



Approches combinatoires pour la reconstruction d'une voie de biosynthèse chez la levure : variation des niveaux d'expression et analyse fonctionnelle d'une étape clé de la voie

Marie Carquet

► To cite this version:

Marie Carquet. Approches combinatoires pour la reconstruction d'une voie de biosynthèse chez la levure : variation des niveaux d'expression et analyse fonctionnelle d'une étape clé de la voie. Biotechnologies. INSA de Toulouse, 2015. Français. NNT : 2015ISAT0024 . tel-01233865

HAL Id: tel-01233865

<https://theses.hal.science/tel-01233865>

Submitted on 25 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par

Institut National des Sciences Appliquées de Toulouse

Discipline ou spécialité : *Ingénieries Microbienne et Enzymatique*

Présentée et soutenue par
Marie CARQUET

Le 24 Mars 2015

Titre : Approches combinatoires pour la reconstruction d'une voie de biosynthèse chez la levure: variation des niveaux d'expression et analyse fonctionnelle d'une étape clé de la voie.

JURY

M. GEISELMANN H.
M. DAIGNAN-FORNIER B.
Mme. WERCK-REICHHART D.
M. KURRAS L.
M. JAY-ALLEMAND C.
M. TRUAN G.

Professeur d'Université, Grenoble
Directeur de recherche CNRS
Directrice de recherche CNRS
Chargé de recherche CNRS
Professeur d'Université, Montpellier
Chargé de recherche CNRS

Rapporteur
Rapporteur
Rapporteur
Examineur
Examineur
Directeur de thèse

Ecole doctorale : *Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)*
Unité de recherche : *Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP)*
UMR 5504, UMR 792, CNRS, INRA, INSA. 135 Avenue de Rangueil-31077 Toulouse, France
Directeur de Thèse : *Gilles TRUAN*
Rapporteurs : *Geiselmann H., Daignan-Fornier B., et Werck-Reichhart D.*



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par

Institut National des Sciences Appliquées de Toulouse

Discipline ou spécialité : *Ingénieries Microbienne et Enzymatique*

Présentée et soutenue par
Marie CARQUET

Le 24 Mars 2015

Titre : Approches combinatoires pour la reconstruction d'une voie de biosynthèse chez la levure: variation des niveaux d'expression et analyse fonctionnelle d'une étape clé de la voie.

JURY

M. GEISELMANN H.
M. DAIGNAN-FORNIER B.
Mme. WERCK-REICHHART D.
M. KURRAS L.
M. JAY-ALLEMAND C.
M. TRUAN G.

Professeur d'Université, Grenoble
Directeur de recherche CNRS
Directrice de recherche CNRS
Chargé de recherche CNRS
Professeur d'Université, Montpellier
Chargé de recherche CNRS

Rapporteur
Rapporteur
Rapporteur
Examineur
Examineur
Directeur de thèse

Ecole doctorale : *Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)*
Unité de recherche : *Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP)*
UMR 5504, UMR 792, CNRS, INRA, INSA. 135 Avenue de Rangueil-31077 Toulouse, France
Directeur de Thèse : *Gilles TRUAN*
Rapporteurs : *Geiselmann H., Daignan-Fornier B., et Werck-Reichhart D.*

AUTEUR : Marie CARQUET

Année : 2015

Thèse dirigée par le Dr Gilles TRUAN

TITRE : Approches combinatoires pour la reconstruction d'une voie de biosynthèse chez la levure: variation des niveaux d'expression et analyse fonctionnelle d'une étape clé de la voie.

RESUME :

Afin d'optimiser la production d'un composé d'intérêt tout en évitant les conséquences néfastes sur la viabilité cellulaire, un défi majeur de l'ingénierie métabolique est d'obtenir un équilibre entre les flux métaboliques endogènes de la cellule et le flux consommé par une nouvelle voie de biosynthèse. Dans ce contexte d'optimisation, les stratégies combinatoires génèrent une diversité de voies métaboliques et de modes de régulation rassemblant de précieuses informations quant aux choix d'orientations stratégiques à faire.

Notre étude s'inscrit dans un projet visant à produire les molécules responsables de l'arôme, de la couleur et du parfum du safran (*Crocus sativus*) chez *Saccharomyces cerevisiae*. Une approche combinatoire a été adoptée pour moduler les niveaux d'expression de trois gènes menant à la synthèse de leur précurseur commun : la zéaxanthine. Cette stratégie nous a permis de décrire des biais inattendus dans la régulation des niveaux d'expression des gènes exprimés sur plasmide. Nous avons détecté une forte interférence transcriptionnelle entre les gènes dans notre système, ainsi qu'une influence de la nature des séquences codantes. Ces éléments, identifiés comme critiques, imposent sur les niveaux d'expression des trois gènes une régulation plus importante que la force des promoteurs qui les contrôlent. Afin de poursuivre le projet vers son objectif final, la réaction de clivage de la zéaxanthine menant à la synthèse des composés d'intérêt du safran a fait l'objet d'une analyse fonctionnelle détaillée. Une absence d'activité de l'enzyme décrite dans la littérature comme responsable de cette réaction nous a conduits à proposer des perspectives d'ingénierie pour atteindre l'objectif final du projet.

MOTS-CLES : Ingénierie métabolique combinatoire, *Saccharomyces cerevisiae*, régulation transcriptionnelle, caroténoïdes, dioxygénases, arômes.

AUTHOR: Marie CARQUET

Year: 2015

Thesis directed by Dr Gilles TRUAN.

TITLE: Combinatorial approaches to rebuild a biosynthetic pathway in yeast: variation of expression levels and functional analysis of a key step in the pathway.

ABSTRACT:

To optimize the production of a value added compound while avoiding toxic consequences on the cell viability, a challenge in the metabolic engineering field is to balance the endogenous metabolic fluxes and the newly consumed fluxes. In this optimization context, combinatorial strategies can generate several variants of synthetic metabolic pathways. This strategy gives precious strategic information on the right combinations of function and regulation choices to be made in the ultimate pathway reconstruction.

Our study aimed at the production of the molecules responsible for aroma, dye, and fragrance of saffron (*Crocus sativus*) in *Saccharomyces cerevisiae*. A combinatorial approach was chosen to modulate expression levels of three genes involved in their common precursor biosynthesis: zeaxanthin. This strategy allowed us to describe some unexpected bias in the regulation of the plasmid-encoded genes expression levels. We detected strong transcriptional interference between the different genes in our system, and the ORF nature also seemed to influence the expression levels. These critical factors imposed a stronger regulation of the three genes expression levels than the promoter strength initially chosen to control them. The project was continued toward its final objective by making a detailed functional analysis of the zeaxanthin cleavage reaction leading to the molecules of interest synthesis. This reaction was described to be catalyzed by a specific enzyme, but no activity was observed in our experiments. This result led us to propose some tools to reach the final goal of the project.

KEY WORDS: Combinatorial metabolic engineering, *Saccharomyces cerevisiae*, transcriptional regulation, carotenoids, dioxygenases, aromas.

Remerciements

Je souhaite tout d'abord remercier la région Midi-Pyrénées et à l'ADRIAF pour avoir co-financé ces travaux. Merci également à Mr Nic Lindley, directeur du LABORATOIRE D'INGENIERIE DES SYSTEMES BIOLOGIQUES ET DES PROCEDES (LISBP) pour m'avoir accueillie dans son unité de recherche.

Je tiens vivement à souligner la collaboration qui a caractérisé ma relation avec mon directeur de thèse, Mr Gilles TRUAN, pendant ces trois années. C'est d'un véritable partenariat dont je me souviendrai, au sein duquel j'ai pu bénéficier d'un encadrement basé sur une discussion scientifique saine et toujours agréable. Je le remercie pour la confiance qu'il m'a accordée, cela m'a permis de grandir et d'évoluer bien au-delà de ce que j'espérais.

Je remercie toute l'EAD11, pour tous les échanges scientifiques et amicaux que l'on a pu avoir pendant ces années. Les soutiens de Denis Pompon, Philippe Urban, Thomas Lautier et Luis Garcia ont souvent été les bienvenus dans les moments difficiles. J'ai été la première doctorante de cette équipe depuis son installation à Toulouse. J'ai vu arriver nombre de personnes, et malheureusement, j'en ai vu repartir beaucoup aussi. Je ne les oublie pas pour autant. Franciane, tes valeurs sont belles et justes, restes comme tu es surtout, ne changes rien. Ton soutien a été essentiel à mes yeux. Frédéric, qu'est-ce que c'était bien d'avoir un tel puits de science si près de nous, dans le même bureau, à la paillasse de derrière, et en face d'un verre de temps en temps ! Merci pour tes si précieux conseils. Yoann, tu es une personne que j'estime beaucoup, humainement et scientifiquement. Je te souhaite sincèrement d'obtenir ce que tu souhaites, tu le mérites par ta sincérité et ta droiture. Arthur, quelques balades très sympas, à vélo ou à pied me reviennent en mémoire, parfois même avec quelques centaines de mètres de vide sous nos pieds... On a quand même bien rit, juste une chose : arrêtes de stresser, ça va bien se passer. Florence, merci pour tes conseils toujours avisés et réfléchis, je garderai de toi le souvenir d'une femme très forte, que j'admire un peu je dois bien l'avouer. Et enfin, je ne peux pas clore cette liste sans parler de celle dont la rencontre laissera dans mon esprit une trace indélébile, celle qui m'a forcée à me découvrir: Romy. Je ne suis pas spécialiste des épanchements sentimentaux. Je n'en dirai donc pas plus, je préfère m'arrêter là plutôt que de coucher sur ce papier des mots trop fades pour te décrire, si tant est que je les trouve...

Je souhaite également remercier le Dr Zya Günata et le Dr Nancy Terrier, pour leur gentillesse et pour avoir été d'un grand secours en me fournissant leur matériel biologique.

J'ai également trouvé un soutien régulier et précieux auprès de Marie-Ange Teste à l'EAD5, ou de Julie Lesage à l'EAD8. Merci d'avoir été là. J'ai eu peu de contact externe à l'équipe pendant mes travaux de rat de labo un peu isolé, mais les rencontres que j'ai faites étaient des rencontres de qualité.

Bien sûr merci à toute ma famille, et particulièrement à Jérôme, qui a su occuper la toute petite place que cette thèse lui a laissé dans ma vie. Merci de m'avoir soutenue, poussée aussi de temps en temps, et surtout, merci de me supporter.

*Je dédie ces
travaux à mes grands-parents,
tous deux décédés pendant cette thèse.*

Table des matières

Introduction21

CHAPITRE 1 : L'ingénierie métabolique dans les biotechnologies - L'implication de *Saccharomyces cerevisiae*.21

I. LES MICROORGANISMES, CATALYSEURS DES BIOTECHNOLOGIES – ESSOR DE LA BIOLOGIE DE SYNTHÈSE.....23

1. Introduction aux biotechnologies.23

2. La biologie de synthèse.....25

II. L'INGENIERIE METABOLIQUE CHEZ *S. CEREVISIAE*.....28

1. Les Applications.....31

1.1 Production de métabolites.31

1.2 Production de protéines recombinantes.34

1.3 Biotransformations *in vivo*.....35

2 Les outils utilisés pour la reconstruction de voies métaboliques chez la levure.....36

2.1 Les constructions plasmidiques.37

2.2 Les YACs.39

2.3 L'intégration chromosomique.40

2.4 L'émergence de nouveaux outils d'assemblage.44

2.5 Les stratégies combinatoires.....50

3 Les stratégies de régulation de l'expression de gènes.....55

3.1 Points de contrôle au niveau de la molécule d'ADN.....56

3.2 Points de contrôle au niveau transcriptionnel.....57

3.3 Les points de contrôle au niveau traductionnel.60

3.4 L'interaction entre biologie synthétique et systémique conduit à la perspective d'une régulation dynamique.61

CHAPITRE 2 : Les caroténoïdes et leur production par ingénierie métabolique chez *Saccharomyces cerevisiae*.65

I. DEFINITION ET PROPRIETES D'INTERET DES CAROTENOÏDES.....67

II. LE CLIVAGE DES CAROTENOÏDES DONNE NAISSANCE A DES AROMES ET DES PARFUMS.....	70
III. LA PRODUCTION DE CAROTENOÏDES CHEZ <i>S. CEREVISIAE</i>.	72
1. Les précurseurs endogènes.	72
2. L'ingénierie métabolique de <i>Saccharomyces cerevisiae</i> pour produire des caroténoïdes.	74

Présentation du projet de thèse83

Résultats et Discussion85

CHAPITRE 1 : Etude des forces de promoteurs constitutifs de *S. cerevisiae*85

I. DONNEES INTRODUCTIVES.	87
II. Transcription interference and ORF nature strongly affect promoter strength in a reconstituted metabolic pathway.	93
Abstract.....	93
1. Introduction.....	94
2. Materials and methods	96
Plasmid constructions	96
Strains, culture and sampling conditions	99
RT-qPCR experiments.....	99
mRNA variation quantification	100
pRS426 copy number determination	100
3. Results.....	101
Zeaxanthin pathway library generation	101
<i>S. cerevisiae</i> strains, culture and sampling.....	102
pRS426 copy number quantification.....	103
Expression profile of the pRS426 marker	105
Controls of plasmid integrity and cell physiology.....	106
Absence of correlation between promoters strengths and mRNA levels	107
Gene orientation and order strongly affect promoters strengths	109

4. Discussion	111
5. Acknowledgement	112
6. Author's contribution	113
7. References	113
Supplementary Material	117
1. Supplementary Data	117
2. Supplementary Figures and tables	118
Supplementary Tables	118
Supplementary Figures	122
3. Reference	125
III. RESULTATS COMPLEMENTAIRES	126
1. Construction de la banque combinatoire.	126
2. Production de caroténoïdes	128
IV. DISCUSSION	131
CHAPITRE 2 : Etude de l'activité des Carotenoid Cleavage Dioxygenases connues chez <i>Crocus sativus</i>.	135
I. DONNEES INTRODUCTIVES.	137
1. Données structurales et mécanisme de clivage des caroténoïdes par les Carotenoid Cleavage Oxygenases (CCOs)	138
2. Le cas de <i>Crocus sativus</i>	141
II. RESULTATS.	145
1. Analyses bioinformatiques des données disponibles sur les CCOs de <i>Crocus sativus</i>	146
1.1 Etude de similarités de séquences.	146
1.3 Analyse structurale.	149
2. Description des constructions de biologie moléculaire utilisées dans la suite de l'étude.	152
3. Expression de la protéine chez <i>E. coli</i>	154
3.1 Validation de l'expression.	154
3.2 Essai de re-solubilisation.	155
3.3 Optimisation des conditions de culture.	156

3.4	Essai de différentes souches recombinantes.....	157
4.	Tests d'activité.	159
4.1	Reproduction des conditions de Bouvier et al., 2003b.	159
4.2	Reproduction de conditions optimisées pour la caractérisation d'une CCO chez <i>Vitis vinifera</i>	160
4.2.1	Test d'activité de <i>VvCCD1</i>	161
4.2.2	Tests d'activité dans les conditions de Mathieu et al. 2005.	163
4.2.3	Essai d'optimisation du test d'activité 1.	166
4.2.4	Essai d'optimisation du test d'activité 2.	168
4.2.5	Test d'activité de <i>CsCCD</i>	170
4.3	Tests d'activité <i>in vitro</i> sur lysats de levures.....	172
4.4	Tests d'activité <i>in vivo</i> chez <i>Saccharomyces cerevisiae</i>	173
III.	DISCUSSION	176

Conclusion et Perspectives181

Matériel et Méthodes187

1.	Analyses <i>in silico</i>.	187
2.	Construction de souches d'expression.	187
2.1	Clonages.....	187
2.2	Transformation de souches.	190
3.	Expression de protéines.	190
4.	SDS-Page	191
5.	Tests d'activité.	192
5.1	Tests d'activité <i>in vitro</i> après expression de protéines chez <i>E. coli</i>	192
5.2	Tests d'activité <i>in vitro</i> après expression de protéine chez <i>S. cerevisiae</i>	193
5.2.1	Préparation des fractions microsomales.....	193
5.2.2	Préparation des liposomes.	193
5.3	Tests d'activité <i>in vivo</i> après expression de protéine chez <i>S. cerevisiae</i>	194
6.	Analyse de métabolites.	195
7.	Perte de plasmide et test cross.	195

Références Bibliographiques	197
ANNEXE 1 : Séquences Codantes des Enzymes Etudiées	215
ANNEXE 2 : Amorces de PCR	223
ANNEXE 3 : Milieux de Culture	225

Liste des abréviations

Acétyl-CoA : Acétyl-Coenzyme A

ADN : Acide DésoxyriboNucléique

ADNc : ADN complémentaire

ARS : Séquence de Réplication Autonome

BET: Bromure d'éthidium

CEN : Centromère

DHAP : Dihydroxyacétone phosphate

[dw] : poids sec

EDTA: Ethylène Diamine Tétra Acétique

FDA: Food and Drug Administration

Glucose-6-P: Glucose-6-Phosphate

GFP: Green Fluorescent Protein

HPLC: High Performance Liquid Chromatography

IPTG: Isopropyl β -D-1-thiogalactopyranoside

Kb : Kilobase

OCDE: Organisation de Coopération et de Développement Economiques

PDB: Protein DataBase

PCR : Polymérase Chain Reaction

PEP : Phosphoénol pyruvate

Liste des figures

- Figure 1:** Schéma représentatif des approches de déconstruction (Top-Down) de systèmes existants, et de construction (Bottom-Up) de nouveaux systèmes vivants.27
- Figure 2:** Représentation schématique d'une voie métabolique. Exemple d'une partie de la voie du mévalonate de *S. cerevisiae*. 1 = Acetyl-CoA C-acetyltransferase; 2 = 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase; 3 = HMG-CoA reductase; 4 = Mevalonate kinase. Mévalonate-5-P = Mévalonate-5-phosphate.28
- Figure 3:** Exemples de projets d'ingénierie métabolique effectués chez *S. cerevisiae*. Les étapes naturellement présentes chez *S. cerevisiae* sont représentées en noir et blanc ; les substrats métabolisés après ingénierie, et les produits obtenus de façon hétérologue sont colorés. Glucose 6-P, Glucose 6-phosphate ; DHAP, Dihydroxyacétone phosphate ; PEP, Phosphoénolpyruvate ; Acétyl-CoA, Acétyl- Coenzyme A.33
- Figure 4:** Eléments composant les constructions plasmidiques utilisées chez *S. cerevisiae*. (A) plasmide monocopie. (B) plasmide multicopies. CEN/ARS, centromère/séquence de réplication autonome ; *URA3*, marqueur de sélection par l'uracile ; ori, origine de réplication de *E. coli* ; *AmpR*, gène de résistance à l'ampicilline ; 2 ori, origine de réplication 2 μ . Les flèches blanches représentent les promoteurs.....38
- Figure 5:** Elements constituant un YAC. CEN/ARS, centromère/séquence de réplication autonome ; TEL, télomère ; *TRP1*, gène Tryptophane1 ; *URA3*, gène Uracile3 (marqueurs de sélection).39
- Figure 6:** (A) Intégration chromosomique via un vecteur d'intégration par simple crossing-over. (B) Intégration chromosomique via un vecteur d'intégration par double crossing-over. (C) Intégration chromosomique d'un fragment d'ADN linéaire (produit de PCR) par double crossing-over. Figure adaptée de Da Silva & Srikrishnan 2012.42
- Figure 7:** Schéma de la méthode de clonage multiple par l'enzyme de restriction *SfiI* (Lu et al., 2007).....46
- Figure 8:** Méthodes d'assemblage de fragments d'ADN. (A) Méthodologie suivie pour assembler des fragments d'ADN par recombinaison en générant des séquences homologues par PCR. EXO, Exonucléase 3' – 5'. (B) Méthode Gibson et al. 2009.48
- Figure 9:** Principe de fonctionnement de la PCR fusion. Plusieurs PCRs sont réalisées en parallèle de façon à obtenir chaque fragment à assembler pourvu de séquences homologues de ses séquences

partenaires. Une deuxième PCR est ensuite réalisée avec tous les fragments à assembler et les amorces s'hybridant aux extrémités de la construction finale.49

Figure 10: Méthode Lund et al. 2014. Sur cette figure, le vecteur dans lequel la séquence d'ADN cible doit être intégrée possède un gène *ccdB* situé entre les deux bornes d'hybridation du premier produit de PCR. Ce gène code pour une protéine « poison » de la bactérie *E. coli*. Si le produit de PCR n'a pas correctement « pris la place » du gène *ccdB*, la cellule de *E. coli* qui portera le plasmide résultant ne poussera pas. Il s'agit d'une stratégie d'évitement des faux positifs.50

Figure 11: Etapes composant la méthode de « genome shuffling ».52

Figure 12: Spectres d'absorption de quatre caroténoïdes entre 250 et 550 nm. (A) Lycopène. (B) β -carotène. (C) Zéaxanthine. (D) Astaxanthine.67

Figure 13: (A) Exemples de molécules entrant dans la catégorie des carotènes ou des xanthophylles. (B) Système de numérotation des atomes de carbone sur le modèle du β -carotène.68

Figure 14: Biosynthèse de deux molécules de rétinol à partir d'une molécule de β -carotène.70

Figure 15: Exemples d'apocaroténoïdes utilisées comme additifs alimentaires ou cosmétiques. (A) Bixine. (B) Crocine.71

Figure 16: Comparaison des molécules de (A) cholestérol (animaux) et de (B) ergostérol (*fungi*).72

Figure 17: Voie de biosynthèse de l'ergostérol chez *S. cerevisiae*. Les gènes codant pour les enzymes responsables de chaque étape sont indiqués en bleu. HMG CoA, 3-hydroxy-3-méthylglutaryl-coenzyme A; IPP, isopentenyl pyrophosphate; DMAPP, Diméthylallyl-pyrophosphate ; GPP, geranyl pyrophosphate ; FPP, farnesyl pyrophosphate.73

Figure 18: Voies de biosynthèse pouvant dériver du farnesyl pyrophosphate (FPP) chez *S. cerevisiae*. Les voies endogènes de *S. cerevisiae* sont représentées en noir, une voie hétérologue en rouge. GGPP, geranylgeranyl pyrophosphate.74

Figure 19: Voie de biosynthèse de l'astaxanthine chez *Xanthophyllomyces dendrorhous*. CrtE, GGPP synthase ; CrtYB, phytoène synthase et lycopène cyclase ; CrtI, phytoène désaturase ; CrtS, P450 (β -carotène hydroxylase) ; CrtR, P450 réductase.79

Figure 20: Présentation de la voie métabolique à implémenter chez la levure *S. cerevisiae*. Les étapes endogènes de *S. cerevisiae* sont indiquées en noir, les étapes hétérologues à implémenter sont colorées.84

Figure 21: Spectres HPLC (450 nm) résultants de l'injection des extraits de caroténoïdes d'un clone blanc (trait bleu) et d'un clone jaune (trait noir). Trait rouge : standard de zéaxanthine. Trait vert : extrait d'une souche contrôle contenant pRS426 et pCM185 vides.	126
Figure 22: Migration sur gel d'agarose 0.8% (w/v) des extraits plasmidiques de 4 clones de levure jaunes et 4 clones blancs (contenant les dérivés de pRS426 et pCM185/ <i>tHMG1</i>) après digestion par <i>Bgl</i> 1. Les tailles de fragments attendues pour pCM185/ <i>tHMG1</i> sont : 3286 bp; 2301 bp; 1646 bp; 1305 bp; 371 bp; 122 bp et 71 bp. Les tailles de fragments attendues pour les dérivées de pRS426 sont : 8245 bp et 4032 bp. Révélation sous lumière UV après coloration au BET. Les bandes en dessous de 250 bp ne sont pas visibles sur ce gel pour une question de quantité d'ADN déposée, mais les bandes visibles correspondent aux profils attendus.	127
Figure 23: Spores issues du croisement d'un clone blanc (A) ou jaune (B) avec une souche <i>Aura3</i> , sur milieu dépourvu d'uracile. (C) Souche contrôle présentant une croissante correcte, sur milieu riche. Ce témoin, ayant eu le même temps de croissance que les tests cross, permet de voir l'absence de croissance en (A) et (B).	128
Figure 24: Analyse HPLC (450 nm) résultants de l'injection des extraits de caroténoïdes des souches I ⁺ H (trait vert), J (trait rouge) et de la souche contrôle contenant pRS426 et pCM185 vides. L'astaxanthine est un standard d'extraction utilisé pour normaliser la quantification des produits par rapport à la quantité de cellules extraites (voir article).	129
Figure 25: Dosages de zéaxanthine et de β -carotène produits par les souches A à J. Les points bleus représentent les souches n'exprimant pas <i>tHMG1</i> ; les points rouges représentent les souches qui expriment <i>tHMG1</i> . Les barres d'erreurs représentent un écart type sur trois valeurs.	130
Figure 26: Voie de biosynthèse de la crocine, la picrocrocine et du safranal chez <i>Crocus sativus</i> à partir de la zéaxanthine.	137
Figure 27: Topologie (A) et structure tridimensionnelle (B) de la CCO bovine RPE65 (code PDB 3FSN). L'ion catalytique Fe ²⁺ est coloré en orange. Les hélices α sont en bleu et les feuillet β sont en vert. La ligne pointillée rouge représente une zone absente du modèle structural. Figure adaptée de Sui et al. 2013.	139
Figure 28: Vue de surface de la structure tridimensionnelle de VP14 (<i>Zea mays</i>). Code PDB 3NPE. Le patch hydrophobe est coloré en jaune. La flèche noire indique l'ouverture du tunnel menant le substrat au site actif. Figure adaptée de Sui et al. 2013.	140
Figure 29: Mécanismes catalytiques des monooxygénases et des dioxygénases. Figure adaptée de Sui et al. 2013.	141
Figure 30: Fleur de <i>Crocus sativus</i> . Les stigmates rouges donnent naissance, après séchage, au safran.	142

- Figure 31:** Alignement (COBALT) de la séquence protéique de CsZCD avec les cinq séquences protéiques les plus proches.....147
- Figure 32:** Alignement (ClustalW) des séquences protéiques de CsZCD (N° d'accension UniProtKB/Swiss-Prot Q84K96) et VP14 (N° d'accension UniProtKB/Swiss-Prot O24592.2). * = acides aminés identiques. : = acides aminés de même famille. . = acides aminés proches du point de vue de l'encombrement.150
- Figure 33:** Structure de VP14 (*Zea mays*) visualisée grâce au logiciel PyMOL. Code PDB : 3NPE. Résolution 3.20 Å. Bleu : résidus qui s'alignent avec ceux de CsZCD ; rouge : résidus manquants dans CsZCD (positions 1 à 237 de VP14) ; vert : ion catalytique Fe²⁺151
- Figure 34:** Migration sur gel SDS-PAGE des fractions solubles et insolubles d'échantillons prélevés dans une culture (37°C) de *E. coli* BL21(De3) transformée par Z/p15. S = surnageant ; C = culot ; Contrôle = souche non induite ; 1h = prélèvement à T1 ; 2h = prélèvement à T2 ; 3h = prélèvement à T3. Taille attendue de CsZCD : 43kDa. La flèche rouge indique CsZCD.....155
- Figure 35:** Migration sur gel SDS-PAGE d'un essai de re-solubilisation de CsZCD contenue dans un culot par ajout de TRITON X-100 1% (w/v). S = surnageant ; C = culot ; 1h = incubation à 30°C pendant 1 heure ; 2h = incubation à 30°C pendant 2 heures ; 3h = incubation à 30°C pendant 3 heures. La flèche rouge indique CsZCD.156
- Figure 36:** Migration sur gel SDS-PAGE des protéines de *E. coli* BL21(De3) transformée avec Z/p15, cultivées en milieu ZYM 5052 à 20°C. C = culot ; S = Surnageant. La flèche rouge indique CsZCD.....157
- Figure 37:** Migration sur gel SDS-PAGE des protéines de *E. coli*, C41(De3) et Origami(De3) transformées avec Z/p15 et cultivées en ZYM 5052 à 20°C. (A) Dépôt des surnageants. (B) Dépôts des culots. T- = souche contrôle transformées avec pET15b vide. CsZCD = souche transformée avec la construction Z/p15. Taille attendue de CsZCD: 43 kDa.....158
- Figure 38:** (A) Migration sur gel SDS-PAGE des différentes fractions obtenues après cassage des cellules de *E. coli* (TOP10) cultivées dans les conditions de (Bouvier et al., 2003b). Expression = souche TOP10 transformée avec la construction Z/pBAD ; Contrôle = souche TOP10 non transformée ; S = surnageant, C = culot. Taille attendue de CsZCD + thiorédoxine = 53 kDa. (B) analyse HPLC (440 nm) résultant de l'injection de tests d'activité effectués dans les conditions de (Bouvier et al., 2003b). Vert : test d'activité effectué sur le surnageant de cassage d'une souche TOP10 exprimant la construction Z/pBAD. Rouge : test d'activité effectué sur le surnageant de cassage de la souche contrôle.....160
- Figure 39:** Réaction de clivage de la zéaxanthine en positions 9,10(9',10') catalysée par VvCCD1 (Mathieu et al., 2005).161

Figure 40: (A) Analyse HPLC (416 nm) résultant de l'injection des produits du test d'activité de VvCCD1 fait dans les conditions de (Mathieu et al., 2005). Tests d'activité effectués sur les surnageants de cassage de souches BL21(De3) : Rouge : V/pGEX. Vert : BL21(De3) non transformée. (B) Spectre d'absorption du produit élué à 2,7 minutes récolté entre 250 nm et 500 nm.....162

Figure 41: Analyse HPLC des produits des tests d'activité de CsZCD, CsZCD^{+N+P} et VvCCD1 dans les conditions de Mathieu et al. 2005. (A) Spectres enregistrés à 416 nm. (B) Spectres enregistrés à 440 nm. Les tests d'activité ont été réalisés sur les fractions solubles issues du cassage de cellules BL21(De3) exprimant : Noir : Z^{+N+P}/p15. Bleu : Z/p15. Vert : V/pGEX. Rouge : pET15b vide.164

Figure 42: Analyse HPLC (416 nm) résultant de l'injection d'un test d'activité de VvCCD1 fait dans les conditions de Mathieu et al. 2005, 24 heures après l'arrêt du test d'activité.165

Figure 43: Analyse HPLC des produits des tests d'activité de CsZCD, CsZCD^{+N+P} et VvCCD1 dans les conditions de Mathieu et al. 2005 optimisée pour le pH et la température de test. (A) Spectres enregistrés à 416 nm. (B) Spectres enregistrés à 440 nm. Les tests d'activité ont été réalisés sur les fractions solubles issues du cassage de cellules BL21(De3) exprimant : Noir : Z^{+N+P}/p15. Bleu : Z/p15. Vert : V/pGEX. Rouge : pET15b vide.....167

Figure 44: Analyse HPLC des produits des tests d'activité de CsZCD, CsZCD^{+N+P} et VvCCD1 dans les conditions de Mathieu et al. 2007. Les spectres VvCCD1 sont enregistrés à 416 nm, les autres à 440 nm. (A) Substrat = zéaxanthine. (B) Substrat = β -carotène. Les tests d'activité ont été réalisés sur les fractions solubles issues du cassage de cellules BL21(De3) exprimant : Noir : V/pGEX ; Bleu : Z^{+N+P}/p15 ; Vert : Z/p15 ; Rouge : pET15b vide.....169

Figure 45: Analyse HPLC (416 nm) résultants de l'injection des produits des tests d'activité de CsCCD et VvCCD1 dans les conditions de Mathieu et al. 2007. Les tests d'activité ont été réalisés sur les fractions solubles issues du cassage de cellules BL21(De3) exprimant : Bleu : C/p15 ; Vert : V/pGEX ; Rouge : pET15b vide.170

Figure 46: Migration sur gel SDS-PAGE des fractions de cassage de souches de *E. coli* BL21(De3) cultivées dans les conditions de Mathieu et al. 2007. Les cellules exprimaient les constructions suivantes : Contrôle : pET15b vide ; VvCCD1 : V/pGEX ; CsCCD : C/p15. Taille attendue de CsCCD : 62 kDa. Taille attendue de VvCCD1 + GST : 88 kDa.171

Figure 47: Analyse HPLC (440 nm) résultants de l'injection de tests d'activité *in vitro* après expression de CsZCD chez *S. cerevisiae*. Tests d'activité effectués sur des lysats de levure CEN.PK. Rose : CEN.PK non transformée ; Bleu : fraction cytosolique de levures exprimant CsZCD ; Vert : fraction microsomale de levures exprimant CsZCD ; Rouge : débris cellulaires et les mitochondries de levures exprimant CsZCD.....173

Figure 48: Analyse HPLC de caroténoïdes extraits de souches de levure YSG50 exprimant CsZCD, CsZCD^{+N}, CsZCD^{+N+P} (spectres enregistrés à 440 nm), CsCCD et VvCCD1 (spectres enregistrés à 416 nm). Les levures YSG50 ont été transformées avec les constructions : Bleu : Z^{+N+P}/V185 ; Rose : Z^{+N}/V185 ; Noir : Z/V185 ; Bleu : C/V185 ; Vert : V/V185. Rouge : pCM185 vide.....174

Figure 49: Imbrication des différentes approches d'un projet d'ingénierie métabolique, depuis l'obtention d'enzymes fonctionnelles, jusqu'à l'intégration de la voie métabolique entière dans les réseaux métaboliques de la cellule hôte. La flèche bleue représente l'évolution des niveaux de complexité.181

Figure 50: Etapes mises en œuvre pour construire pour cloner les séquences CSZCD, CSZCD^{+N} ou CSZCD^{+N+P} dans les vecteurs pET15b, pBAD/TOPO® ThioFusion, YeDP60 ou pCM185. Restriction / ligation : réalisé avec le « Rapid ligation kit » selon les recommandations du fournisseur (Thermo Scientific, Waltham, MA) ; In-Fusion® = Kit commercialisé par Clontech (Mountain View, CA) utilisé selon les recommandations du fournisseur ; GA = Gibson Assembly Cloning Kit (NEB, Ipswich) utilisé en suivant les recommandations du fournisseur. PCM 185 publié dans (Gari et al., 1997).189

Liste des tableaux

Tableau 1: Exemples de composés d'intérêt produits par ingénierie métabolique chez <i>S. cerevisiae</i> . DHAP, Dihydroxyacétone phosphate ; Glucose 6-P, Glucose 6-phosphate ; GGPP, géranylgeranyl pyrophosphate ; Acétyl-CoA, Acétyl- Coenzyme A.....	34
Tableau 2: Exemples d'études d'ingénierie métabolique réalisées à l'aide d'expression de gènes sur plasmides multi-copies chez <i>S. cerevisiae</i>	38
Tableau 3: Analyse bibliographique de la production de caroténoïdes chez <i>Saccharomyces cerevisiae</i>	76
Tableau 4: Exemples de CCOs dans quelques organismes et leurs produits. ABA = acide abscissique.....	138
Tableau 5: Liste des cinq séquences protéiques les plus proches de CsZCD.....	146
Tableau 6: Prédiction de séquence d'adressage au chloroplaste (ChloroP). Length : longueur de la protéine (acides aminés) ; Score : valeur statistique sur laquelle la prédiction est basée ; cTP : prédiction (Y = oui, - = non) ; cTP-length : longueur de la séquence d'adressage identifiée à partir du premier acide aminé. Il doit être noté qu'un nombre apparaît dans la colonne cTP-length, même dans le cas où aucune séquence d'adressage n'est identifiée. Il ne faut alors pas en tenir compte.	148
Tableau 7: Description des constructions de biologie moléculaire utilisées dans la suite de cette étude. Ps, peptide signal ; N-ter, séquence N-terminale, (His) ₆ , tag histidine, Thio, thiorédoxine ; GST, glutathione-S-transférase. Les différentes longueurs de barres représentant les protéines reflètent les tailles respectives des enzymes testées.....	153
Tableau 8: Liste des souches de <i>E. coli</i> utilisées pour optimiser de la production de CsZCD sous forme soluble.....	157
Tableau 9: Méthodes utilisées pour obtenir les différents plasmides d'expression de l'étude. Restriction / ligation : réalisé avec le « Rapid ligation kit » selon les recommandations du fournisseur (Thermo Scientific, Waltham, MA) ; In-Fusion® = Kit commercialisé par Clontech (Mountain View, CA) utilisé selon les recommandations du fournisseur ; GA = Gibson Assembly Cloning Kit (NEB, Ipswich) utilisé en suivant les recommandations du fournisseur. pCM185 : publié dans (Gari et al., 1997). CSZCD a fait l'objet de deux amplifications successives pour être clonée dans pBAD/TOPO® ThioFusion afin d'obtenir des séquences de recombinaison plus longues.	188

Tableau 10: Conditions d’expression des protéines utilisées dans l’étude. Génotype de la souche CEN.PK 113-7D : (Mata his3—1 leu2-3,112 trp1-289 ura3-52 MAL2-8c). La souche YSG50 nous a été fournie par le Dr Huimin Zhao (Urbana, Illinois).	191
Tableau 11: Composition des tests d’activité <i>in vitro</i> réalisés après expression enzymatique chez <i>E. coli</i>	192
Tableau 12: Composition du test d’activité <i>in vitro</i> réalisé sur les différentes fractions de levures récupérées après cassage manuel. Volume final : 200 µL	194
Tableau 13: Amorces utilisées pour la construction de vecteurs bactériens. Rouge : site NdeI. Vert : site Bpu1102I. Bleu : base T ajoutée pour reproduire la construction qui aurait été obtenue avec le pBAD/TOPO Thiofusion ® expression kit (TA cloning). Les séquences correspondant à ces zones de recombinaison pour clonage sont soulignées.	223
Tableau 14: Amorces utilisées pour la construction de vecteurs de levure. Gras : Séquence Kozak (Kozak, 1981). Gras et italique : codon STOP supplémentaire.	223

Introduction

CHAPITRE 1

**L'ingénierie métabolique dans les biotechnologies -
L'implication de *Saccharomyces cerevisiae*.**

I. LES MICROORGANISMES, CATALYSEURS DES BIOTECHNOLOGIES – ESSOR DE LA BIOLOGIE DE SYNTHÈSE.

1. Introduction aux biotechnologies.

Les biotechnologies se situent au carrefour des sciences fondamentales et appliquées. Le préfixe « bio » signifie l'étude du monde vivant, et le terme « technologies » implique le développement et l'utilisation de méthodes et outils. Les biotechnologies sont donc définies comme un ensemble de techniques visant à piloter en laboratoire des applications du vivant. Plus précisément, l'OCDE (Organisation de Coopération et de Développement Economiques) définit les biotechnologies comme « l'application des principes scientifiques et de l'ingénierie à la transformation de matériaux par des agents biologiques pour produire des biens et services ».

Les biotechnologies sont utilisées depuis l'antiquité (fabrication de pain, de bière ou de vin). Cependant, on trouve les origines des biotechnologies telles qu'elles viennent d'être définies lors de la « seconde révolution industrielle », correspondant à l'application de la science aux nouvelles industries, entre 1870 et 1914 (Bud, 2001). La théorie des germes énoncée par le pionnier de la microbiologie, Louis Pasteur, au début des années 1860 a permis de développer des industries articulées autour de processus de fermentation caractérisés, optimisés et intégrés aux industries vinicoles, brassicoles, ou boulangères pour citer les exemples les plus connus. Les procédés biotechnologiques utilisant les microorganismes à une échelle industrielle se sont rapidement diversifiés. Un triste exemple a été donné lors de la première guerre mondiale, quand un procédé biotechnologique a été optimisé pour produire de l'acétone (jusque-là obtenu par distillation de bois) afin d'alimenter la fabrication d'explosifs (William Bunch, 2014). Un peu plus tard le premier antibiotique, la pénicilline, est découvert par Alexander Flemming à la fin des années 1920. Il n'est alors pas encore destiné à une utilisation massive sur l'homme, mais la mise au point de son procédé de purification à la fin des années 1930 par le groupe de Howard Florey, et les besoins en traitements accrus à cause de la seconde guerre mondiale, ont débouché sur sa production par fermentation du *Penicillium*, rendant le médicament disponible à travers le monde à partir de 1944. Ce n'est qu'après la seconde guerre mondiale que le terme de « biotechnologies » est utilisé pour décrire une nouvelle discipline scientifique. Le journal scientifique *Biotechnology and Bioengineering* voit le jour en 1961. L'idée de véritable virage technologique visant la production

de composés d'intérêt chez les microorganismes est alors bien présente dans les esprits scientifiques.

Au début des années 1970 le génie génétique émerge. Les techniques d'ADN recombinant (1973, Stanley Cohen and Herbert Boyer), la découverte des enzymes de restriction et de leurs applications aux problèmes de génétique moléculaire (1978, prix Nobel de médecine décerné à Werner Arber, Daniel Nathans et Hamilton O. Smith), puis l'invention de la Polymerase Chain Reaction, (PCR) par Kary Mullis en 1984 signent le début des biotechnologies « modernes », qui ciblent l'échelle moléculaire.

Les biotechnologies se divisent aujourd'hui en cinq classes selon leur domaine d'application, associées à des couleurs par la Communauté Européenne :

- Les biotechnologies rouges, avec des applications en médecine ;
- Les biotechnologies vertes, appliquées à l'agriculture ;
- Les biotechnologies bleues, appliquées à l'aquaculture ;
- Les biotechnologies jaunes, appliquées à l'assainissement des sols, traitement des eaux, épuration des gaz résiduels et de l'air, le recyclage des déchets ;
- Les biotechnologies blanches, appliquées à l'industrie. Ce sont des biotechnologies modernes permettant la production durable de matériaux, produits et carburants d'utilisation courante. Cela consiste en une transformation de matériaux renouvelables (bois, déchets de l'agriculture, CO₂...) par des organismes vivants et/ou leurs enzymes (Tang & Zhao, 2009).

Les biotechnologies blanches représentent des enjeux majeurs pour des industries aussi diverses que la santé, l'alimentation, l'agriculture ou encore l'énergie (Purnick & Weiss, 2009). Elles sont associées à une réduction de la consommation énergétique, des émissions de gaz à effet de serre et de la génération de déchets. Il est en effet souhaitable de trouver des alternatives durables à une industrie pétrochimique couteuse et en inadéquation avec les préoccupations environnementales actuelles (Stephanopoulos, 2007). De plus, on assiste à l'heure actuelle à un engouement vers une plus grande consommation de produits naturels, ou biologiques, le mot « naturel » étant systématiquement associé à l'idée de « bon pour la santé ». Par opposition, le rejet de tout ce qui peut être qualifié d'« artificiel », de « chimique », ou de « synthétique » est de plus

en plus systématique, car lié à l'idée de dangerosité potentielle du produit. La production d'un composé dans un organisme vivant puis sa purification ne faisant pas appel à des techniques de synthèse chimique, l'ingénierie des microorganismes pour en faire des « usines cellulaires » est couronnée d'un succès grandissant.

Parmi les exemples marquants de production de molécules d'intérêt chez les microorganismes, on trouve la production d'acide glutamique par *Corynebacterium glutamicum* (Gourdon & Lindley, 1999). L'acide citrique, avec un marché annuel mondial excédant les 1.5 milliard de dollars US, est produit par *Aspergillus niger* (Karaffa & Kubicek, 2003). La production de bioéthanol par *Saccharomyces cerevisiae* est à présent très développée, et représente le produit des biotechnologies industrielles le plus répandu dans le monde (Otero & Nielsen, 2010). La co-entreprise DuPont Tate et Lyle commercialise depuis 2006 du 1,3-propanediol (intermédiaire de synthèse de certains polymères utilisés dans le textile ou en cosmétiques) biosourcé, jusque là produit via l'industrie pétrochimique. Le glucose extrait de céréales est fermenté par *S.cerevisiae* pour produire du glycérol. Le glycérol est ensuite fermenté par *Klebsiella pneumoniae* pour former du 1,3 propanediol. On peut aussi donner l'exemple de l'acide succinique, produit dans l'organisme *Mannheimia succiniciproducens*. Ce microorganisme isolé depuis 2002 seulement est devenu, grâce au séquençage de son génome et à la reconstruction de réseaux métaboliques entre autres, un candidat pour une production d'acide succinique à l'échelle industrielle (Hong et al., 2004). De tels succès ont donné naissance à des entreprises comme METabolic Explorer (France, fondée en 1999), Genomatica (USA, fondée en 2000), Fluxome Sciences (Danmark, fondée en 2002), Amyris Biotechnologies (USA, fondée en 2003) ou encore Evolva (Suisse, fondée en 2004). Ces initiatives ont toutes en commun de chercher des réponses durables à des problèmes actuels par les biotechnologies basées sur les microorganismes.

2. La biologie de synthèse.

Stéphane Leduc (biologiste et chimiste français, 1853-1939) a écrit: « Quand on est arrivé à connaître le mécanisme physique de la production d'un objet ou d'un phénomène, (...) il devient possible (...) de reproduire l'objet ou le phénomène, la science est devenue synthétique. La biologie est une science comme les autres, (...) elle doit être successivement descriptive, analytique et

synthétique. » (Lecuc, 1912). En effet, le génie génétique, associé à une caractérisation exponentielle de nouveaux génomes, puis plus tard la biologie computationnelle (construction de modèles enzymatiques/structuraux/métaboliques), l'ingénierie des protéines (amélioration des capacités d'une enzyme pour obtenir le produit désiré), ou encore la génomique fonctionnelle (Otero & Nielsen, 2010; Tang & Zhao, 2009) ont ouvert les portes à la création de ce champ scientifique interdisciplinaire qu'est la biologie de synthèse, en plein essor depuis 2005 environ. Deux principales approches sont utilisées, « Top-Down » et « Bottom-Up » (Lee & Lee, 2013), avec des objectifs différents (Figure 1) :

- Une meilleure compréhension du vivant par la simplification artificielle de son fonctionnement et une observation en conditions contrôlées (de Vos et al., 2013). Il s'agit ici d'une application de la stratégie « Top-Down », qui est une déconstruction de systèmes moléculaires existants de façon à obtenir des organismes vivants simplifiés permettant une étude ciblées des éléments indispensables à la vie d'une cellule. En 2010, un génome entièrement synthétique de *Mycoplasma mycoides*, JCVI-syn1.0, a été construit en ne conservant que les gènes indispensables à la vie de la cellule en dehors de tout stress. Ce génome minimal a été assemblé et introduit dans une souche de *Mycoplasma capricolum*, donnant naissance à la souche *Mycoplasma laboratorium* (Gibson et al., 2010). Il s'agit du premier organisme contenant un génome intégralement reconçu et fabriqué par l'homme.
- L'approche « Bottom-Up », dans laquelle différents composants biologiques sont connectés de façon appropriée au sein d'organismes vivants pour y ajouter une fonction désirée. On utilise alors la notion de modules, ou de briques fondamentales, qui sont des éléments génétiques (séquences codantes, promoteurs, terminateurs, sites de liaison des ribosomes...) isolés. Une fois ces éléments assemblés de manière fonctionnelle dans un organisme, il est possible d'obtenir un système vivant nouveau, présentant les caractéristiques recherchées.

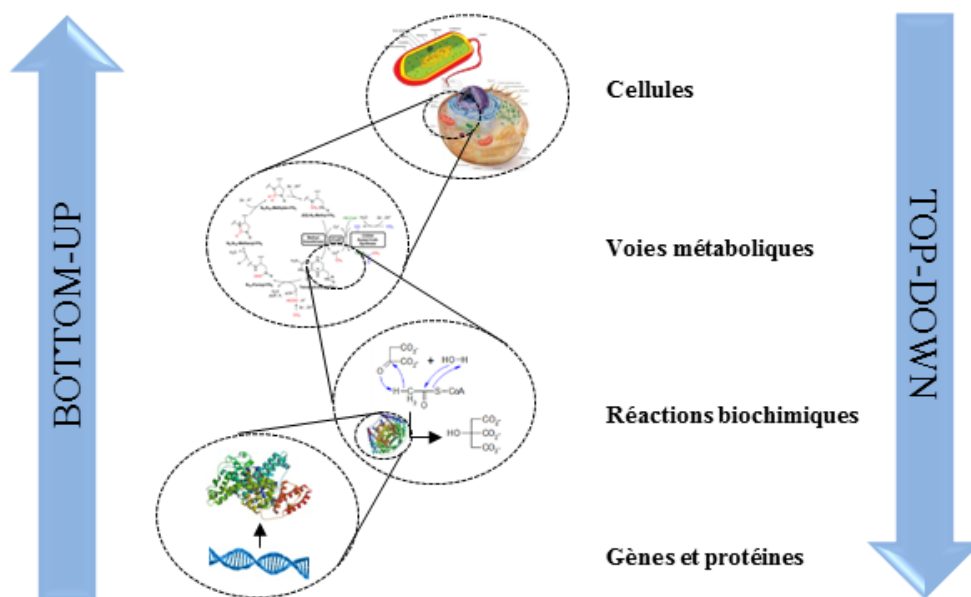


Figure 1: Schéma représentatif des approches de déconstruction (Top-Down) de systèmes existants, et de construction (Bottom-Up) de nouveaux systèmes vivants.

En relation avec l'approche Bottom-Up, le concept de standardisation de la biologie voit le jour entre la fin des années 1990 et les années 2000. Au Massachusetts Institute of Technology (MIT, Boston) et à l'Université de Berkeley Roger Brent, Robert Carlson, Drew Endy ou encore Adam Arkin proposent une biologie « intentionnelle » ou « constructive ». Des méthodes de sciences de l'ingénieur sont appliquées à la conception et à la construction de systèmes biologiques robustes, aux fonctions prévisibles. Le congrès Synthetic Biology 1.0 organisé au MIT en 2004 marquera l'acte de naissance « officiel » de la biologie synthétique contemporaine. A titre d'exemple, la fondation IGEM (International Genetically Engineered Machine, Cambridge, Massachusetts) organise depuis 2003 un concours d'étudiants. Le concept est basé sur un kit de composants biologiques appelés Biobricks, à partir duquel les étudiants doivent mettre au point une stratégie et un design de construction d'un système biologique nouveau et fonctionnel, à l'intérieur d'une cellule vivante.

II. L'INGENIERIE METABOLIQUE CHEZ *S. CEREVISIAE*.

Dans le contexte des biotechnologies blanches, une application phare de la biologie de synthèse est l'ingénierie métabolique. L'ingénierie métabolique a été décrite par Bailey en 1991, comme « une amélioration des activités cellulaires par la manipulation des fonctions enzymatiques, de transport et de régulation de la cellule en utilisant des technologies d'ingénierie génétique » (Bailey, 1991). Cette science permet d'optimiser les capacités naturelles de production d'un microorganisme, ou de rediriger une partie de ses flux métaboliques endogènes pour produire des composés qui lui sont étrangers.

Une voie métabolique est un ensemble de réactions chimiques se déroulant dans une cellule. Ces réactions sont catalysées par une série d'enzymes agissant de manière séquentielle, le produit de l'une étant le substrat de la suivante (Figure 2).

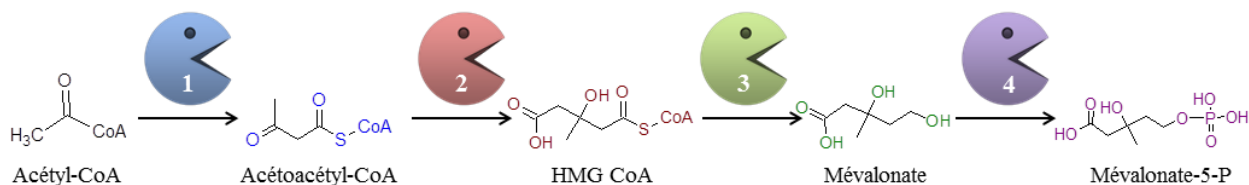


Figure 2 : Représentation schématique d'une voie métabolique. Exemple d'une partie de la voie du mévalonate de *S. cerevisiae*. 1 = Acétyl-CoA C-acétyltransférase; 2 = 3-hydroxy-3-méthylglutaryl-CoA (HMG-CoA) synthase; 3 = HMG-CoA réductase; 4 = Mevalonate kinase. Mévalonate-5-P = Mévalonate-5-phosphate.

Il existe des molécules d'intérêt pour lesquelles les techniques connues de synthèse chimique et/ou d'extraction à partir de producteurs naturels ne sont pas viables économiquement (Xu et al., 2013). Les producteurs naturels peuvent même être des espèces de plantes ou de microorganismes que nous sommes incapables de cultiver à l'heure actuelle. L'ingénierie métabolique offre alors la possibilité d'outrepasser ces limitations : il est possible de reconstruire entièrement une voie métabolique chez un hôte (on parlera alors d'expression hétérologue) sans danger pour l'homme, et dont la génétique et la physiologie sont connues.

Dans les biotechnologies industrielles, différentes contraintes doivent être réunies afin d'aboutir à un organisme viable, et capable de fournir le composé d'intérêt en quantités compatibles avec les besoins du marché. Certaines de ces contraintes s'appliquent au choix de l'hôte, ou « châssis » qui sera utilisé. Il doit présenter une capacité de croissance dans les conditions de culture appliquées, avec un coût maîtrisé (Keasling, 2012). C'est-à-dire qu'il doit être capable de synthétiser tous les éléments nécessaires à sa croissance à partir de simples sources de carbone d'azote, de soufre et de phosphore. Il doit s'agir d'organismes robustes afin de supporter des conditions de culture bien différentes de celles qui sont utilisées en laboratoire (échelle de culture passant souvent de la preuve de concept en fioles, aux fermenteurs de plusieurs centaines voire milliers de m³). Ils doivent enfin être bien caractérisés pour pouvoir bénéficier d'une ingénierie rapide et efficace. *S. cerevisiae* remplit toutes ces conditions.

S. cerevisiae s'avère être un hôte particulièrement prisé de la recherche, intérêt fortement lié au fait qu'il existe une connaissance approfondie de cet organisme tant au niveau physiologique que moléculaire et génétique (Nevoigt, 2008). Il s'agit du premier eucaryote à avoir vu son génome entièrement séquencé, annoté et rendu publiquement disponible (Goffeau et al., 1996). En 1997, la première puce rassemblant les ADNc de *S. cerevisiae* a abouti au développement d'une plateforme commerciale (Affimetrix, Santa Clara, USA) pour une étude approfondie de la régulation de ses gènes dans diverses situations. Un modèle métabolique de *S. cerevisiae* a été construit en 2003 (Forster et al., 2003) grâce à un ensemble de données génomiques, transcriptionnelles et biochimiques. Cette liste de données acquises à propos de *S. cerevisiae* n'est en aucun cas exhaustive, mais les connaissances accumulées au fil des recherches ont été répertoriées dans des bases de données telles que la *Saccharomyces* Genome Database (<http://www.yeastgenome.org/>), ou la Comprehensive Yeast Genome Database (<http://mips.gsf.de/genre/proj/yeast/>). Elles contiennent une quantité de données génétiques, biochimiques et métaboliques inégalée chez tout autre organisme eucaryote. S'ajoutant à ces dernières, d'autres bases mettent à disposition de la communauté scientifique des résultats issus d'études de transcriptome (<http://transcriptome.ens.fr/ymgv/index.php>), ou à propos de réseaux protéiques identifiés (General Repository of Interaction Datasets [<http://www.thebiogrid.org/>]). EUROSCARF (EUROpean *S. Cerevisiae* ARchive for Functional analysis; <http://web.uni-frankfurt.de/fb15/mikro/euroscarf/>) et la YGRC (Yeast Genetic Resource Center) collectent et conservent des souches et outils de biologie

moléculaire, qu'ils mettent à la disposition des généticiens de la levure partout dans le monde. Le « *Saccharomyces* Genome Deletion Project » (Winzeler et al., 1999) a développé une collection de souches inactivée dans 96% de leurs gènes, un par un, fournissant une base unique pour l'étude fonctionnelle du génome de *S. cerevisiae*. Cette base est mise à disposition par EUROSCARF.

L'utilisation répandue de *S. cerevisiae* dans des projets d'ingénierie métabolique est liée à cette connaissance approfondie de l'organisme, et est aussi incontestablement motivée par le fait qu'il s'agit d'un organisme classé GRAS (Generally Regarded As Safe) par la U.S. Food and Drug Administration (FDA). Cela en fait un moyen de production possible de toutes sortes de molécules, depuis les alcools, aux molécules à intérêt thérapeutique ou alimentaire.

Le rôle de *S. cerevisiae* dans les biotechnologies est ancestral avec les procédés de vinification, de panification ou de brassage. Dès l'apparition des biotechnologies blanches, *S. cerevisiae* s'est imposée grâce à sa haute efficacité naturelle à produire de l'éthanol à partir de sucres. En effet, l'éthanol est un carburant potentiel, capable d'apporter une contribution non négligeable aux demandes en énergies pour les transports, à moyen et à long terme. Son obtention par fermentation microbienne de ressources renouvelables est un enjeu majeur de notre ère (Nevoigt, 2008). La recherche en biotechnologies a, certes, su exploiter les capacités naturelles de *S. cerevisiae*, mais elle ne s'arrête pas là. Elle inclue également des projets d'ingénierie métabolique impliquant l'introduction de gènes provenant d'autres organismes de façon à lui conférer des caractéristiques nouvelles (Nielsen et al., 2013; Ukibe et al., 2009; Verwaal et al., 2007).

Les projets d'ingénierie métabolique effectués chez *S. cerevisiae* ciblent deux objectifs principaux: l'assimilation de nouveaux substrats d'une part, et la production de diverses catégories de molécules d'intérêt d'autre part (Ostergaard et al., 2000). Les principales sources de carbone entrant dans les procédés utilisant *S. cerevisiae* sont dérivées de l'amidon de maïs, du sucre de canne ou de betterave, ou de la lignocellulose (composant majeur des parois végétales). Les substrats sont convertis le plus souvent en divers métabolites primaires ou secondaires ou en protéines recombinantes. Un cas un peu particulier consiste à mettre à disposition des cellules une molécule qu'elles peuvent modifier biochimiquement pour ensuite en récupérer le produit : c'est la biotransformation *in vivo* (Mattanovich et al., 2014). Nous nous apprêtons maintenant à détailler

ces différentes applications de l'ingénierie métabolique chez *S. cerevisiae*, ainsi que les approches d'optimisation et de régulation qui peuvent être mises en œuvre.

1. Les Applications.

1.1 Production de métabolites.

La diversité des structures moléculaires synthétisables chez *S. cerevisiae* et leurs précurseurs sont remarquablement bien représentés dans les revues de Nevoigt, 2008, Tang & Zhao 2009, Otero & Nielsen 2010, Hong & Nielsen, 2012, ou encore Liu et al., 2013b. Ces molécules peuvent être des alcools, des dérivés de sucres, des acides organiques, des lipides, des terpènes, des molécules aromatiques ou encore des polycétides. La Figure 3 représente schématiquement différentes voies qui peuvent être exploitées pour produire des composés d'intérêt à partir du métabolisme de *S. cerevisiae*. Nous allons détailler ici les applications potentielles sur quelques exemples de composés ciblés.

L'anti-inflammatoire hydrocortisone (hormone stéroïde humaine) a été produit par l'introduction de non moins de huit gènes de mammifères chez *S. cerevisiae* (Szczebara et al., 2003). Avec l'artémisinine (Paddon et al., 2013), un isoprénoïde antipaludique, ce sont certainement les exemples les plus marquants de production de molécules bioactives chez *S. cerevisiae*. Les deux grandes familles des stérols et des isoprénoïdes sont des cibles privilégiées des biotechnologies chez *S. cerevisiae*. Ils sont utilisés comme colorants alimentaires ou antioxydants (caroténoïdes), arômes et parfums (terpènes), antiparasites et anticancéreux (Wriessnegger & Pichler, 2013). Ils sont dérivés de la voie de l'ergostérol (équivalent du cholestérol chez les animaux) de la levure, dont le précurseur est l'acétyl-Coenzyme A (acétyl-CoA).

Au-delà de l'exploitation de la voie de l'ergostérol, l'acétyl-CoA peut être qualifié de plaque tournante pour la production de molécules d'intérêt chez *S. cerevisiae*. Il est aussi le précurseur de voies métaboliques hétérologues d'acides gras ou d'acides organiques. Des acides gras, apports alimentaires indispensables chez l'homme, tels que l'acide stéaridonique (Kimura et

al., 2009) ou l'acide dihomog- γ -linolénique (Yazawa et al., 2007) ont été produits chez *S. cerevisiae*. Une attention particulière pour la synthèse d'acides organique est portée sur les levures grâce à leur forte tolérance aux pH acides. Cela présente en plus l'avantage de minimiser les risques de contamination des cultures de production par d'autres microorganismes. Par exemple, une souche de *S. cerevisiae* a été mise au point pour produire de l'acide malique (Zelle et al., 2008), utilisé comme conservateur alimentaire ou comme détartrant.

S. cerevisiae produit naturellement de l'éthanol. Le défi de l'ingénierie métabolique réside ici dans le développement de souches plus robustes et utilisant de nouveaux substrats (Mattanovich et al., 2014). Parmi les exemples les plus souvent cités dans la littérature sur le sujet se trouve la production d'éthanol à partir de substrats comme la lignocellulose, composant majeur des parois végétales (90% de la ressource globale de biomasse). Ce procédé repose essentiellement sur l'obtention de souches capables de métaboliser efficacement les pentoses et les hexoses qui composent la lignocellulose, comme le xylose (Matsushika et al., 2009; Van Vleet & Jeffries, 2009; Zaldivar et al., 2001). Cela dit, l'éthanol n'offrant qu'un rendement pauvre en énergie, d'autres alcools constituant de potentiels nouveaux carburants font l'objet d'essais de production chez *S. cerevisiae*. Le butanol et l'isobutanol sont des cibles privilégiées grâce à leur densité d'énergie proche de celle de l'essence, et à une hygroscopicité inférieure à celle de l'éthanol (Lan & Liao, 2013).

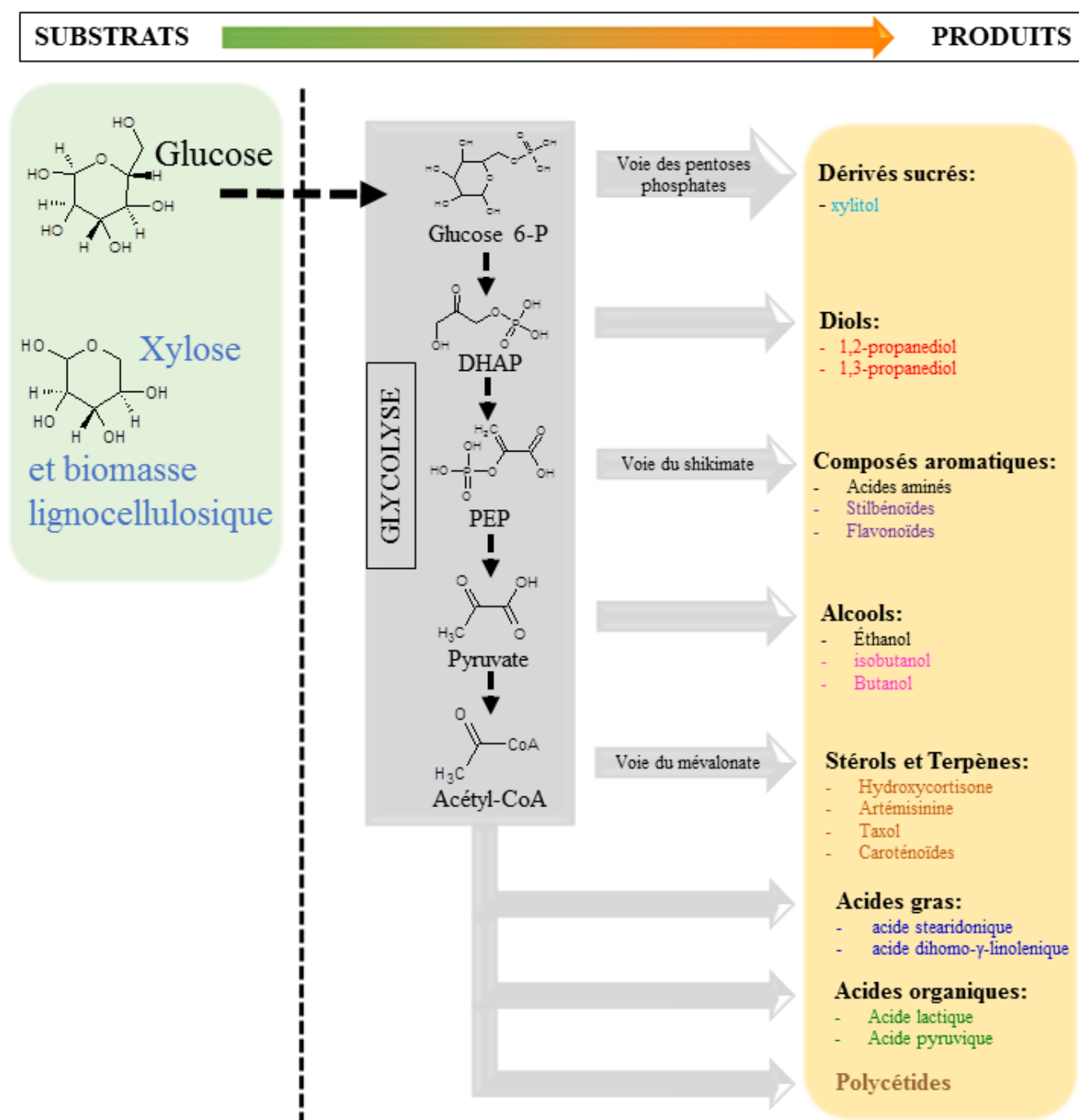


Figure 3: Exemples de projets d'ingénierie métabolique effectués chez *S. cerevisiae*. Les étapes naturellement présentes chez *S. cerevisiae* sont représentées en noir et blanc ; les substrats métabolisés après ingénierie, et les produits obtenus de façon hétérologue sont colorés. Glucose 6-P, Glucose 6-phosphate ; DHAP, Dihydroxyacétone phosphate ; PEP, Phosphoenolpyruvate ; Acétyl-CoA, Acétyl- Coenzyme A.

Nous venons de citer ici les exemples de composés produits par ingénierie métabolique chez *S. cerevisiae* les plus marquants, fondateurs pour certains (hydrocortisone, artémisinine, éthanol) de la discipline chez cet organisme. D'autres exemples sont rassemblés dans le Tableau 1.

Tableau 1: Exemples de composés d'intérêt produits par ingénierie métabolique chez *S. cerevisiae*. DHAP, Dihydroxyacétone phosphate ; Glucose 6-P, Glucose 6-phosphate ; GGPP, géranylgéranyl pyrophosphate ; Acétyl-CoA, Acétyl- Coenzyme A.

Application	Composé	Précurseur	Référence
Résines, détergents, pharmaceutiques, cosmétiques, antigels	1,2-propanediol	DHAP	Lee & Dasilva, 2006
Textiles	1,3-propanediol		Rao et al., 2008
Edulcorant	Xylitol	Glucose 6-P	Chen et al., 2010 (à partir de xylose)
			Toivari et al., 2007 (à partir de glucose)
Antioxydants et arômes	Resvératrol	Acide shikimique	Beekwilder et al., 2006
	Naringénine		Yan et al., 2005
	Vanilline		Hansen et al., 2009
	β -carotène	GGPP	Verwaal et al., 2007
Anticholestérolémians, pigments	Polycétides	Acétyl-CoA	Rugbjerg et al., 2013

1.2 Production de protéines recombinantes.

S. cerevisiae a été la première espèce de levure employée pour faire de la production de protéines recombinantes, au début des années 1980. La production hétérologue de protéines a avant tout un but thérapeutique, avec des applications surtout en immunothérapie. Les dites protéines visent à induire, augmenter ou réprimer une réaction immunitaire chez l'homme (Huang et al.,

2010). La production de fragments d'anticorps de camélidés chez *S. cerevisiae* est à présent un procédé industriel établi (Spadiut et al., 2014). Ces anticorps, ayant la particularité d'être dépourvus de chaîne légère, sont de parfaits anticorps miniaturisés à domaine de reconnaissance unique, beaucoup plus facile à produire chez les microorganismes que les grosses protéines au repliement complexe que sont les anticorps humains. Des protéines recombinantes peuvent également être administrées à des patients touchés par une déficience en une protéine donnée, compensant ainsi le manque en atteignant des concentrations systémiques ou tissulaires normales (Ferrer-Miralles et al., 2009).

Même si sur le marché actuel, une proportion non négligeable des cibles thérapeutiques obtenue par ingénierie dans les microorganismes provient de *S. cerevisiae*, il existe un certain nombre de limitations inhérentes à cet hôte. Il s'agit souvent de rendements faibles, de protéines hyperglycosylées ou encore de difficultés pour sécréter les protéines (Celik & Calik, 2012; Mattanovich et al., 2014). Ces limitations ont conduit l'industrie des biotechnologies à se tourner par la suite vers des hôtes alternatifs comme d'autres levures (*Pichia pastoris*, *Hansenula polymorpha*, *Yarrowia lipolytica*), et même des plantes ou des cellules d'insectes ou de mammifères (Celik & Calik, 2012). Parmi les 151 protéines recombinantes à but pharmaceutique autorisées par la FDA et l'Agence Européenne des Médicaments jusqu'en Janvier 2009, 18,5% étaient produites chez *S. cerevisiae*. Les 81,5% restants sont majoritairement produits chez *Escherichia coli* et dans des cellules de mammifères (Ferrer-Miralles et al., 2009). *S. cerevisiae* tirant tout de même avantage de son classement GRAS, le microorganisme peut être administré à l'homme. Dans le cas où les protéines recombinantes sont immobilisées à la surface de la cellule, *S. cerevisiae* peut devenir un vaccin oral, exprimant des antigènes à sa surface de façon à stimuler le système immunitaire (Schreuder et al., 1996).

1.3 Biotransformations *in vivo*.

Par contraste avec une production de métabolite *de novo* chez un hôte, le principe consiste ici à faire de la transformation de composés organiques ajoutés dans le milieu de culture, grâce aux activités enzymatiques possédées par les cellules, qu'elles soient endogènes ou hétérologues. Cela implique que le composé puisse entrer dans les cellules (nécessitant parfois une perméabilisation

des cellules, ou l'expression de transporteurs spécifiques), ou que les enzymes soient exprimées à leur surface. Des études ont abouti à la production de molécules d'intérêt à l'échelle industrielle par ce type d'approche (Pscheidt & Glieder, 2008).

Suivent ici quelques exemples de ce qui a été récemment fait chez *S. cerevisiae*. En 2001, une équipe est parvenue à synthétiser un biocarburant potentiel à partir d'huile de soja et de méthanol en passant par l'expression intracellulaire d'une lipase de *Rhizopus oryzae*, puis une étape de perméabilisation (Matsumoto et al., 2001). En 2002, la même équipe publiait cette fois un système de biotransformation par immobilisation de surface de la même lipase, parvenant à un résultat similaire en s'affranchissant de l'étape de perméabilisation (Matsumoto et al., 2002). Cette stratégie d'immobilisation de surface a ensuite été utilisée dans d'autres projets de biotransformations, comme pour la synthèse de polyesters (Tanino et al., 2009), ou de phospholipides et d'esters méthylés d'acides gras (Hama et al., 2010). En 2009, de la glucosylation de flavonoïde a été réalisée, par l'expression intracellulaire d'une glucosyltransférase hétérologue (Werner & Morgan, 2009).

2 Les outils utilisés pour la reconstruction de voies métaboliques chez la levure.

Pour produire des métabolites/protéines hétérologues, ou pour métaboliser de nouveaux substrats, la première étape consiste à exprimer les gènes d'intérêt. Plusieurs stratégies sont disponibles pour introduire et exprimer des gènes chez *S. cerevisiae* et elles dépendent du but recherché : expression du/des gènes la plus forte possible ou régulation fine de leur nombre de copies seront des facteurs déterminants du choix de la stratégie d'expression (Da Silva & Srikrishnan, 2012). Nous allons maintenant détailler les principales stratégies utilisées en faisant un point sur leurs avantages et inconvénients.

2.1 Les constructions plasmidiques.

Il s'agit de vecteurs navettes circulaires, capables de se répliquer à la fois dans *S. cerevisiae* et dans *E. coli*. Chez *S. cerevisiae*, pour permettre une expression des gènes transmissible depuis les cellules mères aux cellules filles, et chez *E. coli* pour faciliter toutes les opérations de clonage et d'amplification. La principale différence que l'on va trouver entre les constructions plasmidiques disponibles réside dans leur origine de répllication de levure, qui leur confère différentes caractéristiques. Certaines sont dotées d'une séquence CEN/ARS (Figure 4A), composée d'un centromère (CEN) et d'une origine de répllication (ARS), leur conférant une bonne répartition lors de la division cellulaire. Cela implique qu'ils sont stables au cours de la multiplication des cellules. Ces vecteurs sont retrouvés à 1 ou 2 copies par cellule. Les autres possèdent une origine de répllication de levure basée sur celle qui est trouvée dans le plasmide endogène 2 μ (Figure 4B). Ces vecteurs sont retrouvés à un plus grand nombre de copies par cellule (10 à 40 selon Romanos et al., 1992), mais le nombre exact ne peut pas être contrôlé et peut grandement varier d'une construction plasmidique à l'autre. Cependant, les plasmides 2 μ (multi-copies), grâce à leur nombre de copies élevé, permettent souvent une expression forte des gènes qu'ils portent (Siddiqui et al., 2012). Ils sont donc susceptibles d'offrir des niveaux de production des métabolites/protéines ciblés plus élevés que les plasmides CEN/ARS. Cela amène souvent les chercheurs à utiliser cette stratégie. Des exemples d'études d'ingénierie métabolique réalisées par une approche « plasmide multi-copies » chez *S. cerevisiae* sont rassemblés dans le Tableau 2.

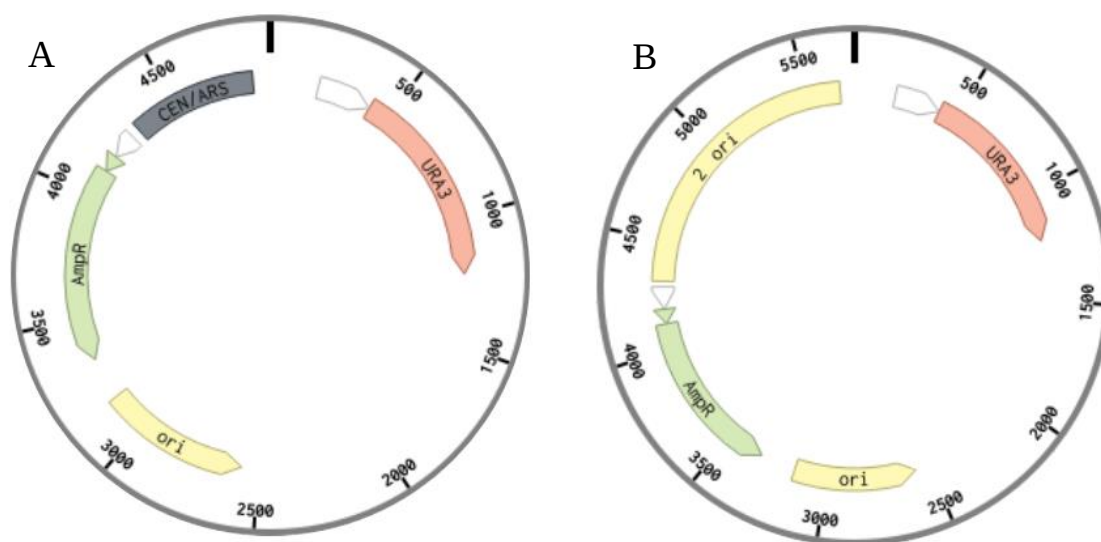


Figure 4: Eléments composant les constructions plasmidiques utilisées chez *S. cerevisiae*. (A) plasmide monocopie. (B) plasmide multicopies. CEN/ARS, centromère/séquence de répllication autonome ; *URA3*, marqueur de sélection par l'uracile ; *ori*, origine de répllication de *E. coli* ; *AmpR*, gène de résistance à l'ampicilline ; *2 ori*, origine de répllication 2 μ . Les flèches blanches représentent les promoteurs.

Une des limitations liées à l'utilisation des constructions plasmidiques est l'impossibilité de travailler dans des milieux de culture complets au profit de milieux sélectifs afin de garantir la conservation du plasmide au fil des générations (grâce à la présence d'un marqueur de sélection, Figure 4). Le coût de ces milieux n'est compatible qu'avec des procédés industriels visant des molécules à forte valeur ajoutée (Hahn-Hagerdal et al., 2005).

Tableau 2: Exemples d'études d'ingénierie métabolique réalisées à l'aide d'expression de gènes sur plasmides multi-copies chez *S. cerevisiae*.

Voie métabolique implémentée	Référence
Biosynthèse de l'acide artémisinique	Ro et al., 2006
Voie du méthylérythriol phosphate (voie des isoprénoides de <i>E. coli</i>)	Maury et al., 2008
Biosynthèse du β -carotène	Verwaal et al., 2007
Biosynthèse de l'astaxanthine	Ukibe et al., 2009
Biosynthèse du n-butanol	Steen et al., 2008
Biosynthèse du poly- β -hydroxybutyrate (polymère biodégradable)	Kocharin et al., 2012

Lorsqu'un projet d'ingénierie métabolique chez *S. cerevisiae* est initié, les auteurs cherchent d'abord à avoir une preuve de l'expression des gènes implémentés et de leur fonctionnalité dans l'hôte. Dans ce contexte, l'approche d'expression sur plasmide multi-copies est la plus simple et rapide à mettre en œuvre, avec un niveau d'expression des gènes potentiellement élevé. Ensuite, lors de futures optimisations des souches en vue d'une application industrielle, des stratégies plus adaptées doivent être implémentées.

2.2 Les YACs.

Les YACs (Yeast Artificial Chromosomes) sont des chromosomes artificiels de levures composés d'une origine de réplication, d'un centromère et de télomères (Figure 5), qui peuvent être linéaires ou circulaires. Décrits pour la première fois par Murray & Szostak (1983), ils ont alors été utilisés pour caractériser les éléments nécessaires à la ségrégation des chromosomes naturels chez la levure. Ils sont retrouvés à bas nombre de copies dans les cellules et se comportent comme des chromosomes naturels lors de la mitose, ce qui présente l'avantage de pouvoir s'affranchir de l'utilisation d'un milieu de culture sélectif. Ces structures de grandes tailles sont capables de contenir des grands fragments d'ADN, c'est pourquoi ils se sont avérés utiles pour reconstruire des voies métaboliques entières, voire des génomes entiers (Da Silva & Srikrishnan, 2012).

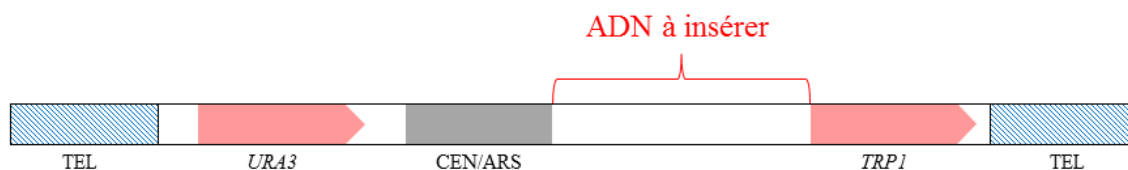


Figure 5: Eléments constituant un YAC. CEN/ARS, centromère/séquence de réplication autonome ; TEL, télomère ; *TRP1*, gène Tryptophane1 ; *URA3*, gène Uracile3 (marqueurs de sélection).

Une des approches de biologie de synthèse les plus marquantes de ces dernières années a vu le jour grâce à cette stratégie. Le génome synthétique de *Mycoplasma genitalium* a été

reconstitué sur un YAC en une seule étape à partir de 25 fragments isolés (Gibson et al., 2008). La capacité des YACs à contenir de grands fragments d'ADN a été exploitée pour isoler des morceaux d'ADN génomique allant jusqu'à 250 kilobases (Kouprina & Larionov, 2008). En 2009, des eYACs (expressible Yeast Artificial Chromosomes) ont été utilisés pour réaliser un assemblage aléatoire des gènes constituant des voies de biosynthèse de flavonoïdes (Naesby et al., 2009). L'équipe ayant effectué ces travaux est parvenue à obtenir des YACs de taille moyenne de 130 Kb, correspondant à environ 50 gènes. Avec une stratégie de biosynthèse *in vivo* totale (molécules d'intérêt synthétisées à partir de précurseurs endogènes de la levure), ils sont parvenus à un rendement maximal de 858 µg/L de naringénine. En comparaison, (Yan et al., 2005) ont implémenté chez *S. cerevisiae* une voie métabolique de biosynthèse de flavonoïdes sur un plasmide multi-copies. Par une stratégie de biotransformation (précurseurs ajoutés au milieu de culture), ils ont obtenu des rendements de 28.3 mg/L de naringénine. En se basant sur cet exemple, la stratégie d'expression d'une voie métabolique sur YAC est moins performante en termes de productivité que sur plasmide multi-copies.

Les YACs sont utiles en recherche, lorsque la taille de la construction finale est très grande (au-delà de 10 à 15 Kb), mais ils ne présentent pas de performances adaptées à la production de métabolites/protéines dans un contexte industriel. De plus, une pression de sélection est tout de même nécessaire pour garantir le maintien d'un YAC sur une culture longue (plusieurs dizaines de générations) (Siddiqui et al., 2012). Il n'y a pas, à notre connaissance d'exemple de production à l'échelle industrielle basée sur une stratégie d'expression sur YAC.

2.3 L'intégration chromosomique.

Le principe est d'intégrer les gènes d'intérêt dans le génome de la levure, en un site unique, ou en les répartissant au niveau de sites multiples. Il n'est plus nécessaire que le/les gène(s) d'intérêt soi(en)t accompagnés d'un « squelette » contenant les éléments nécessaires à la réplication et à la ségrégation de la construction artificielle (tel que c'est le cas sur les plasmides et les YACs) puisqu'ils constituent ici une partie du génome. Pour y parvenir, la machinerie de recombinaison de *S. cerevisiae*, très efficace, est utilisée. Les séquences à intégrer doivent être encadrées par des

séquences homologues du site cible dans le génome dans le but de provoquer des événements de crossing-over. Deux stratégies existent :

- Le/les gènes d'intérêt sont contenus dans un/des vecteurs d'intégration portant déjà des séquences homologues d'une zone dans le génome (Figure 6 A et B);
- Des produits de PCR sont générés par amplification du/des gènes d'intérêt, à la séquence desquels on ajoute les pieds de recombinaison (séquences homologues) voulus (Figure 6C).

Les vecteurs d'intégration ont la particularité de ne pas contenir d'origine de réplication de levure, c'est-à-dire que les séquences à intégrer ne peuvent être conservées dans les futures générations que *via* une intégration chromosomique. L'intégration peut nécessiter un seul événement de crossing-over si le vecteur est préalablement linéarisé à l'intérieur de la séquence homologue (Figure 6A). Une duplication de cette séquence dans la construction finale est alors inévitable. Par conséquent il est possible qu'un nouvel événement de recombinaison homologue entre ces séquences répétées soit observé, résultant en une potentielle instabilité de l'intégration. Le vecteur peut aussi être linéarisé de telle façon qu'un morceau de séquence homologue se trouve de part et d'autre du site de coupure (Figure 6B). Un double crossing-over est cette fois nécessaire, ce qui diminue l'efficacité d'intégration, mais confère à la construction ainsi obtenue une plus grande stabilité. Dans le cas où des produits de PCR sont intégrés, le design des amorces de PCR permet d'ajouter un morceau de séquence homologue de part et d'autre de la séquence à intégrer, nécessitant systématiquement une intégration *via* un double crossing-over (Figure 6C).

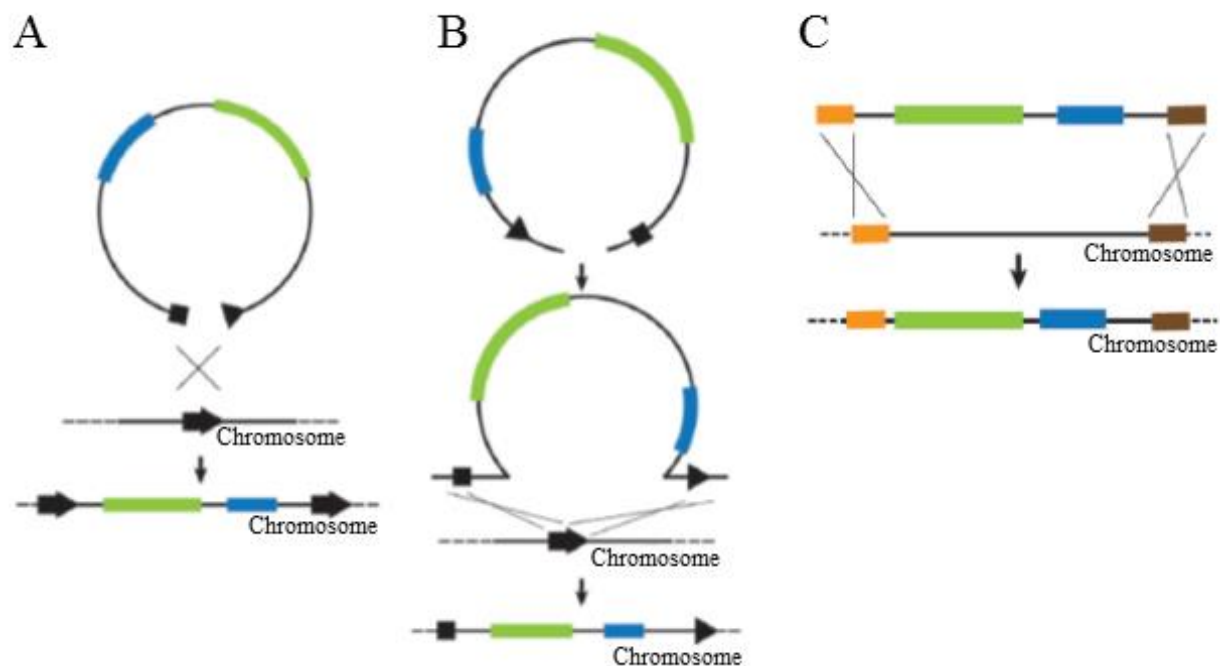


Figure 6: (A) Intégration chromosomique *via* un vecteur d'intégration par simple crossing-over. (B) Intégration chromosomique *via* un vecteur d'intégration par double crossing-over. (C) Intégration chromosomique d'un fragment d'ADN linéaire (produit de PCR) par double crossing-over. Figure adaptée de Da Silva & Srikrishnan 2012.

Une équipe a réalisé en 2007 une comparaison entre la stratégie d'expression sur plasmide multi-copies et la stratégie d'intégration génomique (Verwaal et al., 2007). La voie de biosynthèse du β -carotène (deux gènes) a été implémentée chez *S. cerevisiae* par les deux approches. Sur un plasmide multi-copies, 2 μg par gramme de masse sèche ($\mu\text{g/g}$ [dw]) de β -carotène ont pu être extraits des cellules mais une instabilité du vecteur a été remarquée, avec 10% des cellules qui perdaient leur vecteur au bout de trois jours de culture en milieu non sélectif. Après intégration des mêmes gènes, la stabilité de la voie de biosynthèse fut grandement améliorée (seulement 0,5% de perte après trois jours de culture en milieu non sélectif), mais cela s'accompagnait d'une diminution de la production de β -carotène (chiffre non communiqué par les auteurs).

Cette production amoindrie en condition d'intégration par comparaison avec la stratégie multi-copies peut être liée au nombre de copies des gènes, diminué dans un cas par rapport à l'autre (l'intégration était faite en un site unique). En effet, Lee & Dasilva (2006) ont remarqué une relation linéaire entre le niveau d'expression et le nombre de copies de deux gènes, intégrés au génome de

S. cerevisiae afin de produire du 1,2-propanediol. Un plus grand nombre de copies d'un ou plusieurs gènes intégrés peut résoudre le problème du manque de productivité. Une stratégie d'intégration multiple est basée sur les sites Delta (δ) de *S. cerevisiae*. Il s'agit de séquences terminales longues répétées (LTRs) trouvées aux extrémités des rétrotransposons de familles 1 et 2 de *S. cerevisiae* (Kim et al., 1998), et dispersées dans tout son génome. Cela permet de faire de l'intégration multi-copies, mais pas de maîtriser exactement le nombre de copies intégrées (Sakai et al., 1990; Shiomi et al., 1995), ni leurs positions. La production d'éthanol à partir de cellulose a été réalisée en 1999 en intégrant les gènes codant les enzymes impliquées dans la voie de dégradation de la cellulose aux locus δ (Cho et al., 1999), avec des niveaux d'expression et de production accrus en comparaison avec une expression plasmidique. Plus tard, pour optimiser les niveaux d'expression de gènes impliqués dans la même voie métabolique, Yamada et al. (2010) ont utilisé une approche appelée « cocktail δ -integration ». Trois gènes ont été intégrés aux sites δ de *S. cerevisiae* par trois séries d'intégration successives, en maintenant une pression de sélection sur un milieu contenant de la cellulose. Après une estimation du nombre de copies des gènes intégrés, cette stratégie leur a permis de déterminer le ratio optimal de chacun des gènes dans leurs conditions expérimentales, applicable ensuite à une industrialisation de la souche. Une équipe a intégré au niveau des sites δ du génome de la levure des voies métaboliques entières allant jusqu'à huit gènes en un seul bloc (Shao et al., 2009). Le nombre de copies des voies métaboliques intégré n'a pas été estimé, ni les niveaux de production obtenus mais ils sont parvenus à des souches qui synthétisent de la zéaxanthine (caroténoïde) à partir de xylose.

L'assemblage, puis l'intégration ciblée au génome de *S. cerevisiae* de voies métaboliques à étapes multiples dans un but d'application industrielle reste un défi de la recherche à ce jour (Kuijpers et al., 2013). Kuijpers et al. (2013) ont identifié l'évènement d'intégration d'une voie métabolique composée de sept gènes comme l'étape limitante de la méthode (seulement un transformant sur vingt contenait la voie métabolique entière intégrée au locus cible). Pour améliorer le rendement de cette étape, une stratégie de coupure double-brin de l'ADN génomique au niveau du site cible a été appliquée. Un assemblage de la voie métabolique par recombinaison *in vivo* couplé à l'expression d'une endonucléase site-spécifique a abouti à une efficacité d'intégration à hauteur de 95%. Il s'agit ici du premier succès d'intégration chromosomique site-spécifique d'une voie métabolique complète, en une seule étape, chez *S. cerevisiae*. Aujourd'hui, l'ingénierie des

endonucléases se développe avec des approches autour des « zinc-fingers nucléases » (Kandavelou et al., 2005), des TALENs (Christian et al., 2010), ou plus récemment du système CRISPR/Cas (Cong et al., 2013; DiCarlo et al., 2013). Ces endonucléases peuvent être synthétisées à façon à partir d'une séquence d'ADN cible (voir les activités de l'entreprise CELLECTIS), offrant une très large variété de sites d'intégration potentiels. Ces méthodes mènent à l'intégration d'une seule copie des gènes constituant la voie métabolique d'intérêt. Pour multiplier le nombre de copies, l'utilisation de plusieurs endonucléases serait nécessaire.

2.4 L'émergence de nouveaux outils d'assemblage.

Quelle que soit la stratégie d'expression choisie parmi celles qui viennent d'être décrites, la (re)construction de voies métaboliques commence par l'assemblage des fragments d'ADN d'intérêt. Ces fragments d'ADN d'intérêt sont les gènes à exprimer (inserts), et les vecteurs qui les contiennent si une stratégie d'expression sur vecteur est choisie. Cet assemblage se décompose en trois niveaux :

- La construction des cassettes d'expressions, ou unités transcriptionnelles (3 éléments) : (1) un promoteur contrôlant l'expression d'une (2) séquence codante (ORF), dont la transcription sera stoppée par la présence d'un (3) terminateur.
- L'assemblage des différentes cassettes d'expression, si la voie métabolique ciblée est composée de plusieurs gènes.
- Le clonage de ces dernières dans un/des vecteur(s) si nécessaire.

Aujourd'hui, les nouvelles voies métaboliques implémentées en biotechnologies pour la production de composés d'intérêt sont de plus en plus complexes et constituées d'étapes de plus en plus nombreuses. Il est donc devenu nécessaire de développer des méthodes d'assemblage offrant fiabilité et rapidité du clonage de fragments d'ADN multiples.

Les méthodes classiques de biologie moléculaire consistent à cloner un gène d'intérêt dans un vecteur (contenant souvent un promoteur et un terminateur) *via* un système de restriction/ligation. Ce système se montre aujourd'hui trop limité pour exprimer des voies métaboliques multi-étapes. En effet, le clonage par restriction nécessite que les sites de coupure

utilisés ne soit pas présent dans les inserts, ce qui est une condition fatalement plus difficile à respecter lorsque les inserts sont nombreux (Wang et al., 2012). Des stratégies ont émergé pour surmonter ces difficultés, en développant des vecteurs pourvus de sites de restriction rares (Wakamori et al., 2010). Cependant, plus il y a de gènes à cloner, plus il faut utiliser d'enzymes différentes. Soit les gènes sont clonés un à un, ce qui est consommateur de temps, soit toutes les enzymes sont utilisées en même temps pour cloner tous les gènes en une seule étape, ce qui diminue grandement l'efficacité de clonage. Il a été démontré que l'enzyme de restriction *SfiI* était intéressante pour cloner des fragments multiples grâce à son site cible dégénéré (Lu et al., 2007). La stratégie proposée par ces travaux est de générer des sites de restriction *SfiI* aux extrémités de chaque insert (qui peuvent être différents étant donné le caractère dégénéré de cette séquence) par des clonages intermédiaires dans des vecteurs pourvus de ces sites de restriction, et de les cloner en une seule étape dans un vecteur (Figure 7). Cette méthode permet le clonage multiple avec une seule enzyme, mais oblige à effectuer des clonages intermédiaires.

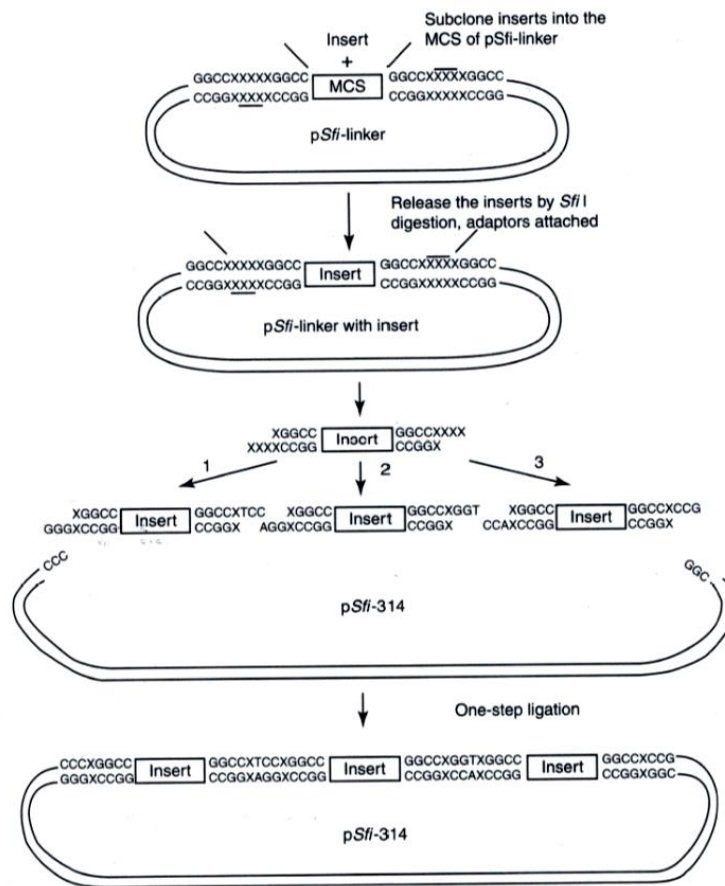


Figure 7: Schéma de la méthode de clonage multiple par l'enzyme de restriction *Sfi*I (Lu et al., 2007).

Les techniques d'assemblage par recombinaison homologue sont aujourd'hui privilégiées. La recombinaison peut se produire en conditions *in vitro* ou *in vivo*. La technique du clonage indépendant de la ligation, basée sur la génération de courts fragments homologues par PCR, a été décrite en 1990 (Aslanidis & de Jong, 1990). Le principe est de générer des séquences homologues aux extrémités des fragments à assembler en les ajoutant aux amorces de PCR. Après un traitement par une exonucléase, des extrémités simples brins sont générées. Les extrémités homologues simples brins peuvent alors s'hybrider pour associer le fragment d'ADN à cloner et le vecteur (Figure 8A). La recombinaison homologue *in vitro* est particulièrement bien illustrée par la méthode isotherme développée par Gibson et al. (2009) (kit commercialisé par New England Biolabs). Cette méthode repose sur la génération de produits de PCR possédant des séquences homologues au vecteur, puis sur l'action concertée d'une exonucléase 5' qui va générer des

extrémités chevauchantes, d'une ADN polymérase et d'une ADN ligase, qui vont ensuite combler les zones manquantes et reconstituer un vecteur circulaire complet (Figure 8B). La méthode SLICE (Zhang et al., 2012b) abouti au même résultat sans que chaque enzyme nécessaire soit purifiée. Un lysat bactérien est utilisé, contenant tous les éléments nécessaires de la recombinaison, jusqu'à la réparation. Dans la technologie In-fusion® (kit commercialisé par Clontech), on utilise l'endonucléase du virus de la vaccine, mais pas de polymérase ni de ligase. La machinerie de réparation de *E. coli* recircularise *in vivo* le vecteur (Irwin et al., 2012).

La PCR permet aussi de fusionner des fragments d'ADN afin de reconstituer des cassettes d'expression entières (Horton et al., 1989). Une première réaction de PCR génère des extrémités homologues grâce aux amorces. Pendant une deuxième PCR, ces extrémités homologues s'hybrident entre elles, et c'est la construction entière qui est amplifiée (Figure 9). C'est la première phase de la méthode de « DNA assembler » qui est basée sur un pré-assemblage de « briques » (une brique = un promoteur + une séquence codante + un terminateur) par PCR, suivi d'un assemblage final des briques générées par recombinaison *in vivo* chez *S. cerevisiae* (Shao et al., 2009). Pour faire de la recombinaison *in vivo*, *S. cerevisiae* est en effet un hôte privilégié grâce à sa machinerie de recombinaison particulièrement efficace (Gibson, 2009; Larionov & Kouprina, 1996; Naesby et al., 2009).

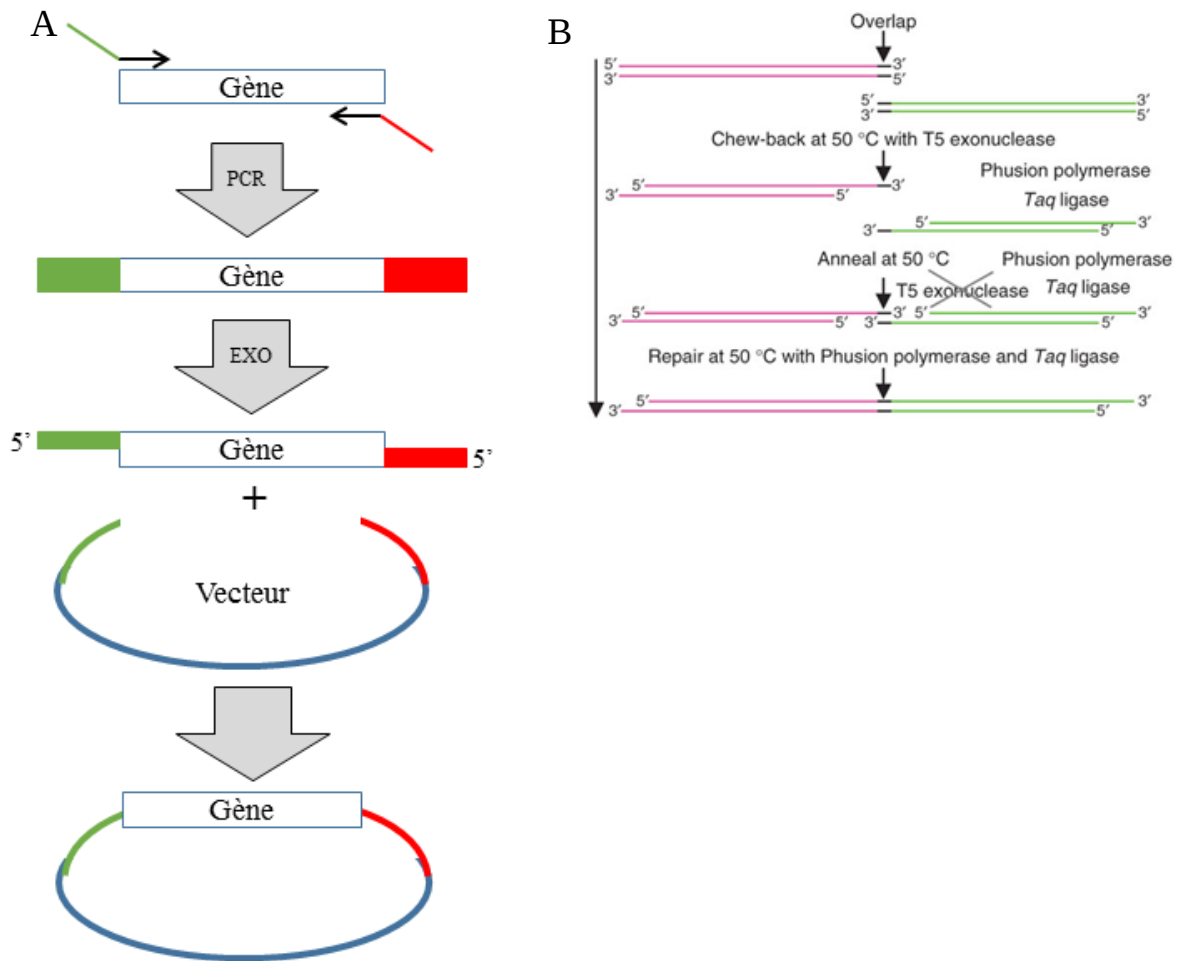


Figure 8: Méthodes d'assemblage de fragments d'ADN. (A) Méthodologie suivie pour assembler des fragments d'ADN par recombinaison en générant des séquences homologues par PCR. EXO, Exonuclease 3' – 5'. (B) Méthode Gibson et al. 2009.

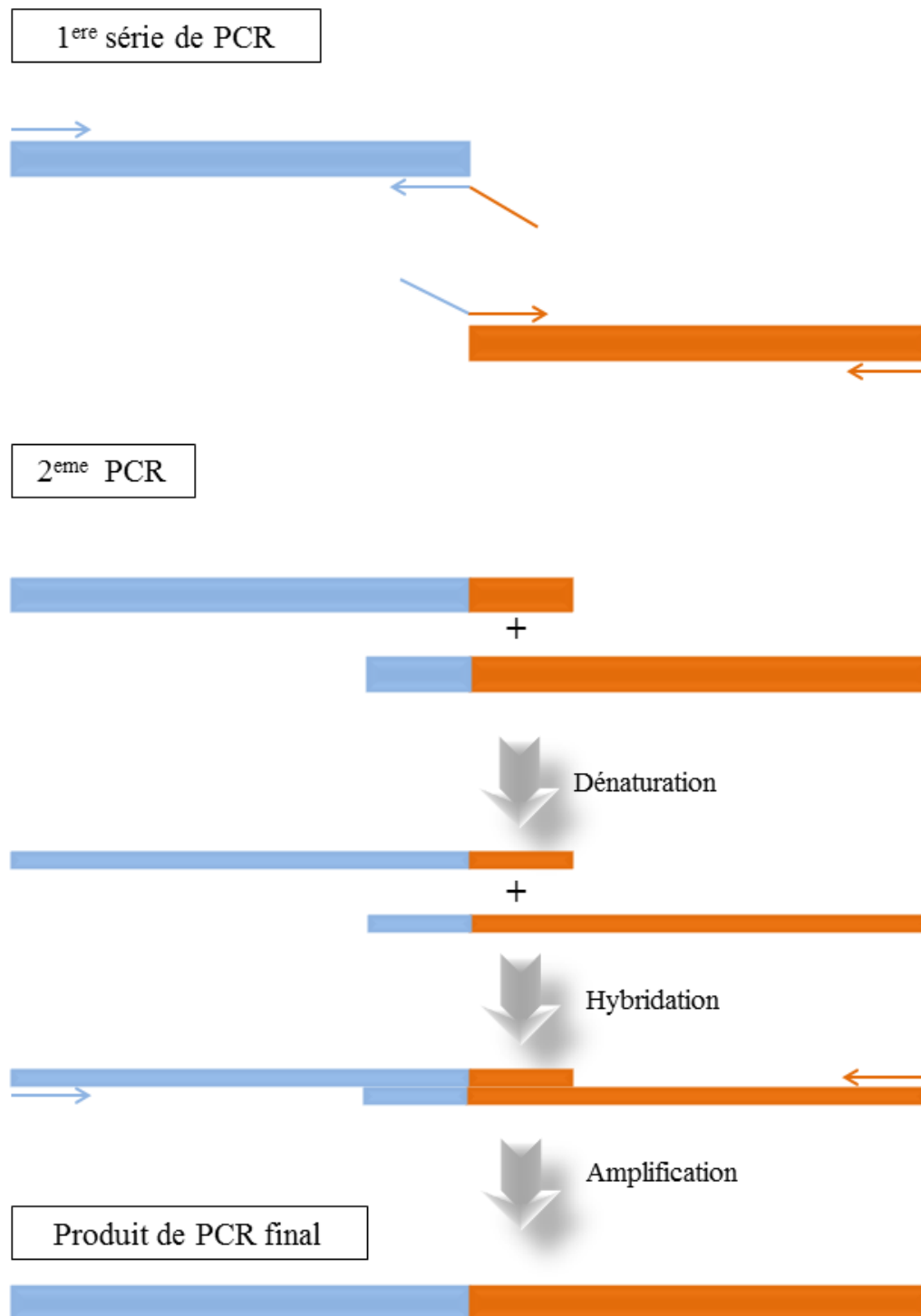


Figure 9: Principe de fonctionnement de la PCR fusion. Plusieurs PCRs sont réalisées en parallèle de façon à obtenir chaque fragment à assembler pourvu de séquences homologues de ses séquences partenaires. Une deuxième PCR est ensuite réalisée avec tous les fragments à assembler et les amorces s'hybridant aux extrémités de la construction finale.

Ces méthodes, offrant un plus grand degré de liberté que la restriction/ligation, nécessitent tout de même une linéarisation du vecteur par restriction. Cependant, une étude a démontré qu'on pouvait éviter cette étape, en procédant ainsi pendant la PCR : une première réaction de PCR génère l'insert allongé de séquences homologues au vecteur, puis une deuxième utilise le produit de la première comme amorce, à partir de laquelle le vecteur va être amplifié, puis refermé *in vivo* dans *E. coli* (Figure 10) (Lund et al., 2014).

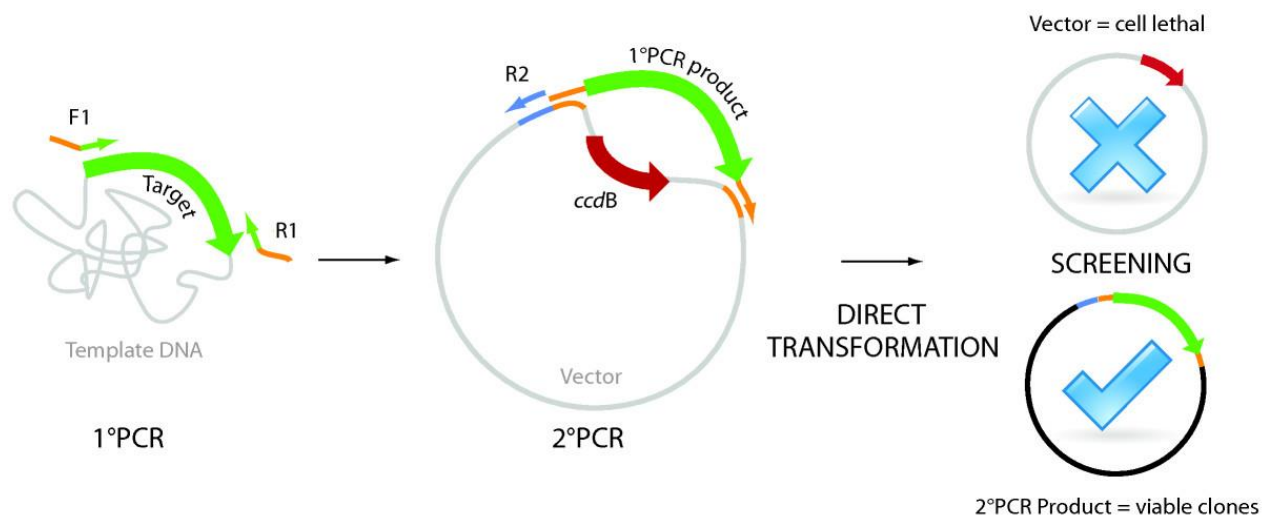


Figure 10: Méthode Lund et al. 2014. Sur cette figure, le vecteur dans lequel la séquence d'ADN cible doit être intégrée possède un gène *ccdB* situé entre les deux bornes d'hybridation du premier produit de PCR. Ce gène code pour une protéine « poison » de la bactérie *E. coli*. Si le produit de PCR n'a pas correctement « pris la place » du gène *ccdB*, la cellule de *E. coli* qui portera le plasmide résultant ne poussera pas. Il s'agit d'une stratégie d'évitement des faux positifs.

2.5 Les stratégies combinatoires.

L'ingénierie qualifiée de rationnelle se base sur les informations disponibles à propos de la voie métabolique concernée par le projet de recherche : les enzymes qui la composent, et leur mode de régulation. Les étapes sont incrémentées de façon séquentielle et itérative, par une suite d'« essai-erreur » de façon à identifier, à chaque étape, la stratégie d'optimisation à mettre en œuvre. Cette méthode est couteuse et chronophage, particulièrement lorsque les étapes de la voie métabolique ciblée sont nombreuses. De plus, le « catalogue » de composés d'intérêt pouvant être produits par ingénierie métabolique *via* cette méthode est fatalement limité par l'état des connaissances à propos de la voie métabolique ciblée. Des stratégies combinatoires, pouvant être

définie de manière large par une génération de diversité, ont été développées pour apporter de nouvelles solutions à la demande de nouvelles voies métaboliques multi-étapes pour la production de composés d'intérêt.

Le « genome shuffling » est une approche d'évolution de souches accélérée pour aboutir à des microorganismes « améliorés » pour le caractère recherché (Stephanopoulos, 2002). Des souches sont d'abord générées par une technique d'amélioration classique comme la mutagenèse aléatoire par les rayons UV, et sélectionnées pour le caractère recherché. Ces souches, dites parentales, sont ensuite croisées et les souches filles résultant de ces croisements sont elles-mêmes sélectionnées pour le caractère recherché. Plusieurs cycles de croisement/sélection sont ainsi répétés (Figure 11) jusqu'à obtenir l'amélioration souhaitée. La méthode a prouvé son efficacité chez *S. cerevisiae*, avec une amélioration de la tolérance à l'éthanol (Hou, 2009; Shi et al., 2009), ou aux inhibiteurs trouvés dans les « SSL » (« Spent Sulfite Liquor »), résidus de l'industrie papetière dont les sucres peuvent être fermentés en éthanol (Pinel et al., 2011). Le « genome shuffling » fait partie des techniques d'amélioration aléatoire : aucun ciblage des modifications à apporter (gènes et positions des mutations) n'est fait en amont, et elles ne peuvent être connues que par un séquençage des souches sélectionnées en aval.

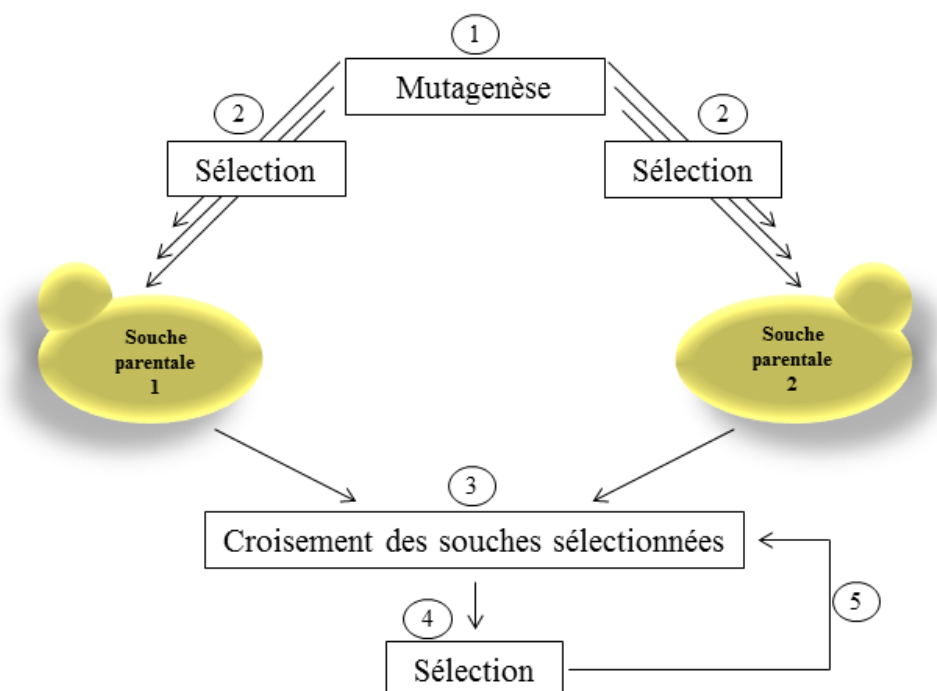


Figure 11: Etapes composant la méthode de « genome shuffling ».

D'autres méthodes combinatoires impliquent la construction de banques de voies métaboliques, en faisant varier certains paramètres. Il est possible de faire varier la nature des gènes implémentés, leur nombre, leur position ou encore la nature des séquences de régulation de l'expression (promoteurs, sites de liaison des ribosomes, facteurs de transcription...). Les nombreuses stratégies d'assemblage de fragments d'ADN précédemment décrites, en particulier les méthodes de recombinaison homologue, ont rendu possible la construction rapide d'un grand nombre de voies métaboliques en parallèle. Cela autorise les chercheurs à cibler le/les paramètre(s) qu'ils souhaitent faire varier un à un pour aboutir à la « version » de la voie métabolique la plus efficace pour le caractère recherché. Naesby et al. (2009) ont implémenté des voies de biosynthèse de flavonoïdes, dont les gènes provenant d'« organismes sources » divers ont été assemblés aléatoirement par recombinaison *in vivo* chez *S. cerevisiae*. Cette combinatoire variant la nature des enzymes leur a permis d'observer de grandes différences en termes de nature des composés produits et de quantités. La méthode du « DNA Assembler » (Shao et al., 2009) a été utilisée pour générer une banque de plus de 8000 combinaisons d'enzymes aux propriétés différentes, afin d'optimiser la voie de consommation du xylose chez *S. cerevisiae* (Kim et al., 2013).

Les séquences de régulation de l'expression des gènes sont également fréquemment ciblées par les auteurs d'études combinatoires. Parmi elles, les promoteurs font souvent l'objet de variations (les objectifs et les stratégies appliquées à propos de la régulation du niveau d'expression par les changements de promoteurs sont détaillés plus loin dans ce document). La technique du clonage par l'enzyme *SfiI* (Lu et al., 2007) a été utilisée pour mettre au point la méthode que les auteurs ont baptisé « multiple-gene-promoter shuffling » (Lu & Jeffries, 2007). Lors de ces travaux, des promoteurs constitutifs de *S. cerevisiae*, de forces connues grâce à une étude transcriptomique (Jin & Jeffries, 2004), ont été utilisés pour construire une voie métabolique à trois gènes par une approche combinatoire. Chaque gène a été associé à chaque promoteur. Ils ont observé qu'une expression forte de tous les gènes impliqués n'était pas souhaitable, mais qu'un équilibre devait plutôt être trouvé entre tous les gènes de façon à consommer le plus efficacement possible le flux métabolique entrant dans la voie. La stratégie « COMPACTER » (Du et al., 2012), basée sur la méthode d'assemblage « DNA Assembler », utilise des promoteurs mutés ayant développé des forces variées. Ils sont placés devant chaque gène d'intérêt de manière combinatoire. L'équipe à l'origine de la méthode a construit une combinatoire dans deux souches différentes de *S. cerevisiae*. Ils ont observé une différence dans les ratios optimums de chaque gène selon la souche d'expression. Lors de travaux plus récents, le niveau d'expression de chaque gène impliqué dans une voie métabolique d'intérêt (ici au nombre de huit) a été régulé en associant une approche combinatoire de promoteurs et une amélioration évolutive des souches (Latimer et al., 2014). La banque métabolique combinatoire construite a été soumise à environ quinze cycles de culture (avec dilution des dilutions de la culture tous les 1 à 4 jours) avec une pression de sélection pour le caractère recherché. Un génotypage des souches a ensuite permis de déterminer quelle condition était la plus adaptée à leurs critères.

La nature des promoteurs n'est qu'un des aspects de la régulation de la transcription d'un gène. Les stratégies d'ingénierie combinatoire peuvent créer de la diversité métabolique en ciblant les facteurs de transcription plutôt que les promoteurs. La technique de gTME (global Transcription Machinery Engineering) (Alper et al., 2006; Alper & Stephanopoulos, 2007) est une reprogrammation de la machinerie de transcription de la cellule basée sur l'introduction de mutations dans la région de liaison à l'ADN d'un facteur de transcription. Cela modifie son affinité pour différents promoteurs, perturbant le profil d'expression de la cellule sur un plan global.

Aujourd'hui, les stratégies combinatoires font preuve d'un niveau de complexité de plus en plus élevé pour aboutir à des souches optimisées de plus en plus performantes, voire à la production de composés non naturels. Une équipe a démontré qu'il était possible de générer chez *S. cerevisiae* une grande diversité de voies de biosynthèse par recombinaison homéologue (entre séquences proches mais non identiques) (Luque et al., 2014). Un assemblage puis une intégration chromosomique de diverses voies de biosynthèse de flavonoïdes a été réalisé *in vivo* en utilisant des gènes de plantes non identiques. Le fait de travailler dans des souches de *S. cerevisiae* déficientes dans leur système de réparation de l'ADN a permis de construire des assemblages de grande taille (23 Kb) à partir de séquences homéologues. Cela a généré des voies métaboliques « mosaïques » présentant des fonctionnalités et des taux de production variés. La génération de diversité chez la levure a récemment été poussée jusqu'à la découverte de nouveaux composés biologiquement actifs (Klein et al., 2014).

Une fois la banque combinatoire créée et implémentée, un criblage adapté au caractère recherché doit être développé pour sélectionner les souches répondant le mieux aux critères demandés. La principale limite des stratégies combinatoires est la nécessité de criblage haut débit à la suite de la génération d'une grande diversité. Les critères de sélection des candidats ne s'adaptent en effet pas toujours à une échelle haut débit. On détecte par exemple souvent la présence d'une molécule d'intérêt par des techniques de biochimie classiques du type HPLC (High Performance Liquid Chromatography) ou GC (Gaz Chromatography), nécessitant une extraction préalable du produit et dont l'utilisation sur une population importante d'échantillons est difficile à mettre en œuvre. Cela amène souvent les auteurs à penser leur stratégie en fonction des possibilités de criblage qui s'offrent à eux, limitant par conséquent les potentialités des stratégies combinatoires. L'équipe de Dueber a récemment proposé de procéder à une première sélection des candidats par une méthode de modélisation, pour considérablement diminuer le nombre d'échantillons à tester et ainsi s'affranchir du besoin de crible haut débit (Lee et al., 2013b). Pour cela, ils ont caractérisé cinq promoteurs de *S. cerevisiae*. Après avoir construit des banques combinatoires, puis génotypé leurs souches, ils ont pu mettre au point un modèle reliant le génotype et les niveaux de production. Ils ont ensuite appliqué ce modèle à la voie de biosynthèse de la violacéine, construisant la première souche de *S. cerevisiae* produisant cette molécule.

L'avantage majeur des stratégies combinatoires en ingénierie métabolique est la possibilité de construire une voie métabolique fonctionnelle dans l'hôte sans bénéficier de toutes les informations concernant la nature des enzymes ou leurs niveaux d'expression optimaux. Ces stratégies permettent de générer une diversité suffisamment grande, en un temps suffisamment limité pour tester toutes sortes d'optimisations de la voie et ensuite sélectionner la meilleure.

3 Les stratégies de régulation de l'expression de gènes.

Le principal défi des biotechnologies réside dans les taux de production de la molécule recherchée, qui doivent répondre aux besoins du marché. Les chercheurs sont donc régulièrement confrontés à la nécessité d'optimiser la productivité de leur système biologique. Dans ce contexte, des stratégies comme la surexpression de précurseurs (Ajikumar et al., 2008) ou la minimisation des voies endogènes compétitrices (Ohto et al., 2010) ont été mises en œuvre. Nous venons également de voir que les stratégies combinatoires constituent une avancée majeure dans ce contexte, avec une attention particulière donnée à la régulation de l'expression des différents gènes implémentés. Il s'agit en effet d'un aspect central de l'ingénierie métabolique.

Une expression non régulée d'enzymes hétérologues dans un hôte implique le détournement d'une partie du flux métabolique naturel de la souche vers la voie synthétique pouvant mener à l'apparition d'un stress, à une surconsommation des ressources cellulaires, et possiblement à une accumulation d'intermédiaires toxiques (Glick, 1995; Neubauer et al., 2003). Il en découle directement la nécessité d'ajuster les niveaux d'expression des enzymes hétérologues, évitant les effets indésirables sur la viabilité cellulaire (Ajikumar et al., 2010; Lu & Jeffries, 2007; Ramon & Smith, 2011). Un des moyens de limiter le stress induit, et donc de construire un organisme qui produit efficacement un composé d'intérêt, est de parvenir à moduler le plus finement possible, et au cas par cas, les niveaux d'expression des enzymes impliquées. En considérant les trois grandes étapes de la synthèse d'une enzyme comme des points d'action potentiels, nous pouvons détailler les différentes stratégies possibles en trois grands groupes de cibles: la molécule d'ADN qui code pour les enzymes, l'étape de synthèse de l'ARN messager qui lui correspond, et enfin la traduction de l'ARN en protéine.

3.1 Points de contrôle au niveau de la molécule d'ADN.

Réguler le niveau d'expression d'un gène *via* sa séquence codante se résume principalement à modifier son nombre de copies (Da Silva & Srikrishnan, 2012; Kim et al., 2012; Siddiqui et al., 2012). Les différentes stratégies d'expression (plasmides, YACs ou intégration chromosomique) décrites auparavant constituent des outils de modulation du nombre copies. Les plasmides multi-copies sont généralement utilisés dans un but initial de garantir un niveau d'expression élevé des gènes qui s'y trouvent. Cependant, une expression constitutive et non régulée peut induire une toxicité (synthèse d'une protéine ou d'un intermédiaire toxique). Dans ce cas, il est possible de constater une diminution du nombre de copies et une instabilité du plasmide en réaction de la souche à cette toxicité (Fang et al., 2011). Cela mène au constat qu'adopter une stratégie d'expression sur un vecteur CEN/ARS ou par intégration chromosomique peut parfois donner de meilleurs niveaux d'expression, et donc de meilleurs taux de production de la molécule ciblée, que sur un vecteur multi-copies (Wittrup et al., 1994). Des stratégies d'augmentation du nombre de copies existent (Chen et al., 2012), en modifiant par exemple les promoteurs placés devant les marqueurs d'auxotrophies pour les rendre plus faibles (Erhart & Hollenberg, 1983; Faulkner et al., 1994; Loison et al., 1989). La souche va ainsi compenser la trop faible quantité de marqueur par une augmentation du nombre de copies du plasmide de façon à compléter efficacement son auxotrophie. Les effets du choix d'un plasmide vont également beaucoup dépendre du contexte génétique de la souche ou encore de son stade de croissance. Pour conclure ce paragraphe, nous pouvons dire que les stratégies de régulation du niveau d'expression *via* le nombre de copies est peu fiable et possède un caractère multi-facteurs dépendant, induisant la nécessité de réaliser une optimisation au cas par cas.

3.2 Points de contrôle au niveau transcriptionnel.

Les stratégies de modification de l'activité transcriptionnelle sont, d'après la littérature, celles qui ont été les plus utilisées en vue de contrôler le niveau d'expression de gènes d'intérêt. Elles peuvent être décomposées en deux groupes : le choix des promoteurs sous le contrôle desquels les gènes sont placés, et le contrôle de la stabilité des ARN messagers après leur synthèse.

Approximativement 6000 régions promotrices ont été recensées chez *S. cerevisiae*, et sont répertoriées dans la *Saccharomyces cerevisiae* Promoter Database (SCPD ; <http://rulai.cshl.edu/SCPD>). Pour exprimer un gène chez la levure, deux types de promoteurs existent : les promoteurs constitutifs et les promoteurs inductibles. L'utilisation des promoteurs constitutifs ne requière aucune intervention du manipulateur pendant la culture, mais elle n'est souhaitable que si le produit généré n'est pas toxique, et si la production est voulue à tous moments de la croissance. Dans le cas contraire, les promoteurs inductibles s'avèrent utiles. En répondant à la présence d'une molécule activatrice ou répressive, ils permettent un contrôle du moment de l'expression. Les limites de ce système sont multiples (Da Silva & Srikrishnan, 2012):

- Une réponse des promoteurs qui n'est pas toujours linéaire en fonction de la concentration de l'inducteur / du répresseur ;
- Une expression « en bruit de fond », ou « fuite d'expression » souvent observée car un verrouillage complet du promoteur en conditions de répression est difficile à obtenir ;
- Le coût des métabolites nécessaires pour effectuer l'induction / la répression qui peut être incompatible avec des contraintes industrielles.
- La possibilité d'une réponse physiologique non désirée de la souche à l'inducteur / au répresseur, qui augmente la complexité du système.

Des systèmes basés sur des promoteurs inductibles ont cependant été utilisés en recherche chez la levure. Les systèmes *GAL1*, *GAL7*, *GAL10*, *CUP1*, *ADH2* ou *MET25* (Da Silva & Srikrishnan, 2012; Siddiqui et al., 2012) sont des promoteurs inductibles naturellement existants chez *S. cerevisiae*. Il existe aussi des composants synthétiques, comme le système *Tet*, répressible par la tétracycline, adapté de la bactérie (Gari et al., 1997). Ce système modulaire présente peu de pléiotropies. Les stratégies reposant sur l'utilisation de promoteurs inductibles sont globalement peu utilisées en ingénierie métabolique, mais elles peuvent s'avérer utiles pour optimiser le niveau

d'expression de certains gènes, avant d'y adapter un promoteur constitutifs conférant aux gènes les niveaux d'expression requis (Siddiqui et al., 2012).

Un éventail de promoteurs constitutifs associés à des niveaux de transcription variés (ou force de promoteurs) est communément utilisé chez la levure. Ils sont souvent issus de la voie de la glycolyse de *S. cerevisiae* : on y trouve les promoteurs des gènes de la phosphoglycérate kinase (*PGK1*), de la pyruvate decarboxylase (*PDC1*), de la triosephosphate isomérase (*TP11*), de l'alcool déshydrogénase I (*ADH1*), ou encore de la glycéraldéhyde-3-phosphate déshydrogénase (*TDH3* ou *GPD*). Le promoteur du facteur d'élongation *TEF1* est également largement utilisé et connu pour conférer un niveau de transcription élevé au gène qui se trouve sous son contrôle. Des projets d'ingénierie métabolique chez la levure ont été couronnés de succès en utilisant des promoteurs constitutifs. Sans que cette liste soit exhaustive, nous pouvons citer les voies métaboliques de consommation du xylose (Shao et al., 2009) ou de l'arabinose (Wisselink et al., 2007), de production d'acide malique (Zelle et al., 2008) ou encore de zéaxanthine (Shao et al., 2009; Sun et al., 2012). De façon à affiner la modulation du niveau d'expression, et ainsi réguler le flux métabolique *via* les promoteurs constitutifs, des travaux ont été effectués pour caractériser les forces respectives de chacun de ces promoteurs, en les associant à des gènes rapporteurs comme des protéines fluorescentes (Fang et al., 2011; Lee et al., 2013a; Sun et al., 2012), le gène LacZ (Mumberg et al., 1995; Partow et al., 2010), ou celui codant pour la β -lactamase (Cartwright et al., 1994). Ces études ont permis une classification des promoteurs constitutifs les plus utilisés en ingénierie métabolique de la levure, dans des conditions modèles. Ces données peuvent maintenant guider les auteurs d'études d'ingénierie métabolique dans le choix de leurs promoteurs (Latimer et al., 2014; Sun et al., 2012). L'équipe de Zhao en 2012 (Sun et al., 2012) a tenu compte du fait que les organismes transcrivent leurs gènes différemment selon les fluctuations des conditions environnementales, et a observé les niveaux de transcription de 14 promoteurs constitutifs de *S. cerevisiae* dans différentes conditions de culture (apport de glucose limité ou non, en erlenmeyer ou en fermenteur). Ils observent toujours la même tendance entre les différents promoteurs, avec un groupe offrant un niveau de transcription plus élevé (promoteur *TEF1p* particulièrement).

Il est donc possible d'imposer une régulation sur une voie métabolique par le choix des promoteurs. Cependant, les promoteurs constitutifs naturels de levure offrant des niveaux d'expression parfois insuffisamment diversifiés selon les besoins des études, certains auteurs ont

choisi de développer des banques de promoteurs synthétiques. L'équipe de G. Stephanopoulos a construit une banque de promoteurs en faisant de la mutagenèse aléatoire sur *TEF1p* (promoteur du gène *TEF1* de *S. cerevisiae*) par PCR, pour obtenir une sélection de 11 mutants dont les forces varient de 8% à 250% du promoteur natif (Alper et al., 2006; Nevoigt et al., 2006). La même technique a ensuite été utilisée pour modifier la force du promoteur *DAN1* chez *S. cerevisiae* (Nevoigt et al., 2007). En créant une banque de promoteurs synthétiques, Jeppsson et al. (2003) sont parvenus à contrôler le niveau d'expression de la glucose-6-phosphate déshydrogénase de 0% à 179% du niveau sauvage. Un jeu combinatoire de sept promoteurs synthétiques a été construit par Murphy et al. (2007), à partir du promoteur *GAL1* de *S. cerevisiae* auquel ont été ajoutés un, deux ou trois sites opérateurs *tetO₂* (site de répression par la tétracycline) dans diverses configurations. Leurs résultats montrent qu'il est possible d'utiliser ces composants synthétiques comme éléments de régulation afin de construire un système d'expression modulable (Murphy et al., 2007). Une approche alternative consiste à utiliser des séquences promotrices d'espèces proches de *S. cerevisiae*. Quatre promoteurs homologues de *MET25p* ou *MET2p* provenant d'une autre espèce de *Saccharomyces* ont été utilisés pour réguler les niveaux d'expression d'enzymes de la voie des flavonoïdes (Naesby et al., 2009). Un état de l'art complet à propos de l'ingénierie des promoteurs est détaillé dans la revue de (Blazeck & Alper, 2013).

Des outils de régulation post-transcriptionnelle basée sur une ingénierie des ARNs existent également. Ils consistent principalement à maîtriser la stabilité des ARN messagers cibles une fois qu'ils sont synthétisés, ou à contrôler l'initiation de la traduction (Siddiqui et al., 2012). Une équipe a par exemple récemment démontré que la demi-vie des ARN messagers peut être modulée par le choix du terminateur placé après la séquence codante (Curran et al., 2013). Ces systèmes sont indépendants du type de promoteur utilisé, et peuvent donc être associés à n'importe lequel d'entre eux. Cela offre donc une régulation à deux niveaux, potentiellement plus précise. La régulation par les 5'-UTRs ou les 3'-UTRs (UnTranslated Regions, ou régions non traduites) est basée sur des protéines qui se lient à l'ARN au niveau de motifs spécifiques. Un système basé sur le fonctionnement de la RNase III de *S. cerevisiae*, *Rnt1p*, a été mis au point (Babiskin & Smolke, 2011b). Cette dernière reconnaît les ARNs avec une structure en épingle à cheveux contenant un consensus AGNN dans une région non traduite (UTRs), et les clive pour une dégradation rapide par une exoribonucléase. Ce type de structure a été placé au niveau du 3' UTR de certains transcrits,

démontrant un système de régulation post-transcriptionnel efficace du niveau d'expression des gènes. Une banque de substrats synthétiques de Rnt1p a été développée, proposant une large gamme de niveaux d'activité de l'enzyme. Cette banque permet de moduler finement les niveaux de protéines *via* une variation prévisible de l'activité de Rnt1p (Babiskin & Smolke, 2011a; Babiskin & Smolke, 2011b). La méthode a été appliquée pour diminuer l'expression de *ERG9*, gène codant pour la squalène synthase chez *S. cerevisiae*, résultant en une régulation à la baisse du flux s'engageant dans la voie des stérols (Babiskin & Smolke, 2011a). D'autres approches de régulation de l'expression par la stabilité des ARNs se basent sur l'utilisation d'ARNs anti-sens, de microARNs (miRNAs) ou d'ARNs interférents (Kim et al., 2012). Les différentes techniques de contrôle de l'expression de gènes par les ARNs sont détaillées de manière plus exhaustive dans la revue (Chang et al., 2012). Très récemment, une étude a utilisé la capacité d'épissage alternatif des eucaryotes pour moduler le niveau d'expression d'une protéine fluorescente rapporteur chez *S. cerevisiae* (Yofe et al., 2014). Différents introns naturels de la levure ont été placés à l'intérieur de la phase codante de la *yfp* (Yellow Fluorescent Protein), pour ensuite observer l'intensité de fluorescence et la quantité d'ARNm dans différentes conditions environnementales. Les auteurs affirment qu'il s'agit d'un système de modulation du niveau d'expression prédictible et peu sensible aux variations environnementales.

3.3 Les points de contrôle au niveau traductionnel.

Le code génétique est tel que plusieurs codons différents codent pour le même acide aminé. Chaque organisme utilise certains codons plus que d'autres, et une protéine identique ne sera souvent pas codée par la même séquence d'ADN dans des organismes différents. Les informations sur les codons les plus fréquemment utilisés par les organismes sont rassemblées dans ce qu'on appelle un biais de codons. Il a été démontré qu'un recodage d'une séquence d'ADN selon ce biais peut mener à une meilleure traductibilité de la protéine, et ainsi à une meilleure activité (Gustafsson et al., 2004). Il s'agit là du premier niveau de contrôle applicable en ingénierie métabolique au niveau traductionnel.

Le contrôle du niveau de protéine hétérologue présent dans la cellule peut aussi être fait *via* le contrôle de la dégradation de cette protéine. Les méthodes utilisées pour y parvenir se servent

d'une machinerie de dégradation préexistante dans la cellule. Cela peut être accompli par l'ajout d'une étiquette de dégradation à une extrémité de la protéine, comme cela a été fait pour diminuer le temps de demi-vie de la GFP chez *S. cerevisiae* (Mateus & Avery, 2000). Chez *S. cerevisiae*, une voie de dégradation programmée des protéines s'appelle la règle du « N-end ». Une séquence en acides aminés spécifique est présente en N-terminal des protéines (N-degron). Il s'agit d'un signal d'adressage au protéasome (Tasaki et al., 2012). Une équipe a choisi d'exploiter cette machinerie de dégradation pour construire une famille de rapporteurs CFP (Cyan Fluorescent Protein) présentant une gamme de demi-vies allant de 5 à 75 minutes (Hackett et al., 2006). Un accroissement de la demi-vie d'une protéine est aussi possible par résistance à l'ubiquitination, signal de dégradation des protéines chez *S. cerevisiae* (Omura et al., 2005). Il existe aussi une ingénierie de la machinerie de dégradation : celle de *E. coli* a été importée chez la levure, donnant naissance à une gamme de répression de la GFP, avec des demi-vies allant de 91 à 22 minutes (Grilly et al., 2007), évitant les effets pléiotropes dans la levure.

Dans la nature, il arrive que les enzymes participant à certaines voies métaboliques s'organisent sous forme de métabolons. Il s'agit de structures multi-protéiques, où les enzymes responsables des étapes successives de transformation du substrat sont étroitement liées de façon à éviter la diffusion des intermédiaires de synthèse dans le milieu intracellulaire. C'est en s'inspirant de ses structures que sont nées les stratégies de construction de scaffolds et de fusions de protéines (Zhang, 2011). Dueber et al. (2009) sont parvenus à construire des scaffolds synthétiques permettant de moduler la stœchiométrie de trois enzymes et de réguler ainsi les flux métaboliques entrant dans la voie hétérologue. Très récemment, la voie métabolique du 2,3-butanediol a été reconstituée chez *S. cerevisiae* sous la forme de scaffolds basés sur l'interaction cohésine/dockérine (Kim & Hahn, 2014).

3.4 L'interaction entre biologie synthétique et systémique conduit à la perspective d'une régulation dynamique.

L'équipe de G. Stephanopoulos fait remarquer dans une revue de 2012 que les différentes méthodes d'ingénierie métabolique sont restées depuis leurs débuts une suite d'élégantes démonstrations, sans jamais devenir une pratique systématique, avec ses outils standardisés (Yadav

et al., 2012). Cela est en partie attribué au fait que beaucoup des outils utilisés pour la manipulation du métabolisme d'une cellule hôte ne sont pas universellement applicables. En effet, ils sont souvent spécifiques d'une voie métabolique ou d'un produit. Ce manque de transférabilité des outils d'ingénierie d'une cible métabolique à l'autre peut lui-même être attribué à une compréhension incomplète des mécanismes de régulation propres de la cellule hôte. Ces mécanismes constituent le sujet d'étude d'un autre champ disciplinaire : la biologie des systèmes, ou biologie systémique. Il s'agit d'une science interdisciplinaire qui étudie les interactions complexes à l'origine du comportement global d'une cellule ou d'un organisme. Grâce à une association d'outils de génomique, de transcriptomique, de protéomique et de métabolomique, c'est l'ensemble d'un système biologique qui est étudié. Cela révèle un profil d'ARNs, de protéines et de métabolites, ainsi que leur dynamique d'interaction au sein de la cellule, qui mettent en lumière des réseaux de régulation complexes (Liu et al., 2013a). Dans la nature en effet, il existe peu de systèmes de régulation statiques, mais les systèmes biologiques utilisent typiquement des réseaux de régulation dynamiques pour contrôler leurs flux métaboliques en réponse aux changements environnementaux (Xu et al., 2013; Zhang & Keasling, 2011). Par exemple, un intermédiaire métabolique peut agir comme une molécule signal pour induire ou réprimer l'expression de l'enzyme responsable de sa synthèse ou de sa fixation. Cette stratégie naturelle permet de moduler les flux métabolique de manière dynamique de façon à ce que les ressources cellulaires soient utilisées plus efficacement en fonction de leur disponibilité et des besoins de la cellule. Dans un but d'optimisation des flux métaboliques pour la production d'un composé d'intérêt, tout en évitant le stress que cela peut induire sur la cellule, l'imitation de cette stratégie de « bio-senseur » apparaît comme une solution essentielle. Par exemple, cette stratégie pourrait être appliquée dans le cas de l'accumulation d'un intermédiaire toxique. Lorsque cet intermédiaire serait détecté par le bio-senseur adapté, une boucle de répression de la voie métabolique responsable de sa synthèse, et une boucle d'induction de la voie qui le consomme seraient déclenchées, évitant ainsi son accumulation (Xu et al., 2013). Un tel système requière un composant sensitif de l'état métabolique de la cellule ou de la présence de métabolites toxiques, une méthode de modulation des flux, et une étroite connexion entre ces deux éléments (Holtz & Keasling, 2010). Des preuves de concept ont déjà été réalisées chez *Escherichia coli*. L'équipe de J.D. Keasling a construit en 2012 une souche de *E. coli* pourvue d'un système de contrôle du flux dans la voie des acides gras grâce à la protéine FadR. Il s'agit d'un facteur de transcription qui se

lie à l'ADN au niveau de séquences spécifiques pour induire la transcription de certains gènes de la voie des acides gras. Grâce au design de promoteurs synthétiques et à une stratégie utilisant de la Red Fluorescent Protein (*rfp*) en tant que rapporteur, l'équipe est parvenue à démontrer la possibilité de réguler une voie métabolique de synthèse par un système sensible de la quantité d'intermédiaire produit (Zhang et al., 2012a). La même équipe a ensuite identifié un certain nombre de promoteurs capables de répondre à des intermédiaires toxiques. Il a grâce à eux été possible de réguler la production d'amorphadiène (également dérivé de la voie des acides gras) chez *E. coli* (Dahl et al., 2013). Ces stratégies sont en plein développement, et elles donnent la direction que prend l'ingénierie métabolique chez les microorganismes aujourd'hui. Elles représentent un défi à relever chez la levure. La construction de circuits synthétiques, permettant un contrôle tenant compte de l'état global de la cellule fait partie des avancées majeures attendues dans le domaine de l'ingénierie métabolique. Ils devront être basés par exemple sur des facteurs de transcription sensibles multiples donnant naissance à de véritables plateformes de production biologiques performantes et robustes (Woolston et al., 2013).

CHAPITRE 2

Les caroténoïdes et leur production par ingénierie métabolique chez *Saccharomyces cerevisiae*.

I. DEFINITION ET PROPRIETES D'INTERET DES CAROTENOÏDES.

Les caroténoïdes forment un groupe de plus de 700 pigments liposolubles qui confèrent aux fleurs, fruits et certains microorganismes une couleur allant du jaune au rouge foncé (Obón & Rivera, 2006). Ils font partie de la grande famille des terpénoïdes, également appelés isoprénoïdes. La présence de doubles liaisons conjuguées est responsable de leur capacité d'absorption dans le visible entre 400 nm et 500 nm, avec plusieurs pics d'absorption caractéristiques de chaque molécule (Figure 12) qui sont largement utilisés pour les identifier. Ils sont synthétisés chez les plantes et certains microorganismes (bactéries, microalgues et champignons) où ils participent au processus photosynthétique, à la protection contre l'oxydation et à la régulation de la rigidité des membranes (Walter & Strack, 2011). Les animaux doivent trouver un apport en caroténoïdes *via* leur régime alimentaire à partir des différentes sources précédemment citées. Cet apport est indispensable : par exemple, environ 50 caroténoïdes différents sont considérés comme de la provitamine A car ils sont les précurseurs du rétinol et de l'acide rétinoïque (Okada et al., 2008) impliqués dans la croissance des os et la synthèse de pigments de l'œil.

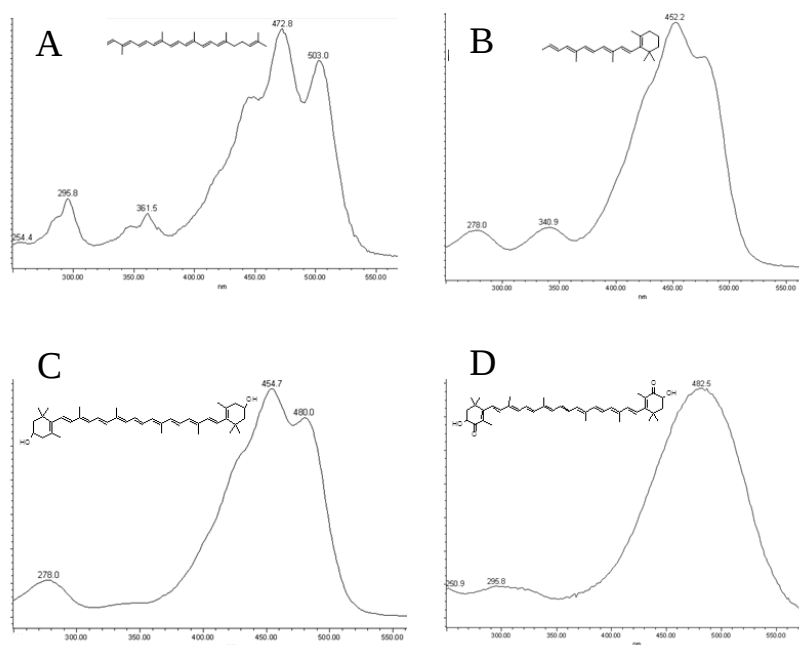


Figure 12: Spectres d'absorption de quatre caroténoïdes entre 250 et 550 nm. (A) Lycopène. (B) β-carotène. (C) Zéaxanthine. (D) Astaxanthine.

Les caroténoïdes forment une classe de composés organiques résultant de la condensation d'unités à cinq carbones (C_5), l'isopentényl-pyrophosphate (IPP). La plupart des caroténoïdes sont des molécules à 40 atomes de carbone (C_{40}), produits de la condensation de 8 molécules d'IPP. Le premier caroténoïde, non coloré, est le phytoène. A partir de ce squelette, des réactions d'hydrogénation, de déshydrogénation, de cyclisation et/ou d'oxydation conduisent à l'immense diversité de structures que l'on rencontre dans la nature. Il existe dans des cas plus rares des caroténoïdes en C_{30} , comme le diaponeurosporène de *Staphylococcus aureus*, ou d'autres en C_{50} comme la flavuxanthine ou la décaprénoxanthine de *Corynebacterium glutamicum* (Lee & Schmidt-Dannert, 2002). Les caroténoïdes se décomposent en deux classes principales : les carotènes (α -carotène, β -carotène, lycopène), exclusivement composés de carbone et d'hydrogène, et les xanthophylles (lutéine, zéaxanthine, astaxanthine) qui contiennent des atomes d'oxygène et portent des fonctions *hydroxy*, *époxy*, *aldéhyde*, *carboxy* ou *méthoxy* (Jaswir et al., 2011; Olson & Krinsky, 1995). La diversité des caroténoïdes et de leurs voies de biosynthèse a fait l'objet d'une revue très complète de (Walter & Strack, 2011). Des exemples et une représentation du système de numérotation des atomes de carbone sont donnés en Figure 13.

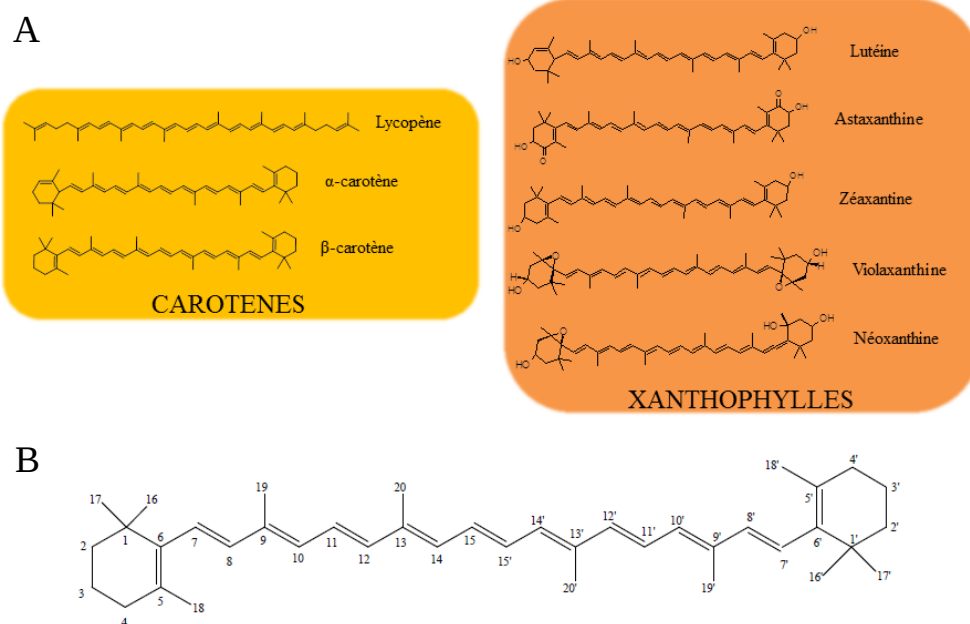


Figure 13: (A) Exemples de molécules entrant dans la catégorie des carotènes ou des xanthophylles. (B) Système de numérotation des atomes de carbone sur le modèle du β -carotène.

Les caroténoïdes possèdent des propriétés colorantes mais de nombreuses études leur ont aussi attribué des propriétés bioactives: anti-cancéreux, antioxydants, implication dans la protection de l'œil ou de la peau, ainsi que dans la lutte contre l'obésité (Jaswir et al., 2011; Stahl & Sies, 2005; Teodoro et al., 2012). Ces propriétés naturelles ont conduit les caroténoïdes à être utilisés dans diverses industries. L'industrie agro-alimentaire en particulier les utilise beaucoup en remplacement de colorants synthétiques, qui sont potentiellement associés à des effets néfastes sur la santé (Lee & Schmidt-Dannert, 2002; Mattea et al., 2009). Au-delà de leur utilisation comme colorants dans la nourriture humaine comme animale, leur propriétés bioactives en font aussi des alicaments, c'est-à-dire des aliments ayant des effets supposés sur la santé. Les industries pharmaceutique et cosmétique exploitent grandement les potentialités des caroténoïdes (Jaswir et al., 2011), le caractère naturel de ces molécules apportant une plus-value aux produits sur le marché.

II. LE CLIVAGE DES CAROTENOÏDES DONNE NAISSANCE A DES AROMES ET DES PARFUMS.

Les caroténoïdes peuvent faire l'objet, aussi bien dans les plantes que dans les microorganismes ou chez les mammifères, d'un clivage oxydatif (impliquant l'oxygène moléculaire O_2) au niveau d'une double liaison C-C par des Carotenoid Cleavage Oxygenases (CCOs) (Auldridge et al., 2006). Les produits qui en résultent, appelés apocaroténoïdes, sont dotés au niveau du site de clivage de groupements aldéhydes ou cétones. Des aldéhyde-oxydases oxydes ensuite souvent des fonctions aldéhydes et cétones en acides, puis des glycosyltransférases y greffent des sucres. Cela leur confère un caractère hydrosoluble qui les différencie de leurs précurseurs caroténoïdes, et qui leur permet d'être probablement stockés dans les vacuoles chez les plantes (Camara & Bouvier, 2004). Chez les animaux, l'apocaroténoïde le plus connu est le rétinol, aussi appelé vitamine A. Cette vitamine fortement impliquée dans le processus de vision est dérivée du β -carotène, clivé par une CCO (Figure 14).

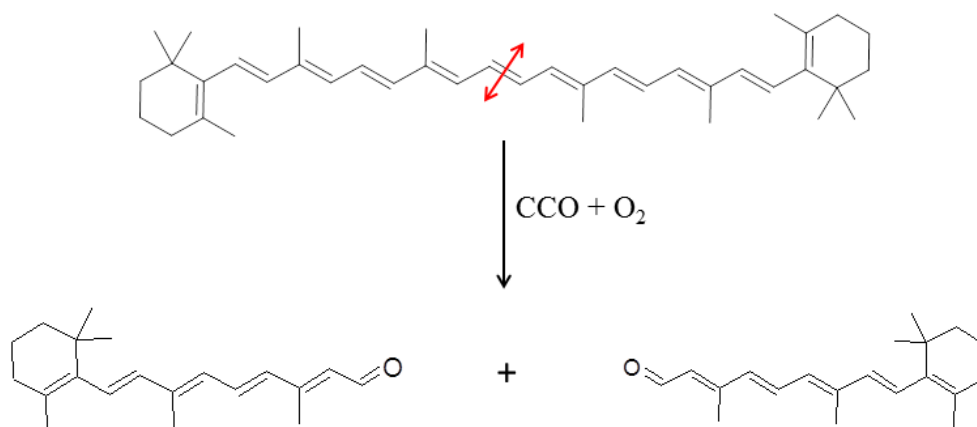


Figure 14: Biosynthèse de deux molécules de rétinol à partir d'une molécule de β -carotène.

Chez les plantes, les apocaroténoïdes peuvent avoir un pouvoir hormonal comme l'acide abscissique (ABA, C_{15}) qui est impliqué dans la tolérance au stress hydrique, le développement de la graine ou encore la sensibilité au sucre (Schwartz et al., 2003). Certains apocaroténoïdes sont des pigments utilisés comme colorants dans les industries agro-alimentaire et cosmétique : la bixine, naturellement produite chez le rocou (*Bixa orellana*) est l'additif E160b, et la crocine, qui

confère aux pistils du crocus safran (*Crocus sativus*) sa couleur rouge est l'additif E164 (Figure 15).

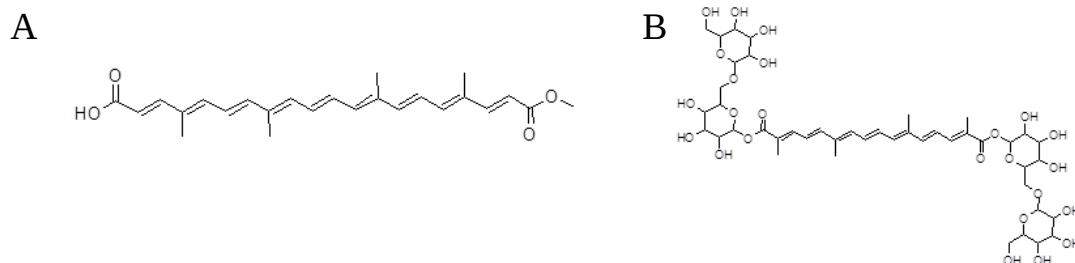


Figure 15: Exemples d'apocaroténoïdes utilisées comme additifs alimentaires ou cosmétiques. (A) Bixine. (B) Crocine.

Le caractère volatile des apocaroténoïdes en C₉ à C₁₃ en fait des parfums et des arômes, qui sont souvent associés à des seuils de détection très bas chez l'homme. Leur distribution dans les feuilles, les huiles essentielles, les fruits, les légumes ou les épices est très étendue. On les retrouve dans des produits transformés tels que le vin, le café ou encore le miel (Winterhalter & Rouseff, 2001). Nous pouvons donner les exemples très connus de composés aromatiques dérivés de caroténoïdes de la β -ionone (odeur de rose) ou de l' α -ionone (odeur de violette), mais leur diversité est très bien illustrées dans des revues comme Winterhalter & Rouseff, 2001, ou Rodriguez-Bustamante & Sanchez, 2007.

L'obtention de ces produits naturels par extraction de plantes puis purification reste aujourd'hui une tâche très difficile. Les producteurs naturels peuvent être des plantes non cultivées, et quand elles le sont, cette méthode est très dépendante des rendements de l'agriculture liés au climat, aux situations politiques ou sociales. De plus, ces composés sont souvent produits en très faibles quantités et généralement à un stade bien particulier du développement de la plante (Bouvier et al., 2003b), ne permettant pas d'obtenir des rendements satisfaisants. Les biotechnologies constituent donc une alternative de production très intéressante pour des industries comme l'alimentation ou les cosmétiques.

III. LA PRODUCTION DE CAROTENOÏDES CHEZ *S. CEREVISIAE*.

1. Les précurseurs endogènes.

Les stérols font partie de la famille des lipides, et sont des composants essentiels des membranes (régulateurs de leur perméabilité et de leur fluidité). Ils sont indispensables à la survie de la cellule. Une revue très complète sur leur biosynthèse et leur fonctions est disponible : Daum et al., 1998. L'ergostérol, stérol spécifique des *Fungi*, se compare au cholestérol des animaux bien qu'il en diffère par la présence d'insaturations et d'un groupement méthyle en au niveau du carbone 24 (Figure 16).

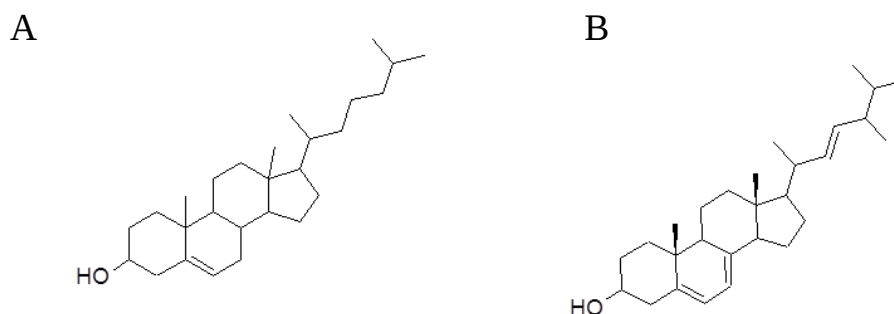


Figure 16: Comparaison des molécules de (A) cholestérol (animaux) et de (B) ergostérol (*fungi*).

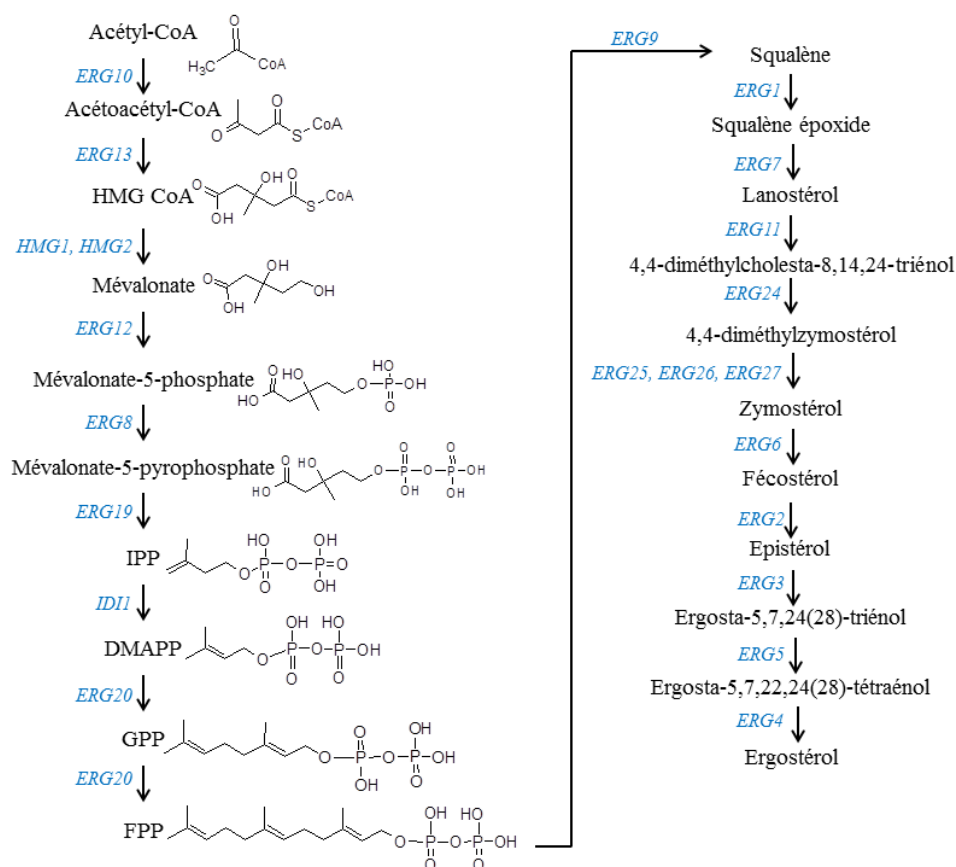


Figure 17: Voie de biosynthèse de l'ergostérol chez *S. cerevisiae*. Les gènes codant pour les enzymes responsables de chaque étape sont indiqués en bleu. HMG CoA, 3-hydroxy-3-méthylglutaryl-coenzyme A; IPP, isopentenyl pyrophosphate; DMAPP, Diméthylallyl-pyrophosphate ; GPP, geranyl pyrophosphate ; FPP, farnesyl pyrophosphate.

Sa voie de biosynthèse chez *S. cerevisiae* se décompose en deux parties : de l'acétyl-CoA (produit de la glycolyse) au farnesyl pyrophosphate (FPP), et du FPP à l'ergostérol (Figure 17). La première partie de la voie est appelée voie du mévalonate, ou voie des isoprénoïdes. Son précurseur est l'acétyl-CoA et neuf étapes enzymatiques sont nécessaires pour produire le FPP. La troisième étape de cette voie, catalysée par l'HMG-CoA réductase, est la plus étudiée de la voie. L'enzyme est codée par deux isoformes du gène : *HMG1* et *HMG2*. *HMG1* a été identifiée comme étant une étape limitante de la voie des stérols chez *S. cerevisiae* (Donald et al., 1997). Le FPP est le précurseur des stérols, mais aussi de la voie de biosynthèse de l'ubiquinone, dans laquelle il entre par une geranylgeranyl diphosphate synthase (BTS1). La réaction catalysée par BTS1 donne naissance au geranylgeranyl pyrophosphate (GGPP) (Figure 18). Une partie du pool de GGPP

synthétisé peut être dévié par ingénierie métabolique pour produire diverses molécules d'intérêt (Wriessnegger & Pichler, 2013) comme les terpènes, dont les caroténoïdes font partie.

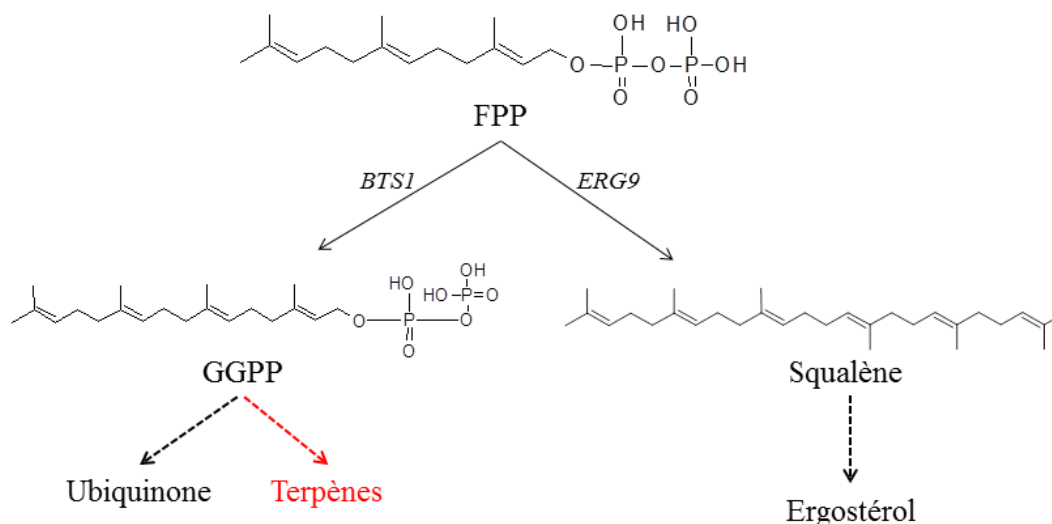


Figure 18: Voies de biosynthèse pouvant dériver du farnesyl pyrophosphate (FPP) chez *S. cerevisiae*. Les voies endogènes de *S. cerevisiae* sont représentées en noir, une voie hétérologue en rouge. GGPP, geranylgeranyl pyrophosphate.

2. L'ingénierie métabolique de *Saccharomyces cerevisiae* pour produire des caroténoïdes.

Un état de l'art de l'ingénierie métabolique chez *S. cerevisiae* pour produire des caroténoïdes (stratégies appliquées et quantités produites) est présenté dans le Tableau 3. Il est néanmoins utile de détailler ici les travaux qui ont pu être effectués sur ces voies métaboliques exprimées de manière hétérologue (*S. cerevisiae* est naturellement non productrice de caroténoïdes), l'étude qui va suivre s'inscrivant dans ce contexte.

L'équipe de Norihiko Misawa (Yamano et al., 1994), a construit en 1994 la première souche de *S. cerevisiae* productrice de caroténoïdes (lycopène et β -carotène) grâce à l'expression sur un plasmide multi-copies de gènes issus de l'organisme procaryote *Erwinia uredovora*. Ils ont détecté dans ces souches une diminution de la teneur en ergostérol de 7 à 10%. Ce résultat suggérait que

les nouvelles molécules synthétisées avaient pour précurseur un intermédiaire de la voie des stérols, redirigeant une partie du flux métabolique vers les caroténoïdes au lieu de l'ergostérol. En 1997, un brevet est déposé sur la production de zéaxanthine et de zéaxanthine glycosylée chez des hôtes bactériens et eucaryotes (dont *S. cerevisiae*) grâce à l'expression hétérologue de gènes de *Erwinia herbicola* (Ausich et al., 1997). La voie de biosynthèse de la zéaxanthine a également été exprimée par implémentation des gènes de *E. uredovora* (5 gènes) chez *S. cerevisiae*, par l'équipe de Huimin Zhao. Les travaux de cette équipe se basent souvent sur la reconstruction de la voie de la zéaxanthine comme preuve de concept de méthodes utiles à l'ingénierie métabolique. Ce fut le cas en 2009 pour démontrer l'efficacité de la technique de « DNA assembler » (Shao et al., 2009), puis en 2012 lors de la mise au point de la méthode du « COMPACTER » (Du et al., 2012), ou encore lorsqu'ils ont caractérisé une série de promoteurs constitutifs de levure pour moduler l'expression de gènes hétérologues (Sun et al., 2012). Un peu plus tard, cette équipe a également testé une méthode d'induction coordonnée des 5 gènes de la voie de la zéaxanthine, en les plaçant sous le contrôle de promoteurs inductibles par le galactose (promoteurs GAL) dans une souche modifiée pour les induire par l'estradiol (Liang et al., 2013).

Tableau 3: Analyse bibliographique de la production de caroténoïdes chez *Saccharomyces cerevisiae*.

Référence	Gènes exprimés	Stratégie d'expression	Produits finaux dosés	Quantités produites
Yamano et al. 1994	<i>CRTE</i> , <i>CRTB</i> et <i>CRTI</i> de <i>E. uredovora</i>	Plasmide 2µ; expression constitutive	Lycopène	113 µg/g[dw]
	<i>CRTE</i> , <i>CRTB</i> , <i>CTI</i> et <i>CRTY</i> de <i>E. uredovora</i>	Plasmide 2µ; expression constitutive	β-carotène	103 µg/g[dw]
Ausich et al. 1997	<i>CRTE</i> , <i>CRTB</i> , <i>CRTI</i> , <i>CRTY</i> et <i>CRTZ</i> de <i>E. herbicola</i>	Plasmide 2µ; expression inductible	zéaxanthine	0,01 % du poids sec
Verwaal et al. 2007	<i>CRTE</i> , <i>CRTI</i> ET <i>CRTYB</i> de <i>X. dendrorhous</i> + <i>tHMG1</i> de <i>S. cerevisiae</i> + autres copie de <i>CRTI</i> de <i>X. dendrorhous</i>	Intégration au génome; expression constitutive	β-carotène	5,9 mg/g[dw]
Ukibe et al. 2009	<i>CRTI</i> et <i>CRTYB</i> de <i>X. dendrorhous</i> + <i>BTS1</i> de <i>S. cerevisiae</i>	Plasmides 2µ; expression constitutive	β-carotène	390 µg/g[dw]
	<i>CRTI</i> , <i>CRTYB</i> , <i>CRTS</i> et <i>CRTR</i> de <i>X. dendrorhous</i> + <i>BTS1</i> de <i>S. cerevisiae</i>	Plasmides 2µ; expression constitutive	Astaxanthine	3 µg/g[dw]
	<i>CRTI</i> et <i>CRTYB</i> de <i>X. dendrorhous</i> + <i>BTS1</i> de <i>S. cerevisiae</i> + <i>CRTW</i> (β-carotène ketolase) de <i>Paracoccus</i> sp. + <i>CRTZ</i> (β-carotène hydroxylase) de <i>Pantoea ananatis</i>	Plasmides 2µ; expression constitutive		29 µg/g[dw]

Référence	Gènes exprimés	Stratégie d'expression	Produits finaux dosés	Quantités produites
Shao et al. 2009	<i>CRTE, CRTB, CRTI, CRTC</i> et <i>CRTZ</i> de <i>E. uredovora</i>	Plasmide 2μ; expression constitutive	Zéaxanthine	Non communiqué
		Intégration au génome; expression constitutive		
Lange & Steinbuchel, 2011	<i>CRTE, CRTI</i> et <i>CRTC</i> de <i>X. dendrorhous</i> + <i>MVA</i> de <i>S. aureus</i>	Plasmide 2μ; expression constitutive	β-carotène	3,897 μg/g[dw]
Ozaydin et al., 2013	<i>CRTI</i> et <i>CRTC</i> de <i>X. dendrorhous</i> + <i>BTS1</i> de <i>S. cerevisiae</i>	Plasmide 2μ; expression constitutive	Caroténoïdes totaux	entre 2000 et 2500 μg/g[dw]
Sun et al. 2012	<i>CRTE, CRTB, CRTI, CRTC</i> et <i>CRTZ</i> de <i>E. uredovora</i>	Plasmide ARS/CEN; expression constitutive	Zéaxanthine	0,74±0,02 mg/L de culture, à partir de xylose
Liang et al. 2013	<i>CRTE, CRTB, CRTI, CRTC</i> et <i>CRTZ</i> de <i>E. uredovora</i>	Plasmide ARS/CEN; expression inductible	Zéaxanthine	Production 50 fois plus élevée que dans Shao et al. 2009
Li et al., 2013	<i>CRTI</i> et <i>CRTC</i> de <i>X. dendrorhous</i> recodés + <i>tHMG1</i> de <i>S. cerevisiae</i>	Intégration au génome des gènes de la voie de caroténoïdes; expression constitutive + <i>tHMG1</i> sur plasmide 2μ	β-carotène	310,8 μg/g[dw]
	<i>CRTI</i> et <i>CRTC</i> de <i>X. dendrorhous</i> recodés + <i>MVA</i> de <i>S. aureus</i>	Intégration au génome des gènes de la voie de caroténoïdes; expression constitutive + <i>MVA</i> sur plasmide 2μ		390 μg/g[dw]

Référence	Gènes exprimés	Stratégie d'expression	Produits finaux dosés	Quantités produites
Xie et al., 2014	<i>CRTE</i> , <i>CRTYB</i> (2 copies) et <i>CRTI</i> (2 copies) de <i>X. dendrorhous</i> + <i>tHMG1</i> de <i>S. cerevisiae</i>	Intégration au génome; expression inductible	Caroténoïdes totaux	1,947 mg/g[dw]
Beekwilder et al., 2014	<i>CRTE</i> , <i>CRTI</i> ET <i>CRTYB</i> de <i>X. dendrorhous</i>	Plasmide ARS/CEN; expression constitutive	β -carotène	0,35±0,01 mg/g[dw]
		Plasmide 2 μ ; expression constitutive		0,55±0,01 mg/g[dw]
Bahieldin et al., 2014	<i>CRTE</i> , <i>CRTB</i> ET <i>CRTI</i> de <i>E. uredovora</i> recodés	Plasmide 2 μ ; expression constitutive	Lycopène	3,3 mg/g[dw]
Reyes et al., 2014	<i>CRTYB</i> , <i>CRTI</i> , <i>CRTE</i> de <i>X. dendrorhous</i>	Intégration au génome; expression constitutive	β -carotène	12 mg/g[dw]

Parmi les organismes naturellement producteurs de caroténoïdes, celui dont les gènes ont été le plus utilisés en ingénierie métabolique est probablement la levure *Xanthophyllomyces dendrorhous* (aussi connue sous le nom de *Phaffia rhodozyma*), productrice d'astaxanthine. Andrewes et al. ont proposé en 1976 une voie métabolique chez cette levure ayant pour précurseur des caroténoïdes le GGPP, obtenu à partir du FPP par l'enzyme CrtE (équivalent de Bts1 chez *S. cerevisiae*). Depuis, une voie de biosynthèse complète de l'astaxanthine chez *X. dendrorhous* a été décrite par Niklitschek et al. (2008) et Alcaïno et al. (2008). Elle se compose d'une enzyme bifonctionnelle possédant les activités de phytoène synthase et de lycopène cyclase (CrtYB), une phytoène désaturase (CrtI) et une β -carotène hydroxylase (CrtS ; P450) fonctionnant en association avec la P450 réductase CrtR (Figure 19).

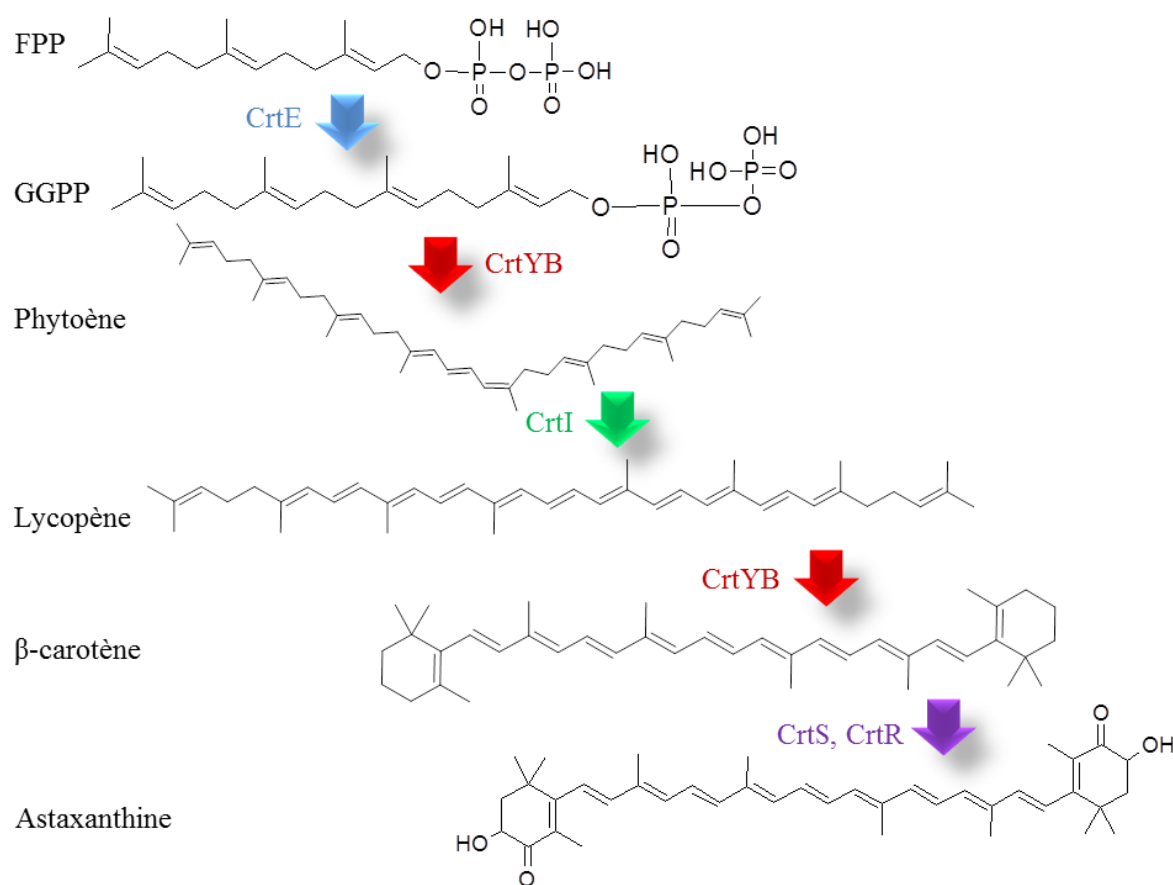


Figure 19: Voie de biosynthèse de l'astaxanthine chez *Xanthophyllomyces dendrorhous*. CrtE, GGPP synthase ; CrtYB, phytoène synthase et lycopène cyclase ; CrtI, phytoène désaturase ; CrtS, P450 (β -carotène hydroxylase) ; CrtR, P450 réductase.

Une partie de ces enzymes a été exprimée chez *S. cerevisiae* pour produire du β -carotène (Verwaal et al., 2007). Différentes combinaisons de gènes, exprimées grâce à différents outils (plasmides 2 μ , CEN/ARS, et intégration chromosomique) ont été testées. Le meilleur rendement a été observé en accompagnant l'expression des gènes de la voie de biosynthèse du β -carotène de *X. dendrorhous* par une surexpression du gène codant la sous-unité catalytique de l'HMG-CoA réductase (*tHMG1*) de *S. cerevisiae*, et d'une deuxième copie du gène de la phytoène désaturase (*CRTI*) de *X. dendrorhous*, le tout intégré au génome de *S. cerevisiae*. Un des plasmides multi-copies construit lors de ces travaux (YEplac195 YB/I/E) a été réutilisé par Lange & Steinbüchel en 2011, en ajoutant aux trois gènes hétérologues déjà présents celui qui code la mévalonate kinase de *Staphylococcus aureus* (*MVAK1*). MvaK1 est responsable chez *S. aureus* de l'étape de phosphorylation du mévalonate, catalysée par Erg12 chez *S. cerevisiae* (Figure 17). L'addition d'une autre copie de mévalonate kinase, sous une forme hétérologue (en supposant que cette forme ne ferait pas l'objet d'une régulation chez *S. cerevisiae*) leur a permis d'augmenter de manière spectaculaire le rendement en β -carotène par rapport à ce qui avait été obtenu par les auteurs de YEplac195 YB/I/E. Dans le même esprit, la sous-unité catalytique de l'HMG-CoA réductase de *S. aureus* (Mva) a été exprimée sur un plasmide multi-copies chez *S. cerevisiae* (Li et al., 2013). Les gènes *CRTI* et *CRTYB* de *X. dendrorhous* recodés selon le biais de codons de *S. cerevisiae* avaient préalablement été intégrés au génome de la souche. L'ajout de Mva de *S. aureus* s'est avéré plus efficace que l'ajout de *tHMG1* de *S. cerevisiae*. En 2009, une autre équipe a aussi exprimé les gènes de la voie de biosynthèse des caroténoïdes de *X. dendrorhous* sur un plasmide multi-copies chez *S. cerevisiae* en ajoutant CrtS, l'astaxanthine synthase (P450) qui catalyse la formation d'astaxanthine à partir du β -carotène. L'expression de CrtR, P450 réductase catalysant le transfert d'électron nécessaire à l'activité catalytique de CrtS, était nécessaire à la production d'astaxanthine (Ukibe et al., 2009). L'implication de certains caroténoïdes dans la tolérance au stress oxydatif chez *S. cerevisiae* a été démontrée dans la même étude, avec une sensibilité moindre au peroxyde d'hydrogène que les souches non productrices de caroténoïdes. Le pouvoir antioxydant des caroténoïdes a été exploité en ingénierie métabolique en faisant de l'évolution dirigée de souches de *S. cerevisiae* modifiées pour la production de caroténoïdes par mise en présence de peroxyde d'hydrogène pour en ressortir des souches plus productives (Reyes et al., 2014). La voie de biosynthèse du β -carotène de *X. dendrorhous* a également fait l'objet de travaux ciblant leur expression de façon contrôlable chez *S. cerevisiae* en utilisant le système GAL (Xie et al., 2014).

Les gènes de la voie de biosynthèse ont été intégrés de manière dispersée dans le génome, sous le contrôle des promoteurs *GAL1* et *GAL10*. L'induction des gènes a été contrôlée par des variations des concentrations en glucose et en glycérol du milieu de culture.

Dans les travaux de l'équipe de J.D. Keasling, les caroténoïdes, grâce à la possibilité de criblage visuel qu'ils offrent *via* l'apparition d'une coloration, ont été transformés en outils de criblage de mutants de *S. cerevisiae* (banque de délétions publiée par Winzeler et al., 1999) offrant une meilleure production de métabolites secondaires dérivés du FPP (Ozaydin et al., 2013). Les mutants ayant été identifiés comme étant de meilleurs producteurs de caroténoïdes ont ensuite été sélectionnés pour produire du bisabolène (biodiésel).

Une récente étude a montré que l'expression des gènes de biosynthèse du β -carotène *via* une construction polycistronique était possible chez *S. cerevisiae* (Beekwilder et al., 2014). Trois gènes de *X. dendrorhous* ont été implémentés sur une même construction plasmidique (centromérique), séparés par des séquences 2A (petits peptides dérivés de polyprotéines virales), permettant une transcription unique des trois gènes, suivie de clivages du transcrit. Les auteurs ont ensuite ajouté un gène de CCO de *Rubus idaeus* (framboise), et sont parvenus synthétiser la β -ionone chez la levure.

Présentation du projet de thèse

Notre étude s'inscrit à l'intérieur un projet d'ingénierie métabolique chez la levure *S. cerevisiae*, visant à terme la production de colorant, d'arôme et de parfum du safran (*Crocus sativus*), dérivés de la zéaxanthine. La voie métabolique hétérologue complète, venant se brancher au niveau du GGPP, est présentée dans sa globalité dans la Figure 20.

Le projet peut être scindé en deux parties, qui sont (Chapitre 1) l'implémentation de la voie métabolique des caroténoïdes menant à la synthèse de la zéaxanthine, et (Chapitre 2) une étude des enzymes susceptibles de catalyser les réactions qui mènent au colorant (la crocine), à la saveur (la picrocrocine) et au parfum (le safranal) du safran à partir de la zéaxanthine.

Dans le premier chapitre des travaux qui vont être exposés dans la suite de ce document, la voie métabolique de la zéaxanthine a été introduite chez *S. cerevisiae* par une approche combinatoire. La stratégie combinatoire a été explorée à deux niveaux. Tout d'abord, l'association de gènes d'origines différentes nous a menés à l'observation d'une production de métabolites. Le deuxième niveau concerne la régulation de cette voie métabolique au niveau transcriptionnel. Les niveaux d'expression des trois gènes ont été modulés un à un, puis des analyses des niveaux d'ARNm correspondants ont été effectuées afin de connaître l'effet réel d'une telle approche combinatoire.

La seconde partie de l'étude présente un certain nombre d'inconnues : seules les étapes catalysées par CsZCD (CCO ; (Bouvier et al., 2003b)) et UGTCS2 (glucosyltransférase ; (Moraga et al., 2004)) sont connues à ce jour (Figure 20). Cela constitue un troisième niveau de combinatoire. En effet, il est intéressant d'adopter une stratégie combinatoire pour la reconstruction de cette voie métabolique, impliquant des enzymes de natures et d'origines différentes, présentant des activités catalytiques proches de celles qui sont recherchées, afin de tester leur activité chez *S. cerevisiae*.

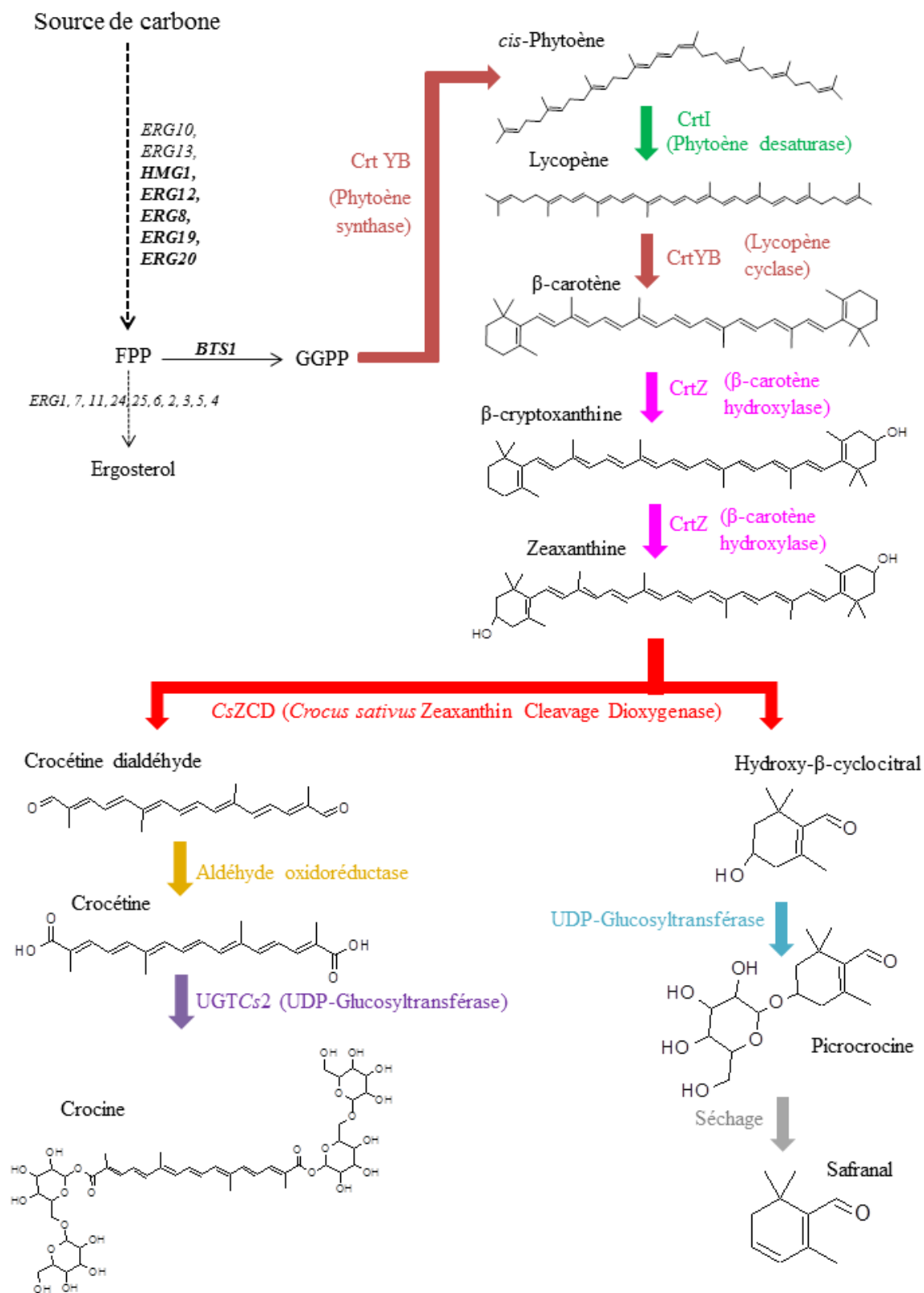


Figure 20: Présentation de la voie métabolique à implémenter chez la levure *S. cerevisiae*. Les étapes endogènes de *S. cerevisiae* sont indiquées en noir, les étapes hétérologues à implémenter sont colorées.

Résultats et Discussion

CHAPITRE 1

Etude des forces de promoteurs constitutifs de *S. cerevisiae*.

I. DONNEES INTRODUCTIVES.

Lee et al. (2013b) sont parvenus à reconstituer la voie métabolique de la violacéine sans connaître la voie naturelle. Cela signifie que l'association de gènes provenant de différentes origines peut être fonctionnelle. Nous avons choisi de construire la voie métabolique la plus courte possible entre le GGPP et la zéaxanthine. Nous avons exprimé chez *S. cerevisiae*, sur un plasmide multi-copies, trois gènes provenant de deux organismes producteurs de caroténoïdes : la levure *Xanthophyllomyces dendrorhous*, et la bactérie *Pantoea ananatis*. Nous avons donc travaillé sur une voie métabolique complètement artificielle, et nous avons observé une production de métabolites.

Comme nous l'avons introduit dans ce document, un défi majeur de l'ingénierie métabolique est de parvenir à une régulation fine du niveau d'expression de chaque gène d'intérêt. Il est en effet connu qu'une expression non régulée de gènes étrangers chez un hôte peut résulter en un fort déséquilibre métabolique, principalement lié à une surconsommation des ressources cellulaires, ou à une accumulation d'intermédiaires toxiques (Glick, 1995). Une optimisation de la répartition des flux métaboliques dans la cellule est donc nécessaire pour éviter des conséquences sur la physiologie cellulaire. Parmi les différentes stratégies de régulation utilisées dans la littérature, nous avons choisi de focaliser notre étude sur la régulation au niveau transcriptionnel, la plus fréquemment utilisée. Une régulation du niveau d'ARN messager issu de chaque gène d'intérêt a été dans certains cas une stratégie fiable de modulation étape-à-étape d'une voie métabolique hétérologue (Kim et al., 2012; Siddiqui et al., 2012). Dans le but d'affiner cette régulation, et de pouvoir l'appliquer de façon plus standardisée dans des projets d'ingénierie métabolique chez *S. cerevisiae*, des équipes ont caractérisé la force de certains promoteurs constitutifs de levure. Les auteurs ont pour cela utilisé des systèmes modèles tels que l'expression de gènes rapporteurs comme *LacZ* qui code pour la β -galactosidase de *E. coli* (Partow et al., 2010), ou codant pour des protéines fluorescentes (Lee et al., 2013b; Sun et al., 2012).

En dosant l'activité *LacZ* provenant du gène de *E. coli* intégré au génome de *S. cerevisiae*, Partow et al. (2010) ont testé les promoteurs correspondant aux gènes de levure *TEF1*, *ADH1*, *TPI1*, *HXT7*, *TDH3*, *PGK1* et *PYK1*. En condition de croissance en erlenmeyer et en consommation de glucose, l'ordre décroissant des forces de promoteurs qu'ils ont observé est :

$$TEF1p \sim PGK1p \sim TDH3p > TPI1p \sim PYK1p > ADH1p > HKT7p$$

En condition de consommation d'éthanol (après que le glucose ait été entièrement consommé), l'ordre était le suivant :

$$TEF1p \sim HXT7p > PGK1p > TPI1p \sim TDH3p > PYK p \sim ADH1p$$

D'après ces résultats, la phase de croissance, et par conséquent la source de carbone, impacte significativement la force des promoteurs considérés.

En 2012, Sun et al. ont caractérisé un jeu de 14 promoteurs constitutifs de *S. cerevisiae* en plaçant le gène de la GFP sous leur contrôle, dans plusieurs conditions de culture. Ils ont exprimé leurs constructions sur un plasmide multi-copies, puis ils ont mesuré l'intensité de fluorescence produite afin de conclure sur les forces respectives des promoteurs testés. En culture en erlenmeyer au bout de 12 heures, aussi bien en conditions aérobies qu'anaérobies, le promoteur le plus fort était *TEF1p*. A l'inverse, les plus faibles étaient *PGI1p* et *ADH1p*. *PGI1p* n'avait pas été testé par Partow et al. (2010), mais les conclusions des deux équipes sont similaires concernant *TEF1p* et *ADH1p*. Dans l'étude de Sun et al. (2012), une corrélation a été recherchée entre l'intensité de fluorescence observée et le niveau d'ARN messager de la GFP mesuré par RT-qPCR (Reverse Transcription quantitative PCR). L'intensité de fluorescence et le niveau d'ARNm étaient généralement corrélés, mais ce résultat n'était pas systématique pour chaque promoteur. Cela suggère que des paramètres indépendants de la nature du promoteur sont à prendre en compte dans le niveau d'expression observé d'un gène.

En 2013, Lee et al. (2013b) ont caractérisé 13 promoteurs constitutifs de *S. cerevisiae* par l'expression de trois protéines fluorescentes différentes, à chaque fois sur un plasmide centromérique. Les cellules étaient ici cultivées en plaques 96 puits, et les mesures de fluorescence ont été effectuées une fois les cultures arrivées à saturation. La force des promoteurs observée dans des conditions standards était conservée quelle que soit la séquence codante qu'ils contrôlaient. Dans ces conditions, *TDH3p* était le promoteur le plus fort, devant *TEF1p*. *PGK1p*, considéré comme de niveau moyen par Sun et al. (2012), conférait ici un niveau de fluorescence équivalant à *TEF1p*.

Ces trois études nous montrent que les forces respectives des promoteurs constitutifs de *S. cerevisiae* dépendent fortement de paramètres tels que les conditions de culture (source de carbone,

oxygénation) ou encore la stratégie d'expression de la séquence codante (intégrée au génome, sur un plasmide centromérique ou multi-copies). Ces études ayant été faites sur les systèmes d'expression et de culture différents, elles sont donc difficilement comparables entre elles. De plus, ces études ont toutes été menées sur des systèmes modèles dans lesquels un seul gène est exprimé, codant pour des protéines produites en grandes quantités chez la levure sans présenter de phénomène de toxicité particulier. Ces conditions sont rarement réunies lorsqu'une voie métabolique complexe est reconstituée. Même si certains auteurs des études présentées ci-dessus ont utilisé les données de leurs caractérisations de promoteurs pour une reconstruction de voie métabolique (zéaxanthine pour Sun et al. 2012, violacéine pour Lee et al. 2013b), les niveaux d'ARNm correspondant à chacun des gènes de la voie n'ont pas été vérifiés afin d'être corrélés aux résultats précédemment obtenus avec les gènes rapporteurs.

Dans notre étude, nous avons voulu savoir si un choix précis des promoteurs constitutifs qui contrôlent les gènes composant une voie métabolique multi-étapes suffisait à imposer une régulation fine de leurs niveaux d'expression respectifs. Sur la base de la caractérisation des forces de promoteurs faite par Sun et al. (2012), nous avons construit une combinatoire de niveaux d'expression des gènes de la voie de la zéaxanthine. Nous avons placé trois gènes sur un plasmide multi-copies, tour à tour sous le contrôle d'un promoteur faible, de niveau moyen, ou fort. Cette stratégie nous a permis de mettre en évidence des éléments critiques à prendre en compte pour réguler les niveaux d'expression des gènes dans ces conditions. Nous avons ensuite analysé la corrélation entre les niveaux d'ARN messagers observés et les quantités de métabolites produites.

Article scientifique soumis à *Frontiers in Bioengineering and Biotechnology*, specialty section *Synthetic Biology*.

Transcription Interference and ORF Nature Strongly Affect Promoter Strength in a Reconstituted Metabolic Pathway.

II. Transcription interference and ORF nature strongly affect promoter strength in a reconstituted metabolic pathway.

Marie Carquet^{1,2,3}, Denis Pompon^{1,2,3}, Gilles Truan^{1,2,3*}

¹Université de Toulouse; INSA, UPS, INP; LISBP, Toulouse, France

²INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, Toulouse, France

³CNRS, UMR5504, Toulouse, France

* **Correspondence:** Gilles Truan, ¹Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France

Gilles.truan@insa-toulouse.fr

Keywords: Metabolic engineering, *Saccharomyces cerevisiae*, Combinatorial metabolic libraries, Expression tuning, Plasmid-encoded metabolic pathway.

Abstract

Fine tuning of individual enzyme expression level is necessary to alleviate metabolic imbalances in synthetic heterologous pathways. A known approach consists of choosing a suitable combination of promoters, based on their characterized strengths in model conditions. We questioned whether each step of a multiple-gene synthetic pathway could be independently tunable at the transcription level. Three open reading frames, coding for enzymes involved in a synthetic pathway were combinatorially associated to different promoters on an episomal plasmid in *Saccharomyces cerevisiae*. We quantified the mRNA levels of the three genes in each strain of our generated combinatorial metabolic library. Our results evidenced that the ORF nature, position, and orientation induce strong discrepancies between the previously reported promoters' strengths and the observed ones. We conclude that, in the context of metabolic reconstruction, the strength of usual promoters can be dramatically affected by many factors. Among them transcriptional interference and ORF nature seem to be predominant.

1. Introduction

Many biotechnological processes involve microorganisms to produce molecules of interest in a sustainable way, sometimes competing efficiently with chemical synthesis (Kim et al., 2013). In this context, one of the typical goals of synthetic biology, and more particularly of metabolic engineering, is to genetically modify known and robust microorganisms to create artificial pathways compatible with industrial processes. However, unregulated expression of foreign enzymes used to rebuild synthetic pathways in heterologous hosts may interfere with endogenous metabolic fluxes, causing metabolic burden, overconsumption of resources or release of toxic intermediates (Glick, 1995; Neubauer et al., 2003). It is thus necessary to properly adjust the levels of all enzymes participating in the heterologous metabolic pathway to maximize metabolic fluxes toward the desired path while avoiding undesirable impacts on cell physiology and viability (Lu and Jeffries, 2007; Ajikumar et al., 2010; Ramon and Smith, 2011). Balancing natural versus synthetic fluxes can be achieved by individually tuning the expression level of each foreign enzyme. In the host *Saccharomyces cerevisiae*, regulated promoters with tunable strengths can be used (Da Silva and Srikrishnan, 2012), but this approach is often limited by the nonlinear responses to the inducer concentration or by the cost of these molecules, prohibitive in an industrial context (Siegele and Hu, 1997; Mnaimneh et al., 2004). Alternatively, collections of natural or engineered constitutive promoters of various strengths allow fine tuning of enzyme expression levels (Alper et al., 2005; Lu and Jeffries, 2007; Du et al., 2012).

The importance of promoter strength and regulation on net protein output has been extensively reviewed (Da Silva and Srikrishnan, 2012; Blazeck and Alper, 2013). Some studies characterized constitutive promoters by placing lacZ (Partow et al., 2010) or fluorescent protein encoding ORFs as reporters (Sun et al., 2012; Lee et al., 2013). Lee et al. characterized a set of *S. cerevisiae* constitutive promoters giving a robust control over gene expression. They used them to build a combinatorial library leading to violacein production without any knowledge of the optimal combination of individual enzymes levels (Lee et al., 2013). However, Karim et al. demonstrated that growth rates and plasmid copy number (parameters known to impact production efficiency) were influenced by additional factors, including the origin of replication and selection marker on episomal plasmids (Karim et al., 2013). One can question whether expression levels measured with

model systems consistently reflect the strengths of the considered promoters, regardless of their direct genetic environment.

While chromosome integration of genes is the preferred technique for final, industrial strains, first optimization steps are usually performed on plasmids (Ro et al., 2006; Steen et al., 2008; Ukibe et al., 2009; Kocharin et al., 2012). Our goal was to assess whether the assigned strengths of promoters, when described in model systems, can be transposable when multiple expression cassettes are inserted on a single episomal plasmid. We therefore used as model a small, heterologous pathway leading to the synthesis of zeaxanthin, and implemented it in *S. cerevisiae*. The heterologous production of carotenoid molecules has previously been described using *S. cerevisiae*, (Verwaal et al., 2007; Ukibe et al., 2009; Sun et al., 2012) demonstrating that rerouting the endogenous terpene pathway to these heterologous metabolites is feasible (Figure 1). Three heterologous ORFs from different carotenoid-producing organisms were used: *CRTI* (phytoene desaturase gene), *CRTYB* (bifunctional phytoene synthase and lycopene cyclase gene) both from *Xanthophyllomyces dendrorhous*, and *CRTZ* (β -carotene hydroxylase gene) from *Pantoea ananatis*. A small metabolic combinatorial library was built using six constitutive promoters that were reported to be strong, medium or weak using the fluorescent protein GFP as a reporter (Sun et al., 2012). mRNA levels of the three abovementioned genes were quantified in various culture conditions.

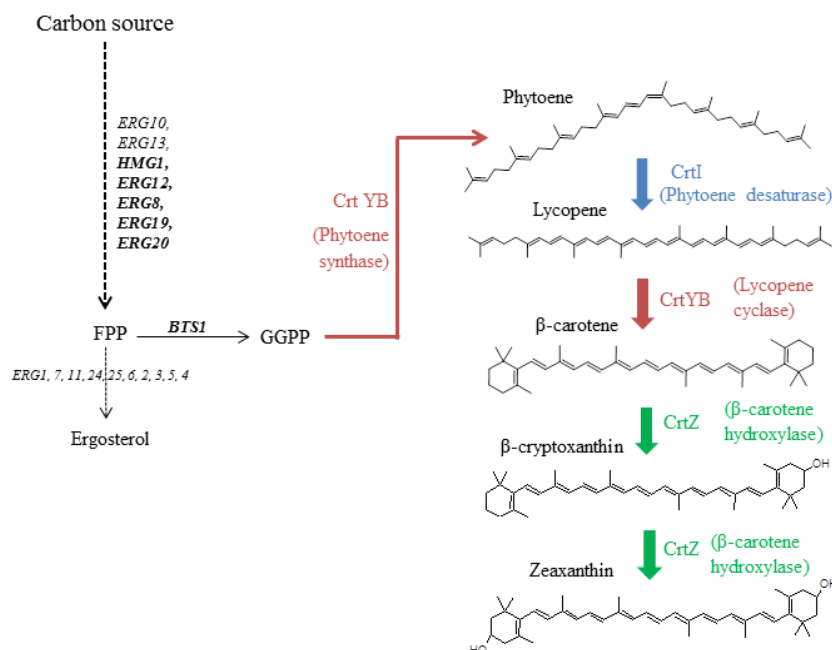


Figure 1: Zeaxanthin reconstruction pathway in the host *S. cerevisiae* from geranylgeranyl pyrophosphate (GGPP). Heterologous steps are colored.

2. Materials and methods

Plasmid constructions

Genes *CRTI* and *CRTYB* from *X. dendrorhous* were amplified from plasmid YEplac195 YB/I (kindly supplied by Prof. Gerhard Sandmann (Verwaal et al., 2007)). *CRTZ* from *P. Ananatis* (GenBank accession number D90087) was chemically synthesized (Genecust, Luxembourg) after *S. cerevisiae* codon bias optimization using Gene Designer 2.0 (See Supplementary data for optimized sequence). *TEF1*, *PDC1*, *PGI1*, *GPD*, *ENO2* and *TEF2* promoters, as well as *ADH1*, *TEF2* and *CYC1* terminators were amplified from genomic DNA of *S. cerevisiae*. Overlaps of 30 bp were designed between adjacent fragments and no spacer sequences were added between the different transcription units (Supplemental Table 1). PCR products were purified and further used for overlapping amplification to construct individual expression cassettes (promoter-ORF-terminator). A last PCR amplification was performed with primers RECpRS426_F and RECpRS426_R to extend (40 bp) the recombination segments between the synthesized fragments and pRS426 (Christianson et al., 1992). The library of promoter-ORF-terminator associations was

subsequently assembled by Gibson cloning method (Gibson et al., 2009) in a *SacI/XhoI* linearized pRS426 (*URA3* selection marker) (Gibson Assembly Cloning Kit, NEB, Ipswich, MA). The whole construction method is detailed in Supplemental Figure 1. The combinatorial library is composed of nine plasmids, named pRS426/A to pRS426/J (Table 1). Empty pRS426 is named H. To avoid recombination event, no repeated sequence was inserted into the plasmids. That explains the absence of reuse of the same promoter in each single construct, and the choice of three different terminators. The catalytic subunit of HMG-CoA reductase gene (*tHMG1*) was amplified from genomic DNA of *S. cerevisiae*, with primers *tHMG1_pCM185_F* and *tHMG1_pCM185_R* (Supplemental Table S1) and inserted in pCM185 (TRP1 selection marker) (Gari et al., 1997), resulting in pCM185/*tHMG1*. All constructions were verified by sequencing.

Table 1: Order of genes in the different pRS426 derivatives. The strong promoter is shown in pink, medium promoters are in orange, and the weak promoter is in green.

Constructs	Transcription unit 1			Transcription unit 2			Transcription unit 3		
A	TEF1p		ADH1t	PDC1p		TEF2t	PGI1p		CYC1t
B	TEF1p			PGI1p			PDC1p		
C	PDC1p			TEF1p			PGI1p		
D	PDC1p			PGI1p			TEF1p		
E	PGI1p			TEF1p			PDC1p		
F	PGI1p			PDC1p			TEF1p		
G	GPDp			ENO2p			TEF2p		
H	Empty pRS426								
I	PDC1p		CYC1t	PGI1p		ADH1t	TEF1p		TEF2t
J	PGI1p		ADH1t	TEF2t		TEF1p	PDC1p		CYC1t

Strains, culture and sampling conditions

The generated vectors were transformed in the *S. cerevisiae* CEN.PK 113-7D strain (*Mata his3—1 leu2-3,112 trp1-289 ura3-52 MAL2-8c*) by the lithium acetate method (Gietz et al., 1992), giving birth to strains A to J or A^{+H} to I^{+H}, according to the presence of empty pCM185 or pCM185/*tHMG1*, respectively. Recombinant strains were grown on CSM-URA-TRP plates (MP Biomedicals, Santa Ana, CA) from a glycerol stock at 30°C during approximately three days. A single colony was then grown in 50 mL of complete synthetic, selective medium (Supplemental Table S2) at 30°C and 200 rpm during 24 hours. 100 mL of the same medium was further inoculated at a final OD₆₀₀ of 0.5, in 500 mL shake flasks. 2 mL samples were drawn and immediately cooled to 4°C and centrifuged at 3,200 g (4°C). Supernatants were filtered over 0.2 µm and quenched at -80°C prior to their use for NMR quantification of extracellular metabolites. The cell-containing pellets were washed once with 1 mL of ice cold ultrapure water, rapidly centrifuged at 4°C, and quenched in liquid nitrogen before being stored at -80°C for subsequent nucleic acids extraction. At the end of the 72 hours of culture, cells were harvested by centrifugation (8,200 g, 2 minutes), washed once with deionized water, frozen at -80°C, and lyophilized with a CHRIST Alpha 2-4 LD plus freeze-drier. All culture experiments were performed in triplicates. For real independency of triplicates, shake-flasks cultures were performed neither simultaneously, nor from the same inoculum.

RT-qPCR experiments

Frozen cells were mechanically disrupted using a Tissue Lyser II (Qiagen, Venlo, Netherland) during 3 minutes at a frequency of 30 Hz. Total nucleic acids were extracted using the GeneJET RNA purification kit (Thermo Scientific, Waltham, MA). Genomic DNA contamination was eliminated by an additional DNase treatment performed with the RNase-free DNase I (Thermo Scientific, Waltham, MA). RNA quality was assayed on a Bioanalyzer 2100 with the RNA 6000 Nano LabChip kit (Agilent Technologies, Santa Clara, CA). One microgram of RNA was reverse-transcribed into cDNA in a 20 µL reaction using the Maxima first strand cDNA synthesis kit (Thermo Scientific, Waltham, MA). 4 µL of a 10 folds dilution of the previously synthesized cDNA were added to the qPCR Master Mix to perform cDNA relative quantification. Quantitative PCR (qPCR) reactions were performed in triplicates, in 20 µL final volume and in 96-well plates (Bio-

Rad, Hercules, CA) using the Maxima SYBR Green/Fluorescein qPCR Master Mix (2X) from Thermo Scientific (Waltham, MA), in the MyIQ real-time PCR system from Bio-Rad. A blank (No Template Control) was also incorporated in each assay. Primers were designed using Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>), and their specificity was verified using the PrimerBLAST tool from NCBI (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). qPCR primer sequences are given in Supplemental Table 3. The PCR efficiency of each primer pair (Eff) was evaluated by the dilution series method using a mix of sampled cDNAs as template. The absence of contaminant genomic DNA in RNA preparations was verified using RNA as a template in real-time PCR assays (minus RT control). The thermocycling program consisted of one hold at 95°C for 10 min, followed by 40 cycles of 15 s at 95°C and 60 s at 60°C. After completion of these cycles, melting-curve data were collected to verify PCR specificity, contamination and absence of primer dimers.

mRNA variation quantification

Biological samples were prepared in independent triplicates, and qPCR analyses were performed in technical triplicates, bringing to 6 the number of analysis performed for each tested condition. Four internal control genes were used: *ALG9*, *TFC1*, *UBC6* and *TAF10*, chosen on their demonstrated expression stability over culture conditions (Teste et al., 2009). All data were calculated using the gene expression module of the BIORAD iQ5 software, using the $2\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). Presented data are average of biological triplicates, normalized to the internal control genes values, with standard deviation. Results were then expressed in terms of mRNA normalized variations, relative to the appropriate reference strain for the observed condition (H or H^H). Although these reference strains do not contain the genes *CRTI*, *CRTYB* and *CRTZ*, some nonspecific amplification was detectable in RT-qPCR when the corresponding cDNAs were targeted. Results were normalized by arbitrarily set at 1 the level of signal observed in the reference strains, and expressing mRNA levels in the tested conditions in fold changes relative to the appropriate reference.

pRS426 copy number determination

Nucleic acids were extracted as described previously. RNA contamination was eliminated by a DNase-free RNase A treatment (Thermo Scientific, Waltham, MA) at 37°C during 10 minutes,

followed by 10 minutes at 45°C and 10 minutes at 60°C. After DNA quantification, 20 ng were added to the qPCR Master Mix targeting *URA3* gene to make absolute quantification of pRS426, based on a standard curve made of known quantities of pRS426 added to DNA extracted from a native CEN.PK 113-7D strain.

3. Results

Zeaxanthin pathway library generation

We wanted to understand if promoters of reported strengths (Sun et al., 2012) can be used to impose a precise transcriptional regulation of expression level of genes involved in a synthetic pathway in yeast. Individual ORFs encoding a synthetic zeaxanthin pathway were combinatorially placed under the control of yeast constitutive promoters having different strengths: TEF1p (strong), PGI1p (weak), PDC1p, GPDp, ENO2p and TEF2p (medium), and three independent terminators (see Materials and Methods). An *in vitro* recombination method was used to quickly and reliably obtain the various individual constructs (Supplementary Figure 1).

The production, in yeast, of mevalonate by HMG-CoA reductase is known to be rate-limiting for the metabolic flux leading to sterols (Donald et al., 1997) and carotenoids (Verwaal et al., 2007; Yan et al., 2012). Therefore, to faithfully reproduce the deregulation strategy usually applied when producing terpene compounds, and to potentially visualize an effect of precursor boost on heterologous genes expression, a gene encoding the catalytic subunit of HMG-CoA reductase (*tHMG1*) from *S. cerevisiae* was cloned on pCM185 (pCM185/*tHMG1*).

Seven strains were generated, 6 upon transformation with the 6 plasmids harboring various promoter strengths and one corresponded to the transformation with the 7th plasmid harboring medium strengths promoters. The reference strain was transformed with the empty pRS426. These 8 generated strains (Table 1) were further transformed with pCM185 or pCM185/*tHMG1*, leading to 16 different strains named A to H when transformed with empty pCM185 (named set 1 throughout the rest of this document) or A⁺H to H⁺H when transformed with pCM185/*tHMG1* (named set 2 throughout the rest of this document).

***S. cerevisiae* strains, culture and sampling**

We first determined that no significant differences existed between the growth curves the zeaxanthin producing strains and the reference strains (Supplementary Figure 2). During batch growth on glucose, yeast cells shift from fermentation to respiration (diauxic shift) after exhausting the primary carbon source, causing extensive reprogramming of the transcription machinery (Galdieri et al., 2010). *GPD*, *PDC1*, *PGI1* and *ENO2* genes are involved in both glucose and ethanol assimilation processes. Consequently, the associated promoters are potentially sensitive to the cell metabolic state. We analyzed by NMR the supernatant of the culture medium of the strains A, C, D and H (Figure 2). After 15 hours of growth, the medium contained glucose as well as ethanol produced by fermentation. Interestingly, the glucose concentration was slightly lower in strain H than in any other strain harboring the zeaxanthin pathway, although the difference between the growth curves measured for strain H compared to the zeaxanthin strain was within error range. However, it is possible that the reference strain H consumed glucose slightly faster than any of the zeaxanthin producing strains. On the contrary, after 40 hours of growth, only ethanol could be found in the medium. We therefore determined that at 15 hours of growth, the strains were still fermenting glucose (preferred carbon source) while at 40 hours of growth, they had shifted to respiratory conditions (consuming ethanol). mRNA extraction and quantification were then performed either in fermentative context (15 hours, glucose) or in respiratory context (40 hours, ethanol).

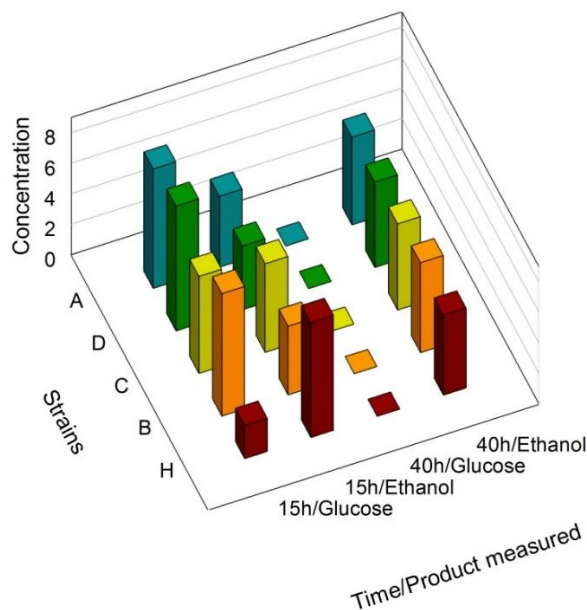


Figure 2: NMR determination of glucose and ethanol concentrations in supernatants of strains A, B, C and D and H after 15 hours and 40 hours of growth.

pRS426 copy number quantification

Prior to analyze the transcription variation of the combinatorial metabolic library genes, we investigated whether the copy number of the pRS426 derivatives changed in the different strains. We determined that strain H contained approximately 28 ± 0.1 copies of the empty pRS426, a regular value for a 2μ based plasmid (Romanos et al., 1992) (Figure 3). Using strain F as a representative for all zeaxanthin producing strains, we determined that the pRS426/F copy number was 2.0 ± 0.2 copies per cell only. It has been demonstrated that in case of deleterious gene products, a decrease in the plasmid copy number was observed (Fang et al., 2011). Zeaxanthin biosynthesis may lead to a certain toxicity, resulting in a positive selection pressure on cells harboring a reduced number of the pRS426 derivatives. Interestingly, adding a second, monocopy plasmid (pCM185/*tHMG1*) in these two strains (F and H, generating F^{+H} and H^{+H} respectively) promoted a systematic decrease of the copy number of pRS426 derivatives (6.0 ± 0.3 copies for H^{+H} compared to 28 ± 0.1 for strain H and about 0.5 ± 0.3 copy for F^{+H} compared to 2.0 ± 0.2 copies in strain F). This suggests a possible interference between the two plasmids, as already observed in *Escherichia coli*

(Gruber et al., 2008). As a conclusion, the copy number of pRS426 derivatives was affected by the zeaxanthin pathway expression and the presence of pCM185/*tHMG1*.

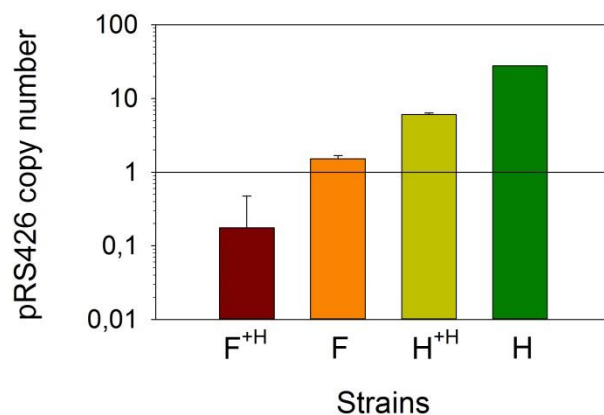


Figure 3: Absolute quantification of the copy number of pRS426/F in strains F and F^{+H}, compared with pRS426 in reference strains H and H^{+H}.

Expression profile of the pRS426 marker

Given the discrepancies in the copy number of pRS426 derivatives when *tHMG1* or carotenoid genes were expressed, we wanted to confirm if this variation was reflected in mRNA levels of the *URA3* gene present in the pRS426 backbone (Supplementary Figure 3). mRNA levels of *URA3* in each set of strains (set 1 or set 2) were averaged, and standardized to the levels calculated from the reference strain H or H⁺H, respectively (Figure 4). First, the fairly minor standard deviation showed that the variation of the *URA3* mRNA levels was minimal between the different strains. Secondly, the mean values of mRNA levels were almost constant over the different conditions (small variations being within error range), but dramatically reduced (50 to 100 folds) in zeaxanthin producing strains, when compared to the reference. As mentioned earlier, the copy number of pRS426 derivatives decreased down to approximately 1 in the strains harboring the zeaxanthin pathway compared to approximately 30 copies in the reference strains. Consequently, the strong decrease in the *URA3* mRNA levels could hardly be explained by the sole reduction of the copy number of the pRS426 derivatives. The three heterologous genes were placed far away and downstream the *URA3* gene in pRS426 (Supplementary Figure 3), theoretically limiting possible transcriptional effects between them. Nonetheless, transcription of *CRTI*, *CRTYB* and *CRTZ* may interfere in some way with *URA3* promoter activity either due to some competition between the different promoters or possible changes in the plasmid DNA structure, hampering a correct DNA unwinding prior transcription. A regulation of endogenous *ura3-52* gene present in CEN.PK may also take part in the *URA3* mRNA level reduction. Indeed, the addition of a plasmid containing the *URA3* gene may modify transcriptional regulation of endogenous uracil pathway.

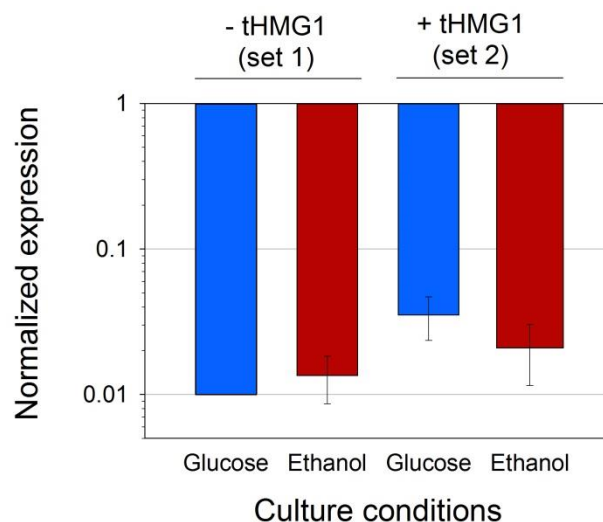


Figure 4: Average *URA3* mRNA level in the four studied conditions: with (set 2) or without (set 1) *tHMG1* expression (indicated on the top of the graph), in glucose (blue bars) or ethanol (red bars) metabolism. Results were normalized with *URA3* signal in strain H for strains from set 1, and with *URA3* signal in strain H⁺H for strains from set 2. Error bars represent one standard deviation.

Controls of plasmid integrity and cell physiology

In order to ascertain reproducibility of our quantifications, several controls experiments were conducted. To ensure that no recombination events occurred between one of the plasmids and *S. cerevisiae* genome, or between the two plasmids, we verified their integrity after transformation and cultivation. Restriction analyses (Supplementary Figure 4), performed on extracted plasmids, demonstrated that the specific bands of pRS426 and pCM185 derivatives corresponded to normal restriction profiles. We then verified if the expression of the different pRS426 derivatives could lead to changes in the transcriptional regulation of some specific genes. mRNA levels of a restricted number of genes involved in the production of endogenous carotenoids precursors were evaluated in the 14 engineered strains. *BTS1*, *ERG20*, *ERG19*, *ERG8*, *ERG12* and *HMG1* were chosen as targets because of their implication in the production of GGPP, the precursor involved in the first step of the heterologous biosynthetic pathway (Figure 1). For *HMG1* mRNA quantification, primers were designed to avoid detection of transcripts from *tHMG1*. There was almost no impact of the carotenoid pathway on the transcriptional response of the 6 chosen genes when *tHMG1* was not expressed (level changes are within the error range) (Figure 5A). However, when *tHMG1* was expressed, there was a slight tendency to overexpress the GGPP pathway genes (Figure 5B). In any

case, the effects were minor and equivalent between all producing strains (no more than 3x the reference level) compared to non-producing strains H and H⁺. Therefore, the production of zeaxanthin in our conditions did probably not strongly impact the pool of precursors and no physiological adaptation was therefore seen at the level of GGPP pathway regulation.

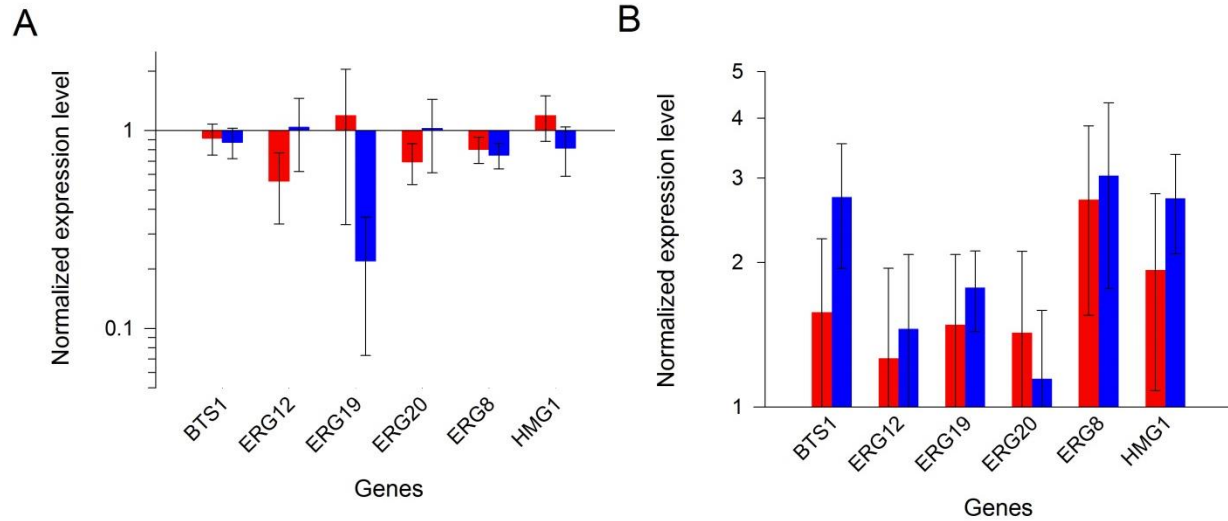


Figure 5: Expression profiles of endogenous GGPP pathway genes. Bars represent the mean value of the different gene expressions in the strains from set 1 (panel A, red and blue bars for glucose or ethanol metabolism, respectively), or in the strains from set 2 (panel B, red and blue bars for glucose or ethanol metabolism, respectively). Error bars represent one standard deviation.

Absence of correlation between promoters strengths and mRNA levels

We analyzed the normalized mRNA levels of *CRTI*, *CRTYB* and *CRTZ* genes for each strain and in every growth context. We again emphasize that only the promoter nature (and strength) in front of each ORF is different in strains from set 1 or set 2, independently (Table 1). Surprisingly, Figure 6 highlights one major profile in these sets of strains that can be described as follows: *CRTI* is the most expressed gene, followed by *CRTYB* and finally by *CRTZ*. The general tendency shows lower levels of mRNAs in set 2 (in accordance with the pRS426 derivatives copy number decrease when *tHMG1* was expressed) or in respiratory conditions. However, in strain G⁺, *CRTZ* mRNA level (in glucose) was not impacted by *tHMG1* and even augmented compared to the one measured in strain G. Furthermore, in strain G⁺, *CRTZ* mRNA level was not modified between glucose and ethanol conditions, unlike *CRTI* and *CRTYB* mRNA levels. In construct G, TEF2p, controlling *CRTZ* transcription, is the only promoter not to be involved in glucose or ethanol assimilation. This

may be related to the *CRTZ* mRNA level stability over culture conditions. This results in *CRTYB*, *CRTI* and *CRTZ* mRNA levels variation less than 15-fold in strain G⁺H, a feature that is in good accordance with the expected mRNA levels imposed by the chosen promoters (medium strength promoters for all three genes) (Sun et al., 2012).

In conclusion, in our experimental conditions, mRNA levels did not reflect the previously described strengths of promoters (Sun et al., 2012). This promoter-independent control may result from several different mechanisms. The ORF nature, *via* the sequence related variation of mRNA half-life, may strongly influence the mRNA level. Occurrence of cryptic promoters in ORFs (Pedersen et al., 1999) might also promote or inhibit transcription independently of the promoter-induced regulation. Gene terminators are also known to influence mRNA stabilization (Curran et al., 2013; Yamanishi et al., 2013). Curran et al. (Curran et al., 2013) reported that the terminator of *CYC1* gene (*CYC1t*) was a poor expression enhancing terminator. Therefore, the use of *CYC1t* behind *CRTZ* may explain the low levels of its corresponding mRNA. Overall, the quasi absence of correlation between the strengths of the different promoters and their corresponding mRNA levels strongly suggests that the relative ORF positions on the plasmid may drive transcription efficiency.

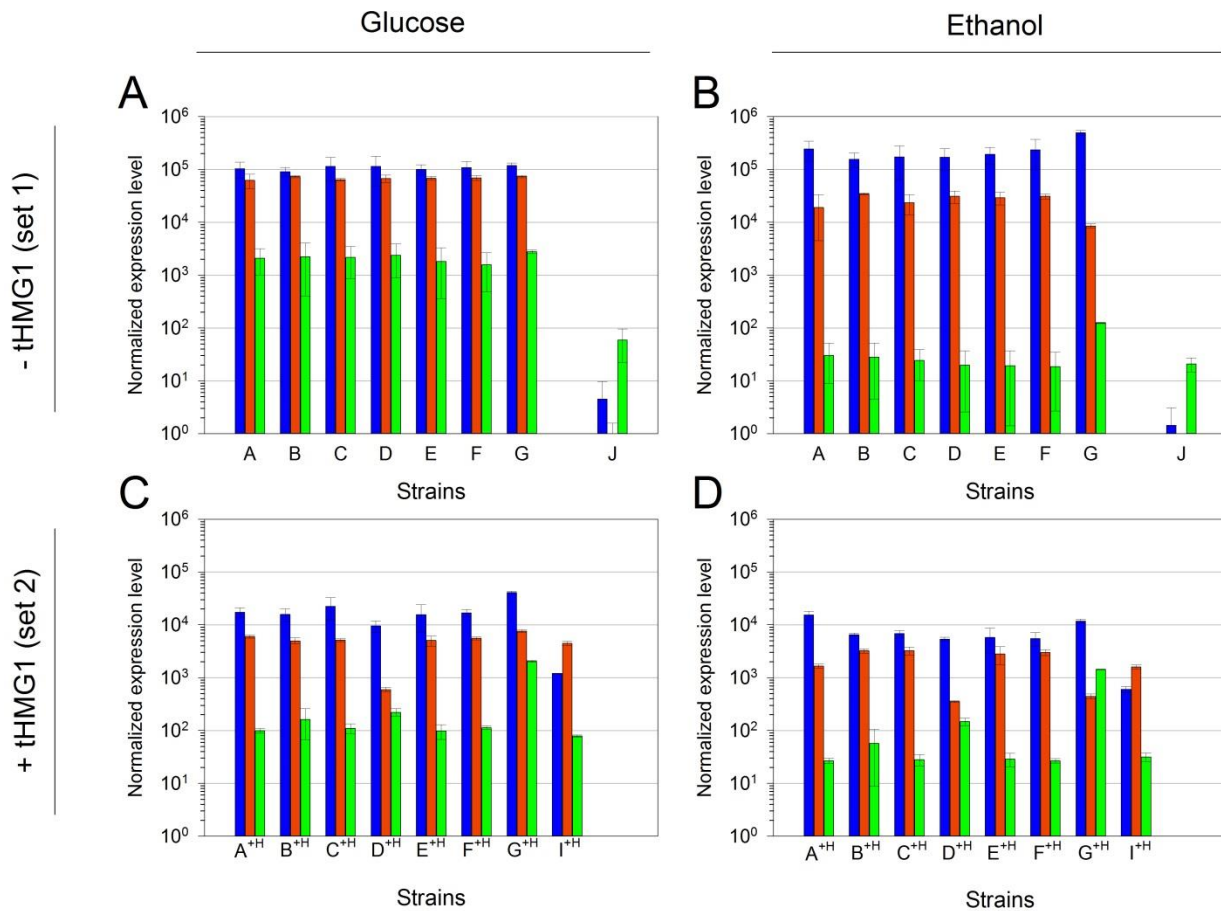


Figure 6: Quantitative RT-PCR analysis of mRNA levels of the three implemented carotenoid genes. (A) mRNA levels in strains A to J, in glucose metabolism (normalized with strain H); (B) mRNA levels in strains A to J, in ethanol metabolism (normalized with strain H); (C) mRNA levels in strains A^H to I^H , in glucose metabolism (normalized with strain H^H); (D) mRNA levels in strains A^H to I^H , in ethanol metabolism (normalized with strain H^H). Bar colors are as follows: *CRTI*, blue; *CRTYB*, orange and *CRTZ*, green. Error bars represent one standard deviation. mRNA levels for strain J in glucose metabolism were calculated with 2 values. Error bars represent maximum values.

Gene orientation and order strongly affect promoters strengths

We tested hypothesis that mRNA levels in our system was genes position dependent. We designed two new plasmids, keeping the promoter-ORF-terminator associations of pRS426/E, but swapping the positions of the three genes (pRS426/I) or reversing the *CRTYB* transcription unit orientation (pRS426/J) (Table 1). Due to very poor transformation efficiency of CEN.PK with these new plasmids, only strains I^H and J were generated.

Compared to strain E^{+H}, *CRTI* mRNA level in strain I^{+H} was decreased by a factor of 10 while the mRNA levels of *CRTYB* and *CRTZ* were constant (Figure 6 C and D). This phenomenon was not dependent on culture conditions, as it was true in glucose and in ethanol metabolisms. A failure of *CRTYB* transcription termination may explain the observed diminution of *CRTI* mRNA level. Although TEF2t (terminator associated to *CRTYB*) was not tested in Curran et al.'s study (2013), its efficiency may not be sufficient to promote the normal release of the entire transcription machinery at the termination site. This phenomenon may hinder the RNA polymerase fixation at *CRTI* promoter. Therefore, changing the *CRTI* transcription unit position could alter the promoter strength, demonstrating that the construction structure and more precisely the relative gene position interfered with its expression level.

Relative quantifications of mRNA levels in strain J pointed a strong decrease of *CRTI* and *CRTYB* mRNAs (Figure 6 A and B). In this strain, *CRTYB* mRNA was absent and *CRTI* mRNA level was strongly decreased, while the *CRTZ* mRNA level was lowered only by a factor of ten. To ascertain that no recombination events could have modified the plasmid harboring the zeaxanthin pathway, we verified (restriction and sequence analysis) that all three genes were still present and correctly ordered on the pRS426/J extracted from strain J (Supplementary Figure 4). In pRS426/J, *CRTI* and *CRTYB* were in a convergent configuration, potentially leading to a collision between the two opposite RNA polymerases. This phenomenon may render difficult their progress toward the terminator. Furthermore, during the transcriptional process, the RNA polymerases progress on DNA making positive supercoils on both strands simultaneously (Liu and Wang, 1987). This may create a region of hyper-supercoiling that could possibly prevent further advancement in either direction on the template, leading to the suppression of the transcriptional processes of *CRTI* and *CRTYB*.

Hence, in our experiments, *CRTI* mRNA level seemed to be affected by the gene position and orientation while *CRTYB* mRNA level was impacted only by the gene orientation. *CRTYB* mRNA level stability over tested positions in the plasmid tends to reveal that *CRTYB* ORF nature imposed a regulation of its transcription, and that the observed annihilation of *CRTI* and *CRTYB* transcription in strain J was probably due to poor terminator (TEF2t) efficiency. *CRTZ* mRNA level was the most stable one over the different tested conditions. *CRTZ* ORF nature or terminator (CYC1t), or both, seemed to impose a strong transcriptional regulation. As CYC1t was described

by Curran et al. (2013) as a non-enhancing mRNA stability terminator, *CRTZ* ORF nature was probably involved in its mRNA level stability.

4. Discussion

Parts of the effects seen in these two last constructions might be attributed to transcription interference (TI). TI is defined as the suppressive influence of one transcriptional process, directly and in cis, on a second transcriptional process (Shearwin et al., 2005). This phenomenon was well-studied with native *S. cerevisiae* genes (Prescott and Proudfoot, 2002; Wang et al., 2014) but only two studies addressed the case of TI in heterologous expression, demonstrating transcription inhibition effects only in the case of divergent genes (Bae et al., 2008; Lee et al., 2014). It is known that the genome of *S. cerevisiae* is highly compact (an ORF every 2 kb in average) and less than 6% of the convergent adjacent genes show a pattern of co-expression. This suggests that a counter-selection pressure for co-transcribed convergent genes exists (Dujon, 1996). Prescott and Proudfoot (Prescott and Proudfoot, 2002) described a drastic effect on the transcription processes of two convergent genes when intergenic sequences containing genes terminators were removed. They interpreted this phenomenon by a collision, in the absence of terminators, between RNA polymerases simultaneously transcribing the two genes. These effects were abolished when the two transcriptional processes were not simultaneous, confirming the strong TI in the case of simultaneous transcription. Contrarily to previous studies (Bae et al., 2008; Lee et al., 2014), our results provide a clear indication that relative positioning of three contiguous transcription units and their orientations probably induce TI on a multicopy plasmid. Large TI effects demonstrated in our design may be related to terminators efficiency, and are probably accentuated by the absence of extra-space between our three transcription units.

We show here that the strengths of chosen promoters, as given by the available literature data, are not systematically conserved when used for the reconstruction of foreign pathways. TI, ORF nature, and possibly terminator may also counterbalance the initially imposed regulation. In strains from set 1, we observed a factor of 1000 between *CRTI* and *CRTZ* mRNA levels, a surprising result considering that the expected regulation of mRNA levels by promoters strengths is

supposedly a factor of ten between the strongest and the weakest promoters (Sun et al., 2012). Overall, we propose that combinatorial metabolic libraries (similar to COMPACTER method (Du et al., 2012)), associating various promoters/terminators to heterologous genes and varying their positions is probably a quick and reliable method to generate diversity at the level of transcription regulation. However, due to possible unexpected transcriptional effects, we propose that a systematic analysis of all heterologous genes mRNA levels should be conducted to link the desired metabolite production level to the real pattern of transcription. This may allow researchers to select the best combination according to the production of the desired molecule.

Abbreviations

GGPP, geranylgeranyl pyrophosphate; **GFP**, green fluorescent protein; **qPCR**, quantitative PCR; **RT-qPCR**, reverse transcription quantitative PCR; **tHMG1**, truncated HMG-CoA reductase.

5. Acknowledgement

This work was supported by the grant 11050480 from Région Midi-Pyrénées (financial support for “accueil d’équipes d’excellence”). The PhD grant for M.C. was funded by the Région Midi-Pyrénées and ADRIAF. The authors wish to thank Jean-Luc Parrou and Marie-Ange Teste from the LISBP-INSA Toulouse for their precious help with qPCR design. This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors. A statement about the principal source of funding should be inserted if appropriate, including grant numbers if appropriate.

6. Author's contribution

M.C. designed the study, made experiments and wrote this manuscript, G.T. and D.P. designed the study and wrote this manuscript.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7. References

- Ajikumar, P.K., Xiao, W.H., Tyo, K.E., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B., and Stephanopoulos, G. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* 330, 70-74. doi: 10.1126/science.1191652.
- Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. (2005). Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences, USA* 102, 12678-12683. doi: 10.1073/pnas.0504604102.
- Bae, J.Y., Laplaza, J., and Jeffries, T.W. (2008). Effects of gene orientation and use of multiple promoters on the expression of *XYL1* and *XYL2* in *Saccharomyces cerevisiae*. *Appl Biochem Biotechnol* 145, 69-78. doi: 10.1007/s12010-007-8076-0.
- Blazeck, J., and Alper, H.S. (2013). Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnology Journal* 8, 46-58. doi: 10.1002/biot.201200120.
- Christianson, T.W., Sikorski, R.S., Dante, M., Shero, J.H., and Hieter, P. (1992). Multifunctional yeast high-copy-number shuttle vectors. *Gene* 110, 119-122.
- Curran, K.A., Karim, A.S., Gupta, A., and Alper, H.S. (2013). Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab Eng* 19, 88-97. doi: 10.1016/j.ymben.2013.07.001.
- Da Silva, N.A., and Srikrishnan, S. (2012). Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 12, 197-214. doi: 10.1111/j.1567-1364.2011.00769.x.
- Donald, K.A., Hampton, R.Y., and Fritz, I.B. (1997). Effects of overproduction of the catalytic domain of 3-hydroxy-3-methylglutaryl coenzyme A reductase on squalene synthesis in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 63, 3341-3344.
- Du, J., Yuan, Y., Si, T., Lian, J., and Zhao, H. (2012). Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res* 40, e142. doi: 10.1093/nar/gks549.
- Dujon, B. (1996). The yeast genome project: what did we learn? *Trends in Genetics* 12, 263-270.
- Fang, F., Salmon, K., Shen, M.W., Aeling, K.A., Ito, E., Irwin, B., Tran, U.P., Hatfield, G.W., Da Silva, N.A., and Sandmeyer, S. (2011). A vector set for systematic metabolic engineering in *Saccharomyces cerevisiae*. *Yeast* 28, 123-136. doi: 10.1002/yea.1824.

- Galdieri, L., Mehrotra, S., Yu, S., and Vancura, A. 2010. Transcriptional regulation in yeast during diauxic shift and stationary phase. *OMICS* [Online], 14. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20863251> [Accessed Dec].
- Gari, E., Piedrafita, L., Aldea, M., and Herrero, E. (1997). A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast* 13, 837-848. doi: 10.1002/(SICI)1097-0061(199707)13:9<837::AID-YEA145>3.0.CO;2-T.
- Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343-345. doi: 10.1038/nmeth.1318.
- Gietz, D., St Jean, A., Woods, R.A., and Schiestl, R.H. (1992). Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res* 20, 1425.
- Glick, B.R. (1995). Metabolic load and heterologous gene expression. *Biotechnol Adv* 13, 247-261.
- Gruber, D.F., Pieribone, V.A., Porton, B., and Kao, H.T. (2008). Strict regulation of gene expression from a high-copy plasmid utilizing a dual vector system. *Protein Expression and Purification* 60, 53-57. doi: 10.1016/j.pep.2008.03.014.
- Karim, A.S., Curran, K.A., and Alper, H.S. (2013). Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast Res* 13, 107-116. doi: 10.1111/1567-1364.12016.
- Kim, H.J., Turner, T.L., and Jin, Y.S. (2013). Combinatorial genetic perturbation to refine metabolic circuits for producing biofuels and biochemicals. *Biotechnology Advances* 31, 976-985. doi: 10.1016/j.biotechadv.2013.03.010.
- Kocharin, K., Chen, Y., Siewers, V., and Nielsen, J. (2012). Engineering of acetyl-CoA metabolism for the improved production of polyhydroxybutyrate in *Saccharomyces cerevisiae*. *AMB Express* 2, 52. doi: 10.1186/2191-0855-2-52.
- Lee, M.E., Aswani, A., Han, A.S., Tomlin, C.J., and Dueber, J.E. (2013). Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res* 41, 10668-10678. doi: 10.1093/nar/gkt809.
- Lee, T.J., Parikh, R.Y., Weitz, J.S., and Kim, H.D. (2014). Suppression of expression between adjacent genes within heterologous modules in yeast. *G3 (Bethesda)* 4, 109-116. doi: 10.1534/g3.113.007922.
- Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A* 84, 7024-7027.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402-408. doi: 10.1006/meth.2001.1262.
- Lu, C., and Jeffries, T. (2007). Shuffling of promoters for multiple genes to optimize xylose fermentation in an engineered *Saccharomyces cerevisiae* strain. *Appl Environ Microbiol* 73, 6072-6077. doi: 10.1128/AEM.00955-07.
- Mnaimneh, S., Davierwala, A.P., Haynes, J., Moffat, J., Peng, W.T., Zhang, W., Yang, X., Pootoolal, J., Chua, G., Lopez, A., Trochesset, M., Morse, D., Krogan, N.J., Hiley, S.L., Li, Z., Morris, Q., Grigull, J., Mitsakakis, N., Roberts, C.J., Greenblatt, J.F., Boone, C., Kaiser, C.A., Andrews, B.J., and Hughes, T.R. (2004). Exploration of essential gene functions via titratable promoter alleles. *Cell* 118, 31-44. doi: 10.1016/j.cell.2004.06.013.
- Neubauer, P., Lin, H.Y., and Mathisizik, B. (2003). Metabolic load of recombinant protein production: inhibition of cellular capacities for glucose uptake and respiration after

- p>induction of a heterologous gene in
- Escherichia coli*
- .
- Biotechnol Bioeng*
- 83, 53-64. doi: 10.1002/bit.10645.
- Partow, S., Siewers, V., Bjorn, S., Nielsen, J., and Maury, J. (2010). Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast* 27, 955-964. doi: 10.1002/yea.1806.
- Pedersen, A.G., Baldi, P., Chauvin, Y., and Brunak, S. (1999). The biology of eukaryotic promoter prediction--a review. *Comput Chem* 23, 191-207.
- Prescott, E.M., and Proudfoot, N.J. (2002). Transcriptional collision between convergent genes in budding yeast. *Proceedings of the National Academy of Sciences, USA* 99, 8796-8801. doi: 10.1073/pnas.132270899.
- Ramon, A., and Smith, H.O. (2011). Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. *Biotechnol Lett* 33, 549-555. doi: 10.1007/s10529-010-0455-x.
- Ro, D.K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C., Withers, S.T., Shiba, Y., Sarpong, R., and Keasling, J.D. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940-943. doi: 10.1038/nature04640.
- Romanos, M.A., Scorer, C.A., and Clare, J.J. (1992). Foreign gene expression in yeast: a review. *Yeast* 8, 423-488. doi: 10.1002/yea.320080602.
- Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference--a crash course. *Trends Genet* 21, 339-345. doi: 10.1016/j.tig.2005.04.009.
- Siegele, D.A., and Hu, J.C. (1997). Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proceedings of the National Academy of Sciences, USA* 94, 8168-8172.
- Steen, E.J., Chan, R., Prasad, N., Myers, S., Petzold, C.J., Redding, A., Ouellet, M., and Keasling, J.D. (2008). Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. *Microb Cell Fact* 7, 36. doi: 10.1186/1475-2859-7-36.
- Sun, J., Shao, Z., Zhao, H., Nair, N., Wen, F., Xu, J.H., and Zhao, H. (2012). Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 109, 2082-2092. doi: 10.1002/bit.24481.
- Teste, M.A., Duquenne, M., Francois, J.M., and Parrou, J.L. (2009). Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomyces cerevisiae*. *BMC Mol Biol* 10, 99. doi: 10.1186/1471-2199-10-99.
- Ukibe, K., Hashida, K., Yoshida, N., and Takagi, H. (2009). Metabolic engineering of *Saccharomyces cerevisiae* for astaxanthin production and oxidative stress tolerance. *Appl Environ Microbiol* 75, 7205-7211. doi: 10.1128/AEM.01249-09.
- Verwaal, R., Wang, J., Meijnen, J.P., Visser, H., Sandmann, G., Van Den Berg, J.A., and Van Ooyen, A.J. (2007). High-level production of beta-carotene in *Saccharomyces cerevisiae* by successive transformation with carotenogenic genes from *Xanthophyllomyces dendrorhous*. *Appl Environ Microbiol* 73, 4342-4350. doi: 10.1128/AEM.02759-06.
- Wang, L., Jiang, N., Wang, L., Fang, O., Leach, L.J., Hu, X., and Luo, Z. 2014. 3' Untranslated regions mediate transcriptional interference between convergent genes both locally and ectopically in *Saccharomyces cerevisiae*. *PLoS Genet* [Online], 10. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24465217> [Accessed Jan].

- Yamanishi, M., Ito, Y., Kintaka, R., Imamura, C., Katahira, S., Ikeuchi, A., Moriya, H., and Matsuyama, T. (2013). A genome-wide activity assessment of terminator regions in *Saccharomyces cerevisiae* provides a "terminatome" toolbox. *ACS Synthetic Biology* 2, 337-347. doi: 10.1021/sb300116y.
- Yan, G.L., Wen, K.R., and Duan, C.Q. (2012). Enhancement of beta-carotene production by over-expression of HMG-CoA reductase coupled with addition of ergosterol biosynthesis inhibitors in recombinant *Saccharomyces cerevisiae*. *Current Microbiology* 64, 159-163. doi: 10.1007/s00284-011-0044-9.

Supplementary Material

Transcription interference and gene nature strongly affect promoter strength in a reconstituted metabolic pathway.

Marie Carquet^{1,2,3}, Denis Pompon^{1,2,3}, Gilles Truan^{1,2,3*}

¹Université de Toulouse; INSA, UPS, INP; LISBP, Toulouse, France

²INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, Toulouse, France

³CNRS, UMR5504, Toulouse, France*

* **Correspondence:** Gilles Truan, ¹Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France

gilles.truan@insa-toulouse.fr

1. Supplementary Data

CRTZ ORF sequence optimized according to *Saccharomyces cerevisiae* codon bias:

```
1      ATGTTATGGA TCTGGAACGC TTTGATCGTT TTCGTTACTG TTATCGGTAT GGAAGTTATC
61     GCTGCTTTGG CCCACAAGTA CATTATGCAC GGTTGGGGTT GGGGTTGGCA CTTATCCCAC
121    CACGAACCAC GTAAGGGTGC CTTCGAAGTT AACGACTTGT ACGCCGTTGT TTTCGCCGCT
181    TTGAGCATCT TGTTGATTTA CTTGGGTAGC ACCGGTATGT GGCCATTACA ATGGATCGGT
241    GCTGGTATGA CCGCCTACGG TTTATTGTAC TTCATGGTTC ACGACGGTTT AGTTCACCAA
301    CGTTGGCCAT TCCGTTACAT TCCACGTAAG GGTTACTTAA AGCGTTTGTA CATGGCCCAC
361    AGAATGCACC ACGCCGTTTC TGGTAAAGAA GGTTGTGTTA GTTTCGGTTT CTTGTACGCC
421    CCACCATTAT CCAAGTTGCA AGCTACTTTG AGAGAACGTC ACGGTGCTAG AGCCGGTGCC
481    GCCAGAGATG CTCAAGGTGG TGAAGATGAA CCAGCCAGCG GTAAGTGA
```

2. Supplementary Figures and tables

Supplementary Tables

Supplementary Table 1 : Primer sequences used to assemble the varying promoter zeaxanthin pathways.

Target	final construct	Primer name ^{a)}	Primer sequence (5'→ 3') ^{b)}
CRTI	TEF1p_CRTI	pRS426_TEF1p_F	<u>ATTGGGTACCGGGCCCCCCC</u> ATAGCTTCAAAATGTTTCTA
		TEF1p_CRTI_R	<u>TTGTTCTTTTCCCATTTT</u> GTAATTAAACTTAGAT
		CRTI_TEF1p_F	<u>AGTTTTAATTACAAA</u> ATGGGAAAAGAACAAGATCA
	PDC1p_CRTI	pRS426_PDC1p_F	<u>ATTGGGTACCGGGCCCCCCC</u> ATGCGACTGGGTGAGCATA
		PDC1p_CRTI_R	<u>TTGTTCTTTTCCCATTTT</u> GATTGATTTGACTGTGT
		CRTI_PDC1p_F	<u>GTCAAATCAATCAAA</u> ATGGGAAAAGAACAAGATCA
	PGI1p_CRTI	pRS426_PGI1p_F	<u>ATTGGGTACCGGGCCCCCCC</u> TAACAAAAATCACGATCTGG
		PGI1p_CRTI_R	<u>TTGTTCTTTTCCCATTTT</u> AGGCTGGTATCTTGAT
		CRTI_PGI1p_F	<u>GATACCAGCCTAAAA</u> ATGGGAAAAGAACAAGATCA
	GPDp_CRTI	pRS426_GPDp_F	<u>ATTGGGTACCGGGCCCCCCC</u> AGTTTATCATTATCAATACT
		GPDp_CRTI_R	<u>TTGTTCTTTTCCCATAT</u> CCGTCGAACTAAGTTCT
		CRTI_GPDp_F	<u>TTAGTTTCGACGGATAT</u> GGGAAAAGAACAAGATCA
	CRTI_ADH1t	CRTI_ADH1t_R	<u>CATAAGAAATTTCGCTT</u> CAGAAAGCAAGAACACCAA
		ADH1t_CRTI_F	<u>GTTCTTGCTTTCTGA</u> AGCGAATTTCTTATGATTTA
	Primers designed to overlap with next cassette	ADH1t_TEF1p_R	<u>ACATTTTGAAGCTAT</u> GGCATGCCGGTAGAGGTGTG
		ADH1t_PDC1p_R	<u>TCACCCAGTCGCAT</u> GGGCATGCCGGTAGAGGTGTG
		ADH1t_PGI1p_R	<u>TCGTGATTTTTGTTA</u> GGCATGCCGGTAGAGGTGTG
		ADH1t_ENO2_R	<u>CCGCAGCGTCGACAC</u> GGCATGCCGGTAGAGGTGTG
CRTYB	TEF1p_CRTYB	TEF1p_ADH1t_F	<u>CTCTACCGGCATGCC</u> ATAGCTTCAAAATGTTTCTA
		TEF1p_CRTYB_R	<u>TGCGAGAGCCGTCAT</u> TTTGTAATTAAACTTAGAT
		CRTYB_TEF1p_F	<u>AGTTTTAATTACAAA</u> ATGACGGCTCTCGCATATTA
	PDC1p_CRTYB	PDC1p_ADH1t_F	<u>CTCTACCGGCATGCC</u> ATGCGACTGGGTGAGCATA
		PDC1p_CRTYB_R	<u>TGCGAGAGCCGTCAT</u> TTTGATTGATTTGACTGTGT
		CRTYB_PDC1p_F	<u>GTCAAATCAATCAAA</u> ATGACGGCTCTCGCATATTA
	PGI1p_CRTYB	PGI1p_ADH1t_F	<u>CTCTACCGGCATGCC</u> TAACAAAAATCACGATCTGG
		PGI1p_CRTYB_R	<u>TGCGAGAGCCGTCAT</u> TTTTAGGCTGGTATCTTGAT
		CRTYB_PGI1p_F	<u>GATACCAGCCTAAAA</u> ATGACGGCTCTCGCATATTA
	ENO2p_CRTYB	ENO2p_ADH1t_F	<u>CTCTACCGGCATGCC</u> GTGTCGACGCTGCGGGTATA
		ENO2p_CRTYB_R	<u>TGCGAGAGCCGTCAT</u> TATTATTGTATGTTATAGTA
		CRTYB_ENO2p_F	<u>TAACATACAATAATA</u> ATGACGGCTCTCGCATATTA
	CRTYB_TEF2t	CRTYB_TEF2t_R	<u>CAATAATTATTACTCT</u> TACTGCCCTTCCCATCCGC
		TEF2t_CRTYB_F	<u>TGGGAAGGGCAGTAA</u> GAGTAATAATTATTGCTTCC
	Primers designed to overlap with next cassette	TEF2t_TEF1p_R	<u>ACATTTTGAAGCTAT</u> GATGAGGCCGTCTTTTGTTG
		TEF2t_PDC1p_R	<u>TCACCCAGTCGCAT</u> GGATGAGGCCGTCTTTTGTTG
		TEF2t_PGI1p_R	<u>TCGTGATTTTTGTTA</u> GATGAGGCCGTCTTTTGTTG
		TEF2t_TEF2p_R	<u>TAAGTATACGGCCCC</u> GATGAGGCCGTCTTTTGTTG

Target	final construct	Primer name ^{a)}	Primer sequence (5'→ 3') ^{b)}
CRTZ	TEF1p_CRTZ	TEF1p_TEF2t_F	<u>AAAGACGGCCTCATC</u> ATAGCTTCAAAATGTTTCTA
		TEF1p_CRTZ_R	<u>CCAGATCCATAACAT</u> TTTTGTAATTAACCTTAGAT
		CRTZ_TEF1p_F	<u>AGTTTTAATTACAAA</u> ATGTTATGGATCTGGAACGC
	PDC1p_CRTZ	PDC1p_TEF2t_F	<u>AAAGACGGCCTCATC</u> CATGCGACTGGGTGAGCATA
		PDC1p_CRTZ_R	<u>CCAGATCCATAACAT</u> TTTTGATTGATTTGACTGTGT
		CRTZ_PDC1p_F	<u>GTCAAATCAATCAAA</u> ATGTTATGGATCTGGAACGC
	PGI1p_CRTZ	PGI1p_TEF2t_F	<u>AAAGACGGCCTCATC</u> TAACAAAAATCACGATCTGG
		PGI1p_CRTZ_R	<u>CCAGATCCATAACAT</u> TTTTAGGCTGGTATCTTGAT
		CRTZ_PGI1p_F	<u>GATACCAGCCTAAAA</u> ATGTTATGGATCTGGAACGC
	TEF2p_CRTZ	TEF2p_TEF2t_F	<u>AAAGACGGCCTCATC</u> GGGGCCGTATACTTACATAT
		TEF2p_CRTZ_R	<u>CCAGATCCATAACAT</u> GTTTAGTTAATTATAGTTCG
		CRTZ_TEF2p_F	<u>TATAATTAACTAAAC</u> ATGTTATGGATCTGGAACGC
	CRTZ_CYC1t	CRTZ_CYC1t_R	<u>CATGATGCGGCCCTC</u> TACTTACCGCTGGCTGGTT
		CYC1t_CRTZ_F	<u>GCCAGCGGTAAGTG</u> AGAGGGCCGCATCATGTAATT
		CYC1t_pRS426_R	<u>CTAAAGGGAACAAA</u> AGCTGGGGCCGCAAATTAAAGCCTTC
		RECpRS426_F	<u>TACGACTCACTATAG</u> GGCGAATTGGGTACCGGGCCCCCCC
		RECpRS426_R	<u>AAGCGCGCAATTA</u> ACCCTCACTAAAGGGAACAAAAGCTGG
<i>tHMG1</i>	pCM185/ <i>tHMG1</i>	<i>tHMG1</i> _pCM185_F	<u>ATTACCGGATCAAT</u> TCGGGGATGGACCAATTGGTGAAAAGTGAAG
		<i>tHMG1</i> _pCM185_R	<u>CAGGCCTGTTTAA</u> ACGGATCTTAGGATTTAATGCAGGTGACGGAC

a) Primer names are constructed as follow: target sequence_overlapping sequence_F/R as it is a Forward or a Reverse primer.

b) Overlapping sequences are shown in underlined letters.

Supplementary Table 2: Culture medium composition.

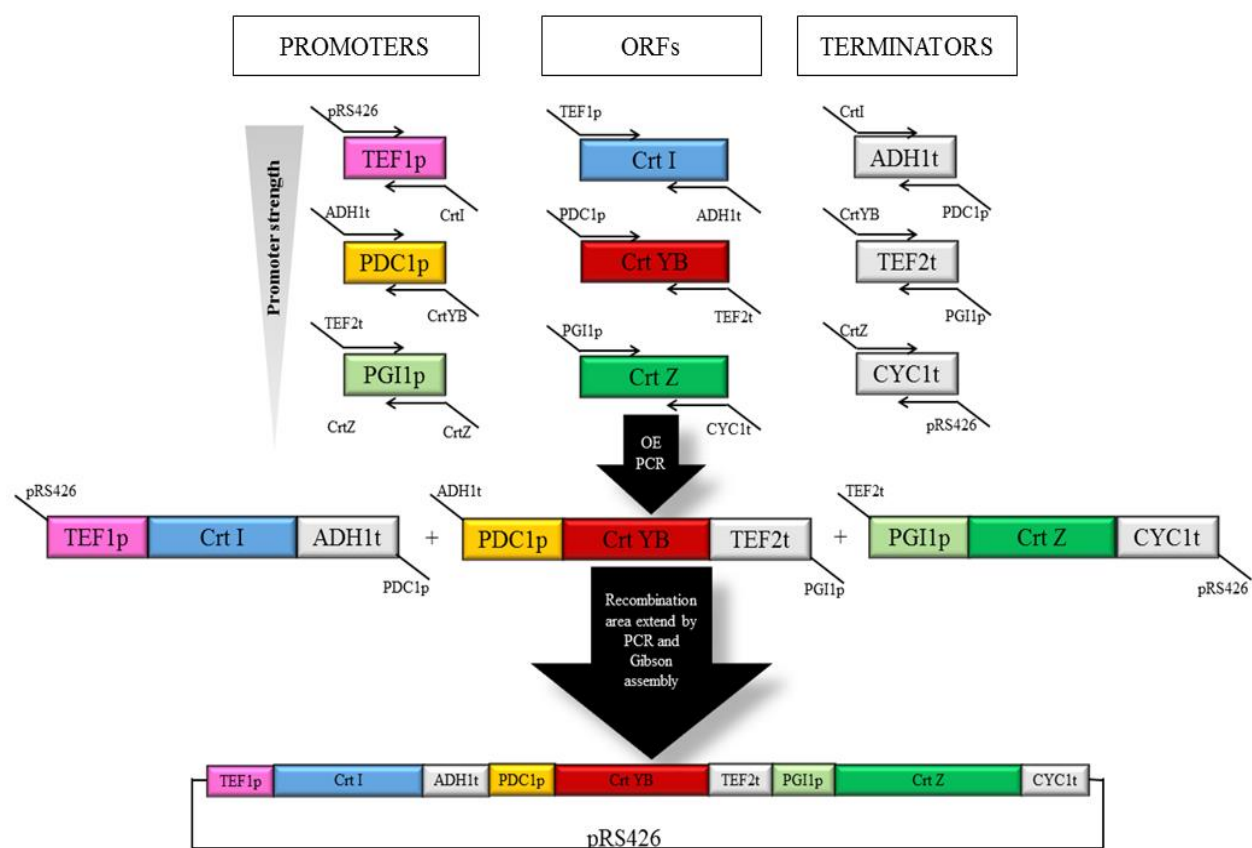
Salts	H3PO4 85%	0,66%
	KH2PO4	21 mM
	MgSO4, 7H2O	2,4 mM
	MnSO4, H2O	94,7 µM
	CuSO4, 5H2O	1,60 µM
	ZnSO4, 7H2O	52,2 µM
	CoCl2, 6H2O	11,8 µM
	Na2MoO4, 2H2O	10,3 µM
	H3BO3	121,4 µM
	Citric acid	2,60 µM
	KI	6 µM
	NiSO4, 6H2O	15,4 µM
	Trisodium citrate, 2H2O	85 µM
Vitamins	Thiamine HCl	6 µM
	Pyridoxine HCl	24,3 µM
	nicotinic acid	65 µM
	D-Biotin	0,20 µM
	Ca-D-Panthothenate	21 µM
	Myo-inositol	444,2 µM
Other components	CaCl2, 2H2O	448 µM
	FeCl3, 7H2O	173,4 µM
	Glucose	20 g/L
	Casaminoacids	10 g/L
	Leucine	0,01% final
	Histidine	0,01% final

Supplementary Table 3: Primers used for RT-qPCR analyses.

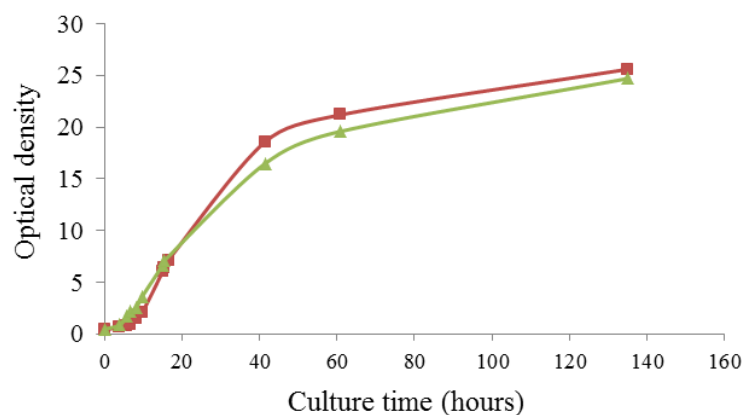
Target	forward Primer 5' > 3'	Reverse Primer 5' > 3'
<i>URA3</i>	AACAGTATAGAACCGTGGATGATGT	CTGTAACGTTACCCCTCTACCTTAG
<i>CRTI</i>	GTTACAGTCCTGCAGAAGAAGT	CTCTTGTCAGATAGACTCGAAGG
<i>CRTYB</i>	CGAACTCCTTAGAGGATACACCAC	GAGCCTGCTACACATAGACCATAG
<i>CRTZ</i>	TACTTGGGTAGCACCGGTATGT	GGTGAAGTAAACCGTCGTGAAC
<i>BTS1</i>	ATAGGGGACAAGGCTTGGAT	GAGACAGCGCTTCCATGAGT
<i>ERG20</i>	ACACTCCAGGCGGTAAGCTA	CCTGCAACAAGTCAATGCAC
<i>ERG19</i>	CATGCCACATGTTTGGACTC	CAACACAGCATTTGGACCTG
<i>ERG8</i>	TCGAACCTCCCGTACAAACT	TAGCGGTTTGAGCCCTAAGA
<i>ERG12</i>	TGGACTGCTTGTCTCAATCG	AGCACCGGTAAGTTTTGTGG
<i>HMG1</i> ^(a)	AGACCCGTTTGACGTCCTTA	GTAGAGGCGCTCAACCAAAA

(a) The HMG1 primer couple targets the 5' HMG-CoA reductase sequence, before the truncation giving birth to tHMG1 occurs. Consequently it targets only the endogenous copy of HMG-CoA reductase mRNA.

Supplementary Figures

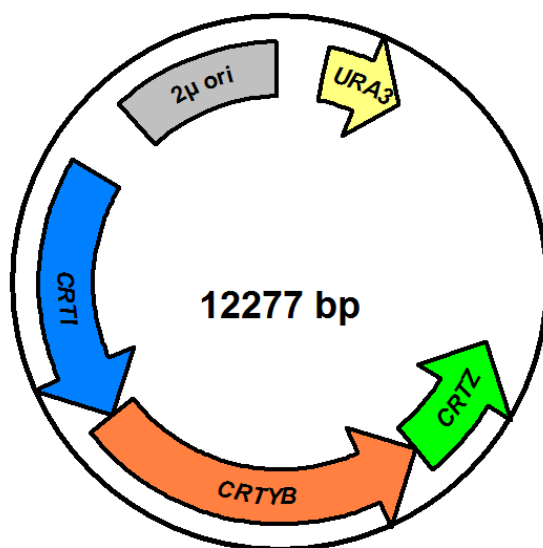


Supplementary Figure 1: Example of construct pRS426/A. Construction method: each element (promoters, genes, and terminators) was PCR amplified with overlapping ends in the primers shown here with arrows. OE PCR permitted to obtain the different expression cassettes. In a single step, three of them were then assembled by Gibson method (Gibson et al., 2009) in a vector pRS426. Strong promoter is shown in pink, medium promoters are in orange and weak one is in green. The example of construct A is shown here.

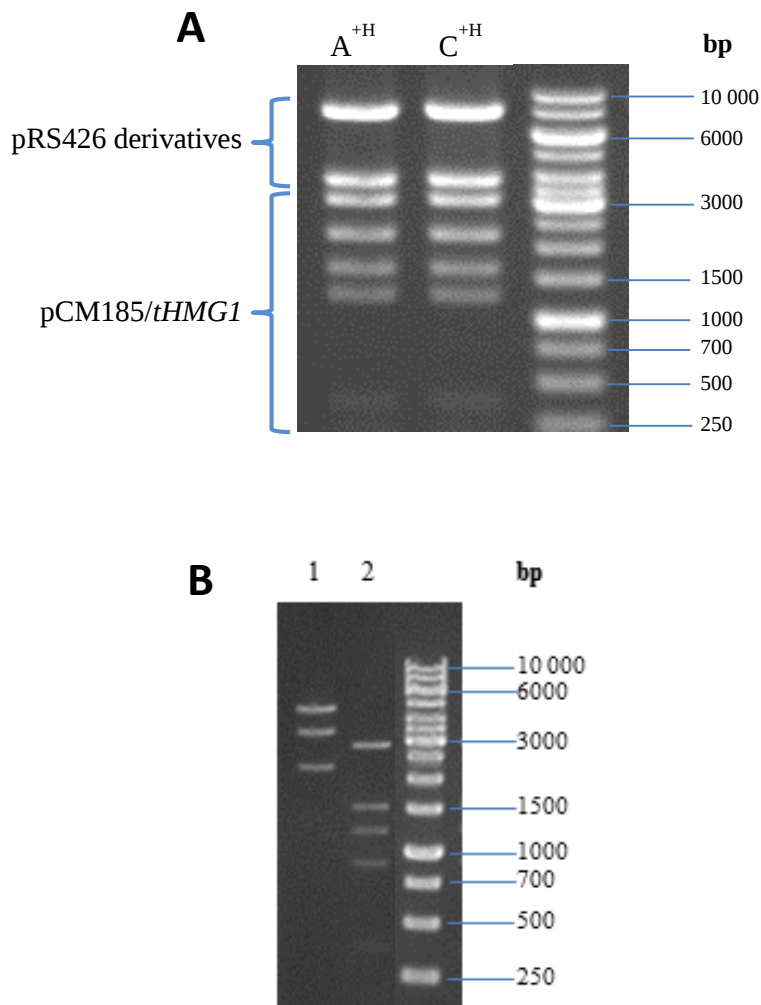


Supplementary Figure 2: Growth parameters of the zeaxanthin producing strains. Few strains are taken here for example, but every strain behaved the same.

Growth curves followed by optical density at 600 nm, with either the zeaxanthin producing strain C⁺H (red line, squares) or with the reference strain H (green line, triangles).



Supplementary Figure 3: pRS426 derivative map. Example of construct A is shown here. 2μ ori: replication origin 2μ. URA3: selection marker.



Supplementary Figure 4: Restriction profiles of plasmids recovered from zeaxanthin producing strains and *BglII* digested. Migration on 0.8% agarose gel.

(A) Plasmids recovered from strains A⁺H and C⁺H are shown here as examples, but result was the same in every zeaxanthin producing strain. Expected bands for the pRS426 derivatives are 8245 bp and 4032 bp. Expected bands for plasmid pCM185/*tHMG1* are 3286 bp; 2301 bp; 1646 bp; 1305 bp; 371 bp; 122 bp and 71 bp. (B) Plasmids recovered from strain J. 1 = pRS426/J ; 2 = pCM185. Expected bands for pRS426/J are : 5701 bp ; 4032 bp et 2544 bp. Expected bands for pCM185 are : 3286 bp ; 1646 bp ; 1305 bp ; 929 bp ; 371 bp ; 122 bp et 71 bp.

DNA bands were BET stained and revealed under UV light. We cannot see the smallest bands in this gel because of insufficient amounts of DNA, but visible bands correspond to expected profiles.

3. Reference

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343-345. doi: 10.1038/nmeth.1318.

III. RESULTATS COMPLEMENTAIRES

1. Construction de la banque combinatoire.

Nous avons effectué des transformations de levure individuelles afin d'obtenir les souches A^(+H) à G^(+H) ainsi que I^(+H) et J. La transformation des cellules avec les dérivés de pRS426 a été généralement peu efficace : entre 10 et 20 transformants pour les souches A à F, 5 transformants pour obtenir la souche G, et seulement un clone obtenu pour la souche I^(+H) ainsi que la souche J. Environ 2/3 de ces transformants développaient au bout de quelques jours de croissance sur boîte une coloration jaune, alors que les autres conservaient leur couleur blanchâtre. Après perte du plasmide multi-copies (dérivés de pRS426), les clones jaunes perdaient leur couleur, alors que les blancs restaient blancs. Le développement de la coloration dans certains clones est donc lié à la présence du plasmide multi-copies. De façon à confirmer que cette couleur jaune correspondait bien à une production de zéaxanthine, nous avons extrait les caroténoïdes de clones blancs et jaunes (Figure 21).

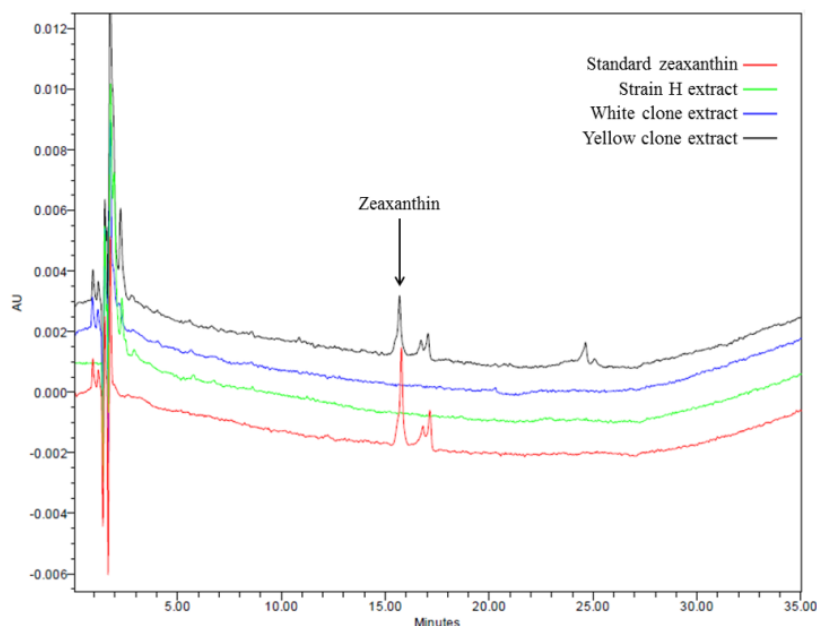


Figure 21: Spectres HPLC (450 nm) résultants de l'injection des extraits de caroténoïdes d'un clone blanc (trait bleu) et d'un clone jaune (trait noir). Trait rouge : standard de zéaxanthine. Trait vert : extrait d'une souche contrôle contenant pRS426 et pCM185 vides.

Seuls les clones jaunes produisaient de la zéaxanthine, alors que le profil de l'extrait issu d'un clone blanc était parfaitement identique à celui de la souche témoin transformée avec le pRS426 vide. La coloration développée dans les clones jaunes était donc bien liée à la production de zéaxanthine. Nous avons ensuite analysé les plasmides extraits des levures (Figure 22).

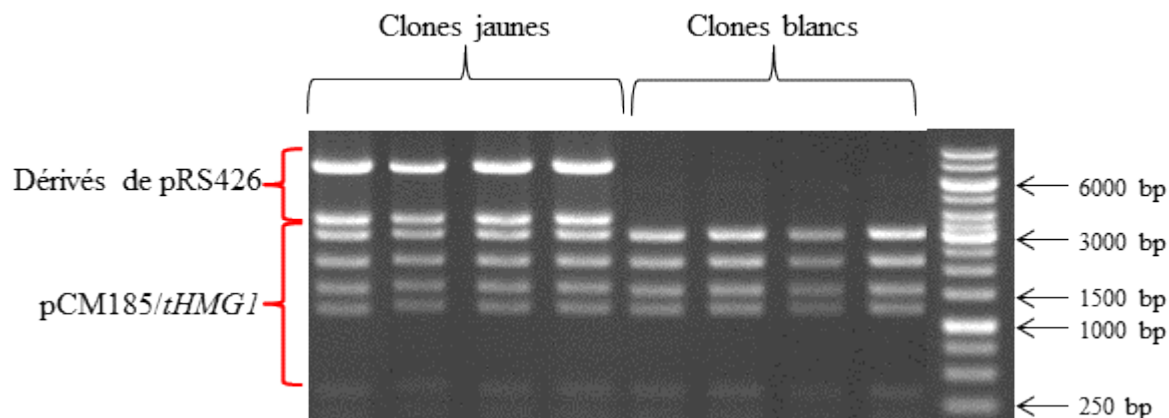


Figure 22: Migration sur gel d'agarose 0.8% (w/v) des extraits plasmidiques de 4 clones de levure jaunes et 4 clones blancs (contenant les dérivés de pRS426 et pCM185/tHMG1) après digestion par *Bgl*I. Les tailles de fragments attendues pour pCM185/tHMG1 sont : 3286 bp; 2301 bp; 1646 bp; 1305 bp; 371 bp; 122 bp et 71 bp. Les tailles de fragments attendues pour les dérivées de pRS426 sont : 8245 bp et 4032 bp. Révélation sous lumière UV après coloration au BET. Les bandes en dessous de 250 bp ne sont pas visibles sur ce gel pour une question de quantité d'ADN déposée, mais les bandes visibles correspondent aux profils attendus.

Le plasmide monocopie (pCM185/tHMG1) était présent dans tous les clones testés, mais les dérivés de pRS426 n'étaient retrouvés que dans les clones jaunes, avec le profil de restriction attendu. Ce résultat suggère une perte des dérivés de pRS426 dans les clones blancs. Cependant, les clones blancs sont toujours capables de croître sur un milieu sans uracile. La seule hypothèse compatible avec une perte de plasmide serait un événement d'intégration au locus *ura3-52* de CEN.PK, rendant la souche prototrophe pour l'uracile. Nous avons testé cette hypothèse en réalisant un test-cross, c'est-à-dire en croisant un clone blanc avec une souche de génotype *ura*-(SK1 : Δ flo8, Δ his4, Δ ura3, -leu, -trp, GAL2-). Après croisement, les diploïdes obtenus ont été mis à sporuler. Les spores ont été disséquées sur milieu non sélectif (YPD). Les colonies ainsi obtenues

sur milieu non sélectif ont fait l'objet de répliques sur milieu sélectif pour l'uracile afin de tester leur génotype URA (Figure 23). Aucune des spores n'a présenté de croissance.

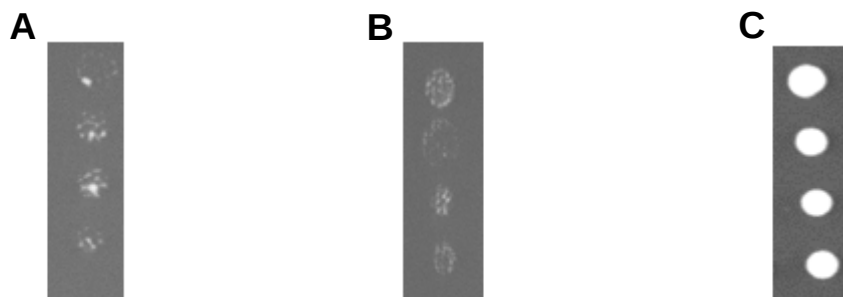


Figure 23: Spores issues du croisement d'un clone blanc (A) ou jaune (B) avec une souche *Δura3*, sur milieu dépourvu d'uracile. (C) Souche contrôle présentant une croissance correcte, sur milieu riche. Ce témoin, ayant eu le même temps de croissance que les tests cross, permet de voir l'absence de croissance en (A) et (B).

Nous n'observons pas la ségrégation 2/2 ([URA⁺], [URA⁻]) attendue si un évènement d'intégration avait eu lieu au locus *ura3-52* de CEN.PK. Cela semble indiquer que les cellules perdent rapidement les plasmides (pRS426) sur le milieu non sélectif sur lequel les spores ont débuté leur croissance après dissection, ce qui signifie que même les clones blancs possèdent un plasmide portant le gène *URA3*.

La dichotomie (deux populations présentant des phénotypes colorés différent) observée peut être expliquée par des réarrangements qui ont probablement eu lieu dans les clones blancs, induisant une perte d'au moins une partie de l'origine de réplication de *E. coli* (expliquant l'incapacité à retrouver le plasmide), et probablement une partie de la voie de biosynthèse de la zéaxanthine (expliquant l'absence de production).

2. Production de caroténoïdes.

Nous avons quantifié la production de zéaxanthine et de β -carotène dans toutes les souches que nous avons construite dans notre banque combinatoire. Les autres intermédiaires de la voie

métabolique (phytoène et β -cryptoxanthine) n'ont pas pu être détectés dans nos extraits. La souche J ne produit aucun caroténoïde (Figure 24, Figure 25), ce qui peut être relié à l'absence d'expression des gènes *CRTI* et *CRTYB* due à leur orientation convergente. La souche J mise à part, deux profils de production sont principalement mis en évidence dans la Figure 25. Les cellules qui expriment *tHMG1* produisent généralement plus de β -carotène que de zéaxanthine (points rouges sur la Figure 25), et celles qui n'expriment pas *tHMG1* produisent généralement plus de zéaxanthine que de β -carotène (points bleus sur la Figure 25). Seules deux souches constituent des exceptions à cette règle : les souches E et G. La production de zéaxanthine de ces dernières augmente lorsqu'on exprime *tHMG1*, alors que le β -carotène reste au même niveau. Globalement, dans nos conditions, l'expression de *tHMG1* semble être une stratégie efficace d'augmentation de la production de β -carotène, mais pas de zéaxanthine.

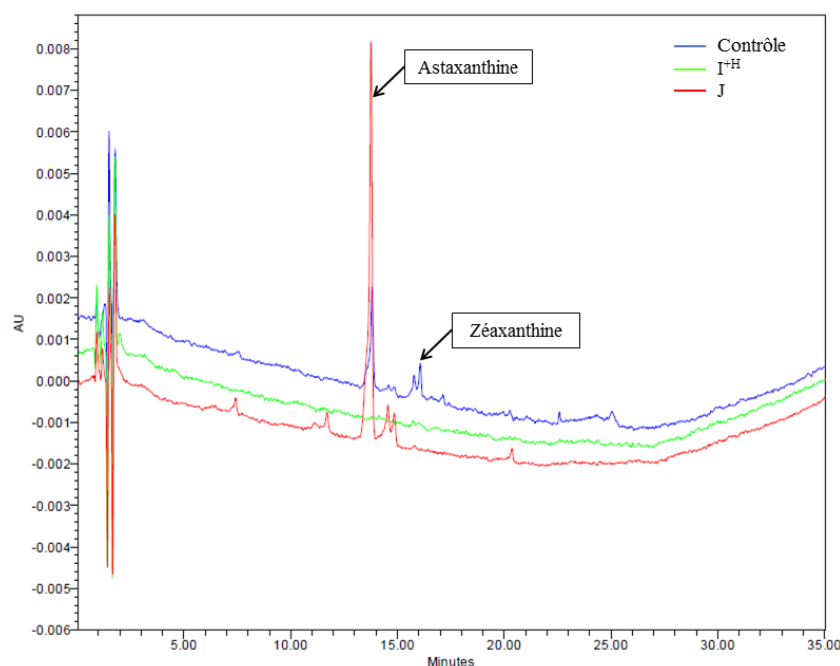


Figure 24: Analyses HPLC (450 nm) résultant de l'injection des extraits de caroténoïdes des souches I^{+H} (trait vert), J (trait rouge) et de la souche contrôle contenant pRS426 et pCM185 vides. L'astaxanthine est un standard d'extraction utilisé pour normaliser la quantification des produits par rapport à la quantité de cellules extraites (voir article).

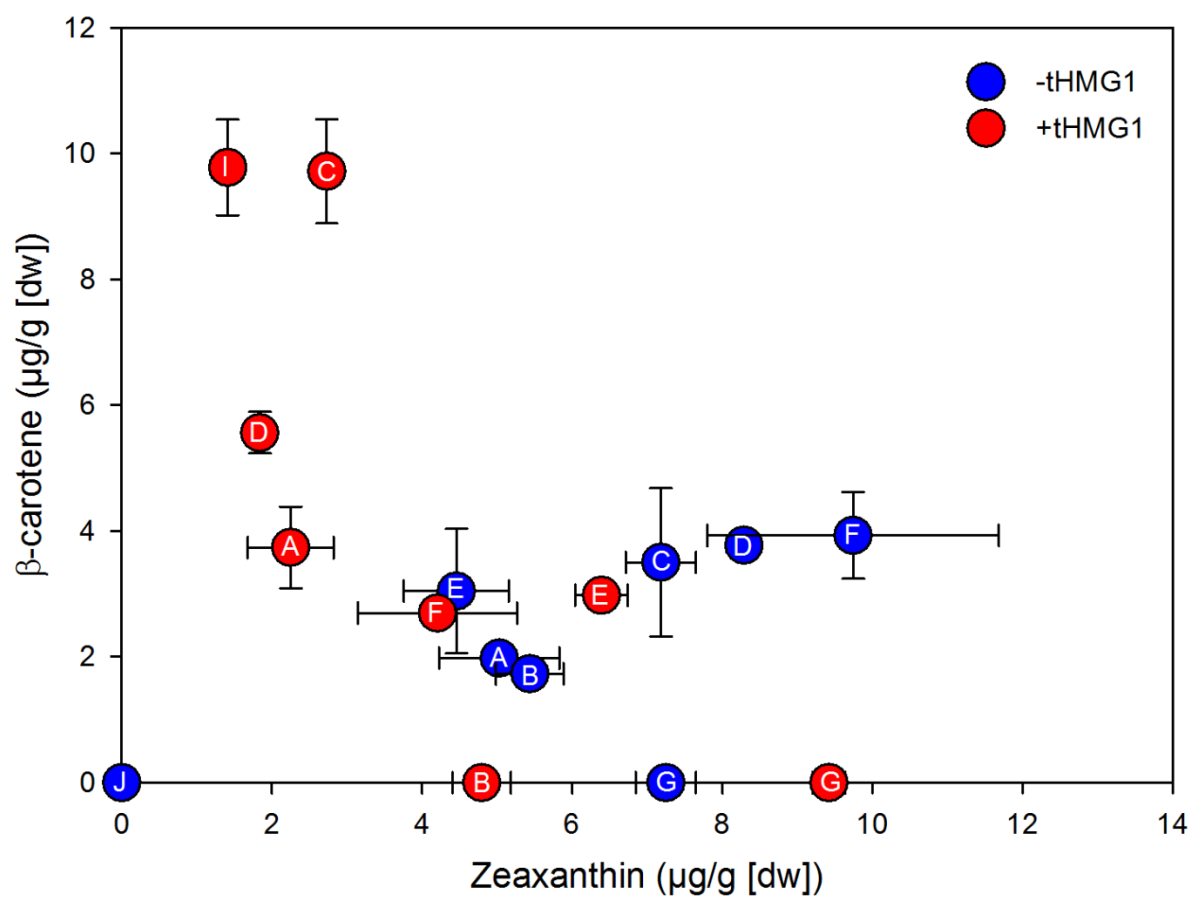


Figure 25: Dosages de zéaxanthine et de β -carotène produits par les souches A à J. Les points bleus représentent les souches n'exprimant pas *tHMG1* ; les points rouges représentent les souches qui expriment *tHMG1*. Les barres d'erreurs représentent un écart type sur trois valeurs.

IV. DISCUSSION

Nous avons reproduit une méthode d'optimisation des niveaux d'expression des gènes classiquement utilisée en ingénierie métabolique. Une banque combinatoire basée sur une voie métabolique à trois gènes (voie de biosynthèse de la zéaxanthine) a été créée sur un plasmide multi-copies, en faisant varier la nature des promoteurs. Les promoteurs constitutifs de *S. cerevisiae* choisis ont été précédemment caractérisés à l'aide d'un système modèle : l'expression de la GFP (Sun et al., 2012). Par des mesures des niveaux d'ARNm correspondant aux trois gènes, nous avons pu étudier les biais existant dans cette méthode d'optimisation, et mettre en évidence des facteurs impactant fortement la régulation imposée par les promoteurs lors du design de départ. L'interférence transcriptionnelle et la nature des ORFs sont déterminants dans les niveaux d'ARNm que nous avons observé.

Ces travaux ont volontairement été effectués sur un plasmide multi-copies dans le but de rester fidèle à ce qui est fait en recherche pour optimiser une voie de biosynthèse. Même si nous avons vérifié grâce à *URA3* que les niveaux de transcrits dans les différentes constructions n'était pas significativement différents, nous savons aussi d'après la littérature qu'un gène ne s'exprime pas de manière identique s'il est sur plasmide ou intégré au génome : Verwaal et al. (2007) ont observé de meilleurs niveaux de production de β -carotène lorsque les gènes hétérologues étaient intégrés au génome que lorsqu'ils étaient exprimés sur plasmide multi-copies. Les influences respectives de l'interférence transcriptionnelle en condition d'intégration et sur plasmide seraient intéressantes à comparer.

Les différentes voies métaboliques souhaitées (de A à J) ont été obtenues en fusionnant par PCR des éléments de base (promoteurs, ORFs, terminateurs) pour obtenir différentes unités transcriptionnelles, puis ces unités individuelles ont été clonées simultanément trois par trois à l'aide d'une technique de recombinaison *in vitro* (Gibson et al., 2009). Cette méthode nous a permis d'obtenir l'ensemble des voies métaboliques que nous souhaitions tester de manière fiable et dans un temps court (après optimisation de la méthode, une semaine suffisait pour obtenir les plasmides définitifs). En revanche, nous avons rencontré des difficultés pour transformer les levures avec nos constructions. Ces difficultés peuvent être mises en relation avec le faible nombre de copies des dérivés de pRS426 calculé dans les souches productrices de zéaxanthine (Figure 3 de l'article) et

le faible niveau de transcription observé sur le gène *URA3* (Figure 4 de l'article). En effet, un faible nombre de copies de plasmide, associé une transcription faible du marqueur de sélection peut logiquement amener à des transformants dont l'auxotrophie n'est pas efficacement compensée par l'expression du gène marqueur. Dans ce cas, l'avantage majeur des stratégies combinatoires qu'est la rapidité de construction d'une banque est atténué par la nécessité de réaliser de nombreuses transformations de levures pour finalement obtenir la souche désirée.

Outre les difficultés de transformation, deux phénotypes colorés apparaissaient parmi les clones obtenus, et nous avons montré que seuls les clones développant une coloration jaune contenaient les dérivés de pRS426 dans leur intégralité. L'apparition de deux phénotypes colorés chez les souches transformées a déjà été remarquée lors de précédents travaux visant à produire des caroténoïdes chez *S. cerevisiae*. Liang et al. (2013) ont comptabilisé, selon la souche utilisée, 20% ou 35,5% de clones blancs après transformation, le reste développant la couleur jaune caractéristique de l'accumulation de caroténoïdes. Suite à un test de stabilité de leur plasmide, ils ont cependant conclu que les clones jaunes ne perdaient pas leur couleur après environ 63 générations. Beekwilder et al. (2014) ont observé seulement 50% de clones devenant oranges après quelques jours de croissance. Cette coloration s'est également avérée stable après 24 heures de culture suivies d'un nouvel étalement sur boîtes. Dans nos données, le nombre de pRS426/F mesuré dans la souche F^{+H} (inférieur à 1) révèle une certaine instabilité du plasmide (Figure 3 de l'article). Cela signifie qu'au sein de la population totale de cellules, certaines ont perdu le plasmide. On peut alors penser qu'elles survivent en milieu sélectif soit grâce à un stock d'uracile hérité de la cellule mère, soit par « cross-feeding », c'est-à-dire en se nourrissant d'uracile synthétisé par les autres cellules autour d'elles. Cela dit, aucune étude publiée à ce jour ne peut appuyer ces hypothèses.

Les difficultés à transformer, la présence de clones blancs, et une instabilité du plasmide constituent un faisceau d'informations suggérant une forte toxicité de la voie de biosynthèse de la zéaxanthine. Malgré cela, nous sommes parvenus à obtenir des souches productrices de zéaxanthine, à des niveaux différents. Les niveaux de production varient d'une souche à l'autre (de $1,41 \pm 0,21 \mu\text{g/g[dw]}$ à $9,83 \pm 0,47 \mu\text{g/g[dw]}$ de zéaxanthine produite) bien que les niveaux d'ARNm des gènes *CRTI*, *CRTYB* et *CRTZ* soient très similaires entre elles. Une régulation traductionnelle intervient donc vraisemblablement, induisant une relation non linéaire entre la quantité de transcrits et de métabolites. Les protéines CrtI, CrtYB et CrtZ n'ont pas été quantifiées dans cette étude mais

leur régulation traductionnelle et post-traductionnelle semblent jouer un rôle important dans la quantité de métabolite d'intérêt obtenue. Malgré l'absence de données de protéomique, nos données de quantifications d'ARNm nous permettent tout de même de dévoiler une condition qui semble plus adaptée à la production de zéaxanthine. En effet, la souche G^{+H} présente des profils d'ARNm et de métabolites différents des autres souches. Elle est la seule à présenter trois niveaux d'ARNm presque équivalents en condition respiratoire (Figure 6D de l'article), elle ne produit pas de β -carotène (Figure 25), et la stratégie d'expression de *tHMG1* est efficace pour augmenter sa production de zéaxanthine. La synthèse de la zéaxanthine à partir du β -carotène est catalysée par CrtZ. Bien qu'à notre connaissance il n'y ait pas de données biochimiques disponibles au sujet de CrtZ, nous pouvons émettre l'hypothèse qu'il s'agisse d'une étape limitante de la voie de biosynthèse. Un flux de précurseur augmenté par l'expression de *tHMG1* est potentiellement consommé par CrtI et CrtYB, mais il est possible que les niveaux trop bas de *CRTZ* dans les souches $A^{(+H)}$ à $F^{(+H)}$ ne permettent pas d'absorber ce flux. Cela résulterait en une accumulation du substrat de CrtZ, le β -carotène, dans les souches $A^{(+H)}$ à $F^{(+H)}$. Cependant, comme les niveaux d'ARNm de *CRTI* et *CRTYB* sont diminués dans la souche G^{+H} , mais pas celui de *CRTZ* (Figure 6D de l'article), nous pouvons imaginer que les quantités plus faibles de β -carotène synthétisées puissent être entièrement consommées par CrtZ. Ainsi, le flux métabolique entrant dans la voie de biosynthèse hétérologue pourrait être entièrement transformé en zéaxanthine, sans accumulation de β -carotène. Si cette hypothèse était juste, cela signifierait que la voie de biosynthèse de la zéaxanthine chez *S. cerevisiae* requière un niveau d'expression équivalent de chacun des trois gènes impliqués.

Au regard des résultats obtenus, il apparaît essentiel d'éloigner les gènes les uns des autres sur le plasmide, afin d'éviter au maximum l'interférence transcriptionnelle. En effet, nous savons maintenant que des facteurs tels que le décrochage de l'ARN polymérase au niveau du terminateur, ou encore les contraintes topologiques de l'ADN sont déterminants dans la régulation transcriptionnelle sur un plasmide multi-copies. L'effet des contraintes topologiques sur la transcription par l'ARN polymérase II a été étudié chez des souches de *S. cerevisiae* mutées dans les topoisomérases (Garcia-Rubio & Aguilera, 2012). Les topoisomérases I et II sont les enzymes qui contrôlent le niveau d'enroulement de l'ADN. Leur action est requise pour désenrouler la molécule d'ADN afin d'autoriser les processus de répllication et de transcription. Les auteurs ont en effet observé que la transcription était grandement impactée par l'absence de topoisomérases,

ce qui signifie que l'ARN polymérase II est incapable de transcrire un gène efficacement si le niveau d'enroulement de l'ADN est trop élevé. De plus, la toxicité de la zéaxanthine est à présent fortement soupçonnée. Dans ce contexte, une stratégie d'expression contrôlée, comme celle qui a été mise en œuvre par Xie et al. (2014) (expression sous promoteurs *GAL* contrôlée par le ratio entre glucose et glycérol dans le milieu de culture) permettrait probablement d'augmenter la production. Ces perspectives d'évolution de la stratégie de production attestent de l'utilité de la réalisation de banques combinatoire. En effet, sans mettre en place une stratégie combinatoire, nous n'aurions pas pu voir de façon aussi claire et ciblée ces pistes d'amélioration.

CHAPITRE 2

Etude de l'activité des Carotenoid Cleavage Dioxygenases connues chez *Crocus sativus*.

I. DONNEES INTRODUCTIVES.

Suite aux travaux exposés dans le Chapitre 1, nous disposons de souches de *S. cerevisiae* productrices de zéaxanthine. Ce second chapitre concerne la fin de la voie de biosynthèse ciblée par le projet : celle qui mène à la synthèse des colorant (crocine), parfum (safranal) et arôme (picrocrocine) du safran (*Crocus sativus*) à partir de la zéaxanthine (Figure 26). Cette voie métabolique est initiée par le clivage de la zéaxanthine en positions 7,8 et 7',8' par une dioxygénase (Bouvier et al., 2003b; Pfander & Schurtenberger, 1982). Nous allons débiter ce chapitre par une description de la littérature concernant le mécanisme de clivage des caroténoïdes par les CCOs (Carotenoid Cleavage Oxygenases), et plus particulièrement de celui qui existe chez la plante *Crocus sativus*.

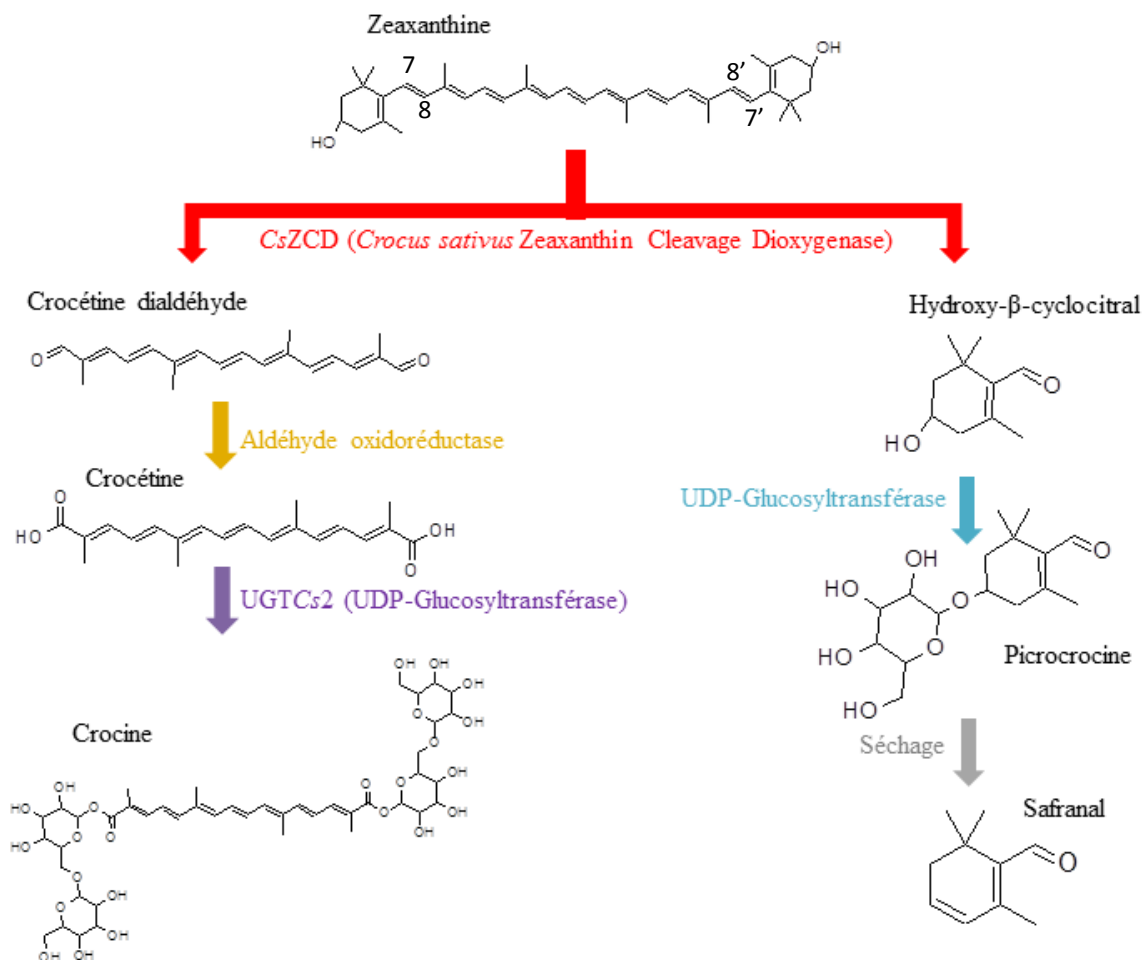


Figure 26: Voie de biosynthèse de la crocine, la picrocrocine et du safranal chez *Crocus sativus* à partir de la zéaxanthine.

1. Données structurales et mécanisme de clivage des caroténoïdes par les Carotenoid Cleavage Oxygenases (CCOs).

Des CCOs ont été isolées chez des cyanobactéries, des algues, des plantes supérieures ou encore des mammifères. Le Tableau 4 donne quelques exemples des CCOs connues à ce jour. Une revue plus exhaustive est disponible dans (Kloer & Schulz 2006).

Tableau 4: Exemples de CCOs dans quelques organismes et leurs produits. ABA = acide abscissique.

Nom de l'enzyme	Organisme source	substrats	produit
VP14	<i>Zea mays</i>	9-cis vialxanthine, - néoxanthine, - zéaxanthine, - antheraxanthine, luthéoxanthine 8'R, luthéoxanthine 8'S	ABA
BCOI	<i>Drosophila melanogaster</i>	β -carotène	Rétinal
	<i>Mus musculus</i>		
	<i>Homo sapiens</i>		
CsZCD	<i>Crocus sativus</i>	zéaxanthine	Safranal Crocétine
BoLCD	<i>Bixa orellana</i>	lycopène	Bixine

Les CCOs clivent leurs substrats au niveau des doubles liaisons C-C. Elles présentent une forte spécificité envers la position de la double liaison qu'elles clivent, mais peuvent souvent prendre pour substrats plusieurs caroténoïdes différents (Auldridge et al., 2006). On distingue plusieurs familles de CCOs : les 15,15' mono-oxygénases présentes dans le règne animal pour générer du rétinol ; les CCD1 et CCD7 chez les plantes coupent la double liaison entre les carbones en position 9 et 10 de leurs substrats ; et les NCEDs (9-cis-epoxycarotenoid dioxygénases) qui coupent entre les positions 11 et 12. D'autres spécificités, plus rares, ciblent d'autres doubles liaisons. Il a été montré que BoLCD de *Bixa orellana* est spécifique de la double liaison 5,6 de la zéaxanthine (Bouvier et al. 2003a), et CsZCD de *Crocus sativus* coupe le même substrat en position 7,8 (Bouvier et al. 2003b).

Ces enzymes nécessitent un ion Fe^{2+} (fer ferreux) non hémique, coordonné par quatre résidus histidine très conservés (Sui et al., 2013). Des études de mutagenèse dirigée sur les acides aminés responsables de la coordination de l'ion Fer indiquent que cet ion est absolument nécessaire à l'activité des CCOs (Kloer & Schulz, 2006). Le rôle du Fe^{2+} est d'activer l'oxygène moléculaire O_2 pour le clivage du substrat (Sui et al., 2013).

Depuis 2005, les structures tridimensionnelles de trois membres de cette famille ont été résolues : une apocaroténoïde oxygénase (ACO) de *Synechocystis* (Kloer et al., 2005), RPE65 de *Bos taurus* (Kiser et al., 2009), et VP14 de *Zea mays* (Messing et al., 2010). Leur analyse a révélé un repliement conservé en β -propeller à 7 lames renfermant le site actif (Figure 27), qui confère aux CCOs une structure rigide en tonneau pouvant être considérée comme une signature structurale de cette famille.

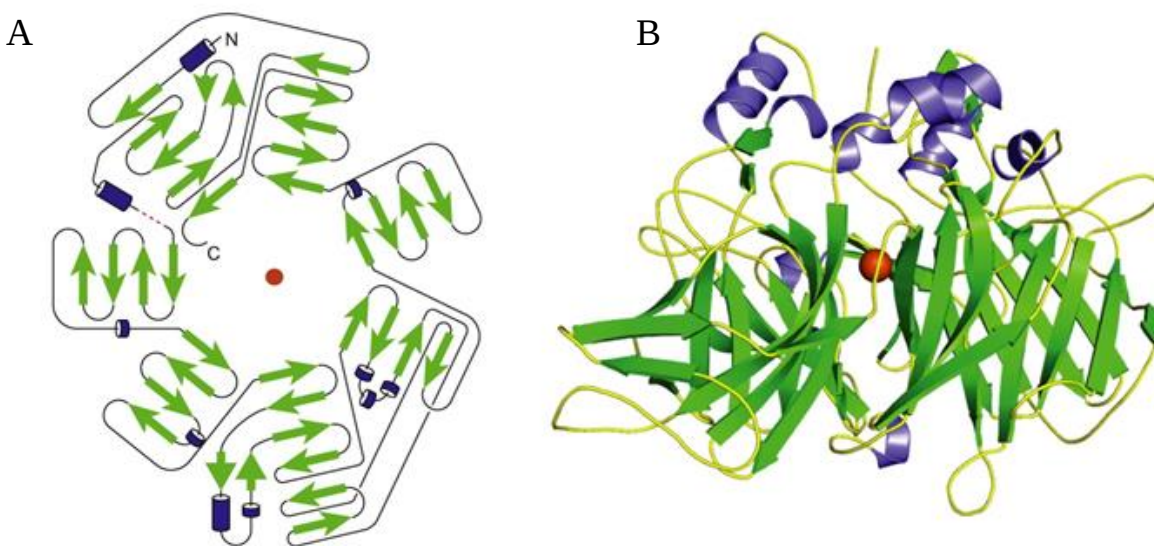


Figure 27: Topologie (A) et structure tridimensionnelle (B) de la CCO bovine RPE65 (code PDB 3FSN). L'ion catalytique Fe^{2+} est coloré en orange. Les hélices α sont en bleu et les feuillets β sont en vert. La ligne pointillée représente une zone absente du modèle structural. Figure adaptée de Sui et al. 2013.

Leurs substrats, les caroténoïdes, se caractérisent par leur liposolubilité. C'est pourquoi on les retrouve dans les environnements où ils sont à l'abri du contact avec l'eau, les membranes étant leur zone de stockage préférentielle. Les CCOs doivent donc interagir avec ces membranes pour y trouver leurs substrats. L'analyse de surface des trois CCOs cristallisées a mis en évidence des

patches hydrophobes à leur surface, leur permettant probablement une interaction directe avec les lipides membranaires et les caroténoïdes. Ensuite, le substrat est dirigé vers un tunnel hydrophobe jusqu'au site actif de l'enzyme (Figure 28). Ces éléments structuraux sont inhérents à la nature des substrats, et ils sont bien conservés au sein de cette famille d'enzyme (Sui et al., 2013).

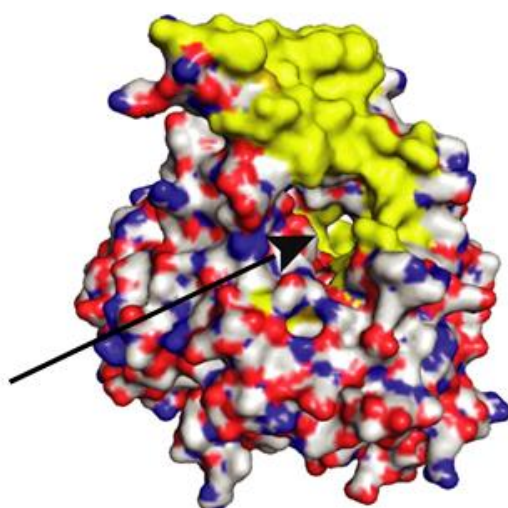


Figure 28: Vue de surface de la structure tridimensionnelle de VP14 (*Zea mays*). Code PDB 3NPE. Le patch hydrophobe est coloré en jaune. La flèche noire indique l'ouverture du tunnel menant le substrat au site actif. Figure adaptée de Sui et al. 2013.

Malgré l'obtention de données structurales, les détails de leur mécanisme d'action restent à clarifier. Les données biochimiques actuelles ne permettent pas de déterminer avec certitude s'il s'agit de mono-oxygénases ou de dioxygénases. Dans le cas de monooxygénases, un des intermédiaires catalytiques serait un époxyde, ne faisant participer qu'un atome d'oxygène sur les deux impliqués (Figure 29). Dans le cas de dioxygénases, les deux atomes d'oxygène participeraient à la réaction en attaquant la double liaison avec formation d'un intermédiaire dioxétane (Figure 29). Des données en faveur de l'hypothèse que les CCOs fonctionnent comme des dioxygénases ont été obtenues par incorporation d'isotopes radioactifs de l'oxygène (Schmidt et al., 2006).

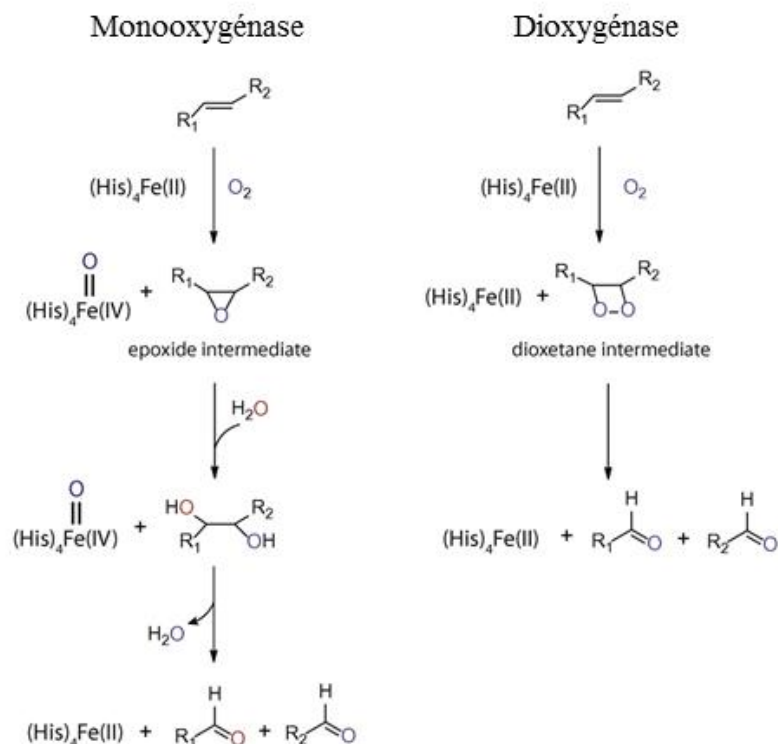


Figure 29: Mécanismes catalytiques des monooxygénases et des dioxygénases. Figure adaptée de Sui et al. 2013.

Malgré cela, de plus amples informations restent encore à obtenir pour définitivement conclure à propos du mécanisme catalytique des CCOs. Dans la suite de ce document, un grand nombre de CCOs seront également nommées CCDs, pour Carotenoid Cleavage Dioxygenase. Ces noms leur ont été attribués par les auteurs qui les ont décrites, compte tenu de la nomenclature adoptée dès la caractérisation des premières enzymes de ce type.

2. Le cas de *Crocus sativus*.

Crocus sativus est une plante vivace de la famille des iridacées qui fleurit de Septembre à Octobre (Figure 30).



Figure 30: Fleur de *Crocus sativus*. Les stigmates rouges donnent naissance, après séchage, au safran.

Les stigmates de la fleur (au nombre de trois) sont récoltés manuellement puis séchés pour devenir l'épice la plus chère du monde, surnommée l'Or Rouge, le safran. *Crocus sativus* est une plante adaptée aux terres semi-arides à arides, ce qui explique que la production mondiale provienne principalement des pays bordant la Méditerranée. Le marché mondial est dominé par l'Iran, l'Inde, la Grèce, le Maroc, l'Azerbaïdjan, l'Afghanistan, et l'Espagne, principaux producteurs dans cet ordre décroissant. L'Iran possède 90% du marché mondial. On trouve cependant de petites exploitations de *Crocus sativus* en France, principalement dans les régions du Quercy et du Gâtinais. La production mondiale avoisine les 200 tonnes par an, avec un prix moyen de 30 000 euros par kilogramme (correspondant à la récolte de 75 000 fleurs), avec des variations selon sa provenance et sa qualité, répondant à des critères décrits par la norme ISO 3632. La qualité du safran dépend de sa concentration en trois métabolites secondaires: la crocine, responsable de sa couleur, la picrocrocine qui lui confère son goût légèrement amer, et le safranal, responsable de son parfum (Lage & Cantrell, 2009). Au-delà de son attractivité pour ses qualités gustatives, on a découvert au safran de nombreuses propriétés bioactives. Il est depuis très longtemps utilisé dans la médecine traditionnelle chinoise comme analgésique ou tranquillisant. Depuis quelques années, on lui découvre des facultés anticancéreuses (Abdullaev, 2002; Bajbouj et al., 2012; Bathaie et al., 2007), de traitement de la maladie d'Alzheimer (Akhondzadeh et al., 2010), de protection de la rétine (Fernandez-Sanchez et al., 2012), de traitement de l'arthrose *via* ses propriétés anti-inflammatoires (Ding et al., 2013), ou encore de neuroprotection (Linardaki et al., 2013). Un

éventail aussi large d'utilisations potentielles fait des molécules à forte valeur ajoutée contenues dans le safran des cibles de choix pour les biotechnologies.

Pfander & Schurtenberger ont proposé en 1982 que la zéaxanthine soit le précurseur des métabolites secondaires responsables de la couleur, de l'arôme et du parfum du pistil de *Crocus sativus*. A l'exception d'une lycopène- β -cyclase (CstLcyB2a) (Ahrazem et al., 2010), la voie de biosynthèse menant à la zéaxanthine n'est pas connue chez *Crocus sativus*. En revanche, la voie de biosynthèse des arômes et parfums a été étudiée entre autres par l'équipe du Professeur Bilal Camara à l'Institut de Biologie Moléculaire des Plantes de Strasbourg. Ces travaux ont fait l'objet d'une publication en 2003 (Bouvier et al., 2003b). En se basant sur la séquence peptidique de VP14 (*Zea mays*) ainsi que sur des séquences de *Synechocystis* et du piment, l'équipe a pu dessiner un couple d'oligonucléotides ciblant les CCOs et ainsi sélectionner toutes les séquences correspondantes dans une banque d'ADNc de *Crocus sativus*. Deux enzymes ont été mises en évidence, baptisées CsZCD (pour *Crocus sativus* Zeaxanthin Cleavage Dioxygenase) et CsCCD (pour *Crocus sativus* Carotenoid Cleavage Dioxygenase). CsCCD est membre de la famille des 9,10(9',10')-Carotenoid Cleavage Dioxygenases, et l'étude a montré qu'elle s'exprime constitutivement dans tous les tissus de la plante. En revanche, CsZCD est, d'après cette étude, l'enzyme chromoplastique qui catalyse spécifiquement le double clivage de la zéaxanthine en positions 7,8 et 7',8' (Figure 26). Il a par la suite été démontré que CsZCD était très fortement exprimée dans les stigmates au moment où le plus haut niveau de zéaxanthine et d'apocaroténoïdes y sont détectés (Castillo et al., 2005). Le clivage de la zéaxanthine par CsZCD donne naissance aux deux apocaroténoïdes précurseurs des molécules d'intérêt : une molécule de zéaxanthine permet d'obtenir une molécule de crocétine dialdéhyde (correspondant à la chaîne carbonée centrale) et deux molécules d'hydroxy- β -cyclocitral (correspondant aux cycles des extrémités). Des oxydations et des glucosylations font suite à ce clivage afin d'obtenir la crocine et la picrocrocine (Figure 26). Une UDP-glucosyltransférase, UGTcs2, a été isolée d'une banque d'ADNc du stigmate de *Crocus sativus* (Moraga et al., 2004). Ces travaux ont montré qu'elle est capable de doublement glucosyler la crocétine, suggérant que cette enzyme est responsable de la synthèse de la crocine finale. Le safranal naît de la déglycosylation de la picrocrocine suivie d'une déshydratation qui surviennent lors de l'étape de séchage après cueillette (Lozano et al., 2000).

Des travaux ultérieurs (Rubio et al., 2008) présentent des résultats différents de la littérature. L'équipe concernée a isolé 4 gènes de CCOs dans une banque d'ADNc de *Crocus sativus* : CsCCD1a, CsCCD1b, CsCCD4a et CsCCD4b. Après analyse de similarité de séquences, il s'avère que CsCCD1a est à 100% identique à CsCCD isolée par Bouvier et al. (2003b). CsCCD4a (580 acides aminés) et CsCCD4b (569 acides aminés), de localisation chromoplastique, sont des variants alléliques présentant respectivement 100% et 98% de similarité avec les 369 acides aminés de CsZCD. Les CsCCD1 (a et b) montrent une activité de clivage du β -carotène et de la zéaxanthine en positions 9,10 et 9',10'. La même activité a été observée avec CsCCD4 (a et b), sur le β -carotène uniquement. CsZCD (Bouvier et al., 2003b) n'a pas pu être isolée de leur banque d'ADNc, mais une forme tronquée de CsCCD4a (CsCCD4a-211) a été obtenue en utilisant les oligonucléotides publiés par Bouvier et al. (2003b). Les tests d'activité effectués sur cette protéine n'ont révélé aucun produit de clivage. Rubio et al. (2008) suggèrent qu'il manque à CsZCD, à l'instar de CsCCD4a-211, des résidus important pour l'activité de clivage.

II. RESULTATS.

Nous sommes parvenus dans le Chapitre 1 de cette étude à construire une souche de *S. cerevisiae* productrice de zéaxanthine, précurseur des colorants, parfums, et arômes du safran (Figure 20). Quatre étapes enzymatiques supplémentaires sont nécessaires pour parvenir à produire ces molécules. Cependant, les données de la littérature au sujet de la biosynthèse de la crocine, la picrocrocine et du safranal sont peu nombreuses. Seules deux études décrivent à ce jour des enzymes impliquées dans la voie métabolique menant à leur synthèse chez la plante *Crocus sativus* :

- Bouvier et al., 2003b : les auteurs ont mis en évidence l'enzyme CsZCD, qu'ils décrivent comme étant la CCO responsable du double clivage oxydatif de la zéaxanthine (en positions 7,8(7',8')) pour donner naissance à l'hydroxy- β -cyclocitral et à la crocétine dialdéhyde.
- Moraga et al., 2004 : les auteurs ont isolé l'enzyme UGTcs2, une UDP-glucosyltransférase responsable de la glucosylation de la crocétine pour obtenir la crocine.

Une enzyme capable de catalyser le clivage de la zéaxanthine en positions 7,8 et 7',8' doit être exprimée dans une souche productrice de zéaxanthine afin d'initier la synthèse de nos molécules d'intérêt. L'enzyme CsZCD a été mise en évidence d'une part (Bouvier et al., 2003b), mais les résultats de l'étude de Rubio et al. (2008) publiée en 2008 contestent l'existence de cette enzyme chez *Crocus sativus* et son implication dans le clivage de la zéaxanthine d'autre part.

Les analyses qui font l'objet de ce Chapitre 2 ont pour but initial de confirmer, ou d'infirmer, l'hypothèse selon laquelle CsZCD catalyse le clivage de la zéaxanthine en positions 7,8 et 7',8'. Les résultats présentés dans ce document sont constitués d'analyses bibliographiques et bioinformatiques, ainsi que de tests d'activité afin enrichir les données disponibles sur la fonctionnalité de cette enzyme encore mal connue.

1. Analyses bioinformatiques des données disponibles sur les CCOs de *Crocus sativus*.

1.1 Etude de similarités de séquences.

Dans un premier temps, afin de comparer CsZCD (Bouvier et al. 2003b) à l'ensemble des CCOs connues, nous avons fait une recherche de séquences homologues par BLAST, en nous basant sur la séquence protéique de CsZCD (numéro d'accèsion UniProtKB/Swiss-Prot Q84K96). Cinq séquences de protéines ayant des niveaux de similarité élevés ont ensuite été alignées grâce au programme COBALT de NCBI (Tableau 5 ; Figure 31).

Tableau 5: Liste des cinq séquences protéiques les plus proches de CsZCD.

Nom de l'enzyme	Organisme source	Numéro d'accèsion GenBank	Pourcentage d'identité	Référence
BoLCD	<i>Bixa orellana</i>	CAD33263	97%	Bouvier et al., 2003a
CsCCD4a	<i>Crocus sativus</i>	ACD62476	98%	Rubio et al., 2008
CsCCD4b	<i>Crocus sativus</i>	ACD62477	98%	Rubio et al., 2008
CsCCD4c	<i>Crocus sativus</i>	AEO50759	66%	Non publiée
CCD4b	<i>Vitis vinifera</i>	AGT63322	58%	Lashbrooke et al., 2013

CsZCD	-----	-----	
BoLCD	-----	-----	
CsCCD4a	1 MDYRLSSSSLFHF-PSPGNRI	FLKQSQVIAfQNPQSHQDHPITTKGKSISINKGGSISRNR---SLAAVFCDALDDL	72
CsCCD4b	1 MEYRLSSS-LFHF-PSPGNRI	FLK-----HYPSHQDHPITTKGKSISINKGGSISRNR---SLAAVFCDALDDL	63
CsCCD4c	1 MDAFSSSSFLSSS---FVPLL[35]NVKNDGH1QI-HLPFHQKATTDKSEKKKKKITSNAPSvwpSLPATFCNILDVL		107
CCD4b	1 MNPLFCP-FLSST1PHPKPLV[21]AIRNVED-KL-HSTIFYATPTTSQFPEIPTTITAKKRFV--pSLLVTIFNALDDF		92
CsZCD	-----	-----	
BoLCD	-----	-----	
CsCCD4a	73 ITRHsFDPDALHPSVDPHRVLRGNFAPVSELPPTPCRVVRGTIIPALAGGAYIRNGFNPpqYLPFGAHLFEGDGLHS		152
CsCCD4b	64 ITRHsFDPDALHPSVDPHRVLRGNFAPVSELPPTPCRVVRGTIIPALAGGAYIRNGFNPq---HLPSGAHLFEGDGLHS		141
CsCCD4c	108 INDF-VDPFVLRPSVDPRVLHNDNFSVVELPPTPCFVIRGSIPCLRGTYIRNGFNPQ---HLPRGPHHLFDGDGLHS		184
CCD4b	93 INNf-IGPP-LPPSIDPKHVLSGNFAPVDELPPTECEVIEGSLPPCLDG-AYIRNGFNPQ---FYPRGPHHLFDGDGLHS		167
CsZCD	1 -----	-----MQVDPITKGIGLANTISLQFS	19
BoLCD	1 -----	-----MQVEPTIRGIGLANTISLQFS	19
CsCCD4a	153 LLLPSSEGGRA-AIFSSRFVETKYLVITAKSRQAIFLSVFSGLCGFTGIARALVFFFRFLTMQVDPITKGIGLANTISLQFS		231
CsCCD4b	142 LLLPSSEGGRA-AIFSSRFVETKYLVITAKSRQAIFLSVFSGLCGFTGIARALVFFFRFLTMQVDPITKGIGLANTISLQFS		220
CsCCD4c	185 LLLPSGGGTApATLSSRFVQTYKYLVREAAAPVLPNVFSGFHGVGGIARGAVTAARVLIGQMNPAGVGLANTISLLFF		264
CCD4b	168 IRIS---HGR--PIFCSRYVKTYKYIIIEKRAGSPVPIPNLFSSYR---SFARSAVAIARLLTGQDFVNGVGLANTISVAFF		239
CsZCD	20 NGRHALCEYDLPYVVRSLPEDGDIStVGRIENNVS-TKSTIAHPKIDFVTGETFSFSYGPIQPYVTSRYDCDGKKS GP		98
BoLCD	20 NGRHALCEYDLPYVVRSLPEDGDIStVGRIENNVS-TKSTIAHPKIDFVTGETFSFSYGPIQPYVTSRYDCDGKKS GP		98
CsCCD4a	232 NGRHALCEYDLPYVVRSLPEDGDIStVGRIENNVS-TKSTIAHPKIDFVTGETFSFSYGPIQPYVTSRYDCDGKKS GP		310
CsCCD4b	221 NGRHALCEYDLPYVVRSLPEDGDIStVGRIENNVS-TKSTIAHPKIDFVTGETFSFSYGPIQPYVTSRYDCDGKKS GP		299
CsCCD4c	265 AGQLYALGESDLPYAVHL-DDNGDMSTLGRSDWGGKIFMGMTAHPKKDPITGEMFAFRYGFPVPIITYFRFDRNGNKLGP		343
CCD4b	240 CGHLYALAESDLPYAVRL-TPDGDIKTLGRYDFDGKLSMSMTAHPKIDPSTGEAFARYSFVRPFLLTYFRFDAQGKKQ-P		317
CsZCD	99 DVPIFSFKEPSFVHDFAITHEYAVFPDIQIVMKPAEIVRGRM-IGPDLEKVPRLGLLPRYATSDSEMRNFVDPGFIRMVH		177
BoLCD	99 DVPIFSFKEPSFVHDFAITHEYAVFPDIQIVMKPAEIVRGRM-IGPDLEKVPRLGLLPRYATSDSEMRNFVDPGFIRMVH		177
CsCCD4a	311 DVPIFSFKEPSFVHDFAITHEYAVFPDIQIVMKPAEIVRGRM-IGPDLEKVPRLGLLPRYATSDSEMRNFVDPGFIRMVH		389
CsCCD4b	300 DVPIFSFKEPSFVHDFAITHEYAVFPDIQIVMKPAEIVRGRM-IGPDLEKVPRLGLLPRYATSDSEMRNFVDPGFIRMVH		378
CsCCD4c	344 DVPIFSVTQPSFLHDFAVTEHYAIFPDIQIVMKPMQVGLGGSpVGSDRGKVPRLGLLPRNASSDSEMRNFVDPGFIRMH		423
CCD4b	318 DVPIFSLSCPSFFHDFAITNRYAIFPDIQIVMNPVQMIIRGGSpVGTDPITKVPVVGIIIPRYAKDESEMRNIDVPGFNIIH		397
CsZCD	178 VVNAWEEEGEVVVIVAPNVSPiENAI DRFDLLHVSVMARIELKSGSVSRILLSAENLDFGVHHRGYSGRKSRAYLGV		257
BoLCD	178 VVNAWEEEGEVVVIVAPNVSPiENAI DRFDLLHVSVMARIELKSGSVSRILLSAENLDFGVHHRGYSGRKSRAYLGV		257
CsCCD4a	390 VVNAWEEEGEVVVIVAPNVSPiENAI DRFDLLHVSVMARIELKSGSVSRILLSAENLDFGLIHRGYSGRKSRAYLGV		469
CsCCD4b	379 VVNAWEEEGEVVVIVAPNVSPiENAI DRFDLLHVSVMARIELKSGSVSRILLSAENLDFGLIHRGYSGRKSRAYLGV		458
CsCCD4c	424 SINAWEEEGGVVVVAPNVMSIEHALERTELHSCVEMVGDILRNGIVSRTPLSAGNLGFGVHHPGYLRQNRAYLGV		503
CCD4b	398 AINAWDEEDGDAIVMVA PNILPIEHALERMDLVHSGLEKVRIDLKTGTITRHRLSQWNLEFAVINPGYLKGKQIRYVYSAV		477
CsZCD	258 GDMPKIRGVVKVDFELAGRG---ECVVARREFGVGCFGGEFFVPASSKKSGGEEDDGYVVSYLHDEGKGESSFVMDA		334
BoLCD	258 GDMPKIRGVVKVDFELAGRG---ECVVARREFGVGCFGGEFFVPASSKKSGGEEDDGYVVSYLHDEGKGESSFVMDA		334
CsCCD4a	470 GDMPKIRGVVKVDFELAGRG---ECVVARREFGVGCFGGEFFVPAS-EGSGGEEDDGYVVSYLHDEGKGESSFVMDA		545
CsCCD4b	459 GDMPKIRGVVKVDFELAGRG---ECVVARREFGVGCFGGEFFVPAS-EGSGGEEDDGYVVSYLHDEGKGESSFVMDA		534
CsCCD4c	504 GDMPKISGVVKLDLNLVGTIcsDCVAKRDFGPRCFGGEFFVPADND--EEEDKGYVMSVHNESGESSFVMDA		581
CCD4b	478 GDPLPKISGIVKLDVSRSDRR--qECIVAKRMEPGCYGGEFFVAKEDPNPAAEEDDGYVLSVHDEQSGKSRFVMDA		555
CsZCD	335 RSPELEIAEVVLPFRVPYGFHGLFVTEAELLSQQ---	369	
BoLCD	335 RSAQLEIAEVVLPFRVPYGFHGLFVTDKDLNQA---	369	
CsCCD4a	546 RSPELEVVAEVVLPFRVPYGFHGLIVTEAELLSQQ---	580	
CsCCD4b	535 RSPELEVVAEVVLPFRVPYGFHGLIVTEAELLSQQ---	569	
CsCCD4c	582 QSPELEIAEVVLPFRVPYGFHGLFVTAELCSQRr1	619	
CCD4b	556 QSPDLDIVAAVKLPTRVPYGFHGLFVKGCDLMD----	589	

Figure 31: Alignement (COBALT) de la séquence protéique de CsZCD avec les cinq séquences protéiques les plus proches.

La similarité entre les séquences de CsZCD et BoLCD (97% d'identité) est très surprenante étant donné l'éloignement phylogénique et géographique entre les deux espèces concernées. En effet, *Bixa orellana* (le Roucou) est une plante d'Amérique tropicale non cultivée de la classe des Magnoliopsida, alors que *Crocus sativus*, cultivée principalement en Afrique du Nord, appartient à la classe des Liliopsida. CsZCD (ainsi que BoLCD) est plus courte que les autres dioxygénases dont elle est proche, d'environ 200 acides aminés. De façon à attribuer une fonction probable à ces acides aminés manquants, nous avons effectué plusieurs analyses. Tout d'abord, des séquences d'adressage au chloroplaste ont été recherchées. Ensuite, des analyses structurales ont permis de localiser les éléments de structure manquants chez CsZCD.

1.2 Recherche de séquences d'adressage au chloroplaste.

Les CCOs de plantes sont souvent localisées dans les plastes (Camara & Bouvier, 2004). Une localisation chromoplastique est d'ailleurs proposée dans le cas de CsZCD (Bouvier et al., 2003b). Les chromoplastes sont des organites observés dans les cellules des organes végétaux colorés de jaune à rouge. Il s'agit de structures cellulaires proches des chloroplastes, qui se différencient de ces derniers par la nature des pigments qu'ils contiennent. Ne contenant pas de chlorophylle, ils ne participent pas à la photosynthèse. Nous avons effectué une prédiction de la présence d'une séquence d'adressage aux chloroplastes grâce à l'outil ChoroP (Tableau 6) sur CsZCD et les deux autres séquences proches identifiées chez *Crocus sativus* : CsCCD4a et CsCCD4b (Rubio et al., 2008).

Tableau 6: Prédiction de séquence d'adressage au chloroplaste (ChloroP). Length : longueur de la protéine (acides aminés) ; Score : valeur statistique sur laquelle la prédiction est basée ; cTP : prédiction (Y = oui, - = non) ; cTP-length : longueur de la séquence d'adressage identifiée à partir du premier acide aminé. Il doit être noté qu'un nombre apparaît dans la colonne cTP-length, même dans le cas où aucune séquence d'adressage n'est identifiée. Il ne faut alors pas en tenir compte.

Name	Length	Score	cTP	CS-score	cTP-length
CsZCD	369	0.440	-	2.877	36
CsCCD4a	580	0.515	Y	0.974	48
CsCCD4b	569	0.477	-	0.974	39

D'après cette analyse, seule CsCCD4a est munie d'une séquence d'adressage au chloroplaste, avec un peptide signal qui s'étend jusqu'à la position 48 de la protéine. L'équipe du Pr Bilal Camara précise dans son étude (Bouvier et al., 2003b) que, malgré l'absence de peptide d'adressage clivable, leurs données immunologiques attestent d'une compartimentalisation de CsZCD au niveau des plastes.

1.3 Analyse structurale.

Nous avons réalisé une recherche d'homologues structuraux dans la PDB (Protein Data Bank). La structure disponible présentant la plus grande similarité de séquence avec CsZCD est celle de Viviparous 14 (VP14, code PDB 3NPE), isolée chez le maïs (Messing et al., 2010), avec 43% d'identité. Les séquences protéiques de CsZCD et de VP14 ont été alignées (Figure 32), et VP14 possède 237 acides aminés de plus que CsZCD au niveau de son extrémité N-terminale. Ces résidus ont été identifiés, puis localisés sur la structure de VP14 (Figure 33).

CsZCD VP14	----- MQGLAPPTSVSIHRHLPARSRARASNSVRFSPRAVSSVPPAECLQAPFHKPVDLPAPSR 60
CsZCD VP14	----- KPAAIAVPGHAAAPRKAEGGKKQLNLFQRAAAAALDAFEEGFVANVLERPHGLPSTADPA 120
CsZCD VP14	----- VQIAGNFAPVGERPPVHELFPVSGRIPPFIDGVYARNGANPCFDPVAGHHLFDGDGMVHAL 180
CsZCD VP14	-----MQVD 4 RIRNGAAESYACRFTETARLQERAIGRPVFPKAIGELHGHSGIARLALFYARAACGLVD 240 **
CsZCD VP14	PTKGIGLANTSLQFSNGRLHALCEYDLPYVVRSLPEDGDISTVGRIENN-VSTKSTIAHP 63 PSAGTGIVANAGLVYFNGRLLAMSEDDLPHYVRVAD-DGDLETVGRYDFDGLGCAMIAHP 299 *: * *:*. * : **** *:.* **** **: : **:*.*** : : : **
CsZCD VP14	KTDFVTGETFSFSYGPIQ-PYVTYSRYDCDGKKSGPDVPIFSFKEPSFVHDFAITHEHYAV 122 KLDPATGELHALSYDVIKRPYLKYFYFRPDGTSKD-DVEIP-LEQPTMIHDFAITENLVV 357 * **.* ** .:***. *: **:.* : **.*. ** * :::::*****: .*
CsZCD VP14	FPDIQIVMKPAEIVRGRRMIGPDLEKVPRLGLLPRYATSDSEMRWFDVPGFNMVHVNAW 182 VPDHQVVFVKLQEMLRGGSPVVLDEKTSRFGVLPKHAADASEMAWVDVPCFCFHLWNAW 417 .* ***:.* *:*** : * **.*:***:*. *** *.*. .*: **
CsZCD VP14	EEEGGEVVIVAPNVSPHENAIKRD-LLHVSVMARIELKSG-SVSRILLS---AENLD 237 EDEATGEVVVIGSCMTPADSIFNESDERLESVLTEIRLDARTGRSTRRAVLPPSQVNLE 477 *:.* ***:.. :.* :. :. * *. : **: :.* *. *:*. **:
CsZCD VP14	FGVIHRGYSGRKSRYAYLGVGDPMPKIRGVVKVDFELAGRGEVVARREFGVGCFGGEFF 297 VGMVNRNLLGRETRYAYLVAEPWPVKVSGFAKVDLSTG-----ELTKFEYGEGRFGGEPC 532 .*:*. ***:*****.*:.* **: *.***. . :*: :.* * *****
CsZCD VP14	FVPASSKKSGGE-EDDGYVVSYLHDEGKGESSFVMDARSPELEILAEVVLPRVVPYGFH 356 FVPMDPAAAHPRGEDDGYVLTFFVHDERAGTSELLVNA--ADMRLAETVQLPSRVPPGFH 590 *** .. : . *****:*** * *:***:*. :*: * * * ***:***
CsZCD VP14	GLFVTEAELLSQQ- 369 GTFITGQELEAQAA 604 * *: * ** :*

Figure 32: Alignement (ClustalW) des séquences protéiques de CsZCD (N° d'accèsion UniProtKB/Swiss-Prot Q84K96) et VP14 (N° d'accèsion UniProtKB/Swiss-Prot O24592.2). * = acides aminés identiques. : = acides aminés de même famille. . = acides aminés proches du point de vue de l'encombrement.

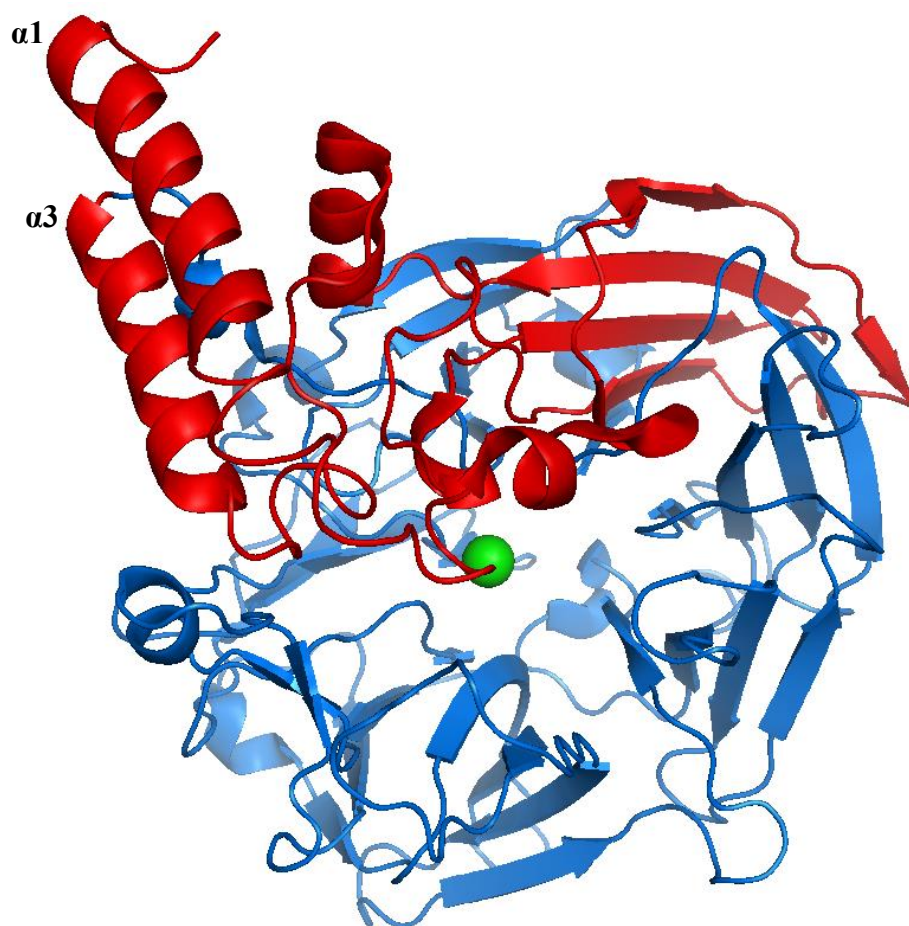


Figure 33: Structure de VP14 (*Zea mays*) visualisée grâce au logiciel PyMOL. Code PDB : 3NPE. Résolution 3.20 Å. Bleu : résidus qui s'alignent avec ceux de CsZCD ; rouge : résidus manquants dans CsZCD (positions 1 à 237 de VP14) ; vert : ion catalytique Fe^{2+} .

Les acides aminés manquants dans CsZCD par rapport à VP14 correspondent à plusieurs hélices α et à un feuillet β . Dans leur étude, Messing et al. (2010) décrivent les hélices $\alpha 1$ et $\alpha 3$ situées en N-terminal de la protéine comme étant responsables de son ancrage dans la membrane, étape indispensable à la fixation de son substrat. Le feuillet β fait partie des sept feuillets qui composent la structure en β -propeller à sept lames caractéristique de la famille des CCOs, indispensable à son activité catalytique d'après la littérature.

2. Description des constructions de biologie moléculaire utilisées dans la suite de l'étude.

Nous avons cherché à approfondir les données disponibles sur CsZCD, et tenter de « lever le voile » sur les résultats qui opposent les équipes ayant travaillé sur les CCOs chez *Crocus sativus* (Bouvier et al., 2003b; Rubio et al., 2008). Les tests réalisés dans ce but, qui composent la suite de ce chapitre, ont nécessité la construction d'une série de vecteurs d'expression (Tableau 7).

L'activité des enzymes suivantes a été testée :

- CsZCD (Bouvier et al., 2003b), identifiée comme la dioxygénase spécifique du clivage en positions 7,8(7',8') de la zéaxanthine ;
- CsCCD (Bouvier et al. 2003b), deuxième CCO caractérisée dans la même étude, identifiée comme une 9,10(9',10') Carotenoid Cleavage Dioxygenase ;
- CsZCD^{+N+P} : les informations apportées par les alignements de séquences et les analyses structurales précédemment décrites nous ont amenés à penser que la séquence publiée de CsZCD pourrait être une version tronquée de l'enzyme (ce qui serait en accord avec les conclusions de Rubio et al., 2008). Nous avons alors construit CsZCD^{+N+P}, constituée de CsZCD à laquelle nous avons ajouté une séquence N-terminale (N) et un peptide signal (P), dont les séquences correspondent aux acides aminés supplémentaires retrouvés dans CsCCD4a et CsCCD4b (Rubio et al., 2008).
- CsZCD^{+N} : cette séquence a été construite dans le cadre de tests d'activité chez *S. cerevisiae*. Chez la levure, les protéines interagissant avec les membranes sont dirigées au réticulum endoplasmique, à la mitochondrie, ou à la membrane plasmique. C'est la nature de la séquence d'adressage (ou encore peptide signal) qui détermine la destination d'une protéine (Rapoport, 2007). Souhaitant éviter un adressage de CsZCD à la mitochondrie, nous avons construit CsZCD a à laquelle nous avons ajouté uniquement la séquence N-terminale de CsCCD4b (N), sans peptide signal.
- VvCCD1, une CCO qui catalyse le clivage de la zéaxanthine en positions 9,10(9',10') (Mathieu et al. 2005). Nous l'avons utilisée en tant que témoin positif de nos expériences afin de vérifier le bon déroulement des différents protocoles de tests d'activité que nous avons appliqué, et pouvoir conclure à l'issue de ces tests.

Tableau 7: Description des constructions de biologie moléculaire utilisées dans la suite de cette étude. Ps, peptide signal ; N-ter, séquence N-terminale, (His)₆, tag histidine, Thio, thiorédoxine ; GST, glutathione-S-transférase. Les différentes longueurs de barres représentant les protéines reflètent les tailles respectives des enzymes testées.

Nom de la construction	Protéine synthétisée			Vecteur	Inducteur (I) / Répresseur (R)	Hôte
	Tag N-ter	CCO	Tag C-Ter			
Z/p15	(His) ₆	CsZCD		pET15b	IPTG (I)	<i>E. coli</i>
Z ^{N+P} /p15	(His) ₆	Ps N-ter CsZCD				
C/p15	(His) ₆	CsCCD				
Z/pBAD	Thio	CsZCD		pBAD/TOPO® Thiofusion	Arabinose (I)	
V/pGEX	GST	VvCCD1	(His) ₆	pGEX-4T2	IPTG (I)	
Z/V185		CsZCD		pCM185	Doxycycline (R)	<i>S. cerevisiae</i>
Z ^N /V185		N-ter CsZCD				
Z ^{N+P} /V185		Ps N-ter CsZCD				
C/V185		CsCCD				
V/V185		VvCCD1				
Z/V60		CsZCD		YeDP60	Galactose (I)	

3. Expression de la protéine chez *E. coli*.

Afin de réaliser une étude structurale, fonctionnelle ou biochimique d'une protéine, la production de la protéine sous une forme recombinante dans l'hôte *E. coli* est privilégiée. Deux problèmes inhérents à ce mode d'expression peuvent être rencontrés : une expression trop faible, voire indétectable de la protéine, ou une expression de la protéine sous une forme insoluble. Lorsque le problème de la solubilité se pose, aucune solution universelle n'existe, mais des stratégies au cas par cas doivent être testées, comme changer de vecteur d'expression, changer de souche-hôte de *E. coli*, modifier les conditions de culture et d'induction de la transcription, ou encore recoder la séquence codante de la protéine selon le biais de codon de l'hôte (Gopal & Kumar, 2013). Bouvier et al. (2003b) signalent dans leur étude qu'ils ont eu des difficultés à obtenir CsZCD sous une forme soluble. Ils ont pu surmonter ce problème en abaissant la température de culture des cellules à 20°C. La diminution de la température de culture a pour conséquence de ralentir la croissance de *E. coli*, laissant ainsi plus de temps disponible pour le repliement de la protéine à chaque génération. De plus, des « Cold Shock Proteins » sont induites lorsque *E. coli* est soumise à une diminution de température, il a été montré qu'elles peuvent avoir un effet sur le repliement des protéines (Yamanaka, 1999). Un repliement correct de la protéine permet une meilleure solubilité, et est indispensable à l'activité catalytique de l'enzyme.

En vue de réaliser des tests d'activité, nous avons tout d'abord confirmé l'expression de CsZCD dans notre système, puis nous avons effectué des tests d'optimisation de la production de la protéine sous une forme soluble.

3.1 Validation de l'expression.

Nous avons débuté nos tests en confirmant la production de CsZCD chez *E. coli*. Pour cela, nous avons utilisé la construction Z/p15, exprimée dans la souche de *E. coli* BL21(De3). Différents prélèvements de cellules ont été effectués pendant la culture : au moment de l'inoculation (T0), puis un prélèvement par heure après l'induction à l'IPTG, pendant trois heures (T1, T2, T3). Les fractions solubles et insolubles des cellules ont ensuite été déposées sur un gel de polyacrylamide pour visualiser la production de la protéine (Figure 34).

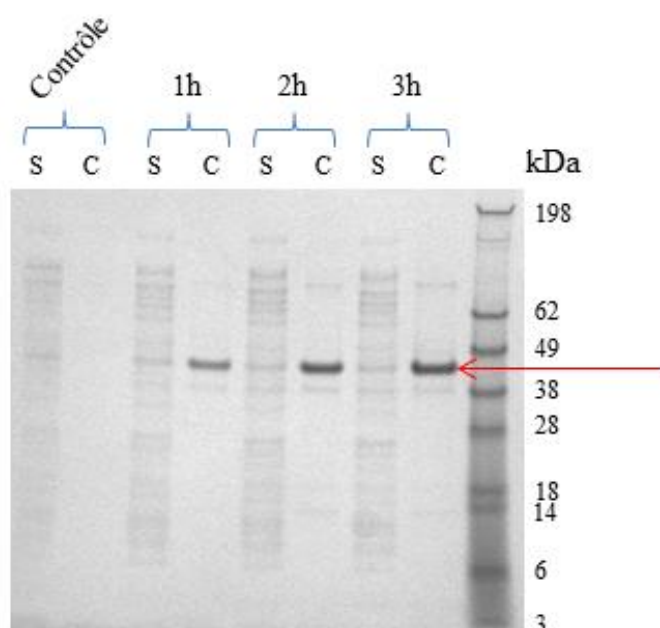


Figure 34: Migration sur gel SDS-PAGE des fractions solubles et insolubles d'échantillons prélevés dans une culture (37°C) de *E. coli* BL21(De3) transformée par Z/p15. S = surnageant ; C = culot ; Contrôle = souche non induite ; 1h = prélèvement à T1 ; 2h = prélèvement à T2 ; 3h = prélèvement à T3. Taille attendue de CsZCD : 43kDa. La flèche rouge indique CsZCD.

La souche BL21(De3) produit bien CsZCD à partir d'une heure après induction à l'IPTG. De plus, la bande correspondant à CsZCD sur le gel s'intensifie entre les prélèvements T1 et T3, ce qui signifie que cette protéine s'accumule dans les cellules à partir du moment où l'induction à l'IPTG a été effectuée. Cependant, la quasi-totalité de la protéine est retrouvée dans la fraction insoluble (culot). Cela est en accord avec les difficultés de production que Bouvier et al. (2003b) expliquent avoir rencontrées.

3.2 Essai de re-solubilisation.

Dans un premier temps, nous avons tenté de re-solubiliser la protéine contenue dans la fraction insoluble en incubant un culot précédemment obtenu avec du TRITON X-100 1% (w/v). Après plusieurs heures d'incubation, Les fractions solubles et insolubles ont à nouveau été déposées sur un gel de polyacrylamide (Figure 35).

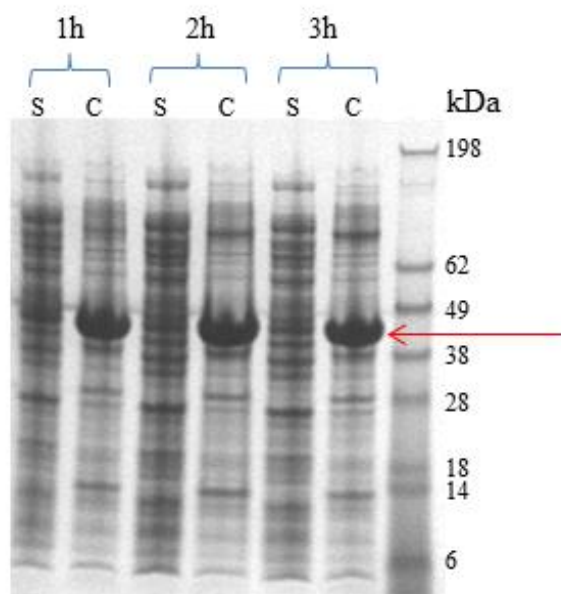


Figure 35: Migration sur gel SDS-PAGE d'un essai de re-solubilisation de CsZCD contenue dans un culot par ajout de TRITON X-100 1% (w/v). S = surnageant ; C = culot ; 1h = incubation à 30°C pendant 1 heure ; 2h = incubation à 30°C pendant 2 heures ; 3h = incubation à 30°C pendant 3 heures. La flèche rouge indique CsZCD.

Même après trois heures d'incubation, CsZCD est toujours dans la fraction insoluble. Ne parvenant pas à re-solubiliser CsZCD après production, nous avons donc essayé d'optimiser les conditions d'expression de CsZCD pour améliorer sa solubilité.

3.3 Optimisation des conditions de culture.

Pour améliorer les conditions d'expression de CsZCD sous forme soluble chez la souche *E. coli* BL21(De3), deux conditions ont été testées : la baisse de la température à 20°C (en accord avec Bouvier et al., 2003b), et le changement de milieu de culture. Le milieu ZYM 5052 est un milieu auto-inductible ((Studier, 2005) ; Annexe 3). Sa composition mixte de glucose et de lactose permet une induction lente et progressive de la T7 ARN polymérase chez BL21(De3). La synthèse plus progressive de la protéine qui en résulte pourrait être plus adaptée à son repliement, et donc permettre de retrouver une plus grande fraction de CsZCD dans la fraction soluble (Figure 36).

Cependant, même avec ce milieu, la protéine est toujours très majoritairement retrouvée dans la fraction insoluble.

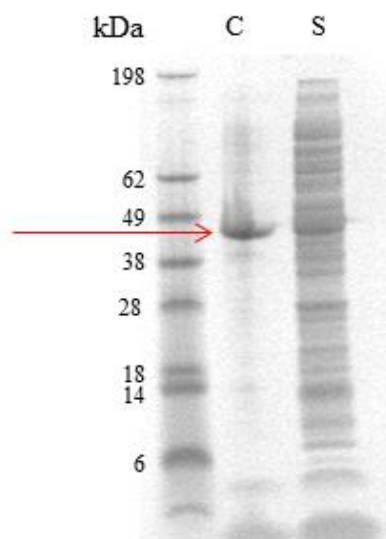


Figure 36: Migration sur gel SDS-PAGE des protéines de *E. coli* BL21(De3) transformée avec Z/p15, cultivées en milieu ZYM 5052 à 20°C. C = culot ; S = Surnageant. La flèche rouge indique CsZCD.

3.4 Essai de différentes souches recombinantes.

Une des stratégies connues pour optimiser l'expression sous forme soluble d'une protéine recombinante chez *E. coli* est de tester plusieurs fonds génétiques différents (Gopal & Kumar, 2013). En effet, les souches de *E. coli* optimisées pour la production de protéines recombinantes peuvent être plus adaptées à la production d'une protéine ou d'une autre. Nous avons décidé de tester deux autres souches (Tableau 8) : C41(De3) et Origami(De3) (Figure 37).

Tableau 8: Liste des souches de *E. coli* utilisées pour optimiser de la production de CsZCD sous forme soluble.

Souche (<i>E. coli</i>)	Utilisation
C41(De3)	Souche optimisée pour la production de protéines membranaires et toxiques.
Origami(De3)	Souche optimisée pour la formation des ponts disulfures.

Sachant que nos analyses de séquences et celles de la littérature ont montré que les CCOs contiennent probablement une zone d'ancrage membranaire, nous avons testé des souches adaptées à la production de ce type de protéines.

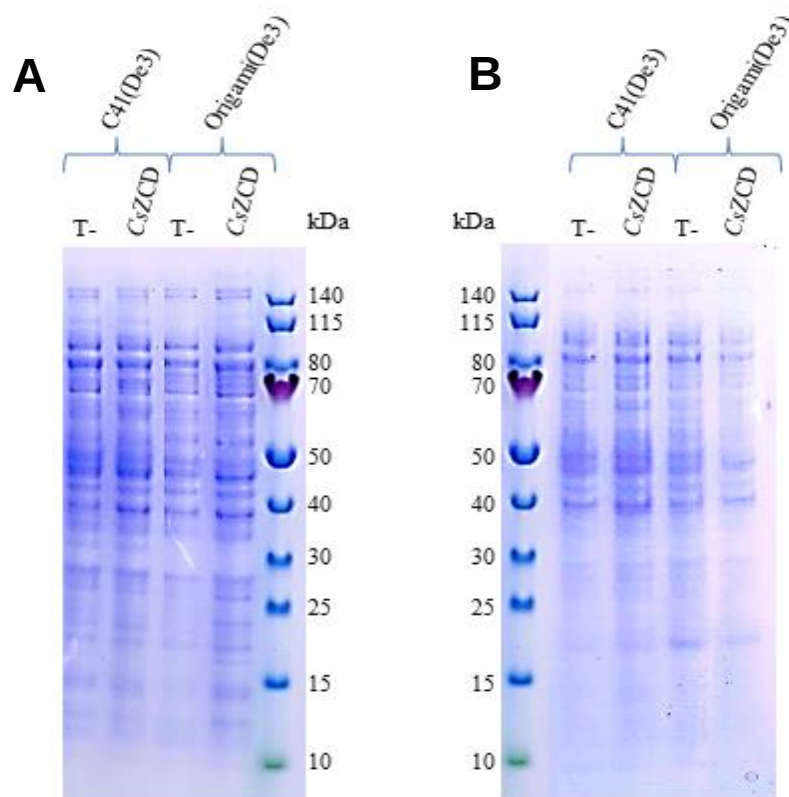


Figure 37: Migration sur gel SDS-PAGE des protéines de *E. coli*, C41(De3) et Origami(De3) transformées avec Z/p15 et cultivées en ZYM 5052 à 20°C. (A) Dépôt des surnageants. (B) Dépôts des culots. T- = souche contrôle transformée avec pET15b vide. CsZCD = souche transformée avec la construction Z/p15. Taille attendue de CsZCD: 43 kDa.

CsZCD n'est pas présente dans les échantillons déposés (pas de bande supplémentaire entre les dépôts témoins et les dépôts d'expression). La protéine n'est pas produite dans ces conditions. Les souches C41(De3) et Origami(De3) ne sont donc pas adaptées à la production de CsZCD.

L'ensemble des tests d'expression de CsZCD chez *E. coli* que nous avons réalisé mettent en évidence une enzyme produite de façon quasi exclusive dans la fraction insoluble. Les essais d'optimisation qui ont été faits pour surmonter ce problème se sont révélés infructueux. Cependant, il est possible qu'une petite proportion d'enzyme soluble soit suffisante pour pouvoir observer une activité catalytique. Les auteurs d'études de caractérisation de CCOs présentes dans la littérature ont utilisé un système de production chez *E. coli* (Bouvier et al., 2003b; Mathieu et al., 2007; Messing et al., 2010). La reproduction fidèle de leurs protocoles d'expression et de tests d'activité pourrait nous permettre d'observer un produit de clivage de la zéaxanthine.

4. Tests d'activité.

4.1 Reproduction des conditions de Bouvier et al., 2003b.

CsZCD ayant été isolée et caractérisée par l'équipe du Pr Bilal Camara en 2003, nous avons reproduit strictement les conditions publiées. Les conditions de culture, d'induction et de cassage des cellules ont été reproduite telles que décrites dans Bouvier et al., 2003b, avec la construction Z/pBAD. Les fractions solubles et insolubles récupérées après le cassage des cellules ont été déposées sur un gel de polyacrylamide (Figure 38A). CsZCD est détectée dans les cellules exprimant Z/pBAD mais pas dans le contrôle. Elle est produite très majoritairement dans la fraction insoluble. Etant donnée la diversité de protéines présentes dans la fraction soluble, il est difficile de dire si la protéine en question se trouve aussi dans le surnageant déposé. Un test d'activité sur ce surnageant a été réalisé dans les conditions décrites par Bouvier et al. (2003b). Les produits de ce test ont été séparés par HPLC et détectés en sortie de colonne à 440 nm, longueur d'onde d'absorption du dialdéhyde attendu (Figure 38B). Nous n'avons détecté dans ces conditions aucun produit de clivage de la zéaxanthine.

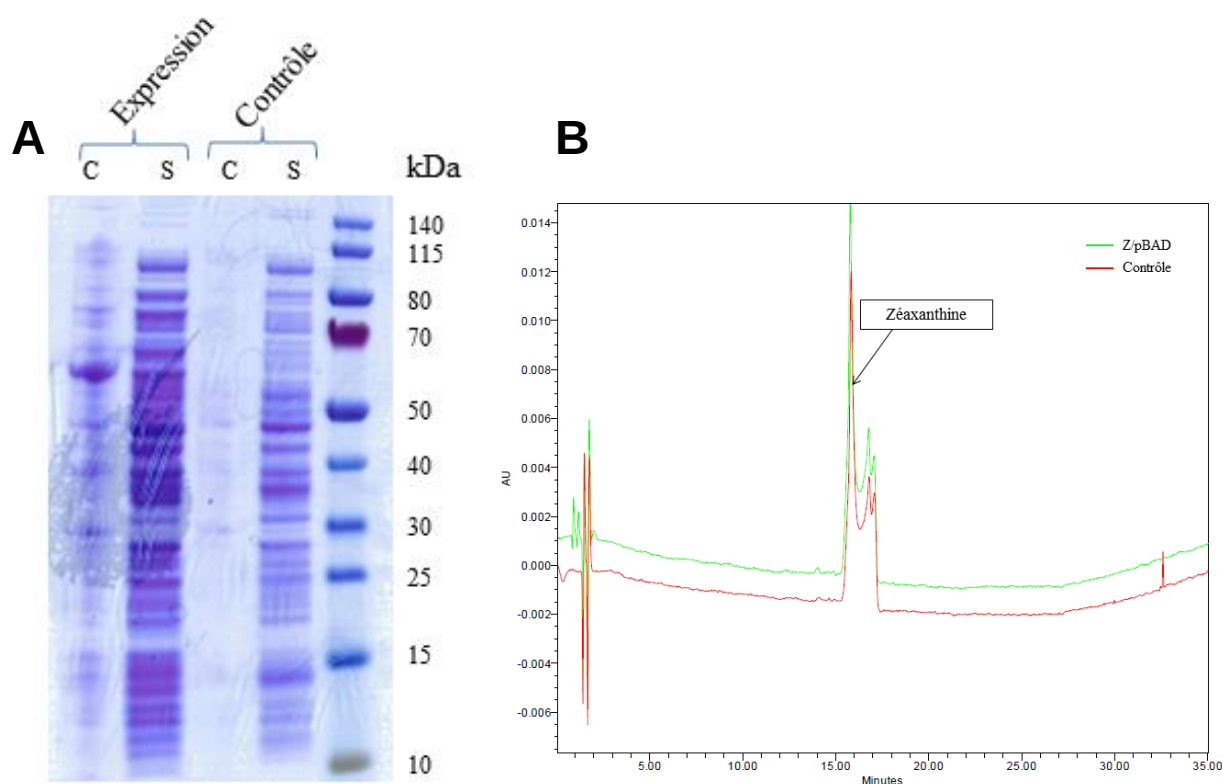


Figure 38: (A) Migration sur gel SDS-PAGE des différentes fractions obtenues après cassage des cellules de *E. coli* (TOP10) cultivées dans les conditions de (Bouvier et al., 2003b). Expression = souche TOP10 transformée avec la construction Z/pBAD ; Contrôle = souche TOP10 non transformée ; S = surnageant, C = culot. Taille attendue de CsZCD + thiorédoxine = 53 kDa. (B) analyse HPLC (440 nm) résultant de l'injection de tests d'activité effectués dans les conditions de (Bouvier et al., 2003b). Vert : test d'activité effectué sur le surnageant de cassage d'une souche TOP10 exprimant la construction Z/pBAD. Rouge : test d'activité effectué sur le surnageant de cassage de la souche contrôle.

4.2 Reproduction de conditions optimisées pour la caractérisation d'une CCO chez *Vitis vinifera*.

Etant donnés nos essais non conclusifs pour détecter une activité de CsZCD, nous avons cherché à éliminer l'hypothèse que nos conditions d'expression de la protéine et de test d'activité *in vitro* puissent être inadaptées. Nous avons donc reproduit les conditions d'une autre étude (Mathieu et al., 2005) ayant caractérisé une CCO chez *Vitis vinifera* (vigne) : VvCCD1. VvCCD1 clive la zéaxanthine et la lutéine en positions 9,10(9',10') pour donner naissance à la 3-hydroxy- β -ionone et à un dialdéhyde en C14 (Figure 39) qui possède un maximum d'absorption à 416 nm.

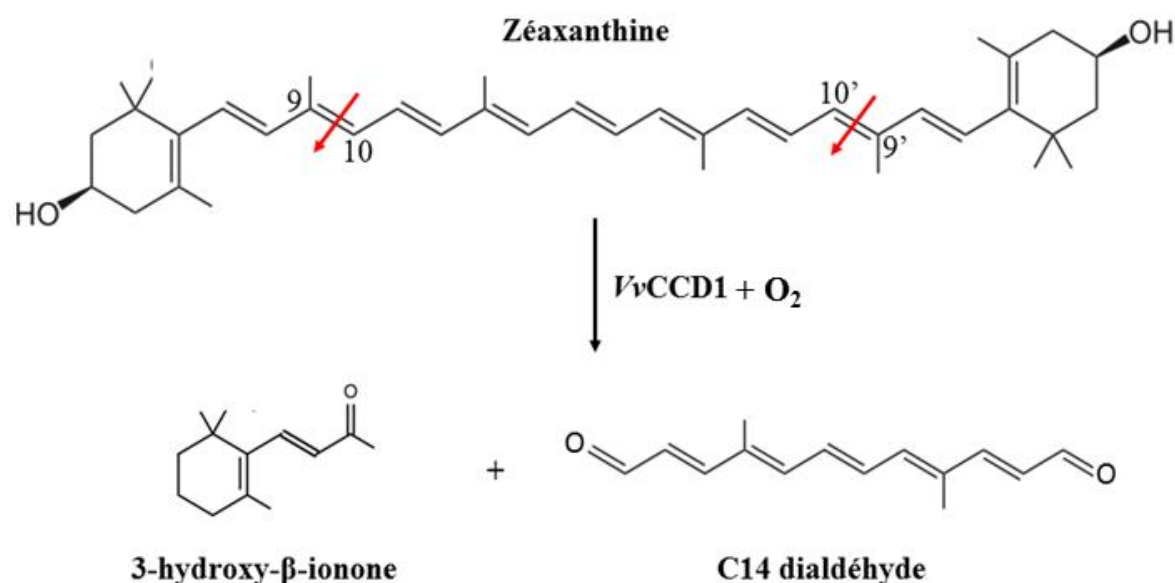


Figure 39: Réaction de clivage de la zéaxanthine en positions 9,10(9',10') catalysée par VvCCD1 (Mathieu et al., 2005).

4.2.1 Test d'activité de VvCCD1.

Pour caractériser VvCCD1, les auteurs (Mathieu et al., 2005) l'ont exprimée sous la forme d'une protéine fusion avec la GST (expression dans vecteur de la série pGEX). Le Dr Nancy Terrier, auteur de l'étude, nous a fourni la construction V/pGEX dans la souche BL21(De3). Dans un premier temps, les conditions décrites par Mathieu et al. (2005) ont été strictement appliquées pour vérifier que leur reproduction nous menait bien à la détection d'un produit de clivage de la zéaxanthine par VvCCD1 (Figure 40).

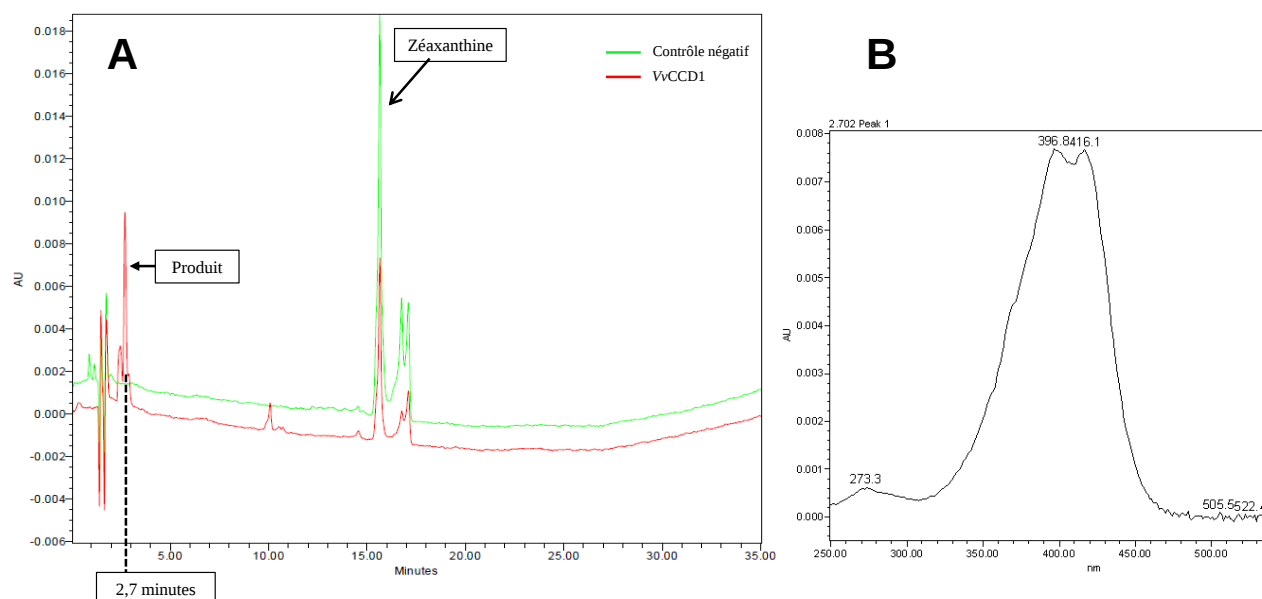


Figure 40: (A) Analyse HPLC (416 nm) résultant de l'injection des produits du test d'activité de VvCCD1 fait dans les conditions de (Mathieu et al., 2005). Tests d'activité effectués sur les surnageants de cassage de souches BL21(De3) : Rouge : V/pGEX. Vert : BL21(De3) non transformée. (B) Spectre d'absorption du produit élué à 2,7 minutes récolté entre 250 nm et 500 nm.

Nous observons un pic d'absorption à 416 nm, 2.7 minutes après l'injection. Ce pic est absent du test d'activité réalisé avec la souche contrôle. Mathieu et al. (2005) décrivent un produit avec des maxima d'absorption à 396 nm et 414 nm pour le dialdéhyde en C₁₄. Le spectre du produit que nous observons est de forme identique et possède des maxima d'absorption très proches (398 nm et 417.3 nm). Le pic observé à 10 minutes possède un spectre d'absorption qui ne correspond pas au spectre attendu d'un des produits de la réaction. Nous observons donc bien un produit de clivage de la zéaxanthine par VvCCD1 dans les conditions d'expression et de test d'activité *in vitro* décrites par Mathieu et al. (2005). Dans la suite de l'étude, nous avons donc appliqué ce protocole aux différentes enzymes que nous avons clonées sur des vecteurs d'expression chez *E. coli*, afin de détecter leurs produits.

4.2.2 Tests d'activité dans les conditions de Mathieu et al. 2005.

Les conditions d'expression, d'induction, et de test d'activité *in vitro* décrites par Mathieu et al. (2005) ont été appliquées à des cultures de la souche de *E. coli* BL21(De3) transformées avec les constructions suivantes :

- Z/p15b ;
- Z^{+N+P}/p15 ;
- V/pGEX (contrôle positif);
- pET15b vide (contrôle négatif).

Les produits extraits suite au test d'activité *in vitro* ont été séparés par HPLC et détectés à 416 nm (longueur d'onde d'absorption maximale du produit de VvCCD1) et à 440 nm (longueur d'onde d'absorption maximale du produit de CsZCD) (Figure 41).

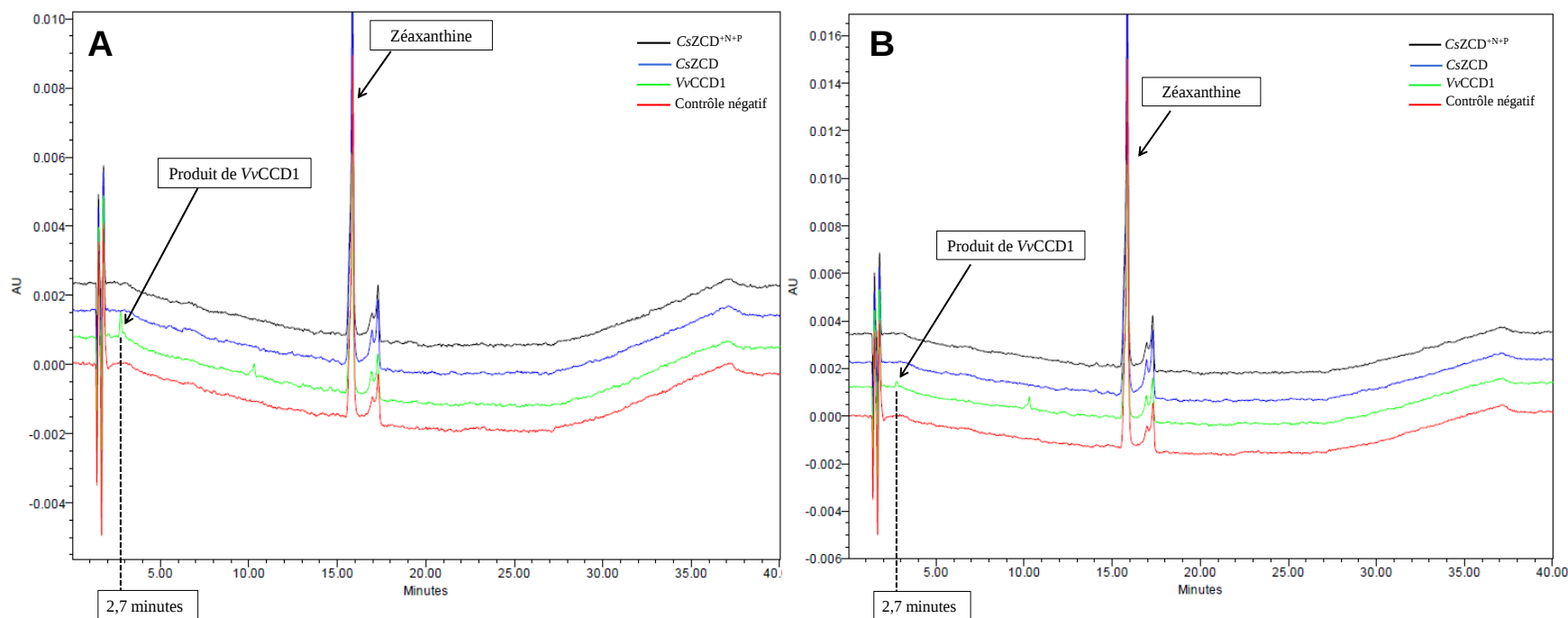


Figure 41: Analyse HPLC des produits des tests d'activité de CsZCD, CsZCD^{N+P} et VvCCD1 dans les conditions de Mathieu et al. 2005. (A) Spectres enregistrés à 416 nm. (B) Spectres enregistrés à 440 nm. Les tests d'activité ont été réalisés sur les fractions solubles issues du cassage de cellules BL21(De3) exprimant : Noir : Z^{N+P}/p15. Bleu : Z/p15. Vert : V/pGEX. Rouge : pET15b vide.

Aucun produit de clivage de la zéaxanthine n'a pu être détecté, à 416 nm ou à 440 nm. Nous pouvons également remarquer que le pic de produit de VvCCD1 observé à 416 nm (Figure 41A) est moins intense que celui qui a été présenté en Figure 40 lors du premier test. De façon à tester la stabilité du produit (dialdéhyde) de VvCCD1, nous avons à nouveau analysé les produits du test d'activité que nous venons de décrire après leur conservation 24 heures à 4°C (Figure 42).

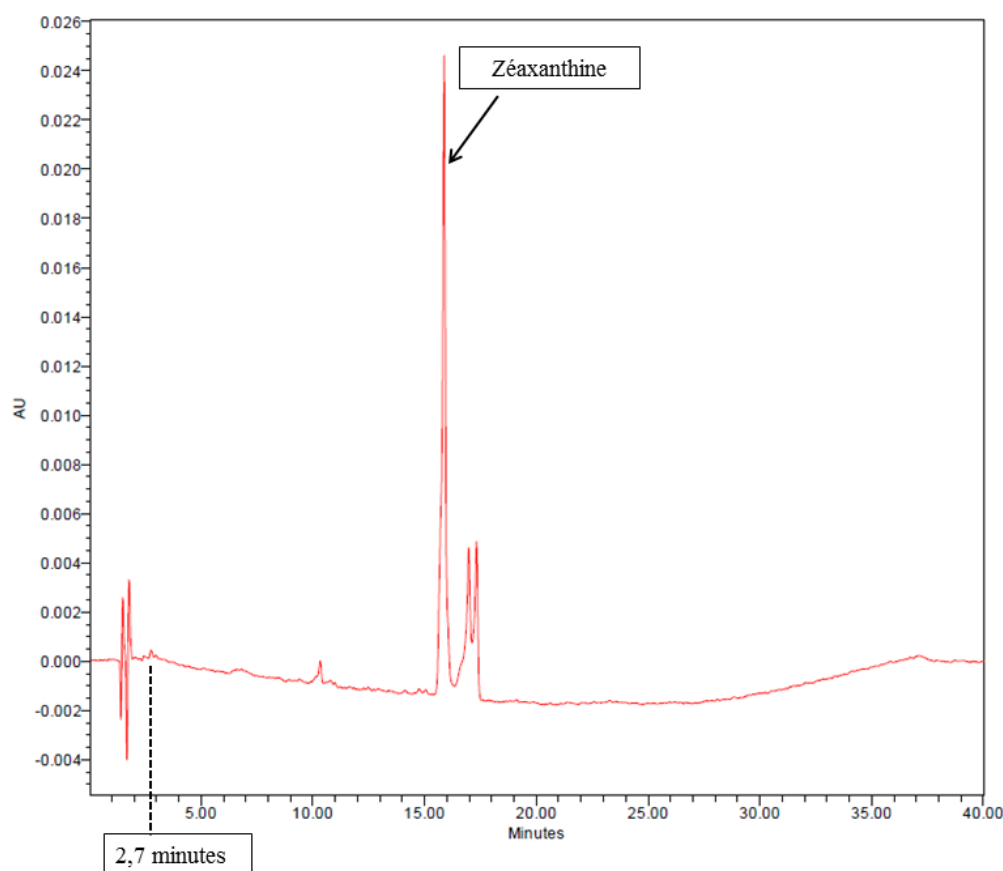


Figure 42: Analyse HPLC (416 nm) résultant de l'injection d'un test d'activité de VvCCD1 fait dans les conditions de Mathieu et al. 2005, 24 heures après l'arrêt du test d'activité.

Le produit élué à 2,7 minutes n'est à présent presque plus détectable. Nous en déduisons que ce produit est peu stable. Les caroténoïdes et les apocaroténoïdes sont des antioxydants, c'est à dire qu'ils sont particulièrement sensibles à l'oxydation. Les aldéhydes en particulier sont des molécules très instables. Cette instabilité du produit est probablement liée à des facteurs comme

l'oxydation par l'air ou la lumière. Le temps de manipulation apparait donc comme un élément critique de la réussite de ce test d'activité, et il est possible par conséquent que l'absence de produits de CsZCD et CsZCD^{+N+P} soit liée à cela. En effet, il suffirait que ces enzymes soient moins actives que VvCCD1, et/ou que leurs produits soient un peu plus instable encore pour que nous ne puissions pas les observer dans nos conditions. Dans les expériences qui suivent, nous avons cherché à améliorer les conditions du test d'activité pour une observation optimale des produits recherchés.

4.2.3 Essai d'optimisation du test d'activité 1.

Une recherche de meilleures conditions pour l'observation du produit de VvCCD1 a été effectuée au sein du laboratoire où elle a été caractérisée (équipe du Dr Zya Günata, Université Montpellier II). Les conditions de pH et de température du test d'activité *in vitro* ont été optimisées. Une augmentation de 45% de la dégradation du substrat par l'enzyme dans un temps donné a pu être observée (communication personnelle). Un test d'activité (Figure 43) a été réalisé dans les conditions optimisées après avoir exprimé les enzymes chez *E. coli* BL21(De3) transformée avec les constructions suivantes :

- Z/p15 ;
- Z^{+N+P}/p15 ;
- V/pGEX (contrôle positif) ;
- pET15b vide (contrôle négatif).

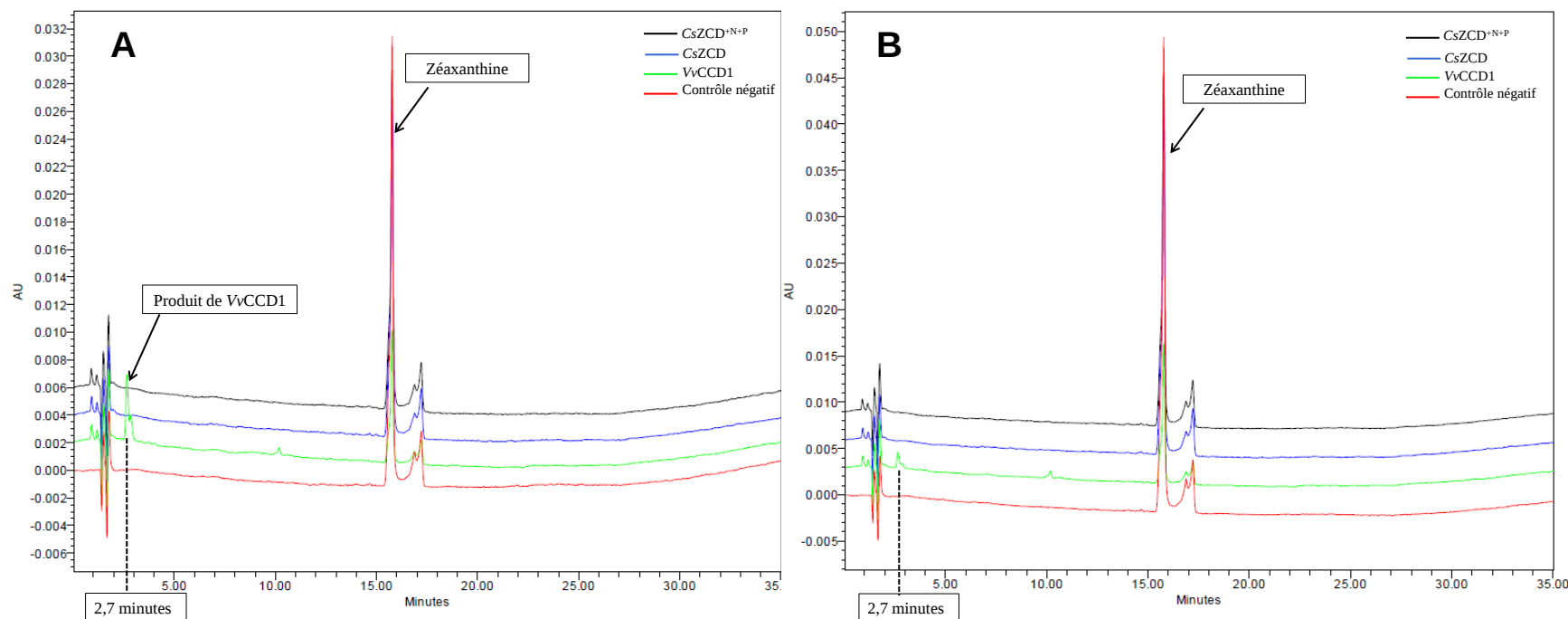


Figure 43: Analyse HPLC des produits des tests d'activité de CsZCD, CsZCD^{N+P} et VvCCD1 dans les conditions de Mathieu et al. 2005 optimisée pour le pH et la température de test. (A) Spectres enregistrés à 416 nm. (B) Spectres enregistrés à 440 nm. Les tests d'activité ont été réalisés sur les fractions solubles issues du cassage de cellules BL21(De3) exprimant : Noir : Z^{N+P}/p15. Bleu : Z/p15. Vert : V/pGEX. Rouge : pET15b vide.

Nous observons bien le produit de clivage de la zéaxanthine par VvCCD1 à 416 nm. Nous ne détectons de nouveau aucun produit de clivage, ni avec CsZCD ni avec CsZCD^{+N+P}.

4.2.4 Essai d'optimisation du test d'activité 2.

Dans une étude de 2007 (Mathieu et al., 2007), la même équipe a publié un travail d'optimisation des conditions de culture de *E. coli* recombinante, et de test d'activité *in vitro* pour observer l'activité de VvCCD1. Ces conditions ont été appliquées aux enzymes CsZCD et CsZCD^{+N+P}. De plus, les résultats de tous les tests présentés précédemment ayant été négatifs, nous avons émis l'hypothèse que la zéaxanthine ne serait pas le bon substrat pour ces enzymes. Nous avons donc réalisé ce test d'activité sur deux substrats différents : la zéaxanthine et le β -carotène (Figure 44). Aucun produit de clivage de CsZCD ou de CsZCD^{+N+P} n'a pu être détecté, quel que soit le substrat testé.

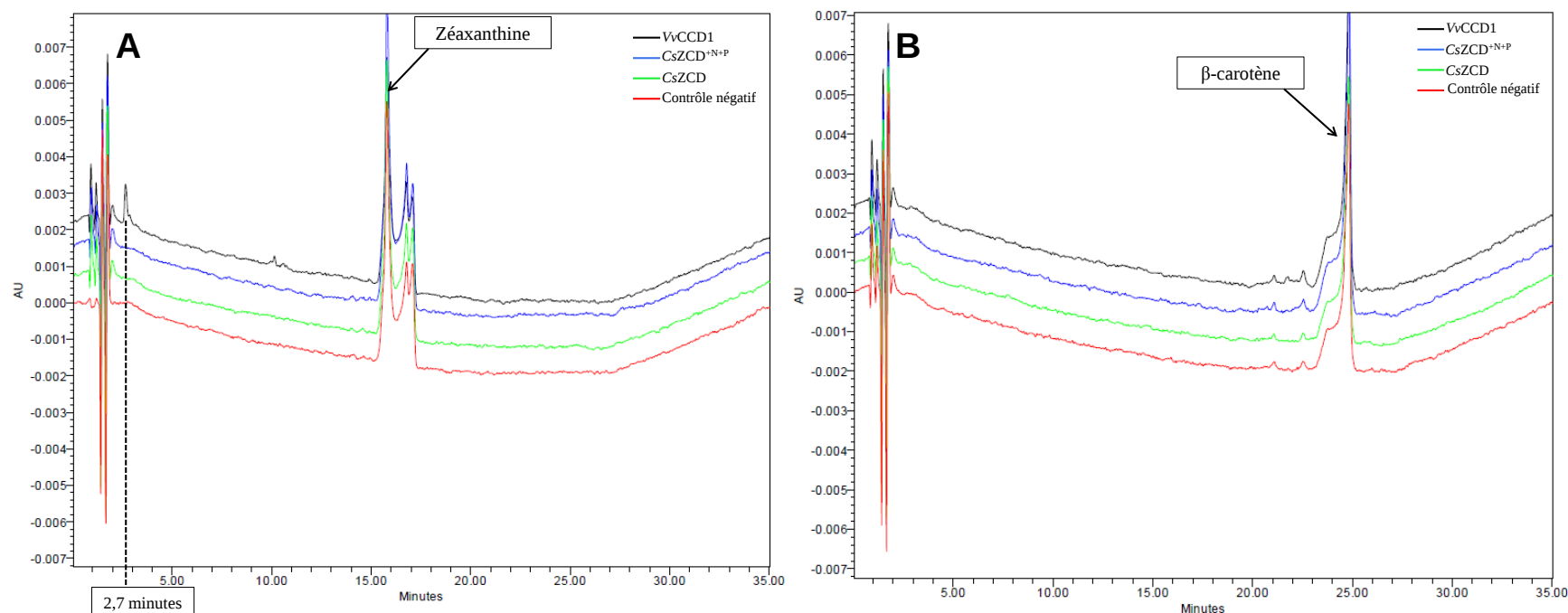


Figure 44: Analyse HPLC des produits des tests d'activité de CsZCD, CsZCD^{N+P} et VvCCD1 dans les conditions de Mathieu et al. 2007. Les spectres VvCCD1 sont enregistrés à 416 nm, les autres à 440 nm. (A) Substrat = zéaxanthine. (B) Substrat = β -carotène. Les tests d'activité ont été réalisés sur les fractions solubles issues du cassage de cellules BL21(De3) exprimant : Noir : V/pGEX ; Bleu : Z^{N+P}/p15 ; Vert : Z/p15 ; Rouge : pET15b vide.

4.2.5 Test d'activité de CsCCD.

D'après Bouvier et al. (2003b), CsCCD clive la zéaxanthine, et d'autres caroténoïdes, en positions 9,10 et 9',10', exactement comme le fait VvCCD1. Nous devrions donc, dans des conditions où ces deux protéines sont correctement produites et actives, observer le même produit (dialdéhyde C₁₄). Les conditions de culture, d'induction et de test d'activité de Mathieu et al. (2007) ont été appliquées pour réaliser un test d'activité de CsCCD sur la zéaxanthine (Figure 45). L'activité de CsCCD sur la zéaxanthine n'a pas été observée dans ces conditions.

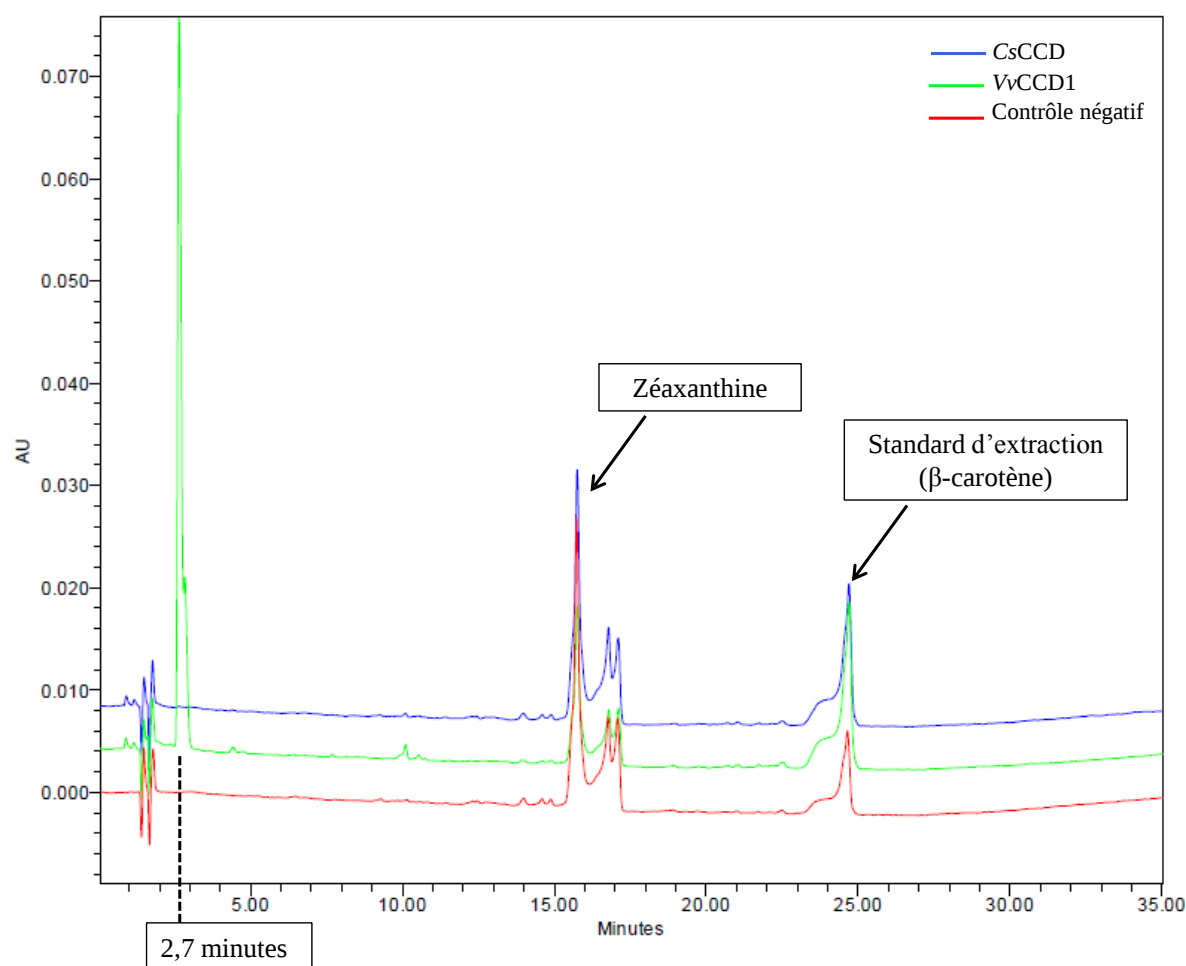


Figure 45: Analyse HPLC (416 nm) résultants de l'injection des produits des tests d'activité de CsCCD et VvCCD1 dans les conditions de Mathieu et al. 2007. Les tests d'activité ont été réalisés sur les fractions solubles issues du cassage de cellules BL21(De3) exprimant : Bleu : C/p15 ; Vert : V/pGEX ; Rouge : pET15b vide.

Pour vérifier la production de la protéine, une migration sur gel de polyacrylamide a été effectuée en déposant les surnageants et les culots issus du cassage des cultures de BL21(De3) ayant servi à faire ce test (Figure 46).

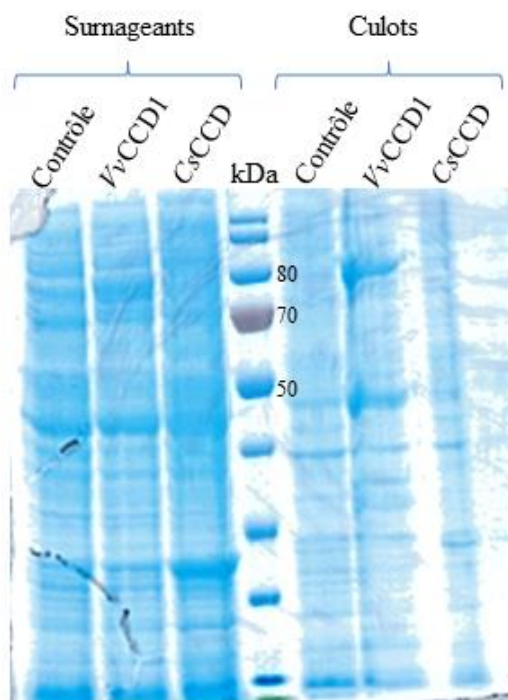


Figure 46: Migration sur gel SDS-PAGE des fractions de cassage de souches de *E. coli* BL21(De3) cultivées dans les conditions de Mathieu et al. 2007. Les cellules exprimaient les constructions suivantes : Contrôle : pET15b vide ; VvCCD1 : V/pGEX ; CsCCD : C/p15. Taille attendue de CsCCD : 62 kDa. Taille attendue de VvCCD1 + GST : 88 kDa.

Il est difficile de conclure au sujet d'une production d'enzymes solubles sur ce gel. En effet, si VvCCD1 ou CsCCD sont présentes dans les surnageants, elles sont en quantité trop faibles pour être visibles de façon évidente sur le gel. En revanche, VvCCD1 est visible dans le culot, ce qui n'est pas le cas de CsCCD. Cela suggère que les conditions d'expression appliquées ici à CsCCD ne sont pas adaptées. Nous rappelons que VvCCD1 est exprimée sous la forme d'une protéine fusion avec la GST, ce qui n'est pas le cas de CsCCD. Il est possible que cette construction soit déterminante pour la production de VvCCD1, et qu'il en soit de même pour celle de CsCCD. Nous pouvons alors imaginer que cette remarque soit valable aussi pour CsZCD.

4.3 Tests d'activité *in vitro* sur lysats de levures.

Le fait de ne pas être parvenus à observer une activité dioxygénase en adoptant une stratégie d'expression chez *E. coli* peut être lié à un mauvais repliement de l'enzyme dans l'hôte procaryote, peu performant si un système de modification post-traductionnelle, ou des structures cellulaires spécifiques sont nécessaires. Dans le but de vérifier cette hypothèse, nous avons utilisé une stratégie d'expression dans l'hôte eucaryote chez lequel la voie métabolique hétérologue complète doit être implémentée: *S. cerevisiae*. L'enzyme CsZCD a été exprimée dans la souche CEN.PK via la construction Z/V60. Après avoir induit l'expression de la protéine, les fractions microsomales ont été récoltées. Il s'agit d'un ensemble de membranes internes, contenant en particulier des fractions du réticulum endoplasmique, dans lequel on peut souvent retrouver des protéines membranaires. Lors de la purification des microsomes, nous avons récupéré à la suite d'une première centrifugation les débris cellulaires, les membranes externes et les mitochondries, puis, suite à une précipitation des microsomes et une seconde centrifugation, les microsomes sont récupérés dans le culot, et le surnageant correspond au cytosol des cellules cassées. Ne sachant pas quelle sera la localisation de CsZCD une fois exprimée dans la levure, un test d'activité *in vitro* a été réalisé sur chacune de ces trois fractions (Figure 47). Quelle que soit la fraction cellulaire testée, aucun produit de clivage de la zéaxanthine n'a pu être détecté.

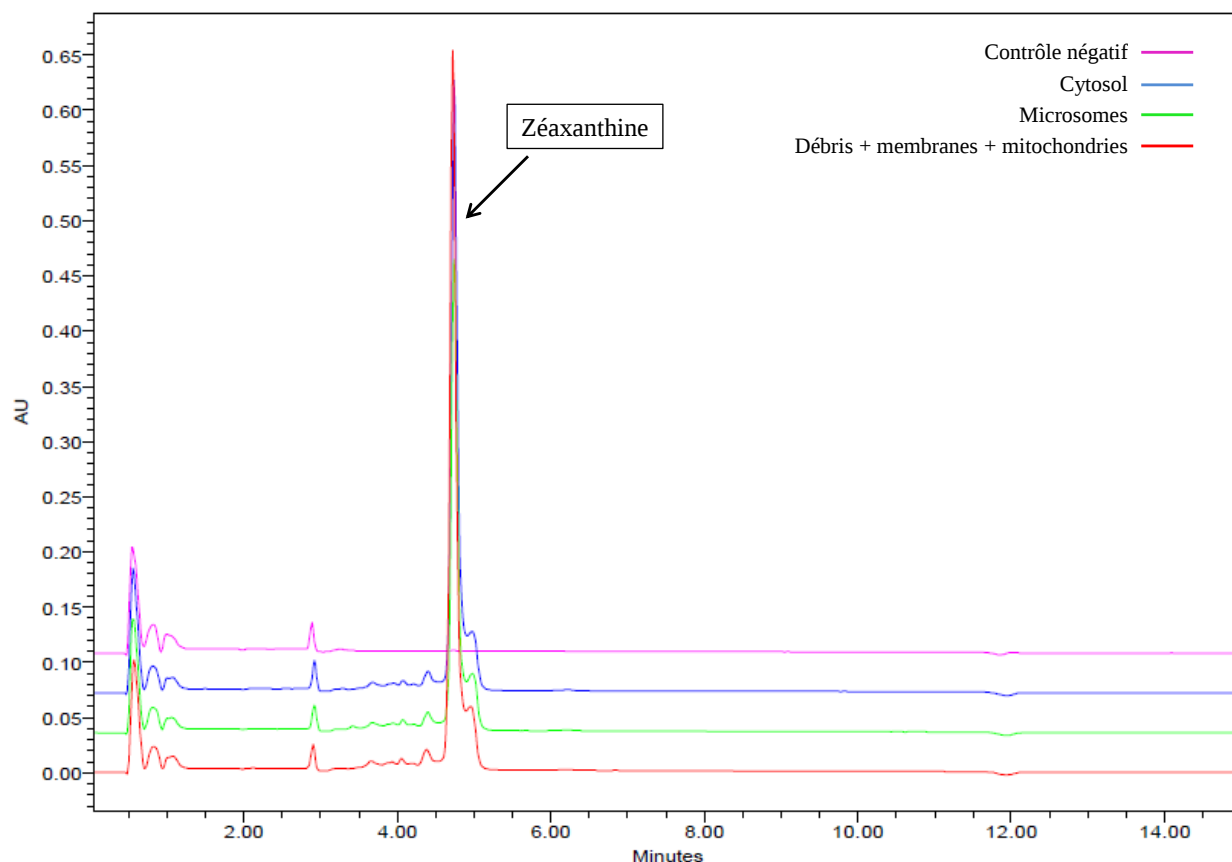


Figure 47: Analyse HPLC (440 nm) résultants de l'injection de tests d'activité *in vitro* après expression de CsZCD chez *S. cerevisiae*. Tests d'activité effectués sur des lysats de levure CEN.PK. Rose : CEN.PK non transformée ; Bleu : fraction cytosolique de levures exprimant CsZCD ; Vert : fraction microsomale de levures exprimant CsZCD ; Rouge : débris cellulaires et les mitochondries de levures exprimant CsZCD.

4.4 Tests d'activité *in vivo* chez *Saccharomyces cerevisiae*.

La souche de *S. cerevisiae* YSG50 ayant été modifiée afin de produire de la zéaxanthine nous a été fournie par l'équipe du Dr Huimin Zhao qui l'a mise au point (Shao et al., 2009). Nous avons transformé cette souche avec les constructions suivantes :

- Z/V185 ;
- Z^{+N+P} /V185 ;
- Z^{+N} /V185 ;
- C/V185 ;
- V/V185.

La construction $CsZCD^{+N+P}$ contient un peptide d'adressage aux plastides végétaux. Cet organite n'existant pas chez la levure, nous ne savions pas si la protéine $CsZCD^{+N+P}$ pouvait être dirigée vers une structure cellulaire particulière ou non. Nous avons donc testé les configurations « présence » ($Z^{+N+P}/V185$) et « absence » ($Z^{+N}/V185$) du peptide signal. Une culture en milieu sélectif liquide a été réalisée, d'abord en présence de doxycycline de façon à ce que les cellules accumulent de la zéaxanthine sans qu'elle soit dégradée par les CCOs. La transcription des CCOs a ensuite été induite. Les caroténoïdes ont été extraits des différentes souches transformées, puis analysés par HPLC (Figure 48).

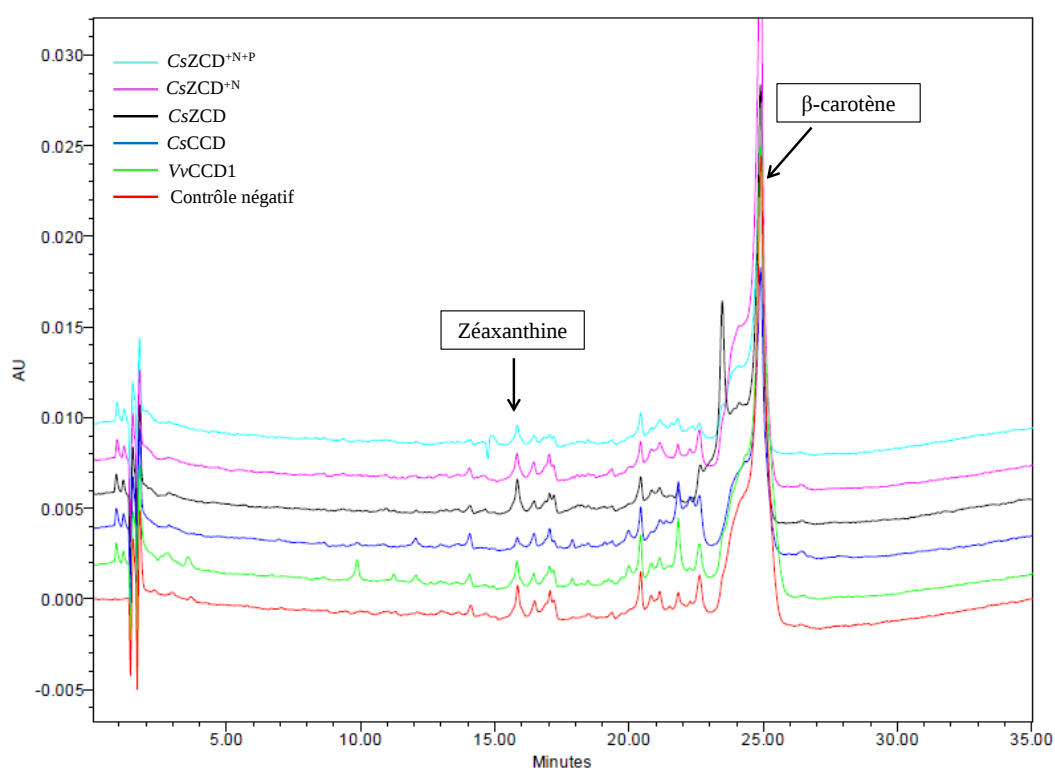


Figure 48: Analyse HPLC de caroténoïdes extraits de souches de levure YSG50 exprimant $CsZCD$, $CsZCD^{+N}$, $CsZCD^{+N+P}$ (spectres enregistrés à 440 nm), $CsCCD$ et $VvCCD1$ (spectres enregistrés à 416 nm). Les levures YSG50 ont été transformées avec les constructions : Bleu : $Z^{+N+P}/V185$; Rose : $Z^{+N}/V185$; Noir : $Z/V185$; Bleu : $C/V185$; Vert : $V/V185$. Rouge : pCM185 vide.

Aucun produit de clivage de la zéaxanthine n'est observé quelle que soit la condition. Les pics visibles entre 10 et 15 minutes sur les spectres correspondant à $VvCCD1$ et à $CsCCD$ possèdent

des spectres d'absorption qui ne correspondent pas à celui qui est attendu pour les dialdéhydes recherchés.

III. DISCUSSION

Les analyses qui ont composé ce Chapitre 2 avaient pour but initial d'étudier l'hypothèse selon laquelle l'enzyme CsZCD initie la formation des molécules responsables de l'arôme, de la couleur et du parfum du safran. Le mécanisme proposé est un clivage de la zéaxanthine en positions 7,8 et 7',8' donnant naissance à un dialdéhyde composé d'une chaîne à 16 carbones (Bouvier et al., 2003b). Différentes analyses, incluant des comparaisons de séquences et de structures *in silico*, des tests d'activité *in vivo* et *in vitro* ont été réalisées.

Aucun produit de clivage de la zéaxanthine n'a pu être observé dans nos conditions : expression de CsZCD chez *E. coli* suivie de tests d'activité *in vitro*, ou expression de CsZCD chez *S. cerevisiae* suivie de tests d'activité *in vivo* et *in vitro*. Nous pouvons expliquer ces résultats de différentes manières. Les conditions d'expression de la protéine que nous avons appliquée chez *E. coli* ne sont peut-être pas adaptées. Cependant, nous avons appliqué ces conditions à l'enzyme VvCCD1, caractérisées chez *Vitis vinifera* et ayant une activité très proche de CsZCD sur la zéaxanthine (Mathieu et al., 2005). Nous avons observé une activité de VvCCD1 dans tous nos essais, avec toutefois une variabilité de la quantité de produit formé d'un test à l'autre. Cela nous indique qu'il est possible d'observer un produit de clivage de la zéaxanthine dans les conditions d'expression et de tests *in vitro* que nous avons appliqué, mais que ces conditions ne sont pas optimales. Nous avons d'ailleurs pu voir que dans des conditions identiques d'expression, VvCCD1 était produite chez *E. coli*, mais pas CsCCD (Figure 46). Il est donc possible qu'une amélioration des conditions de production de l'enzyme permette d'observer son activité. L'expression de CsZCD sous une forme soluble est en effet problématique, et aucun des tests d'optimisation des conditions d'expression que nous avons effectué dans cette étude ne s'est montré efficace. VvCCD1 a la particularité d'avoir été exprimée chez *E. coli* sous la forme d'une protéine fusion avec la GST (Mathieu et al., 2005). L'utilisation de cette stratégie est peut-être une condition nécessaire pour augmenter la solubilité de CsZCD (et de CsCCD) et ainsi visualiser une activité.

Il est également possible que nous n'ayons pas testé le bon substrat. En effet, les produits chimiquement synthétisés qui sont utilisés en recherche en tant que substrats pour les tests enzymatiques sont parfois d'une pureté moyenne, et peuvent contenir des traces d'intermédiaires de synthèse par exemple. On peut imaginer que c'est ce type d'intermédiaire que CsZCD a pu

prendre pour substrat lors des tests d'activité réalisés par Bouvier et al. en 2003. Il serait donc intéressant de tester un éventail de substrats plus large pour déterminer la spécificité de CsZCD.

Nous sommes partis du principe qu'un repliement plus efficace de la protéine était possible en l'exprimant chez *S. cerevisiae*, organisme eucaryote pourvu d'une machinerie de traduction des protéines capable d'effectuer certaines modifications post-traductionnelles. Nos tests d'activité, *in vitro* et *in vivo* se sont avérés négatifs sur toutes les enzymes testées. Il est possible que ce résultat soit lié à un problème d'expression des protéines. Il faudrait donc exprimer à nouveau les protéines dans les mêmes conditions, cette fois pourvues d'une étiquette visible par immunodétection, et réaliser ensuite une analyse en western blot. De plus, un mauvais repliement peut toujours être la cause de l'absence d'activité. Chez la levure, la nature de la séquence d'adressage d'une protéine qui possède une interaction membranaire, et en particulier son hydrophobicité, est déterminante. Elle contient une information sur la voie de transport qu'elle va emprunter (Ng et al., 1996). Le transport de protéines membranaires chez *S. cerevisiae* peut être co-traductionnel ou post-traductionnel et, contrairement aux eucaryotes supérieurs, le mécanisme post-traductionnel est souvent préféré chez les organismes à croissance rapide tels que les bactéries, les archées, ou les levures (Rapoport, 2007). Nous pouvons imaginer que le repliement fonctionnel de CsZCD, à l'instar de VvCCD1 ou encore CsCCD, nécessite une prise en charge par le système de repliement co-traductionnel. Leurs séquences d'adressage en revanche, peuvent correspondre à l'inverse à un signal de transport post-traductionnel *S. cerevisiae*. Pour tester cette hypothèse, il serait envisageable d'exprimer ces enzymes chez une levure alternative comme *Yarrowia lipolitica*, connue pour utiliser un système de transport des protéines membranaires plus proche de celui des eucaryotes supérieurs (Boisrame et al., 1996; He et al., 1990). De plus, une ingénierie existe à l'heure actuelle pour produire des caroténoïdes chez *Y. lipolitica* (Matthaus et al., 2014), laissant entrevoir la possibilité de faire des tests d'activité *in vivo* en y produisant la zéaxanthine. On peut aussi les exprimer chez une plante modèle telle que le tabac (*Nicotiana tabacum*). Un test d'activité *in vitro* sur différents substrats après expression transitoire dans les feuilles pourrait apporter des réponses à propos de l'activité de CsZCD dans des conditions de production de l'enzyme plus proches du système naturel.

Même si toutes les hypothèses énumérées ci-dessus peuvent expliquer l'absence de produit de CsZCD et doivent être testées, les analyses *in silico* effectuées lors de cette étude apportent des

informations suggérant que de CsZCD est tronquée. La séquence primaire de CsZCD est plus courte que la grande majorité des séquences de CCOs disponibles. La seule CCO présentant une séquence primaire de taille comparable à celle de CsZCD a été caractérisée par la même équipe en 2003 également, chez l'organisme *Bixa orellana* : BoLCD (Bouvier et al., 2003a). Une telle similarité de séquence malgré leur éloignement phylogénétique et géographique, est très surprenante. Etant donné que ces deux enzymes ont été découvertes dans le même laboratoire à la même période, et connaissant les difficultés inhérentes à la manipulation très délicate des ARNs, nous pouvons imaginer qu'une contamination ait eu lieu entre ces échantillons d'organismes différents. Cela augmente malheureusement le niveau d'incertitude qui existe à propos des conclusions apportées par Bouvier et al. (2003b).

Une analyse de la structure tridimensionnelle de VP14 (structure 3D disponible présentant la structure primaire la plus proche de CsZCD) nous a permis d'identifier la position et la fonctionnalité potentielle des résidus manquants de CsZCD (environ 200 acides aminés N-terminaux). Il s'agit d'une part de plusieurs hélices α identifiées chez VP14 comme étant responsables de son interaction membranaire, et d'autre part d'un feuillet β participant à la formation du β -propeller à sept lames (structure en tonneau) renfermant le site actif de l'enzyme. Cela suggère fortement que l'absence d'éléments structuraux indispensables rende CsZCD non-fonctionnelle. Cependant, nous avons ajouté en N-terminal de CsZCD des résidus identifiés sur d'autres CCOs de *Crocus sativus* (Rubio et al., 2008) dans l'espoir de restaurer son activité, et cela n'a pas permis non plus d'observer un produit de dégradation de la zéaxanthine. Cela suggère que la seule hypothèse consistant à penser que CsZCD est une enzyme tronquée ne suffit pas à expliquer l'absence d'activité de CsZCD *in vitro*.

En résumé, nos analyses de génomique fonctionnelle et nos tests d'activité semblent indiquer que la structure primaire de CsZCD (Bouvier et al., 2003b) est erronée. Cela est en accord avec l'étude de Rubio et al. (2008), qui ne retrouvent pas la séquence de CsZCD dans leur banque d'ADNc de *Crocus sativus*, et qui concluent leur article en proposant que les résultats de Bouvier et al. (2003b) soient reconsidérés. Cependant, nous avons détecté le produit de clivage de la zéaxanthine par VvCCD1 (utilisée comme témoin positif dans nos tests d'activité) à des niveaux parfois très faibles, ce qui suggère qu'une optimisation des conditions d'expression et de test

d'activité est nécessaire afin d'être capables de faire la lumière sur la nature de l'enzyme responsable de la formation de l'arôme, du colorant et du parfum du safran.

Conclusion et Perspectives

Un projet d'ingénierie métabolique, au stade de la recherche fondamentale, peut se décomposer en plusieurs approches principales, de niveaux de complexité différents (Figure 49). Tout d'abord, il est nécessaire de vérifier la fonctionnalité, dans l'hôte, de chaque étape enzymatique à implémenter. Le second niveau consiste à réguler individuellement le niveau d'expression de chaque enzyme afin d'optimiser la consommation des flux métaboliques par chacune de ces étapes. Enfin, cette nouvelle voie de biosynthèse optimisée doit être intégrée à l'intérieur de l'ensemble des réactions métaboliques de l'hôte, et le flux de précurseur endogène doit y être optimisé sans impacter la viabilité cellulaire.

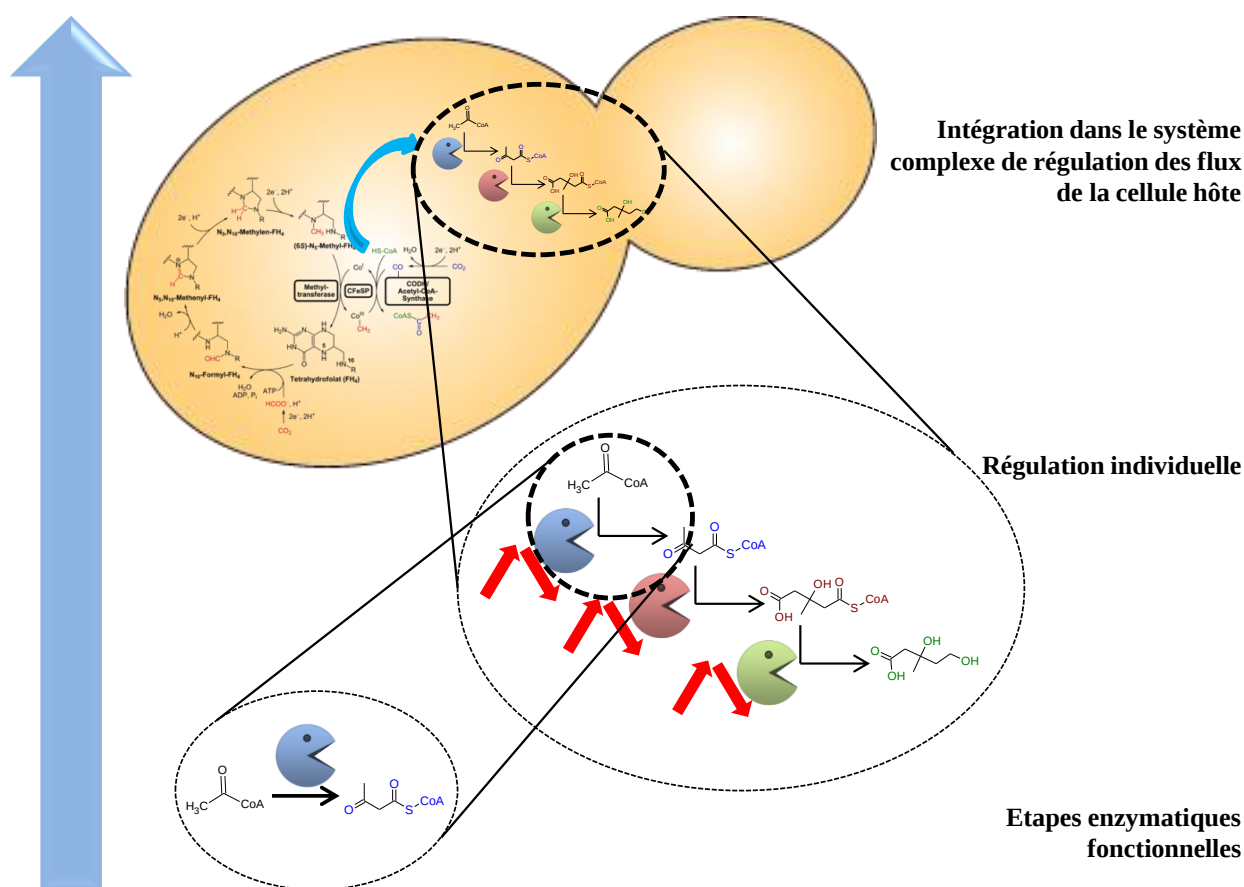


Figure 49: Imbrication des différentes approches d'un projet d'ingénierie métabolique, depuis l'obtention d'enzymes fonctionnelles, jusqu'à l'intégration de la voie métabolique entière dans les réseaux métaboliques de la cellule hôte. La flèche bleue représente l'évolution des niveaux de complexité.

Le but de ce travail de thèse était de reconstruire une voie métabolique complexe menant à la biosynthèse des arômes, colorants et parfums du safran. Nous avons choisi de scinder notre projet en deux parties : la reconstruction de la voie de biosynthèse de leur précurseur commun, la zéaxanthine, et l'analyse fonctionnelle d'un groupe d'enzymes candidates pour catalyser le clivage oxydatif de ce précurseur. La production de zéaxanthine chez *S. cerevisiae* avait déjà été effectuée en amont de ce projet (Shao et al., 2009; Sun et al., 2012). Cela constituait une bonne indication de la fonctionnalité de cette voie de biosynthèse hétérologue. Nous avons donc abordé cette partie du projet au niveau du second point de complexité décrit ci-dessus : la régulation des niveaux d'expression de chacune des étapes hétérologues. En revanche, la seconde partie du projet présentait une incertitude quant à la fonctionnalité de l'enzyme responsable de l'initiation de la synthèse des molécules d'intérêt (Rubio et al., 2008). C'est pourquoi nous avons approché cette partie du projet à un niveau d'avancement plus en amont, c'est-à-dire la recherche de la fonction désirée.

La production de zéaxanthine chez *S. cerevisiae* a été précédemment décrite en exprimant cinq gènes provenant de l'organisme *Erwinia uredovora* (Shao et al., 2009; Sun et al., 2012). Dans notre cas, nous avons reconstruit une voie de biosynthèse, la plus courte possible (trois gènes), entre le précurseur endogène (GGPP) et le précurseur des métabolites ciblés (zéaxanthine). En effet, dans la première partie du projet, nous avons observé une production de zéaxanthine en utilisant des gènes issus des organismes *Xanthophyllomyces dendrorhous* et *Pantoea ananatis*. L'utilisation de cette stratégie était inédite pour produire de la zéaxanthine, mais elle est classique en ingénierie métabolique pour diminuer l'impact énergétique de la voie à reconstruire (Hansen et al., 2009; Szczepara et al., 2003). Pour analyser les effets de la variation du niveau d'expression de ces trois gènes, nous avons construit une banque combinatoire de promoteurs de forces différentes.

La quantification des ARNm associée a révélé que la force des promoteurs n'impose pas de régulation individuelle dans notre système, mais que les niveaux d'expression de chaque gène sont fortement dépendants de l'ordre dans lequel les gènes sont placés sur le plasmide ainsi que de l'orientation de leur transcription. Nous avons relié ces deux facteurs critiques à une forte interférence transcriptionnelle d'un gène sur l'autre (Shearwin et al., 2005), ainsi qu'à une efficacité inégale des différents terminateurs (Curran et al., 2013). La nature de certaines séquences codantes semble également imposer une régulation de l'expression, probablement *via* la stabilité

des ARNm. Nos données mettent en lumière le fait que la force connue des promoteurs n'est pas forcément constante, mais dépend de nombreux facteurs. Faire varier la nature des promoteurs ne suffit pas, et les stratégies combinatoires en vue d'optimiser de nouvelles voies métaboliques doivent aussi faire varier d'autres paramètres comme l'espace entre les gènes ou l'orientation de leur transcription.

Ces données ont été obtenues en adoptant une stratégie d'expression plasmidique. La question qui se pose maintenant est de savoir si les biais et limitations identifiés grâce à cette stratégie interviennent également dans le cas d'une stratégie d'expression par intégration au génome de *S. cerevisiae*. Tout d'abord, nous avons détecté une compétition entre les deux plasmides présents dans une souche. Afin d'observer l'interférence transcriptionnelle existante sans ce biais supplémentaire, il serait intéressant de commencer par intégrer *tHMG1* au génome des différentes souches. Il a de plus déjà été remarqué que l'expression d'un même gène pouvait différer d'un système plasmidique à un système intégré (Verwaal et al., 2007). Dans un second temps, il est donc envisageable de tester la stratégie d'intégration chromosomique de tous les gènes implémentés. Lorsqu'une stratégie d'intégration est appliquée, les gènes sont souvent intégrés en bloc au même site dans le génome (Shao et al., 2009; Verwaal et al., 2007). En revanche, dans de récents travaux de production de β -carotène chez *S. cerevisiae* (Xie et al., 2014), l'intégration de tous les gènes a été effectuée de façon décentralisée, c'est-à-dire que les gènes ont été disséminés dans le génome. Cette stratégie atténue potentiellement les phénomènes d'interférence due à l'absence d'espace entre les différentes unités transcriptionnelles. Cette équipe a choisi d'utiliser un système de régulation existant chez la levure pour la voie hétérologue : tous les gènes impliqués dans la voie métabolique hétérologue ont été placés sous le contrôle du système d'induction au galactose / répression par le glucose (Johnston et al., 1994; Lohr et al., 1995). Le moment de l'induction a ensuite été optimisé de façon à ce que les ressources énergétiques des cellules soient entièrement utilisées pour la croissance en début de culture, et qu'elles soient ensuite entièrement dédiées à la production de caroténoïdes, permettant aux cellules de rester viables et de garder une croissance normale. Cela suggère qu'il est plus judicieux d'ajuster le moment de l'induction plutôt que les niveaux individuels de chaque gène exprimés constitutivement.

Les preuves de concept de la production de zéaxanthine chez *S. cerevisiae* apportées par l'équipe du Dr H. Zhao utilisaient les gènes de la voie de biosynthèse de la zéaxanthine de

l'organisme *Erwinia uredovora* (Shao et al., 2009; Sun et al., 2012). Il serait donc intéressant d'inclure ces gènes dans la combinatoire afin de déceler éventuellement des combinaisons de gènes plus efficaces en termes de production de métabolites, mais aussi identifier certains verrous dans leur régulation transcriptionnelle spécifique, qui pourraient ensuite devenir des cibles d'optimisation.

Suite aux observations qui ont été faites dans notre étude, la variation de la nature des terminateurs apparaît aussi comme une stratégie importante dans l'optimisation des flux métaboliques à l'intérieur d'une voie hétérologue. La construction d'une banque combinatoire faisant varier la nature des terminateurs fait donc également partie des perspectives que nos résultats nous permettent de mettre en évidence.

Enfin, dans le contexte de la régulation des flux métaboliques en biologie de synthèse, la régulation transcriptionnelle doit être accompagnée d'une approche protéomique. En effet, une régulation traductionnelle intervient entre les étapes de synthèse de l'ARNm et de synthèse de la protéine correspondante. La quantité d'enzyme active présente dans la cellule dépend finalement autant de la régulation transcriptionnelle que de la régulation traductionnelle étant intervenues en amont. Une quantification des protéines cibles du projet est donc nécessaire afin d'attribuer rigoureusement les effets dus à une régulation transcriptionnelle ou à une régulation traductionnelle sur la quantité de métabolites finale.

L'analyse fonctionnelle de CSZCD que nous avons effectué montre des lacunes dans la description de la nature de l'enzyme responsable de la formation des précurseurs des molécules d'intérêt du safran. Plusieurs enzymes candidates ont été testées, mais nous n'avons pas observé de formation des produits attendus. Prenant en considération le fait que CsZCD semble non fonctionnelle, différentes stratégies sont à présent à examiner afin de parvenir à la synthèse des molécules désirées : crocine, picrocrocine et safranal. Il est possible de réaliser une analyse des transcrits de la plante par RNAseq par exemple, afin de faire une recherche des enzymes de la famille des CCOs exprimées chez *Crocus sativus* en période de synthèse des molécules recherchées. Une telle analyse, potentiellement plus puissante que la collection d'ADNc, pourrait permettre d'identifier de nouveaux candidats pour l'activité recherchée.

Nous avons observé une activité de VvCCD1, qui clive la zéaxanthine en positions 9,10 et 9',10'. Grâce à l'analyse bibliographique et aux alignements (séquences et structures) qui ont été effectués au cours de ces travaux, nous sommes aujourd'hui capables de cibler les acides aminés responsables de la spécificité de l'enzyme. Par une stratégie de mutation dirigée de VvCCD1, il est envisageable de modifier sa spécificité de clivage et d'obtenir ainsi une CCO qui clive la zéaxanthine en positions 7,8 et 7',8'.

Les CCOs sont présentes dans un grand nombre d'organismes (microorganismes, plantes, mammifères), et certaines présentent des activités identiques ou proches de celle qui est recherchée. Il est concevable de faire une recherche d'enzymes candidates (présentant une activité similaire) parmi les CCOs connues et de construire une banque combinatoire de souches faisant varier la nature de la CCO exprimée, qui serait ensuite criblée pour l'activité de clivage recherchée.

Les travaux présentés peuvent fournir des outils pour adopter une stratégie plus fiable de reconstruction métabolique, notamment pour des voies complexes comme la transposition de la voie de biosynthèse des arômes, colorants et parfum du safran chez la levure. Le choix de scinder le projet en deux parties selon leurs états de l'art respectifs s'est avéré avantageux par rapport à une stratégie de reconstruction totale. En effet, une reconstruction totale ne nous aurait pas permis d'identifier précisément l'étape non fonctionnelle à laquelle attribuer l'absence de production de métabolites finaux. La reconstruction métabolique peut donc utiliser les principes de la biologie de synthèse et de l'ingénierie : considérer les bases présentes dans la littérature comme des briques élémentaires, à assembler par la suite en respectant les différents niveaux de reconstruction. Des constructions peuvent ensuite être menées en parallèle afin de trouver l'assemblage le plus efficace pour le caractère recherché.

Matériel et Méthodes

1. Analyses *in silico*.

Les recherches de séquences protéiques similaires proches ont été effectuées grâce à l'outil « blastp » de NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Pour réaliser des alignements de séquences nucléotidiques ou protéiques, deux outils en ligne ont été successivement utilisés : COBALT de NCBI (http://www.st-vi.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi), ou ClustalW (Larkin et al., 2007).

La recherche de séquences d'adressage au chloroplaste a été effectuée grâce à l'outil en ligne ChloroP (<http://www.cbs.dtu.dk/services/ChloroP/>).

La structure tridimensionnelle de VP14 a été visualisée et interprétée en utilisant le logiciel PyMOL.

2. Construction de souches d'expression.

2.1 Clonages.

Toutes les séquences codantes des protéines qui ont été utilisées dans cette étude sont présentées en Annexe 1. Les séquences CSZCD, CSZCD^{+N}, CSZCD^{+N+P} et CSCCD ont toutes fait l'objet d'un recodage selon le biais de codons de *S. cerevisiae* par le logiciel Gene Designer 2.0 (DNA 2.0), et d'une synthèse réalisée par GenCust (Luxembourg). Les séquences de toutes les amorces de PCR utilisées dans cette étude sont présentées en Annexe 2. Toutes les séquences clonées à la suite d'une amplification par PCR ou d'une restriction ont fait l'objet d'une purification sur gel d'agarose coloré en SYBR® Safe DNA gel stain (Life Technologies, Carlsbad, CA), grâce au « GeneJET Gel Extraction Kit » (Thermo Scientific, Waltham, MA). La méthode suivie pour chacun des clonages utile à cette étude est détaillée dans le Tableau 9, et les étapes ayant servi à obtenir les différentes enzymes CsZCD, CsZCD^{+N} et CsZCD^{+N+P} dans différents vecteurs sont représentées en Figure 50.

Tableau 9: Méthodes utilisées pour obtenir les différents plasmides d'expression de l'étude. Restriction / ligation : réalisé avec le « Rapid ligation kit » selon les recommandations du fournisseur (Thermo Scientific, Waltham, MA) ; In-Fusion® = Kit commercialisé par Clontech (Mountain View, CA) utilisé selon les recommandations du fournisseur ; GA = Gibson Assembly Cloning Kit (NEB, Ipswich) utilisé en suivant les recommandations du fournisseur. pCM185 : publié dans (Gari et al., 1997). CSZCD a fait l'objet de deux amplifications successives pour être clonée dans pBAD/TOPO® ThioFusion afin d'obtenir des séquences de recombinaison plus longues.

Séquence à cloner	Méthode de clonage	Amorces de PCR amplification 1	Amorces de PCR amplification 2	Vecteur	Construction finale obtenue
CS ZCD	Restriction / ligation Nde1 + Bpu11021			pET15b linéarisé Nde1 + Bpu11021	Z/p15
	In-Fusion®	Cs ZCD_pBAD_F1	Cs ZCD_pBAD_F2	pBAD/TOPO® ThioFusion amplifié par PCR à l'aide des amorces pBAD_F et pBAD_R	Z/pBAD
		Cs ZCD_pBAD_R1	Cs ZCD_pBAD_R2		
	In-Fusion®	Cs ZCD_V60_F		YeDP60 linéarisé BamH1 + EcoR1	Z/V60
		Cs ZCD_V60_R			
	GA	Cs ZCD_pCM185_F		pCM185 linéarisé BamH1	Z/V185
		Cs ZCD_pCM185_R			
N-TER (b)	GA	N-ter(b)_pET15b_F		Z/p15 digéré Nde1	Z ^{+N} /p15
		N-ter(b)_pET15b_R			
PS(a)	GA	Ps(a)_pET15b_F		Z ^{+N} /p15 amplifié par PCR à l'aide des amorces Cs ZCD+N/pET15b_F et CsZCD+N/pET15b_R	Z ^{+N+P} /p15
		Ps(a)_pET15b_R			
CS CCD	Restriction / ligation Nde1 + Bpu11021	Cs CCD_pET15b_F		pET15b linéarisé Nde1 + Bpu11021	C/p15
		Cs CCD_pET15b_R			
	GA	Cs CCD_pCM185_F		pCM185 linéarisé BamH1	C/V185
		Cs CCD_pCM185_R			
CS ZCD^{+N}	GA	Cs ZCD ^{+N} _pCM185_F		pCM185 linéarisé BamH1	Z ^{+N} /V185
		Cs ZCD ^{+N} _pCM185_R			
CS ZCD^{+N+P}	GA	Cs ZCD ^{+N+P} _pCM185_F		pCM185 linéarisé BamH1	Z ^{+N+P} /V185
		Cs ZCD ^{+N+P} _pCM185_R			
Vv CCD1	GA	Vv CCD1_pCM185_F		pCM185 linéarisé BamH1	V/V185
		Vv CCD1_pCM185_R			

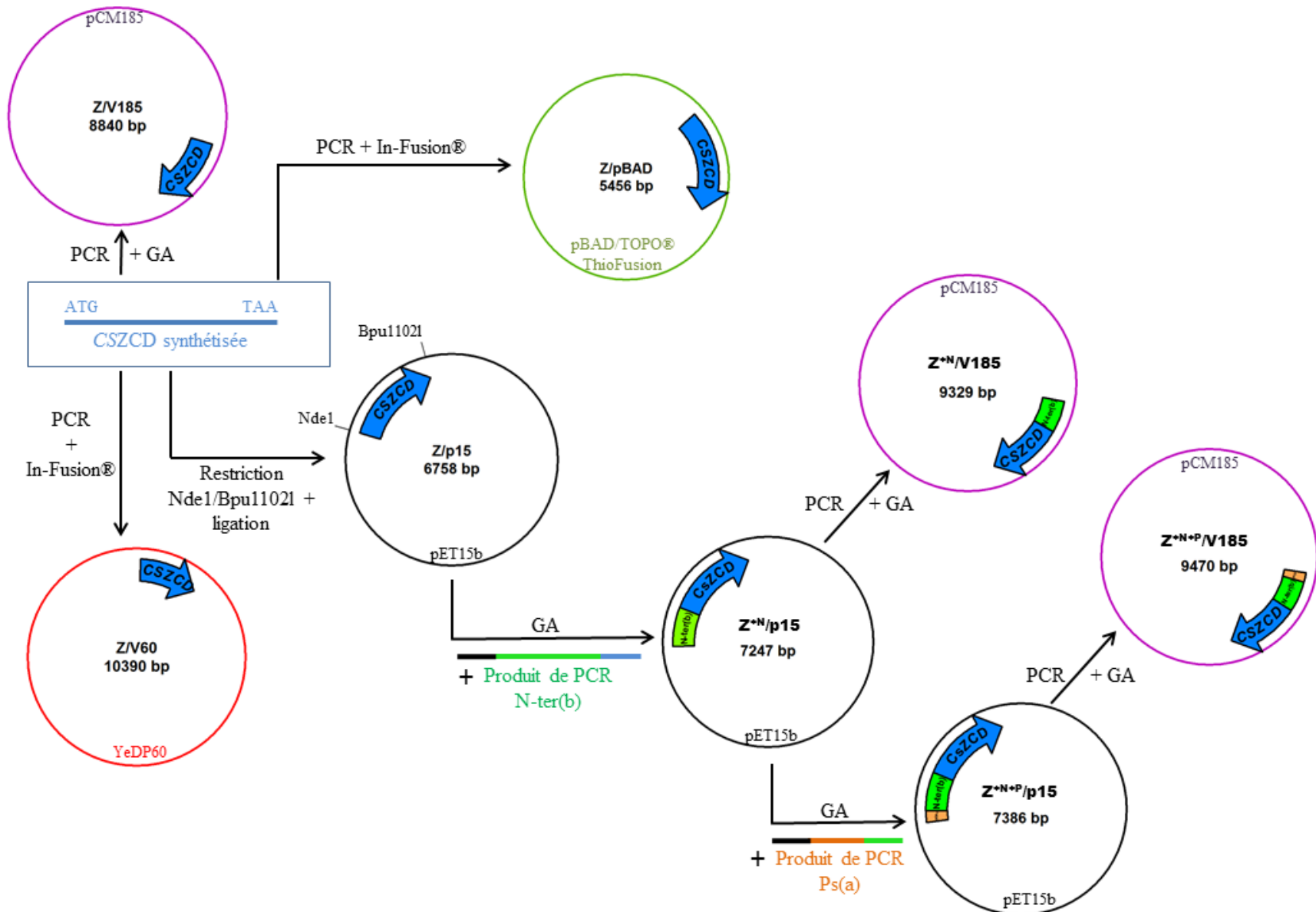


Figure 50: Etapes mises en œuvre pour construire pour cloner les séquences CSZCD, CSZCD^N ou CSZCD^{N+P} dans les vecteurs pET15b, pBAD/TOPO® ThioFusion, YeDP60 ou pCM185. Restriction / ligation : réalisé avec le « Rapid ligation kit » selon les recommandations du fournisseur (Thermo Scientific, Waltham, MA) ; In-Fusion® = Kit commercialisé par Clontech (Mountain View, CA) utilisé selon les recommandations du fournisseur ; GA = Gibson Assembly Cloning Kit (NEB, Ipswich) utilisé en suivant les recommandations du fournisseur. PCM 185 publié dans (Gari et al., 1997).

2.2 Transformation de souches.

Les clonages présentés précédemment ont été obtenus dans *E. coli* DH5a. Toutes les constructions utilisées ont fait l'objet d'une vérification par séquençage (GATC Biotech, Constance, Allemagne).

Les souches d'expression de *E. coli* (BL21(De3), C41(De3), Origami(De3) et TOP10) étaient disponibles au laboratoire, et ont été transformées après avoir été rendues compétentes par la méthode au chlorure de calcium. Une culture de 100 mL était faite en milieu LB pendant environ 3h, puis arrêtée sur la glace. Les cellules étaient récupérées par centrifugation à 2050 g à 4°C pendant 10 minutes, et reprises dans 10 mL de CaCl₂ 0.1 M froid. Après une nouvelle centrifugation à 2050 g à 4°C pendant 10 minutes, les cellules étaient reprises dans 2 mL de CaCl₂ 0.1M froid. A cette étape, les cellules étaient conservées à 4°C pendant une nuit. Le lendemain, 200 µL de cellules étaient transformés avec environ 50 ng d'ADN par une incubation sur la glace de 30 minutes, puis un choc thermique à 42°C pendant 90 secondes, et enfin un retour sur la glace de 2 minutes. 800 µL de milieu LB était ajoutés, puis les cellules étaient incubées 1 heure à 37°C avant d'être étalées sur un milieu sélectif LBA gélosé.

Les souches de levures ont été transformées par la méthode à l'acétate de lithium suivi d'un choc thermique (Gietz & Akio, 1988).

3. Expression de protéines.

Les différentes conditions d'expression des protéines chez *E. coli* sont rassemblées dans Tableau 10.

Tableau 10: Conditions d'expression des protéines utilisées dans l'étude. Génotype de la souche CEN.PK 113-7D : (Mata his3—1 leu2-3,112 trp1-289 ura3-52 MAL2-8c). La souche YSG50 nous a été fournie par le Dr Huimin Zhao (Urbana, Illinois).

	Souche	Milieu de culture	Température avant induction	Induction	Température / temps de culture après induction
Validation de l'expression	BL21(De3)	LBA	37°C	1mM IPTG à DO ₆₀₀ 0,5	37°C / 3h
Optimisation des conditions de culture	BL21(De3)	ZYM 5052	20°C pendant 24 heures		
Optimisation de la souche d'expression	BL21(De3); C41(De3); Origami (De3)	ZYM 5052	20°C pendant 24 heures		
Bouvier et al. 2003b	TOP10	LBA	20°C	0,02% arabinose (w/v) à DO ₆₀₀ 0,5	20°C / une nuit
Mathieu et al. 2005	BL21(De3)	2xYT	37°C	125 µM IPTG à DO ₆₀₀ 1,6	18°C / une nuit
Expression chez <i>S. cerevisiae</i>	CEN.PK 113-7D	YPGE	30°C	2% galactose (w/v) au bout de 24h de culture	30°C / 24h
	YSG50	CSM -URA-TRP + doxycycline 2µg/mL	30°C	Milieu changé au bout de 4 jours pour du CSM-URA-TRP sans doxycycline	30°C / 48h

4. SDS-Page

Les cellules étaient cassées par sonication 3 x 30 secondes (sonicateur VibraCell 72434 BIOBLOCK SCIENTIFIC, 40% de puissance) dans un tampon Tris-HCl pH7,6 100 mM. Les fractions solubles et insolubles étaient séparées par une centrifugation à 15 500 g à 4°C. 7.5 µL de chaque surnageant et de chaque culot (dilué au 1/20^e) étaient prélevés et mélangés à 2.5µL de LDS Buffer (Life Technologies, Carlsbad, CA). L'ensemble était chauffé à 70°C pendant 10 minutes, puis déposé sur gel de polyacrylamide (Novex NuPAGE 4-12% Bis-Tris, Life Technologies, Carlsbad, CA). La migration était effectuée en tampon MES pendant 35 minutes à 200V.

5. Tests d'activité.

5.1 Tests d'activité *in vitro* après expression de protéines chez *E. coli*.

Les cellules étaient cassées par sonication 3 x 30 secondes (sonicateur VibraCell 72434 BIOBLOCK SCIENTIFIC, 40% de puissance) dans différents tampons selon les conditions appliquées. La nature des tampons et les conditions de tests d'activité *in vitro* sont détaillées dans le Tableau 11.

Tableau 11: Composition des tests d'activité *in vitro* réalisés après expression enzymatique chez *E. coli*.

	Bouvier et al. 2003b	Mathieu et al. 2005	Mathieu et al. 2005 + optimisation	Mathieu et al. 2007
Tampon de cassage des cellules (5 mL final pour 50 mL de culture)	Tris-HCl pH 7,6 50 mM	PBS + DTT 4mM + TRITON X-100 1%	PBS + DTT 4mM + TRITON X-100 1%	PBS + DTT 4mM + TRITON X-100 1%
Volume final (µL)	200	1500	375	375
Tampon	Tris-HCl pH 7,6 50 mM	Tris-HCl pH 7,6 100 mM	KCl 0,1M H ₃ BO ₃ 0,1M pH9	Tris-HCl pH 7,6 100 mM
Substrat (µM)	100	35	35	35
DTT (mM)	1	1	1	1
FeSO ₄ (µM)	5	5	5	5
Octyl-β-glucoside (%; w/v)	0,2			
TRITON X-100 (%; v/v)		1	1	0,11
Acétone (%; v/v)			10	10
Surnageant de cassage (µL)	21	500	21	21
Température / temps d'incubation	30°C/1h	30°C / 3h	30°C / 1h	30°C / 1h
Arrêt de la réaction	200 µl acetone + 300µL dichloromethane / methanol (1/1; v/v)	4,5 mL de pentane / dichloromethane (2/1; v/v)	4,5 mL de pentane / dichloromethane (2/1; v/v)	4,5 mL de pentane / dichloromethane (2/1; v/v)

5.2 Tests d'activité *in vitro* après expression de protéine chez *S. cerevisiae*.

5.2.1 Préparation des fractions microsomales.

Les cellules étaient centrifugées à 6300 g, lavées dans 25 mL d'eau osmosée et à nouveau centrifugées à 6300 g. Le culot cellulaire était repris dans 1 à 2 mL de tampon TES (TRIS-HCl pH 7.4 50 mM ; EDTA 1mM ; Sorbitol 0.6 M). A partir de cette étape, les tubes étaient manipulés sur la glace. Des billes de verre (0.5 mm de diamètre) étaient ensuite ajoutées aux culots de cellules. Les cellules étaient ensuite cassées manuellement en effectuant des mouvements amples de haut en bas, régulièrement et avec force, 4 fois pendant 2 minutes 30 avec refroidissements sur la glace entre chaque cycle de cassage. 5 mL de tampon TES étaient ajoutés. Après décantation, le surnageant était prélevé et transvasé dans un autre tube. Ces deux dernières étapes étaient répétées jusqu'à ce que le surnageant devienne limpide. Le « fond de tube restant » était centrifugé à 15 550 g pendant 15 minutes, et le culot séparé du surnageant. Le culot contient ici les débris cellulaires, les membranes plasmiques et les mitochondries. 0,4 mL de NaCl 5 M, puis 2,6 mL de PEG 4000 50% étaient ajoutés pour 10 mL de surnageant. Un précipité correspondant aux microsomes apparaissait au bout de 20 minutes d'incubation sur la glace. Une centrifugation de 10 minutes à 12 850 g permettait de récupérer d'un côté les microsomes (culot) et de l'autre le cytosol des cellules (surnageant). Les microsomes étaient repris dans 1,5 mL de tampon TEG (TRIS-HCl pH 7.4 50 mM ; EDTA 1mM ; Glycérol 20% (v/v)). 50 µL de chacune des fractions récupérées ont été utilisés pour réaliser un test d'activité (volume final 200 µl) en présence de zéaxanthine encapsulées dans des liposomes (composition en Tableau 12).

5.2.2 Préparation des liposomes.

10 mg de phosphatidylcholine (lécithine d'œuf, SIGMA) étaient solubilisés dans 1 mL de chloroforme, auquel étaient ajoutés 200 µL de zéaxanthine 1 mM dans le chloroforme. Le mélange était séché sous gaz neutre (azote ou argon), puis sous vide (dessicateur en verre relié à une pompe à vide) pendant environ 2 heures. 2 mL de tampon (TRIS-HCl pH 7.4 ; NaCl 100 mM ; FeSO₄ 5µM ; Asorbate 5 mM) étaient ensuite ajoutés, et le mélange était incubé environ 30 minutes sur

la glace. Le mélange était ensuite soniqué (Vibra-Cell 72434, 70% de puissance) jusqu'à ce qu'il devienne limpide. La solution était ensuite passée au LipoFast™ (Avestin ; MacDonald et al. 1991) sur membrane de 50 nm (20 passages). 150 µL de liposomes ainsi calibrés étaient ajoutés aux 50 µL de fraction de cassage de levures pour réaliser le test d'activité.

Tableau 12: Composition du test d'activité *in vitro* réalisé sur les différentes fractions de levures récupérées après cassage manuel. Volume final : 200 µL

Réactifs	Concentration finale
Zéaxanthine dans les liposomes	400 µM
TRIS-HCl pH 7,4	50 mM
NaCl	100 mM
FeSO ₄	5 µM
Ascorbate	5 mM

5.3 Tests d'activité *in vivo* après expression de protéine chez *S. cerevisiae*

Des cultures étaient faites en milieu sélectif et en présence de doxycycline (Sigma, 2µg/mL) pendant 4 jours, puis le milieu était changé pour un milieu sans doxycycline, et la culture était relancée pendant 48 heures. Après un lavage à l'eau osmosée, les cellules étaient congelées une nuit à -80°C, puis lyophilisées avec un lyophilisateur CHRIST Alpha 2-4 LD plus. 2 g de billes de verre étaient ensuite ajoutés aux cellules sèches. Les culots de cellules secs étaient vortexés avec les billes pendant 2 minutes pour séparer les cellules agglomérées, puis 5 mL d'acétone étaient ajoutés. La suspension de cellules dans l'acétone était passée au vortex 3 x 2 minutes avec des intervalles de refroidissement sur la glace entre chaque cycle. Les cellules cassées étaient ensuite incubées pendant 3 heures dans l'acétone, sur la glace et à l'abri de la lumière. Après une étape de centrifugation à 8 200 g pendant 2 minutes à 4°C, les surnageants étaient conservés dans la glace. Les débris cellulaires et les billes de verres étaient lavés 3 fois avec 5 mL d'acétone. Les surnageants de chaque lavage étaient poolés avec le premier. Les extraits étaient ensuite séchés sous un flux de gaz neutre (azote ou argon), et repris dans 1 mL d'isopropanol avant injection à l'HPLC pour analyse des produits.

6. Analyse de métabolites.

Les produits des tests d'activité ainsi que les caroténoïdes extraits des souches de *S. cerevisiae* exprimant la voie de biosynthèse de la zéaxanthine ont été analysés comme décrit ici. 10 µL de chaque échantillon étaient injectés dans une colonne Agilent ZORBAX SB-C18, et analysés avec un module HPLC Waters 2690 couplé à un détecteur UV-visible. Le gradient d'élution était le suivant, avec un flux de 1 mL/min : 0–3 min, 50% (v,v) H₂O-25% isopropanol-25% acetonitrile (MeCN); 3–20 min, gradient linéaire à partir de 50% H₂O-25% isopropanol-25% MeCN pour arriver à 10% H₂O-70% isopropanol-20% MeCN; 20–25 min, 10% H₂O-70% isopropanol-20% MeCN; 25–35 min, gradient linéaire à partir de 10% H₂O-70% isopropanol-20% MeCN pour arriver à 50% H₂O-25% isopropanol-25% MeCN. La température de la colonne était fixée à 60°C. La détection des produits se faisait en sortie de colonne à 416 nm pour les produits de VvCCD1 et CsCCD, à 440 nm pour les produits de CsZCD, CsZCD^{+N} et CsZCD^{+N+P}, à 450 nm pour les caroténoïdes extraits de levures. Les standards analytiques (β-carotène, zéaxanthine et astaxanthine) ont été achetés chez Extrasynthèse France (Lyon).

7. Perte de plasmide et test cross.

La perte des plasmides multi-copies (dérivés de pRS426) a été effectuée en mettant les souches sur un milieu minimum gélosé (MP Biomedicals, Santa Ana, CA) contenant de l'acide 5-fluoroorotic 1g/L (Sigma-Aldrich, St-Louis, MO).

Des souches de signes sexuels opposés ont été croisées sur milieu YPD gélosé (Annexe 3). Une fois les diploïdes formés (vérifié sous microscope), leur sporulation a été provoquée sur un milieu gélosé d'acétate de potassium 1%. Après une incubation d'environ deux jours à 30°C, les tétrades ont été disséquées à l'aide d'un micromanipulateur (SINGER instruments MSM system).

Références Bibliographiques

- Abdullaev, F.I. 2002. Cancer chemopreventive and tumoricidal properties of saffron (*Crocus sativus* L.). *Exp Biol Med (Maywood)*, **227**(1), 20-5.
- Ahrazem, O., Rubio-Moraga, A., Lopez, R.C., Gomez-Gomez, L. 2010. The expression of a chromoplast-specific lycopene beta cyclase gene is involved in the high production of saffron's apocarotenoid precursors. *J Exp Bot*, **61**(1), 105-19.
- Ajikumar, P.K., Tyo, K., Carlsen, S., Mucha, O., Phon, T.H., Stephanopoulos, G. 2008. Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol Pharm*, **5**(2), 167-90.
- Ajikumar, P.K., Xiao, W.H., Tyo, K.E., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B., Stephanopoulos, G. 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science*, **330**(6000), 70-4.
- Akhondzadeh, S., Shafiee Sabet, M., Harirchian, M.H., Togha, M., Cheraghmakani, H., Razeghi, S., Hejazi, S.S., Yousefi, M.H., Alimardani, R., Jamshidi, A., Rezazadeh, S.A., Yousefi, A., Zare, F., Moradi, A., Vossoughi, A. 2010. A 22-week, multicenter, randomized, double-blind controlled trial of *Crocus sativus* in the treatment of mild-to-moderate Alzheimer's disease. *Psychopharmacology (Berl)*, **207**(4), 637-43.
- Alcaino, J., Barahona, S., Carmona, M., Lozano, C., Marcoleta, A., Niklitschek, M., Sepulveda, D., Baeza, M., Cifuentes, V. 2008. Cloning of the cytochrome p450 reductase (*crtR*) gene and its involvement in the astaxanthin biosynthesis of *Xanthophyllomyces dendrorhous*. *BMC Microbiol*, **8**, 169.
- Alper, H., Moxley, J., Nevoigt, E., Fink, G.R., Stephanopoulos, G. 2006. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science*, **314**(5805), 1565-8.
- Alper, H., Stephanopoulos, G. 2007. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng*, **9**(3), 258-67.
- Andrewes, A.G., Herman, J.P., Mortimer, P.S. 1976. Carotenoids of *Phaffia rhodozyma*, a red-pigmented fermenting yeast. *Phytochemistry*, **15**(6), 1003-7.
- Aslanidis, C., de Jong, P.J. 1990. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res*, **18**(20), 6069-74.
- Auldrige, M.E., McCarty, D.R., Klee, H.J. 2006. Plant carotenoid cleavage oxygenases and their apocarotenoid products. *Curr Opin Plant Biol*, **9**(3), 315-21.

- Ausich, R.L., Brinkhaus, F.L., Mukharji, I., Proffitt, J.H., Yarger, J.G., Yen, H.C.B. 1997. Biosynthesis of zeaxanthin and glycosylated zeaxanthin in genetically engineered hosts, Google Patents.
- Babiskin, A.H., Smolke, C.D. 2011a. A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. *Mol Syst Biol*, **7**, 471.
- Babiskin, A.H., Smolke, C.D. 2011b. Synthetic RNA modules for fine-tuning gene expression levels in yeast by modulating RNase III activity. *Nucleic Acids Res*, **39**(19), 8651-64.
- Bahieldin, A., Gadalla, N.O., Al-Garni, S.M., Almehdar, H., Noor, S., Hassan, S.M., Shokry, A.M., Sabir, J.S., Murata, N. 2014. Efficient production of lycopene in *Saccharomyces cerevisiae* by expression of synthetic crt genes from a plasmid harboring the ADH2 promoter. *Plasmid*, **72**, 18-28.
- Bailey, J.E. 1991. Toward a science of metabolic engineering. *Science*, **252**(5013), 1668-75.
- Bajbouj, K., Schulze-Luehrmann, J., Diermeier, S., Amin, A., Schneider-Stock, R. 2012. The anticancer effect of saffron in two p53 isogenic colorectal cancer cell lines. *BMC Complement Altern Med*, **12**, 69.
- Bathaie, S.Z., Bolhasani, A., Hoshyar, R., Ranjbar, B., Sabouni, F., Moosavi-Movahedi, A.A. 2007. Interaction of saffron carotenoids as anticancer compounds with ctDNA, Oligo (dG.dC)₁₅, and Oligo (dA.dT)₁₅. *DNA Cell Biol*, **26**(8), 533-40.
- Beekwilder, J., van Rossum, H.M., Koopman, F., Sonntag, F., Buchhaupt, M., Schrader, J., Hall, R.D., Bosch, D., Pronk, J.T., van Maris, A.J., Daran, J.M. 2014. Polycistronic expression of a beta-carotene biosynthetic pathway in *Saccharomyces cerevisiae* coupled to beta-ionone production. *J Biotechnol*, **192 Pt B**, 383-92.
- Beekwilder, J., Wolswinkel, R., Jonker, H., Hall, R., de Vos, C.H., Bovy, A. 2006. Production of resveratrol in recombinant microorganisms. *Appl Environ Microbiol*, **72**(8), 5670-2.
- Blazeck, J., Alper, H.S. 2013. Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol J*, **8**(1), 46-58.
- Boisrame, A., Beckerich, J.M., Gaillardin, C. 1996. Sls1p, an Endoplasmic Reticulum Component, Is Involved in the Protein Translocation Process in the Yeast *Yarrowia lipolytica*. *Journal of Biological Chemistry*, **271**(20), 11668-11675.
- Bouvier, F., Dogbo, O., Camara, B. 2003a. Biosynthesis of the food and cosmetic plant pigment bixin (annatto). *Science*, **300**(5628), 2089-91.
- Bouvier, F., Suire, C., Mutterer, J., Camara, B. 2003b. Oxidative remodeling of chromoplast carotenoids: identification of the carotenoid dioxygenase CsCCD and CsZCD genes involved in *Crocus* secondary metabolite biogenesis. *Plant Cell*, **15**(1), 47-62.
- Bud, R. 2001. History of biotechnology. *eLS*, 1-6.

- Camara, B., Bouvier, F. 2004. Oxidative remodeling of plastid carotenoids. *Arch Biochem Biophys*, **430**(1), 16-21.
- Cartwright, C., Li, Y., Zhu, Y., Kang, Y., Tipper, D. 1994. Use of beta-lactamase as a secreted reporter of promoter function in yeast. *Yeast*, **10**, 497-508.
- Castillo, R., Fernandez, J.A., Gomez-Gomez, L. 2005. Implications of carotenoid biosynthetic genes in apocarotenoid formation during the stigma development of *Crocus sativus* and its closer relatives. *Plant Physiol*, **139**(2), 674-89.
- Celik, E., Calik, P. 2012. Production of recombinant proteins by yeast cells. *Biotechnol Adv*, **30**(5), 1108-18.
- Chang, A., Wolf, J., Smolke, C. 2012. Synthetic RNA switches as a tool for temporal and spacial control over gene expression. *Current opinion in biotechnology*, **23**, 679-688.
- Chen, X., Jiang, Z.H., Chen, S.t., Qin, W. 2010. Microbial and bioconversion production of D-xylitol and its detection and application. *Int J Biol Sci*, **6**(7), 834-44.
- Chen, Y., Partow, S., Scalcinati, G., Siewers, V., Nielsen, J. 2012. Enhancing the copy number of episomal plasmids in *Saccharomyces cerevisiae* for improved protein production. *FEMS Yeast Res*, **12**(5), 598-607.
- Cho, K., Yoo, Y., Kang, H. 1999. Delta-Integration of endo/exo-glucanase and beta-glucosidase genes into the yeast chromosomes for direct conversion of cellulose to ethanol. *Enzyme Microb Technol*, **25**, 23-30.
- Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., Voytas, D.F. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**(2), 757-61.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**(6121), 819-23.
- Curran, K.A., Karim, A.S., Gupta, A., Alper, H.S. 2013. Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab Eng*, **19**, 88-97.
- Da Silva, N.A., Srikrishnan, S. 2012. Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS Yeast Res*, **12**(2), 197-214.
- Dahl, R.H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T.S., Redding-Johanson, A.M., Petzold, C.J., Mukhopadhyay, A., Lee, T.S., Adams, P.D., Keasling, J.D. 2013. Engineering dynamic pathway regulation using stress-response promoters. *Nat Biotechnol*, **31**(11), 1039-46.
- Daum, G., Lees, N.D., Bard, M., Dickson, R. 1998. Biochemistry, cell biology and molecular biology of lipids of *Saccharomyces cerevisiae*. *Yeast*, **14**(16), 1471-510.

- de Vos, M.G., Poelwijk, F.J., Tans, S.J. 2013. Optimality in evolution: new insights from synthetic biology. *Curr Opin Biotechnol*, **24**(4), 797-802.
- DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., Church, G.M. 2013. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res*, **41**(7), 4336-43.
- Ding, Q., Zhong, H., Qi, Y., Cheng, Y., Li, W., Yan, S., Wang, X. 2013. Anti-arthritic effects of crocin in interleukin-1 β -treated articular chondrocytes and cartilage in a rabbit osteoarthritic model. *Inflamm Res*, **62**(1), 17-25.
- Donald, K.A., Hampton, R.Y., Fritz, I.B. 1997. Effects of overproduction of the catalytic domain of 3-hydroxy-3-methylglutaryl coenzyme A reductase on squalene synthesis in *Saccharomyces cerevisiae*. *Appl Environ Microbiol*, **63**(9), 3341-4.
- Du, J., Yuan, Y., Si, T., Lian, J., Zhao, H. 2012. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res*, **40**(18), e142.
- Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A.V., Prather, K.L., Keasling, J.D. 2009. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol*, **27**(8), 753-9.
- Erhart, E., Hollenberg, C.P. 1983. The presence of a defective LEU2 gene on 2 μ DNA recombinant plasmids of *Saccharomyces cerevisiae* is responsible for curing and high copy number. *J Bacteriol*, **156**(2), 625-35.
- Fang, F., Salmon, K., Shen, M.W., Aeling, K.A., Ito, E., Irwin, B., Tran, U.P., Hatfield, G.W., Da Silva, N.A., Sandmeyer, S. 2011. A vector set for systematic metabolic engineering in *Saccharomyces cerevisiae*. *Yeast*, **28**(2), 123-36.
- Faulkner, J.D., Anson, J.G., Tuite, M.F., Minton, N.P. 1994. High-level expression of the phenylalanine ammonia lyase-encoding gene from *Rhodospiridium toruloides* in *Saccharomyces cerevisiae* and *Escherichia coli* using a bifunctional expression system. *Gene*, **143**(1), 13-20.
- Fernandez-Sanchez, L., Lax, P., Esquivia, G., Martin-Nieto, J., Pinilla, I., Cuenca, N. 2012. Safranin, a saffron constituent, attenuates retinal degeneration in P23H rats. *PLoS ONE*, **7**(8), e43074.
- Ferrer-Miralles, N., Domingo-Espin, J., Corchero, J.L., Vazquez, E., Villaverde, A. 2009. Microbial factories for recombinant pharmaceuticals. *Microb Cell Fact*, **8**, 17.
- Forster, J., Famili, I., Fu, P., Palsson, B.O., Nielsen, J. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*, **13**(2), 244-53.
- Garcia-Rubio, M.L., Aguilera, A. 2012. Topological constraints impair RNA polymerase II transcription and causes instability of plasmid-borne convergent genes. *Nucleic Acids Res*, **40**(3), 1050-64.

- Gari, E., Piedrafita, L., Aldea, M., Herrero, E. 1997. A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast*, **13**(9), 837-48.
- Gibson, D.G. 2009. Synthesis of DNA Fragments in Yeast by One-Step Assembly of Overlapping Oligonucleotides. *Nucleic Acids Res*, **37**, 6984-6990.
- Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O., Hutchison, C.A. 2008. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc Natl Acad Sci U S A*, **105**(51), 20404-20409.
- Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E.A., Young, L., Qi, Z.Q., Segall-Shapiro, T.H., Calvey, C.H., Parmar, P.P., Hutchison, C.A., Smith, H.O., Venter, J.C. 2010. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, **329**(5987), 52-56.
- Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**(5), 343-345.
- Gietz, R.D., Akio, S. 1988. New yeast-*Escherichia coli* shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. *Gene*, **74**(2), 527-534.
- Glick, B. 1995. Metabolic load and heterologous gene expression. *Biotechnology adv*, **13**, 247-261.
- Goffeau, A., Barrell, B., Bussey, H., Davis, R. 1996. Life with 6000 genes. *Science*, **274**, 546-567.
- Gopal, G.J., Kumar, A. 2013. Strategies for the production of recombinant protein in *Escherichia coli*. *Protein J*, **32**(6), 419-25.
- Gourdon, P., Lindley, N.D. 1999. Metabolic analysis of glutamate production by *Corynebacterium glutamicum*. *Metab Eng*, **1**(3), 224-31.
- Grilly, C., Stricker, J., Pang, W.L., Bennett, M.R., Hasty, J. 2007. A synthetic gene network for tuning protein degradation in *Saccharomyces cerevisiae*. *Mol Syst Biol*, **3**, 127.
- Gustafsson, C., Govindarajan, S., Minshull, J. 2004. Codon bias and heterologous protein expression. *Trends Biotechnol*, **22**(7), 346-53.
- Hackett, E.A., Esch, R.K., Maleri, S., Errede, B. 2006. A family of destabilized cyan fluorescent proteins as transcriptional reporters in *S. cerevisiae*. *Yeast*, **23**(5), 333-49.
- Hahn-Hagerdal, B., Karhumaa, K., Larsson, C.U., Gorwa-Grauslund, M., Gorgens, J., van Zyl, W.H. 2005. Role of cultivation media in the development of yeast strains for large scale industrial use. *Microb Cell Fact*, **4**, 31.

- Hama, S., Yoshida, A., Nakashima, K., Noda, H., Fukuda, H., Kondo, A. 2010. Surfactant-modified yeast whole-cell biocatalyst displaying lipase on cell surface for enzymatic production of structured lipids in organic media. *Appl Microbiol Biotechnol*, **87**(2), 537-43.
- Hansen, E.H., Moller, B.L., Kock, G.R., Bunner, C.M., Kristensen, C., Jensen, O.R., Okkels, F.T., Olsen, C.E., Motawia, M.S., Hansen, J. 2009. De novo biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). *Appl Environ Microbiol*, **75**(9), 2765-74.
- He, F., Yaver, D., Beckerich, J.M., Ogrydziak, D., Gaillardin, C. 1990. The yeast *Yarrowia lipolytica* has two, functional, signal recognition particle 7S RNA genes. *Current Genetics*, **17**, 289-292.
- Holtz, W.J., Keasling, J.D. 2010. Engineering static and dynamic control of synthetic pathways. *Cell*, **140**(1), 19-23.
- Hong, K.K., Nielsen, J. 2012. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell Mol Life Sci*, **69**(16), 2671-90.
- Hong, S.H., Kim, J.S., Lee, S.Y., In, Y.H., Choi, S.S., Rih, J.K., Kim, C.H., Jeong, H., Hur, C.G., Kim, J.J. 2004. The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*. *Nat Biotechnol*, **22**(10), 1275-81.
- Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., Pease, L.R. 1989. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, **77**, 61-68.
- Hou, L. 2009. Novel methods of genome shuffling in *Saccharomyces cerevisiae*. *Biotechnol Lett*, **31**(5), 671-7.
- Huang, C.J., Lowe, A.J., Batt, C.A. 2010. Recombinant immunotherapeutics: current state and perspectives regarding the feasibility and market. *Appl Microbiol Biotechnol*, **87**(2), 401-10.
- Irwin, C., Farmer, A., Willer, D., Evans, D. 2012. In-fusion® cloning with vaccinia virus DNA polymerase. *Methods Mol Biol*, **890**, 23-35.
- Jaswir, I., Noviendri, D., Hasrini, R.F., Octavianti, F. 2011. Carotenoids: Sources, medicinal properties and their application in food and nutraceutical industry. *Journal of Medicinal Plants Research*, **5**, 7119-7131.
- Jeppsson, M., Johansson, B., Jensen, P.R., Hahn-Hagerdal, B., Gorwa-Grauslund, M.F. 2003. The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant *Saccharomyces cerevisiae* strains. *Yeast*, **20**(15), 1263-72.
- Jin, Y.S., Jeffries, T.W. 2004. Stoichiometric network constraints on xylose metabolism by recombinant *Saccharomyces cerevisiae*. *Metab Eng*, **6**(3), 229-38.

- Johnston, M., Flick, J.S., Pexton, T. 1994. Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **14**(6), 3834-41.
- Kandavelou, K., Mani, M., Durai, S., Chandrasegaran, S. 2005. ‘Magic’ scissors for genome surgery. *Nature*, **435**, 646-51.
- Karaffa, L., Kubicek, C.P. 2003. *Aspergillus niger* citric acid accumulation: do we understand this well working black box? *Appl Microbiol Biotechnol*, **61**(3), 189-96.
- Keasling, J.D. 2012. Synthetic biology and the development of tools for metabolic engineering. *Metab Eng*, **14**(3), 189-95.
- Kim, B., Du, J., Eriksen, D.T., Zhao, H. 2013. Combinatorial design of a highly efficient xylose-utilizing pathway in *Saccharomyces cerevisiae* for the production of cellulosic biofuels. *Appl Environ Microbiol*, **79**(3), 931-41.
- Kim, I.K., Roldao, A., Siewers, V., Nielsen, J. 2012. A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Res*, **12**(2), 228-48.
- Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A., Voytas, D.F. 1998. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res*, **8**(5), 464-78.
- Kim, S., Hahn, J.S. 2014. Synthetic scaffold based on a cohesin-dockerin interaction for improved production of 2,3-butanediol in *Saccharomyces cerevisiae*. *J Biotechnol*, **192PA**, 192-196.
- Kimura, K., Tomita, N., Uemura, H., Aki, T., Ono, K., Kamisaka, Y. 2009. Improvement of Stearidonic acid production in Oleaginous *Saccharomyces cerevisiae*. *Biosci Biotechnol Biochem*, **73**(6), 1447-9.
- Kiser, P.D., Golczak, M., Lodowski, D.T., Chance, M.R., Palczewski, K. 2009. Crystal structure of native RPE65, the retinoid isomerase of the visual cycle. *Proc Natl Acad Sci U S A*, **106**(41), 17325-30.
- Klein, J., Heal, J.R., Hamilton, W.D., Boussemghoune, T., Tange, T.O., Delegrange, F., Jaeschke, G., Hatsch, A., Heim, J. 2014. Yeast synthetic biology platform generates novel chemical structures as scaffolds for drug discovery. *ACS Synth Biol*, **3**(5), 314-23.
- Kloer, D.P., Ruch, S., Al-Babili, S., Beyer, P., Schulz, G.E. 2005. The structure of a retinal-forming carotenoid oxygenase. *Science*, **308**(5719), 267-9.
- Kloer, D.P., Schulz, G.E. 2006. Structural and biological aspects of carotenoid cleavage. *Cell Mol Life Sci*, **63**(19-20), 2291-303.
- Kocharin, K., Chen, Y., Siewers, V., Nielsen, J. 2012. Engineering of acetyl-CoA metabolism for the improved production of polyhydroxybutyrate in *Saccharomyces cerevisiae*. *AMB Express*, **2**(1), 52.

- Kouprina, N., Larionov, V. 2008. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat Protoc*, **3**(3), 371-7.
- Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res*, **9**(20), 5233-5252.
- Kuijpers, N.G., Chroumpi, S., Vos, T., Solis-Escalante, D., Bosman, L., Pronk, J.T., Daran, J.M., Daran-Lapujade, P. 2013. One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*. *FEMS Yeast Res*, **13**(8), 769-81.
- Lage, M., Cantrell, C.L. 2009. Quantification of saffron (*Crocus sativus* L.) metabolites crocins, picrocrocin and safranal for quality determination of the spice grown under different environmental Moroccan conditions. *Scientia Horticulturae*, **121**(3), 366-373.
- Lan, E.I., Liao, J.C. 2013. Microbial synthesis of n-butanol, isobutanol, and other higher alcohols from diverse resources. *Bioresour Technol*, **135**, 339-49.
- Lange, N., Steinbuchel, A. 2011. beta-Carotene production by *Saccharomyces cerevisiae* with regard to plasmid stability and culture media. *Appl Microbiol Biotechnol*, **91**(6), 1611-22.
- Larionov, V., Kouprina, N. 1996. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc Natl Acad Sci U S A*, **93**(1), 491-496.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**(21), 2947-8.
- Lashbrooke, J.G., Young, P.R., Dockrall, S.J., Vasanth, K., Vivier, M.A. 2013. Functional characterisation of three members of the *Vitis vinifera* L. carotenoid cleavage dioxygenase gene family. *BMC Plant Biol*, **13**, 156.
- Latimer, L.N., Lee, M.E., Medina-Cleghorn, D., Kohnz, R.A., Nomura, D.K., Dueber, J.E. 2014. Employing a combinatorial expression approach to characterize xylose utilization in *Saccharomyces cerevisiae*. *Metab Eng*, **25**, 20-9.
- Lecuc, S. 1912. *La Biologie Synthétique*, Paris.
- Lee, K.S., Kim, J.S., Heo, P., Yang, T.J., Sung, Y.J., Cheon, Y., Koo, H.M., Yu, B.J., Seo, J.H., Jin, Y.S., Park, J.C., Kweon, D.H. 2013a. Characterization of *Saccharomyces cerevisiae* promoters for heterologous gene expression in *Kluyveromyces marxianus*. *Appl Microbiol Biotechnol*, **97**(5), 2029-41.
- Lee, M.E., Aswani, A., Han, A.S., Tomlin, C.J., Dueber, J.E. 2013b. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res*, **41**(22), 10668-78.
- Lee, P.C., Schmidt-Dannert, C. 2002. Metabolic engineering towards biotechnological production of carotenoids in microorganisms. *Appl Microbiol Biotechnol*, **60**(1-2), 1-11.

- Lee, S.J., Lee, D.W. 2013. Design and development of synthetic microbial platform cells for bioenergy. *Front Microbiol*, **4**, 92.
- Lee, W., Dasilva, N.A. 2006. Application of sequential integration for metabolic engineering of 1,2-propanediol production in yeast. *Metab Eng*, **8**(1), 58-65.
- Li, Q., Sun, Z., Li, J., Zhang, Y. 2013. Enhancing beta-carotene production in *Saccharomyces cerevisiae* by metabolic engineering. *FEMS Microbiol Lett*, **345**(2), 94-101.
- Liang, J., Ning, J.C., Zhao, H. 2013. Coordinated induction of multi-gene pathways in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, **41**(4), e54.
- Linardaki, Z.I., Orkoulas, M.G., Kokkosis, A.G., Lamari, F.N., Margaritis, M. 2013. Investigation of the neuroprotective action of saffron (*Crocus sativus* L.) in aluminum-exposed adult mice through behavioral and neurobiochemical assessment. *Food Chem Toxicol*, **52**, 163-70.
- Liu, D., Hoynes-O'Connor, A., Zhang, F. 2013a. Bridging the gap between systems biology and synthetic biology. *Front Microbiol*, **4**, 211.
- Liu, L., Redden, H., Alper, H.S. 2013b. Frontiers of yeast metabolic engineering: diversifying beyond ethanol and *Saccharomyces*. *Curr Opin Biotechnol*, **24**(6), 1023-30.
- Lohr, D., Venkov, P., Zlatanova, J. 1995. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. *FASEB J*, **9**(9), 777-87.
- Loison, G., Vidal, A., Findeli, A., Roitsch, C., Balloul, J.M., Lemoine, Y. 1989. High level of expression of a protective antigen of schistosomes in *Saccharomyces cerevisiae*. *Yeast*, **5**(6), 497-507.
- Lozano, P., Delgado, D., Gomez, D., Rubio, M., Iborra, J.L. 2000. A non-destructive method to determine the safranal content of saffron (*Crocus sativus* L.) by supercritical carbon dioxide extraction combined with high-performance liquid chromatography and gas chromatography. *J Biochem Biophys Methods*, **43**(1-3), 367-78.
- Lu, C., Jeffries, T. 2007. Shuffling of promoters for multiple genes to optimize xylose fermentation in an engineered *Saccharomyces cerevisiae* strain. *Appl Environ Microbiol*, **73**(19), 6072-7.
- Lu, C., Mansoorabadi, K., Jeffries, T. 2007. Comparison of multiple gene assembly methods for metabolic engineering. *Appl Biochem Biotechnol*, **137-140**(1-12), 703-10.
- Lund, B.A., Leiros, H.K., Bjerga, G.E. 2014. A high-throughput, restriction-free cloning and screening strategy based on ccdB-gene replacement. *Microb Cell Fact*, **13**(1), 38.
- Luque, A., Sebai, S.C., Santiago-Schubel, B., Le Coz, Y., Jenot, D., Ramaen, O., Sauveplane, V., Pandjaitan, R. 2014. In vivo evolution of metabolic pathways by homeologous recombination in mitotic cells. *Metab Eng*, **23**, 123-35.

- Mateus, C., Avery, S.V. 2000. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast*, **16**(14), 1313-23.
- Mathieu, S., Bigey, F., Procureur, J., Terrier, N., Gunata, Z. 2007. Production of a recombinant carotenoid cleavage dioxygenase from grape and enzyme assay in water-miscible organic solvents. *Biotechnol Lett*, **29**(5), 837-41.
- Mathieu, S., Terrier, N., Procureur, J., Bigey, F., Gunata, Z. 2005. A carotenoid cleavage dioxygenase from *Vitis vinifera* L.: functional characterization and expression during grape berry development in relation to C13-norisoprenoid accumulation. *J Exp Bot*, **56**(420), 2721-31.
- Matsumoto, T., Fukuda, H., Ueda, M., Tanaka, A. 2002. Construction of Yeast Strains with High Cell Surface Lipase Activity by Using Novel Display Systems Based on the Flo1p Flocculation Functional Domain *Appl Environ Microbiol*, **69**(9), 4517-4522.
- Matsumoto, T., Takahashi, S., Kaieda, M., Ueda, M., Tanaka, A., Fukuda, H., Kondo, A. 2001. Yeast whole-cell biocatalyst constructed by intracellular overproduction of *Rhizopus oryzae* lipase is applicable to biodiesel fuel production. *Appl Microbiol Biotechnol*, **57**(4), 515-520.
- Matsushika, A., Inoue, H., Kodaki, T., Sawayama, S. 2009. Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives. *Appl Microbiol Biotechnol*, **84**(1), 37-53.
- Mattanovich, D., Sauer, M., Gasser, B. 2014. Yeast biotechnology: teaching the old dog new tricks. *Microb Cell Fact*, **13**(1), 34.
- Mattea, F., Martín, Á., Cocero, M.J. 2009. Carotenoid processing with supercritical fluids. *J Food Eng*, **93**(3), 255-265.
- Matthaus, F., Ketelhot, M., Gatter, M., Barth, G. 2014. Production of lycopene in the non-carotenoid-producing yeast *Yarrowia lipolytica*. *Appl Environ Microbiol*, **80**(5), 1660-9.
- Maury, J., Asadollahi, M.A., Moller, K., Schalk, M., Clark, A., Formenti, L.R., Nielsen, J. 2008. Reconstruction of a bacterial isoprenoid biosynthetic pathway in *Saccharomyces cerevisiae*. *FEBS Lett*, **582**(29), 4032-8.
- Messing, S.A., Gabelli, S.B., Echeverria, I., Vogel, J.T., Guan, J.C., Tan, B.C., Klee, H.J., McCarty, D.R., Amzel, L.M. 2010. Structural insights into maize viviparous14, a key enzyme in the biosynthesis of the phytohormone abscisic acid. *Plant Cell*, **22**(9), 2970-80.
- Moraga, A.R., Nohales, P.F., Perez, J.A., Gomez-Gomez, L. 2004. Glucosylation of the saffron apocarotenoid crocetin by a glucosyltransferase isolated from *Crocus sativus* stigmas. *Planta*, **219**(6), 955-66.
- Mumberg, D., Muller, R., Funk, M. 1995. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, **156**(1), 119-22.

- Murphy, K.F., Balazsi, G., Collins, J.J. 2007. Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci U S A*, **104**(31), 12726-31.
- Murray, A., Szostak, J. 1983. Construction of artificial chromosomes in yeast. *Nature*, **305**, 189-193.
- Naesby, M., Nielsen, S.V., Nielsen, C.A., Green, T., Tange, T.O., Simon, E., Knechtle, P., Hansson, A., Schwab, M.S., Titiz, O., Folly, C., Archila, R.E., Maver, M., van Sint Fiet, S., Boussemghoune, T., Janes, M., Kumar, A.S., Sonkar, S.P., Mitra, P.P., Benjamin, V.A., Korrapati, N., Suman, I., Hansen, E.H., Thybo, T., Goldsmith, N., Sorensen, A.S. 2009. Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in *Saccharomyces cerevisiae*. *Microb Cell Fact*, **8**, 45.
- Neubauer, P., Lin, H.Y., Mathisizik, B. 2003. Metabolic load of recombinant protein production: inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in *Escherichia coli*. *Biotechnol Bioeng*, **83**(1), 53-64.
- Nevoigt, E. 2008. Progress in metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev*, **72**(3), 379-412.
- Nevoigt, E., Fischer, C., Mucha, O., Matthaus, F., Stahl, U., Stephanopoulos, G. 2007. Engineering promoter regulation. *Biotechnol Bioeng*, **96**(3), 550-8.
- Nevoigt, E., Kohnke, J., Fischer, C.R., Alper, H., Stahl, U., Stephanopoulos, G. 2006. Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*. *Appl Environ Microbiol*, **72**(8), 5266-73.
- Ng, D.T., Brown, J.D., Walter, P. 1996. Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J Cell Biol*, **134**(2), 269-78.
- Nielsen, J., Larsson, C., van Maris, A., Pronk, J. 2013. Metabolic engineering of yeast for production of fuels and chemicals. *Curr Opin Biotechnol*, **24**(3), 398-404.
- Niklitschek, M., Alcaíno, J., Barahona, S., Sepúlveda, D., Lozano, C., Carmona, M., Marcoleta, A., Martínez, C., Lodato, P., Baeza, M., Cifuentes, V. 2008. Genomic organization of the structural genes controlling the astaxanthin biosynthesis pathway of *Xanthophyllomyces dendrorhous*. *Biol Res*, **41**, 93-108.
- Obón, C., Rivera, D. 2006. Plant Pigments and their manipulation. *Economic Botany*, **60**(1), 92-92.
- Ohto, C., Muramatsu, M., Obata, S., Sakuradani, E., Shimizu, S. 2010. Production of geranylgeraniol on overexpression of a prenyl diphosphate synthase fusion gene in *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol*, **87**(4), 1327-34.
- Okada, T., Nakai, M., Maeda, H., Hosokawa, M., Sashima, T., Miyashita, K. 2008. Suppressive effect of neoxanthin on the differentiation of 3T3-L1 adipose cells. *J Oleo Sci*, **57**(6), 345-51.

- Olson, J.A., Krinsky, N.I. 1995. Introduction: the colorful, fascinating world of the carotenoids: important physiologic modulators. *FASEB J*, **9**(15), 1547-50.
- Omura, F., Fujita, A., Miyajima, K., Fukui, N. 2005. Engineering of yeast Put4 permease and its application to lager yeast for efficient proline assimilation. *Biosci Biotechnol Biochem*, **69**(6), 1162-71.
- Ostergaard, S., Olsson, L., Nielsen, J. 2000. Metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev*, **64**(1), 34-50.
- Otero, J.M., Nielsen, J. 2010. Industrial systems biology. *Biotechnol Bioeng*, **105**(3), 439-60.
- Ozaydin, B., Burd, H., Lee, T.S., Keasling, J.D. 2013. Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production. *Metab Eng*, **15**, 174-83.
- Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D.R., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K.W., Fickes, S., Galazzo, J., Gaucher, S.P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L.F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P.S., Keasling, J.D., Reiling, K.K., Renninger, N.S., Newman, J.D. 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, **496**(7446), 528-32.
- Partow, S., Siewers, V., Bjorn, S., Nielsen, J., Maury, J. 2010. Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast*, **27**(11), 955-64.
- Pfander, H., Schurtenberger, H. 1982. Biosynthesis of C20-carotenoids in *Crocus sativus*. *Phytochemistry*, **21**(5), 1039-1042.
- Pinel, D., D'Aoust, F., del Cardayre, S.B., Bajwa, P.K., Lee, H., Martin, V.J. 2011. *Saccharomyces cerevisiae* genome shuffling through recursive population mating leads to improved tolerance to spent sulfite liquor. *Appl Environ Microbiol*, **77**(14), 4736-43.
- Pscheidt, B., Glieder, A. 2008. Yeast cell factories for fine chemical and API production. *Microb Cell Fact*, **7**, 25.
- Purnick, P.E., Weiss, R. 2009. The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol*, **10**(6), 410-22.
- Ramon, A., Smith, H.O. 2011. Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. *Biotechnol Lett*, **33**(3), 549-55.
- Rao, Z., Ma, Z., Shen, W., Fang, H., Zhuge, J., Wang, X. 2008. Engineered *Saccharomyces cerevisiae* that produces 1,3-propanediol from D-glucose. *J Appl Microbiol*, **105**(6), 1768-76.

- Rapoport, T.A. 2007. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature*, **450**(7170), 663-9.
- Reyes, L.H., Gomez, J.M., Kao, K.C. 2014. Improving carotenoids production in yeast via adaptive laboratory evolution. *Metab Eng*, **21**, 26-33.
- Ro, D.K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C., Withers, S.T., Shiba, Y., Sarpong, R., Keasling, J.D. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, **440**(7086), 940-3.
- Rodriguez-Bustamante, E., Sanchez, S. 2007. Microbial production of C13-norisoprenoids and other aroma compounds via carotenoid cleavage. *Crit Rev Microbiol*, **33**(3), 211-30.
- Romanos, M.A., Scorer, C.A., Clare, J.J. 1992. Foreign gene expression in yeast: a review. *Yeast*, **8**(6), 423-88.
- Rubio, A., Rambla, J.L., Santaella, M., Gómez, M.D., Orzaez, D., Granell, A., Gómez-Gómez, L. 2008. Cytosolic and Plastoglobule-targeted Carotenoid Dioxygenases from *Crocus sativus* Are Both Involved in β -Ionone Release. *J Biol Chem*, **283**, 24816-24825.
- Rugbjerg, P., Naesby, M., Mortensen, U.H., Frandsen, R.J. 2013. Reconstruction of the biosynthetic pathway for the core fungal polyketide scaffold rubrofusarin in *Saccharomyces cerevisiae*. *Microb Cell Fact*, **12**, 31.
- Sakai, A., Shimizu, Y., Hishinuma, F. 1990. Integration of heterologous genes into the chromosome of *Saccharomyces cerevisiae* using a delta sequence of yeast retrotransposon Ty. *Appl Microbiol Biotechnol*, **33**(3), 302-306.
- Schmidt, H., Kurtzer, R., Eisenreich, W., Schwab, W. 2006. The carotenase AtCCD1 from *Arabidopsis thaliana* is a dioxygenase. *J Biol Chem*, **281**(15), 9845-51.
- Schreuder, M.P., Mooren, A.T., Toschka, H.Y., Verrips, C.T., Klis, F.M. 1996. Immobilizing proteins on the surface of yeast cells. *Trends Biotechnol*, **14**(4), 115-20.
- Schwartz, S.H., Tan, B.C., McCarty, D.R., Welch, W., Zeevaart, J.A. 2003. Substrate specificity and kinetics for VP14, a carotenoid cleavage dioxygenase in the ABA biosynthetic pathway. *Biochim Biophys Acta*, **1619**(1), 9-14.
- Shao, Z., Zhao, H., Zhao, H. 2009. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res*, **37**(2), e16.
- Shearwin, K.E., Callen, B.P., Egan, J.B. 2005. Transcriptional interference--a crash course. *Trends Genet*, **21**(6), 339-45.
- Shi, D.J., Wang, C.L., Wang, K.M. 2009. Genome shuffling to improve thermotolerance, ethanol tolerance and ethanol productivity of *Saccharomyces cerevisiae*. *J Ind Microbiol Biotechnol*, **36**(1), 139-47.

- Shiomi, N., Fukuda, H., Murata, K., Kimura, A. 1995. Improvement of S-adenosylmethionine production by integration of the ethionine-resistance gene into chromosomes of the yeast *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol*, **42**(5), 730-3.
- Siddiqui, M.S., Thodey, K., Trenchard, I., Smolke, C.D. 2012. Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res*, **12**(2), 144-70.
- Spadiut, O., Capone, S., Krainer, F., Glieder, A., Herwig, C. 2014. Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol*, **32**(1), 54-60.
- Stahl, W., Sies, H. 2005. Bioactivity and protective effects of natural carotenoids. *Biochim Biophys Acta*, **1740**(2), 101-7.
- Steen, E.J., Chan, R., Prasad, N., Myers, S., Petzold, C.J., Redding, A., Ouellet, M., Keasling, J.D. 2008. Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. *Microb Cell Fact*, **7**, 36.
- Stephanopoulos, G. 2007. Challenges in engineering microbes for biofuels production. *Science*, **315**(5813), 801-4.
- Stephanopoulos, G. 2002. Metabolic engineering by genome shuffling. *Nat Biotechnol*, **20**(7), 666-8.
- Studier, F.W. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*, **41**(1), 207-34.
- Sui, X., Kiser, P.D., Lintig, J., Palczewski, K. 2013. Structural basis of carotenoid cleavage: from bacteria to mammals. *Arch Biochem Biophys*, **539**(2), 203-13.
- Sun, J., Shao, Z., Zhao, H., Nair, N., Wen, F., Xu, J.H. 2012. Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in *Saccharomyces cerevisiae*. *Biotechnol Bioeng*, **109**(8), 2082-92.
- Szczebara, F.M., Chandelier, C., Villeret, C., Masurel, A., Bourot, S., Duport, C., Blanchard, S., Groisillier, A., Testet, E., Costaglioli, P., Cauet, G., Degryse, E., Balbuena, D., Winter, J., Achstetter, T., Spagnoli, R., Pompon, D., Dumas, B. 2003. Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nat Biotechnol*, **21**(2), 143-9.
- Tang, W.L., Zhao, H. 2009. Industrial biotechnology: tools and applications. *Biotechnol J*, **4**(12), 1725-39.
- Tanino, T., Aoki, T., Chung, W.Y., Watanabe, Y., Ogino, C., Fukuda, H., Kondo, A. 2009. Improvement of a *Candida antarctica* lipase B-displaying yeast whole-cell biocatalyst and its application to the polyester synthesis reaction. *Appl Microbiol Biotechnol*, **82**(1), 59-66.
- Tasaki, T., Sriram, S.M., Park, K.S., Kwon, Y.T. 2012. The N-end rule pathway. *Annu Rev Biochem*, **81**, 261-89.

- Teodoro, A.J., Oliveira, F.L., Martins, N.B., Maia Gde, A., Martucci, R.B., Borojevic, R. 2012. Effect of lycopene on cell viability and cell cycle progression in human cancer cell lines. *Cancer Cell Int*, **12**(1), 36.
- Toivari, M.H., Ruohonen, L., Miasnikov, A.N., Richard, P., Penttila, M. 2007. Metabolic engineering of *Saccharomyces cerevisiae* for conversion of D-glucose to xylitol and other five-carbon sugars and sugar alcohols. *Appl Environ Microbiol*, **73**(17), 5471-6.
- Ukibe, K., Hashida, K., Yoshida, N., Takagi, H. 2009. Metabolic engineering of *Saccharomyces cerevisiae* for astaxanthin production and oxidative stress tolerance. *Appl Environ Microbiol*, **75**(22), 7205-11.
- Van Vleet, J.H., Jeffries, T.W. 2009. Yeast metabolic engineering for hemicellulosic ethanol production. *Curr Opin Biotechnol*, **20**(3), 300-6.
- Verwaal, R., Wang, J., Meijnen, J.P., Visser, H., Sandmann, G., van den Berg, J.A., van Ooyen, A.J. 2007. High-level production of beta-carotene in *Saccharomyces cerevisiae* by successive transformation with carotenogenic genes from *Xanthophyllomyces dendrorhous*. *Appl Environ Microbiol*, **73**(13), 4342-50.
- Wakamori, M., Umehara, T., Yokoyama, S. 2010. A series of bacterial co-expression vectors with rare-cutter recognition sequences. *Protein Expr Purif*, **74**(1), 88-98.
- Walter, M.H., Strack, D. 2011. Carotenoids and their cleavage products: biosynthesis and functions. *Nat Prod Rep*, **28**(4), 663-92.
- Wang, T., Ma, X., Zhu, H., Li, A., Du, G., Chen, J. 2012. Available methods for assembling expression cassettes for synthetic biology. *Appl Microbiol Biotechnol*, **93**(5), 1853-63.
- Werner, S.R., Morgan, J.A. 2009. Expression of a *Dianthus* flavonoid glucosyltransferase in *Saccharomyces cerevisiae* for whole-cell biocatalysis. *J Biotechnol*, **142**(3-4), 233-41.
- William Bunch, A. 2014. How Biotechnology Helped Maintain the Supply of Acetone for the Manufacture of Cordite During World War I. *The International Journal for the History of Engineering & Technology*, **84**(2), 211-226.
- Winterhalter, P., Rouseff, R. 2001. Carotenoid-Derived Aroma Compounds: An Introduction. in: *Carotenoid-Derived Aroma Compounds*, Vol. 802, American Chemical Society, pp. 1-17.
- Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., Chu, A.M., Connelly, C., Davis, K., Dietrich, F., Dow, S.W., El Bakkoury, M., Foury, F., Friend, S.H., Gentalen, E., Giaever, G., Hegemann, J.H., Jones, T., Laub, M., Liao, H., Liebundguth, N., Lockhart, D.J., Lucau-Danila, A., Lussier, M., M'Rabet, N., Menard, P., Mittmann, M., Pai, C., Rebischung, C., Revuelta, J.L., Riles, L., Roberts, C.J., Ross-MacDonald, P., Scherens, B., Snyder, M., Sookhai-Mahadeo, S., Storms, R.K., Veronneau, S., Voet, M., Volckaert, G., Ward, T.R., Wysocki, R., Yen, G.S., Yu, K., Zimmermann, K., Philippsen, P., Johnston, M., Davis, R.W. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, **285**(5429), 901-6.

- Wisselink, H.W., Toirkens, M.J., del Rosario Franco Berriel, M., Winkler, A.A., van Dijken, J.P., Pronk, J.T., van Maris, A.J. 2007. Engineering of *Saccharomyces cerevisiae* for efficient anaerobic alcoholic fermentation of L-arabinose. *Appl Environ Microbiol*, **73**(15), 4881-91.
- Wittrup, K., Robinson, A., Parekh, R., Forrester, K. 1994. Existence of an Optimum Expression Level for Secretion of Foreign Proteins in Yeast. *Ann N Y Acad Sci.*, **745**, 321-330.
- Woolston, B.M., Edgar, S., Stephanopoulos, G. 2013. Metabolic engineering: past and future. *Annu Rev Chem Biomol Eng*, **4**, 259-88.
- Wriessnegger, T., Pichler, H. 2013. Yeast metabolic engineering--targeting sterol metabolism and terpenoid formation. *Prog Lipid Res*, **52**(3), 277-93.
- Xie, W., Liu, M., Lv, X., Lu, W., Gu, J., Yu, H. 2014. Construction of a controllable beta-carotene biosynthetic pathway by decentralized assembly strategy in *Saccharomyces cerevisiae*. *Biotechnol Bioeng*, **111**(1), 125-33.
- Xu, P., Bhan, N., Koffas, M.A. 2013. Engineering plant metabolism into microbes: from systems biology to synthetic biology. *Curr Opin Biotechnol*, **24**(2), 291-9.
- Yadav, V.G., De Mey, M., Lim, C.G., Ajikumar, P.K., Stephanopoulos, G. 2012. The future of metabolic engineering and synthetic biology: towards a systematic practice. *Metab Eng*, **14**(3), 233-41.
- Yamada, R., Taniguchi, N., Tanaka, T., Ogino, C., Fukuda, H., Kondo, A. 2010. Cocktail delta-integration: a novel method to construct cellulolytic enzyme expression ratio-optimized yeast strains. *Microb Cell Fact*, **9**, 32.
- Yamanaka, K. 1999. Cold shock response in *Escherichia coli*. *J Mol Microbiol Biotechnol*, **1**(2), 193-202.
- Yamano, S., Ishii, T., Nakagawa, M., Ikenaga, H., Misawa, N. 1994. Metabolic Engineering for Production of B-Carotene and Lycopene in *Saccharomyces*. *Biosci Biotechnol Biochem*, **58**, 1112-1114.
- Yan, Y., Kohli, A., Koffas, M.A. 2005. Biosynthesis of natural flavanones in *Saccharomyces cerevisiae*. *Appl Environ Microbiol*, **71**(9), 5610-3.
- Yazawa, H., Iwahashi, H., Kamisaka, Y., Kimura, K., Aki, T., Ono, K., Uemura, H. 2007. Heterologous production of dihomogamma-linolenic acid in *Saccharomyces cerevisiae*. *Appl Environ Microbiol*, **73**(21), 6965-71.
- Yofe, I., Zafir, Z., Blau, R., Schuldiner, M., Tuller, T., Shapiro, E., Ben-Yehzekel, T. 2014. Accurate, model-based tuning of synthetic gene expression using introns in *S. cerevisiae*. *PLoS Genet*, **10**(6), e1004407.
- Zaldivar, J., Nielsen, J., Olsson, L. 2001. Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Appl Microbiol Biotechnol*, **56**(1-2), 17-34.

- Zelle, R.M., de Hulster, E., van Winden, W.A., de Waard, P., Dijkema, C., Winkler, A.A., Geertman, J.M., van Dijken, J.P., Pronk, J.T., van Maris, A.J. 2008. Malic acid production by *Saccharomyces cerevisiae*: engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. *Appl Environ Microbiol*, **74**(9), 2766-77.
- Zhang, F., Carothers, J.M., Keasling, J.D. 2012a. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat Biotechnol*, **30**(4), 354-9.
- Zhang, F., Keasling, J. 2011. Biosensors and their applications in microbial metabolic engineering. *Trends Microbiol*, **19**(7), 323-9.
- Zhang, Y., Werling, U., Edelmann, W. 2012b. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res*, **40**(8), e55.
- Zhang, Y.H. 2011. Substrate channeling and enzyme complexes for biotechnological applications. *Biotechnol Adv*, **29**(6), 715-25.

ANNEXE 1 : Séquences Codantes des Enzymes Etudiées

Séquence CSZCD recodée selon le biais de codons de *S. cerevisiae* :

```
1      ATGCAGGTGGATCCAACAAAAGGCATTGGTCTTGCGAATACGAGCCTGCAGTTTTCAAAT
1      M  Q  V  D  P  T  K  G  I  G  L  A  N  T  S  L  Q  F  S  N

61     GGCCGCCTCCACGCCTTGTGTGAGTACGACCTGCCGTACGTAGTTCGGTTAAGCCCCGAG
21     G  R  L  H  A  L  C  E  Y  D  L  P  Y  V  V  R  L  S  P  E

121    GATGGTGATATCTCGACCGTTGGTAGGATCGAAAATAACGTCTCAACAAAAAGCACAACA
41     D  G  D  I  S  T  V  G  R  I  E  N  N  V  S  T  K  S  T  T

181    GCTCATCCGAAGACTGATCCAGTTACCGGTGAACTTTTCAGTTTTTCGTACGGTCCAATT
61     A  H  P  K  T  D  P  V  T  G  E  T  F  S  F  S  Y  G  P  I

241    CAGCCATACGTTACGTATAGCCGCTATGACTGTGACGGGAAGAAGAGTGGTCCAGATGTG
81     Q  P  Y  V  T  Y  S  R  Y  D  C  D  G  K  K  S  G  P  D  V

301    CCTATATTCTCTTTCAAAGAGCCATCGTTTGTTCACGACTTTGCGATTACCGAGCATTAT
101    P  I  F  S  F  K  E  P  S  F  V  H  D  F  A  I  T  E  H  Y

361    GCTGTGTTCCCCGATATTTCAGATCGTCATGAAACCGGCAGAGATAGTTAGAGGACGTCGC
121    A  V  F  P  D  I  Q  I  V  M  K  P  A  E  I  V  R  G  R  R

421    ATGATCGGACCGGATTTGGAAAAAGTCCCACGTCTGGGCTTATTGCCCCGTTATGCGACT
141    M  I  G  P  D  L  E  K  V  P  R  L  G  L  L  P  R  Y  A  T

481    AGTGATTCCGAAATGCGTTGGTTTGATGTTCTGCGTTTAATATGGTCCATGTTGTCAAT
161    S  D  S  E  M  R  W  F  D  V  P  G  F  N  M  V  H  V  V  N

541    GCTTGGGAAGAAGAAGGCGGCGAAGTAGTTGTGATCGTTGCCCCGAATGTTAGTCCAATT
181    A  W  E  E  E  G  G  E  V  V  V  I  V  A  P  N  V  S  P  I

601    GAGAACGCTATCGACAGATTTGATTTACTGCACGTAAGCGTAGAAATGGCACGGATTGAG
201    E  N  A  I  D  R  F  D  L  L  H  V  S  V  E  M  A  R  I  E

661    CTGAAATCCGGGAGTGTTTCCAGAACACTACTGTCTGCGGAGAACTTGGATTTTCGGTGTG
221    L  K  S  G  S  V  S  R  T  L  L  S  A  E  N  L  D  F  G  V

721    ATCCATAGAGGTTATTTCGGGCCGTAAATCACGCTATGCTTATCTGGGTGTTGGTGATCCA
241    I  H  R  G  Y  S  G  R  K  S  R  Y  A  Y  L  G  V  G  D  P

781    ATGCCTAAAATCCGAGGTGTCGTTAAAGTTGATTTTGAAGTGGCTGGCCGTGGCGAGTGC
261    M  P  K  I  R  G  V  V  K  V  D  F  E  L  A  G  R  G  E  C

841    GTGGTAGCCCGCAGAGAGTTTGGTGTGGATGTTTTGGCGGTGAACCGTTTTTTGTTCCA
281    V  V  A  R  R  E  F  G  V  G  C  F  G  G  E  P  F  F  V  P

901    GCATCATCGAAGAAGTCTGGTGGTGAAGAAGACGACGGTTATGTTGTCTCATATCTGCAT
301    A  S  S  K  K  S  G  G  E  E  D  D  G  Y  V  V  S  Y  L  H
```

```

961      GATGAAGGTAAGGGAGAATCCTCATTTGTTGTAATGGATGCTAGGAGCCCAGAACTGGAA
321      D E G K G E S S F V V M D A R S P E L E

1021     ATTCTCGCCGAGGTTGTTCTGCCGCGACGTGTTCCGTATGGTTTTACGGTCTGTTTGTC
341      I L A E V V L P R R V P Y G F H G L F V

1081     ACCGAAGCGGAGTTACTGTCTCAACAATAA
361      T E A E L L S Q Q

```

Séquence CSZCD^{+N} recodée selon le biais de codons de *S. cerevisiae* :

Jaune : séquence N-terminale, devant laquelle un codon ATG a été ajouté (souligné) pour initier la transcription.

```

1      ATGATTAACAAGGGAGGATCTATTTTCGAGAAACAGATCTTTGGCTGCTGTTTTTTGCGAT
1      M I N K G G S I S R N R S L A A V F C D

61     GCTCTTGATGATTTAATAACTAGACATTCATTCGACCCAGACGCTTTGCACCCCTCTGTT
21     A L D D L I T R H S F D P D A L H P S V

121    GATCCACATAGGGTGTTAAGGGATAACTTCGCTCCGGTCAGTGAATTACCGCCTACTCCA
41     D P H R V L R D N F A P V S E L P P T P

181    TGCAGGGTTGTTAGGGGCACCATAACCCTCCGCTCTAGCCGGAGGCGCCTATATTAGAAAT
61     C R V V R G T I P S A L A G G A Y I R N

241    GGTCCAAACCCAAATCATCTACCATCCGGTGCACATCATCTATTTGAAGGAGATGGCATG
81     G P N P N H L P S G A H H L F E G D G M

301    TTGCATAGCTTATTATTACCTTCTTCCGAAGGTGGCAGAGCCGCTATATTCTCCTCTAGA
101    L H S L L L P S S E G G R A A I F S S R

361    TTCGTTGAGACATATAAATATCTTGTTACAGCTAAATCTCGCCAAGCTATATTTTAAAGT
121    F V E T Y K Y L V T A K S R Q A I F L S

421    GTCTTTTCGGGTCTATGCGGATTACCCGGCATAGCCCGCGCTCTTGCTTCTTTTTTAGA
141    V F S G L C G F T G I A R A L V F F F R

481    TTCTTGACCATGCAGGTGGATCCAACAAAAGGCATTGGTCTTGCGAATACGAGCCTGCAG
161    F L T M Q V D P T K G I G L A N T S L Q

541    TTTTCAAATGGCCGCCTCCACGCCTTGTGTGAGTACGACCTGCCGTACGTAGTTCGGTTA
181    F S N G R L H A L C E Y D L P Y V V R L

601    AGCCCCGAGGATGGTGATATCTCGACCGTTGGTAGGATCGAAAATAACGTCTCAACAAAA
201    S P E D G D I S T V G R I E N N V S T K

661    AGCACAACAGCTCATCCGAAGACTGATCCAGTTACCGGTGAAACTTTCAGTTTTTTCGTAC
221    S T T A H P K T D P V T G E T F S F S Y

721    GGTCCAATTACGCCATACGTTACGTATAGCCGCTATGACTGTGACGGGAAGAAGAGTGGT
241    G P I Q P Y V T Y S R Y D C D G K K S G

```

781 CCAGATGTGCCTATATTCTCTTTCAAAGAGCCATCGTTTGTTCACGACTTTGCGATTACC
 261 P D V P I F S F K E P S F V H D F A I T

 841 GAGCATTATGCTGTGTTCCCCGATATTCAGATCGTCATGAAACCGGCAGAGATAGTTAGA
 281 E H Y A V F P D I Q I V M K P A E I V R

 901 GGACGTCGCATGATCGGACCGGATTTGGAAAAAGTCCCACGTCTGGGCTTATTGCCCCGT
 301 G R R M I G P D L E K V P R L G L L P R

 961 TATGCGACTAGTGATTCCGAAATGCGTTGGTTTGATGTTTCCTGGGTTTAATATGGTCCAT
 321 Y A T S D S E M R W F D V P G F N M V H

 1021 GTTGTCAATGCTTGGGAAGAAGAAGGCGGCGAAGTAGTTGTGATCGTTGCCCCGAATGTT
 341 V V N A W E E E G G E V V V I V A P N V

 1081 AGTCCAATTGAGAACGCTATCGACAGATTTGATTTACTGCACGTAAGCGTAGAAATGGCA
 361 S P I E N A I D R F D L L H V S V E M A

 1141 CGGATTGAGCTGAAATCCGGGAGTGTTTCCAGAACACTACTGTCTGCGGAGAACTTGGAT
 381 R I E L K S G S V S R T L L S A E N L D

 1201 TTCGGTGTGATCCATAGAGGTTATTCGGGCCGTAAATCACGCTATGCTTATCTGGGTGTT
 401 F G V I H R G Y S G R K S R Y A Y L G V

 1261 GGTGATCCAATGCCTAAAATCCGAGGTGTCGTTAAAGTTGATTTTGAAGTGGCTGGCCGT
 421 G D P M P K I R G V V K V D F E L A G R

 1321 GGCGAGTGCGTGGTAGCCCGCAGAGAGTTTGGTGTGGATGTTTTGGCGGTGAACCGTTT
 441 G E C V V A R R E F G V G C F G G E P F

 1381 TTTGTTCCAGCATCATCGAAGAAGTCTGGTGGTGAAGAAGACGACGGTTATGTTGTCTCA
 461 F V P A S S K K S G G E E D D G Y V V S

 1441 TATCTGCATGATGAAGGTAAGGGAGAATCCTCATTTGTTGTAATGGATGCTAGGAGCCCA
 481 Y L H D E G K G E S S F V V M D A R S P

 1501 GAACTGGAAATTCTCGCCGAGGTTGTTCTGCCGCGACGTGTTCCGTATGGTTTTACGGT
 501 E L E I L A E V V L P R R V P Y G F H G

 1561 CTGTTTGTCAACGAAGCGGAGTTACTGTCTCAACAATAA
 521 L F V T E A E L L S Q Q

Séquence CSZCD^{+N+P} signal recodée selon le biais de codons de *S. cerevisiae* :

Jaune : séquence N-terminale.

Bleu : Peptide signal.

```
1      ATGGATTACAGATTATCCTCGTCATCCTTATTTCACTTCCCTTCTCCAGGCAACCGTATC
1      M  D  Y  R  L  S  S  S  S  L  F  H  F  P  S  P  G  N  R  I

61     TTCTTAAAACAGTCACAAGTCTTAGCTTTTCAAACCAGCCTTCTCATCAAGACCATCCT
21     F  L  K  Q  S  Q  V  L  A  F  Q  N  Q  P  S  H  Q  D  H  P

121    ACAACAAAAAAGAAATCCATATCGATTAACAAGGGAGGATCTATTTGAGAAACAGATCT
41     T  T  K  K  K  S  I  S  I  N  K  G  G  S  I  S  R  N  R  S

181    TTGGCTGCTGTTTTTTGCGATGCTCTTGATGATTTAATAACTAGACATTCATTGACCCA
61     L  A  A  V  F  C  D  A  L  D  D  L  I  T  R  H  S  F  D  P

241    GACGCTTTGCACCCCTCTGTTGATCCACATAGGGTGTTAAGGGATAACTTCGCTCCGGTC
81     D  A  L  H  P  S  V  D  P  H  R  V  L  R  D  N  F  A  P  V

301    AGTGAATTACCGCCTACTCCATGCAGGGTTGTTAGGGGCACCATACCCTCCGCTCTAGCC
101    S  E  L  P  P  T  P  C  R  V  V  R  G  T  I  P  S  A  L  A

361    GGAGGCGCCTATATTAGAAATGGTCCAAACCCAAATCATCTACCATCCGGTGCACATCAT
121    G  G  A  Y  I  R  N  G  P  N  P  N  H  L  P  S  G  A  H  H

421    CTATTTGAAGGAGATGGCATGTTGCATAGCTTATTATTACCTTCTTCCGAAGGTGGCAGA
141    L  F  E  G  D  G  M  L  H  S  L  L  L  P  S  S  E  G  G  R

481    GCCGCTATATTCTCCTCTAGATTGCTTGAGACATATAAATATCTTGTTACAGCTAAATCT
161    A  A  I  F  S  S  R  F  V  E  T  Y  K  Y  L  V  T  A  K  S

541    CGCCAAGCTATATTTTTTAAGTGTCTTTTCGGGTCTATGCGGATTACCGGCATAGCCCCGC
181    R  Q  A  I  F  L  S  V  F  S  G  L  C  G  F  T  G  I  A  R

601    GCTCTTGTCTTCTTTTTTTAGATTCTTGACCATGCAGGTGGATCCAACAAAAGGCATTGGT
201    A  L  V  F  F  F  R  F  L  T  M  Q  V  D  P  T  K  G  I  G

661    CTTGCGAATACGAGCCTGCAGTTTTCAAATGGCCGCCTCCACGCCTTGTGTGAGTACGAC
221    L  A  N  T  S  L  Q  F  S  N  G  R  L  H  A  L  C  E  Y  D

721    CTGCCGTACGTAGTTCGGTTAAGCCCCGAGGATGGTGATATCTCGACCGTTGGTAGGATC
241    L  P  Y  V  V  R  L  S  P  E  D  G  D  I  S  T  V  G  R  I

781    GAAAATAACGTCTCAACAAAAAGCACAACAGCTCATCCGAAGACTGATCCAGTTACCGGT
261    E  N  N  V  S  T  K  S  T  T  A  H  P  K  T  D  P  V  T  G

841    GAAACTTTTCAGTTTTTTCGTACGGTCCAATTTCAGCCATACGTTACGTATAGCCGCTATGAC
281    E  T  F  S  F  S  Y  G  P  I  Q  P  Y  V  T  Y  S  R  Y  D

901    TGTGACGGGAAGAAGAGTGGTCCAGATGTGCCTATATTCTCTTTCAAAGAGCCATCGTTT
301    C  D  G  K  K  S  G  P  D  V  P  I  F  S  F  K  E  P  S  F

961    GTTCACGACTTTGCGATTACCGAGCATTATGCTGTGTTCCCCGATATTTCAGATCGTCATG
321    V  H  D  F  A  I  T  E  H  Y  A  V  F  P  D  I  Q  I  V  M
```

1021 AAACCGGCAGAGATAGTTAGAGGACGTCGCATGATCGGACCGGATTTGGAAAAAGTCCCA
341 K P A E I V R G R R M I G P D L E K V P

1081 CGTCTGGGCTTATTGCCCCGTTATGCGACTAGTGATTCCGAAATGCGTTGGTTTGATGTT
361 R L G L L P R Y A T S D S E M R W F D V

1141 CCTGGGTTTAAATATGGTCCATGTTGTCAATGCTTGGGAAGAAGAAGGCGGCGAAGTAGTT
381 P G F N M V H V V N A W E E E G G E V V

1201 GTGATCGTTGCCCGAATGTTAGTCCAATTGAGAACGCTATCGACAGATTTGATTTACTG
401 V I V A P N V S P I E N A I D R F D L L

1261 CACGTAAGCGTAGAAATGGCACGGATTGAGCTGAAATCCGGGAGTGTTTCCAGAACACTA
421 H V S V E M A R I E L K S G S V S R T L

1321 CTGTCTGCGGAGAACTTGGATTTCCGGTGTGATCCATAGAGGTTATTCGGGCCGTAAATCA
441 L S A E N L D F G V I H R G Y S G R K S

1381 CGCTATGCTTATCTGGGTGTTGGTGATCCAATGCCTAAAATCCGAGGTGTCGTTAAAGTT
461 R Y A Y L G V G D P M P K I R G V V K V

1441 GATTTTGAAGTGGCTGGCCGTGGCGAGTGCGTGGTAGCCCGCAGAGAGTTTGGTGTGGA
481 D F E L A G R G E C V V A R R E F G V G

1501 TGT TTTGGCGGTGAACCGTTTTTTTGTTCAGCATCATCGAAGAAGTCTGGTGGTGAAGAA
501 C F G G E P F F V P A S S K K S G G E E

1561 GACGACGGTTATGTTGTCTCATATCTGCATGATGAAGGTAAGGGAGAATCCTCATTGT
521 D D G Y V V S Y L H D E G K G E S S F V

1621 GTAATGGATGCTAGGAGCCCAGAACTGGAATTTCTCGCCGAGGTTGTTCTGCCGCGACGT
541 V M D A R S P E L E I L A E V V L P R R

1681 GTTCCGTATGGTTTTTACGGTCTGTTTGTACCCGAAGCGGAGTTACTGTCTCAACAATAA
561 V P Y G F H G L F V T E A E L L S Q Q

Séquence de CSCCD recodée selon le biais de codons de *S. cerevisiae* :

1 ATGGGTGAAGTTGCTAAGGAAGAAGTTGAAGAACGTAGAAGTATTGTTGCTGTGAACCTT
1 M G E V A K E E V E E R R S I V A V N P

61 CAACCTAGCAAAGGCTTGGTTTCGTCTGCAGTTGACTTGATTGAAAAAGCGGTTGTATAC
21 Q P S K G L V S S A V D L I E K A V V Y

121 TTGTTTCATGACAAGTCAAAGCCTTGTTCATTACCTAAGCGGTAATTTGCTCCAGTCGTG
41 L F H D K S K P C H Y L S G N F A P V V

181 GATGAAACCCACCCCTGTCTTGACCTTCCTGTAAGGGGACATCTTCCTGAATGTCTGAAC
61 D E T P P C P D L P V R G H L P E C L N

241 GGTGAATTTGTGAGGGTGGGCCCAAATCCAAAATTCATGCCCCGTAGCGGGATACCACTGG
81 G E F V R V G P N P K F M P V A G Y H W

301 TTTGACGGTGATGGTATGATTCATGGCATGAGGATTAAAGACGGTAAAGCGACCTATGCG

101 F D G D G M I H G M R I K D G K A T Y A
 361 TCCAGGTACGTCAAAACCTCGAGATTGAAGCAAGAGGAGTACTTCGAAGGACCAAAATTT
 121 S R Y V K T S R L K Q E E Y F E G P K F
 421 ATGAAAATCGGAGATCTGAAAGGATTTTTCGGCTTATTTATGGTTCAGATGCAGTTATTG
 141 M K I G D L K G F F G L F M V Q M Q L L
 481 AGGGCAAAGTTAAAAGTTATAGATGTATCCTACGGCGTAGGAACCGGTAACACGGCCTTG
 161 R A K L K V I D V S Y G V G T G N T A L
 541 ATCTATCATCACGGTAAATTGCTTGCCTTATCCGAGGCTGATAAACCATACGTTGTTAAG
 181 I Y H H G K L L A L S E A D K P Y V V K
 601 GTTTTGAAGATGGGGACCTCCAAACACTCGGACTACTGGACTATGATAAGCGCCTGTCT
 201 V L E D G D L Q T L G L L D Y D K R L S
 661 CACTCTTTTACTGCTCACCCAAAAGTCGATCCTTTCACCGATGAAATGTTACATTTGGA
 221 H S F T A H P K V D P F T D E M F T F G
 721 TATGCGCATACACCACCATACGTGACATACAGGGTAATATCAAAGATGGAGTAATGAGG
 241 Y A H T P P Y V T Y R V I S K D G V M R
 781 GATCCAGTTCCTATCACGATACCTGCTAGCGTGATGATGCATGATTTTGCAATCACAGAG
 261 D P V P I T I P A S V M M H D F A I T E
 841 AACTATTCTATTTTCATGGATCTTCCTTTGTATTTCCAACCAAAGAAATGGTCAAAGGA
 281 N Y S I F M D L P L Y F Q P K E M V K G
 901 GGTAACATAATCTTTTCATTCGATGCTACTAAAAAGCACGATTTGGTGTCTCCCTAGA
 301 G K L I F S F D A T K K A R F G V L P R
 961 TACGCTAAGGATGATTCCTTAATCAGATGGTTTGAATTACCTAATTGTTTTATTTTCCAT
 321 Y A K D D S L I R W F E L P N C F I F H
 1021 AATGCAAACGCTTGGGAGGAAGGTGACGAAGTTGTACTAATCACTTGTAGGTTGGAAAAT
 341 N A N A W E E G D E V V L I T C R L E N
 1081 CCTGACCTAGACATGGTAAATGGAGCAGTGAAGGAAAACTAGAGAATTTCAAAAATGAA
 361 P D L D M V N G A V K E K L E N F K N E
 1141 CTGTACGAAATGAGGTTTAAACATGAAAACAGGGGCGGCATCTCAAAAACAACTTTCCGTG
 381 L Y E M R F N M K T G A A S Q K Q L S V
 1201 AGCGCAGTAGATTTCCCCAGAATAAATGAATCTTACACCACAAGGAAGCAAAGATATGTT
 401 S A V D F P R I N E S Y T T R K Q R Y V
 1261 TATGGAACCATCTTGGACAATATTACAAAAGTGAAAGGCATTATAAAATTCGATTTGCAT
 421 Y G T I L D N I T K V K G I I K F D L H
 1321 GCTGAACCTGAAGCTGGTAAGAAGAAGCTGGAAGTTGGTGGCAACGTTTCAGGGGATCTTT
 441 A E P E A G K K K L E V G G N V Q G I F
 1381 GATCTAGGTCCGGGCCGCTATGGAAGCGAAGCTGTATTTGTCCCGAGAGAAAGAGGTATA
 461 D L G P G R Y G S E A V F V P R E R G I
 1441 AAATCAGAGGAAGATGATGGTTACTTGATATTTTTCGTCCATGATGAAAACACTGGAAAA

481 K S E E D D G Y L I F F V H D E N T G K
1501 TCTGAAGTAAATGTCATCGACGCTAAAACAATGTCAGCCGAACCCGTGGCAGTGGTAGAA
501 S E V N V I D A K T M S A E P V A V V E
1561 TTACCAAATAGAGTTCCTATGGTTTCCATGCGTTCTTCGTAAACGAAGAACAATTGCAA
521 L P N R V P Y G F H A F F V N E E Q L Q
1621 TGGCAACAACTGATGTCTAA
541 W Q Q T D V

Séquence de VVCCD1 donnée par le Dr Nancy Terrier (Mathieu et al. 2005):

1 ATGGCGGAGAAGGAGGAGCAAGGCGGCGCAGGAGTAGTGGTGGTGGATCCGAAGCCGAGT
1 M A E K E E Q G G A G V V V V D P K P S
61 AAGGGGTTCACTTCGAAGGCGGTGGATTGGCTGGAGAAGCTGATAGTGAAGTTGATGCAC
21 K G F T S K A V D W L E K L I V K L M H
121 GATTCTTCACAGCCTCTCCACTACTTGTCTGGAAACTTCGCTCCGGTTCGCGACGAGACT
41 D S S Q P L H Y L S G N F A P V R D E T
181 CCGCCGTGCAAGAACCTTCCGGTCATTGGGTATCTTCCGGAGTGCCTGAATGGAGAGTTT
61 P P C K N L P V I G Y L P E C L N G E F
241 GTCAGGGTTGGCCCCAACCCTAAGTTTTCCCTGTGGCTGGATATCATTGGTTTGATGGA
81 V R V G P N P K F S P V A G Y H W F D G
301 GATGGCATGATTCATGGTTTGCACATCAAAGATGGAAAAGCTACATATGTCTCACGTTAT
101 D G M I H G L H I K D G K A T Y V S R Y
361 GTTAGGACATCACGTCTTAAACAAGAAGAGTATTTTGGAGGAGCAAAATTCATGAGGATT
121 V R T S R L K Q E E Y F G G A K F M R I
421 GGTGATCTTAAGGGCCTGTTTGGATTGCTTATGGTTAACATGCAAATGCTGAGAGCAAAA
141 G D L K G L F G L L M V N M Q M L R A K
481 TTGAAAATATTGGATGTTTCATATGGAACCTGGGACAGGTAATACAGCTCTAGTGTTCAC
161 L K I L D V S Y G T G T G N T A L V F H
541 CATGGTAAGCTTTTGGCACTTTTCGGAGGCTGATAAACCTTATGTCCTAAAAGTTTGGAA
181 H G K L L A L S E A D K P Y V L K V L E
601 GATGGAGATCTGCAAACCCTGGGCATGCTGGATTATGATAAGAGATTGACCCATTCCTTC
201 D G D L Q T L G M L D Y D K R L T H S F
661 ACTGCTCATCCAAAGGTTGACCCCTTCACTGGTGAAATGTTTAGCTTTGGCTATTACAT
221 T A H P K V D P F T G E M F S F G Y S H
721 ACACCTCCTTATATTACATACAGAGTTATCTCGAAGGATGGTTTCATGCATGAGCCAGTG
241 T P P Y I T Y R V I S K D G F M H E P V
781 CCGATAACAATTTTCAGATCCTATCATGATGCATGATTTTGCCATTACCGAGAATTATGCA

261 P I T I S D P I M M H D F A I T E N Y A
 841 ATTTTCATGGATCTTCCTTTGTATTTTCAGACCAAAGGAAATGGTGAAAGAAAAAAGCTA
 281 I F M D L P L Y F R P K E M V K E K K L
 901 ATATTACATTTGATGCAACAAAAAGGCTCGTTTTGGTGTACTTCCTCGATATGCAAAG
 301 I F T F D A T K K A R F G V L P R Y A K
 961 AATGAGCTCCACATTAAATGGTTTGAGCTTCCAAATTGCTTCATATTCCACAATGCCAAT
 321 N E L H I K W F E L P N C F I F H N A N
 1021 GCTTGGGAGGAGGAGGACGAAGTTGTTTTAATCACCTGCCGCCTTGAAAATCCAGATTTA
 341 A W E E E D E V V L I T C R L E N P D L
 1081 GACCTGGTTGGTGGGGATGTTAAAGAAAAGCTTGAAAATTTTGGAATGAGCTGTATGAG
 361 D L V G G D V K E K L E N F G N E L Y E
 1141 ATGAGATTCAACATGAAAACGGGTATTGCGTCACAGAGAAAACATCAGCATCTTCTGTG
 381 M R F N M K T G I A S Q R K L S A S S V
 1201 GATTTTCCAAGGGTGAATGAGAGCTACACTGGCAGGAAACAACGATATGTATATGGAACC
 401 D F P R V N E S Y T G R K Q R Y V Y G T
 1261 ATACTGGACAGCATAGCCAAGGTGACGGGCATTATCAAGTTTGATCTGCATGCTGAACCC
 421 I L D S I A K V T G I I K F D L H A E P
 1321 GACACCGGAAAATCAAAGCTTGAAGTTGGAGGGAACGTTCAAGGCATTTTTGACTTGGGG
 441 D T G K S K L E V G G N V Q G I F D L G
 1381 GTCGGTAGATTTGGTTTCAGAAGCTGTCTTTGTCCCTCGTGAGCCTGGTATCACCTCTGAA
 461 V G R F G S E A V F V P R E P G I T S E
 1441 GAAGATGATGGCTACTTGATATTCTTCGTCCATGACGAGAAAACCTGGAAAATCATATGTG
 481 E D D G Y L I F F V H D E K T G K S Y V
 1501 AATGTAATCGATGCAAAAACGATGTCGCCCCGATCCAATTGCCATTGTTGAACTACCCAAT
 501 N V I D A K T M S P D P I A I V E L P N
 1561 AGAGTTCCATACGGTTTTTCATGCCTTCTTTGTCACAGAGGAGCAACTCAAAGAACAAGCA
 521 R V P Y G F H A F F V T E E Q L K E Q A
 1621 AAAC TTGA
 541 K L

ANNEXE 2 : Amorces de PCR

Tableau 13: Amorces utilisées pour la construction de vecteurs bactériens. Rouge : site NdeI. Vert : site Bpu1102I. Bleu : base T ajoutée pour reproduire la construction qui aurait été obtenue avec le pBAD/TOPO Thiofusion® expression kit (TA cloning). Les séquences correspondant à ces zones de recombinaison pour clonage sont soulignées.

Nom de l'amorce	Séquence 5'→3'
N-Ter(b)_pET15b_F	<u>CCGCGCGGCAGCCAT</u> ATTAACAAGGGAGGATCTAT
N-Ter(b)_pET15b_R	TGGATCCACCTGCATGGTCAAGAATCTAAAAAAGA
Ps(a)_pET15b_F	<u>CCGCGCGGCAGCCAT</u> ATGGATTACAGATTATCCTC
Ps(a)_pET15b_R	<u>TCCTCCCTTGTTAAT</u> CGATATGGATTTCCTTTTTTG
CsZCD ^{+N} _pET15b_F	ATTAACAAGGGAGGATCTATTTTCGAGAAACAGATCTTTGG
CsZCD ^{+N} _pET15b_R	CGGCCTGGTGCCGCGCGGCAGCCAT
CsCCD_pET15b_F	GCGCATATGGGTGAAGTTGCTAAGGA
CsCCD_pET15b_R	GCGGCTCAGCTTAGACATCAGTTTGTTGCCATT
pBAD_F	AGGGCGAGCTTGAAGGTAAGCCTAT
pBAD_R	GGTGATGACGATGACAAGCTCGCCCTT
CsZCD_pBAD_F1	<u>GACAAGCTCGCCCTT</u> ATGCAGGTGGATCCAACAAAAG
CsZCD_pBAD_R1	<u>GGCTTACCTTCAAGCTCGCCCT</u> TTTATTGTTGAGACAGTAACTC
CsZCD_pBAD_F2	<u>GGCTCTGGATCCGGTGATGACGATGACAAGCTCGCCCTT</u>
CsZCD_pBAD_R2	<u>CGAGGAGAGGGTTAGGGATAGGCTTACCTTCAAGCTCGCCCT</u>

Tableau 14: Amorces utilisées pour la construction de vecteurs de levure. Gras : Séquence Kozak (Kozak, 1981). Gras et italique : codon STOP supplémentaire.

Nom de l'amorce	Séquence 5'→3'
CsZCD_pCM185_F	<u>ATTACCGGATCAATTCGGGG</u> ATGCAGGTGGATCCAACAAA
CsZCD ^{+N} _pCM185_F	<u>ATTACCGGATCAATTCGGGG</u> ATGATTAACAAGGGAGGATC
CsZCD ^{+N+P} _pCM185_F	<u>ATTACCGGATCAATTCGGGG</u> ATGGATTACAGATTATCCTC
CsZCD_pCM185_R	<u>CAGGCCTGTTTAAACGGATCTTATTGTTGAGACAGTAACT</u>
CsCCD_pCM185_F	<u>ATTACCGGATCAATTCGGGG</u> ATGGGTGAAGTTGCTAAGGA
CsCCD_pCM185_R	<u>CAGGCCTGTTTAAACGGATCTT</u> AGACATCAGTTTGTTGCC
VvCCD1_pCM185_F	<u>ATTACCGGATCAATTCGGGG</u> ATGGCGGAGAAGGAGGAGCAAG
VvCCD1_pCM185_R	<u>CAGGCCTGTTTAAACGGATCTCAAAGTTTTGCTTGTTCTTTG</u>
CsZCD_V60_F	<u>ACACACACTAAATTACCAAAATGGCTGTGGATCCAACAAAAGGCAT</u>
CsZCD_V60_R	<u>GAGACATGGGAGATCCCCGTTTATTATTGTTGAGACAGTAACT</u>
V60_F	<u>ACTTCTATAGACACGCAAACACAAATACACACACTAAATTACC</u>
V60_R	<u>AAGAAGCACCACCACAGTAGAGACATGGGAGATCCCCCG</u>

ANNEXE 3 : Milieux de Culture

LB (Sigma): Tryptone: 10 g/L
Yeast Extract: 5 g/L
NaCl: 10 g/L

LBA: milieu LB + Ampicilline (Sigma) 100µg/mL

2xYT (Difco): Tryptone: 16 g/L
Yeast Extract: 10 g/L
NaCl: 5 g/L

YPD (Sigma): Yeast Extract: 10 g/L
Peptone: 20 g/L
Dextrose (=glucose): 20 g/L

YPGE:

	Concentration
Yeast Extract (Difco) (1%)	10 g/L
Peptone (Difco) (2%)	20 g/L
Glucose (0,5%)	5 g/L
Stérilisation par autoclave 20 minutes à 121°C	
éthanol (2%) (extemporanément)	20 ml/L

ZYM 5052:

	Composants	Concentration
ZY	Tryptone (Difco)	10 g/L
	Yeast Extract (Difco)	5 g/L
M	Na ₂ HPO ₄	25 mM
	KH ₂ PO ₄	25 mM
	NH ₄ Cl	50 mM
	Na ₂ SO ₄	5 mM
5052	Glycérol	0,5%
	Glucose	0,05%
	α -lactose	0,2%
Métaux	FeCl ₃	50 μ M
	FeCl ₃	50 μ M
	CaCl ₂	20 μ M
	MnCl ₂	10 μ M
	ZnSO ₄	10 μ M
	CoCl ₂	2 μ M
	CuCl ₂	2 μ M
	NiCl ₂	2 μ M
	Na ₂ MoO ₄	2 μ M
	Na ₂ SeO ₃	2 μ M
	H ₃ BO ₃	2 μ M
	MgSO ₄	2 mM

Le milieu était préparé à partir des stocks suivants conservés à +4°C:

5052 50X
M 50X
Métaux 1000X
MgSO₄ 1M

Thèse dirigée par le Dr Gilles TRUAN

TITRE : Approches combinatoires pour la reconstruction d'une voie de biosynthèse chez *S. cerevisiae* : variation des niveaux d'expression et analyse fonctionnelle d'une étape clé de la voie.

RESUME :

Afin d'optimiser la production d'un composé d'intérêt tout en évitant les conséquences néfastes sur la viabilité cellulaire, un défi majeur de l'ingénierie métabolique est d'obtenir un équilibre entre les flux métaboliques endogènes de la cellule et le flux consommé par une nouvelle voie de biosynthèse. Dans ce contexte d'optimisation, les stratégies combinatoires génèrent une diversité de voies métaboliques et de modes de régulation rassemblant de précieuses informations quant aux choix d'orientations stratégiques à faire.

Notre étude s'inscrit dans un projet visant à produire les molécules responsables de l'arôme, de la couleur et du parfum du safran (*Crocus sativus*) chez *Saccharomyces cerevisiae*. Une approche combinatoire a été adoptée pour moduler les niveaux d'expression de trois gènes menant à la synthèse de leur précurseur commun : la zéaxanthine. Cette stratégie nous a permis de décrire des biais inattendus dans la régulation des niveaux d'expression des gènes exprimés sur plasmide. Nous avons détecté une forte interférence transcriptionnelle entre les gènes dans notre système, ainsi qu'une influence de la nature des séquences codantes. Ces éléments, identifiés comme critiques, imposent sur les niveaux d'expression des trois gènes une régulation plus importante que la force des promoteurs qui les contrôlent. Afin de poursuivre le projet vers son objectif final, la réaction de clivage de la zéaxanthine menant à la synthèse des composés d'intérêt du safran a fait l'objet d'une analyse fonctionnelle détaillée. Une absence d'activité de l'enzyme décrite dans la littérature comme responsable de cette réaction nous a conduit à proposer des perspectives d'ingénierie pour atteindre l'objectif final du projet.

MOTS-CLES : Ingénierie métabolique combinatoire, *Saccharomyces cerevisiae*, régulation transcriptionnelle, caroténoïdes, dioxygénases, arômes.