



**Etude fonctionnelle de gènes homologues à une
DOPA-dioxygénase de *Portulaca grandiflora*
chez des plantes sans bétalaïne**

Travail de diplôme

Présenté à la Faculté des Sciences de l'Université de Lausanne

Institut d'écologie – Laboratoire de phytogénétique cellulaire

Par

Frédéric Burdet

Directeur de diplôme : Prof. J.-P. Zryd

Lausanne, 2003

Résumé

Le gène de synthèse des bétalaïnes, pigments jaunes et violets, vient d'être isolé chez *Portulaca grandiflora* (pourpier). Il s'agit d'une dihydroxyphénylalanine dioxygénase. Des homologues à ce gène existent chez les plantes avec et sans bétalaïne. Afin de découvrir quelle était la fonction de ce gène chez les plantes sans bétalaïne, nous avons tenté de faire un knock-out chez la mousse *Physcomitrella patens*. Il n'est pas sûr que nous en ayons produit un knock-out, en tous cas aucun phénotype n'est visible parmi les clones isolés.

Nous avons exprimé l'homologue d'*Arabidopsis thaliana* dans des pétales de pourpier par complémentation biochimique. D'après nos observations, ce gène n'est pas capable de synthétiser des bétalaïnes, mais doit avoir une fonction importante étant donné sa haute conservation parmi les plantes.

Des séquences homologues chez les plantes à bétalaïnes ont été obtenues, permettant de prédire, par des alignements et des modélisations tridimensionnelles, plusieurs résidus qui seraient impliqués dans le mécanisme catalytique. Un consensus de trois acides aminés a pu être défini chez les plantes avec et sans bétalaïne à partir de l'alignement.

Table des matières

Introduction	4
Résultats	12
Etude fonctionnelle du gène d' <i>Arabidopsis thaliana</i>	12
Obtention de séquences de la DOD chez d'autres plantes	13
Différence fonctionnelle entre le gène DOD et ses homologues	15
Knock-out du gène L.6 chez <i>Physcomitrella patens</i>	24
Conclusions et perspectives	27
Matériel et méthodes	29
Références	33
Annexes	35

Introduction

Les bétalaïnes sont une classe structurellement et biosynthétiquement distincte de pigments naturels, et se répartissent en deux types : les bétaxanthines jaunes et les bétacyanines violettes. On avait observé depuis longtemps que ces pigments étaient différents des anthocyanes et contenaient de l'azote. Ce sont donc des chromoalcaloïdes et c'est dans les années 1960 que leur structure a été élucidée (Wilcox et al., 1965).

Les plantes contenant des bétalaïnes appartiennent toutes à l'ordre des Caryophyllales. Ce sont souvent des plantes grasses. A part les familles des Caryophyllacées et les Molluginacées, elles ne contiennent jamais d'anthocyanes, de même, ces dernières ainsi que toutes les plantes à anthocyanes ne contiennent jamais de bétalaïnes. Les bétalaïnes se trouvent aussi chez les champignons, des genres *Amanita* et *Hygrocybe*. Ces derniers contiennent des bétalaïnes généralement différentes que celles contenues dans les plantes : le chromophore est le même, mais les chaînes latérales sont différentes (cf. schéma de synthèse fig. 1).

Les bétalaïnes sont composées d'un chromophore, l'acide bétalamique. Il est formé par le clivage de la L-DOPA (dihydroxyphénylalanine) entre les positions 4 et 5 (dit extradiol) et la recyclisation qui s'en suit, formant l'anneau dihydropyridine caractéristique. L'enzyme catalysant cette réaction est donc une DOPA-dioxygénase, qui ajoute une molécule de dioxygène entre les positions 4 et 5 de la molécule de DOPA. Ce chromophore est conjugué par une liaison imino à un acide aminé (bétaxanthines) ou à un résidu cyclo-DOPA qui peut être glycosylé (bétacyanines). Le chromophore et le cyclo-DOPA glycosylé dérivent d'une molécule de tyrosine.

Des dérivés provenant du clivage du cycle de la tyrosine ont été observés chez beaucoup de plantes. Les seuls autres enzymes connues chez les plantes supérieures qui utilisent la L-DOPA comme substrat préféré de clivage et de recyclisation sont la stizolobic acid synthase et stizolobinic acid synthase de *Stizolobium spp.*, de la famille des Légumineuses (Saito and Komamine 1976).

La fonction des bétalaïnes chez les plantes est ignorée à ce jour, mais elles pourraient jouer un rôle dans la pollinisation par les insectes. Certains ont postulé qu'elles pourraient avoir une fonction de défense, par exemple de protection contre les UV en les absorbant ou de résistance au stress hydrique (contre-ions stockés dans la vacuole).

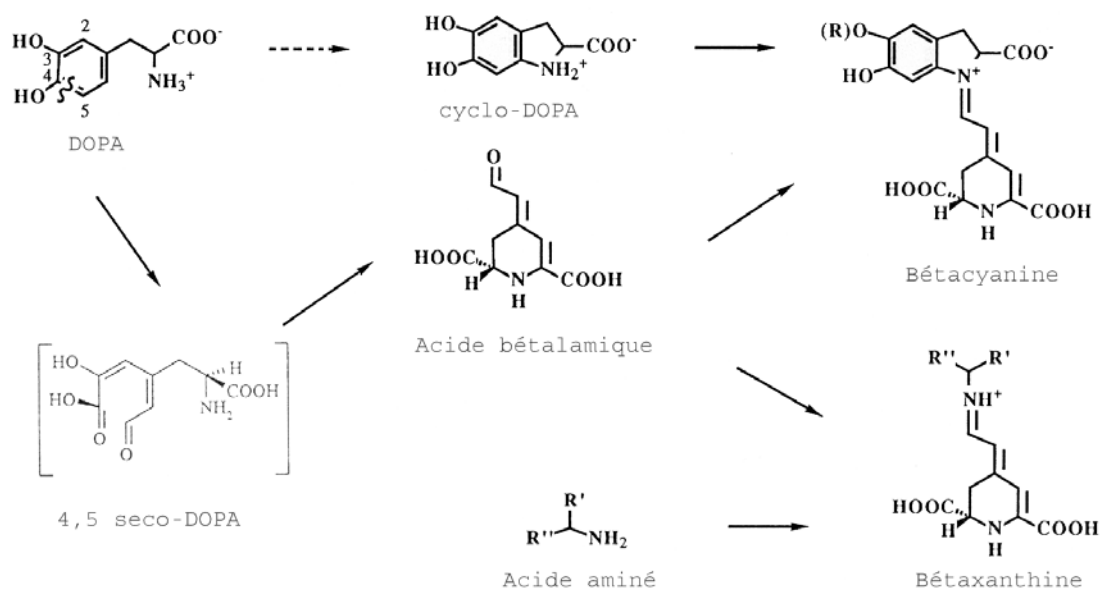


Figure 1 : Schéma de synthèse des bétalaïnes. La DOPA est clivée sur la liaison 4,5 par la dioxygénase, générant de l'acide bétalamique, avec la 4,5 seco-DOPA comme intermédiaire. La cyclo-DOPA dériverait également de la DOPA. Les bétacyanines (violette) résultent de la conjugaison de l'acide bétalamique et de la cyclo-DOPA. Les bétaxanthines dérivent de la conjugaison de l'acide bétalamique avec un acide aminé.

Les bétalaïnes ne sont pas seulement produites dans les fleurs et les fruits, mais aussi lors de blessures, infections par des pathogènes, sénescence et dans les jeunes pousses. La présence de flavonoïdes chez les Caryophyllales produisant des bétalaïnes suggère que ces dernières pourraient avoir perdu un gène important pour la synthèse des anthocyanes (Bittrich and Amaral 1991)

Une des hypothèses pour expliquer l'apparition des bétalaïnes est due à Ehrendorfer (1976). Un ancêtre du groupe des plantes à bétalaïnes pourrait avoir perdu la faculté de produire des anthocyanes dans un habitat aride à semi-aride pauvre en pollinisateurs, à l'époque du Crétacé, par un abaissement de la pression sélective. Les bétalaïnes pourraient être apparues plus tard lorsque des pollinisateurs se seraient adaptés à l'environnement aride ou lorsque les plantes elles-mêmes se seraient adaptées à un environnement moins aride, plus riche en pollinisateurs. Cette hypothèse part du principe que les Chenopodiinae (sous ordre des Caryophyllales regroupant les plantes à bétalaïnes) sont monophylétiques, ce que des études phylogénétiques moléculaires semblent confirmer (Cuénoud et al., 2002).

Une autre hypothèse suggère que les anthocyanes et les bétalaïnes puissent également avoir « cohabité » chez une même espèce pendant une période. Cette espèce hypothétique aurait ensuite perdu la faculté de synthétiser des anthocyanes (Clement, 1996). La similarité

dans le développement, la distribution dans les tissus et la fonction entre les anthocyanes et les bétalaïnes indiquent que des gènes comparables (ou les mêmes) en régulent la biosynthèse, même si les voies de synthèse sont différentes (Stafford 1994).

Dans le laboratoire du professeur Zrýd, on cherchait depuis plusieurs années une enzyme, en principe une dioxygénase, qui devait catalyser la transformation de la DOPA (dihydroxyphénylalanine) en acide bétalamique, par ouverture du cycle aromatique en 4,5 (cf. fig. 1).

Chez le champignon *Amanita muscaria*, Girod et Zrýd ont purifié et partiellement caractérisé une enzyme transformant la DOPA en acide bétalamique et identifiée comme la DOPA-4,5-dioxygénase d'*A. muscaria* (Girod and Zrýd 1990). Dans les conditions qu'ils ont testées, elle coupait exclusivement la double liaison en 4,5 et produisait uniquement de l'acide bétalamique. Cette enzyme fongique n'a pas d'homologues chez les plantes (Fivaz 1994). Les méthodes pour isoler le gène chez les plantes, utilisant des homologies de séquence ou des anticorps, se sont donc révélées inefficaces.

Les extraits bruts de *Portulaca grandiflora* (pourpier) et *Beta vulgaris* (betterave) ne contiennent pas d'activité « DOPA-dioxygénase » détectable. On s'est donc intéressé à leurs cDNAs, afin de tenter d'isoler le gène de la DOPA-dioxygénase des plantes. Plusieurs candidats avaient été identifiés, mais aucun ne semblait correspondre à une dioxygénase, malgré les comparaisons répétées avec les bases de données existantes. Toutefois, le clone « L.6 », cDNA de *Portulaca grandiflora* isolé par Zaiko à partir d'une banque soustractive, a montré une homologie avec le domaine LigB récemment défini chez des dioxygénases bactériennes : pFam 02900. L'homologie globale n'est que de 20% environ, c'est pourquoi on ne la retrouve pas au moyen d'un blast classique. Ce clone est exprimé spécifiquement dans les tissus pigmentés (pétales, tiges) du pourpier, mais pas dans les non colorés (Christinet, 2002, non publié). Toutefois, le test de complémentation par biolistique (Mueller, Hinz et al. 1997) n'a pas fonctionné (Zaiko 2000).

Après avoir aligné la séquence protéinique avec celle de la bactérie *Sphingomonas* (AAA17728), Christinet s'est aperçu que les acides aminés catalytiques (définis à partir de la structure 3D) et ceux impliqués dans la fixation du cofacteur Fe^{2+} par exemple, étaient largement conservés. Il y avait donc des raisons de penser que le clone L.6 était une dioxygénase de même type. La construction pour la complémentation a donc été refaite, cette fois directement à partir du DNA génomique (ce gène ne contient pas d'intron chez *Portulaca grandiflora*), et la complémentation chez le pourpier avec ce nouveau gène a effectivement donné des taches sur les pétales (Christinet 2002, non publié), prouvant la fonction de

dioxygénase. De plus, l'analyse par HPLC des pigments synthétisés dans les cellules transformées a montré qu'il s'agissait bien des bétalaïnes trouvées dans les pétales naturellement colorés. Le cDNA précédent contenait apparemment des erreurs. Le gène L.6 de pourpier est donc une DOPA-dioxygénase. Toutefois le même gène chez d'autres plantes à bétalaïnes n'ayant pas encore été testé, on ne peut donc pas encore généraliser et dire que ce gène a pour fonction la synthèse de l'acide bétalamique chez toutes les plantes à bétalaïnes. Un Northern blot sur du RNA de *Beta vulgaris* (Christinet, 2003, non publié) montre que le profil d'expression de l'homologue L.6 chez cette plante est le bon : le gène est spécifiquement exprimé chez les cellules colorées (cultivées in vitro).

En comparant la séquence nucléotidique du pourpier aux bases de données, on se rend compte que des homologues (un ou deux par espèces) existent chez beaucoup d'autres plantes sans bétalaïne (surtout des EST). Les gènes sont globalement homologues à 70%, avec des zones très conservées. On peut donc supposer, vu sa haute conservation, que ce gène a également un rôle important chez les plantes sans bétalaïne, l'enzyme ayant certainement un substrat différent de celle des plantes avec bétalaïnes.

Famille	Caractéristiques	Produits / Type de réaction
Lipoxygénases	Fer (II ou III) Incorporation d'O ₂ dans des acides gras	Acide jasmonique Certains arômes
2-oxoacid dépendant dioxygénases	Fer (II ou III) Hydroxylations Désaturations	Flavonoïdes Gibbérélines Alcaloïdes
Dioxygénases intradiol (bactéries)	Fe (III) Clivent en ortho par rapport au groupe OH	Dégradation de composés aromatiques (ouverture du cycle aromatique)
Dioxygénases extradiol (bactéries)	Fe (II) Clivent en meta par rapport au groupe OH	Dégradation de composés aromatiques (ouverture du cycle aromatique)

Tableau 1 : Classification des différentes dioxygénases non-héminiques connues, avec leurs caractéristiques principales et les types de réactions qu'elles catalysent.

Afin d'essayer de comprendre quel pourrait être le rôle de ce gène chez les plantes sans bêtaïne, intéressons-nous aux dioxygénases connues. Elles font partie des oxygénases, qui sont une classe d'enzymes très répandues. Les oxygénases catalysent l'incorporation de O à partir d'O₂ dans un substrat organique. Elles utilisent souvent le fer comme cofacteur, qui appartient soit à un groupe hème, ou qui est sous une forme non-héminique.

Parmi les oxygénases les dioxygénases incorporent les deux atomes d'une molécule d'oxygène dans le substrat. On trouve différents types de dioxygénases, dont la classification selon (Prescott and John 1996) est indiquée dans le tableau 1. Elles sont présentes dans tous les types d'organismes vivants. Elles jouent aussi un rôle crucial dans la dégradation de composés aromatiques comme les drogues, les insecticides et les carcinogènes. Elles ont souvent pour substrat des cycles où deux carbones adjacents portent chacun un groupe hydroxyl.

Chez les bactéries, les dioxygénases sont impliquées dans des processus anaboliques et cataboliques. Parmi ces derniers, on note en particulier la dégradation de composés aromatiques (naturels ou artificiels) : hydroxylation et clivage du cycle aromatique de métabolites catécholiques qui possèdent des groupes hydroxyl sur 2 atomes de carbone adjacents. Les dioxygénases intradiol clivent en ortho par rapport au substituant hydroxyl et contiennent du Fe(III) non-héminique, alors que les dioxygénases extradiol clivent en meta par rapport aux substituants hydroxyl et dépendent de Fe(II) non-héminique. Les membres de ces deux classes d'enzymes ont des structures complètement différentes et utilisent des mécanismes distincts. (Eltis and Bolin 1996)

Chez les dioxygénases extradiol, après avoir considéré qu'il existait deux différentes classes, une troisième classe a été définie au vu de nouvelles séquences qui ne s'alignaient ni avec la classe I ni avec la classe II. Il n'y a pas de lien évolutif entre classe II et III, on pense donc que ce serait une évolution convergente (Spence, Kawamukai et al. 1996). Toutefois, des études spectroscopiques ont suggéré que les enzymes de classe III partagent beaucoup de similarités au niveau du site actif avec les enzymes de classe II (Arciero, Lipscomb et al. 1983; Arciero, Orville et al. 1985; Arciero and Lipscomb 1986).

Dans cette troisième classe, les ligands pour l'ion métallique sont deux histidines et un glutamate, et deux histidines du site actif sont conservées. D'autres résidus sont conservés dans cette séquence, et un consensus pour les séquences bactériennes a été défini avec le domaine LigB. La structure tridimensionnelle d'une de ces protéines bactériennes, la protocatechuate 4,5-dioxygénase de *Sphingomonas paucimobilis*, a été résolue (Sugimoto,

Senda et al. 1999). C'est une enzyme clé dans la voie de la dégradation de la lignine de *S. paucimobilis*.

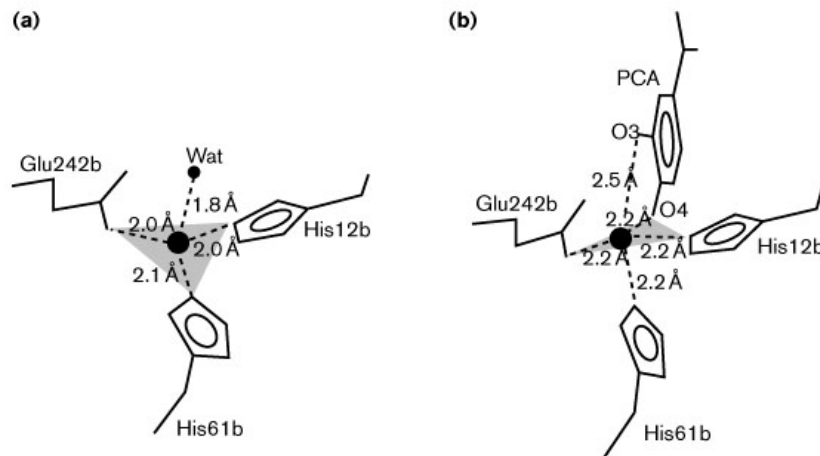


Figure 2 : Géométrie des sphères de coordination (a) sans substrat et (b) avec substrat. Tiré de l'article de Sugimoto et al.

Voici les informations que les auteurs ont tirées de cette structure : l'enzyme LigAB est un hétérodimère ($\alpha_2\beta_2$). Toutes les expériences de cet article ont été faites en conditions aérobiques, ce qui implique que le fer était sans doute oxydé dans un état non physiologique de Fe(III) (ferrique). C'est pour cette raison que les auteurs pensent qu'ils ont pu observer l'enzyme complexée avec un substrat intact (cristallisation avec et sans substrat). L'enzyme LigAB a été inactivée car le Fe a passé sous forme ferrique. Si les cristaux étaient redissous, ils ne montraient plus d'activité, qui était restaurée en ajoutant du Fe ferreux ou de l'ascorbate. Il y a une forte interaction entre les sous-unités α et β , moins forte entre les deux $\alpha\beta$.

Le site actif est situé dans un renforcement sur la partie supérieure de la sous-unité β . Cette poche est en grande partie couverte par la sous-unité α , ce qui fait que le site actif est protégé, environ à 15 Å sous la surface de l'hétérodimère $\alpha\beta$. Dans la forme sans substrat, le fer est directement coordonné par His12b, His61b et Glu242b. Une molécule d'eau peut s'insérer près du fer. La sphère de coordination peut être décrite comme une pyramide trigonale distordue avec la molécule d'eau en position axiale et trois ligands protéiques en positions équatoriales (voir fig. 2, tirée de l'article).

Dans la sphère de coordination du complexe LigAB-PCA, les auteurs supposent que les deux groupes hydroxyles du substrat se coordinent directement à l'ion Fe. La sphère de coordination peut être décrite comme une bipyramide tétragonale avec His12b, Glu242b et O4 du PCA qui sont les ligands équatoriaux, et His61b et O3 du PCA qui sont les ligands

axiaux (voir fig. 3, tirée de l'article). Il y a un site inoccupé qui pourrait être celui du dioxygène nécessaire pour la réaction catalytique.

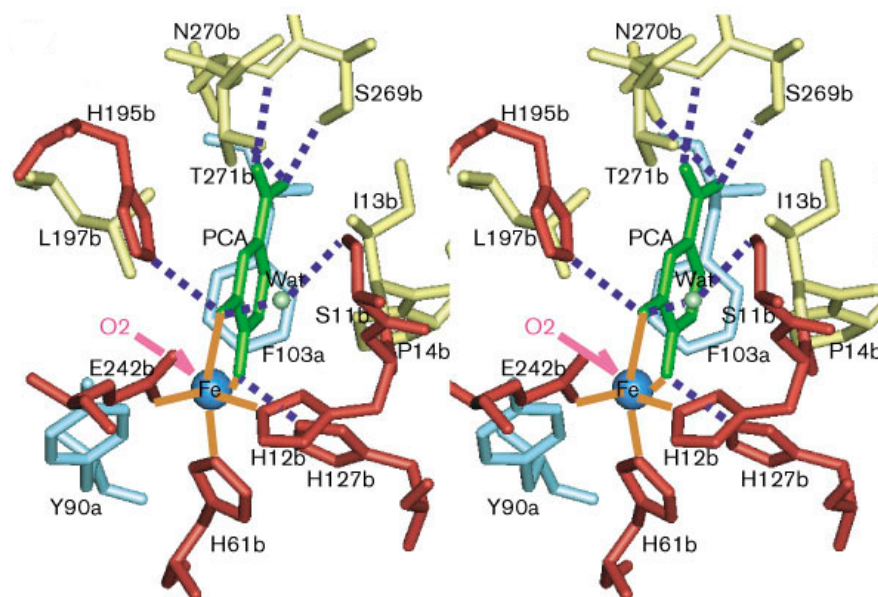


Figure 3 : Le site actif du complexe LigAB-PCA en stéréo. Les résidus de la sous-unité α sont en bleu. Le substrat PCA (en vert) fait six ponts hydrogène (en pointillés) avec les acides aminés environnants. Les liens de coordination avec le Fe sont en orange. Le site potentiel de liaison de l'O₂ (flèche rose) serait le quatrième ligand équatorial dans la bipyramide tétragonale de la sphère de coordination. Tiré de l'article de Sugimoto et al.

Les distances du fer avec ses ligands augmentent légèrement lors de la liaison du substrat. Deux hélices de la sous-unité α forment un couvercle qui ferme l'ouverture de la poche, ce qui fait que le PCA lié est complètement inaccessible au solvant. Cette inaccessibilité semble être essentielle pour l'activité de cette classe d'enzymes. Dans d'autres enzymes bactériennes similaires qui n'ont pas de sous unité α , une longue insertion de 42 à 44 acides aminés est souvent observée entre les résidus 203 et 204 (numérotation de LigAB). Cette insertion serait pliée et située au sommet de la cavité et aurait le même rôle que la sous-unité α chez LigAB. L'anneau aromatique du protocatéchuate est en contact avec plusieurs résidus hydrophobes, Ile13b, Pro14b, Leu197b, Phe103a, His195b et Thr271b. La Phe103a de l'unité α est l'obstacle qui empêche l'accès du solvant au C5 du PCA, site de l'attaque par un dioxygène activé.

Les O3 et O4 du PCA font des ponts hydrogène avec le N ϵ de l'His195b et le N ϵ de l'His127b. Des deux oxygènes du groupe carboxyl du PCA, un fait un pont hydrogène sur le N de la chaîne principale de l'Asn270b et l'autre plusieurs ponts hydrogène bifurqués sur le O γ de la Ser269b et du N δ de l'Asn270b. La présence du groupe carboxyl en C1 sur le

substrat semble être essentiel pour la catalyse par LigAB en jouant le rôle d'ancre via des forces de van der Waals et des ponts hydrogène, sans être suffisant pour que la catalyse ait lieu.

Lors de la liaison du substrat, le fer monte d'environ 1.4 Å, ce qui lui permet de se coordonner aux groupes hydroxyl du PCA.

Le dioxygène doit se lier à l'ion Fe pour être activé. La chaîne latérale de la Tyr90a se déplace lors de la liaison du PCA. Ce déplacement créerait un espace permettant à l'oxygène de venir se fixer. Toutefois, ce mécanisme n'apparaît pas chez d'autres enzymes ayant un site actif similaire.

En superposant les sites actifs de LigAB et BphC (extradiol, classe II), on s'aperçoit qu'il y a une grande similarité entre les deux. En particulier, dans les deux cas, le groupe hydroxyl attaché à une des extrémités de la liaison C-C à cliver est le liant équatorial au fer. Toujours en comparant LigAB avec BphC, on peut supposer que le N de l'imidazole de l'His195b serait la base catalytique.

Afin de faire une étude fonctionnelle comparée de ce gène de dioxygénase dans les plantes avec et sans bétalaïne, les expériences suivantes ont été prévues :

- Etude fonctionnelle du gène d'*Arabidopsis thaliana* homologue à celui de *Portulaca grandiflora* par complémentation biolistique sur des pétales blancs de *P.grandiflora*.
- Obtention de la séquence du gène de la dioxygénase L.6 chez d'autres plantes à bétalaïnes.
- Identification des spécificités fonctionnelles entre les gènes des plantes à bétalaïnes et les plantes sans bétalaïne par alignements multiples et modélisation tridimensionnelle.
- Knock-out du gène homologue chez la mousse *Physcomitrella patens*.

Résultats

Etude fonctionnelle du gène d'*Arabidopsis thaliana*

La première question que nous nous sommes posée était de savoir si le gène homologue à la DOD de *P. grandiflora* des plantes sans bétalaïne pouvait, dans des conditions adéquates, catalyser la réaction de conversion de la dihydroxyphénylalanine en acide bétalamique. Pour cela, nous avons utilisé la complémentation biochimique par biolistique sur des pétales blancs de *P. grandiflora* (Mueller, Hinz et al. 1997). La complémentation a déjà été montrée avec le gène DOD fongique et la DOD de *Portulaca grandiflora* nouvellement isolée par Christinet (non publié). Le protocole a été légèrement modifié : à la place du gène Gus comme contrôle positif, nous avons utilisé le gène DS-Red. La protéine codée par ce gène émet une fluorescence rouge quand elle est excitée avec une lumière verte.

Les observations se font 12 à 24 heures après le tir, et si le résultat est positif, on observe des cellules isolées qui ont synthétisé du pigment. Les expériences menées avec le gène *Pgdod* ont montré que dans tous les cas, une cellule transformée avec le gène DsRed2 est aussi transformée avec le gène DOD, ce qui limite le nombre de faux positifs pouvant éventuellement être observés, en raison de taches dues à la variégation (en général plusieurs cellules groupées), à des nécroses ou de débris de particules d'or que l'on peut confondre avec des cellules synthétisant des bétalaïnes. Les pétales blancs contiennent tous les précurseurs nécessaires à la synthèse des bétalaïnes, en particulier de la DOPA, mais la DOD n'y est pas exprimée en raison de la présence probable d'un transposon.

Nous avons inséré le gène d'*Arabidopsis thaliana* dans le vecteur d'expression pNco 35S, puis nous avons bombardé les pétales avec ce plasmide. L'expérience a été répétée trois fois, et à chaque fois aucune tache colorée sur les pétales n'a été observée. Le contrôle DS-Red était exprimé sur un nombre raisonnable de cellules, ce qui exclut un échec du bombardement - les billes d'or ont effectivement pu pénétrer dans les cellules et y être exprimées - et un mauvais état des pétales, qui étaient aptes à synthétiser des gènes.

On peut donc conclure de cette expérience que le gène d'*Arabidopsis thaliana* n'est pas capable de synthétiser des bétalaïnes, même placé dans un environnement où de la DOPA est disponible. On peut donc raisonnablement supposer que les gènes d'autres plantes sans bétalaïne sont dans le même cas. Le rôle de ce gène reste par ailleurs inconnu : la fonction de dioxygénase est certainement la même, mais la spécificité pour le substrat différente.

Obtention de séquences du gène de la DOD chez d'autres plantes à bétalaïnes

Très peu d'homologues de la DOD chez d'autres espèces de Caryophyllales sont disponibles sur les bases de données, étant donné qu'il n'existe pas de programme de séquençage de ces plantes. Afin de pouvoir comparer les séquences des plantes avec et sans bétalaïne, il était souhaitable d'obtenir d'autres séquences chez des plantes à bétalaïnes.

L'obtention d'une séquence à partir de cDNA est souvent fastidieuse : il faut déjà fabriquer ce dernier à partir de mRNA. Puis il faut connaître la séquence du début, faire des amplifications avec des amorces aspécifiques à la fin du gène. Il était impossible de se lancer dans un tel projet dans un délai aussi court.

Nous avons donc choisi de définir des primers « universels », c'est-à-dire choisis sur des régions ou les séquences des Caryophyllales que l'on connaît sont particulièrement conservées. Nous nous sommes basés sur les séquences de *Portulaca grandiflora*, *Beta vulgaris* et deux séquences de *Mesembryanthemum crystallinum*. Nous avons défini un primer sense et deux primers antisense, le primer As2 donnant le plus long fragment (520 bp contre 410 pour As1). Ces primers contenaient plusieurs bases dégénérées.

Il y a deux désavantages à cette méthode : d'une part, on ne peut obtenir qu'une partie du gène, car le début et la fin ne sont pas très bien conservés. D'autre part, les séquences obtenues sont génomiques, c'est-à-dire que les introns vont aussi être amplifiés, clonés et séquencés. En cas de longs introns, il faut imaginer que plusieurs séquençages pourraient être nécessaires pour obtenir la séquence codante. Néanmoins, cette méthode est beaucoup plus rapide que celle passant par le cDNA. D'ailleurs, cela peut constituer une base si l'on veut par la suite isoler le cDNA d'une de ces plantes.

Les plantes suivantes ont été choisies pour être analysées :

- *Beta vulgaris*, cultivar différent que celui dont la séquence est connue
- *Silene sp.*
- *Dianthus caryophyllus*
- *Phytolacca americana*
- *Celosia sp.*
- *Mammillaria sp.*
- *Iresine sp.*
- *Mesembryanthemum crystallinum*

- *Amaranthus sp.*
- *Bougainvillea sp.*

Elles représentent un choix représentatif de la classification des Caryophyllales selon Mabry (Clement and Mabry 1996)

Les PCR sur *Beta vulgaris*, *Mesembryanthemum* et *Mammillaria* ont donné une seule bande dans les conditions standard. Il a été facile de purifier ce fragment et de le cloner dans un vecteur pGem.

Pour les autres espèces, l'amplification a été plus difficile. *Silene*, *Dianthus caryophyllus*, *Phytolacca americana* et *Iresine* ont donné plusieurs bandes. Au moment de la purification sur gel, il était impossible de choisir la bande qui pouvait correspondre au gène. En règle générale, c'est la bande la plus intense qui a été choisie, et quand deux bandes étaient fortes, elles ont été prises les deux. La taille des fragments était elle-même imprévisible, étant donné qu'on ne connaît pas la taille des introns, qui paraît varier passablement d'une espèce à l'autre. Comme les deux combinaisons d'amorces ont à chaque fois été testées, si on observait une différence d'environ 100 entre les deux bandes correspondant à chaque combinaison, on a pu supposer que cela pouvait être les bonnes. Mais cela restait très approximatif. Les conditions d'annealing de la PCR, ainsi que les concentrations de chlorure de magnésium ont été modifiées pour optimiser l'amplification. Les fragments de DNA isolés chez *Silene* et *Dianthus* que l'on a fait séquencer donnent des séquences ne présentant aucune homologie avec le gène de la dioxygénase. Il s'agit d'une amplification aspécifique. *Silene* et *Dianthus* sont des Caryophyllacées ne produisant pas de bétalaïnes, il était donc prévisible que cela ne marche pas. Mais cela nous aurait donné une indication sur la séquence du gène de plantes entre les plantes à bétalaïnes et les plantes phylogénétiquement proches, mais sans bétalaïne.

Celosia donnait plusieurs bandes très faibles qu'il n'a pas été possible de cloner. De même pour *Amaranthus* et *Bougainvillea*.

On a également testé la PCR sur du DNA de *Physcomitrella patens*, pour voir si nos primers pouvaient également amplifier le gène de plantes sans bétalaïne. Cela n'a pas fonctionné.

Il est intéressant de noter que ces primers, basés en grande partie sur la séquence de *P. grandiflora*, semblent être des « indicateurs phylogénétiques », en ce sens qu'ils n'amplifient efficacement que les séquences proches de *P. grandiflora*, et de plus en plus mal ou pas du tout à mesure que l'on s'éloigne.

Finalement, nous avons pu obtenir les séquences de *Beta vulgaris*, *Phytolacca americana*, *Iresine* et *Mammillaria*. Apparemment, il y a un intron de taille variable entre les 2 exons où nous avons défini les amorces. Par chance, un séquençage de chaque côté du fragment a suffi à avoir la séquence complète entre les deux primers, avec quelques nucléotides superposés entre les deux séquences.

La séquence de *Beta vulgaris* était bien la même que celle de l'autre cultivar que nous avons déjà.

Différences fonctionnelles entre le gène de synthèse des bétalaïnes et ses homologues

Un alignement avec un choix significatif de séquences de plantes a été réalisé. Les séquences nucléotidiques suivantes ont été choisies : *Portulaca grandiflora*, *Beta vulgaris*, *Iresine sp.*, *Mammillaria sp.*, *Phytolacca americana* et *Physcomitrella patens* (nos propres séquences, voir annexe 1) et *Mesembryanthemum crystallinum* (2 séquences, 11551280 et 8578568), *Arabidopsis thaliana* (AY050948), *Glycine max* (soja, 9899215), *Medicago truncatula* (11910252), *Lotus japonicus* (7740336), ces trois dernières étant des Légumineuses, *Lycopersicon esculentum* (tomate, 5602563), *Zea mays* (maïs, BZ405169), *Oryza sativa* (riz, PO471B04), *Triticum aestivum* (blé, 11117560), *Pseudomonas putida* (NP_744024, séquence protéique) et *Sphingomonas paucimobilis* (AAA17728, séquence protéique), cette dernière ayant été cristallisée et sa structure tridimensionnelle résolue.

Pour certaines espèces, plusieurs EST contenant des fragments du gène ont été combinés.

La protéine de *Pseudomonas putida* et ses homologues chez plusieurs autres bactéries représentent une nouvelle famille de protéines bactériennes très similaires à celles de plantes (environ 35% d'identité et 54% de similarité), ce qui est plus élevé que la protéine cristallisée de *Sphingomonas paucimobilis*. Toutefois, aucune de ces protéines bactériennes nouvellement définies n'a encore été cristallisée et ne peut donc pas nous donner des informations quant à la structure tridimensionnelle.

Il est difficile de savoir combien de copies d'homologues à L.6 existent par espèce, comme on dispose rarement de la séquence complète du génome d'une espèce. Pour les espèces *Oryza sativa*, *Lycopersicon esculentum*, *Solanum tuberosum*, *Mesembryanthemum crystallinum* il y a clairement 2 gènes présentant, chez une espèce, environ 70% d'homologie. Chez *Arabidopsis thaliana*, on ne trouve qu'un homologue sur le chromosome 4 parmi les séquences connues. Un petit fragment du gène L.6 semble s'être dupliqué sur le chromosome 3.

Sur l'alignement en annexe 2, on peut voir, sur la première ligne, un index de A à Z, pour indiquer précisément les positions conservées ou importantes. En dessous, les H en rouge indiquent les régions correspondant aux hélices α de la structure de *Sphingomonas paucimobilis* telle que décrites par (Sugimoto, Senda et al. 1999). De même, S en vert pour les feuillet β . En dessous, la séquence de la sous-unité β de LigAB, superposée à celle de *Portulaca grandiflora*. En dessous de cette dernière se trouvent les prédictions de structure secondaire réalisées par le serveur de PredictProtein. Sur la première ligne, les prédictions de PROF: Profile network prediction Heidelberg (Rost and Sander 1993). Sur la deuxième ligne, une sélection parmi ces mêmes prédictions avec une précision de plus de 82%. On trouve ensuite toutes les séquences précédemment citées, alignées. Une prédiction de la structure secondaire a également été faite sur la séquence de *Physcomitrella patens*, et se trouve en bas de la figure.

Les prédictions de structure secondaire faites sur les séquences de *Portulaca grandiflora* et *Physcomitrella patens* correspondent en général bien, sur l'alignement avec la séquence de *Sphingomonas paucimobilis*, avec les structures secondaires réelles de cette dernière. Ne correspondent pas : l'hélice sur la séquence de *S. paucimobilis* entre 20 et 30 (numérotation de l'alignement) non prédite sur les deux autres, une hélice sur ces dernières aux alentours de 100 inexistante sur *S. paucimobilis* et juste après un petit feuillet β non prédit sur les deux séquences, une hélice à 210 inexistante sur *S. paucimobilis*. Sinon, toutes les autres structures ont été prédites, quelques décalages mis à part.

Cette bonne correspondance entre les structures observées sur la protéine de *S. paucimobilis* et celles prédites sur les séquences de *Portulaca grandiflora* et *Physcomitrella patens* justifient le fait d'utiliser la première comme "template" pour modéliser la structure tridimensionnelle des deux autres, malgré la faible homologie de séquence (env. 20%).

Il était intéressant de regarder où se trouvent, sur la structure tridimensionnelle, les résidus conservés entre la protéine bactérienne et les séquences de plantes. Certains, comme A, F, G, I, J, K, L, O, P, Q, R, U, V se trouvent soit à l'intérieur de la protéine, ou loin du site actif. Ils ne sont donc certainement pas impliqués dans le mécanisme de catalyse, mais importants pour des contraintes structurales.

D'autres se trouvent à proximité du site actif. L'histidine en B lie le résidu Fe^{2+} , elle est conservée chez toutes les espèces. En C et D, l'isoleucine et la proline font des interactions hydrophobes avec l'anneau aromatique du PCA. On ne retrouve pas l'Ile chez les plantes, par

contre une proline conservée est à proximité, mais décalée de un résidu, sans compter le gap introduit par l'alignement (mauvais à cet endroit là).

En E, il y a une glycine conservée chez toutes les espèces. Elle ne semble pas interagir directement avec le substrat, mais se trouve proche de l'histidine en B et pourrait être importante pour l'orientation de cette chaîne, en empêchant un encombrement stérique.

En H se trouve la deuxième Histidine interagissant avec le Fe^{2+} , qui est également conservée chez toutes les espèces.

L'histidine en N fait des ponts hydrogène avec le substrat (mentionné par les auteurs). L'aspartate en M n'est pas mentionné, mais il pourrait également avoir une importance vue sa conservation et sa proximité avec l'histidine.

L'histidine en S et la leucine en T font des interactions hydrophobes avec l'anneau aromatique du PCA. Pour l'histidine, conservée, il pourrait donc en être de même avec les substrats inconnus des autres enzymes. La leucine est en général conservée chez les plantes sans bétalaïne.

Chez les plantes, aux alentours de 220, on observe une insertion par rapport à la séquence bactérienne. Or les auteurs de la structure tridimensionnelle ont observé qu'une insertion existe à cet endroit là chez des bactéries qui n'ont pas de sous-unité α (recouvre le site actif), cette insertion étant toutefois beaucoup plus grande que celle observée chez les plantes. Il est toutefois possible que cette insertion joue également le rôle de « couvercle » sur le site actif. Juste après l'histidine en S, on observe un motif très conservé chez les plantes sans bétalaïne, NLR. Chez les plantes à bétalaïnes, il y a PxxxTP. Ce dernier consensus a été défini à partir de seulement quatre séquences de trois espèces différentes, les séquences d'Iresine, Mammillaria, Phytolacca s'arrêtant avant, leur séquence ayant été obtenue avec les primers dégénérés précédemment cités. Ces patterns différents entre les plantes avec et sans bétalaïne pourraient expliquer la différence de substrat entre les deux groupes d'enzymes. Nous le verrons plus bas sur les modélisations tridimensionnelles.

Le glutamate en W est, chez *S. paucimobilis*, le troisième ligand du Fe^{2+} . Il ne semble pas être conservé chez les plantes. Par contre, on trouve chez ces dernières une histidine strictement conservée, superposée au glutamate sur l'alignement, si celui-ci est correct à cet endroit. Il est possible que le fer soit coordonné par trois histidines, comme cela a déjà été observé chez une lipoxigénase de soja (Boyington, Gaffney et al. 1993). Ou alors, le résidu se trouvant juste avant, qui est soit un glutamate ou un aspartate, pourrait jouer ce rôle. Il y a une exception dans une des séquences de *Mesembryanthemum crystallinum*, mais il se pourrait que ce soit un homologue ayant encore un autre rôle. En effet, on a observé chez certaines

plantes plusieurs homologues au clone L.6. présentant parfois de grandes différences entre eux. Qu'il s'agisse de l'histidine ou de l'aspartate ou glutamate, il y a un résidu à cet endroit qui peut jouer le rôle de troisième ligand pour le fer. Nous avons admis dans la modélisation que c'était l'histidine.

Les trois derniers résidus en X, Y et Z, qui sont, chez *S. paucimobilis* une sérine, asparagine et thréonine font des ponts hydrogène avec le groupe carboxylique du PCA. Ils ne se retrouvent pas chez les plantes. Il n'y a aucun résidu conservé chez les plantes qui pourrait remplacer ces derniers.

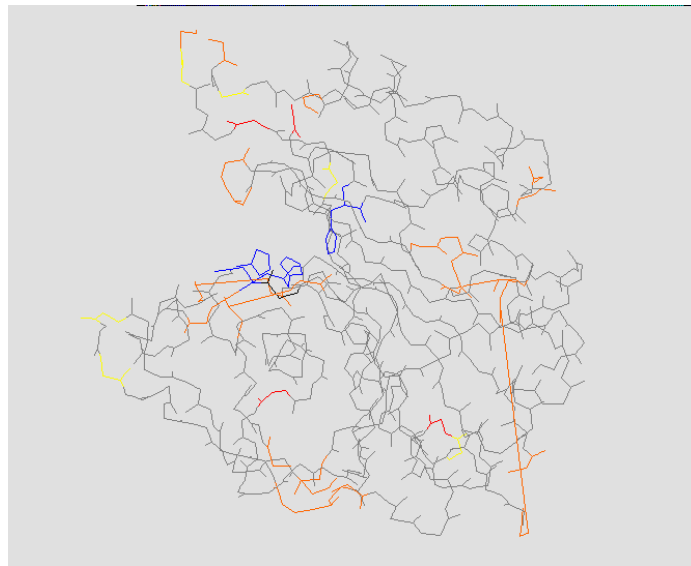


Figure 4 : Vue globale du modèle de la DOPA-dioxygénase de *Portulaca grandiflora*, coloré selon les problèmes de la protéine : rouge pour les prolines ayant un angle de torsion incorrect pour une proline, jaune pour les autres résidus ; orange pour des résidus internes sans ponts H. En rose, les chaînes latérales faisant des « collisions ». En bleu, les trois histidines liant le Fe.

Des modèles tridimensionnels des protéines de *Portulaca grandiflora* et *Physcomitrella patens* ont été créés en utilisant comme "template" la protéine de *S. paucimobilis*, entrée PDB 1B4U, en utilisant le logiciel Swiss-PdbViewer (Guex and Peitsch 1997). Les séquences à modéliser ont été alignées sur la structure connue selon le même alignement que cité plus haut (annexe 2). En raison d'insertions de tailles importantes entre la séquence de *S. paucimobilis* et celles des plantes, des erreurs grossières ont été générées, le programme de modélisation n'ayant pas d'autre choix, lors d'un tel cas, que d'allonger artificiellement la longueur de la liaison peptidique. Par exemple, entre 180 et 190 environ, une hélice a été supprimée du modèle, et remplacée par une « chaîne carbonée » de la même taille. Toutefois, nous voyons que l'hélice semble également exister dans nos deux protéines à modéliser, mais juste avant. Sans autres informations, il est impossible de la modéliser. Si ces erreurs étaient loin du site

actif, elles n'ont pas été corrigées, même si elles étaient grossières. Sur la figure 4, on peut voir une vue globale de la chaîne carbonée du modèle de *Portulaca grandiflora* colorée selon les problèmes sur la protéine : rouge pour les prolines ayant un angle de torsion incorrect pour une proline, jaune pour les autres résidus ; orange pour des résidus internes sans ponts H. En rose, les chaînes latérales faisant des « collisions ». On peut voir qu'il y a pas mal de problème. Les trois histidines liant le fer ont été colorées en bleu clair, afin de localiser le site actif. On observe notamment un grand trait en bas à droite, correspondant à l'hélice manquante précédemment citée. Comme elle est loin du site actif, on suppose que ce n'est pas important pour les observations que l'on désire faire ici.

Nous sommes partis de l'hypothèse que l'alignement aux sites importants est suffisamment bon pour que les résidus proches du site actif soient correctement placés. De toute façon, une homologie aussi faible ne permet pas d'étudier les mécanismes précis de l'action catalytique, mais seulement de se faire une idée sur la structure générale du site actif, et d'émettre des hypothèses, notamment pour définir des sites de mutagenèse, permettant de les vérifier.

Il est relativement difficile de distinguer quelque chose sur une image tridimensionnelle contenant tous les résidus d'une protéine. Nous avons donc construit, à partir des trois modèles de protéines, des structures où n'ont été gardés que les résidus proches du site actif, pouvant avoir une fonction catalytique ou une interaction avec le substrat.

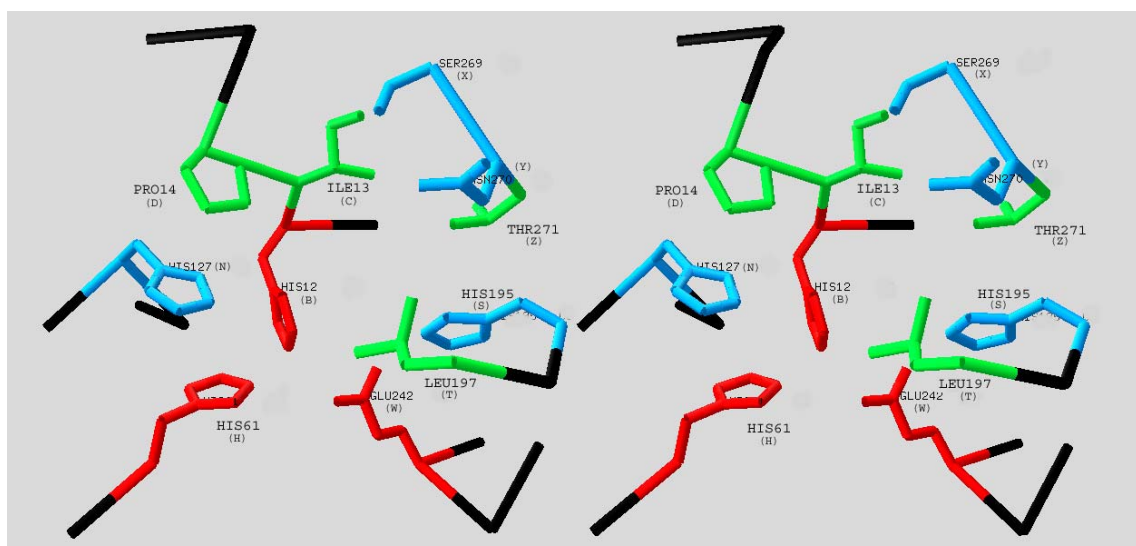


Figure 5 : Vue du site actif de la protéine 1B4U construite à partir du fichier PDB. En rouge, les histidines liant le Fe (sur l'alignement : A, H et W), en bleu les acides aminés faisant des ponts hydrogène avec le substrat (N, S, X, Y), en vert les acides aminés faisant des interactions hydrophobes avec le substrat (C, D, S, T, Z). Les numérotations sont celles de la séquence 1B4U.

Sur la figure 5, on voit une image stéréo construite à l'aide de spdbviewer, à partir de la structure tridimensionnelle 1B4U de *Pseudomonas*, avec les résidus importants : en rouge, les trois résidus liant le fer (sur l'alignement : A, H et W), en bleu, les résidus faisant des ponts H avec le PCA (N, S, X, Y) et en vert les résidus faisant des liaisons hydrophobes avec l'anneau aromatique du PCA (C, D, S, T, Z).

Les résidus précédents qui étaient conservés sur les séquences de *Physcomitrella patens* et *Portulaca grandiflora* (figures 6 et 7, respectivement) ont été indiqués par les mêmes couleurs, ceux qui pourraient avoir une importance pour le mécanisme catalytique ont été colorés en rose. L'histidine située à 262 sur l'alignement a été colorée en jaune pour des raisons de visibilité. Elle est conservée chez toutes les séquences de plantes, étant remplacée par une glycine chez la bactérie. Suivant l'orientation de sa chaîne latérale, elle pourrait interagir avec le substrat. Comme elle faisait une collision avec une chaîne voisine, un meilleur rotamère a été choisi.

L'angle de vue des sites actifs des protéines de plantes est différent, également pour des raisons de visibilité. Sur la figure de *Physcomitrella patens*, on voit les trois histidines liant le fer, qui se trouve au milieu (non représenté). Les deux histidines en bleu feraient des ponts hydrogène avec le substrat inconnu. La leucine (T) est conservée chez les plantes sans bétalaïne et chez la bactérie, elle a donc été colorée en vert (fait des interactions hydrophobes avec l'anneau aromatique). Juste avant se trouve une asparagine, conservée chez toutes les plantes sans bétalaïne. Après la leucine se trouve une arginine également conservée. Ces résidus pourraient interagir avec le substrat et être en partie responsable pour la spécificité du substrat, inconnu, de cette enzyme des plantes sans bétalaïne.

En haut à droite, on observe une insertion de cinq acides aminés qui n'existe pas dans la protéine bactérienne. Pour la modéliser, nous avons utilisé une fonction de spdbviewer qui permet de chercher, selon la séquence, dans une base de donnée de boucles déjà observées. Dans notre cas, aucune des solutions proposées n'était satisfaisante, nous avons choisi la « moins mauvaise ». Par ailleurs, il n'y a aucune indication pour faire un modèle valable. Notamment, l'orientation de cette boucle pourrait être complètement différente et recouvrir le site actif comme observé chez les bactéries si l'on changeait les angles de la structure carbonée. Les distances des chaînes latérales en rose et vert par rapport au substrat pourraient être plus petites que sur le modèle. Si on colore les résidus de cette boucle selon les problèmes, comme sur la figure 4, plusieurs acides aminés présentent des angles de torsion incorrects, dans les deux modèles.

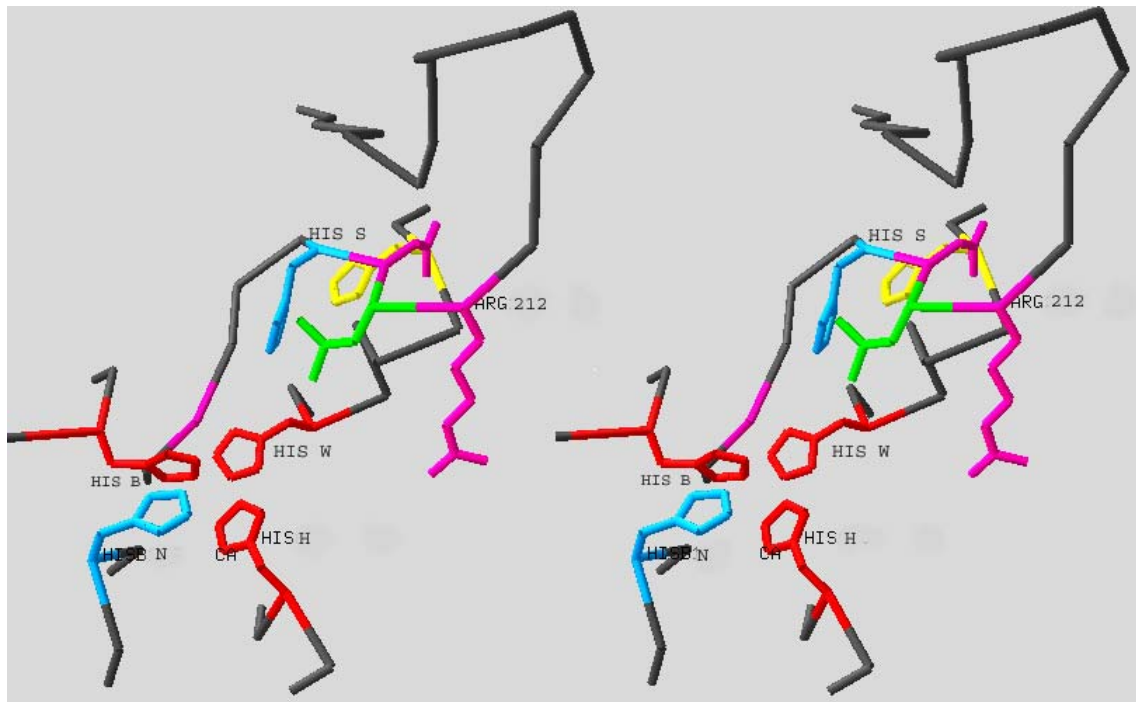


Figure 6 : Modélisation en stéréo du site actif de la dioxygénase « L.6 » de *Physcomitrella patens*, avec en rouge les trois histidines liant potentiellement le fer (B, H, W), en bleu les deux histidines qui feraient des ponts H avec le substrat (N, S), et en rose et vert, sur la grande boucle, asparagine, leucine et arginine (les trois après S), qui pourraient ancrer le substrat. En jaune l’histidine qui pourrait éventuellement interagir avec le substrat (262 sur l’alignement).

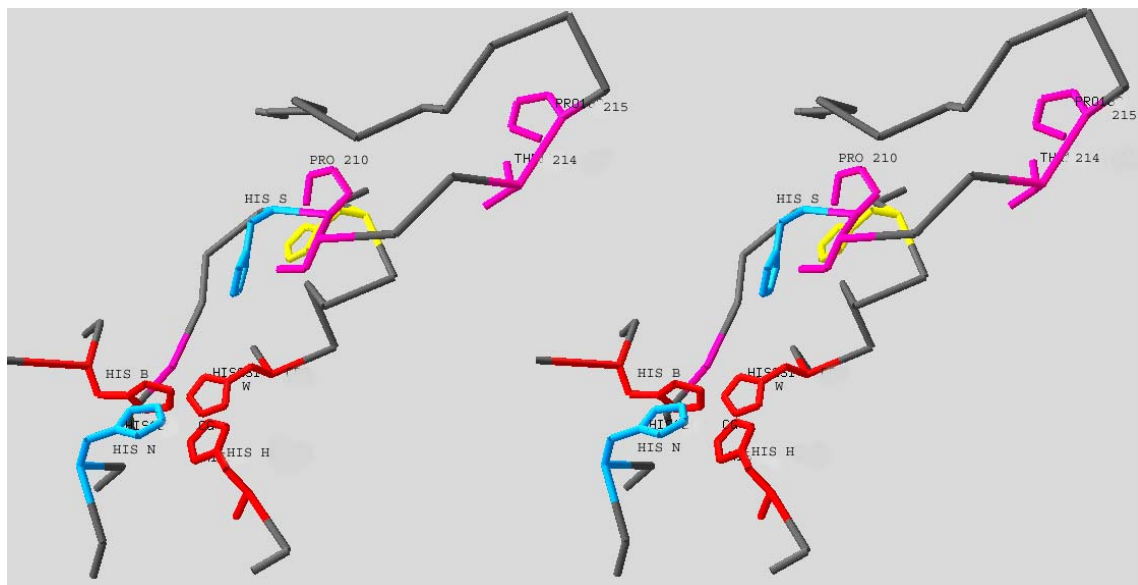


Figure 7 : Modélisation du site actif de la DOPA-dioxygénase de *Portulaca grandiflora*, avec en rouge les trois histidines liant potentiellement le fer (B, H, W), en bleu les deux histidines qui feraient des ponts H avec le substrat (N, S), et en rose, sur la grande boucle, proline, asparagine ou serine, puis plus haut tyrosine et proline (après S), qui pourraient ancrer le substrat. En jaune l’histidine qui pourrait éventuellement interagir avec le substrat (262 sur l’alignement).

Il est intéressant de noter que la glycine en 203, conservée chez toutes les espèces et indiquée en rose sur la figure (derrière les histidines), passe tout près du site de liaison du fer. Si le résidu à cet endroit là avait une chaîne latérale, cela pourrait perturber la coordination du fer. Le fait que cette glycine se trouve à cet endroit valide la modélisation faite à partir de l'alignement. On observe la même chose sur la figure 7, modélisation de la protéine de *Portulaca grandiflora*, à l'exception de la boucle qui contient des résidus différents, avec également un consensus : après l'histidine conservée (S), proline, puis, quatre résidus plus loin, thréonine et proline. Là aussi, l'orientation de la boucle est certainement fautive, et la proline et l'arginine en haut à droite pourraient se trouver sur le site actif, et interagir avec le substrat. Elles se retrouvent en tous cas chez les 4 séquences connues de plantes à bétalaïnes.

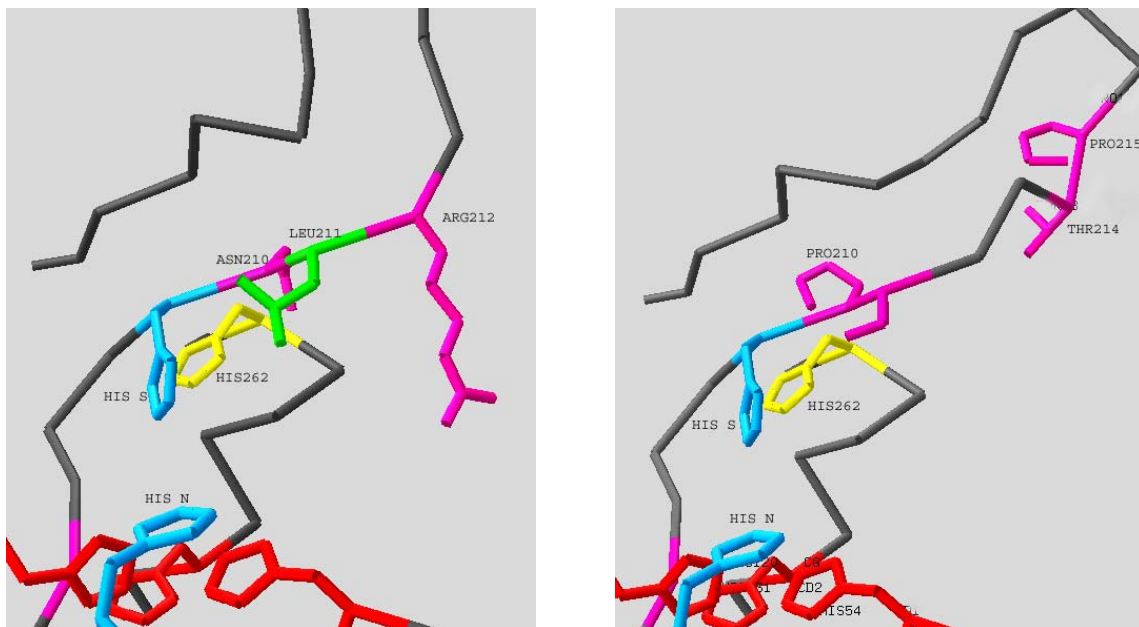


Figure 8 : Détails des sites actifs modélisés de *Physcomitrella patens* et *Portulaca grandiflora*, sous un autre angle de vue que précédemment. Lorsque la position n'était pas indiquée par une lettre, la numérotation selon l'alignement a été indiquée.

Sur la figure 8, on voit des vues de détail des 2 modèles, sous un autre angle que précédemment, avec la boucle et ses résidus spécifiques.

Il est difficile de faire des prédictions quant aux éventuelles interactions du substrat inconnu des plantes sans bétalaïne avec la dioxygénase modélisée. Par contre, pour la DOPA-dioxygénase de *P. grandiflora*, on peut faire des suppositions, si l'orientation du substrat est également conservée par rapport à la dioxygénase bactérienne. Si l'on compare les deux substrats (fig. 9), on peut voir que la chaîne de la DOPA est plus longue d'un carbone, et

contient un groupe $-NH_3$, pouvant avoir une charge positive. Les résidus de l'enzyme interagissant avec le $-COOH$ ou le $-NH_3$ peuvent donc être légèrement plus éloignés du fer, comme ce serait le cas sur le modèle. La thréonine et la proline en 214 et 215 sur l'alignement pourraient orienter le substrat par des ponts H (thréonine) ou par interaction hydrophobe (proline). De même pour la proline juste après l'histidine (210) et la sérine, apolaire ou alanine, polaire juste après. Ces deux dernières n'ayant pas les mêmes propriétés, on peut se demander si elles jouent réellement le rôle de ligand du substrat à cet endroit. De même pour l'asparagine ou l'aspartate juste après, qui n'ont pas été indiqués sur le modèle.

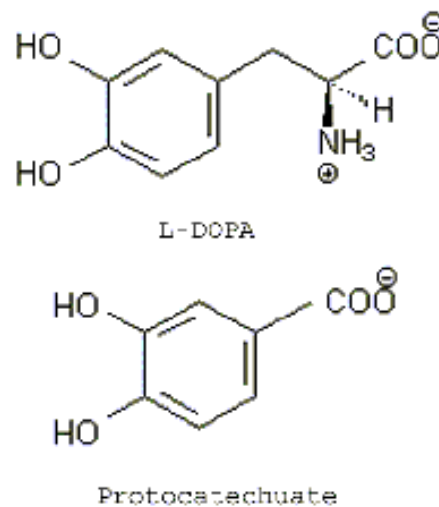


Figure 9 : Comparaison entre la L-DOPA et le protocatechuate. La chaîne de la L-DOPA est plus longue d'un carbone et contient un groupe $-NH_3$.

Si l'on se reporte à une étude faite sur la DOPA-dioxygénase d'*Amanita muscaria* purifiée avec laquelle différents substrats ont été testés, il apparaît de ces études que le groupe amino est nécessaire pour une bonne activité de clivage. Au contraire, le groupe $-COOH$ seul serait plutôt inhibiteur. La chaîne latérale du substrat positionnerait donc ce dernier correctement sur le site actif et pourrait également moduler la réaction via des effets inductifs à travers l'anneau aromatique. La séquence d'*Amanita* n'ayant aucune similitude avec celle de *Portulaca*, nous ne peut pas directement appliquer ces conclusions à notre modèle. Toutefois, le produit étant le même et la complémentation de *Amdod* dans les pétales de *Portulaca* étant possible, il y a une probabilité pour que le mécanisme catalytique soit identique.

Knock-out du gène chez *Physcomitrella patens*

Afin d'étudier la fonction de l'homologue à L.6 chez des plantes sans bétalaïne, un knock-out de ce gène a été fait chez la mousse *Physcomitrella patens*, seule plante connue à ce jour à faire des recombinaisons homologues de DNA (Schaefer and Zryd 1997). Un plasmide contenant un fragment du DNA génomique, avec introns, a été intégré dans des protoplastes de mousse. Dans ce système, le fragment homologue du type sauvage va être remplacé par le DNA homologue du plasmide qui contient un gène de résistance et qui va s'intégrer au milieu du gène ciblé, l'inactivant par la même. Le schéma d'une des insertions possibles est montré en figure 10. Il est fréquent que le plasmide s'intègre plusieurs fois. Nous avons fait un schéma avec 2 intégrations du plasmide au milieu du gène de type sauvage.

Après transformation et deux sélections successives sur milieu contenant l'antibiotique G418, une vingtaine de clones résistants ont été isolés. Aucun d'entre eux ne montraient de phénotype particulier à première vue.

Plusieurs faux positifs peuvent être observés. Premièrement, le plasmide peut ne pas s'intégrer mais rester cytoplasmique, tout en conférant la résistance à l'antibiotique. Les sélections successives sont là pour « diluer » les éventuels plasmides. Deuxièmement, le DNA peut s'intégrer de manière non homologue, conférant également la résistance sans inactiver le gène. Il se peut également qu'il y ait plusieurs copies du gène chez *Physcomitrella patens*, et que l'on ne les ait pas toutes inactivées. Le Southern blot sur ce gène n'ayant pas encore été fait, on ne sait pas encore si ce gène existe en copie unique chez cette mousse. Une amplification par PCR du fragment génomique semble indiquer qu'il n'y a qu'une copie, mais cela n'est pas une preuve.

Pour vérifier si l'intégration a effectivement eu lieu au bon endroit, le moyen le plus fiable est le Southern blot. En comparant avec du DNA de type sauvage, on peut être sûr que le clone étudié est vraiment un knock-out pour le gène en question. Pour des questions de temps et de quantité de matériel végétal, nous n'avons pas eu le temps de faire cette expérience.

En lieu et place, nous avons testé une nouvelle méthode de détection par PCR. Le principe est d'utiliser un primer sur le gène sauvage, non remplacé après la délétion, et un autre sur le DNA inséré, sur le gène de résistance par exemple. Ainsi, seuls les clones positifs montreront des fragments amplifiés par PCR. Plusieurs essais ont été réalisés, et l'un d'entre eux, qui montrait quelques amplifications, est montré sur la figure 11. Les résultats ne correspondent pas à nos attentes, qui seraient environ 1 kb pour p1-T3, 2 kb pour T7-p4. Ce seraient les seules amplifications au cas où une seule copie du plasmide se serait intégrée. Si plusieurs

copies s'intègrent, la combinaison T3-T7, antisenses sur le plasmide, peut donner une amplification, d'environ 2,5 kb dans le cas de deux insertions. Si le gène n'a pas été ciblé, p1-p4 devrait donner une bande d'environ 1,3 kb.

Or sur le gel de la PCR, on n'observe des bandes que chez le clone 1 :

- T3-T7 (4) une bande à 200 une à 1,5 et une à env. 3,5 kb (difficile à voir sur l'image numérisée)
- p1-p4 (5) bandes à 600, 1100, 1300, 2500
- T3-p4 (6) bandes à 500, 750, 1100
- p4 (8) bandes à 500, 750, 1100

On peut voir qu'apparemment le primer p4 peut générer, seul, des amplifications aspécifiques comme en 8, mais pas pour chaque échantillon. On retrouve ces mêmes bandes en 6. L'amplification en 5, qui ne devrait avoir lieu que si le clone 1 n'est pas un knock-out pour L.6, est certainement due à p4. S'il s'agissait d'une amplification spécifique, on devrait la retrouver chez le type sauvage. Les fragments amplifiés par T3-T7 pourraient indiquer une insertion multiple chez ce clone, toutefois les tailles ne correspondent pas à celles attendues.

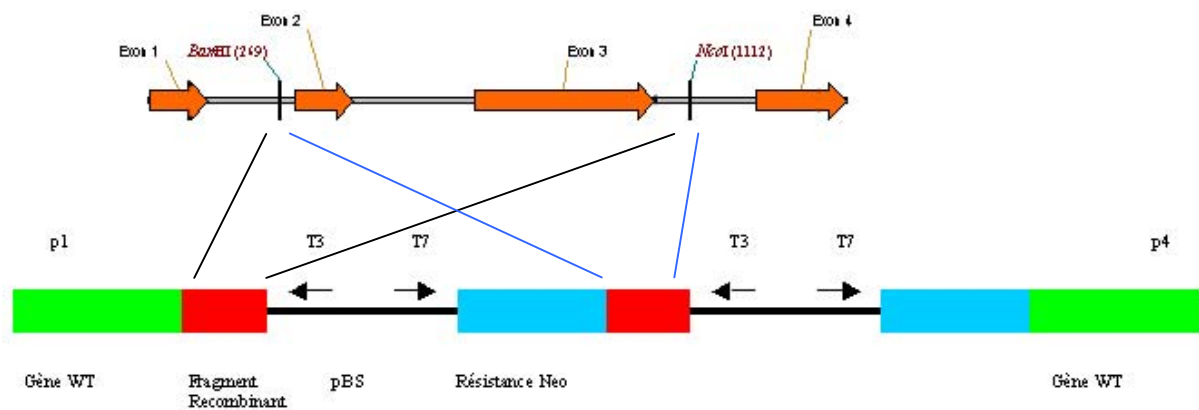


Figure 10 : Schéma de la recombinaison homologue chez *Physcomitrella patens* avec le gène L.6. Le gène de type sauvage (WT) est représenté en vert. Le fragment en rouge est le gDNA inséré dans le plasmide d'insertion, qui va remplacer le DNA de type sauvage. Au dessus, le schéma du gène avec les 2 sites de restriction BamHI(269) et NcoI(1112) générant le fragment inséré dans pKO. La séquence du plasmide bluescript est représentée en noir. En bleu le gène de résistance au G418. Il est très fréquent de trouver des insertions multiples des plasmides, c'est pourquoi la séquence du plasmide pKO ori1 Pp gDNA a été indiquée deux fois. En dessus, les emplacements des primers pour la PCR ont été indiqués.

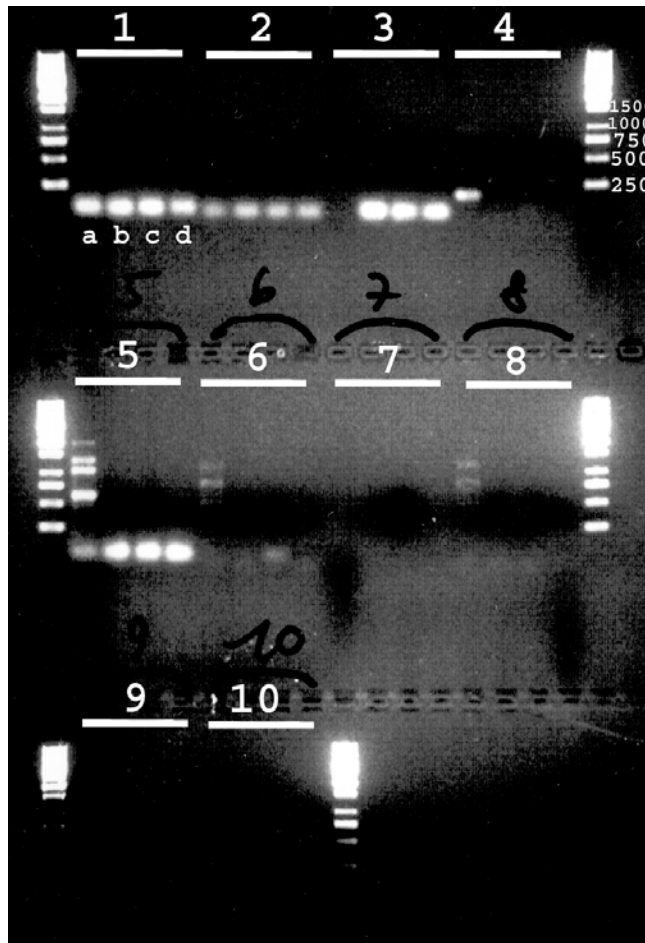


Figure 11 : Photographie du gel de migration des produits PCR avec les différentes combinaisons d'amorces indiqués sur la figure 10. Les différents DNA testés sont : clone 1 (a), 2 (b), souche sauvage (c). Un contrôle sans DNA a été réalisé pour chaque combinaison d'amorces (d). Pour chaque combinaison d'amorces, les DNA ont été chargés dans cet ordre. Les différentes combinaisons sont : 1 :T7-p4 ; 2 : p1-T3 ; 3 : p1-T7 ; 4 : T3-T7 ; 5 : p1-p4 ; 6 : T3-p4 ; 7 : p1 ; 8 : p4 ; 9 : T3 ; 10 : T7.

Il faudrait prendre une autre amorce que p4 qui, dans ces conditions, peut faire des amplifications a-spécifiques. La méthode de préparation du DNA (simple lyse des mousses dans un tampon et utilisation du lysat directement comme "template") ne semble pas convenir pour ce choix d'amorce.

Ainsi nous ne pouvons pas affirmer qu'un knock-out de L.6 a vraiment été produit. Toutefois, sur 20 clones résistants, il y a de fortes probabilités pour qu'il y en ait au moins un. Des vérifications par Southern blot nous le diront.

Conclusions et perspectives

Dans les expériences ci-dessus, nous avons pu montrer que la L.6-dioxygénase d'*Arabidopsis* avait une spécificité différente de celle de *Portulaca*. Cela permet de faire des hypothèses quant à l'évolution de ce gène. Les séquences bactériennes très similaires, représentées dans l'alignement par *Pseudomonas putida*, montrent que ce gène est largement conservé chez des espèces très éloignées. Sa haute conservation indique qu'il doit avoir une fonction importante. Si l'hypothèse émise par Ehrendorfer (1976) est correcte, on peut la compléter et imaginer que ce gène a muté lors de l'apparition des plantes à bétalaïnes, et que sa spécificité s'est modifiée en devenant capable de cliver la DOPA. Il pourrait s'agir d'une duplication de gène, puisque deux copies du gène ont été observées chez certaines plantes avec et sans bétalaïne.

Les trois nouvelles séquences de plantes à bétalaïnes obtenues dans ce travail nous ont permis de vérifier la conservation de certains sites importants, mais leur suite manque cruellement à l'endroit de l'insertion, à proximité du site actif. Il sera facile, en connaissant la séquence sur 500 paires de bases au milieu du gène, d'isoler et de séquencer le reste du gène de ces plantes pour voir si le consensus défini est vraiment valable, et pour définir des expériences de mutagenèse. On pourrait mutagéniser les trois ou quatre acides aminés prédits d'être à proximité du site actif sur la séquence de *Portulaca*, et voir si l'on inactive l'enzyme. Parallèlement, on pourrait introduire le consensus des plantes à bétalaïnes dans la séquence de *Physcomitrella* ou d'*Arabidopsis* et voir si l'enzyme devient capable de cliver la DOPA. Pour tester une telle mutation, le seul test à disposition pour l'instant est la complémentation par biolistique. C'est une méthode rapide et efficace qui donne des résultats 24h après le bombardement. On peut ensuite analyser le pigment produit dans les pétales par HPLC, mais on ne peut pas faire de supposition quant aux éventuels substrats possible. Il faudrait donc, comme cela a déjà été fait pour l'enzyme d'*Amanita*, mettre au point un test enzymatique in vitro à partir d'enzyme exprimée dans une bactérie, afin de pouvoir étudier les différents substrats possibles des enzymes mutées ou non. Une telle mise au point demande de tester beaucoup de conditions différentes afin de déterminer les optimales. Le tampon utilisé pour les tests in vitro de l'enzyme d'*Amanita* ne fonctionnait pas avec le gène de *Portulaca* (Christinet, 2002, non publié). Il n'y avait toutefois pas d'ascorbate, alors que plusieurs

auteurs le mentionnent comme indispensable pour certaines dioxygénases bactériennes, afin de réduire le fer de l'état (III) à l'état (II).

Pour découvrir la fonction du gène chez les plantes sans bétalaïne, on pourra, une fois qu'on sera sûr qu'un des clones produits chez *Physcomitrella* est bien un knock-out, étudier un éventuel phénotype dans différentes conditions. En effet, même sur milieu minimal, dans des bonnes conditions de lumière et de température, il ne semble pas y avoir de phénotype. On pourrait donc imaginer de faire pousser ces mutants à des températures différentes, des conditions de lumière différentes ou en présence de toxiques à structure aromatique afin de mettre en évidence un éventuel phénotype. Ce gène de dioxygénase ne semble pas être impliqué dans le métabolisme primaire, sinon il est probable que tous les clones seraient morts, mais pourrait avoir une fonction de détoxification de composés produits en présence de facteurs de stress comme la température basse ou élevée, ou en présence de toxiques du milieu extérieur.

On retrouve ce gène dans les bases de données de cDNA chips d'*Arabidopsis*, mais son niveau d'expression ne semble pas varier de façon significative dans les expériences réalisées. Récemment, un mutant par insertion d'*Arabidopsis* a été identifié, les graines peuvent être obtenues et il serait intéressant de les analyser.

Si ces différentes expériences, dont certaines découlent du présent travail, donnent des résultats intéressants, on pourra certainement élucider les relations entre les dioxygénases « L.6 » des différentes espèces et leurs fonctions respectives.

Matériel et méthodes

Isolation du gène homologue à L.6 chez *Arabidopsis thaliana*

Nous disposons de la séquence complète du mRNA (AY050948) et d'une séquence incomplète d'un EST que nous avons pu obtenir directement dans un plasmide. Pour amplifier le gène, nous avons dessiné deux primers, L6AtcDNA, à partir de la séquence de l'EST, et L6AtcDNarev à partir de la séquence du mRNA. Leurs séquences étaient les suivantes :

L6AtcDNA : CGT CTA GAC CAC CAT GGA GAA AGT GAA TC

L6AtcDNarev : GGA GCT GCA GAT TGA AAT GTT TC

Les sites XbaI et PstI ont été introduits dans ces primers pour pouvoir facilement cloner le gène dans notre vecteur d'expression. Pour l'amplification par PCR, nous avons utilisé de la Taq polymérase Hifi (Roche) afin d'introduire le minimum d'erreurs dans la séquence.

Le vecteur d'expression a été préparé à partir du vecteur d'expression pNcoPgdod (Christinet) contenant le gène de *Portulaca grandiflora*. Le vecteur contient un promoteur 35S et un terminateur CaMVTer (voir figure 12) Le gène de *P. grandiflora* a été excisé avec les enzymes de restriction XbaI et PstI et le vecteur a été purifié sur gel au moyen du kit MinElute de Qiagen.

Les ligations ont été faites avec le kit pGem T-easy de Promega.

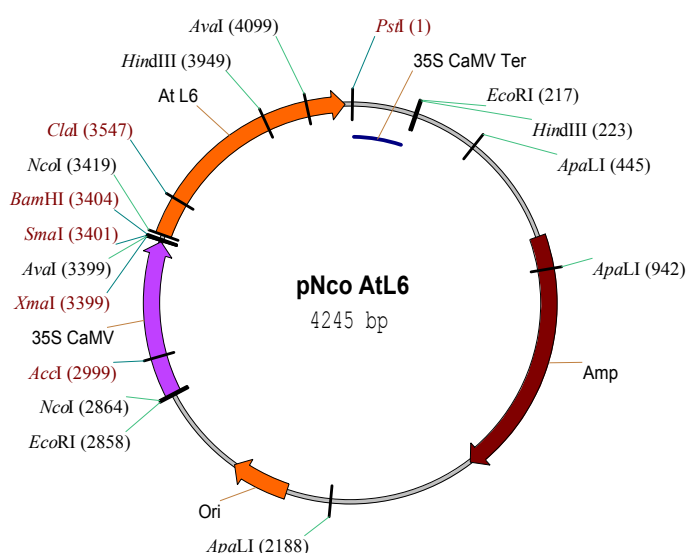


Figure 12 : Schéma du plasmide pNco AtL.6, contenant le gène L.6 d'*Arabidopsis thaliana* entouré d'un promoteur et d'un terminateur 35S.

Après vérification des transformants par analyse de restriction, un plasmide a été envoyé à séquencer à MicroSynth GmbH afin de vérifier sa séquence en vue de l'expression fonctionnelle. (pour la séquence, voir annexe 1)

Transformation transitoire par biolistique

Le plasmide contenant le gène d'*A.thaliana* pour la transformation transitoire a été amplifié et purifié par midipreps (kit Qiagen, midiprep Highspeed). Comme contrôle de transformation, nous avons utilisé un plasmide contenant le gène codant pour la protéine DS-Red, fourni par A. Finka.

Après quantification aux UV, la concentration des plasmides a été ajustée à 1 µg/µl, concentration requise par le protocole de la biolistique selon (Mueller, Hinz et al. 1997)

Isolation des homologues à L.6 chez *Physcomitrella patens*

Nous avons utilisé une banque de cDNA fournie par D. Schaefer. Nous connaissions le début de la séquence d'après une banque d'EST. Deux primers ont été fabriqués :

L6PpP1 : GAG TCG AGG TTT AGT GTG AGG TTT C

L6PpP2 : AGT TGG TTG TGC AAG CCT CAG ACG CA

Les extensions ajoutées à l'extrémité 3' des cDNA contiennent la séquence du primer T7, qui a été utilisé comme deuxième primer, pour faire deux amplifications successives, L6PpP1 T7 puis L6PpP2-T7, P2 se trouvant en aval de P1.

L'amplification a tout d'abord été faite avec de la Taq polymérase normale (Invitrogen), le fragment a été séquencé. Puis deux nouveaux primers ont été fabriqués à l'extrémité 3' du gène :

L6PpP4 : GTG AGC ATC CAC TCA AGC AGC ACA

L6PpP5 : ACC AAC AAC TAC CAT ACT AGT TTT GAG GAT

Puis les amplifications ont été refaites à partir de la banque de cDNA, avec les combinaisons L6PpP1-L6PpP4 et L6PpP2-L6PpP5, avec de la Taq Hifi (Roche), afin d'avoir le moins d'erreurs possible dans la séquence.

La même procédure a été suivie à partir du DNA génomique.

Les introns et exons ont été prédits sur la séquence génomique grâce aux consensus de début et de fin d'introns, ainsi qu'à l'aide de la séquence de cDNA. Il y a trois introns et quatre exons, comme montré sur la figure 13.

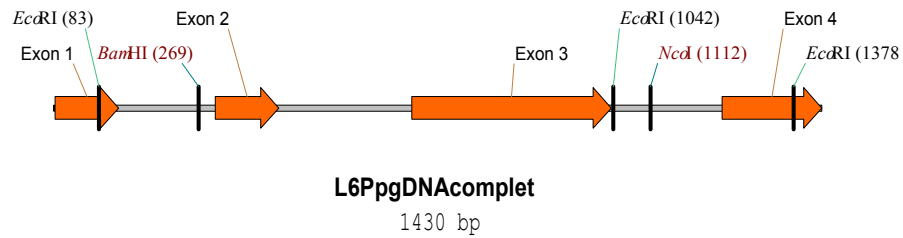


Figure 13 : Schéma de la séquence génomique du gène L.6 de *Physcomitrella patens*. Les flèches rouges représentent les exons, le reste les introns. Les sites BamHI et NcoI utilisés pour construire le plasmide pour le knock-out sont indiqués.

Culture et transformation de *Physcomitrella patens*

Le protocole de D. Schaefer (Principles and protocols for the moss *Physcomitrella patens*) a été suivi.

Le plasmide pour le knock-out est une cassette d'insertion. Il s'agit d'un plasmide pBluescript SK- dans lequel a été inséré le gène de résistance à la néomycine, entouré d'un promoteur et d'un terminateur 35S. Deux de tels plasmides ont été fournis par D. Schaefer, pKO ori1 et ori2, dans lesquels le sens du gène de résistance est inversé. Ils ont été digérés par l'enzyme de restriction XbaI, puis complétés avec de l'enzyme de Klenow, puis digérés par BamHI.

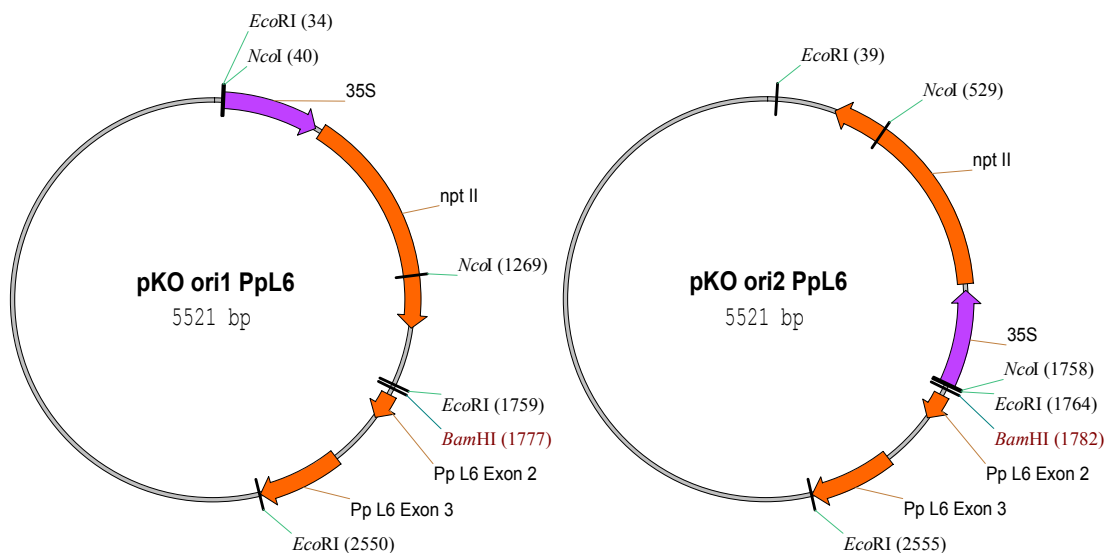


Figure 14 : Schéma des deux plasmides utilisés pour le knock-out de *Physcomitrella patens*. La différence entre ori1 et ori2 est l'orientation du gène de résistance au G418 ainsi que du promoteur et du terminateur. Le fragment du gène de *Physcomitrella patens* contient deux exons et un intron complets, et deux fragments d'introns à chaque bout.

Le plasmide pGem contenant le gDNA de *P. patens* a été digéré par NcoI, complétés avec de l'enzyme de Klenow, puis digérés par BamHI.

La ligation a ensuite été réalisée avec le pGem T-easy de Promega. Le plasmide a été amplifié selon la méthode de Maniatis (Purification of plasmid DNA, après une lyse alcaline).

Amplification de séquences génomiques chez plusieurs Caryophyllales

Des primers ont été fabriqués à partir des régions très conservées de l'alignement :

L6univSense : CGT TCT TCA TAT CTC ATG GRA CTC CA

L6univAs1 : ACA AAC YGG GAT GTC AGC YTC TGG

L6univAs2 : CCA GAR CCA ATG ATG AGR ACA CCT TC

Ces primers ont été indiqués en couleur sur la séquence de *Portulaca* (annexe 1).

PCR pour tester le mutant de *Physcomitrella patens*

Le protocole défini par Schween (2002) a été suivi. Les fragments de mousse ont été légèrement écrasés après les 15 minutes de lyse à 45°C. 10 mM de β -mercaptoéthanol ont été ajoutés au tampon de lyse. Les primers suivants ont été utilisés : L6PpP1, L6PpP4, T3 et T7, ces deux derniers existant dans le plasmide pBluescript SK- utilisé. La température d'annealing était de 50°C.

Alignements

Les alignements ont été générés avec le programme T-coffee, et ensuite retravaillés avec l'application Java Jalview.

Références

- Arciero, D. M. and J. D. Lipscomb (1986). "Binding of ^{17}O -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." J Biol Chem **261**(5): 2170-8.
- Arciero, D. M., J. D. Lipscomb, et al. (1983). "EPR and Mossbauer studies of protocatechuate 4,5-dioxygenase. Characterization of a new Fe^{2+} environment." J Biol Chem **258**(24): 14981-91.
- Arciero, D. M., A. M. Orville, et al. (1985). "[^{17}O]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." J Biol Chem **260**(26): 14035-44.
- Bittrich, V. and M. C. E. Amaral (1991). "Proanthocyanidins in the testa of centrospermous seeds." Biochem. Syst. Ecol(19): 319-321.
- Boyington, J. C., B. J. Gaffney, et al. (1993). "The three-dimensional structure of an arachidonic acid 15-lipoxygenase." Science **260**(5113): 1482-6.
- Clement, J. S. and T. J. Mabry (1996). "Pigment evolution in the Caryophyllales: a systematic overview." Bot. Acta(109): 360-367.
- Cuénoud, P., Chase M. W. and al. (2002). "Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences." American Journal of Botany **89**(1):132-144
- Eltis, L. D. and J. T. Bolin (1996). "Evolutionary relationships among extradiol dioxygenases." J Bacteriol **178**(20): 5930-7.
- Ehrendorfer (1976). "Closing remarks: Systematics and evolution of centrospermous families." Plant Syst Evol 126:99-106
- Fivaz, J. (1994). "Origine des bétalaïnes." Travail de diplôme.
- Girod, P. and J.-P. Zrýd (1990). "Biosynthesis of betalains: purification and partial characterisation of DOPA-4,5-dioxygenase from *Amanita muscaria*." Phytochemistry **00**: 001-006.
- Guex, N. and M. C. Peitsch (1997). "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling." Electrophoresis **18**(15): 2714-23.

- Mueller, L. A., U. Hinz, et al. (1997). "Biochemical complementation of the betalain biosynthetic pathway in *Portulaca grandiflora* by a fungal 3,4-dihydroxyphenylalanine dioxygenase." Planta **203**(2): 260-263.
- Prescott, A. G. and P. John (1996). "DIOXYGENASES: Molecular Structure and Role in Plant Metabolism." standard.dtl Annu. Rev. Plant Physiol. Plant Mol. Biol. **47**(1): 245-271.
- Rost, B. and C. Sander (1993). "Prediction of protein secondary structure at better than 70% accuracy." J Mol Biol **232**(2): 584-99.
- Saito, K. and A. Komamine (1976). "Biosynthesis of stizolobinic acid and stizolobic acid in higher plants. An enzyme system(s) catalyzing the conversion of dihydroxyphenylalanine into stizolobinic acid and stizolobic acid from etiolated seedlings of *Stizolobium hassjoo*." Eur J Biochem **68**(1): 237-43.
- Schaefer, D. G. and J. P. Zryd (1997). "Efficient gene targeting in the moss *Physcomitrella patens*." Plant J **11**(6): 1195-206.
- Spence, E. L., M. Kawamukai, et al. (1996). "Catechol dioxygenases from *Escherichia coli* (MhpB) and *Alcaligenes eutrophus* (MpcI): sequence analysis and biochemical properties of a third family of extradiol dioxygenases." J Bacteriol **178**(17): 5249-56.
- Stafford, H. A. (1994). "Anthocyanins and betalains - Evolution of the mutually exclusive pathways." Plant Science(101): 91-98.
- Sugimoto, K., T. Senda, et al. (1999). "Crystal structure of an aromatic ring opening dioxygenase LigAB, a protocatechuate 4,5-dioxygenase, under aerobic conditions." Structure Fold Des **7**(8): 953-65.
- Wilcox, M. E., Wyler, H., Mabry, T. J., Dreiding, A. S. (1965). "Die Struktur Des Betanins .7. Uber Die Konstitution Des Randenfarbstoffes Betanin." Helvetica Chimica Acta **48**(1):272-
- Zaiko, M. (2000). "Colour-specific genes from betalain producing plants." Travail de thèse.

Annexes

Annexe 1: Séquences obtenues

Séquences protéiques:

>Portulaca grandiflora

MVGKEVSFKESFFLSHGPNPAMLADESFIARNFLLGWKKNVFPVKPKSILVVSAAHWETDVPCV
SAGQYPNVIYDFTEVPASMFQMKYPAPGCPKLAKRVQELLIAGGFKSACLDEERGFHDSSWV
PLSMMCPEADI PVCQLSVQPGLDATHHFVGRALAPLKGEVLFIFSGGAVHPSDDTPHWFD
GVAPWAAEFDQWLEDALLEGRYEDVNNYQTKAPEGWKLAHP IPEHFLPLHVAMGAGGEKSKA
ELIYRTWDHGTLGYASYKFTSI

>Beta vulgaris

MGSEDNIKETFFISHGTPMMAIDDSKPSKKFLESWREKIFSKKPKAILVISAHWETDQPS
VNVVDINDTIYDFRGGFPARLYQFKYSAPGSPELANRIQDLLAGSGFKSVNTDKKRGLDHG
AWVPLMLMYPEADI PVCQLSVQSHLDGTHHYKLGQALAPLKDEGVLI IIGSGSATHPSNGT
PPCSDGVAPWAAAFDSWLETALTNNGSYEEVNKYETKAPNWKLAHPWPEHFYPLHVAMGAA
GENSKAELIHNSWDGGIMSYGSYKFTST

>Mammillaria sp.

KMSIDDSIPARKFFQEWKEKVYSKRPKAILVISAHWETNVPVAVNAVNHSDLIYDFRGGFPAIM
YQLKYPVPGAPNLRARVEELLTASGFSCVVDKNRGLDHGWSWVPLMLMYPEADI PVCQLSVQS
HLDGTHHYNLGKALAPLK

>Phytolacca americana

MAIDKSVQARPFLEGWREKVLSKKPKSILMISAHWETDVPTVNAVHSDLVYDFYGFAPMY
QLKYPAPGAPHLARRIQEVLTASGLKCAVDKKRGLDHGWSWVPLRLMYPEASIPVCQLSVQSN
LDATHHYNLGRALAPLK

>Iresine sp.

MSIDESIPARRFLEEWKDKVYSKRPSILVITAHWLTVVPTVSAIDHSDLIYDFGGFPACMY
QIKYTAPGSPDLAKRVQELVTGSGFECVDTKRGLDHGWSWVPLMLMYPEASIPVCQLSVQPH
LDGKHHDYDLGRALAPLK

>Arabidopsis thaliana

MEKVNQTFFLSHGSPTLSIDDSLEARQFFKSWTQKVLQPKPKSILVISAHWDTKFPSVNTVL
RNNTIHDGSGFPDPMYKLYEAPGAIELGKRVKELLMKEGGMKRVEDTKRGLDHGAWVPLM
LMYPEADI PICQLSVQSNQNGSYHYNMGKALASLKDEGVLI IIGSGSATHNLRKLDNFNITDGS
PVPWALEFDHWLRDSSLQGRYGDVNEWEKAPNAKMAHPWPEHLYPLHVVMGAAGDAKAEQ
IHTSWQLGTLSSYSSTSS

>Physcomitrella patens

MATSAGLSTFYVSHGSPMMPLEDTPIREFFSTWTERYPTRPKAILAISAHWDTREPAVNAV
QNSTIHDGFGFPRELYQLQYTPPGAPDVAKRVTSLKDGAGFKTVLEDNKRGLDHGAWTPLML
MYPNADIPVLQVSIQSNKDGHLHYQLGRALAPLKDEGVLI FASGTTVHNLREIDFSAKKPTV

WAKAFDGLWLTDLVLLNSKHKEAMEWEKAPYASKAHPHPDHFLPVLVGLGAAGEQCQAEKIYEE
FAYGLALSCFAFHPQN

Séquences génomiques

>Physcomitrella L.6 gDNA

ATGGCAACTTCAGCTGGTTTGGAGTACTTTTTATGTATCGCATGGGTCGCCTATGATGCCACT
TGAGGATACTCCCATACGAGAATTCTTCTCCACTTGGACGGAGCGCTATCCCACAAG
GTACTGCGTCAATCTTCTTTCTGATTACATCGATTGAAGCAAATAAACTCAGGTCAAGGAC
GCGAAATCACATCAATTCAATACAGCCGCGATAAGTTTTGAAAATTGCTCATCACAGAAGCA
ACCTACTTCTATCAAATAATATTGGATCCTATTGCATGTCCGCGGCTTTGATCAG
GCCAAAGGCCATTCTCGCAATTTAGCTCACTGGGATACTCGAGAGCCTGCCGTTAATGCTG
TCAGCCAGAACAGCACTATTCAGCACTTTTATGGCTTCCCTCGCGAGCTATATCAA
GTAATTCCTGCACTAACACACCACTACTTTACCCAGATTCTGCGCTACACAAGTTTTTGTTC
ACTGATTGGTACCTTGATTCATTTGTTTTCTCCTTAGTATCTGAATCTTTTTTTTCATTCTG
GTCACGATTACGTGACATATTTATAGATTCTCTGACAAATTCAAGGTCTACTTATATAACA
CCAGTCTTTTCTACGTCGTTTGTGCTCATGAGATAATTGTTTTCTCCTGTGTCTCGTCAG
TTGCAATACACGCCTCCAGGGGCTCCAGACGTTGCAAAGA
GGGTGACATCTCTCCTCAAAGACGCTGGCTTCAAGACCGTCCCTCGAGGACAACAAGAGAGGG
CTCGACCACGGAGCATGGACGCCCTGATGCTAATGTACCCCAACGCTGACATTCCGTGTCCT
CCAGGTCTCAATCCAAAGCAACAAAGATGGGCTTACCATTACCAGCTCGGCCGGGCTCTTG
CACCGCTCAAGGACGAAGGAGTGTTAATTTTTGCCTCTGGAACCTACAGTCCACAATTTGAGA
GAGATAGATTTTTCTGCCAAGAAGCCAACCTGTGTGGGCCAAGGCATTCGACGGATGGTTGAC
TGATGTGCTACTCAACAGCAAG
TAAGAATTCTTCCATTCCCTCCGTTTTCTTATCCATTTGAAGAGCCTGTCCACAAAGATCAGAG
TTCATAAATTCCCATGGATCATTTTACTGCATAATCTCTGTGGTTTTGATTTGTGACGGTTT
TTAGGGTTGTTTGAAGCTCTTTTCTTTCTTTCTTTTGTCTTTGTGAAGCTGACGTAGT
GTTAAACGACATCGATGCAGG
CACAAGGAAGCAATGGAATGGGAGAAAGCACCATATGCTTCCAAAGCTCACCCATCCAGA
TCATTTCCCTGCCTGTGTGGTTGGGTTGGGAGCTGCGGGAGAGCAATGCCAAGCTGAGAAGA
TATATGAGGAATTCGCATATGGTTTGGCGCTTTCATGTTTTGCTTTCATCCTCAAACCTAG

Les introns sont indiqués en rouge

>Portulaca L.6 cDNA

ATGGGTGTTGGGAAGGAAGTGTGCTTTAAGGAGAGTTTCTTCTTGTCTCATGGGAATCCAAGC
CATGTTGGCGGATGAATCCTTTATAGCGAGGAACCTTCCTTTTGGGATGGAAAAAGAATGTGT
TTCCTGTGAAACCAAGTCCATTTTAGTAGTCTCTGCTCACTGGGAGACTGATGTCCCTTGT
GTATCTGCTGGTCAGTACCCTAATGTTATCTACGATTTCACTGAAGTCCCTGCTTCCATGTT
TCAGATGAAGTACCAGCTCCAGGGTGCCCAAAGCTGGCAAAAAGAGTGCAGGAACCTGCTGA
TAGCGGGAGGTTTCAAGAGCGCGAAGCTAGATGAGGAGCGGGTTTCGACCACAGCTCATGG
GTGCCACTGAGCATGATGTGC CCGGAGGCGGACATCCCGGTGTGC CAGCTTTCGGTGCAGCC
TGGGCTTGACGCGACCCACCCTTCAACGTGGGGCGAGCGTTGGCCCCACTCAAAGGGGAAG
GTGTCTCTTTCATTGGCTCCGGTGGGGCTGTCCACCCTTCTGATGACACCCTCATTGGTTT
GATGGTGTGCTCCCTGGGCTGCTGAGTTTGATCAATGGCTTGAGGATGCTCTCCTAGAAGG
AAGGTACGAAGATGTGAATAACTATCAAACAAAAGCACCAGAAGGGTGAAGCTAGCACATC
CAATTCAGAACATTTTCTACCATTGCATGTAGCCATGGGTGCAGGTGGTGAAGAAATCAAAG
GCAGAGCTTATTTATCGTACTTGGGATCATGGTACTCTTGGCTACGCTCTTACAAGTTTAC
TTCCATATGA

L6univSense
L6univAs1
L6univAs2

>Arabisopsis L.6 cDNA

```
ATGGAGAAAGTGAATCAAACCTTCTTCTTATCTCATGGATCTCCAACCTTGAGCATAGATGA
TTCCTTGGAAGCGAGACAATTCTTCAAATCATGGACTCAAAAAGTTCTTCCTCAAAAACCCA
AATCGATTTTAGTCATCTCTGCTCATTGGGACACCAAATTCCCATCTGTCAACACTGTTCTT
CGCAATAACACGATCCACGATTTCTCTGGCTTCCCTGATCCTATGTACAAGTTGAAATATGA
AGCACCAGGAGCAATTGAATTAGGGAAAAGGGTTAAAGAGTTGTTAATGAAAGAAGGAGGAA
TGAAACGTGTTGATGAAGATACAAAGAGAGGACTTGATCATGGAGCTTGGGTTCCCTCTTATG
TTGATGTATCCAGAAGCTGATATAACCCATTTGTCAACTCTCTGTTCAATCAAATCAAATGG
GAGTTATCATTACAACATGGGTAAAGCATTGGCTTCATTGAAAGATGAAGGTGTTCTTATCA
TTGGCTCTGGAAGTGCTACTCATAACTTGAGGAAGCTTGACTTTAATATCACTGATGGCTCA
CCCGTTCCCTTGGGCCTTGGAGTTTGATCATTGGCTTAGAGATTCTCTACTTCAAGGAAGGTA
TGGAGATGTGAATGAGTGGGAAGAGAAAGCTCCGAATGCGAAAATGGCGCATCCGTGGCCCG
AGCATTGTATCCGTTACATGTTGTGATGGGTGCAGCAGGTGGAGATGCAAAGGCAGAGCAA
ATCCATACAAGTTGGCAACTTGGTACCTTGTCTTATTCTTCTTATAGCTTCACTTCTTCTCT
TTGA
```