd, J.P. (1990b) *Portulaca grandiflora*: A model system for the study of the biochemistry and genetics of betalain synthesis. Acta Horticulturae **280:** 581-585

Tudzynski B, Holter K (1998) Gibberellin biosynthetic pathway in *Gibberella fujikuroi*: Evidence for a gene cluster. Fungal Genetics and Biology **25:** 157-170

Vogt T (2002) Substrate specificity and sequence analysis define a polyphyletic origin of betanidin 5- and 6-O-glucosyltransferase from *Dorotheanthus bellidiformis*. Planta **214:** 492-495

Vogt T, Grimm R, Strack D (1999a) Cloning and expression of a cDNA encoding betanidin 5-*O*-glucosyltransferase, a betanidin- and flavonoid-specific enzyme with high homology to inducible glucosyltransferases from the Solanaceae. Plant Journal **19:** 509-519

Vogt T, Ibdah M, Schmidt J, Wray V, Nimtz M, Strack D (1999b) Light-induced betacyanin and flavonol accumulation in bladder cells of *Mesembryanthemum crystallinum*. Phytochemistry **52:** 583-592

Vogt T, Zimmermann E, Grimm R, Meyer M, Strack D (1997) Are the characteristics of betanidin glucosyltransferases from cell-suspension cultures of *Dorotheanthus bellidiformis* indicative of their phylogenetic relationship with flavonoid glucosyltransferases? Planta **203:** 349-361

von Ardenne R, Döpp H, Musso H, Steglich W (1974) Über das Vorkommen von Muscaflavin bei *Hygrocyben* (Agaricales) und seine Dihydroazepin-Struktur. Zeitschrift der Naturforschung **29c:** 637-639

Weiss MR (1995) Floral colour-change - A widespread functional convergence. American Journal of Botany **82:** 167-185

Wilcox ME, Wyler H, Dreiding AS (1965) Stereochemie von betanidin und isobetanidin. Helvetica Chimica Acta **48:** 1134-1147

Wohlpart A, Mabry TJ (1968) On light requirement for betalain biogenesis. Plant Physiology **43:** 457-459

Wuthrich B (1993) Adverse reactions to food-additives. Annals of Allergy **71:** 379-384

Wybraniec S, Platzner I, Geresh S, Gottlieb HE, Haimberg M, Mogilnitzki M, Mizrahi Y (2001) Betacyanins from vine cactus *Hylocereus polyrhizus*. Phytochemistry **58:** 1209-1212

Wyler H, Dreiding AS (1961) Uber Betacyane, die Stickstoffhaltigen Farbstoffe der Centrospermen - Vorlaufige Mitteilung. Experientia **17:** 23-25

Wyler H, Mabry TJ, Dreiding AS (1963) Zur Struktur des Betanidins. Helvetica Chimica Acta **46:** 188-189

Yamamoto K, Kobayashi N, Yoshitama K, Teramoto S, Komamine A (2001) Isolation and purification of tyrosine hydroxylase from callus cultures of *Portulaca grandiflora*. Plant and Cell Physiology **42:** 969-975

Yang XH, Zhang XT, Zhou RH (2001) Determination of L-DOPA content and other significant nitrogenous compounds in the seeds of seven *Mucuna* and *Stizolobium* species in China. Pharmaceutical Biology **39:** 312-316

Yashi K (1920) Genetical studies in *Portulaca grandiflora*. Botanical Magazine Tokyo **34:** 55-65

Zaiko M (2000) Colour-specific genes from betalain producing plants. PhD Thesis. University of Lausanne, Switzerland, Lausanne

Zakharova NS, Petrova TA (1998) Relationships between the structure and antioxidant activity of certain betalains. Applied Biochemistry and Microbiology **34:** 182-185