



Conception de nouveaux biocatalyseurs par fusion de domaines catalytiques

Mamy Hery Ny Aina Rabeharindranto

► To cite this version:

Mamy Hery Ny Aina Rabeharindranto. Conception de nouveaux biocatalyseurs par fusion de domaines catalytiques. Microbiologie et Parasitologie. INSA de Toulouse, 2019. Français. NNT : 2019ISAT0015 . tel-02443294

HAL Id: tel-02443294

<https://theses.hal.science/tel-02443294>

Submitted on 17 Jan 2020

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 l'Institut National des Sciences Appliquées de
Toulouse**

**Présentée et soutenue par
Hery RABEHARINDRANTO**

Le 8 juillet 2019

**Conception de nouveaux biocatalyseurs par fusion de domaines
catalytiques**

Ecole doctorale : **SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et
Bioingenieries**

Spécialité : **Ingénieries microbienne et enzymatique**

Unité de recherche :
LISBP - Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés

Thèse dirigée par
GILLES TRUAN

Jury

Mme Muriel Gondry, Rapporteuse
Mme Laura BACIOU, Rapporteuse et Présidente du Jury
M. Vincent Courdavault, Examinateur
M. Lautier Thomas, Examinateur
M. Gilles TRUAN, Directeur de thèse

Thèse dirigé par le Pr Gilles TRUAN

Titre : Conception de nouveaux biocatalyseurs par fusion de domaines catalytiques

Résumé

La production de molécules d'intérêt par voie microbienne peut être améliorée par différentes stratégies d'ingénierie. L'ingénierie enzymatique tient une place centrale dans l'obtention de souches de production économiquement viables. L'efficacité des voies de biosynthèse dépend en effet de l'efficacité de chaque enzyme de la voie. Les stratégies d'ingénierie enzymatique peuvent s'inspirer de solutions observées dans la nature. De plus, la technologie de la synthèse d'ADN permet aujourd'hui la création de nouvelles configurations d'enzymes à façon. La question est de savoir quelles configurations d'enzymes synthétiques sont efficaces et quelles sont les mises au point nécessaires pour confirmer leur efficacité. L'obtention de souches de production efficaces peut enfin s'appuyer sur l'utilisation d'outils issus de la biologie systémique.

La colocalisation des enzymes d'une voie de biosynthèse pourrait aider à la transformation rapide des intermédiaires d'une voie de biosynthèse et donc favoriser la production des molécules d'intérêt. Des protéines multidomaines regroupant au sein d'une même protéine plusieurs activités enzymatiques sont décrites dans la nature. Ce travail de thèse s'est focalisé en partie sur la création de fusions synthétiques d'enzymes au sein de la voie de biosynthèse du beta carotène chez *Saccharomyces cerevisiae*. Différents types de fusions et d'organisations enzymatiques ont été testés sur les enzymes caroténogéniques. L'étude a mis en évidence une configuration de fusion efficace réduisant l'accumulation d'intermédiaires réactionnels en faveur de la production du beta carotène. La mesure précise des concentrations des caroténoïdes produits, couplée à des méthodes de quantification des enzymes ont permis de caractériser finement l'efficacité de chaque construction enzymatique. D'autres stratégies de colocalisation spatiale ont également été testées. Nous avons notamment utilisé les domaines d'interaction cohésine dockérine et la protéine oligomérique CcmK2 pour cela. Il a été montré que les enzymes caroténogéniques sont fonctionnelles dans ces conditions et ces différentes stratégies de colocalisation permettent aussi de modifier le flux de la voie de production en caroténoïde. Du au changement en concentration des intermédiaires de la voie, on observe un changement dans la nature du produit final. De manière intéressante, l'étude permet d'envisager une régulation plus fine des niveaux d'activité des enzymes au sein des différentes stratégies de colocalisation pour choisir la nature du produit final. Les études sur des stratégies de colocalisation d'enzymes, notamment membranaires, sont peu décrites dans la littérature, ce travail de thèse a contribué à montrer des possibilités dans ce domaine.

Mots clés : Ingénierie métabolique, ingénierie enzymatique, colocalisation d'enzyme, linker, domaines d'interaction, caroténoïdes, enzymes membranaires, composés insolubles, *Saccharomyces cerevisiae*

Thesis directed by Pr Gilles TRUAN

Titre : Design of biocatalysts by domain fusion and scaffolding

Abstract :

Microbial production of molecules of interest could be improved by several engineering strategies. Enzymatic engineering has a central role in the conception of efficient host because pathway's efficiency depends in first place on the efficiency of enzymes. Enzymatic engineering can take inspiration from strategies observed in nature. Today, DNA synthesis technology allows the conception of potentially any kind of enzyme configuration. Challenges are today to know which synthetic enzymes conceptions might be efficient and what parameters need to be tested to characterize the best ones. Creating efficient microbial host can take advantage of the use of several tools from fields inside systemic biology.

Enzyme colocalisation inside metabolic pathway might improves the production of final molecule of interest by allowing rapid biotransformation of intermediates of the pathway. Multidomain proteins regrouping several enzymatic activities are described in the literature. This work focused on the creation of synthetic fusion of carotenogenic enzymes for the production of beta carotene in *Saccharomyces cerevisiae*. Different types of enzymatic fusions and configurations have been tested. The study allowed the creation of an efficient tripartite enzyme fusion which produces less intermediates and more beta carotene. Precise measurement of each carotenoid's concentration coupled with quantification of enzymes allows the characterisation of the efficiency of each synthetic enzyme. Other strategies for enzyme spatial colocalisation have also been tested using domains of interaction like cohesine-dockerine or the oligomeric protein CcmK2. Carotenogenic enzymes are still functional inside those configurations. Some strategies have changed the metabolic flux of carotenoids inside the pathway. Concentration of intermediates which are substrates of last enzymatic step changes and consequently the nature of the final product changes as well. Interestingly, a fine control of activity of enzyme might allow a fine control of the nature of the final carotenoid. Other strategies for colocalisation of those enzymes could be envisaged. Colocalisation of membrane enzymes are less described in the literature compared with soluble enzymes, and this study gave some result in this direction.

Keywords: Metabolic engineering, enzymatic engineering, enzyme colocalisation, linker, interaction domains, carotenoids, membrane enzymes, insoluble molecules, *Saccharomyces cerevisiae*

Remerciements

Je voudrais tout d'abord remercier grandement mon directeur de thèse, Gilles TRUAN, directeur du laboratoire d'Ingénierie Moléculaire et Métabolique, pour m'avoir encadré durant ces travaux de thèse. Je suis ravi d'avoir travaillé en sa compagnie. Son appui scientifique et ses conseils ont été précieux au cours de l'élaboration de cette thèse.

Je remercie Claude Marange et les personnes au sein de l'école doctorale pour leur aide durant ces années de doctorat.

J'adresse tous mes remerciements à Madame Muriel Gondry, directrice du laboratoire d'Enzymologie et Biosynthèse Peptidique Non Ribosomale, ainsi qu'à Madame Laura Baciou, directrice du laboratoire de Chimie Physique, de l'honneur qu'elles m'ont fait en acceptant d'être rapporteurs de cette thèse.

J'exprime ma gratitude à Monsieur Vincent Courdavault, Enseignant Chercheur au Département de biologie et physiologie végétales, et à Monsieur Thomas Lautier, chercheur CNRS au laboratoire d'Ingénierie Moléculaire et Métabolique, pour leur rôle d'examineurs.

Je tiens à remercier particulièrement les membres de l'équipe I2M, Luis Garcia Alles, Philippe Urban, Denis Pompon, Sara Castano Cerezo pour les discussions et les conseils tout au long de ces travaux de thèse.

Je remercie toutes les personnes avec qui j'ai partagé mes études et notamment ces années de thèse et qui ont tous contribué d'une manière ou d'une autre à la réalisation de ce travail de thèse.

Je remercie papa et maman, ma famille et mes amis, pour leur amour et leur soutien.

Et en tout et pour tout, je remercie Dieu qui me guide, me protège et me fortifie dans son amour.

Hery

LISTE DES ABBREVIATIONS

A : Alanine
ADN : Acide désoxyribonucléique
ARNt : acides ribonucléiques de transfert
ARNm : acide ribonucléique messenger
CBM : Carbohydrate binding module
CRISPR : courtes répétitions palindromiques groupées et régulièrement espacées
E : Acide glutamique
FAD : Flavine adénine dinucléotide
G : Glycine
G418 : Généticine
GeLC-MS : in gel digestion masse spectrometry
GRAS : Generally recognized as safe
IPTG : Isopropyl β -D-1-thiogalactopyranoside
K : Lysine
LB : Lysogeny broth
LBA : Lysogeny broth avec Ampicilline
OD600 : Densité optique à 600nm
ORF: Open reading frame
PBS : Phosphate buffered saline
PCR : Polymerase chain reaction
PEG : Polyéthylène glycol
PMSF : Fluorure de phénylméthylsulfonyl
RE : Réticulum endoplasmique
S : Serine
SDS : Laurylsulfate de sodium
MES : Acide 2-(N-morpholino)éthanosulfonique
TALEN : Nucléases effectrices de type activateur de transcription
TB : Terrific broth
YPD : Yeast extract peptone dextrose
YPG : Yeast extract peptone galactose
ZFN : Nucléases à doigts de zinc

TABLE DES MATIERES

I	INTRODUCTION.....	13
I-1	Les apports de l'ingénierie métabolique	14
I.1.1	La biologie de synthèse.....	17
I.1.2	La biologie des systèmes.....	18
I.1.3	Evolution adaptative en laboratoire	19
I-2	Les différentes étapes de construction d'une usine cellulaire	20
I.2.1	Choix de l'hôte de production	20
I.2.2	Design et construction de la voie.....	21
I.2.3	Optimisation d'une voie de biosynthèse	24
I.2.4	Optimisation du châssis	28
I-3	Colocalisation d'enzymes.....	31
I.3.1	Stratégies naturelles pour colocaliser des enzymes	31
I.3.2	Ingénierie de la colocalisation d'enzymes	36
I-4	Ingénierie métabolique pour la production de caroténoïdes.....	44
I.4.1	Les isoprénoïdes.....	44
I.4.2	Les caroténoïdes	46
I.4.3	La voie de biosynthèse du β -carotène	48
I.4.4	La production de β -carotène chez la levure <i>S. cerevisiae</i>	54
I.4.5	Les enzymes caroténogéniques de <i>X. dendrorhous</i>	55
I-5	Présentation du projet de thèse	59
II	MATERIEL ET METHODES.....	60
II-1	Matériel.....	60
II.1.1	Souches de levure	60
II.1.2	Souches de bactéries	60
II-2	Méthodes expérimentales	60
II.2.1	Milieux de culture	60
II.2.2	Méthodes concernant les acides nucléiques	60
II.2.3	Méthodes concernant les protéines	62
II.2.4	Méthodes concernant les caroténoïdes	65
II.2.5	Calcul de la quantité de caroténoïdes produit dans un échantillon	66

III	RESULTATS	68
III-1	Analyses bioinformatiques des enzymes CrtI et CrtB	68
III.1.1	Modélisation par homologie de CrtI	68
III.1.2	Modélisation par homologie de CrtB.....	71
III.1.3	Modélisation de CrtY.....	74
III-2	Stratégie d'expression.....	74
III-3	Colocalisation avec des domaines d'interaction.....	75
III.3.1	Stratégie de colocalisation avec les modules cohésine dockérine	75
III.3.2	Construction et analyse des souches.....	76
III.3.3	Perspectives	77
III-4	Colocalisation par utilisation d'une protéine oligomérique	77
III.4.1	Construction et analyse des souches	78
III.4.2	Perspective.....	78
III-5	Colocalisation par fusion directe	78
III.5.1	Stratégie de colocalisation des enzymes par fusion directe	78
III.5.2	Construction des souches	79
III.5.3	Domaines séparés.....	80
III.5.4	Bifusions.....	81
III.5.5	Fusion tripartite	83
III-6	Tests de production et de purification des enzymes CrtB et CrtI chez <i>E. coli</i>	85
IV	DISCUSSION ET PERSPECTIVES.....	88
V	RÉFÉRENCES BIBLIOGRAPHIQUES.....	93
VI	ANNEXE: article scientifique publié	115

LISTE DES FIGURES

- Figure 1 : Exemples de molécule obtenues par ingénierie métabolique décrits dans la littérature. 15
- Figure 2 : Le développement de nouvelles technologies accélère le passage de l'échelle de laboratoire à l'échelle industrielle. 16
- Figure 3 : Différentes cibles pour améliorer les capacités d'une « usine cellulaire ». 20
- Figure 4 : Représentation des paramètres régulant le flux métabolique. 25
- Figure 5 : Représentation des différentes approches pour équilibrer les activités enzymatiques dans les voie de biosynthèse. 25
- Figure 6 : Un exemple de protéine retrouvée sous forme soluble et parfois en multidomaine: les cytochromes P450. 32
- Figure 7 : Canalisation d'un substrat au sein d'un tunnel intramoléculaire. a. La structure cristallographique de la tryptophane synthase. 33
- Figure 8 : Canalisation d'un substrat par guidage électrostatique. 33
- Figure 9 : Canalisation d'un substrat par utilisation d'un bras oscillant chimique. 33
- Figure 10 : Représentation schématique d'un complexe cellulosomique. 34
- Figure 11 : Encapsulation de réactions dans un microcompartiment bactérien formé par une structure protéique. 35
- Figure 12 : Des exemples de méthodes biochimiques pour évaluer le channeling du substrat. 36
- Figure 13 : Des interactions moléculaires possibles lors de la fusion de deux enzymes. 37
- Figure 14 : Stratégies pour la construction de systèmes multienzymatiques basés sur les interactions coh-doc. 41
- Figure 15 : Une approche de scaffolding pour optimiser la colocalisation des enzymes d'une voie de biosynthèse. 42
- Figure 16 : Stratégie de colocalisation utilisant une fusion avec une protéine formant un complexe oligomérique naturel. 44
- Figure 17 : Les voies de synthèse des précurseurs des isoprénoïdes. 45
- Figure 18 : La voie des précurseurs des différentes classes d'isoprénoides. 46

Figure 19 : Les différentes formes actives de la vitamine A obtenues à partir du β -carotène.	47
Figure 20 : Comparaison des réactions catalysées par la phytoène synthase, CrtM et la squalène synthase humaine ainsi que des structures de CrtM et de la squalène synthase.	49
Figure 21 : Les caractéristiques structurales de la squalène synthase humaine, son orientation face à la membrane et ses séquences consensus avec une phytoène synthase.	50
Figure 22 : Les étapes de la synthèse de lycopène chez les phytoène désaturases de bactéries à partir du 15-cis phytoène. Isomérisation suivie de 4 désaturations successivement à des positions précises	51
Figure 23 : Etapes de la synthèse du lycopène à partir du 15-cis phytoène chez les organismes photosynthétiques.	52
Figure 24 : Détails structuraux de CrtI et de la PDS.	53
Figure 25 : Etapes de la synthèse du β -carotène à partir du lycopène.	54
Figure 26 : Organisation en complexe des enzymes caroténogéniques formant le β -carotène chez les fungi proposés par des études ultérieures.	57
Figure 27 : Illustration schématique du modèle hypothétiques de complexes multienzymatiques caroténogéniques dans le membranes et stroma de thylakoïde de plantes et d'algues.	58
Figure 28 : Les différentes étapes de biologie moléculaire pour l'obtention d'une bifusion.	61
Figure 29 : Voie de biosynthèse d' γ -caroténoïdes, tels que le myxol et la saproxanthine dans laquelle agit CrtD par désaturation en position C3'-C4' du 1-OH- γ -caroténoïde.	68
Figure 30 : Alignement de séquence de CrtI de <i>X. dendrorhous</i> , de la gamma désaturase de <i>Nonlabens dokdonensis</i> , de la PDS de <i>O. sativa</i> et de CrtI de <i>P. ananatis</i> .	69
Figure 31 : Structure tridimensionnelle de CrtD(4REP).	69
Figure 32 : Localisations des extrémités N et C-terminale des désaturases.	70
Figure 33 : Modèle de CrtI à partir de CrtD et sur lequel l'hélice en C terminale non modélisable est représentée.	71
Figure 34 : Alignement de séquences pour la délimitation de la séquence du linker dans CrtYB.	71
Figure 35 Alignement de séquence de CrtB de <i>X. dendrorhous</i> avec d'autres séquences.	73
Figure 36 Modèles par homologie de CrtB obtenus avec le serveur Phyre	Erreur ! Signet non défini.
Figure 37 Optimisation du flux de la voie en amont du GGPP et stratégie d'expression des différents gènes dans cette étude.	75

Figure 38 Quantification des caroténoïdes produits par le système d'enzymes fusionnées aux domaines cohésine et dockerine.	76
Figure 39 Représentation schématique d'un auto assemblage de CrtI lorsque l'enzyme est fusionnée à un domaine Ccmk2.	77
Figure 40 Caractéristiques des linker utilisés dans cette étude.	79
Figure 41 Liste des souches construites dans la partie de l'étude concernant la stratégie de fusion d'enzymes.	79
Figure 42 Expression d'enzymes séparées mesurée par western blot.	80
Figure 43 Quantification des caroténoïdes produits par les souches portant des enzymes individuelles B et I.	81
Figure 44 Quantification des caroténoïdes produits par les souches portant des enzymes individuelles B, I et Y.	81
Figure 45 Expression de bifusions détectée par western blot.	82
Figure 46 Quantification des caroténoïdes produits par les souches portant des enzymes individuelles.	82
Figure 47 Expression de la fusion tripartite détectée par western blot.	84
Figure 48 Detection des peptides CrtY(yb)B(ek)I par analyse GeLC-MS.	84
Figure 49 Quantification des caroténoïdes produits par la souche portant l'enzyme tripartite.	85
Figure 50 Tests de production et de la solubilité initiaux des enzymes CrtB et CrtI.	86
Figure 51 Tests de production et de la solubilité de CrtB après optimisation de codon.	86
Figure 52 Utilisation de détergents pour solubiliser la fraction insoluble de CrtI.	87
Figure 53 Utilisation d'agents chaotropiques pour solubiliser la fraction insoluble de CrtI.	87

LISTE DES TABLEAUX

Tableau 1 : Les attractivités pour la bioproduction par voie microbienne et les challenges associés.	14
Tableau 2 : Descriptif de différentes méthodes d'assemblage d'ADN par recombinaison homologue. d'après (Li and Borodina, 2015)	22

I INTRODUCTION

La bioproduction de molécules d'intérêt par voie microbienne constitue une ressource alternative pour la génération d'énergie et pour la production des molécules nécessaires à notre mode de vie contemporain. Le développement de microorganismes en « usines cellulaires » efficaces produisant carburants, produits chimiques, produits pharmaceutiques, composants de matériaux, nutraceutiques ingrédients alimentaires et autres molécules, est nécessaire pour un changement vers la bioéconomie. Pour être viable économiquement, la bioproduction microbienne doit être efficace. Les atouts de l'utilisation d'usines cellulaires pour la bioproduction sont : une réduction considérable d'émission de déchets et de consommation d'énergie, l'utilisation de ressources renouvelables, la possibilité d'accéder à de nouvelles structures de molécules à façon et qui sont impossibles à obtenir par la chimie organique. Pour être compétitif avec les processus pétrochimiques, il est essentiel que les bioconversions soient exécutées avec une productivité et un rendement élevés. Dans ce cadre, l'ingénierie métabolique concerne "l'amélioration des activités cellulaires" par la manipulation des fonctions enzymatiques, de transport et des fonctions de régulations cellulaires, à travers l'utilisation de la technologie de l'ADN recombinant " (Bailey, 1991). Elle manipule le métabolisme microbien afin d'améliorer la production de composés d'intérêt (Keasling, 2010). Malgré de grandes avancées, la majeure partie des plates-formes microbiennes artificielles utilisées pour la bioproduction restent au stade de preuve de concept en raison d'un faible rendement. Aujourd'hui, d'autres disciplines comme la biologie synthétique, la biologie des systèmes ou l'ingénierie évolutive, aident à concevoir rapidement les usines cellulaires et à augmenter leur efficacité. Les efforts d'ingénierie métabolique se concentrent donc sur le développement d'outils et d'approches visant à optimiser l'efficacité des usines cellulaires microbiennes.

Une stratégie développée par les ingénieurs métaboliques concerne la colocalisation des enzymes de la voie de biosynthèse pour contrôler la concentration locale en enzymes et en intermédiaires réactionnels afin d'optimiser le flux de la voie. Il est nécessaire de comprendre les mécanismes et les paramètres clés à maîtriser pour profiter au maximum de l'efficacité de cette stratégie. La voie de biosynthèse pour la production de β -carotène est un modèle de voie de biosynthèse hétérologue implémenté chez *S. cerevisiae*. Dans cette voie se succède la biosynthèse des caroténoïdes phytoène, phytofluène, ζ -carotène, neurosporène, lycopène, γ -carotène puis le β -carotène. Ce travail de thèse a cherché à implémenter des stratégies de colocalisation d'enzymes de la voie du β -carotène pour comprendre les possibilités offertes par cette stratégie.

L'introduction du manuscrit de thèse présentera la biologie de synthèse avec notamment les outils qu'elle apporte à l'ingénierie métabolique. Différents axes d'approche pour améliorer les souches de production conçues en ingénierie métabolique seront ensuite exposés. Cela sera suivi de la présentation des stratégies possibles de rapprochement d'enzymes en ingénierie enzymatique. Enfin, la voie de biosynthèse des carotènes et les résultats en ingénierie métabolique sur cette voie seront exposés. Après avoir décrit le matériel et méthodes utilisés durant ce travail, les résultats obtenus seront ensuite décrits suivis d'une partie discussion et perspectives suite à ce travail de thèse.

I-1 Les apports de l'ingénierie métabolique

Face à la raréfaction des sources d'énergie fossiles et à la nécessité de préserver notre planète, nous avons besoin de développer des alternatives efficaces aux méthodes traditionnelles qui utilisent des sources d'énergies fossiles. Une transition vers une production durable, stable et efficace de produits chimiques est donc nécessaire et cela est possible (Keasling, 2010, 2012; Liu et al., 2017). La production de molécules d'intérêt par voie microbienne présente plusieurs avantages (tableau 1).

Attractivités	Avantages	Défis
Écologique et potentiellement bénéfique pour l'environnement (utilisation direct du CO ₂)	Alternatif à l'utilisation du pétrole, à la synthèse chimique et à l'extraction à partir de plante Condition de production écologique Réutilisation du biocatalyseur plusieurs fois Croissance rapide dans un milieu de culture simple et non polluant Produits naturels	Obtenir des usines cellulaires efficaces en production (titre, taux de production, rendement) Créer des souches capables d'utiliser le CO ₂
Production de molécules complexes, d'énantiomère biologiquement actif. Obtention de molécules avec des structures complexes	Haute efficacité catalytique des enzymes, auto régénération des cofacteurs, cascade enzymatique Capacité à produire une large gamme de molécules (même non naturels) en utilisant et en créant les enzymes adéquates	Optimisation de la voie de biosynthèse pour augmenter le flux de production, diminuer les produits non désirés, créer de nouvelles voies de biosynthèse, créer de nouvelles enzymes
Capacité de production optimisable	Augmentation rapide de la population bactérienne productrice en milieu de fermentation contrôlé Grand volume de production	Avoir des enzymes qui agissent de manière optimale au sein des voies de biosynthèse Optimisation de la capacité de la souche à croître à partir de la biomasse végétale, à tolérer les produits toxiques Etablir des souches robustes pour les conditions de fermentation

Tableau 1 : Les attractivités pour la bioproduction par voie microbienne et les challenges associés.

Les microorganismes sont utilisés traditionnellement par l'homme pour la production de produits comme le pain, le vin, la bière, le yogourt entre autre. L'utilisation de microorganismes pour la bioproduction de molécules d'intérêt existe depuis plus de cent ans (Ligon, 2004). Par exemple, le genre de champignon *Penicillium* produit naturellement la pénicilline. Ce producteur naturel a été optimisé et industrialisé par l'homme depuis plusieurs décennies pour une bioproduction à grande échelle. Des microorganismes modèles tels que *Saccharomyces cerevisiae*, *Escherichia coli*, *Corynebacterium glutamicum*, *Bacillus subtilis*, *Aspergillus niger* sont utilisés depuis longtemps pour la recherche en

laboratoire, et peuvent également être utilisés pour produire des substances d'intérêt. Ces microorganismes sont utilisés dans la production de molécules d'intérêt pharmaceutique comme des antibiotiques, des hormones ou des agents anticancéreux... Plus récemment, les microorganismes sont également utilisés pour produire des biocarburants et différentes molécules chimiques (Figure 1).

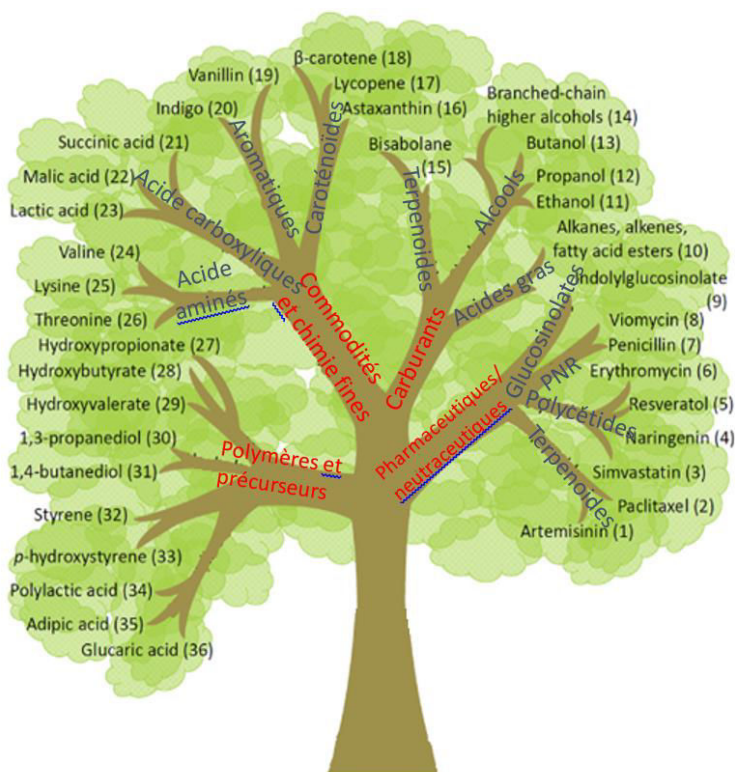


Figure 1 : Exemples de molécule obtenues par ingénierie métabolique décrits dans la littérature. Tirée de (Curran and Alper, 2012). En rouge sont écrits différentes catégories d'utilisation des molécules, en bleu les classes de molécule et en noir des molécules produites par ingénierie métabolique.

L'ingénierie métabolique concerne la manipulation de voies de biosynthèse endogènes d'un microorganisme, la conception de nouvelles voies de biosynthèse, la construction et de l'optimisation des voies de biosynthèse de produits naturels et non naturels au sein du microorganisme. Des activités hétérologues, c'est-à-dire non natif au microorganisme, peuvent être apportées au sein de celui-ci dans l'objectif d'améliorer les propriétés cellulaires nécessaires pour des besoins industriels. Aujourd'hui, l'ingénierie métabolique est un domaine scientifique hautement multidisciplinaire, avec de nombreux excellents articles de revue et des livres publiés (Kholodenko and Westerhoff, 2004; Stephanopoulos et al., 1998).

Malgré d'énormes avancées en ingénierie métabolique pour le développement d'une « usine cellulaire », la synthèse chimique conventionnelle et les produits issus du pétrole restent de loin aujourd'hui le socle des industries chimiques et pharmaceutiques. L'efficacité de production des usines cellulaires n'est souvent pas assez élevée pour avoir une production économiquement viable. Les raisons pour lesquelles l'industrie ne migre pas totalement vers la bioproduction sont les coûts élevés et le temps nécessaire pour le développement d'une « usine cellulaire » efficace (figure 2). Les processus commercialisés

concernent souvent encore actuellement la production de métabolites natifs (acides aminés, acides organiques, acides gras, alcools, antibiotiques...). Dans ces exemples, l'hôte possède une bonne aptitude à tolérer une concentration élevée du produit natif (Van Dien, 2013). Pour les productions de métabolites hétérologues, tels que les biocarburants, les produits d'intérêt pharmaceutique, les biopolymères, les précurseurs ou produits de la chimie fine, d'avantage d'efforts et de technologies sont donc à mettre en œuvre pour atteindre les niveaux métriques nécessaires (figure 2). La montée en échelle depuis la preuve de concept de laboratoire à la souche productrice en usine demande beaucoup plus de temps et de moyens financiers comme le montre l'exemple du développement du bioprocédé pour l'artémisinine. Ce dernier a nécessité 50 millions de dollars, et a impliqué le travail annuel de 150 personnes pendant 10 ans, (Keasling, 2012). La majorité des bioproductions par fermentation microbienne reste donc souvent au stade de l'échelle de laboratoire (Sun and Alper, 2015; Van Dien, 2013).

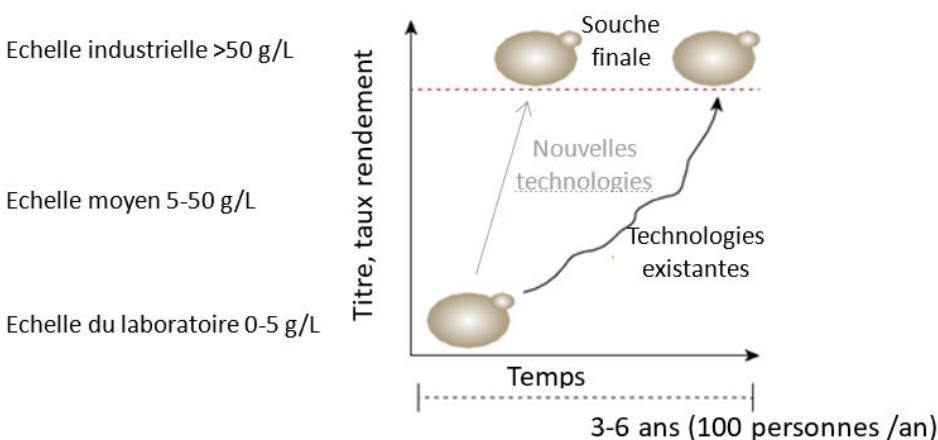


Figure 2 : Le développement de nouvelles technologies accélère l'industrialisation. De nouvelles technologies sont développées pour diminuer le temps, l'argent et l'effectif humain nécessaires pour l'obtention de souches de production atteignant des niveaux métriques industriellement viables. Le rendement est le rapport entre ce que produit le microorganisme et ce qu'elle consomme (produit/substrat). Le titre de production mesure la quantité de produit durant un temps donné (produit/temps). Le taux de production mesure la quantité de produit pour une quantité de producteur (produit/producteur).

Pour accélérer le processus de développement des souches industrielles, l'augmentation de la productivité, la génération de nouveaux bioproduits ou l'utilisation des matières premières alternatives par les microorganismes producteurs, il a été nécessaire d'intégrer des approches complémentaires pour manipuler le métabolisme cellulaire et les réseaux de régulation. Durant les deux décennies suivant le début de l'ingénierie métabolique (Bailey, 1991), une période appelée «ère post-génomique», un volume important de données sur les gènes, les protéines et les métabolites a été accumulé, et les coûts de la synthèse des oligonucléotides ont diminué de façon exponentielle. Le développement de techniques à haut débit pour déchiffrer des génomes, des transcriptomes, des protéomes, des métabolomes et des fluxomes, avec des outils de calcul, ont permis d'accumuler des données sur le comportement global de différentes espèces de microorganismes, dans des conditions variées. Grâce à ces données descriptives globales, l'ingénierie métabolique s'est progressivement éloignée des approches réduites de perturbation de voies individuelles pour commencer à manipuler la totalité du métabolisme cellulaire (Lee and Kim, 2015). L'ingénierie métabolique des systèmes étend le concept de

l'ingénierie métabolique classique (localisé) au niveau des systèmes (considérant la cellule en entière) en incorporant les techniques utilisées en biologie des systèmes (Chae et al., 2017), en biologie synthétique (Stephanopoulos, 2012), en ingénierie enzymatique (Eriksen et al., 2014), et en ingénierie évolutive (Dragosits and Mattanovich, 2013). Nous allons introduire ces différentes disciplines.

I.1.1 La biologie de synthèse

La biologie synthétique a comme objectif de créer (« synthétiser ») des systèmes biologiques artificiels pour comprendre les systèmes biologiques naturels (Cameron et al., 2014). Elle se réfère à une variété d'approches expérimentales qui cherchent soit à modifier soit à imiter les systèmes biologiques naturels. L'application de la biologie synthétique à la recherche fondamentale suit l'idée que « ce que je ne peux pas créer, je ne le comprends pas » (Richard Feynman prix Nobel de physique). La biologie synthétique cherche à standardiser les éléments constitutifs du vivant pour les réassocier à façon, en suivant les principes issus de l'ingénierie. Elle se différencie de l'ingénierie génétique par une approche de construction de systèmes biologiques en combinant des modules biologiques simples pour tendre vers des systèmes biologiques de plus en plus complexes suivant des cycles itératifs de modélisation par ordinateur, d'assemblage biologique, de test et d'apprentissage. On a cherché à codifier ces cycles itératifs en incorporant des principes d'ingénierie à la biologie tels que la modularité, la standardisation et l'abstraction (Andrianantoandro et al., 2006; Endy, 2005). En biologie synthétique, des éléments moléculaires classés en « parties » constituent le plus bas niveau de la hiérarchie. Les « parties » peuvent être utilisées pour construire des dispositifs de pièces assemblées produisant une fonction souhaitée. Ces dernières peuvent à leur tour être combinées au sein de systèmes biologiques. Les systèmes biologiques peuvent avoir des caractéristiques modulaires, à différentes échelles : l'ADN, l'ARN, les protéines, les voies métaboliques, les cellules entières et consortia microbien. Ils sont modulaires car ont des fonctions séparables (discrètes) et peuvent être réorientés et réorganisés dans de nombreux contextes différents tout en conservant une fidélité de fonction. D'autre part, la standardisation consiste en l'utilisation des blocs de construction avec des propriétés et des modes de connectivité bien définis, ce qui est important en ingénierie. La biologie synthétique travaille à identifier, caractériser, modifier, voire créer puis réutiliser ces modules biologiques. Elle les utilise ensuite pour construire des dispositifs dotés de nouvelles fonctions utiles et pour étudier des systèmes artificiels qui peuvent reproduire des caractéristiques des systèmes naturels (oscillateurs et commutateurs synthétiques, systèmes d'interaction artificielle entre cellule, ingénierie de la transduction de signal, ingénierie des voies métaboliques...). La gamme existante de modules biologiques peut être complétée en modifiant les modules naturels ou en créant de nouveaux modules lorsque les modules naturels ne sont pas adéquats pour un objectif donné. La construction des modules a été facilitée par le développement des technologies de synthèse d'ADN qui fut un catalyseur de la biologie synthétique. Enfin, la notion d'abstraction consiste à ignorer, dans les systèmes complexes composés de nombreux composants, le détail du fonctionnement de ses constituants (car bien caractérisés) pour faciliter l'ingénierie. En développant des outils et des méthodes pour la conception, la construction et la caractérisation de systèmes biologiques, la biologie synthétique est devenue une plateforme technologique pour d'autres domaines (Kelwick et al., 2014; Lu et al., 2009; Purnick and Weiss, 2009; Drubin et al., 2007). Des outils de biologie synthétique permettent par exemple de réduire le temps de développement d'« usine cellulaire ». Ces outils permettent d'obtenir des gènes synthétiques par la

technologie de la synthèse de l'ADN, l'assemblage rapide de fragment d'ADN pour le clonage, l'utilisation de CRISPR Cas9 pour la construction des voies de biosynthèse. Des modules caractérisés de biologie synthétique permettent de contrôler avec précision des réactions natives et hétérologues à l'intérieur de la cellule (promoteur, terminateur, riboswitch...). Des outils de biologies synthétiques comme les biosenseurs sont utilisés pour sélectionner les souches de production efficaces (Rogers et al., 2016). La plupart de ces outils ont été développés dans le cadre de la biologie synthétique. Ils permettent la construction d'usines cellulaires efficaces de manière rapide, quantitative et prédictive.

I.1.2 La biologie des systèmes

La biologie des systèmes est un domaine récent de la biologie qui étudie les organismes vivants comme les systèmes qu'ils sont en réalité, par opposition aux approches historiques qui tendent à décomposer l'étude à tous les niveaux. Le développement de la bioinformatique et de la génomique fonctionnelle a non seulement permis de comprendre ou de modifier les systèmes biologiques existants, mais aussi de créer de nouveaux systèmes biologiques à des fins spéciales. Le résultat naturel d'une telle avancée est la biologie synthétique, décrite précédemment, qui traite de la conception et de l'assemblage de parties et systèmes biologiques prévisibles et robustes, et de la biologie des systèmes, qui visent à la compréhension des systèmes biologiques au niveau du système. Développer des outils de biologie synthétique efficaces pour l'ingénierie des souches dépend largement de la modélisation computationnelle qui constitue le noyau de la biologie des systèmes. La biologie des systèmes permet une description quantitative d'un système biologique afin de prédire ses propriétés dynamiques et permettre une conception intuitive des réseaux biologiques synthétiques.

Les avancées de la biologie des systèmes conduisent à une large collection d'outils de calcul qui aident à l'ingénierie rationnelle du métabolisme cellulaire (Fletcher Eugene et al., 2015). La biologie des systèmes s'intéresse à l'intégration et à la synthèse des ensembles de données cellulaires obtenues par des méthodes expérimentales à haut débit ou par la modélisation mathématique, dans le but de comprendre et de modéliser des cellules au niveau global. Lorsqu'il est combiné avec les objectifs d'ingénierie métabolique, le domaine de l'ingénierie métabolique systémique émerge avec le potentiel de comprendre et d'améliorer les systèmes cellulaires (Blazeck et Alper, 2010 ; Lee et al., 2007). Les outils de la modélisation et de la biologie des systèmes sont utiles pour générer la conception initiale de la souche, mais aussi pour comprendre les données expérimentales dans un contexte cellulaire global durant le processus d'amélioration de la souche. Le développement des outils de génomique fonctionnelle tels que le séquençage, la protéomique, la transcriptomique, la métabolomique, et les techniques d'analyse de cellule unique (microfluide) permettent à l'ensemble du paysage cellulaire d'être résolu au plus fin des détails, résultant en une incroyable richesse de données. Un grand défi est celui de synthétiser ces données sous forme d'un modèle cellulaire complet pouvant ensuite être utilisé en ingénierie métabolique afin d'améliorer le titre des voies connues, concevoir des voies de biosynthèse de novo (Wang et al., 2017), comprendre et détourner les réseaux de régulation... Les modèles informatiques intègrent les ensembles de données génomiques, dans le but de représenter le métabolisme cellulaire entièrement in silico. Ces prédictions in silico ont un potentiel considérable car elles permettent un grand nombre de changements combinatoires de gènes à tester in silico, réduisant les interventions humaines pendant le développement de la souche. En analysant les données omiques,

et en utilisant les modèles métaboliques à l'échelle du génome, divers algorithmes d'analyse de flux sous contrainte ont été développés pour prédire des cibles d'altérations génétiques telles que des cibles de knockout ou d'amplification de gènes pouvant améliorer la capacité de production en redirigeant le flux métabolique (Guo et al., 2015; O'Brien et al., 2015).

I.1.3 Evolution adaptative en laboratoire

Une approche non rationnelle cette fois ci en ingénierie métabolique est l'évolution adaptative en laboratoire (Winkler and Kao, 2014). Elle peut générer des systèmes de production microbienne potentiellement robustes et optimisés. Contrairement à l'ingénierie rationnelle et à la modification dirigée d'enzymes spécifiques, l'évolution adaptative en laboratoire a l'avantage de permettre à des mutations bénéfiques, non intuitives, de se produire chez divers gènes en parallèle (Portnoy et al., 2011). Pour cela, la souche candidate est exposée à des conditions sous optimales rencontrées en fermentation industrielle et qui provoque des adaptations génétiques de certaines cellules. Ensuite, grâce au développement rapide des technologies analytiques à haut débit en génomique fonctionnelle (Bro and Nielsen, 2004) ainsi que des technologies omiques, les corrélations entre phénotype et génotype peuvent être obtenues en comparant les études analytiques des souches améliorées et en reséquençant leur génome. Les mutations génétiques bénéfiques identifiées peuvent être trop complexes ou non intuitives pour être obtenues par des approches rationnelles. Ces mutations bénéfiques peuvent ensuite être réintroduites dans la souche parentale pour retrouver le phénotype désiré. Les applications biotechnologiques pour cette méthode comprennent l'activation des voies latentes, l'amélioration de l'utilisation de substrats non natifs ou la production de produits non naturels, l'augmentation de la capacité ou du taux de substrats utilisables, tels que les sucres (Çakar Z. Petek et al., 2011), ainsi que l'amélioration du taux de croissance cellulaire et de titres de produits (Hong et al., 2011). Ainsi que l'adaptation de souches à des conditions environnementales comme l'augmentation de la tolérance au stress qui se produisent couramment dans les conditions de fermentation industrielle (Çakar et al., 2005; Caspeta et al., 2014).

Face à la complexité du vivant et à notre compréhension incomplète du métabolisme microbien et de sa régulation (Endy, 2005; Kwok, 2010), il est difficile de les concevoir et de les contrôler avec précision. La biologie systémique utilise des cycles de conception-construction-test-analyse (design, build, test, learn) que l'on utilise alors aussi pour optimiser les systèmes de production microbiens. Une ingénierie métabolique systémique peut être définie comme l'ingénierie métabolique au niveau des systèmes intégrant les techniques « omiques » et de calcul de la biologie des systèmes (design), les capacités de conception affinée de la biologie synthétique et les méthodes de mutagenèse rationnelle et aléatoire de l'ingénierie par évolution (build). L'efficacité des souches obtenues sont mesurées (test). Les données générées après chaque cycle d'optimisation devraient être analysées bioinformatiquement. Puis avec l'aide de modèles mathématiques couplés à la grande masse de données présentent par exemple dans la littérature scientifique ou les bases de données, une conception plus rationnelle des propriétés des cellules ou des processus de production peut être envisagée (learn) (Koutinas et al., 2012; Almquist et al., 2014).

I-2 Les différentes étapes de construction d'une usine cellulaire

Il existe diverses possibilités d'optimisation d'une souche de production. Les faibles efficacités de production d'un microorganisme peuvent avoir plusieurs origines et constituent autant de cibles génétiques à modifier pour augmenter l'efficacité de production d'une molécule donnée par un microorganisme (Paddon and Keasling, 2014) (Figure 3).

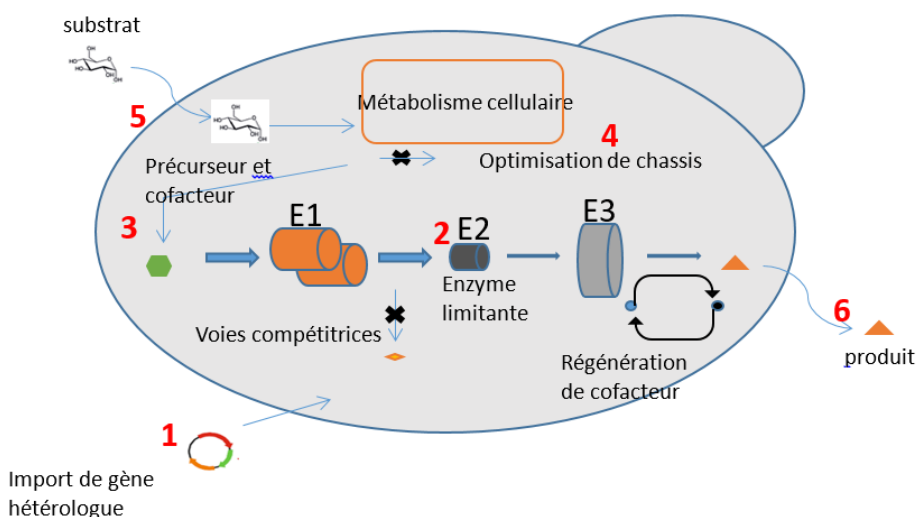


Figure 3 : Différentes cibles pour améliorer les capacités d'une « usine cellulaire ». (1) un assemblage efficace et rapide de la voie de biosynthèse dans l'hôte (2) une optimisation de la voie de biosynthèse (3) une optimisation de la cellule hôte pour l'approvisionnement efficace en précurseurs, en cofacteurs, et en énergie pour le processus (4) la création de souches robustes adaptées aux conditions de fermentation (5) développement de souche capable d'utiliser des substrats peu onéreux comme la biomasse végétale ou (6) d'exporter le produit.

I.2.1 Choix de l'hôte de production

Choisir le type d'hôte cellulaire permet de minimiser les efforts d'ajustement métabolique (Fisher et al., 2014). Certains hôtes possèdent de meilleures capacités métaboliques pour la production d'une catégorie donnée de produit. Par exemples, un producteur naturel possède la voie des précurseurs du produit ciblé. Certains hôtes cellulaires possèdent une grande surface membranaire idéale pour la production de produits liposolubles. *S. cerevisiae* présente des avantages pour les voies de biosynthèse impliquant des cytochromes P450. Des outils de biologie des systèmes pour la modélisation et la simulation métaboliques in silico à l'échelle du génome aident au choix d'un organisme pour la production d'une molécule donnée (Purdy and Reed, 2017). L'utilisation d'organismes industriels modèles tels que *E. coli* et *S. cerevisiae* donnent plusieurs avantages (van Dijken et al., 2000; Hong and Nielsen, 2012; Liu et al., 2013b). Les outils génétiques pour manipuler ces organismes modèles sont par exemple nombreux et performants. Bien qu'un organisme producteur naturel puisse posséder différentes capacités innées pour la production d'une molécule, des réseaux de régulations cellulaires naturelles existent pour limiter la production de ces molécules. Les microorganismes ne sont en effet pas programmés pour produire une molécule aux fortes concentrations recherchées par l'industrie. Au contraire, la production naturelle répond à un contexte environnemental précis et est régulée et

sporadique. Un vrai challenge est celui de contourner ces mécanismes pour obtenir un flux optimal. Parfois, il peut être préférable de choisir une cellule hôte hétérologue dans laquelle ces mécanismes de régulation sont probablement absents.

1.2.2 Design et construction de la voie

Nous allons voir des méthodes pour la conception et la construction des voies de biosynthèse (sélection des enzymes constituant la voie, régulations et équilibres). Avec le développement rapide de la technologie de séquençage, ainsi que des analyses « omique » et des technologies de criblage à haut débit, des milliers de gènes codant pour des isoformes enzymatiques impliqués dans des voies de biosynthèse d'intérêt ont été identifiés ces dernières années (Medema and Osbourn, 2016). Le génome mining ou le métagénome permettent d'identifier des enzymes candidates (Shin et al., 2013). Des outils de conception assistée par ordinateur (CAO) ont été développés pour générer des voies hétérologues in silico (Fernández-Castané et al., 2014). Ces derniers permettent d'évaluer la faisabilité de diverses voies basées sur des critères comme l'utilisation maximale de réactions natives, un nombre minimal d'étapes, la maximisation du rendement, la balance en cofacteur, la faisabilité thermodynamique... (Long et al., 2015). De plus, lorsque certaines enzymes de la voie de biosynthèse ne sont pas identifiées, ou lorsque de nouvelles structures de produits chimiques sont à obtenir, l'utilisation d'outils de conception computationnelle de protéines est utile pour évoluer ou même repenser des enzymes en modifiant par exemple leur spécificité de substrat. Cela permet ainsi de compléter de nouvelles voies de biosynthèse (Eriksen et al., 2014; Pleiss, 2011).

Une grande variété de techniques permet d'implémenter une voie de biosynthèse dans l'organisme hôte. Des outils de biologie synthétique améliorent la flexibilité, la vitesse et la précision du clonage des gènes dans les vecteurs d'expression. Certains sont basés sur l'approche classique restriction/ligation, d'autres utilisent des recombinases ciblant des séquences homologues et d'autres par l'utilisation d'oligonucléotides (Chao et al., 2015; Li and Borodina, 2015). Plus récemment, le clonage de plusieurs fragments d'ADN indépendamment de site de restriction peut se faire par exemple in vitro, avec des kits commerciaux (Throop and LaBaer, 2015) ou in vivo, par la méthode DNA assembler (Shao and Zhao, 2012). Le kit In-Fusion HD Cloning permet de cloner de manière directionnelle un ou plusieurs fragments dans n'importe quel vecteur en utilisant 15 pb d'homologie (tableau 2).

Méthodes d'assemblage	Taille des fragments homologues	Nombre typique de fragments assemblés	Taille de construction (kb) démontrée	Efficacité	Références
Gibson	40 pb	≈ 4	900	90% (pour 3 fragments)	(Gibson et al., 2009)
In-fusion	>15 pb	2-3	≈ 5	>60% (pour 2 fragments) <40% (pour 3 fragments)	(Sleight et al., 2010)
USER	7-12 pb, doit contenir une base d'uracile	3-7	≈ 8	>90% (pour 7 fragments)	(Jensen et al., 2014) (Lund et al., 2014)
CPEC	15-25 pb	>4	≈ 8	95-100%	(Quan and Tian, 2009)
MoClo	4 pb et un site de reconnaissance pour l'enzyme de restriction de type IIS	≈ 10	33 (en trois fois)	>90%	(Engler et al., 2009) (Weber et al., 2011)
LCR	Les fragments doivent être 5'-phosphorillés, des oligos de liaison de 60 à 90 pb	>10	20	>90% (jusqu'à 6 fragments) ≈ 75% (12 fragments)	(Kok et al., 2014)
Recombinaison in vivo	>40 pb	>10	>20	>90% (jusqu'à 6 fragments) ≈ 75% (12 fragments)	(Kok et al., 2014)

Tableau 2 : Descriptif de différentes méthodes d'assemblage d'ADN par recombinaison homologue. d'après (Li and Borodina, 2015)

L'expression des gènes d'enzymes se fait sur plasmides ou par intégration des gènes dans le génome. Des séries de vecteurs pré-implémentés pour des applications biotechnologiques ont été développées avec un polylinker flanqué d'un promoteur constitutif ou inductible et un terminateur transcriptionnel. Des séries de vecteurs sont aussi disponibles pour l'expression de plusieurs gènes sur un même plasmide (Ishii et al., 2014). Des vecteurs ont été construits pour permettre un recyclage des marqueurs de sélection. Plusieurs séries de plasmides centromériques, épisomaux (maintenus extrachromosomiquement) ou intégratifs existent chez *S. cerevisiae* avec plusieurs choix de marqueurs auxotrophiques ou dominants. Les systèmes plasmidiques multicopies permettent d'introduire les gènes d'intérêt dans les cellules. Cela présente néanmoins des désavantages tels que la charge métabolique pour le maintien des plasmides (Pitera et al., 2007). Les variations de ségrégations plasmidiques créées aussi une hétérogénéité de la population cellulaire ne facilitant pas l'analyse de résultats d'ingénierie.

L'approche par expression sur plasmide multicopie permet par exemple une expression élevée et donc de vérifier la fonctionnalité d'enzymes avec une activité difficilement détectable. L'utilisation d'un vecteur monocopie permet quant à lui d'éliminer les problèmes de ségrégation de plasmide et permet aussi le transfert rapide d'une voie de biosynthèse, assemblée sur un même plasmide, d'une souche à une autre si nécessaire dans une étude. Des chromosomes artificiels de levure (YACS) peuvent être utiles pour introduire plusieurs gènes simultanément dans le cas de longues voies de biosynthèse en raison de leur capacité à porter de grands fragments d'ADN (Naesby et al., 2009). Une intégration ciblée des gènes dans le génome chez *S. cerevisiae* permet une expression plus stable. L'intégration génomique s'effectue par recombinaison homologue de fragments d'ADN contenant les régions d'homologie (60 à 500pb), la modification souhaitée, et un marqueur de sélection. L'insert peut être obtenu par PCR ou par un plasmide intégratif. Certains sites d'intégrations dans le génome comme ceux proches de gènes importants par exemple ont été caractérisés comme stables, c'est-à-dire que les gènes d'intérêt introduits au niveau de ces régions sont moins susceptibles d'être perdus par des événements de recombinaisons non désirés. Pour accélérer la construction de souches productrices, des plasmides bidirectionnels existent pour insérer deux gènes en une intégration et permettent donc d'insérer 4 gènes en recyclant une seule fois la cassette de sélection. Des cycles itératifs d'insertion et de recyclage du marqueur peuvent être réalisés grâce à des outils tels que la recombinaison LoxP-CreA (Fang et al., 2011; Güldener et al., 1996; Xie et al., 2014). L'expression hétérologue de recombinases permet d'éliminer le fragment d'ADN entre deux sites de reconnaissance présents initialement sur le plasmide. Une copie du site de reconnaissance est laissée comme cicatrice dans le génome, pouvant conduire à des réarrangements chromosomiques lorsque la même recombinase est utilisée de nombreuses fois. Des intégrations à des sites éloignés du génome limitent ces événements de recombinaison non désirés (Xie et al., 2014). Une approche différente pour intégrer les voies multigéniques consiste en leur assemblage par recombinaison homologue in vivo sur un plasmide ou directement par intégration simultanée des différents fragments d'ADN au niveau d'un même site dans le génome (Shao et al., 2009). Cette stratégie, entièrement basée sur la PCR, a une efficacité de 10 à 70% de souche correctement construite en utilisant les sites δ . Les promoteurs et les terminateurs doivent cependant être différents pour chaque gène afin d'éviter des recombinaisons non désirées. Les fragments d'ADN obtenus par PCR ne sont pas vérifiés par séquençage lorsque la construction est intégrée dans le génome. Mais l'efficacité de la méthode permet la création rapide de banques de constructions. Des technologies récentes permettent d'augmenter l'efficacité d'intégration dans le génome. Enfin, l'efficacité d'intégration est grandement améliorée par l'introduction d'une coupure double brin au niveau du site d'intégration grâce à l'activation de la machinerie de réparation d'ADN (Carroll, 2004; Kuijpers et al., 2013). Pour obtenir ces coupures, des endonucléases telles que des méganucléases sont utilisées. Les méganucléases nécessitent l'intégration au préalable d'un site de reconnaissance de coupure dans le chromosome, ou bien elles peuvent être ingénierées pour reconnaître une séquence choisie. Les méganucléases telles que TALEN et ZFN reconnaissent des sites de coupure grâce à des domaines protéiques de liaisons à l'ADN que l'on peut manipuler pour reconnaître un site particulier (Beurdeley et al., 2013). D'autres nucléases, comme CRISPR/Cas9, sont des outils encore plus flexibles pour cliver sur mesure presque tous les sites du génome et ainsi éditer efficacement le génome. Cet outil est basé sur le système bactérien CRISPR de type-II de *Streptococcus pyogenes*, et utilise la nucléase Cas9 et un complexe d'ARN, synthétisable artificiellement à façon et à moindre coût, servant de guide vers une

séquence d'ADN spécifique, qui est alors clivé. CRISPR/Cas9 peut accélérer considérablement la construction de voie de biosynthèse. Des variantes permettent l'insertion de plusieurs gènes à différents locus en une seule étape de transformation, sans l'utilisation de marqueur de sélection et sans laisser de cicatrice dans le génome (Ronda et al., 2015; Shi et al., 2016; Wang et al., 2016).

1.2.3 Optimisation d'une voie de biosynthèse

L'introduction au sein d'un microorganisme des enzymes d'une voie de biosynthèse ne garantit pas un fonctionnement optimal de la voie de biosynthèse. Une fois qu'une souche hôte est capable de produire un produit chimique, la prochaine étape consiste donc à améliorer la production du produit cible par optimisation de la voie de biosynthèse. Un déséquilibre entre les activités enzymatiques entraîne inévitablement l'accumulation d'intermédiaires réactionnels. On cherche à limiter l'accumulation des intermédiaires dans les cellules pour plusieurs raisons. Les produits étant dérivés d'une source de carbone, typiquement le glucose, l'accumulation d'intermédiaires constitue une perte d'efficacité carbone. Un maximum de carbone utilisé devrait être converti en produit. Des études antérieures ont montré que l'accumulation d'intermédiaires peut être toxique pour la cellule et a un impact élevé sur le rendement des voies (Keasling, 2010; Martin et al., 2003). L'accumulation d'intermédiaires favorise leur consommation par des enzymes concurrentielles de la cellule. Certains intermédiaires peu stables se dégradent rapidement et devraient donc être rapidement biotransformés. Enfin, l'accumulation d'intermédiaires complique les processus d'extraction finaux.

L'efficacité de la voie de biosynthèse est mesurée par la circulation de matière au sein de la voie appelée le flux. Deux considérations importantes pour le flux de la voie sont la concentration en substrat (l'apport en substrat) et en produits (l'enlèvement du produit) dans la cellule, et les niveaux d'activité enzymatique, plus précisément celle de l'enzyme avec la plus faible activité (figure 4). La vitesse d'une réaction métabolique dépend de plusieurs variables telles que les concentrations en enzyme, en substrat, en produit et les molécules régulatrices ($v = v(ce, cs, cp, cr)$). Comprendre les régulations au niveau local (ce, cs, cp) et au niveau global est nécessaire pour contrôler les flux à travers la voie. La concentration des métabolites est donc fonction de la thermodynamique et de la cinétique des réactions. Le flux est une résultante de ces deux paramètres. In vivo, les réactions enzymatiques sont modulées par différents mécanismes comme la modulation de l'activité enzymatique par des régulations allostériques telles que des inhibitions, des contrôles par inhibition ou par activation, par la phosphorylation ou par d'autres modifications post traductionnelles. Enfin, la concentration des enzymes dépend de plusieurs paramètres tels que leur taux de synthèse, leur dégradation, leur inhibition irréversible (figure 4).

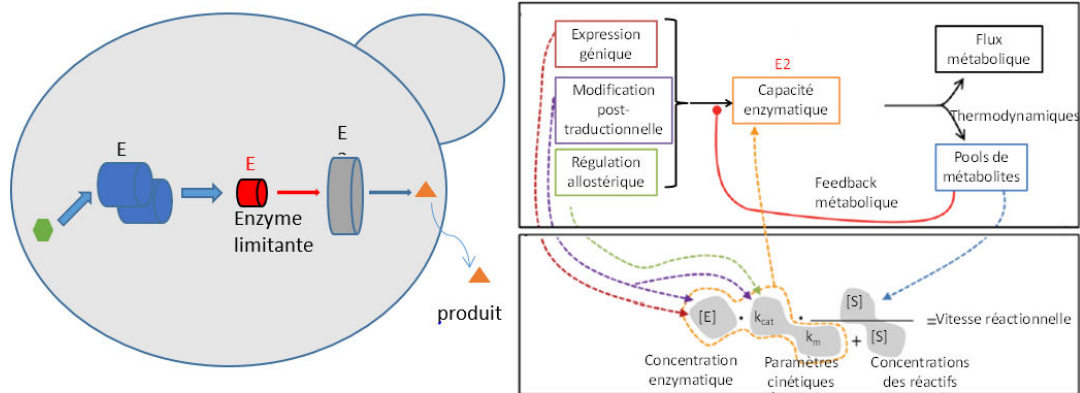


Figure 4 : Représentation des paramètres régulant le flux métabolique. Le flux métabolique dépend de plusieurs paramètres : la capacité enzymatique et les concentrations en métabolites. Les concentrations en métabolites dépendent de la cinétique des réactions enzymatiques et de la thermodynamique. L'enzyme E2 limitante impose le flux maximal de la voie de biosynthèse. La capacité enzymatique de E2 dépend de sa concentration enzymatique, qui est liée au taux d'expression génique (au turnover de la protéine et aux inhibitions irréversibles) ou aux modifications post traductionnelles. Elle dépend aussi de ses paramètres cinétiques qui sont définis par son activité intrinsèque ainsi que par les modifications post- traductionnelles (phosphorylations, ...) et les régulations allostériques (inhibition, rétroinhibition ou rétroactivation).

A partir de là, les stratégies d'optimisation de la voie peuvent être catégorisées en stratégie statique (Jones et al., 2015) et en stratégie dynamique (Cress et al., 2015). Les approches statiques permettent d'implémenter un niveau d'activité enzymatique fixe. Ce dernier n'est pas conçu pour réagir ou s'ajuster en fonction des concentrations intercellulaires en métabolites. Les méthodes d'équilibrage dynamique utilisent elles des biosenseurs pour détecter et réagir aux changements concentrations de métabolites intracellulaires pour par exemple augmenter le taux d'expression d'un gène donné (figure 5).

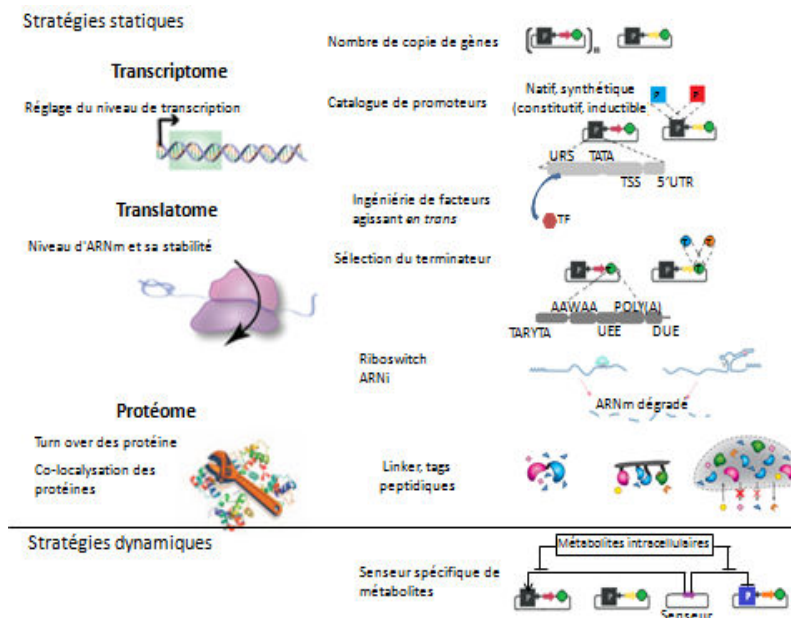


Figure 5 : Représentation des différentes approches pour équilibrer les activités enzymatiques dans des voies de biosynthèse. Du haut vers le bas : approches par des stratégies statiques au niveau du transcriptome, du translatome et du protéome. Approches par des stratégies dynamiques.

1.2.3 1 Ingénierie au niveau du transcriptome et du translatome

La première stratégie utilisée pour l'optimisation de la voie est l'équilibrage par le nombre de copies de gènes (Da Silva and Srikrishnan, 2012). Elle consiste par exemple en l'utilisation de plasmides avec des nombres de copie différents pour chaque enzyme afin d'équilibrer leur niveau d'activité (Lian et al., 2016). Cela est idéal par exemple pour des gènes pour lesquels un niveau d'expression très élevé est nécessaire pour avoir une activité enzymatique adéquate (Liao et al., 2012). Pourtant, à chaque division cellulaire, le nombre de plasmides par cellule fille est souvent différent. In fine, cela peut rendre difficile l'analyse des résultats. En intégrant plusieurs fois un gène par des cycles d'intégration successifs, on peut aussi varier le nombre de copies de gène. Cela prend plus de temps expérimental mais le nombre de copie de gènes par cellule est connu et est fixe. Le nombre d'intégration maximal est limité car cela augmente la probabilité d'événements de recombinaison de séquence homologue *in vivo* non désirés.

Les ingénieurs métaboliques peuvent réguler la force de l'expression des gènes de la voie afin de varier le niveau de l'ARN produit dans les cellules. Cette régulation au niveau de la transcription des gènes se fait par le choix des promoteurs utilisés (Du et al., 2012). La biologie de synthèse fournit divers promoteurs synthétiques (Curran et al., 2014; Leavitt and Alper, 2015). Des promoteurs constitutifs ou inductibles natifs offrent un large éventail de niveaux d'expression (Blazeck et al., 2012; Da Silva and Srikrishnan, 2012; Xie et al., 2015). Des bibliothèques de promoteurs synthétiques ont été développées (Curran et al., 2014) avec des niveaux d'expression élevés et une gamme d'expression pouvant varier de 200 fois (Portela et al., 2017). L'utilisation de facteurs agissant *in trans* sur les promoteurs permet aussi de réguler le niveau de transcription des gènes. De larges gammes de niveaux de transcription ont été obtenus par l'ingénierie de facteurs de transcription artificiels et de régions *cis* régulatrices synthétiques de promoteur leur correspondant (Naseri et al., 2017). Aussi, le choix d'un terminateur impacte sur le niveau et la stabilité de l'ARNm (Curran et al., 2013). Enfin le niveau d'ARNm peut être contrôlé post-transcriptionnellement par l'utilisation de l'interférence ARN (Crook et al., 2014) ou par l'introduction de séquence riboswitch en 3'UTR de l'ARNm (McKeague et al., 2016). Des approches computationnelles peuvent aider à comprendre les mécanismes influençant l'expression de gènes (Singh et al., 2016).

Le niveau d'expression génique peut ne pas corrélérer avec le niveau d'activité protéique attendu. Des événements posttranscriptionnels peuvent changer le niveau d'activité protéique. Puisque les protéines sont les effecteurs finaux du processus biologique, une régulation de la quantité de protéines constitue un autre niveau de régulation.

1.2.3 2 Ingénierie au niveau du protéome

Certains goulots d'étranglement dans l'expression hétérologue d'enzyme peuvent être attribués à des facteurs intrinsèques de l'hôte, tels que la vitesse de traduction, l'abondance des ARNt et les protéines chaperonnes ou la phase de croissance durant laquelle la protéine est produite (Plotkin and Kudla, 2011). Plusieurs variables peuvent affecter la traduction d'un gène, telles que la stabilité de l'ARNm, la vitesse de traduction et la précision, le repliement co-translationnel, et la localisation co-translationnelle. Il est par exemple souvent nécessaire d'adapter les codons synthétiques pour une expression protéique correcte chez un hôte donné (Lanza et al., 2014). L'interdépendance entre l'utilisation des codons, le repliement co-translationnel, et l'intégrité des protéines peut grandement affecter l'activité catalytique de protéines hétérologues (Agashe et al., 2013; Kimchi-Sarfaty et al., 2007; Zhou et al., 2013). Les

activités enzymatiques au sein de la voie de biosynthèse dépendent du turnover des enzymes. Le turnover des enzymes peut être modulé par l'utilisation d'étiquettes peptidiques contrôlant leur taux de dégradation et permettant ainsi de réguler la quantité de protéine (Papagiannakis et al., 2017).

I.2.3.2.1 Ingénierie protéique pour la colocalisation d'enzymes

Puisque les concentrations locales en réactifs, en intermédiaires et en enzymes dictent les taux et les rendements des réactions enzymatiques, contrôler leur concentration locale peut être un autre moyen d'optimiser le flux de la voie. Le transfert efficace et rapide des intermédiaires réactionnels d'une enzyme à une autre pourrait optimiser le flux de la voie. Par une colocalisation spatiale des enzymes, le flux de réaction peut être amélioré. Cette stratégie augmente la concentration locale des enzymes et des métabolites, minimise les réactions croisées (Siu et al., 2015) et réduit les limitations par transfert de masse (Lee et al., 2012). Différentes stratégies d'ingénierie protéique permettent de rapprocher spatialement les enzymes d'une voie de biosynthèse (Lee et al., 2012; Zhang, 2011). Chaque stratégie présente des avantages et des inconvénients. Cette approche de colocalisation d'enzyme, sur laquelle est basée ce travail de thèse sera présentée en détail plus loin dans le manuscrit.

A l'opposé du concept de rapprochement spatial des enzymes, l'isolement spatial des enzymes d'une voie de biosynthèse des autres composants cellulaires peut être bénéfique. Pour cela, le ciblage des enzymes dans des compartiments subcellulaires (Chen and Silver, 2012; Hammer and Avalos, 2017) par ajout d'une séquence d'adressage au niveau du gène protéique peut fournir un environnement optimal à la voie de biosynthèse. Cet environnement optimal peut consister en une concentration accrue en cofacteurs ou en précurseurs, au confinement d'intermédiaires toxiques du reste de la cellule, à des concentrations locales en enzyme supérieures au sein du compartiment, à un potentiel redox favorable, à l'absence d'interférence avec des voies de biosynthèses concurrentes... (Hammer and Avalos, 2017; Huttanus Herbert M. and Feng Xueyang, 2017).

I.2.3.2.2 Ingénierie de l'activité enzymatique ou utilisation d'enzymes homologues.

L'enzyme avec l'activité catalytique la plus faible de la voie définit le flux maximal de cette dernière. C'est l'étape limitante de la voie. Elle peut être améliorée en remplaçant l'enzyme limitante par un homologue avec une activité plus élevée ou en améliorant les propriétés catalytiques de l'enzyme par ingénierie protéique (Chen Zhen and Zeng An-Ping, 2013; Marcheschi et al., 2013; Pirie et al., 2013). Le criblage combinatoire de plusieurs homologues de gènes provenant de différentes sources peut être réalisé (Kim et al., 2013; Santos et al., 2011; Thodey et al., 2014). Le coût décroissant de la synthèse d'ADN a considérablement réduit le coût de cette démarche et l'optimisation des codons pour ajuster l'utilisation des codons à celle de l'hôte peut améliorer l'expression d'enzymes hétérologues. L'amélioration de l'efficacité catalytique d'une enzyme permet de ne pas introduire une charge d'expression supplémentaire à l'hôte microbien comme dans le cas des stratégies d'optimisation d'expression. Des exemples d'application récentes d'ingénierie enzymatique pour l'ingénierie métabolique concernent entre autre l'élimination d'inhibition allostérique native, la mise en place de régulation allostérique non naturelle, l'ingénierie des protéines pour modifier le pH préférentiel d'une enzyme, la suppression de sous-produits non ciblés en améliorant la sélectivité, la génération de nouvelles activités enzymatiques (Chen and Zeng, 2016). Les approches d'ingénierie enzymatique par des techniques de conception assistée par ordinateur sont généralement longues, dépendantes aussi de

la disponibilité d'un criblage ou d'une sélection et souffrent également de la spécificité de l'enzyme modifiée résultante.

Les stratégies présentées ci-dessus permettent d'améliorer l'efficacité intrinsèque des voies de biosynthèse. Elles ne couvrent pas les mécanismes de régulation dans un contexte métabolique global. Les voies dérégulées, même si localement optimisées, pourraient conduire à une perturbation métabolique non favorable dans le microorganisme hôte, et par conséquent donner des titres et des rendements sous-optimaux du produit désiré. Un contrôle dynamique des niveaux d'enzymes et des activités enzymatiques peut par exemple être utilisé pour changer la concentration de métabolites cibles et le flux de la voie en réponse à l'évolution de l'environnement intracellulaire et extracellulaire (Venayak et al., 2015).

I.2.3.2.3 Contrôle dynamique

Un équilibrage dynamique du flux utilise des mécanismes de régulation pour modifier activement les niveaux ou les activités des enzymes de la voie pendant les phases de croissance et de production (Cress et al., 2015). Un système de contrôle dynamique comprend deux éléments: un élément senseur qui réagit à un métabolite cible (par exemple, un facteur de transcription); et un élément régulateur, dont l'activité est modulé par l'élément senseur, contrôle la production / consommation d'un ou de plusieurs métabolites cibles dans la voie (par exemple, une séquence promotrice / opérateur qui lie le facteur transcription et contrôle à son tour l'expression des enzymes de la voie). Les biosenseurs utilisant les facteurs de transcription (Skjoedt et al., 2016; Xu et al., 2014; Yuan and Ching, 2015) et les biosenseurs à base de riboswitch permettent la détection in vivo pour une régulation active et réagissent à une large gamme de molécules cibles. Une méthode notable pour la découverte de senseur natif utilise la transcriptomique pour détecter les changements dans l'expression des ARNm sur des souches soumises à des concentrations effectrices de métabolites cibles (Shi et al., 2017). Cette stratégie est efficace pour sélectionner des voies de biosynthèses améliorées en permettant un criblage à haut débit et spécifique des réseaux de signalisation cellulaire.

I.2.4 Optimisation du châssis

Notre connaissance de la régulation du métabolisme est limitée et les microorganismes ne sont pas programmés pour synthétiser à grande échelle un produit particulier (Nicolaou et al., 2010). Leur réseau métabolique et de régulation de gènes ainsi que de signalisation ont évolué pour utiliser de manière optimale les ressources disponibles pour la croissance cellulaire. A peu près 1000 réactions anaboliques permettent la synthèse des macromolécules dédiées au fonctionnement cellulaire. 11 molécules intermédiaires du métabolisme primaire et les cofacteurs ATP, NADH et NADPH sont obtenus grâce au métabolisme primaire. La cellule maintient un flux maximum de carbone au sein de son métabolisme primaire par des régulations à plusieurs niveaux pour garantir l'utilisation des métabolites précurseurs pour la biosynthèse des acides aminés, des lipides, des nucléotides et des glucides nécessaires à la croissance cellulaire, l'homéostasie ou l'entretien cellulaire (Fox and Howlett, 2008). Des métabolites secondaires, sont quant à eux produits par la cellule de manière régulée et sporadique en quantité limitée et faible, pour répondre à des conditions environnementales spécifiques. L'introduction d'une voie de biosynthèse hétérologue pour la production d'un composé d'intérêt à partir des métabolites primaires recevra donc par essence un flux limité en carbone. Un défi de l'ingénierie métabolique est de

rediriger un maximum de flux de carbone du métabolisme central vers la production de molécule d'intérêt, tout en limitant les impacts sur la croissance cellulaire qui doit être préservée pour avoir un bon rendement global final (Pandit et al., 2017; Schmitz Alexander C. et al., 2017; Soma et al., 2014). Différentes approches sont possibles.

1.2.4 1 Elimination et réutilisation de circuits de régulation

La régulation du métabolisme central du carbone a évolué pour assurer une production de composants cellulaires équilibrée avec la production et la consommation d'énergie. Cela permet aux cellules de maintenir l'homéostasie métabolique lorsqu'elles sont exposées à des conditions environnementales et nutritionnelles difficiles et limite une accumulation de métabolites au détriment de la croissance cellulaire. Les mécanismes de rétrocontrôle sont donc présents naturellement dans les cellules et constituent des freins au flux vers la voie de biosynthèse. La connaissance de ces mécanismes de rétrocontrôle peut aider à identifier des goulots d'étranglement pour la surproduction d'un produit d'intérêt. Les souches de production qui sont soumises à la mutagenèse pour augmenter la tolérance au produit résultent souvent de l'élimination d'une rétro-inhibition. L'ingénierie de la régulation est de plus en plus effectuée avec les mécanismes de régulation liés à la formation de divers produits qui sont découverts (Alam et al., 2017; Gonçalves et al., 2017; Österlund et al., 2015). Les régulations transcriptionnelles et traductionnelles du métabolisme peuvent, en principe, être explorées de manière exhaustive en utilisant les ensembles de données et les méthodes à haut débit disponibles (Campbell et al., 2017; Moutsatsos and Parker, 2016). Le métabolisme central du carbone est largement régulé à la fois par voie transcriptionnelle (atténuation transcriptionnelle) et par voie post-transcriptionnelle (régulation allostérique). Dans le premier cas, on peut remplacer les promoteurs d'enzymes métaboliques par des promoteurs constitutifs ou faire un knock out du facteur de transcription impliqué. Dans le second cas la protéine régulatrice qui détecte l'accumulation du métabolite peut être modifiée par ingénierie protéique pour désactiver le mécanisme moléculaire de régulation lorsqu'il est connu (Polakowski et al., 1998). Le remplacement d'une enzyme ou d'une voie soumise à une régulation cellulaire par un homologue naturel permet aussi de les contourner. La compréhension des systèmes de régulation permet aussi de les utiliser en faveur de la voie de biosynthèse. Par exemple, de nombreuses voies de synthèse utilisent des méthodes de prévention de l'accumulation d'intermédiaires nocifs en utilisant la régulation des gènes. La production d'acides gras pour le biodiesel a été rendue plus efficace lorsque le régulateur transcriptionnel naturel FapR a été réutilisé pour contrôler dynamiquement les expressions géniques impliquées dans la production et la consommation de malonyl-CoA (Xu et al., 2014). Dans une autre étude, une expérience de microarray a été réalisée pour identifier une liste de promoteurs natifs qui réagissent à l'intermédiaire toxique accumulé, le farnésyl pyrophosphate (FPP). La voie de régulation dynamique a utilisé des promoteurs FPP-sensibles lui empêchant d'atteindre des concentrations inhibitrices de FPP, et a conduit à une croissance cellulaire améliorée et à une augmentation de la production de l'isoprénoïde amorphadiène (Dahl et al., 2013).

1.2.4 2 Réacheminement des flux pour optimiser la disponibilité en cofacteur et en précurseur

L'équilibre énergétique et en cofacteur doivent aussi être pris en compte pour l'optimisation d'une souche de production. La présence d'une voie de biosynthèse au sein d'un microorganisme hôte lui soutire des précurseurs et des cofacteurs initialement dédiés au métabolisme cellulaire. Les cofacteurs

sont des composés non protéiques qui se lient aux protéines et qui sont nécessaires à la fonctionnalité catalytique normale des protéines. L'ATP est la principale source d'énergie dans les cellules. Le NADH et le NADPH sont les transporteurs d'électron les plus importants, et fournissent le pouvoir réducteur à divers processus de biosynthèse. Ils sont essentiels dans les voies de biosynthèses anabolique et sont souvent les facteurs limitants pour l'augmentation du rendement, du titre et du taux de production du produit cible (Wu et al., 2015). Parce que le carbone intracellulaire et la distribution des ressources énergétiques dans les cellules ont été « optimisés » pour les besoins cellulaires, détourner des ressources cellulaires pour la surexpression de voie de biosynthèse, la maintenance de plasmide, et la synthèse du produit peut perturber les processus cellulaires normaux et entraîner par exemple une baisse de croissance (Bu et al., 2017; Heyland et al., 2011; Kok et al., 2012). L'une des tâches difficiles de l'ingénierie métabolique est d'équilibrer les cofacteurs redox et les métabolites énergétiques des cellules hôtes (Dhamankar and Prather, 2011; Hou et al., 2010; Wang et al., 2013). La charge métabolique est définie comme la proportion des ressources cellulaires qui est utilisée pour construire et faire fonctionner la voie de biosynthèse (Wu et al., 2016). Elle doit être minimisée tout en maximisant l'efficacité de production.

1.2.4 3 Ingénierie à l'échelle globale

En plus de l'optimisation du flux interne de la voie de biosynthèse d'intérêt, des stratégies classiques d'ingénierie métabolique consistent à surexprimer les voies menant vers le produit désiré et à supprimer ou à diminuer les voies concurrentes pour assurer un certain flux au sein de la voie de biosynthèse. Alors que des stratégies comme la suppression de voies concurrentes peuvent être prises intuitivement, des méthodes de calcul plus sophistiquées sont nécessaires pour révéler des stratégies moins évidentes pouvant conduire à un flux accru vers la voie d'intérêt (Misra et al., 2013). Des stratégies d'intervention métabolique sont destinées à reconfigurer le métabolisme de l'hôte pour canaliser des flux supplémentaires vers la voie choisie et éliminer les pertes de carbone et de cofacteurs redox vers des produits indésirables. Les modèles métaboliques à l'échelle du génome mènent à la prédiction de plusieurs cibles d'ingénierie pour l'amélioration de la production (Bordbar et al., 2014; Copeland et al., 2012; Kerkhoven et al., 2015). Les progrès dans le séquençage du génome et le développement de nouvelles approches de biologie synthétique et systémique ont conduit à de nouvelles approches de génomiques telles que des systèmes d'ingénierie à haut débit à l'échelle du génome (Csörgő et al., 2016; Lewis et al., 2012; Wang et al., 2016). Des méthodes de modifications par CRISPR / Cas9 peuvent être multiplexées (jusqu'à 5 modifications à la fois) (Jakočiūnas et al., 2015) et permettre l'introduction et la modification de plusieurs étapes des voies ou de nouvelles cibles génomiques à la fois. Des méthodes comme le « trackable multiplex recombineering » augmentent et diminuent systématiquement l'expression de tous les gènes dans le génome, et utilisent la sélection pour identifier les gènes qui influencent le fitness cellulaire (Si et al., 2017). Une fois qu'une combinatoire optimale de knockdown des cibles est identifiée en utilisant une stratégie sur plasmide, les souches industrielles peuvent être développées en intégrant la combinaison dans le chromosome pour une souche plus stable. La transcriptomique (profil d'expression génique global) fournit une variété d'informations concernant la physiologie de la cellule. Le profil d'expression peut identifier des étapes métaboliques clés insuffisamment exprimées ou diminuées au fil du temps. Il peut aussi prédire des imprévus comme des voies concurrentes par lesquelles les gènes sont induits, ou identifier les réponses

cellulaires dues au stress global. D'autres données 'omics' (protéomiques, métabolomiques, fluxomiques) en conjonction avec des outils de modélisation peuvent constituer une approche puissante pour comprendre la physiologie cellulaire tout au long du développement de la souche (Boyle and Gill, 2012; Roldão et al., 2012). L'analyse de flux métabolique de l'isotope ^{13}C combinant une approche expérimentale à une approche computationnelle permet aussi une estimation du flux intracellulaire (Zamboni et al., 2009). Outre les modifications dirigées décrit ci-dessus, les performances de la souche peuvent être améliorées en utilisant l'évolution adaptative en laboratoire (Dragosits and Mattanovich, 2013). Si la production du produit chimique désiré est couplée à la croissance, c'est-à-dire que les cellules doivent produire la molécule chimique lorsqu'elles se développent, on peut lier les améliorations dans la croissance de l'organisme pour sélectionner une amélioration de la production de la molécule désirée. Grâce à la disponibilité de systèmes de séquençage performants et à biologie systémiques, il est possible d'identifier ensuite les mutations responsables des phénotypes souhaitables. Ajouté à l'analyse du transcriptome et / ou du métabolome, on peut ainsi cartographier les mécanismes moléculaires sous-jacents aux phénotypes acquis.

I-3 Colocalisation d'enzymes

I.3.1 Stratégies naturelles pour colocaliser des enzymes

La plupart des réactions cellulaires successives ne sont pas catalysées par des enzymes isolées et solubles. Les enzymes de voies de biosynthèse forment par exemple des complexes multienzymatiques appelés métabolon. Si un certain nombre de métabolons ont été proposés, la preuve qu'ils fonctionnent pour améliorer le flux métabolique est rare. La colocalisation spatiale d'enzymes pourrait offrir plusieurs avantages par rapport aux enzymes isolées solubles (Conrado et al., 2008; Jørgensen et al., 2005). Les sites actifs des enzymes de la voie séquentielle peuvent être rapprochés, l'efficacité catalytique globale est accrue notamment si un transfert direct d'intermédiaires d'une enzyme à l'autre est observé. Lorsque l'intermédiaire réactionnel ne s'équilibre pas dans le milieu extérieur aux sites actifs, on parle de channeling du substrat (Milani et al., 2003; Schmitt and An, 2017; Wheeldon et al., 2016; Wilson Miles, 2001). Les systèmes cellulaires contiennent des concentrations élevées en macromolécules, avec les protéines occupant approximativement 40 % de ce volume. Cela impacte sur la fonction des protéines (Chen and Silver, 2012; Conrado et al., 2008; Kuznetsova et al., 2014; Long et al., 2005). Il a été montré que dans les cellules, cet encombrement macromoléculaire pourrait favoriser les associations enzymatiques et une organisation structurale dans le cytoplasme (Mourão et al., 2014; Spitzer, 2011). La fluidité dans le cytoplasme est alors étroitement liée à l'activité métabolique (Parry et al., 2014) et les environnements *in vivo* peuvent influencer significativement la cinétique enzymatique (Minton, 2001). La nécessité d'une organisation spatiale et temporelle des protéines dans les voies de signalisation et de métabolisme est évidente lorsqu'on examine ce milieu encombré de macromolécules à l'intérieur des cellules ainsi que les nombreuses réactions concurrentes qui se déroulent simultanément (Sweetlove and Fernie, 2013).

La structure globale des métabolons n'est souvent pas définie ou mise en évidence et leur durée de vie est largement inconnue en raison de limitations technologiques. Ces structures complexes sont souvent des assemblages transitoires, s'appuyant sur des changements locaux délicats dans les solutés et les éléments structuraux ou éventuellement sur des protéines d'échafaudage (An et al., 2008; Laursen et al.,

2015). Contrairement à ces complexes protéiques transitoires, le phénomène de channeling a été démontré expérimentalement dans des associations d'enzymes très stables qui présentent un channeling statique. Ces complexes enzymatiques, par exemple la tryptophane synthase bactérienne, ont une constante de dissociation si faible qu'il est possible de déterminer la structure de l'ensemble du complexe par cristallographie aux rayons X (Banik et al., 1995). La formation de métabolons peut également être contrôlée par la liaison au substrat. Par exemple, un changement conformationnel dans les enzymes cytochrome P450 résulte de la liaison au substrat et intervient dans la fixation de la cytochrome P450 réductase (Paine et al., 2005). Etudier la formation de métabolons constitue une thématique de recherche importante (Bassard and Halkier, 2017) car les connaissances obtenues en démêlant les mécanismes de régulation de l'assemblage et de démontage sont cruciales pour comprendre le flux métabolique et pour l'optimisation réussie de voies de biosynthèse de produits naturels.

Approximativement 40% à 65% des protéines procaryotes contiennent plusieurs domaines et la proportion est encore plus élevée chez les protéines eucaryotes avec 65% à 80% des protéines. Le terme domaine est défini comme un composant structural, fonctionnel, et évolutif des protéines, qui peut souvent être exprimé en tant qu'unité unique (Murzin et al., 1995). Les domaines dans les protéines multidomaines ont souvent des interfaces significatives et peuvent soit être attachés par des linkers courts, structurés ou plus longs et flexibles. Les linkers eux-mêmes peuvent être importants pour la fonction de la protéine (Gokhale and Khosla, 2000). Par exemple la fusion bidomaine MsrA-MsrB est présente sous la forme fusionnée chez E coli et sous la forme séparée chez la plupart des autres organismes (Han et al., 2016). La structure du bidomaine enzymatique a été obtenue et des tests biochimiques ont été réalisés indiquant que non seulement les deux enzymes sont non fonctionnelles lorsqu'elles sont séparées mais en plus, le linker jouerait un rôle essentiel pour la catalyse. L'organisation modulaire des domaines protéiques observée dans les protéines de fusion naturelle est alors considéré comme un phénomène évolutif important (Long, 2000). Certaines activités enzymatiques sont portées par des protéines individuelles chez certaines espèces et retrouvées dans des protéines multidomaines chez d'autres comme illustré par l'exemple des Cytochrome P450 (Sadeghi and Gilardi, 2013) (figure 6). La compréhension des mécanismes d'assemblage au sein de ces protéines multidomaines peut aider à la conception de protéines multidomaines synthétiques.

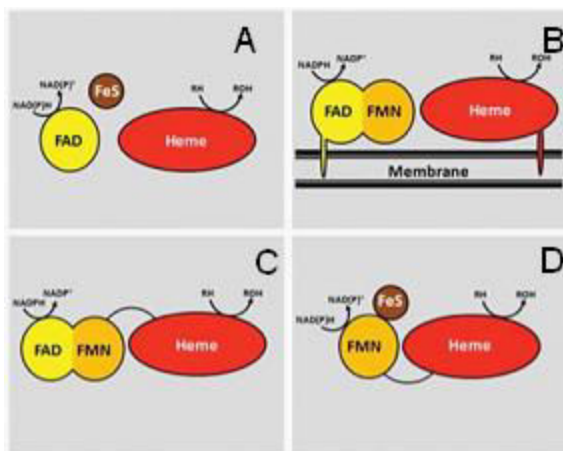


Figure 6 : Un exemple de protéine retrouvée sous forme soluble et parfois en multidomaine: les cytochromes P450. Le domaine / protéine contenant l'hème est commun mais peut-être associé de manière différente à des domaines / protéines réductases. La réaction catalysée par les enzymes P450 implique l'activation de l'oxygène moléculaire, habituellement par transfert d'électrons séquentiel à partir du NAD(P)H à l'hème du P450 via un partenaire(s) redox. A. classe I-trois protéines distinctes: la réductase contenant une FAD, la ferrédoxine (groupe [2Fe-2S]) et l'hème; B. classe II-deux protéines membranaires: la diflavine P450 réductase (FAD et FMN) et l'hème; C. classe VIII-fusion naturelle de la diflavine réductase (FAD et FMN) et de l'hème via un linker; et D. classe VII - fusion naturelle d'une phthalate dioxygénase réductase (FMN et cluster [2Fe-2S]) et de l'hème via un linker d'acide aminé. Figure tirée de (Sadeghi and Gilardi, 2013)

La nature a établie d'autres stratégies aussi sophistiquées les unes que les autres pour tirer profit d'une colocalisation des enzymes participant à une même voie de biosynthèse: les tunnels au sein des protéines (figure 7), le guidage par interactions électrostatiques le long de surface d'enzymes multifonctionnelles (figure 8), la liaison covalente des intermédiaires dans une structure multienzymatique (figure 9), les métabolons (Wheeldon et al., 2016), ou encore les protéines contenant plusieurs domaines (figure 6)(Hawkins and Lamb, 1995; Sadeghi and Gilardi, 2013; Wheeldon et al., 2016).

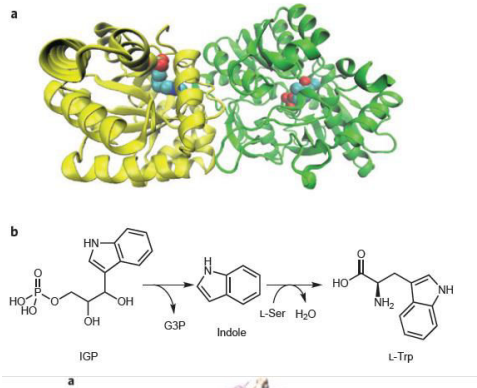


Figure 7: Canalisation d'un substrat au sein d'un tunnel intramoléculaire. a. La structure cristallographique de la tryptophane synthase. La sous-unité α est représentée en jaune et la sous-unité β en vert. L'Indole-3-glycérol-phosphate (IGP) est converti en indole et en glycéraldéhyde-3-phosphate (G3P) dans le premier site actif. L'intermédiaire indole est canalisé dans un tunnel intramoléculaire au sein de la protéine dimère vers le second site actif où il est condensé avec une sérine (L-Ser) pour produire du tryptophane (L-Trp). b. Le schéma réactionnel. Figure tirée de (Wheeldon et al., 2016).

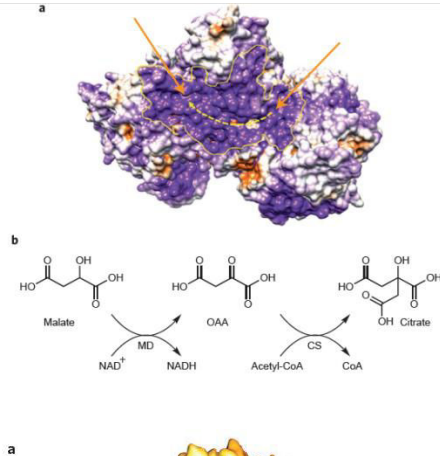


Figure 8: Canalisation d'un substrat par guidage électrostatique. a. La structure cristalline du malate-déshydrogénase (MD) -citrate synthase (CS) dans le métabolite de l'acide tricarboxylique cyclique (TCA). Les charges positives sont indiquées en violet, les charges négatives en orange et les résidus neutres en blanc. La surface protéique chargée positivement entre les sites actifs du MD et du CS est délimitée en jaune. La flèche jaune en pointillé indique le chemin de diffusion approximatif du MD au site actif du CS. b. Au site actif du MD, le malate est oxydé en oxaloacétate (OAA) produisant un équivalent réducteur sous la forme de NADH. L'intermédiaire oxaloacétate est canalisé le long de la surface de la structure vers le site actif du CS où il est converti en citrate. Figure tirée de (Wheeldon et al., 2016).

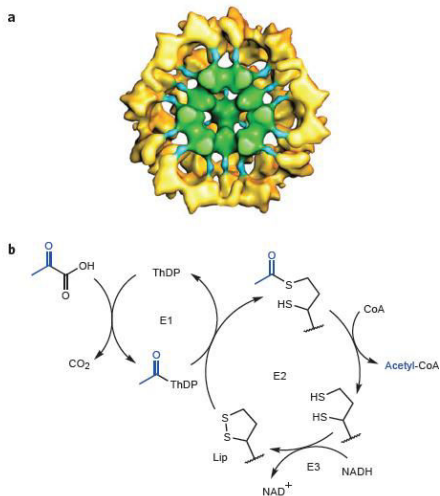


Figure 9: Canalisation d'un substrat par utilisation d'un bras oscillant chimique. a. L'enzyme E1 et la structure E2 du complexe eucaryote de la pyruvate déshydrogénase. La structure du noyau E2 est colorée en vert, les enzymes E1 associées sont représentées en jaune et le lien entre E1 et E2 est représenté en bleu. b. E1 décarboxyle le pyruvate et transfère le groupe acétyle résultant vers le bras oscillant lipoamide (Lip) sur E2. ThDP, thiamine diphosphate. E2 transfère ensuite le groupe acétyle du bras oscillant à la coenzyme A libre (CoA) et le bras oscillant est réactivé par E3. Figure tirée de (Wheeldon et al., 2016).

Le contrôle ou l'aide au transfert des réactifs entre les sites actifs d'enzymes d'une voie de biosynthèse améliore la diffusion du substrat et l'équilibre de la réaction, protège les intermédiaires instables,

augmente le flux en augmentant la concentration locale d'intermédiaires, diminue la concentration d'enzymes nécessaires pour atteindre un flux donné, dirige les produits d'une voie vers une localisation subcellulaire spécifique, minimise la fuite des réactifs intermédiaires, élimine la concurrence avec d'autres substrats ou l'inhibition entre différentes voies, et maintient un microenvironnement enzymatique stable (Conrado et al., 2008; Laursen et al., 2015; Schmitt and An, 2017; Srere, 1987; Zhang, 2011). Grâce à la réduction de la distance de diffusion et du temps de transit, la concentration locale efficace des intermédiaires de la voie de biosynthèse est plus élevée autour du complexe enzymatique comparée au reste de la cellule (Lee et al., 2012). Cette augmentation des concentrations locales de métabolites empêche les réactions secondaires non spécifiques et donc les interférences métaboliques, favorise la cinétique de la réaction et dirige le flux intracellulaire de métabolites vers la synthèse du produit désiré. Cependant, ce modèle n'est vrai que pour les enzymes à diffusion limitée (Lee et al., 2012; Sweetlove and Fernie, 2013). La majorité des enzymes ne sont pas limitées en diffusion, ce qui signifie que la catalyse est plus lente que la vitesse de diffusion. Pour les enzymes non limitées par la diffusion, les avantages les plus importants du channeling métabolique sont une réduction du temps d'équilibre et un meilleur contrôle de la spécificité de la réaction et de la régulation des branches métaboliques (Lee et al., 2012; Sweetlove and Fernie, 2013). Aussi, la libération d'intermédiaires toxiques et / ou instables est limitée par de tels complexes multi-enzymes qui fonctionnent comme des pipelines contrôlant strictement le flux métabolique.

1.3.1.1 Colocalisation par interaction de domaines partenaires

Un autre exemple d'organisation spatiale augmentant l'efficacité des enzymes chez les procaryotes est représenté par le complexe multienzymatique cellulosique ou cellulosome retrouvé chez *Clostridium thermocellum* (Chen and Silver, 2012). De nombreuses enzymes différentes requises pour la dégradation des parois cellulaires végétales sont organisées comme des nanomachines extracellulaires sur un scaffold à la surface cellulaire, grâce à l'interaction spécifique des domaines protéiques dockérine et cohésine (figure 10) (Bayer et al., 2004; Mazzoli et al., 2012).

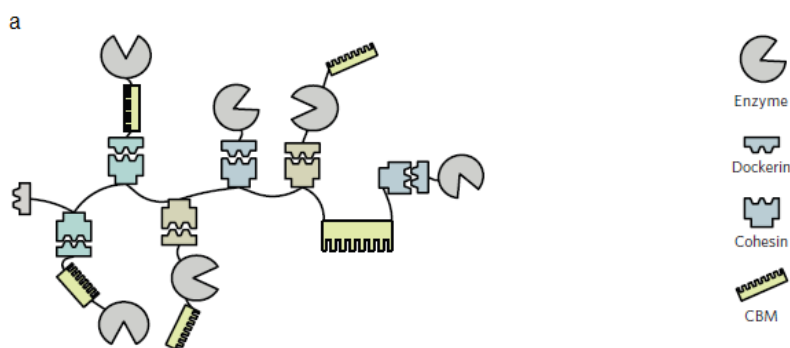


Figure 10 : Représentation schématique d'un complexe cellulosomique. Les enzymes sont fusionnées avec une cohésine. Chaque cohésine interagit spécifiquement avec son domaine dockérine partenaire. Les CBM facilite l'arrimage des enzymes aux substrats polysaccharidiques. Tirée de (Agapakis et al., 2012)

Ainsi des cellulases, xylanases, mannases sont chacune fusionnées à un domaine dockérine et interagissent avec des modules de cohésine au sein d'un scaffold protéique pour former un complexe de cellulosome (Bayer et al., 1994). L'organisation et la co-localisation des différentes enzymes hydrolytiques et la proximité du substrat fournissent une stratégie synergique efficace pour dégrader la cellulose et l'hémicellulose (Morais et al., 2012). L'interaction Coh-doc est une interaction protéine-protéine de haute affinité avec une constante de dissociation allant de 10^{-9} à 10^{-12} M (Liu et al., 2013a) qui peut être, dans un but d'ingénierie métabolique, réutilisée pour guider l'interaction d'autres enzymes.

1.3.1.2 La compartimentation dans les organelles et les structures spécialisées

Les organelles, qui sont une caractéristique des cellules eucaryotes, contiennent un certain sous-ensemble d'enzymes métaboliques qui effectuent des réactions biologiques distinctes (Polka et al., 2016). La séparation physique de réactions biologiques du reste des composants cellulaires augmente l'efficacité métabolique globale et permet même à des réactions incompatibles ou contradictoires telles que la synthèse (anabolisme) et la dégradation (catabolisme) du glucose, ou l'oxydation et la réduction de se dérouler en même temps dans la même cellule. La compartimentation isole des intermédiaires toxiques du reste de la cellule, et permet également la détoxification des intermédiaires toxiques de la voie sans nuire à la cellule. Dans les peroxysomes, la molécule H_2O_2 cytotoxique pour la cellule est efficacement dégradée par la catalase présente spécifiquement dans ces organites. Par proximité imposée du substrat et de l'enzyme, la concentration effective locale de catalase dans les peroxysomes est élevée et le H_2O_2 libéré est ainsi dégradé instantanément et efficacement. Le H_2O_2 qui est dangereux pour la cellule n'a pas le temps de s'y propager.

Les bactéries organisent également spatialement leur milieu intérieur pour des fonctions spécialisées (Chen and Silver, 2012), mais ne possèdent pas d'organites membranaires. L'intérieur des procaryotes peut contenir divers compartiments à base de protéines pour séparer physiquement des réactions enzymatiques distinctes, souvent critiques (Boyle and Silver, 2012). Ces compartiments bactériens sont composés de plusieurs protéines qui s'assemblent pour former des enveloppes protéiques minces et encapsulent typiquement des enzymes séquentielles. L'incorporation des enzymes dans le complexe est supposée avoir lieu pendant l'assemblage des protéines de l'enveloppe. Les petits pores ($\sim 0,5$ nm) formés par l'organisation protéiques permettent un échange sélectif de métabolites par exclusion de taille (Lee et al., 2012). Les carboxysomes, un exemple de microcompartiments bactériens trouvé chez les cyanobactéries et autres procaryotes autotrophes (environ 100 nm de diamètre) contiennent 270 molécules de ribulose-1,5-bisphosphate carboxylase / oxygénase (RuBisCO), l'enzyme clé de la fixation du carbone (Chen and Silver, 2012) (figure 11).

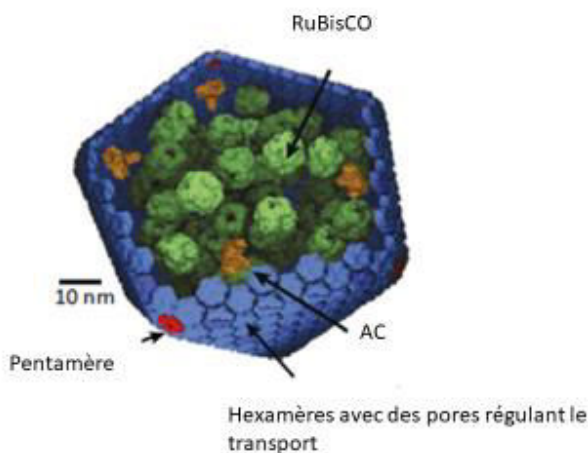


Figure 11: Encapsulation de réactions dans un microcompartiment bactérien formé par une structure protéique. Le carboxysome contient plusieurs ribulose 1,5-bisphosphate carboxylase oxygénase (RuBisCO en vert) et plusieurs anhydrase carbonique (orange). La coquille comprend des pentamères (rouge) et des hexamères (bleu), dont certains contiennent des pores qui sont

Dans un carboxysome, la concentration locale en CO₂ est considérablement augmentée au voisinage de la RuBisCO, augmentant ainsi son activité catalytique et empêchant son oxydation.

1.3.2 Ingénierie de la colocalisation d'enzymes

La conception d'un système de channeling parfait artificiel semble encore difficile. Le channeling du substrat en lui-même par des enzymes en cascade est difficile à mesurer in vivo (Wu and Minter, 2015). Le channeling a été démontré par exemple pour des enzymes du cycle de Krebs in vivo (Lindbladh et al., 1994; Sumegi et al., 1990). Le channeling du substrat est plus souvent déduit de tests in vitro par des études de la thermodynamique d'une réaction, de la modélisation du transport / de la réaction, de l'analyse de la diffusion moléculaire et des interactions protéiques, ou de l'étiquetage isotopique dynamique ou stationnaire (figure 12) (Abernathy et al., 2017).

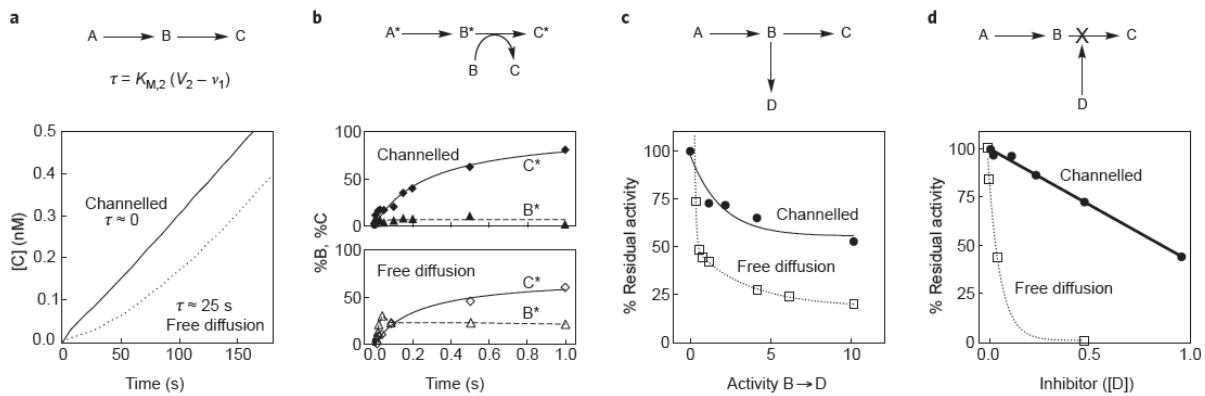


Figure 12 : Des exemples de méthodes biochimiques pour évaluer le channeling du substrat. Des schémas de réaction et des exemples de données sont présentés pour a. l'analyse du temps transitoire τ , b. la dilution isotopique et enrichissement, c. la mise en concurrence par une réaction secondaire compétitrice. d. la mise en concurrence par un inhibiteur en cascade. Figure tirée de (Wheeldon et al., 2016)

Sur la base des nombreux exemples naturels de colocalisation enzymatique, l'organisation spatiale enzymatique peut constituer une variable clé dans la conception de systèmes enzymatiques. Dans les applications d'ingénierie métabolique, les voies hétérologues ont typiquement une efficacité inférieure par rapport aux voies natives. Contrairement aux voies naturelles, les voies d'ingénierie pourraient être incapables de surmonter par exemples les effets de surpopulation intracellulaire, les limitations de diffusion, les réactions de promiscuité ou le contrôle par une régulation optimale. La possibilité de construire des complexes enzymatiques synthétiques a été discutée dans un certain nombre de revues (Agapakis et al., 2012; Conrado et al., 2008; Lee et al., 2012). Au fil des années, les chercheurs ont ainsi imité les stratégies naturelles pour colocaliser les enzymes. Dans le cadre d'une ingénierie métabolique, le moyen le plus direct pour déterminer l'utilité du rapprochement spatiale des enzymes est de l'introduire dans des cellules et d'évaluer son effet sur le flux ou la formation de produit.

1.3.2.1 La fusion protéique

1.3.2.1.1 Exemples de création de fusions enzymatiques

La manière la plus simple d'assembler les complexes protéiques est de fusionner les enzymes métaboliques qui catalysent des réactions successives. Les protéines de fusion sont une classe de protéines avec deux domaines protéiques ou plus intégrés dans une molécule. Des protéines de fusion artificielles ont été créées pour diverses applications. Dans le domaine de l'ingénierie enzymatique, les premières tentatives ont été faites pour créer des fusions de protéines soit pour augmenter la vitesse de réaction enzymatique consécutive soit pour générer des enzymes bifonctionnelles. La première enzyme de fusion artificielle a été produite par introduction de mutations ponctuelles dans le second gène *hisD*, de l'opéron histidine, chez *Salmonella typhimurium*. Cela a permis la création de la fusion bifonctionnelle d'une histidinol déshydrogénase avec l'aminotransférase (Yournon et al., 1970). L'étude évoque que la fusion pourrait conduire à la formation d'agrégats protéiques car les deux enzymes forment initialement des multimères. Différentes formes d'agrégats pourraient alors être possibles en fonction de la flexibilité du linker utilisés (figure 13).

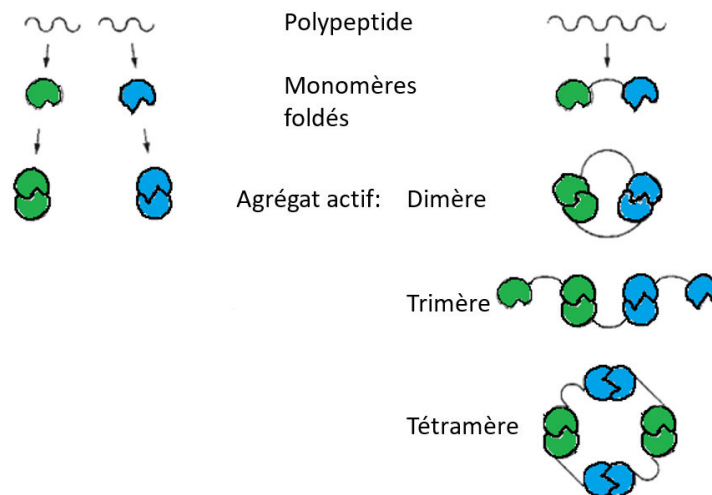


Figure 13 : Des interactions moléculaires possibles lors de la fusion de deux enzymes. A noter qu'en plus de ce qui est représenté sur cette figure tirée de (Yournon et al., 1970), les agrégats peuvent être formés aussi par interaction entre domaines protéiques différents.

Cette étude pionnière a ouvert la route pour la production de systèmes enzymatiques bi- et multifonctionnels pour différentes applications biotechnologiques (Elleuche, 2015; Grünberg and Serrano, 2010; Schmidt, 2013).

Une autre étude a conçu une fusion directe entre le gène de la bêta-galactosidase d'*E. coli* (LacZ), une enzyme tétramérique, et le gène codant pour la galactokinase d'*E. coli* (GalK), une enzyme monomérique, pour la conversion en deux étapes du lactose en galactose-1-phosphate. La protéine de fusion résultante montre l'activité enzymatique des produits des deux gènes, mais avec une réduction significative par rapport aux activités individuelles (Bülow et al., 1985). Un inconvénient de la fusion de deux enzymes est donc que leur liaison non naturelle peut avoir des effets préjudiciables sur leur fonctionnalité. Le même groupe créa ensuite une fusion directe entre LacZ et la galactose

déshydrogénase (GalDH) de *Pseudomonas fluorescens*, une enzyme dimérique, afin d'hydrolyser séquentiellement le lactose puis d'oxyder le galactose, formant la lactone correspondant (Ljungcrantz et al., 1989). L'enzyme bifonctionnelle possède cette fois-ci des avantages cinétiques notamment une double réduction de la constante de saturation, le K_m pour le lactose, et l'atteinte de l'état stationnaire de deux à quatre fois plus rapidement en solution par rapport au système natif en particulier à de faibles concentrations de substrat. Cela a suggéré que la proximité conférée par la fusion artificielle a permis une certaine quantité de synergie pour le couplage des réactions séquentielles (c'est-à-dire le channeling du substrat) absente quand les activités enzymatiques sont séparées. Ces exemples n'utilisant aucune optimisation particulière du linker en raison de l'absence d'outils de biologie moléculaire, montrent les premiers inconvénients mais aussi les avantages que peuvent procurer la fusion de deux enzymes. De même, dans une autre étude chez *E. coli*, une fusion artificielle entre une enzyme catalysant des étapes non successives, à savoir la première étape enzymatique (galactose déshydrogénase) et celle catalysant la troisième étape enzymatique (luciférase), auquel est ajoutée séparément l'enzyme catalysant l'étape intermédiaire (diaphorase), est plus efficace (produit plus de luminescence) que le système enzymatique constitué des trois enzymes séparées (Lindblad et al., 1992). Dans ces exemples, l'amélioration de l'activité observée lors du couplage des activités enzymatiques était maximale lorsque l'activité de la première enzyme était mise en condition limitante, par les conditions de pH non optimales ou de faibles concentrations d'enzyme ou de substrat (Lindblad et al., 1992; Ljungcrantz et al., 1989). Dans ces exemples, l'effet du couplage sur l'activité était modéré, mais néanmoins significatif. De meilleurs résultats pourraient être possibles en optimisant la position des centres catalytiques l'un par rapport à l'autre. Dans une étude par Meynial Salles et ses collègues, une fusion efficace a été obtenue par une approche par pression de sélection en laboratoire en utilisant l'évolution comme force motrice. Les auteurs de l'étude décrivent la combinaison de deux enzymes distinctes de deux gènes de levure impliquées dans la voie du glycérol en utilisant un seul plasmide pour leur expression chez *Escherichia coli* (Meynial Salles et al., 2007). Une délétion entre les gènes codant ces enzymes a résulté en la production d'une enzyme de fusion présentant à la fois les activités glycérol-3-phosphate déshydrogénase et glycérol-3-phosphate phosphatase avec des performances catalytiques améliorées comparées aux enzymes libres et avec un mécanisme de channeling partiel du glycérol-3-phosphate (Meynial Salles et al., 2007). En vérifiant la séquence de la bifusion créée, 8 résidus en C terminale de la protéine en N terminale de la bifusion ont été éliminés et l'insertion d'un linker de 4 résidus issus du plasmide ont permis d'obtenir cette bifusion efficace.

I.3.2.1.2 Choix du linker dans la conception de fusions protéiques

Des résultats montrent les effets possibles du linker sur les caractéristiques d'une bifusion et les avantages que pourrait conférer une fusion dans des conditions de milieux cellulaire dense. Un possible effet de channeling est proposé pour expliquer l'efficacité d'une fusion artificielle entre la β galactosidase et la galactokinase utilisant un linker de 20 acides aminés (Bülow, 1987). Alors que la bifusion ne présente aucun avantage particulier par rapport aux formes séparées, un effet de proximité est observé en présence de l'enzyme compétitrice pour l'intermédiaire galactose, la galactose déshydrogénase. Selon les auteurs, l'effet de proximité est dû à une concentration locale plus élevée de l'intermédiaire. Cet effet est augmenté en présence de PEG pour imiter la densité du milieu cellulaire. L'étude évoque le rôle important du linker sur les activités catalytiques et la stabilité de la fusion

enzymatique. Afin de vérifier cela, et pour étudier les effets observés in vitro dans les conditions cellulaires, la même fusion enzymatique est utilisée dans leur étude suivante en condition in vivo chez *E. coli* et avec différents types de linker variant en longueur et en propriété (Carlsson et al., 1992). Des résidus prolines et glycines ont été introduits pour donner respectivement de la rigidité ou de la flexibilité au linker. In vivo, une fusion avec le linker le plus court (10 acides aminés) permet un meilleur transfert de l'intermédiaire du premier site actif au suivant. Selon les auteurs, cela limiterait la diffusion de l'intermédiaire dans le milieu et diminuerait ainsi son accessibilité pour l'enzyme compétitrice galactose déshydrogénase co exprimée. Pour confirmer in vitro les résultats observés, un dosage compétitif avec l'intermédiaire radioactif sur le système enzymatique a confirmé les mêmes propriétés conférées par la bifusion avec un linker court par rapport aux autres bifusions. De plus, la longueur et la flexibilité de la séquence de liaison utilisée pour coupler les enzymes peuvent influencer l'activité de la protéine de fusion résultante. Pour déterminer une conception optimale, des linkers de différentes longueurs (3-239 acides aminés) et flexibilités (déterminés par l'encombrement des résidus) ont été testés avec les fusions Erg20-PTS. La production de patchoulol était relativement constante pour différentes flexibilités de linker; cependant, le titre diminue avec l'augmentation de la longueur du linker (Albertsen et al., 2011). Il est probable que l'effet de la longueur du linker et de l'orientation des parties codantes peuvent dépendre des propriétés spécifiques des enzymes dans la protéine de fusion et peuvent nécessiter une optimisation spécifique à chaque enzyme. Le choix du linker peut donc être un facteur décisif pour la construction d'une fusion réussie. Les propriétés, les fonctions et les exemples de linkers de protéines de fusion ont été récemment examinés et discutés (Chen et al., 2013). La longueur du linker, sa composition en acides aminés, son hydrophobicité, la sensibilité aux protéases, sa structure secondaire et les interactions possibles avec des partenaires de fusion méritent un examen attentif. Deux études ont utilisé des bases de données existantes pour compiler et caractériser les linkers dans des protéines multi-domaines naturelles (A George and Heringa, 2002; Argos, 1990). Ces études ont révélé des propensions de séquences d'acides aminés pour des linkers naturels de différentes tailles et longueurs, ainsi que des informations sur la rigidité et la structure secondaire. Cette information a contribué à la conception empirique de linker personnalisés pour des applications particulières. Néanmoins, la sélection d'un linker est principalement ad hoc et reste une zone sous-explorée dans l'ingénierie des protéines de fusion.

D'autres exemples d'améliorations dans les caractéristiques cinétiques enzymatiques ont été amenées par la fusion de gènes dans les biotechnologies, en particulier en ingénierie microbienne et végétale (Kampranis and Makris, 2012). L'approche de fusion enzymatique a été utilisée pour atteindre une amélioration de 15 fois de la production de resvératrol dans la levure et des cellules humaines (Zhang et al., 2006). Dans cette étude, il a été noté que des interactions protéines-protéines pourraient augmenter le channeling métabolique ou placer les sites actifs des enzymes à proximité. La fusion de la FPP synthase et de la farnésène synthase a entraîné une augmentation de la production de farnésène et a éliminé la formation d'un produit secondaire, le farnesol, en comparaison avec les enzymes libres (Wang et al., 2011). Dans un autre projet, pour augmenter le flux de bisabolène chez *S. cerevisiae*, la farnésyl diphosphate synthase endogène (ERG20) a été fusionné à la bisabolène synthase (AgBIS) en utilisant le linker flexible Gly-Gly-Gly-Ser (GGGS), résultant en une augmentation de la production de bisabolène de 2 fois supérieure aux gènes exprimés individuellement (Özaydın et al., 2013).

Étonnamment, la protéine de fusion est significativement mieux exprimée que la bisabolène synthase non fusionnée dans cette étude, ce qui souligne que la fusion peut avoir un impact sur la stabilité ou le niveau de production des protéines fusionnées. L'orientation des ORF impacte aussi l'activité de la protéine de fusion, mais l'effet dépend des propriétés des enzymes. Par exemple, une différence de presque quatre fois a été observée dans la production de GGPP entre les protéines de fusion Erg20-Bts1 et Bts1-Erg20 (Ohto et al., 2010). L'orientation relative de la fusion Erg20-PTS n'a pas eu d'effet significatif sur la production de patchoulol (Albertsen et al., 2011).

Une étude récente impliquant la production hétérologue de cétone de framboise chez *S. cerevisiae* a évalué différentes possibilités de fusion de gènes pour obtenir une combinaison enzyme-linker-enzyme optimisée (Lee et al., 2016). La 4-coumaroyl-CoA ligase (4CL) et la benzalacétone synthase (BAS) ont été fusionnées dans deux orientations différentes avec soit des linkers flexibles ou soit des linkers rigides. Les performances en utilisant deux variantes de 4CL ont été comparées. Cette approche a abouti à une fusion avec un rendement augmenté de plus de 5 fois par rapport aux enzymes monofonctionnelles natives.

Un avantage de l'utilisation des fusions de gènes dans les applications d'ingénierie métabolique est que les deux domaines protéiques (ou plus) sont implicitement co-localisés dans l'organisme hôte. Par exemple les bactéries ne possèdent pas de structures endomembranaires nécessaires pour une performance optimale des Cytochrome P450 liées au RE. D'autant plus que ces dernières nécessitent une seconde enzyme liée au RE, la cytochrome P450 réductase (CPR), pour un transfert d'électrons efficace. La technologie de fusion de gènes a permis l'expression de CYP dans *E. coli* par liaison avec une CPR microbienne ou de plante (Schückel et al., 2012). Combiner des enzymes microbiennes et de plantes représente un troisième avantage de l'application de la fusion de gènes, donnant la capacité de créer des enzymes efficaces n'interagissant pas naturellement dans la nature. L'organisme hôte peut ne pas reconnaître les informations de localisation contenues sur une protéine hétérologue, ou ne pas posséder l'anatomie cellulaire ou subcellulaire idéale pour une performance enzymatique adéquate.

Dans la nature, on observe de nombreux « échecs » de fusions de gènes naturelles, tels que des domaines catalytiquement actifs fusionnés à des domaines inactifs (Compaan and Ellington, 2003; Hiltunen et al., 2012; Moody Suzy C. et al., 2012; Zi et al., 2014) et de nombreuses fusions amenant à des protéines non actives (Li et al., 2016b), montrant que l'ingénierie de fusions de gènes est une entreprise risquée. Même si la fusion n'entraîne pas une perte de fonction enzymatique, une amélioration fonctionnelle importante par rapport aux composantes non fusionnées peut ne pas être obtenue.

1.3.2.2 Des stratégies de colocalisation post traductionnelle

L'ajout de tag d'adressage sur les enzymes permet de les colocaliser au sein de compartiment cellulaire spécifique et peut être avantageux dans certain cas. Cela peut par exemple être un moyen de délocaliser une enzyme pour diriger la spécificité vers certaines réactions. Une augmentation du flux de la voie de biosynthèse par cette stratégie pourrait s'expliquer par exemple par l'isolement des intermédiaires de voies concurrentielles ou une concentration locale élevée en cofacteur ou en précurseur spécifique. Les défis pour l'utilisation des organites membranaires comprennent le ciblage des composants

hétérologues, le contrôle de la stœchiométrie des enzymes encapsulée, la sécrétion des produits (Chen and Silver, 2012). L'encapsulation de voie au sein de compartiments spécifiques comme les carboxysomes concentre également les efforts de plusieurs équipes de recherche. La compréhension du mode d'assemblage de ces structures, du mode d'incorporation des enzymes et des substrats, de la sélectivité des pores pour pouvoir sélectionner les molécules sont nécessaires pour réaliser le potentiel de cette stratégie.

Une méthode alternative pour rapprocher spatialement des enzymes consiste à joindre les protéines après leur expression séparée. La stratégie nécessite la fusion sur chaque enzyme d'un domaine d'interaction protéique. Cette stratégie offre la possibilité de colocaliser deux ou plusieurs centres catalytiques sans synthétiser de longs polypeptides. A cet égard, les domaines d'interaction des complexes cellulolytiques peuvent être manipulés avec une remarquable polyvalence. Les stratégies de construction de système multienzymatiques basées sur les interactions coh-doc sont divisées en deux catégories : (a) une enzyme fusionnée avec une cohésine, et une autre enzyme fusionnée avec la dockérine correspondante; (b) toutes les enzymes sont fusionnées avec la dockérine, et les cohésines sont fusionnées pour former un scaffold chimérique. On parle alors d'une stratégie de scaffolding dans ce cas (figure 14).

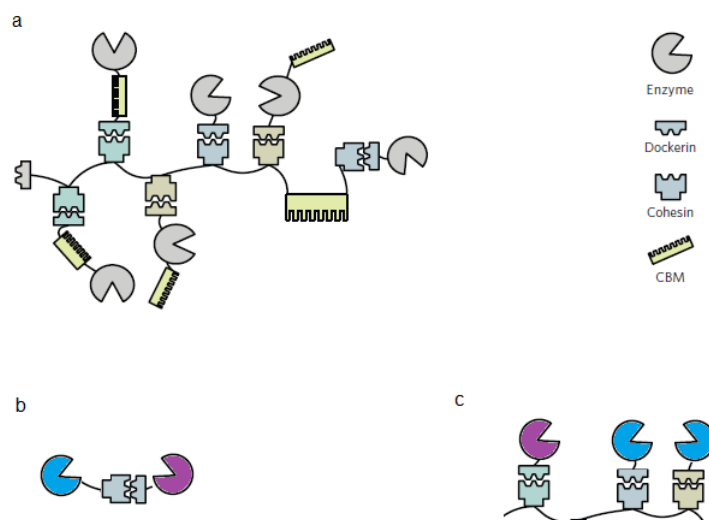


Figure 14 : Stratégies pour la construction de systèmes multienzymatiques basés sur les interactions coh-doc. a. représentation schématique d'un cellulosome, b. représentation schématique d'une enzyme bi-fonctionnelle par interaction cohésine dockérine, c. représentation schématique d'enzymes colocalisées par recrutement au niveau d'un scaffold de cohésines. Tirée et adaptée de (Agapakis et al., 2012)

La première stratégie ne convient que pour un assemblage à deux enzymes alors que la seconde peut être appliquée pour la construction de deux ou plusieurs cascades d'enzymes. La modularité du scaffold synthétique permet alors l'optimisation de la stœchiométrie enzymatique en faisant varier le nombre de domaines de liaison sur l'échafaudage. Des exemples d'assemblages d'enzymes basés sur la cohésine-dockérine sont décrits dans la littérature (Chiang et al., 2014; Liu et al., 2013a; Tsai et al., 2009, 2013; You et al., 2012). Une amidohydrolase-cohésine (AHL-coh) et une hydantoinase-dockérine (HDT-doc) ont permis d'obtenir une enzyme bifonctionnelle, qui présente un taux de réaction 2,6 fois plus élevé que

les enzymes libres (Chiang et al., 2014). Une synthèse métabolique basée sur le système coh-doc avec une amélioration de la vitesse de réaction initiale de 38 fois par rapport à celle obtenue avec le mélange d'enzymes libres avec un système in vitro est décrite (You and Zhang, 2013).

La nature fournit différents autres domaines d'interaction protéique (Whitaker and Dueber, 2011). Un scaffold protéique synthétique a été créé en 2009 en recrutant des enzymes métaboliques de levures au niveau de protéines métazoaires par une interaction naturelle de domaine issus d'une voie de signalisation métazoaires. Cela a amélioré la production du mévalonate de trois jusqu'à plus de 70 fois chez *Escherichia coli*, en fonction des proportions relative d'enzymes exprimées (Dueber et al., 2009). Le titre du produit a augmenté de manière stœchiométrique avec le nombre de domaines peptide-ligand de scaffolding (figure 15).

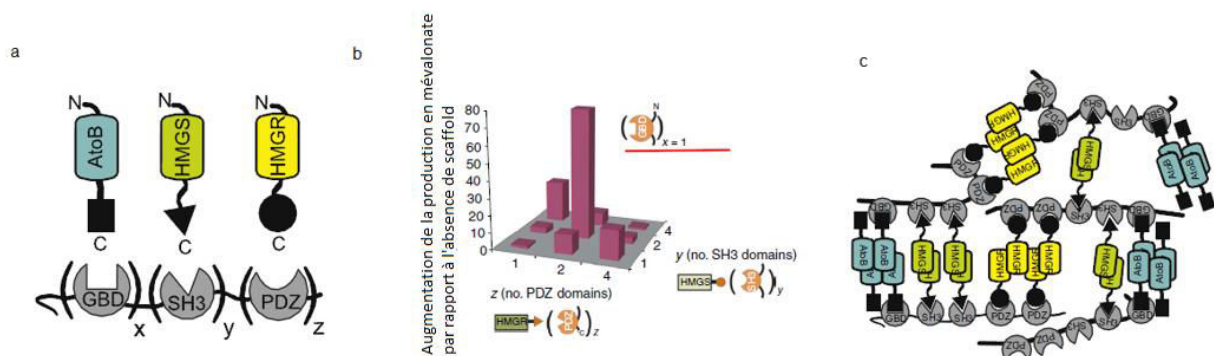


Figure 15 : Une approche de scaffolding pour optimiser la colocalisation des enzymes d'une voie de biosynthèse. Tirée de (Dueber et al., 2009) a. AtoB (acétyl-CoA acétyltransférase), HMGS (3-hydroxy-3-méthylglutaryl CoA synthase) et HMGR (3-hydroxy-3-méthylglutaryl-CoA réductase) ont été fusionnées avec des ligands peptidiques organisées sur un scaffold artificiel formé par GBD, SH3 et PDZ. b. Une matrice représentant neuf échafaudages différents montrant des différences spectaculaires dans les titres du produit mévalonate. c. Les enzymes avec des structures oligomériques peuvent potentiellement être liées avec d'autres scaffold, résultant en un large complexe dont l'organisation n'est pas contrôlée. Tirée de (Lee et al., 2012)

Selon les auteurs, l'amélioration provient d'un flux de voie équilibré, qui a empêché l'accumulation d'un intermédiaire toxique et qui a permis de réduire le niveau d'expression de protéine superflue. Cet effet est perdu lorsque le nombre de domaines connectés dépasse un certain niveau, peut-être à cause de mauvais repliement enzymatique ou d'interactions allostériques non caractérisées causées par un excès de ligand-peptide. Des améliorations beaucoup moins importantes ont été observées lorsque cette approche a été utilisée pour d'autres voies métaboliques (Horn and Sticht, 2015). Le mécanisme réel par lequel les scaffold opèrent pour augmenter le rendement est inconnu et on ne sait pas s'ils donnent lieu à la formation de structures supérieures (Chen and Silver, 2012). Une augmentation supplémentaire de l'efficacité de production pourrait être obtenue grâce à une optimisation des linkers ou l'utilisation de domaines adaptatifs alternatifs, de la variation stœchiométrique des enzymes utilisées (Dueber et al., 2009).

1.3.2.3 Vers un contrôle plus précis et designable du scaffolding.

L'organisation de composants cellulaires interdépendants dans le temps et l'espace est cruciale pour accroître l'efficacité de divers processus cellulaires. Nous avons besoin de blocs de construction modulaires pour la fabrication sur mesure d'architectures subcellulaires qui pourront être programmées

par la suite pour des fonctions précises. L'idée est de pouvoir contrôler de manière prédictive l'architecture du scaffold et de contrôler l'assemblage. Les biomolécules polymérisantes représentent des blocs de construction idéaux parce qu'ils peuvent s'auto-assembler de manière efficace in vivo. A ce jour, la nanotechnologie par hybridation d'ADN (par ex. ADN origami) est peut-être la plate-forme moléculaire la mieux développée et la mieux concevable (Pinheiro et al., 2011). Les architectures d'ADN sont particulièrement flexibles dans des conditions non physiologiques où un nombre illimité d'architecture peuvent être construits de manière prédictive et contrôlée à des échelles approchant le sous-nanomètre (Wilner et al., 2009; Fu et al., 2012; Funke et Dietz, 2016). Traduire cette technologie pour une application intracellulaire est cependant partiellement contraint parce que la concentration d'acide nucléique simple brin nécessaire pour former ces structures est importante, et les propriétés environnementales impactent sur le repliement des acides nucléiques comme par exemple la température ou les ions (Pinheiro et al., 2011).

En raison de leurs propriétés d'auto-assemblage, les protéines formant la coque de microcompartiments bactériens pourraient constituer des blocs de construction prometteurs à réutiliser pour concevoir de nouveaux échafaudages de protéines in vivo, distinctivement de leur propension naturelle à assembler des microcompartiments bactériens. Les principaux constituants des coquilles de microcompartiments bactériens sont typiquement des petites protéines (100 acides aminés) contenant le domaine d'interaction BMC-H qui interagissent pour former un disque hexagonal de 70 Å avec des faces distinctes et un pore circulaire au centre (Kerfeld et al., 2010; Yeates et al., 2010) (figure 16 a et b). On pourrait imaginer que les protéines d'échafaudage à hexamères simples peuvent théoriquement colocaliser jusqu'à 6 protéines par hexamère, de manière semblable aux premières conceptions synthétiques d'échafaudage (Dueber et al., 2009) (figure 16 c). Le challenge est de ne pas perturber les mécanismes d'interaction naturelle de ces complexes protéiques et que cette organisation ne perturbe pas la fonctionnalité des enzymes colocalisées.

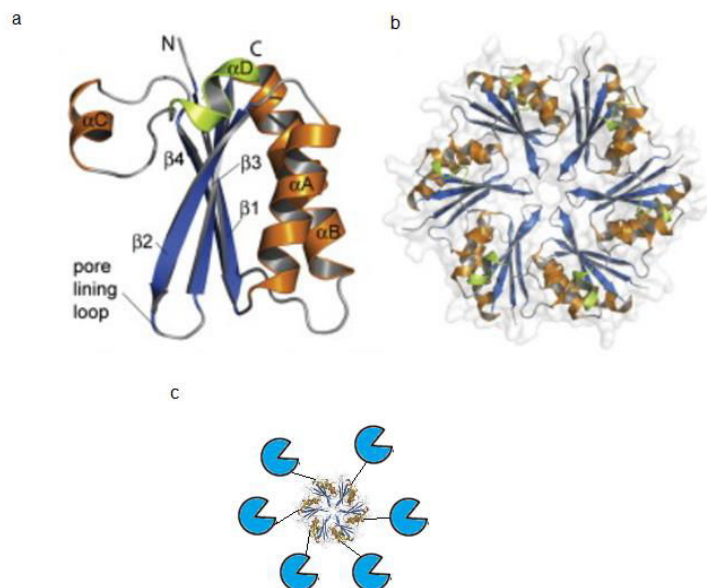


Figure 16 : Stratégie de colocalisation utilisant une fusion avec une protéine formant un complexe oligomérique naturel. a. Structure du protomère CcmK2 de forme 1, avec la structure secondaire indiquée et les extrémités N terminale et C terminale indiquées. b. l'hexamère CcmK2 de forme 1; la surface protéique est représentée avec une couche transparente. Notez que les hélices individuelles dans l'hexamère n'interagissent pas. c. Représentation de la colocalisation d'une enzyme par fusion à une extrémité de CcmK2. Tirée et adapté de (Samborska and Kimber, 2012)

Les stratégies d'ingénierie de colocalisation sont plus ou moins faciles à implémenter. L'efficacité de chaque stratégie peut dépendre du contexte du système enzymatique étudié. La caractérisation des différentes stratégies, la compréhension des mécanismes par lesquels elles peuvent être bénéfiques ainsi que de leur limitation respective sont recherchées.

I-4 Ingénierie métabolique pour la production de caroténoïdes

I.4.1 Les isoprénoïdes

Avec plus de 40.000 composés, les isoprénoïdes, également appelés terpénoïdes ou terpènes, constituent le groupe de molécules le plus diversifié structuellement et fonctionnellement. Ils sont synthétisés dans presque tous les organismes vivants. De nombreux isoprénoïdes sont utilisés en tant que produits pharmaceutiques, nutraceutiques, arômes, produits cosmétiques, additifs alimentaires et biocarburants (Keasling, 2010; Maury et al., 2005; Tippmann Stefan et al., 2013). Les isoprénoïdes sont assemblés à partir de formes activées de molécules à 5 carbones, les isoprènes, à savoir l'isopentényl pyrophosphate (IPP) et son isomère le diméthylallyl pyrophosphate (DMAPP). Ces deux précurseurs sont obtenus par deux voies différentes selon l'organisme : la voie du mévalonate (MVA) (Buhaescu and Izzedine, 2007) chez la plupart des eucaryotes et chez les archées et la voie du méthylérythritol phosphate (MEP) (Wanke et al., 2001) chez les procaryotes et les plastides des organismes photosynthétiques (figure 17).

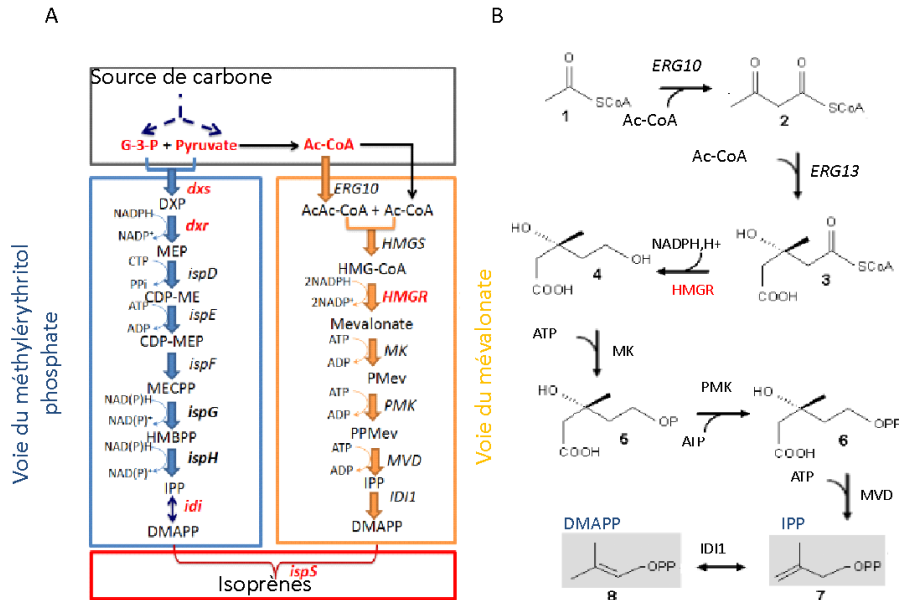


Figure 17 : Les voies de synthèse des précurseurs des isoprénoides. A. La voie du méthylérythritol phosphate (en bleu) et la voie du mévalonate (orange). B. Le détail de la voie du mévalonate : initiée par la condensation de deux acétyl-CoA par une thiolase (ERG10) pour produire de l'acétoacétyl-CoA. Par la suite, un autre acétyl-CoA est condensé avec l'acétoacétyl-CoA pour synthétiser le 3-hydroxy-3- méthyl-glutaryl-CoA (HMG-CoA) par la HMG-CoA synthase. L'acide mévalonique est ensuite formé à partir de l'HMG-CoA en utilisant le NADPH comme cofacteur par l'HMG-CoA réductase. L'HMG-CoA réductase est une enzyme limitante de cette voie. Deux kinases, la mévalonate kinase (MK) et la phosphomévalonate kinase (PMK), catalysent séquentiellement la phosphorylation du mévalonate pour produire du mévalonate 5- diphosphate (Mev-PP). La dernière étape de la voie du mévalonate pour former l'IPP est une décarboxylation ATP-dépendante catalysée par une mévalonate diphosphate décarboxylase. L'isomérisation de l'IPP par l'isopentényl diphosphate isomérase (Idi) conduit à la formation de l'autre précurseur des isoprénoides, le DMAPP.

Le DMAPP sert d'amorce et subit une élongation par condensation avec un ou plusieurs molécules d'IPP catalysées par des prényles synthases comme le géranyl pyrophosphate (GPP) synthase, le farnésyl pyrophosphate (FPP) synthase ou le géranylgéranyl pyrophosphate (GGPP) synthase. Ces terpènes synthases produisent les précurseurs GPP, FPP et GGPP à partir desquels sont issus les différentes classes d'isoprénoides selon le nombre de carbones : hémiterpènes (C5), monoterpènes (C10), sesquiterpènes (C15), diterpènes (C20), triterpènes (C30) et tétraterpènes (caroténoïdes, C40) (figure 18).

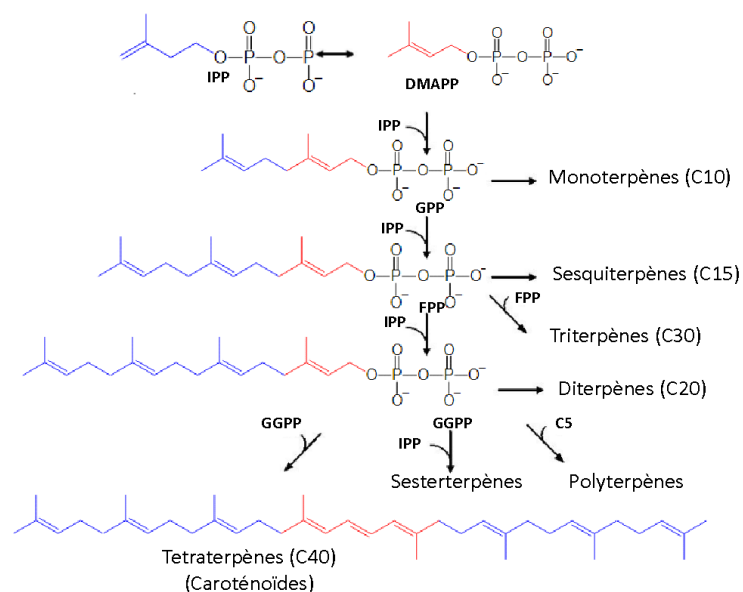


Figure 18 : La voie des précurseurs des différentes classes d'isoprénoides. L'origine des squelettes carbonnés est identifiée par la coloration rouge (DMAPP) ou bleue (IPP). Le nombre de carbone de chaque précurseurs d'isoprénoides est noté entre parenthèse.

Une fois le squelette de base formé, les isoprénoides sont souvent modifiés, par des cyclisations, des désaturations et des ajouts de groupe fonctionnels comme des hydroxylations, pour générer des molécules d'une grande diversité structurale.

I.4.2 Les caroténoïdes

Les caroténoïdes constituent un sous-groupe important d'isoprénoides. Les caroténoïdes appartiennent majoritairement à la classe d'isoprénoides appelée tetraterpène, qui commence par la condensation en tête à tête de deux molécules de GGPP. Ils sont produits par les plantes, les champignons, les algues, les bactéries, et certaines levures (Alcaïno et al., 2016). Les caroténoïdes sont produits sous forme d'hydrocarbures (carotènes) ou sous leur forme dérivée oxygénée (xanthophylles). Au moins 1175 caroténoïdes ont été caractérisés et répartis en caroténoïdes C30, C40 et C50 (Yabuzaki, 2017) en fonction de leur nombre de carbone. Les caroténoïdes sont des pigments naturels liposolubles. Chez les espèces photosynthétiques, les caroténoïdes sont associés au photosystème agissant comme pigments accessoires collecteurs de la lumière, étendant efficacement la gamme de lumière absorbée par l'appareil photosynthétique. Les caroténoïdes fournissent alors aux organismes une coloration allant du jaune vif, au rouge à l'orange. L'absorption de la lumière dépend du nombre de double liaison conjuguées dans la structure du caroténoïde. Dans ces organismes photosynthétiques, les caroténoïdes jouent également un rôle photoprotecteur en désactivant les molécules de chlorophylle à l'état triplet et en piégeant l'oxygène singulet et d'autres espèces oxygénées formées dans le chloroplaste, et en dissipant l'excès d'énergie nuisible (Ruban et al., 2007). Chez les organismes non photosynthétiques, leur fonction principale semble être celle de la (photo)protection. Chez l'homme, qui acquiert les caroténoïdes par l'alimentation, les caroténoïdes possèdent de multiples fonctions physiologiques et

nutritionnelles (Fraser and Bramley, 2004). Les caroténoïdes ont des propriétés antioxydantes (Britton, 1995; Krinsky and Yeum, 2003; Rice-Evans et al., 1997; Stahl and Sies, 2003) associées à leur capacité à lier un oxygène singulet par leur système de doubles liaisons conjuguées, avec une protection maximale pour les caroténoïdes avec plus de neuf double liaisons. Les caroténoïdes ont des propriétés préventives contre certains cancers (Gerster, 1993). Les caroténoïdes qui possèdent un cycle β sont précurseurs de la vitamine A (DellaPenna and Pogson, 2006). Le β carotène possède deux cycles β et constitue le caroténoïde principal précurseur de vitamine A par clivage au centre de la molécule par une β -carotène-15,15'-dioxgénase (figure 19).

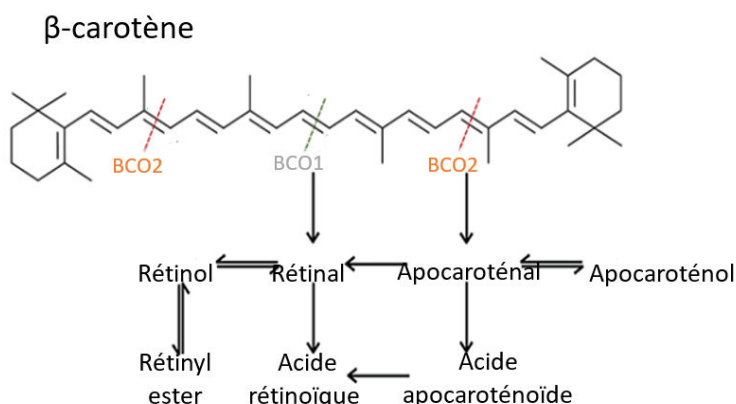


Figure 19 : Les différentes formes actives de la vitamine A obtenues à partir du β -carotène. BCO1 (β -carotène 15,15'-oxygénase 1) catalyse le clivage symétrique du β -carotène au niveau de la double liaison 15,15' pour donner du rétinal. Le rétinal peut être oxydé en acide rétinoïque et réduit en rétinol et convertit encore en rétinyl ester pour le transport et le stockage. Les caroténoïdes avec un cycle β et donc précurseurs de vitamine A peuvent aussi être clivés par BCO2 (β -carotène 9',10'-oxygénase 2) au niveau de la double liaison 9,10 ou 9',10' donnant des apocaroténals.

L'implication de la rétine (vitamine A) comme pigment chromophore visuel dans les yeux est d'une importance primordiale dans le processus de vision. Elle est également importante pour ses fonctions systémiques dans la croissance et l'efficacité reproductive. En raison de sa capacité antioxydante, le β -carotène est également utilisé comme protecteur solaire pour prévenir du photo vieillissement et de la brulure du soleil. Par ingestion, le β -carotène doit notamment être consommé durant plusieurs semaines pour augmenter son contenu dans le plasma sanguin et dans la peau et pour augmenter son effet de protection (Packer et al., 1999). Les caroténoïdes sont largement utilisés comme colorants alimentaires, composant de produits cosmétiques, suppléments alimentaires, dans l'alimentation animale. Le marché mondial des caroténoïdes en 2015 était de 1,23 milliard de dollars, avec une augmentation attendue à 1,81 milliard de dollars d'ici 2022 (März, 2015). Le β -carotène est largement utilisé dans les industries alimentaires, cosmétiques et pharmaceutiques pour ses traits de coloration, ses propriétés antioxydantes et son activité provitamine A. Son utilisation dans les aliments humains et pour animaux a été largement approuvée par les autorités compétentes mondiales. Le β -carotène concerne enfin environ 30% du marché des caroténoïdes avec seulement 2% de β -carotène naturel produit dont la majorité est utilisé comme supplément alimentaire. Les caroténoïdes synthétisés chimiquement représentent 90% du marché total alors que la demande pour des caroténoïdes naturels se développe rapidement en raison de la préoccupation concernant les caroténoïdes synthétiques pour la santé. Comme la plupart des isoprénoïdes ont été découverts à l'origine chez les plantes, leur

l'extraction à partir de matières végétales reste une voie de production majeure. Cependant, il devient de plus en plus difficile de répondre à la demande par leur extraction à partir des plantes en raison des faibles rendements. La production d'isoprénoïdes à partir de leur producteur naturel peut être limitée par la quantité de production et la rentabilité du bioprocédé. Il y a donc beaucoup d'intérêts à développer la production microbienne d'isoprénoïdes (Maury et al., 2005). Les microbes constituent une excellente alternative pour surmonter cette limitation. La bioproduction microbienne nécessite peu de ressources en terres et en eau. Les microbes produisent naturellement les blocs précurseurs de tous les isoprénoïdes que sont l'IPP et le DMAPP. De nouvelles structures de composés bioactifs peuvent être produites en trouvant ou en créant de nouvelles activités enzymatiques (Schmidt-Dannert et al., 2000). Le β -carotène naturel peut être obtenu par des procédés biotechnologiques utilisant des champignons filamenteux, des levures, des bactéries ou des micro-algues (Schweiggert and Carle, 2016). Des équipes de recherches se focalisent sur l'optimisation de la bioproduction du β -carotène en utilisant ces producteurs naturels. D'autres essaient d'implémenter les voies de biosynthèse issues de ces organismes chez des microorganismes hôtes modèles. Nous allons présenter les différentes enzymes impliquées dans la biosynthèse du β -carotène.

I.4.3 La voie de biosynthèse du β -carotène

I.4.3.1 Phytoène synthase

La phytoène synthase catalyse la première étape de la voie de biosynthèse. Des séquences sont conservées entre des phytoène synthases de plantes, d'algues, de cyanobactéries (PSY) et de bactéries (CRTB). Des similitudes de séquences sont aussi observées avec des squalènes synthases (F. X. Cunningham and Gantt, 1998). Le géranylgeranyl pyrophosphate (GGPP) est le précurseur immédiat des caroténoïdes. Deux molécules de GGPP sont condensées pour former l'intermédiaire pré-phytoène pyrophosphate suivi de l'élimination subséquente du groupe diphosphate et l'abstraction stéréospécifique d'un proton pour donner le 15-cis phytoène (C40). Le 15-cis phytoène (7,8,11,12,70, 80, 110, 120-octahydro-cc- carotène) est ainsi un hydrocarbure symétrique incolore contenant trois doubles liaisons conjuguées. Son squelette C40 constitue la structure de base à partir de laquelle tous les caroténoïdes C40 sont dérivés (figure 20 A).

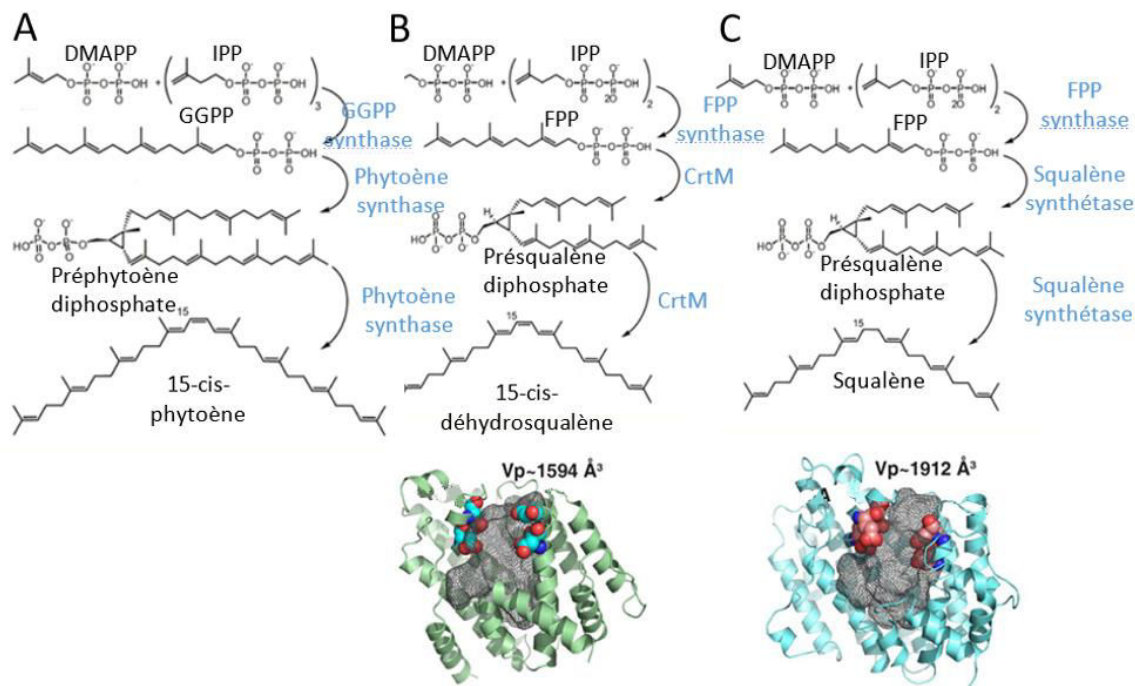


Figure 20 : Comparaison des réactions catalysées par la phytoène synthase, CrtM et la squalène synthase humaine ainsi que des structures de CrtM et de la squalène synthase. Molécules (en noir) et enzymes (en bleu). Sur les structures tridimensionnelles, les cavités formées au centre des protéines contiennent les sites actifs. La cavité de la squalène synthase est légèrement plus grande pour pouvoir incorporer un NADPH nécessaire pour la dernière étape de formation de la squalène. Tirée de (Lawrence et al., 2014; Moise et al., 2014).

Il n'existe pas de structure tridimensionnelle de phytoène synthase. La déhydrosqualène (4-4'-diapophytoène) synthase (CrtM) et la squalène synthase (SQase) possèdent des similitudes avec la phytoène synthase notamment des similitudes réactionnelles (figure 20 B et C). Leurs structures tridimensionnelles ont été résolues (Moise et al., 2014) (figure 20 B et C). L'enzyme CrtM est retrouvée chez des bactéries non phototrophes comme *Staphylococcus aureus* et *Streptococcus faecium* et catalyse la formation du caroténoïde C30-4,4'-diapophytoène (déhydrosqualène) par condensation de deux molécules de FPP. Cette molécule sert de précurseur à une série de rares caroténoïdes C30, les staphyloxanthines, produites par *S. aureus* qui est pathogène pour l'homme. *Staphylococcus aureus* produit des infections nosocomiales et hospitalières, et sa résistance à la méthicilline constitue une grave menace pour la santé publique. Le caroténoïde doré staphyloxanthine favorise la résistance aux espèces réactives de l'oxygène et à la destruction par les neutrophiles chez les hôtes. Les bactéries ne produisant pas ce caroténoïde poussent normalement, mais sont détruites par les espèces réactives de l'oxygène provenant des neutrophiles de l'hôte (Liu et al., 2005). Le blocage de la biosynthèse de staphyloxanthine constitue donc une cible thérapeutique potentielle d'où les intérêts qui ont mené à l'obtention de la structure tridimensionnelle de CrtM (Liu et al., 2008) Dans un système hétérologue d'expression combinant une GGPP synthase, CrtM et une phytoène désaturase bactérienne chez *E. coli*, on a détecté la production de traces de lycopène, rappelant une similarité de réaction entre la phytoène synthase et CrtM (Umeno et al., 2002). Une similarité fonctionnelle avec la squalène synthase humaine (SQS) existe aussi (figure 20) (Blagg et al., 2002). L'importance de la squalène synthase dans le métabolisme du cholestérol a encouragé la recherche sur le mécanisme, la résolution de la structure et

la régulation de l'enzyme (Tansey and Shechter, 2001). Certains inhibiteurs de la squalène synthase humaine inhibent CrtM. Les structures tridimensionnelles de CrtM et de la squalène synthase sont similaires. Elles peuvent être utilisées pour obtenir des modèles par homologie de phytoène synthases. Le mécanisme réactionnel catalysé par la squalène synthétase équivaut à celle de CrtM. Seule la stéréochimie de la dernière étape réactionnelle est différente. Elle est contrôlée par une molécule de NADPH qui permet la réduction et l'inversion de la configuration du 15-cis-dehydrosqualène en squalène (Blagg et al., 2002). La structure de la squalène synthétase comporte un seul domaine formé essentiellement d'hélice α et est caractérisée par un large tunnel central. Les deux sites actifs des deux demi réactions catalysées par la squalène synthétase sont localisés dans ce tunnel. Les extrémités N terminale et C terminale de la protéine ont été tronquées pour obtenir la protéine sous sa forme soluble. Ces deux extrémités se retrouvent du côté de la membrane du réticulum endoplasmique (Pandit et al., 2000) (figure 21).

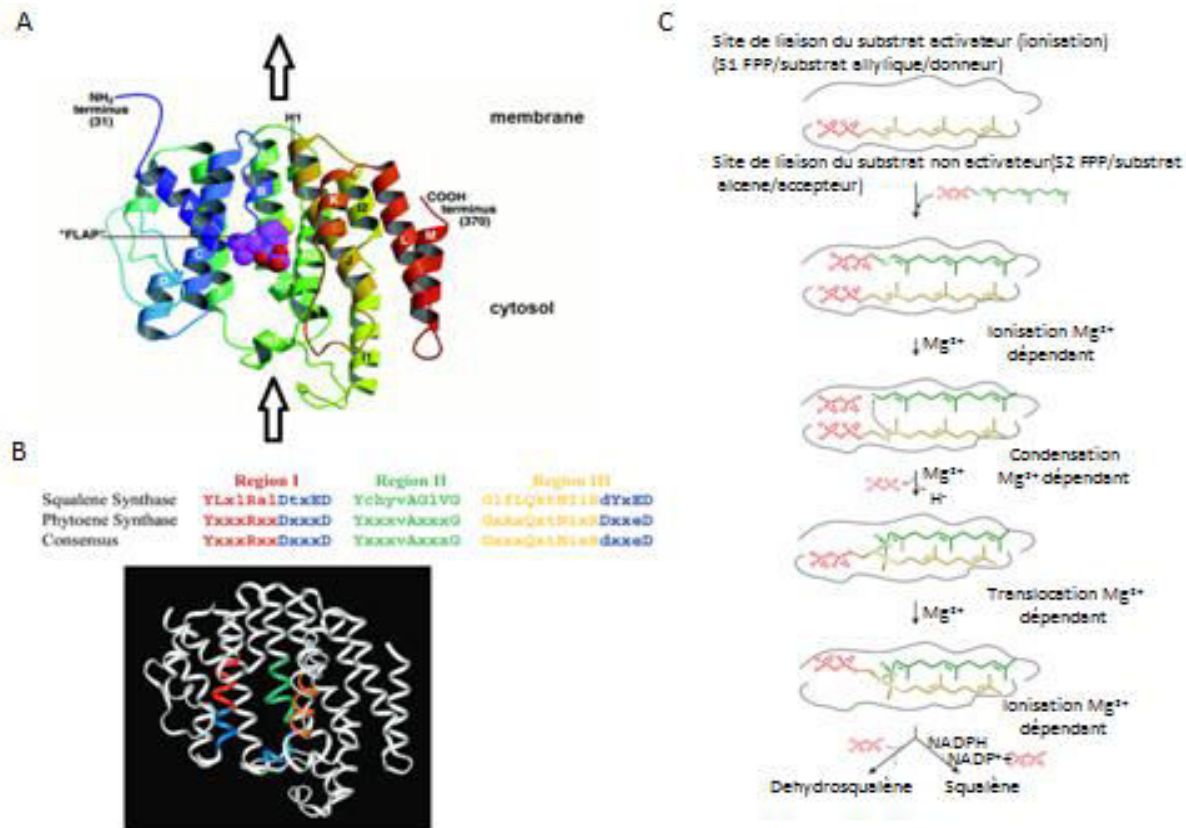


Figure 21 : Caractéristiques structurales de la squalène synthase humaine. A) La structure doublement tronquée à ses extrémités de la squalène synthase humaine est composée essentiellement d'une association d'hélice α et diffère des autres enzymes à NADPH par l'absence d'un domaine structuré en Rosmann fold. L'enzyme se positionne par rapport à la membrane pour accueillir le substrat soluble FPP cytosolique et relargue le squalène du côté membranaire. B) Structure de la squalène synthase humaine montrant le « fold isoprénoidique ». Une cavité centrale entourée de 3 régions conservées constitue le corps hydrophobique des sites actifs. Les séquences conservées riches en Aspartate (80DXXED84) et (219DXXED223) des hélices αC et αH font face à la cavité centrale et avec les groupements carboxyliques éloignés de 11Å. C) Deux sites actifs au sein de la même enzyme accueillent initialement les deux substrats FPP, l'un constituant le donneur allylique et l'autre l'accepteur de prényle. Le groupement phosphate du prényle donneur interagit avec un groupement métallique coordonné par un motif conservé DXXXD. La réaction passe par un intermédiaire réactionnel qui est ensuite réarrangé et qui donne après la perte d'un proton une molécule de configuration cis en position 15. Adapté de (Blagg et al., 2002).

Jusqu'à ce qu'un anticorps anti squalène synthase devienne disponible, les tentatives de localisation de l'enzyme dans la cellule ont été faites par dosages de l'activité enzymatique dans les fractions subcellulaires. Un anticorps généré avec une squalène synthase tronquée purifiée a ensuite été utilisé pour localiser la squalène synthase hépatique de rat *in situ* (Stamellos et al., 1993). La microscopie immunoélectronique de foies de rats normaux a montré que la squalène synthase était associée uniquement à la membrane du réticulum endoplasmique; aucune squalène synthase n'a été détectée dans les peroxysomes. La possibilité de libérer une forme active de squalène synthase à partir de membranes microsomaes par digestion trypsique a suggéré que la squalène synthase est ancrée à la membrane du réticulum endoplasmique avec son domaine catalytique orienté vers le cytosol. Cette orientation permettrait à la squalène synthase d'accepter le FPP soluble et le NADPH provenant du cytosol et de libérer le squalène lipophile dans la membrane du réticulum endoplasmique, reflétant la position clé de l'enzyme dans la biosynthèse du cholestérol pour le passage des intermédiaires de solubles à insolubles (figure 21A).

1.4.3 2 Phytoène désaturase

Il existe plusieurs types de désaturases. La phytoène désaturase de la bactérie *Pantoea ananatis* et celles des champignons et levures sont appelées des désaturases de type CrtI. Le 15-cis phytoène est isomérisé en trans-phytoène par la phytoène désaturase puis une série de quatre deshydrogénations alternativement à gauche et à droite (positions C-11,12 et C-11',12' puis positions C-7,8 et C-7',8') des trois insaturations présentes au centre du phytoène sont réalisées (figure 22) (Fraser et al., 1992). Ces quatre désaturations successives sur le phytoène donnent le phytofluène, le ζ-carotène, le neurosporène et le lycopène.

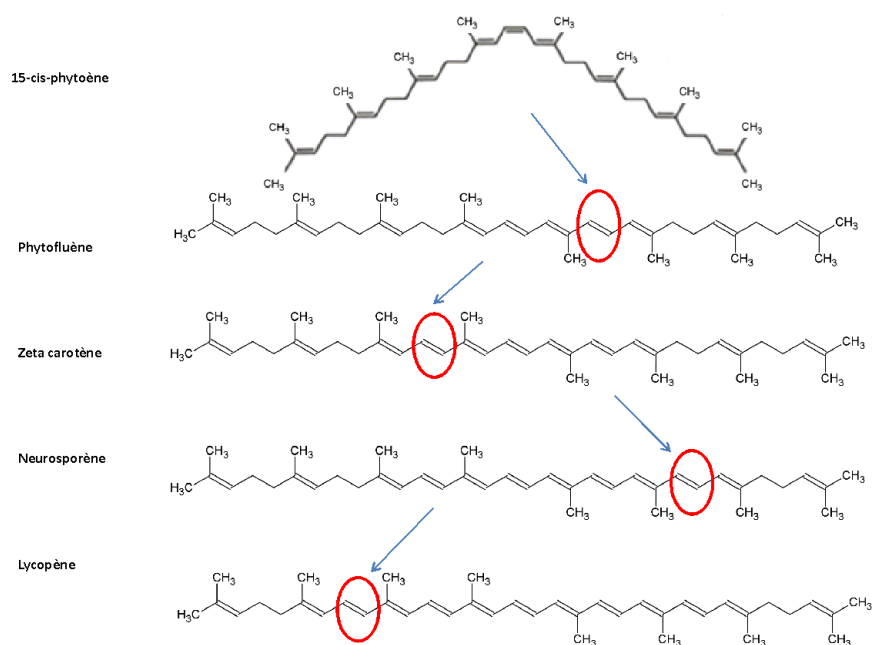


Figure 22 : Les étapes de la synthèse de lycopène chez les phytoène désaturases de bactéries à partir du 15-cis phytoène. Isomérisation suivie de 4 désaturations successivement à des positions précises (cercle en rouge).

Chez les organismes photosynthétiques (les plantes, les algues ou les cyanobactéries), les quatre désaturations formant le lycopène sont introduites par deux désaturases distinctes en deux étapes respectives (figure 23).

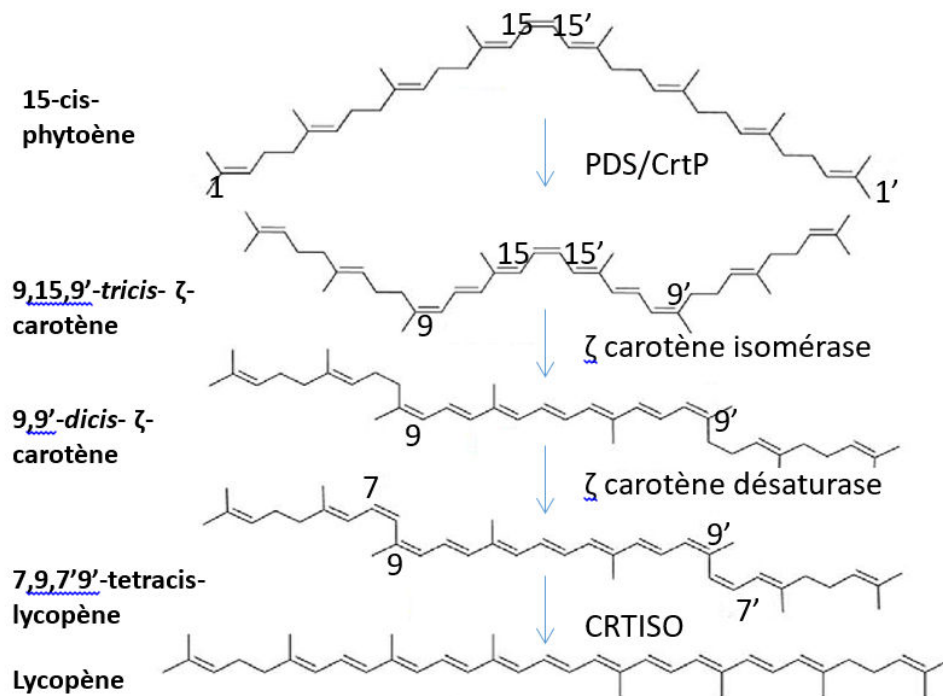


Figure 23 : Etapes de la synthèse du lycopène à partir du 15-cis phytoène chez les organismes photosynthétiques. La synthèse du lycopène nécessite 2 types de désaturases (PDS/CrtP et ζ-carotène désaturase) et deux isomérases (ζ carotène isomérase et CRTISO)

Ces deux phytoènes désaturases sont la désaturase de type PDS/CrtP qui introduit les deux premières désaturations et la ζ carotène désaturase qui introduit les deux dernières désaturations (Bartley et al., 1999). La désaturase de type PDS (chez les plantes) ou de type CrtP (chez les cyanobactéries) agissent sur le 15-cis-phytoène pour ajouter deux insaturations aux positions C-11,12 et C-11',12' en ajoutant à chaque fois deux isomérisations en cis aux positions C-9,8 et C-9',8' et donnant alors le 9,15,9'-tricis- ζ carotène. Une isomérase appelée ζ carotène isomérase agit en position C-15,15' pour donner le 9,9'-dici- ζ carotène qui est ensuite substrat de la ζ carotène désaturase. Cette dernière introduit deux doubles liaisons en cis en position C-8,7 et C-8',7' pour donner le 7,9,7'9'-tetracis- lycopène. L'isomérase CRTISO agit ensuite pour donner le trans-lycopène.

Les structures tridimensionnelles de la phytoène désaturase de la bactérie *Pantoea ananatis* de type CrtI (Schaub et al., 2012a) (figure 24A) et de la phytoène désaturase PDS de l'espèce de riz *Oryza sativa* de type PDS/CrtP (figure 24B) (Brausemann et al., 2017; Gemmecker et al., 2015) ont été résolues.

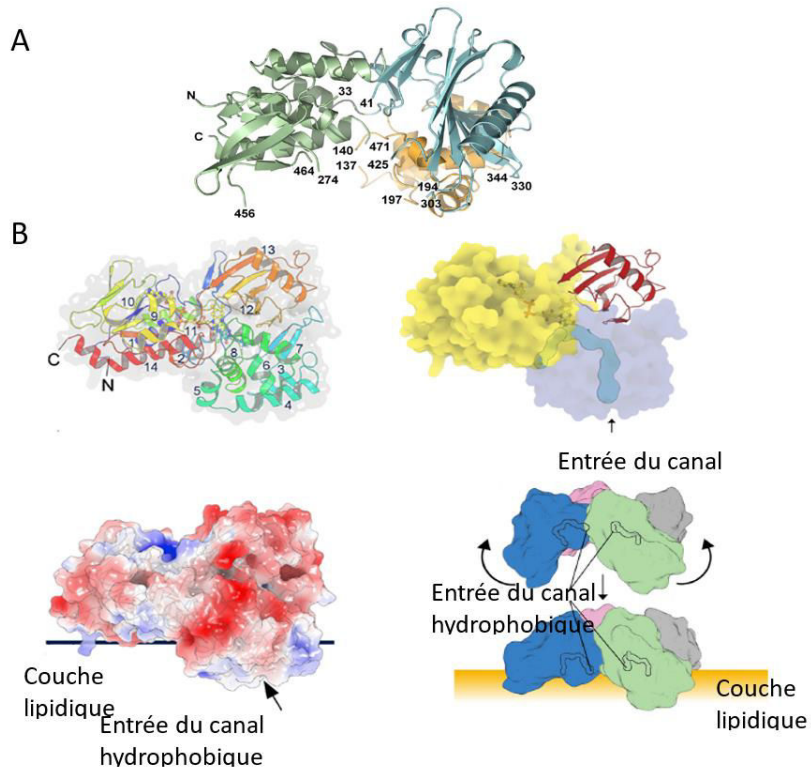


Figure 24 : Détails structuraux de CrtI et de la PDS. A. Structure de CrtI. Les pseudodomains sont colorés en bleu (liaison au substrat), en orange (hélicoïdal non conservé ou de liaison membranaire) et vert (liaison au FAD). B. Structure de PDS. Un monomère est coloré du bleu au rouge de l'extrémité N-terminale à l'extrémité C-terminale dans la représentation en comics avec la représentation de la surface moléculaire en gris clair. Le FAD et le norflurazon sont montrés dans une représentation boules-bâtonnets. Les pseudodomains sont colorés en rouge (domaine Hotdog fold), en bleu (canal hydrophobe du PDS) et jaune (domaine Rossman fold). Représentation du potentiel électrostatique de la PDS et de la possible insertion de l'enzyme dans la membrane. Arrangement tétramérique de la PDS. La PDS est montrée comme un tétramère avec chaque monomère coloré individuellement. Les entrées des canaux hydrophobes pointent vers le centre du tétramère. Aux surfaces membranaires, une rotation de chaque monomère est nécessaire pour positionner l'entrée dans la bicouche lipidique.

Ces deux types de phytoène désaturases diffèrent sur plusieurs aspects : le nombre de doubles liaisons formées, l'implication de différents isomères géométriques, l'utilisation d'accepteurs d'électrons différents. Malgré ces différences et le peu d'homologie de séquence, leur structure montrent un même folding composé de trois domaines (figure 24). Les différences se retrouvent essentiellement au niveau de résidus du site actif susceptibles d'être impliqués dans la réaction catalytique (Brausemann et al., 2017). Avec une meilleure résolution, la structure de la PDS de *Oryza sativa* indique une protéine qui interagit avec la membrane de façon monotopique. Un long canal hydrophobe permet l'entrée, à partir de la membrane, du substrat phytoène et de son cofacteur plastoquinone (figure 24B). Concernant l'organisation des protéines dans le cristal de la PDS, celle-ci indique un assemblage en homotétramère. D'après la structure du substrat dans le site actif, la réaction nécessiterait un homodimère pour être fonctionnelle : le substrat rentre entièrement dans le canal pour être modifié sur la moitié de la molécule, puis sort du canal pour être modifiée sur son autre moitié au sein du canal de l'homodimère d'en face. Le FAD réduit durant la première réaction est réoxydé par une plastoquinone qui rentre par le même tunnel pour régénérer le cofacteur permettant un site actif d'être prêt à nouveau pour commencer un cycle réactionnel.

1.4.3.3 Lycopène cyclase

La cyclisation du lycopène par la lycopène β -cyclase forme un cycle β -ionone à une extrémité pour donner le γ -carotène puis un autre cycle β -ionone à l'autre extrémité pour donner le β -carotène (figure 25).

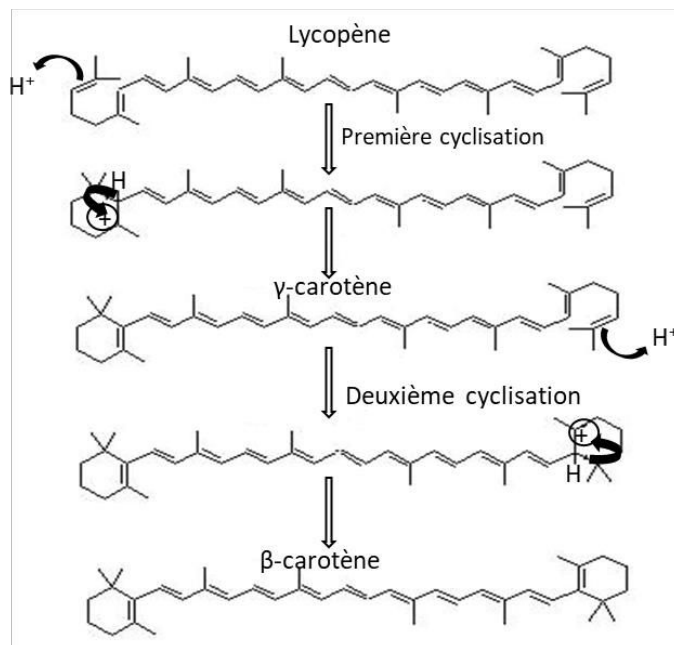


Figure 25 : Etapes de la synthèse du β -carotène à partir du lycopène. Une première cyclisation est catalysée au niveau d'une extrémité suivie d'une deuxième cyclisation à l'autre extrémité de la molécule.

1.4.4 La production de β -carotène chez la levure *S. cerevisiae*

S. cerevisiae est un hôte important en biotechnologie industrielle (Hong and Nielsen, 2012; Nielsen, 2015) (tableau 6). Les connaissances accumulées sur cet organisme modèle de laboratoire ont contribué au fait que *S. cerevisiae* est utilisée depuis longtemps comme organisme châssis pour la production d'isoprénoides (Kampranis and Makris, 2012). Un exemple est la bioproduction de l'artémisinine à un niveau viable économiquement, une référence en ingénierie métabolique. La levure *S. cerevisiae* utilise la voie du mévalonate pour générer les précurseurs IPP et DMAPP à partir de l'acétyl-CoA. Des stratégies d'ingénierie métabolique des caroténoïdes, et des isoprénoïdes en général, consistent à optimiser la voie en amont du précurseur puis à optimiser la voie de biosynthèse caroténogénique hétérologue en elle-même. L'HMG-CoA réductase (HMGR) et la GGPP synthase constituent des enzymes clés pour l'ingénierie de la biosynthèse des isoprénoïdes chez *S. cerevisiae* (Özaydin et al., 2013). HMGR est constituée d'un domaine transmembranaire et d'un domaine catalytique. Une version tronquée de la 3-hydroxy-3-méthyl-glutaryl-CoA réductase (tHMGR) améliore l'efficacité de la voie du mévalonate et le titre des produits correspondants. Grâce à la troncature du domaine transmembranaire, la protéine tronquée est relocalisée dans le cytosol sous forme soluble, et les caractéristiques régulatrices de HMGR n'ont plus été observées. En dérégulant HMGR, on a observé une augmentation significative de la production d'une série de médicaments et de produits chimiques, comme le squalène (Donald et al.,

1997; Rico et al., 2010), l'artémisinine (Ro et al., 2006) et les précurseurs du taxol (Engels et al., 2008) et de l'acide ganodérique (Xu et al., 2012) ainsi que de caroténoïdes (Verwaal et al., 2007).

Pour produire du β -carotène chez *S. cerevisiae*, les gènes caroténogéniques bactériens de *Erwinia uredovora* ont été introduites (Yamano et al., 1994). Une faible production de β -carotène (103 $\mu\text{g/g}$ de cellule sèche) a été observée et 11% de phytoène et 11% de lycopène en plus des 78% de β -carotène ont été mesurés après 3 jours de culture (Yamano et al., 1994). En utilisant des gènes caroténogéniques de la levure *Xanthophyllomyces dendrorhous*, plus de 500 $\mu\text{g/g}$ de cellule sèche à 11mg/g de cellule sèche sont obtenus après des ingénieries d'optimisations (Verwaal et al., 2007). Avec ces gènes caroténogéniques de levure, une souche initiale sans optimisation a donné 508 $\mu\text{g/g}$ de cellule sèche de caroténoïdes totaux avec 94% de phytoène et 3 % de β -carotène (Verwaal et al., 2007). La souche finale avec tHMG1 (HMGR tronqué) et CrtE (GGPP synthase de *X. dendrorhous*) surexprimés, puis une copie de CrtI en plus (deuxième étape dans la voie du β -carotène) a donné 5918 $\mu\text{g/g}$ de cellule sèche de β -carotène avec 5380 $\mu\text{g/g}$ de cellule sèche de phytoène accumulé, soit à peu près 50% du total des caroténoïdes produits non transformés en β -carotène (Verwaal et al., 2007). Une étude plus récente a utilisé les mêmes gènes caroténogéniques de *X. dendrorhous*. Dans cette étude, le groupe de Yu a développé un système de plasmides intégratifs permettant des intégrations de gènes à différents locus pour limiter des événements de recombinaison non désirés. Ils ont appelé la stratégie un système décentralisé (Xie et al., 2014). Leur système utilise un promoteur Gal pour contrôler l'expression des gènes. Leur meilleure souche a produit 4.9 mg/g de cellule sèche de β -carotène pour 8 mg/g de cellule sèche de caroténoïdes totales soit environ 40% d'intermédiaire accumulé.

Dans chacune de ces études, l'accumulation des intermédiaires est observée. Notamment, une importante accumulation de l'intermédiaire phytoène. La désaturation du phytoène semble constituer l'étape limitante dans la production de β -carotène hétérologue par *S. cerevisiae*. L'activité phytoène désaturase détermine donc la vitesse de l'ensemble de la voie métabolique. Les enzymes caroténogéniques de *X. dendrorhous* permettent une meilleure production en caroténoïdes chez la levure *S. cerevisiae* par rapport aux enzymes bactériennes. Une ingénierie enzymatique de ces enzymes pourrait permettre d'améliorer cette production.

1.4.5 Les enzymes caroténogéniques de *X. dendrorhous*

La levure hétérobasidiomycète *Xanthophyllomyces dendrorhous* (forme imparfaite: *Phaffia rhodozyma*) produit comme caroténoïde principal l'astaxanthine, un dérivé oxydé du β -carotène. Les groupements oxydés sont une fonction céto et une fonction hydroxyle sur chaque cycle β ajoutées par une enzyme appelée CrtS. Le torulène, une molécule avec une insaturation en plus par rapport au α -carotène est aussi détecté chez espèce à de faibles quantités. L'astaxanthine est un additif alimentaire largement utilisé dans les aquacultures de saumon, de truite et de crevette (Andrewes and Starr, 1976). L'astaxanthine est le caroténoïde le plus important après β -carotène en terme de marché (Schmidt et al., 2011).

Les champignons et les levures, dont *X. dendrorhous*, possèdent une particularité dans leurs enzymes caroténogéniques. Les activités enzymatiques de phytoène synthase et de lycopène cyclase sont assurées au sein d'un même polypeptide. Cette particularité a été démontrée par des études de

complémentation hétérologues du gène *CrtYB* de *X. dendrorhous* dans des souches recombinantes d'*E. coli* (Verdoes et al., 1999). Dans *CrtYB*, l'activité phytoène synthase réside dans la partie C-terminale, et l'activité lycopène cyclase est restreinte à l'extrémité N-terminale. La partie *CrtY* en N-terminale est constituée de plusieurs segments transmembranaires contrairement à la partie C terminale portant l'activité phytoène synthase. Des gènes similaires codant des enzymes bifonctionnelles ont été ensuite décrits chez d'autres champignons, tels que l'ascomycète *N. crassa* (Schmidhauser et al., 1994) et les zygomycètes *M. circinelloides* (Velayos et al., 2000) et *P. blakesleeana* (Arrach et al., 2001). La large distribution de ce gène bifonctionnel tout au long du règne fongique suggère que cette enzyme bifonctionnelle a été acquise au début de l'évolution des champignons (Krubasik and Sandmann, 2000). Cette fusion naturelle entre *CrtB* et *CrtY* procure elle un avantage réel par rapport à un système séparé comme chez d'autres espèces caroténogéniques? Et dans le contexte d'un environnement hétérologue pour des enzymes, et dans lequel les interactions spécifiques entre les enzymes peuvent être perturbées ou non favorisées, la présence de la fusion naturelle permet-elle de faciliter ces interactions interprotéiques en maintenant une colocalisation des enzymes?

Concernant la phytoène déshydrogénase, le gène *CrtI* de *X. dendrorhous*, a également été cloné par complémentation dans *E. coli* (Verdoes, 1999). Le système hétérologue montre que le gène est suffisant pour catalyser les quatre étapes successives d'oxydation du phytoène au lycopène. Donc deux enzymes *CrtYB* et *CrtI* sont impliquées dans la synthèse du γ -carotène ou du β -carotène chez *X. dendrorhous*. Pour décrire la voie caroténogénique complète chez *X. dendrorhous*, décrivons le système jusqu'à la production de l'astaxanthine. Le lycopène est un point de branchement de la voie caroténogénique, car il agit comme précurseur commun pour des caroténoïdes cycliques tels que le γ -carotène, le β -carotène, le torulène et l'astaxanthine. Chez la levure *X. dendrorhous*, l'astaxanthin synthase (*CrtS*), une enzyme cytochrome P450, en combinaison avec son cytochrome P450 réductase, sont responsables de la biosynthèse de l'astaxanthine (Alcaíno et al., 2012). Dans les levures et autres champignons, des caroténoïdes monocycliques et carboxyliques comme le torulène et le torularhodin sont fréquemment rencontrés (Schweiggert and Carle, 2016). L'existence d'une voie monocyclique divergeant de la voie dicyclique au neurosporène et passant par le β -zeacarotène, le 3,4-didéhydrolycopène, le torulène, le 3-hydroxy-3', 4'-didéhydro- β , ψ -carotène-4-one (HDCO) au produit final 3,3'-dihydroxy- β , ψ -carotène-4,4'-dione (DCD) a été discutée (An et al., 1999). Chez l'espèce *Rhodotorula* par contre, la voie biosynthétique de monocyclisation du lycopène entraîne la formation de γ -carotène, qui est ensuite désaturé pour donner du torulène (Moliné et al., 2012).

1.4.5 1 Localisation et organisation quaternaire des enzymes

Des études récentes sur l'organisation quaternaire des enzymes caroténogéniques sont peu nombreuses alors que la compréhension des mécanismes exacts de fonctionnement de ces enzymes pourrait aider l'ingénierie de la bioproduction de caroténoïdes (Shumskaya and Wurtzel, 2013). Une hypothèse de longue date soutient que la caroténogénèse se produit dans des agrégats d'enzymes associés aux membranes. Les enzymes caroténogéniques sont probablement impliquées dans des complexes protéiques. Cette hypothèse est soutenue par la détection limitée d'intermédiaires de la voie chez les producteurs naturels et par la preuve de channeling in vivo (Candau et al., 1991; Guardia et al., 1971), ainsi que l'observation de complexe de haut poids moléculaire contenant des enzymes biosynthétiques

(Bonk Michael et al., 2004; Lopez Alex B. et al., 2008; Maudinas et al., 1977). Des études sur la localisation subcellulaire des enzymes caroténogéniques ont été réalisées chez les levures *Phycomyces blakesleeanus* (Riley and Bramley, 1976) et *Neurospora crassa* et ont montré que l'essentiel de l'activité caroténogénique se retrouve dans la fraction microsomale, au niveau des membranes du réticulum endoplasmique notamment (Mitzka-Schnabel and Rau, 1981; Rau and Mitzka-Schnabel, 1985).

Une souche sauvage de *Phycomyces blakesleeanus* produit du β carotène. *Phycomyces blakesleeanus* est un champignon saprophyte, hétérothallique bipolaire (+, -) appartenant à l'espèce *Phycomycète* (Mucorales). Des études utilisant des hétérocaryons de ce champignon ont permis l'obtention de différents mélanges d'enzymes caroténogéniques issues de souches sauvages, ou mutées à certaines étapes de la voie de biosynthèse. Il a été montré à partir des résultats d'analyses quantitatives de caroténoïdes issus de différents hétérocaryons que le modèle pouvant expliquer les résultats obtenus était celui d'une biosynthèse se faisant au sein d'un agrégat d'enzyme caroténogénique fonctionnant comme une chaîne d'assemblage dans lequel chaque enzyme reçoit son substrat de la précédente et passe son produit à l'enzyme suivante. Cette chaîne d'assemblage contiendrait deux copies du produit de gène de la cyclase formant un agrégat d'enzymes, et agissant séquentiellement sur le même substrat (Guardia et al., 1971) (figure 26A). Dans le modèle, les intermédiaires et les substrats ne peuvent pas être transférés d'un agrégat à un autre, ni repris à partir du cytoplasme environnant. Il a aussi été démontré que quatre copies de la phytoènes désaturases, assemblés dans un complexe enzymatique, agissent séquentiellement dans la conversion du phytoène au lycopène (Aragón et al., 1976; Eslava and Cerdá-Olmedo, 1974; Sanz et al., 2002). En utilisant encore des hétérocaryons du champignon *Phycomyces blakesleeanus*, une autre étude a mesuré la concentration des intermédiaires réactionnels en présence cette fois-ci d'inhibiteurs. De même, l'étude a proposé que les résultats mesurés peuvent s'expliquer par un fonctionnement en complexe des enzymes caroténogéniques menant au β -carotène (Candau et al., 1991). Un modèle du complexe enzymatique formant le β -carotène a été ensuite proposé par certains auteurs (Iturriaga et al., 2000) (figure 26B).

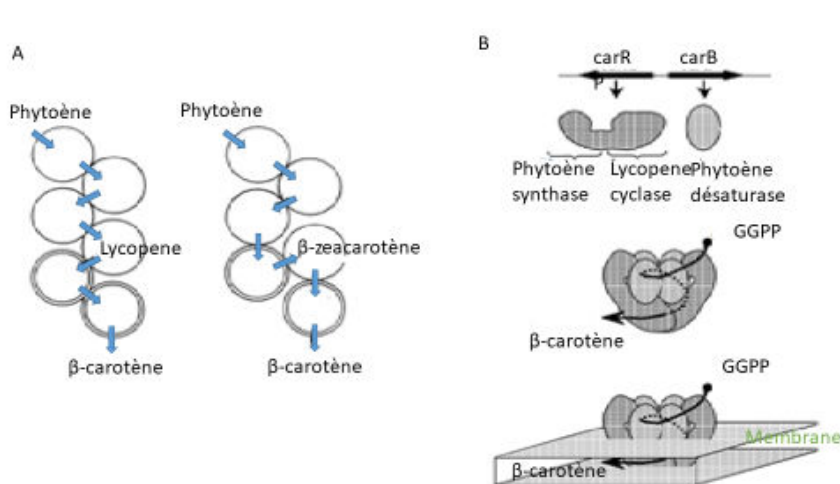


Figure 26 : Organisation en complexe des enzymes caroténogéniques formant le β -carotène chez les fungi proposés par des études ultérieures. A. Le transfert des molécules d'enzyme à enzymes dans le complexe (flèches bleues). B. Représentation du complexe formé par *carR* (phytoène synthase/lycopène cyclase) et *carB* (phytoène désaturase). Le transfert des intermédiaires du GGPP au β -carotène est indiqué.

Chez les plantes, un modèle schématique de la conception de ces complexes caroténogéniques hypothétiques est représenté (figure 27) (F. X. Cunningham and Gantt, 1998). Les trois enzymes "solubles" de la voie qui catalysent la formation de phytoène ont été isolés ensemble dans des complexes multienzymatiques (Camara, 1993; Dogbo and Camara, 1987; Maudinas et al., 1977). La preuve des complexes contenant les désaturases associées à la membrane, les cyclases et les hydroxylases (enzymes suivants dans la voie) est moins directe (Candau et al., 1991; Kleinig, 1989). Le GGPP synthase est un dimère (Dogbo and Camara, 1987). Le complexe contenant la phytoène synthase soluble possède une copie de PSY, deux copies pour la GGPP synthase. Le produit final de ces réactions, le cis-phytoène, est un composé liposoluble qui doit en quelque sorte atteindre les désaturases et/ou la membrane. L'association de PSY-membranaire observée chez la jonquille (Schledz et al., 1996) suggère qu'un tel complexe est effectivement associé aux membranes. Les agrégats d'enzymes postulés de désaturases / cyclases associés à la membrane sont montrés pour contenir deux copies de chaque désaturase et deux sous-unités cyclases. Il peut bien y avoir d'autres composants polypeptidiques permettant de réguler les capacités d'interaction, agissant en tant que cofacteurs, ou qui interagissent avec les sous-unités catalytiques. Deux types de complexes désaturase /cyclase sont affichés pour tenir compte de la possibilité de former l' α -carotène chez les plantes.

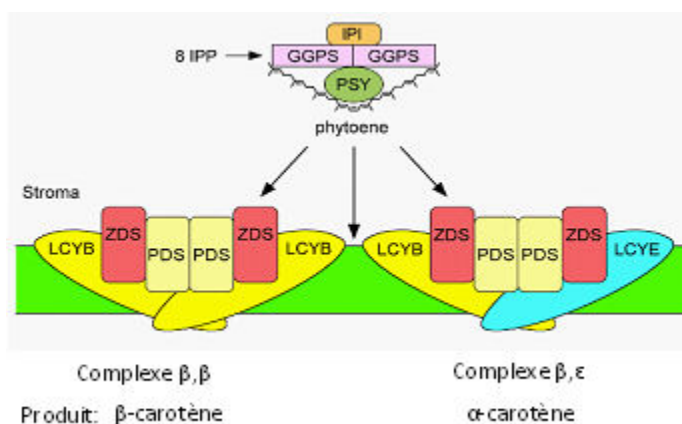


Figure 27: Illustration schématique du modèle hypothétique de complexes multienzymatiques caroténogéniques dans le membranes et stroma de thylakoïde de plantes et d'algues. GGPS, géranylgéranyl pyrophosphate synthase; IPI, isopentényle pyrophosphate isomérase; LCYB, lycopène β cyclase; LCYE, lycopène ϵ la cyclase; PDS, phytoène désaturase; PSY, phytoène synthase; ZDS, ζ -carotène désaturase. Tirée de (F. X. Cunningham and Gantt, 1998)

I-5 Présentation du projet de thèse

L'objectif principal de ce travail de thèse est celui d'étudier différentes stratégies de colocalisation spatiale d'enzymes afin de diminuer l'accumulation d'intermédiaires réactionnels dans une voie de biosynthèse et d'améliorer le flux de production. Dans le système enzymatique étudié, quelle stratégie pourrait être efficace pour colocaliser spatialement les enzymes et comment valider l'efficacité des systèmes artificiels créés.

Une première partie consiste en la présentation du matériel et des méthodes utilisées pour l'étude. La partie résultat débute par une étude bioinformatique sur les enzymes dans laquelle des modèles par homologie ont été obtenus. Une analyse de séquences a été utilisée pour définir la région limite entre les domaines de la fusion naturelle CrtYB.

Une deuxième partie expose les résultats obtenus par une stratégie de colocalisation des enzymes utilisant des domaines d'interaction fusionnés à ces dernières. La stratégie suivante a consisté en la colocalisation d'enzymes en les fusionnant avec une protéine oligomérique.

Dans une troisième partie des résultats, différentes fusions directes entre les enzymes ont été construites en variant plusieurs paramètres tels que le type de linker ou l'orientation des enzymes. Cette approche a été divisée en deux étapes. Dans un premier temps, nous avons séparé le domaine CrtB du domaine CrtY dans la fusion naturelle CrtYB. Cela a permis de mesurer la fonctionnalité de ces enzymes lorsqu'elles sont séparées. De plus, en absence de CrtY, l'étape de biotransformation du lycopène en β -carotène est éliminée et a permis une étude centrée sur l'étape limitante allant du phytoène au lycopène. Dans un second temps, les caractéristiques de bifusions synthétiques ont été caractérisées et la fusion optimale alors obtenue a été réutilisée pour créer une fusion tripartite d'enzyme. L'intégrité des différentes fusions construites a été vérifiée et le niveau d'expression protéique de chaque construction a été mesuré par western blot. Une mesure plus fine du niveau de production de l'enzyme tripartite optimale a été faite par une technique de protéomique.

Une quatrième partie des résultats expose différents tests de production de CrtB et de CrtI chez *E. coli* pour les obtenir sous forme soluble. Après les résultats, la partie suivante concernera une discussion et les perspectives générales.

II MATERIEL ET METHODES

II-1 Matériel

II.1.1 Souches de levure

La souche de levure de départ utilisée dans cette étude a été BY4741 MAT α SUC2 gal2 mal2 mel flo1 flo8-1 hap1 ho bio1 bio6 his3 Δ 1leu2 Δ 0 met15 Δ 0 ura3 Δ 0.

II.1.2 Souches de bactéries

La souche *E. coli* BL21 (DE3) a été utilisée pour les tests de production et de solubilisation des enzymes. F⁻ ompT gal dcm lon hsdSB(rB-mB-) λ (DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+] K-12(λ S). La souche *E. coli* DH5 α F⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 ϕ 80dlacZ Δ M15 Δ (lacZYA-argF)U169, hsdR17(rK-mK+), λ - a été utilisée pour l'amplification des plasmides.

II-2 Méthodes expérimentales

II.2.1 Milieux de culture

Les milieux utilisés pour la culture des bactéries ont été le milieu LB (tryptone 10g/L, chlorure de sodium 10g/L et extrait de levure 5g/L) ou TB (tryptone 20g/L, extrait de levure 24 g/L, glycérol 4mL/L, tampon phosphate 0.017 M KH₂PO₄ 0.072 M K₂HPO₄). Les milieux solides utilisés ont été le LB agar (tryptone 10g/L, chlorure de sodium 10g/L et extrait de levure 5g/L, 15g /L d'agar). Ces milieux sont autoclavés à 121°C pendant 30 min sous une pression de 15 psi. Les milieux ainsi stérilisés ont été supplémentés d'Ampicilline à une concentration de 100 μ g /ml finale. De l'Ampicilline à la même concentration est ajoutée au LB agar autoclavé avant de couler le milieu sur boîte de Petri.

Le milieu YPD ou le milieu YPG (20 g/L peptone, 10 g/L yeast extract, 20 g/L glucose ou galactose respectivement) ont été utilisés pour les cultures de levure. Du YPD ou YPG agar solide (milieu YPD ou YPG + 20 g/L agar) ont été utilisés pour concevoir les milieux solides sur boîte de Petri. Le Geneticin (G418) a été ajouté à 200 μ g/ml dans le milieu YPD pour la sélection des colonies résistantes au Geneticin. Les milieux SD-URA ou SG-URA (1.7 g/L Yeast Nitrogen Base sans acide aminé, 5 g/L sulfate d'ammonium, 20 g/L de glucose (SD) ou de galactose (SG), 0.77 g/L de mix dropout) contenant tous les acides aminés et les acides nucléiques sans l'uracile (MP Biomedicals, Santa Ana, California, USA) et les milieux solides SD ou SG -URA (milieu SD ou SG -URA avec 17 g/L d'agar) ont été utilisés pour sélectionner les colonies URA⁺.

II.2.2 Méthodes concernant les acides nucléiques

Les enzymes de restrictions proviennent de chez Thermo Fisher Scientific. La ligation par In fusion a été réalisée selon le protocole du kit In-Fusion[®] HD Cloning Plus Kit (Ozyme). Les miniprep ont été réalisées avec le kit GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). Les séquençages ont été réalisés chez Eurofins genomics.

II.2.2 1 Construction des fusions

La stratégie générale pour la construction des fusions a été basée sur une stratégie par restriction ou par recombinaison homologue utilisant le kit In fusion. Les étapes pour la construction des fusions d'enzyme sont décrites (Figure 28).

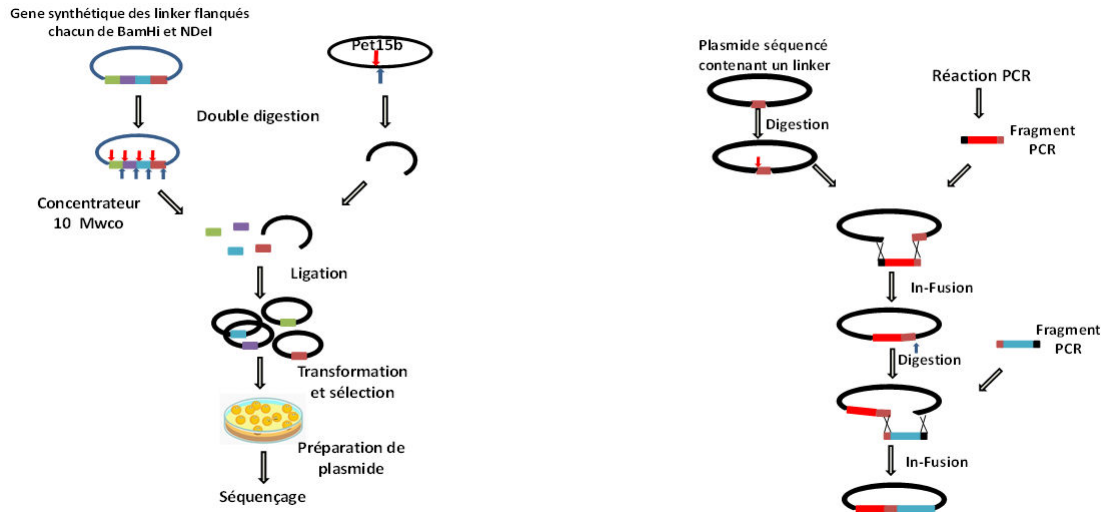


Figure 28 : Etapes de biologie moléculaire pour l'obtention d'une bifusion. Un gène synthétique a été commandé pour obtenir les linkers. Dans le gène synthétique, chaque séquence de linker est entourée des sites de restriction BamHI et NdeI. Après une double digestion par BamHI et NdeI et une migration sur gel, le fragment de gel contenant l'ensemble des différents linkers séparés a été coupé et purifié. Les enzymes ont été dénaturées 5 min à 100°C. L'échantillon contenant les linkers est utilisé pour une ligation dans le plasmide pet15b préalablement ouvert par BamHI et NdeI. Les produits de ligation ont été utilisés pour transformer E. coli. Plusieurs colonies ont été repiquées pour des minicultures afin d'amplifier les plasmides qui ont alors été envoyés en séquençage. Pour chaque linker dans pet15b, une ligation du produit PCR du gène de CrtB ou de CrtI (au niveau du site de BamHI pour insertion en N-terminale ou NdeI pour insertion en C-terminale) dans le plasmide contenant le linker et préalablement digéré par l'enzyme de restriction correspondante est réalisée. Après séquençage, l'autre ORF est également ajouté au gène pour obtenir la bifusion.

II.2.2 2 Transformation de levure

L'intégration des différentes constructions dans le génome a été réalisée par la méthode de transformation de haute efficacité de Gietz (Schiestl and Gietz, 1989) avec des modifications mineures: Inoculer la souche de levure dans 5 ml de milieu liquide YPD et incubé pendant une nuit sur un agitateur rotatif à 200 tr / min et à 30 ° C. Déterminez le titre de la culture de levure en pipetant 10 ml de cellules dans 1,0 ml d'eau dans une cuvette à spectrophotomètre et en mesurant la DO à 600 nm. Pour de nombreuses souches de levure, une suspension contenant 1 x 10⁶ cellules / ml donnera une DO₆₀₀ de 0,1. Transférer 50 ml de YPD préchauffé à 30°C dans un flask et ajouter 2,5 x 10⁸ cellules pour donner 5 x 10⁶ cellules / ml. Incuber dans un agitateur rotatif ou alternatif à 30 ° C et 200 tr / min. Lorsque le titre des cellules est d'au moins 2 x 10⁷ cellules / ml (environ 4 heures), récolter les cellules par centrifugation à 3000 g pendant 5 min, laver les cellules dans 25 ml d'eau stérile et remettre en suspension dans 1 ml d'eau stérile. Transférer la suspension cellulaire dans un tube de 1,5 ml, centrifuger pendant 30 secondes et jeter le surnageant. Ajouter de l'eau jusqu'à un volume final de 1,0 ml et mélanger vigoureusement au vortex pour remettre en suspension les cellules. Pipeter 100 µl d'échantillons dans des tubes à centrifuger de 1,5 ml, un pour chaque transformation, centrifuger à la vitesse maximale pendant 30 secondes et retirer le surnageant. Préparez un mélange de

transformations (240 µL de PEG 3500 50% w/v, 36 µL de Lithium Acétate 1.0 M, 50 µL d'ADN de saumon simple brin) suffisant pour le nombre de transformations planifié plus un extra. Conservez le mélange de transformation dans de la glace / eau. 1 µg d'ADN linéarisés par digestion avec Sfi1 (New England BioLabs, Ipswich, MA, USA) a été utilisé dans un volume maximum de 10 µL. La mixture a été incubée 30 min à 30°C puis les cellules ont subi un choc thermique 20 min dans un bain marie à 42°C. 1ml de YPD est ajouté aux cellules qui ont été ensuite gardées 2 heures à 30°C puis ont été étalées sur du milieu YPD solide contenant du G418 à une concentration de 200 µg/ml.

II.2.2 3 PCR sur cellules

Cette méthode a permis de vérifier la présence d'un insert ou la vérification par séquençage de gènes intégré dans le génome. Aliquoter 20µL d'une solution de NaOH 20mM dans les tubes PCR. En utilisant le bout d'un cône pour pipettes 0.1 à 10µL, piquer une colonie et remuer dans la solution de NaOH pour libérer les cellules. Incuber à 99°C pendant 10 min grâce à une machine PCR. Centrifuger à vitesse maximale pendant 5 secondes. Vortexer 2 fois 15 secondes et centrifuger à vitesse maximum pendant 30 secondes. Utiliser 1µL du surnageant comme matrice ADN dans une réaction PCR de 10µL.

II.2.3 Méthodes concernant les protéines

II.2.3 1 Analyses in silico

Les modèles par homologie ont été obtenus par le portail Phyre2 (Protein Homology/analogy Recognition Engine V 2.0). Les structures ont été analysées en utilisant le logiciel Pymol. Les alignements de séquence ont été réalisés en utilisant le logiciel MAFFT version 7 ou sur ClustalW.

II.2.3 2 Méthode d'expression chez E. coli

Une colonie d'*E. coli* BL21 fraîche transformée avec le vecteur recombinant est repiquée dans 5 ml de préculture (LB) contenant 100µg/ml d'Ampicilline et mis à 37°C pour la nuit sous agitation. La préculture est ensuite utilisée pour inoculer une culture de 50ml (LBA) au millième. La culture est mise à 37°C sous agitation et la croissance bactérienne est mesurée en mesurant la densité optique à 600 nm. A une densité optique de 0.6, on a induit la production de la protéine par ajout d'IPTG (concentration finale de 0.5mM) et la culture est mise à 30°C sous agitation pour la production de la protéine d'intérêt.

II.2.3 3 Test de production et de solubilisation

Pour obtenir le lysat cellulaire, les cellules sont centrifugées pendant 15 minutes à 4500g à 4°C afin de récupérer le culot cellulaire. Après avoir éliminé le surnageant (milieu de culture), le culot est lavé dans du tampon de lyse (20 mM Tris-HCl, pH 8.0, 1 mM PMSF, 1 mg/ml lysozyme, 0.05% Triton X-100) puis est repris dans 30 ml de tampon de lyse et est conservé dans la glace ou stockés à -20°C pour une lyse cellulaire ultérieure.

II.2.3 4 Lyse cellulaire

Le culot cellulaire repris dans le tampon de lyse a été gardé dans la glace et soniqué dans un appareil de type VibraCell 72434 BIOBLOCK SCIENTIFIC. On a soumis les cellules à une seconde d'ultrason intercalée d'une seconde d'arrêt et cela durant trois fois 30 secondes à une puissance de 40%. Le lysat après sonication a été centrifugé à 15500g à 4°C durant 20 min. Le surnageant contenant les protéines

solubles est récupéré et le culot cellulaire contenant la fraction insoluble est repris dans un même volume de tampon de lyse.

II.2.3 5 Electrophorèse SDS-PAGE

12 µl de l'échantillon (fraction soluble ou insoluble) est mélangé avec 3 µl d'une solution de tampon de charge (Tris à 0,20 mol/L, glycérol à 20% en vol, SDS à 4% en masse et du β-mercaptoéthanol à 10% en volume ajouté extra temporairement). L'ensemble a été chauffé à 70°C pendant 10 minutes, puis a été déposé sur gel de polyacrylamide (Novex NuPAGE 4-12% Bis-Tris, Life Technologies, Carlsbad, CA). La migration a été effectuée en tampon MES pendant 35 minutes à 200V.

II.2.3 6 Test de solubilisation de protéines membranaires

Environ 2mg/ml de protéine de la fraction non soluble ont été mis dans des solutions tampons contenant différentes concentrations finales en détergeant (Cholâtes 0.2 %, 0.4 %, 0.6 % et Triton X100 0.1 %, 0.2 %, 0.5 %) et dans un volume de 10 ml pendant 10 min à 4°C. Une fraction est ensuite centrifugée à vitesse max avec une centrifugeuse de paillasse à 4°C et le surnageant et le culot (repris dans le même volume que le surnageant) sont soumis à une électrophorèse sur gel polyacrylamide en présence de SDS.

II.2.3 7 Dénaturation in vitro et refolding des protéines

Une première étape a consisté à isoler les corps d'inclusion. Les cellules complètement lysées et traitées au lysozyme (1/10 vol of lysozyme 10 mg/ml et incubation 10 min à 37 °C). Après une centrifugation, le culot est lavé avec une solution tampon contenant une faible concentration d'un agent chaotrope (urée 0.5 M) ou d'un détergent (Triton X-100 1%). La concentration totale de protéine était d'environ 1mg/ml. La concentration protéique a été obtenue en utilisant le Nanodrop 2000C de Thermo Scientifique pour avoir une estimation de la concentration en mg par ml (en supposant que 1DO280 équivaut à 1 mg/ml). Une deuxième étape a consisté en la resolubilisation des agrégats protéiques. Les corps d'inclusion sont resuspendus et incubés dans une solution tampon (50 mM Tris-HCl, pH 7.5) contenant 20mM de β-mercaptoéthanol et un agent solubilisant (urée 5M ou chlorure de guanidium 1M) pendant 60 min à 30°C. Une centrifugation est réalisée pour éliminer les agrégats non resolubilisés. La troisième étape a consisté au repliement des protéines en enlevant les agents dénaturants par dialyse. Environ 100 µg/ml de protéine est utilisé pour la resolubilisation. L'échantillon est placé dans un boudin de dialyse et mis dans un bain de dialyse de 2L (50 mM Tris-HCl, pH 7.5). Le tout a été mis à 4°C et soumis à une faible agitation. Une fraction est ensuite centrifugée à vitesse max avec une centrifugeuse de paillasse à 4°C et le surnageant et le culot (repris dans le même volume que le surnageant) sont soumis à une électrophorèse sur gel polyacrylamide en présence de SDS.

II.2.3 8 Analyse des protéines par Western Blot

Les extraits protéiques ont été préparés suivant la méthode décrite par Zhang et al (Zhang et al., 2011). Brièvement, les cellules d'un culot cellulaire issu d'une culture de 50 ml ont été resuspendues dans 1 ml de PBS (Invitrogen, Carlsbad, CA, USA) par agitation. Pour normaliser la quantité de cellule avant extraction, la densité cellulaire a été vérifiée par mesure de l'OD à 600 nm. 8 µL de la suspension cellulaire ont été mélangés doucement dans la glace avec 200 µL d'une solution froide de lithium acétate à 2 M, puis ont été laissés 5 min avant une centrifugation 5 min à 5000g à 4°C. Le surnageant a

été éliminé et 200 µL d'une solution de 0.4M de NaOH a été ajoutée. Après une douce resuspension, et 5 min dans la glace, les échantillons ont été centrifugés 5 min à 4°C. Après avoir éliminé le surnageant, les culots ont été vigoureusement vortexés dans 100 µL de bleu de bromophénol supplémentée avec 5 % de β-mercaptoéthanol. Après une dénaturation de 5min à 90°C, 15 µL de chaque échantillon ont été déposés dans un gel de 10% SDS page. Un transfert semi sec a été réalisé sur membrane PVDF (Merck Millipore, Darmstadt, Germany) en utilisant un appareil Trans-Blot® SD Cell BioRad (18V à 600 mA pendant 25 min). 5% de lait bovin dans du TBS a été utilisé comme agent blocquant. Les incubations avec l'anti-Flag primaire (Thermofischer, Waltham, MA, USA), l'anti-c Myc de souris (Invitrogen, Carlsbad, CA, USA), et l'anticorps secondaire d'IgG anti-souris couplé à l'alkaline phosphatase (Thermoscientific, Rockford, IL, USA), dilué selon les recommandations du vendeur, pendant une heure. Les protéines ont été détectées par incubation dans la solution de substrat BCIP/NBT AP (Sigma-Aldrich, St. Louis, MO, USA).

II.2.3 9 Analyse des protéines par GeLC-MS

Les lysats cellulaires ont été dilués dans une solution Laemmli reducteur (Sambrook and Russell, 2006) pour avoir une solution à 3.3 µg/µL et le SDS-PAGE a été réalisé avec des gel gradient 4-15 % pré-castés (Mini-PROTEAN TGX gel; Bio-rad, Munich, Germany). Les gels sont colorés au Coomassie et les bandes protéiques ont été coupées pour une digestion sur gel. Après réduction (20 mM de dithiothreitol) et alkylation (55 mM chloroacetamide), les protéines ont été digérées avec 0.1 µg de Trypsine (Promega, Mannheim, Germany) dans 0.1 M d'une solution de bicarbonate de triéthylammonium durant 16h à 37°C. Les peptides ont été extraits, le solvant séché par lyophilisation, 20 µL de tampon de charge HPLC (acétonitrile 3%, acide trifluoroacétique 0.05% dans de l'eau) ont été réutilisés pour la resuspension. Les extraits peptidiques ont été analysés par LC-MS avec un système Dionex U3000 uHPLC couple à un spectromètre de masse Q Exactive Plus (Thermo Scientific, Dreieich, Germany). Un volume de 4 µL d'échantillon a été injecté dans une pré-colonne C18 100 µ-PepMap (dimensions de la colonne : 300 µm i.d. x 5 mm) avant une séparation sur une colonne Acclaim PepMap RSLC (dimensions de la colonne : 75 µm i.d. x 50 cm ; Thermo Scientific). Un gradient binaire avec un éluant A (0.05% d'acide formique dans l'eau) et un éluant B (8% acétonitrile, 0.04% d'acide formique) avec un débit à 0.3 µL/min a été utilisé. Les peptides ont été élués avec un gradient de 5% de B à 50% d'éluant B pendant 70 min, suivi d'une augmentation jusqu'à 90% de l'éluant B en 6 min. Après une élution isocratique à 90% d'éluant B de 10min, la colonne a été équilibrée 10 min avec 5% de l'éluant B. Le système LC a été directement couplé à un spectromètre de masse Q Exactive Plus. La totalité du scan MS a été acquis en mode ion positif à une résolution de 35,000 avec une région de scan de 240 m/z à 1800 m/z. Les données dépendants MS/MS des 10 ions précurseurs les plus intenses ont été acquises avec une résolution de 17,500 ; les paramètres de scan ont été réglés à une fenêtre d'isolation de 1.6 m/z, une énergie de collision de 27. Les précurseurs avec un état de charge <2 et >6 ont été exclus et les précurseurs ont été exclus pour isolation suivante pendant 10 s. Les fichiers bruts ont été comparés à une base de données incluant celle de *Saccharomyces cerevisiae* (souche CEN.PK113-7D; 5,439 UniProt entries, 10.2018) additionnée des protéines cibles et des contaminants habituels en utilisant l'algorithme de recherche Sequest et le logiciel Proteome discoverer (Thermo Scientific). Des recherches combinant une spécificité protéase full-tryptique ou semi-tryptique, permettant une oxydation variée des méthionines

et une acétylation en N terminale, ont été combinées. Les comparaisons de spectres peptidiques et d'identification de protéines ont été réalisées avec un taux de faux positifs inférieur à 1%.

II.2.3 10 Alignements de séquences

Les alignements ont été réalisés grâce au logiciel Muscle. Les noms et les codes Uniprot des différentes protéines alignées sont les suivantes : C.t., *Chloracidobacterium thermophilum* (G2LGW3); M.sp, *Mycobacterium* sp (A0A1A0Q6A2); N.sp., *Nanosalina* sp. (G0QG73); C.f., *Cystobacter fuscus* (S9QG19); P.a., *Pantoea ananas* (P21687); A.n., *Aspergillus niger* (A2QM49); A.o., *Aspergillus oryzae* (Q2U4X9); A.g., *Arthroderma gypseum* (E4UPP6); *Neurospora crassa* (P37295); *Phaffia rhodozyma* (Q7Z859); *Podospira anserina* (B2ATB0); *Pyrenophora tritici-repentis* (B2WAQ3); *Phaeosphaeria nodorum* (Q0V6M5); *Pantoea ananas* (P21683); *Franconibacter pulveris* (C7C5F2); *Erwinia gerundensis* (A0A0U5LOT8); *Cronobacter turicensis* (C9Y0F0); *Enterobacter agglomerans* (Q47846).

II.2.4 Méthodes concernant les caroténoïdes

II.2.4 1 Procédure d'extraction des caroténoïdes

Les culots cellulaires préalablement congelés ont été lyophilisés puis ont été resuspendus dans 500 µL d'acétone. De l'apocaroténal (24mg/L dans de l'hexane) a été utilisé comme standard interne (100 µL pour 0.025 g de cellules sèches). L'extraction a été réalisée comme suit : 3 cycles de 20s à 0.05 m/s utilisant une Fast Prep FP120 (ThermoFisher, Waltham, MA, USA) avec une minute dans la glace. La phase acétonique a été transférée dans un nouveau tube et le procédé d'extraction est ainsi répété trois fois. Les extraits acétoniques sont rassemblés et séchés sous un flux d'azote. L'extrait sec est resuspendu dans 500 µL ou de 2500 µL d'hexane dépendant de la concentration de l'extrait de caroténoïdes pour analyse par HPLC.

II.2.4 2 Analyse HPLC

La chromatographie du caroténoïde a été réalisée avec un système d'HPLC Waters 2690 couplé à un détecteur UV-visible (996 Waters PDA) opérant sur tout le spectre visible (400-700nm). La colonne utilisée était une Brownlee RP18 100 x 2.1 mm. Le flux a été de 0.3 ml/min à 35°C durant 49 min avec un gradient isocratique de 30 % d'isopropanol et 70 % d'acétonitrile/eau (90:10, v:v) à 55% d'isopropanol et 45 % d'acétonitrile/eau (90:10, v:v). 30 µL d'extrait caroténoïdique a été injecté dans le système HPLC Waters 2690. La séparation a été comme suit : un gradient isocratique composé d'isopropanol et d'acétonitrile/H₂O (90:10, v:v) a été appliqué 49 min avec un flux de 0.3 ml/min dans une colonne Brownlee RP18 Brownlee RP18 à 35°C. Les spectres d'absorbance ont été enregistrés entre 200 nm et 600 nm utilisant un détecteur UV-visible (996 Waters PDA). Les pics ont été caractérisés par leur temps d'élution et leur spectre (Takaichi, 2000). Les pics des caroténoïdes principaux ont été manuellement intégrés : à 282 nm (phytoène), à 475 nm (lycopène), à 454 nm (β-carotène), à 442 nm (neurosporène), à 399 nm (ζ-carotène), à 349 nm (phytofluène), à 430 nm (dihydro-β-carotène), et à 487nm (torulène). Les aires de pic ont été utilisées pour calculer leur quantité basée sur une courbe de calibration obtenue avec des solutions standards de phytoène, lycopène and β-carotène (Sigma-Aldrich, St. Louis, MO, USA) et de neurosporène, zeta-carotène, phytofluène, β-zeacarotène and torulène (Carotenature, Münsingen, Switzerland) et la correction sur le taux d'extraction a été performée sur la quantification d'apocaroténal (Sigma-Aldrich, St. Louis, MO, USA). L'attribution des pics de dihydro-β-carotène a été faite en se référant à l'ordre d'élution décrit dans (Mizoguchi et al., 2015; Takaichi, 2000). Comme il

n'existe pas de standard de dihydro- β -carotène, la quantification a été réalisée avec une courbe standard de β -zeacarotène comme les deux produits contiennent le même nombre de d'insaturations. Les autres caroténoïdes n'ont pas été quantifiés dû à l'absence de standard commerciale. Chaque échantillon extrait a été injecté et quantifié deux fois (duplicat technique). Un second set de batch de culture a été préparé et analysé avec le même protocole pour duplication (duplicat biologique). Les barres d'erreur représentent la déviation standard obtenue de ces différentes valeurs.

II.2.4 3 Courbe de calibration

Les courbes de calibrations ont été générées par injection de solution de références standards. Une gamme de dilutions a été réalisée pour chaque caroténoïde. Les pic d'élution ont été intégrés manuellement à la longueur d'onde d'absorption maximale respective de chaque caroténoïde afin d'obtenir l'aire du pic d'élution. La gamme de concentration respectant une correspondance linéaire entre la dilution et l'aire de pic sont retenues pour réaliser la gamme étalon de chaque standard. Les longueurs d'onde d'absorption maximale de chaque caroténoïde sont respectivement : 282 nm (phytoène), 475 nm (lycopène), 454 nm (β -carotène) (Sigma-Aldrich, St. Louis, MO, USA), 442 nm (neurosporène), 399 nm (ζ -carotène), 349 nm (phytofluène), 430 nm (dihydro- β -carotène), 487nm (torulène) (Carotenature, Münsingen, Switzerland. La courbe de calibration du standard interne apocaroténal (Sigma-Aldrich, St. Louis, MO, USA) a été réalisée de la même manière et intégrer à 475nm. Les courbes de calibrations ont été obtenues dans Excel et les formules obtenues ont été utilisées pour convertir l'aire mesuré de chaque caroténoïde en concentration exprimé en $\mu\text{g/g}$ de masse sèche de cellules.

II.2.5 Calcul de la quantité de caroténoïdes produit dans un échantillon

II.2.5 1 Calcul du facteur de correction sur l'extraction

Le facteur de correction sur l'extraction est obtenu par $F_c = Q_{apm} / Q_{apa}$ avec F_c le facteur de correction, Q_{apm} la quantité de standard apocaroténal mesurée et Q_{apa} la quantité initiale d'apocaroténal que l'on a ajouté dans les échantillons avant le début de l'extraction.

$Q_{apm} = (A_{ap} / \text{Coef}_{ap})$ avec Q_{apm} la quantité de standard apocaroténal mesurée, A_{ap} l'aire mesurée d'apocaroténal injecté et Coef_{ap} le coefficient directeur de la droite obtenue par régression linéaire avec les points de la gamme étalon d'apocaroténal.

$Q_{apa} = C_{api} * V_{ap} * (V_i / V_r)$ avec Q_{apa} la quantité d'apocaroténal attendue, V_{ap} le volume d'apocaroténal ajouté dans les échantillons avant le début de l'extraction, V_i le volume d'injection et V_r le volume de resuspension de l'échantillon.

II.2.5 2 Calcul de la quantité d'un caroténoïde x en pmol, en μg puis en $\mu\text{g/g}$ DCW

La quantité d'un caroténoïde x en pmol est obtenu par :

$Q_{xp} = ((A_x / \text{Coef}_x) * F_c) * (V_r / V_i)$ avec Q_{xp} la quantité d'un caroténoïde x en pmol, A_x l'aire de pic mesurée du caroténoïde x, Coef_x le coefficient directeur de la droite obtenue par régression linéaire avec les points de la gamme étalon du caroténoïde x, F_c le facteur de correction sur l'extraction, V_r le volume de resuspension de l'extrait sec, V_i le volume d'injection.

La quantité du caroténoïde en μg est obtenue par :

$Q_{xm} = Q_{xp} * M_x * 10^{-6}$ avec Q_{xm} la quantité du caroténoïde x en μg , Q_{xp} la quantité du caroténoïde x en pmol, M_x la masse molaire du caroténoïde x.

La quantité du caroténoïde en $\mu\text{g/g DCW}$ est obtenue par :

$Q_x = Q_{xm} / M_c$ avec Q_x la quantité du caroténoïde en $\mu\text{g/g DCW}$, Q_{xm} la quantité du caroténoïde x en μg et M_c la masse cellulaire sèche utilisée pour l'extraction.

III RESULTATS

III-1 Analyses bioinformatiques des enzymes CrtI et CrtB

Les résidus aux extrémités N ou C terminale des enzymes peuvent être importants pour la fonction enzymatique. Ils peuvent être perturbés par une ingénierie enzymatique impliquant une fusion. Pour comprendre la structure tridimensionnelle des enzymes ainsi connaître où sont situés les résidus importants pour la fonction des enzymes d'intérêt, des alignements de séquence et des modèles par homologie ont été réalisés.

III.1.1 Modélisation par homologie de CrtI

La gamma désaturase de *Nonlabens dokdonensis* ou CrtD possède le meilleur pourcentage d'identité avec CrtI (31%). CrtD appartient à une voie de biosynthèse dans laquelle elle catalyse l'ajout d'une insaturation sur des substrats ressemblant au neurosporène, au lycopène ou à l' γ -carotène mais ayant un groupement hydroxyle (Ahn and Kim, 2015) (figure 29).

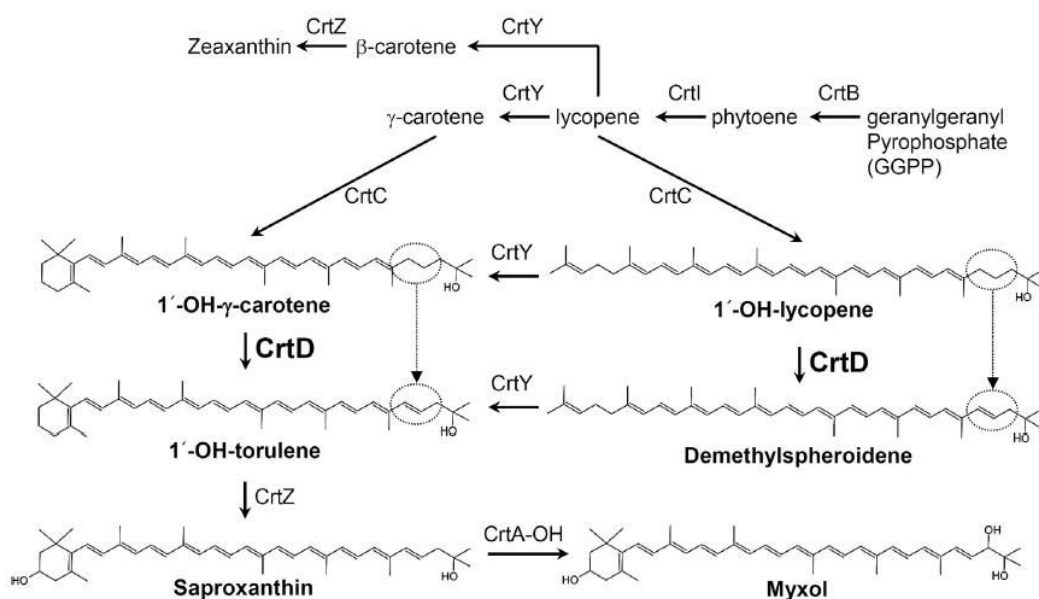


Figure 29 : Voie de biosynthèse d' γ -caroténoïdes, tels que le myxol et la saproxanthine dans laquelle agit CrtD par désaturation en position C3'-C4' du 1'-OH- γ -caroténoïde. La position de l'insaturation catalysée par CrtD sur plusieurs substrats est encadrée. Figure tirée de (Ahn and Kim, 2015).

Des séquences conservées entre la CrtI de *X. dendrorhous*, la CrtD de *Nonlabens dokdonensis*, la PDS de *O. sativa* et la CrtI de *P. ananatis* sont mises en évidence par une comparaison de leur séquence (figure 30).

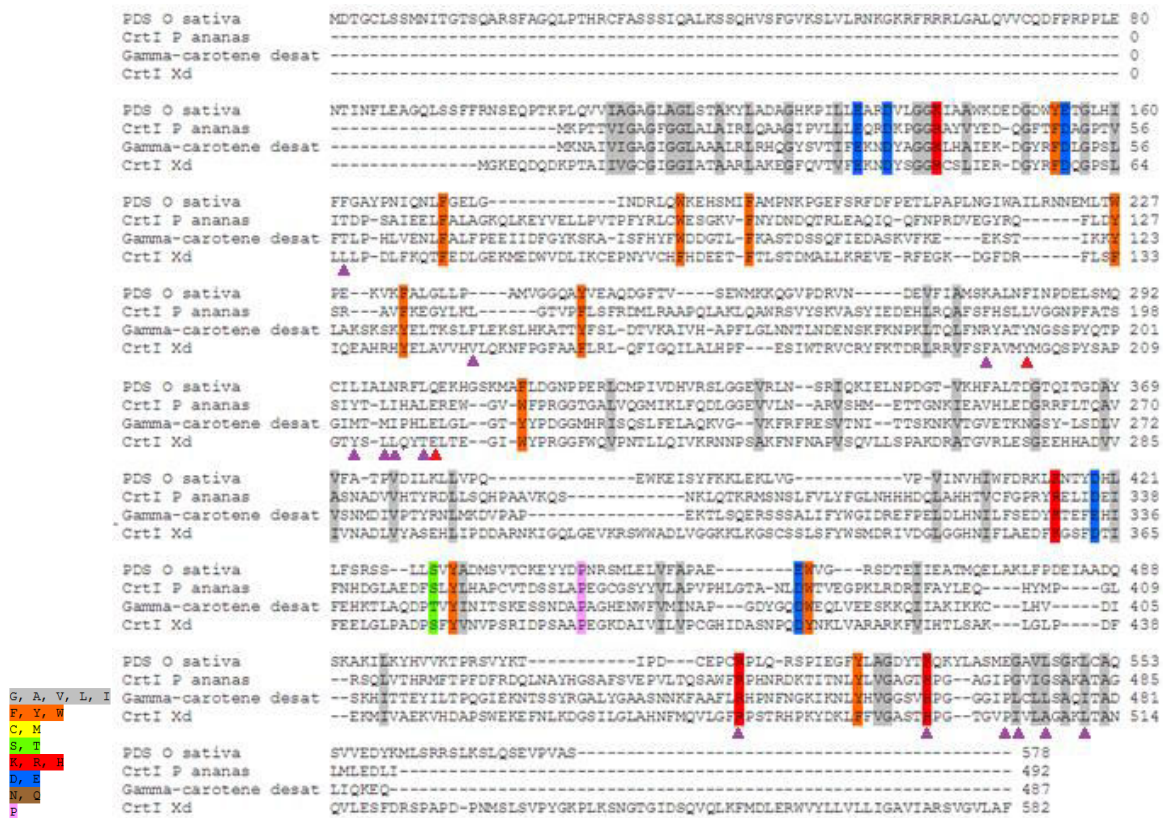


Figure 30 : Alignement de séquence de CrtI de *X. dendrorhous*, de la gamma désaturase de *Nonlabens dokdonensis*, de la PDS de *O. sativa* et de CrtI de *P. ananatis*. Les résidus de même propriété et conservés sont surlignés avec la même couleur. Des résidus conservés chez les CrtD et qui sont soit au niveau de la poche hydrophobique ou sont impliqués dans la catalyse sont représentés par des rectangles violets et rouges respectivement (Ahn and Kim, 2015).

Plusieurs résidus sont conservés aux extrémités N et C terminales des séquences. Des résidus en N terminale sont impliqués dans la liaison au cofacteur (Bartley et al., 1990). Ils forment un motif $\beta\alpha\beta$ (Bartley et al., 1990). Les résidus conservés en C terminale sont retrouvés chez les trois autres types de désaturase. D'autres résidus conservés sont spécifiques aux CrtD et sont répartis ailleurs dans la séquence (figure 30). Ces derniers sont impliqués dans la formation de la poche hydrophobique accueillant le caroténoïde (Ahn and Kim, 2015). Ils sont également impliqués dans le mécanisme catalytique (figure 30). Les différents résidus conservés entre les différentes enzymes sont mis en évidence sur la structure tridimensionnelle de CrtD (figure 31).

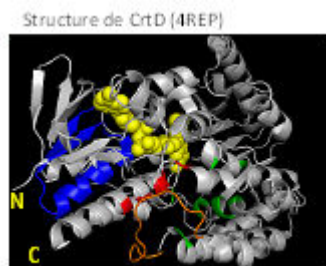


Figure 31 : Structure tridimensionnelle de CrtD(4REP). L'image est vue d'en bas coté membrane. Les extrémités N et C terminales de la structure sont indiquées. Le FAD est représenté en sphère jaune. Le motif $\beta\alpha\beta$ interagissant avec le FAD est représenté en bleu. Les résidus conservés en C terminal conservés chez les CrtD et retrouver chez les 3 autres désaturases sont représentés en rouge. Les autres résidus conservés chez les CrtD et non retrouvé chez les trois autres désaturases sont représentés en vert. Une boucle mobile au niveau de l'entrée du substrat chez les CrtD est représentée en orange.

Une fusion en N terminale pourrait perturber par exemple les résidus impliqués dans l'interaction avec le FAD (en bleu sur la figure 31). La fonction des résidus conservés en C-terminale (en rouge sur la figure 31) pourraient également être perturbée par une fusion en C terminale. Comme indiqué précédemment, les substrats des CrtD sont hydroxylés. Des différences sont retrouvées au niveau du canal hydrophobique de CrtD comparé à celui de la PDS. Spécifiquement à CrtD, une boucle hautement flexible (représentée en orange sur la figure 31) à l'entrée du canal permet d'allonger la longueur de la poche mais aussi d'accueillir le cycle β comme dans le 1'-OH- γ -carotène. Cela oriente correctement le reste de la molécule substrat par rapport au FAD et positionne les carbones en position C3'-C4' pour la désaturation. Malgré ces différences majeures, l'organisation tridimensionnelle globale de CrtD ressemble aux autres structures de désaturases, à savoir une organisation en trois sous domaine (Ahn and Kim, 2015).

Les structures de la CrtI bactérienne *P. ananatis* (Schaub et al., 2012a) et celle de la phytoène désaturase de la plante *O. sativa* se ressemblent (Gemmecker et al., 2015). Lorsqu'on les superpose, les résidus du site actif constituent l'essentiel des différences entre les deux structures (Gemmecker et al., 2015). Cela reflète les mécanismes réactionnels différents (Gemmecker et al., 2015). Un modèle par homologie de CrtI avec la PDS comme modèle est représenté (figure 32B).

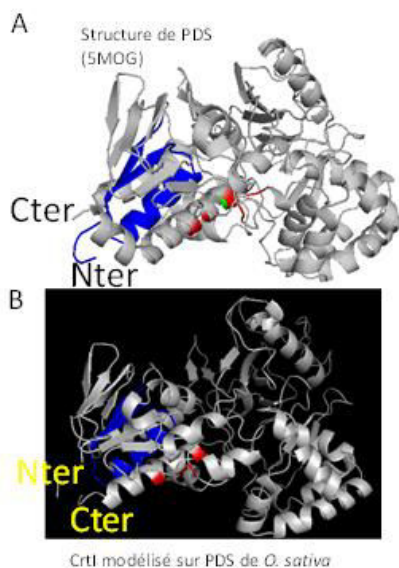


Figure 32 : Localisations des extrémités N et C-terminale des désaturases. A. structure de la PDS avec en bleu, rouge et vert des résidus importants pour la fonction enzymatique B. modèle de CrtI modélisé à partir de PDS de *O. sativa*. Les extrémités N et C terminale des modèles sont indiquées.

Une séquence en C terminale de CrtI de *X. dendrorhous* est non modélisée. En effet par alignement de séquence, on observe que l'enzyme possède une séquence d'une vingtaine de résidus en C terminale prédite pour former une hélice et absente chez les autres enzymes. Une prédiction de la structure secondaire de cette séquence C terminale en hélice par Phyre indique une hélice transmembranaire. Une autre hélice enfouie au sein de la protéine est aussi prédite comme étant transmembranaire par Phyre (figure 33).

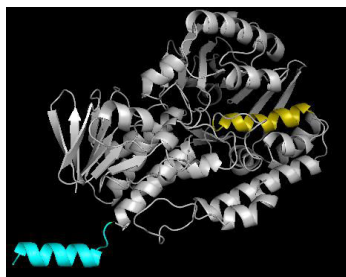


Figure 33 : Modèle de CrtI à partir de CrtD et sur lequel l'hélice en C terminale non modélisable est représentée. L'hélice représentée en bleu ciel n'existe pas dans le modèle par homologie car CrtI possède 62 résidus en plus par rapport à CrtD en C terminale de l'enzyme. En jaune doré, une hélice définie comme transmembranaire par Phyre mais qui se trouve être enfouie au sein de la protéine.

III.1.2 Modélisation par homologie de CrtB

Nous avons premièrement délimité une séparation entre les domaines présents dans CrtYB. Un alignement de séquences de plusieurs séquences de CrtYB fongiques avec des CrtB et CrtY individuelles issues d'autres organismes a été réalisé. Une région d'acides aminés non conservés est mise en évidence et nous permet de définir une séquence linker naturelle dans CrtYB (figure 34).

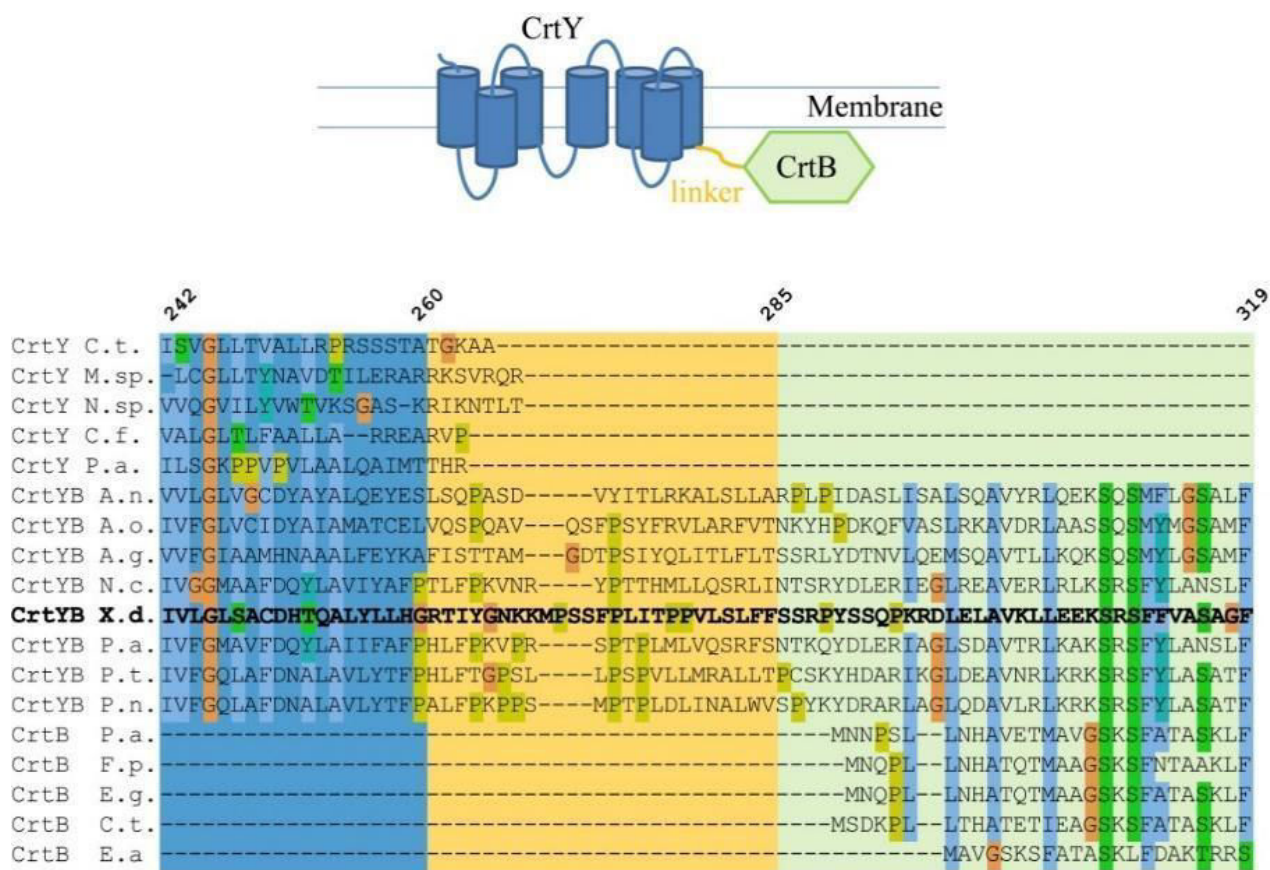
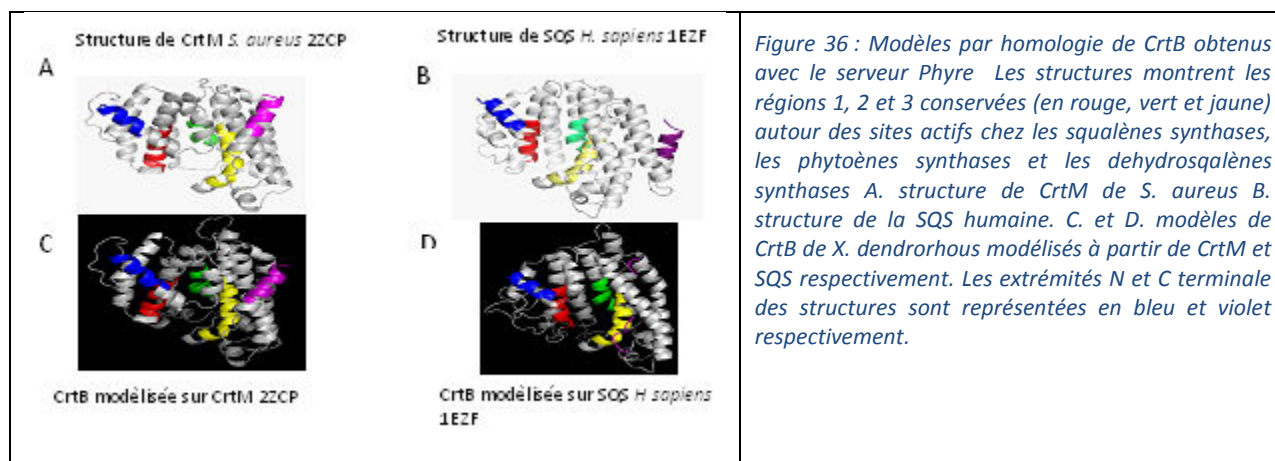


Figure 34 : Alignement de séquences pour la délimitation de la séquence du linker dans CrtYB. Panel du haut : représentation de l'organisation structurale de CrtYB dans la membrane. Les domaines lycopène cyclase, phytoène synthase et le linker sont représentés en bleu, vert et jaune, respectivement. Les lycopène cyclases d'Eucaryotes sont décrites pour avoir 7 domaines transmembranaires (<http://www.uniprot.org/uniprot/Q>) alors que les phytoènes synthases n'ont pas de segments transmembranaires (Schaub et al., 2012b). Panel du bas : alignement partiel entre CrtYB de Xanthophyllomyces dendrorhous (X.d.) (en gras) avec d'autres phytoène synthases/lycopène synthases bifonctionnelles et des phytoène synthases ou lycopène cyclases monofonctionnelles.

Des modèles par homologie de CrtB ont été obtenus. La déhydro-squalène (4-4'-diapophytoène) synthase (CrtM) et la squalène synthase (SQase) sont les premiers modèles proposés pour modéliser CrtB. Un alignement des enzymes montre 3 régions conservées entre CrtB, CrtM et SQS (Liu et al., 2014) (figure 35). Les substrats, produits et séquences peptidiques de CrtM, SQS et CrtB sont similaires (Moise et al., 2014). Les structures de CrtM et SQS sont formées par l'empaquetage de plusieurs hélices α au centre desquelles se positionnent les 3 motifs de séquences conservées. Ces séquences conservées entourent deux poches constituant les sites actifs des enzymes (Blagg et al., 2002; Liu et al., 2014) (figure 36 A et B). Les extrémités N et C terminales sur les structures de CrtM et SQS se retrouvent vers l'extérieur des enzymes et ne sont pas en contact direct avec les résidus du site actif. La structure de la SQS humaine a été obtenue après élimination de résidus transmembranaires à ses extrémités (Pandit et al., 2000). La structure de CrtM a été déterminée sur la forme entière (Lin et al., 2010).

CrtYB X.d	MTALAYYQIHLYITLPIGLGLLLTSPILTKFDIYKISILVFIASFATTPWDSWIIRNGAWTYPSAESGQGVFGTFLDVP	80
SQS homo	-----	0
SQS S.c	-----	0
CrtB P.a	-----	0
CrtM S.a	-----	0
CrtYB X.d	YEEYAFFVIQTVITGLVYVLATRHLLPSLALPKTRSSALSALKALIPLPITYLFTAHPSPSPDPLVTDHYFYMRALSLL	160
SQS homo	-----	0
SQS S.c	-----	0
CrtB P.a	-----	0
CrtM S.a	-----	0
CrtYB X.d	ITPPTMLLAALSGEYAFDWKSGRAKSTIAAIMIPTVYLIWVDYVAVGQDSWSINDEKIVGWRLGGVLPIDEEAMFFLLTNL	240
SQS homo	-----	0
SQS S.c	-----	0
CrtB P.a	-----	0
CrtM S.a	-----	0
CrtYB X.d	MIVLGLSAC-DHTQALYLLHGRTIYGNKKMPSSFFLITPPVLSLFFSSRPYSSQPKRDLELAVKLEKKSSFFVASAGF	319
SQS homo	--MEFVKCLGHPPEFYNLVRF-RIGGKR-----KVMPPM-----DQDSLSSSLKTCYKYLNTSSSFAAVIQAL	61
SQS S.c	--MGKLLQALHPVEMKAALKL-KF--CR-----TPLFSI-----YDQSTSPYLLHCFELLNLTSSSFAAVIREL	60
CrtB P.a	-----MNN-PSLLNHAVETMAVGSKSFATASKLF	28
CrtM S.a	-----MTMDMNFYKCHKIMKKHSSSFYAFDLL	29
CrtYB X.d	PSEVNERLVGLAFQVTDLLISPEVSSNPHTIDMVSDFLTLLFGPPLHPSQPDKILSSPLLPSPHPSRPTGMYPLPP	399
SQS homo	DGEMNAVCIIFYLVLSALDTLEDDMTI-SVE-KKVPLLLHNFHSFL-----YQPDWR-----FMES--KE-----KD	118
SQS S.c	HPELANCVTFLFLLIALDTIEDMSI-EHD-LKIDLLRHFEKL-----LLTKWS-----FDGNAPDV-----KD	119
CrtB P.a	DAKTRSRVLMLYAWCHCDVIDDQTL-GFQ-ARQPALQTPQRIQMQLMKTRQA-----YAGSQ-----MH	88
CrtM S.a	PEDQKAYWAIYAVCKIDSDIVYGDIOFL-NQ--I--KEDIQSI--EKYP-----YEHHHFQ-----SD	83
CrtYB X.d	PPSLSPAELVQFLTERVPVQYHFAFRLLAKLQGLIPRYPLDELLRGYTTDLIFPLSTEAVQARKTPDIETALLDGLCV	479
SQS homo	RQVLEDFPTISLEFRNLAEKYQTVIADICRRMGI-----GMAEFLDK-----H-----VTSEQWDDKYCHYV	175
SQS S.c	RAVLTDFFESILIEFHKLKPEYQEVKEITEKMGN-----GMADYILD-----ENYNLNLGLQTVHYDYVCHYV	182
CrtB P.a	EPAPFAAQEVA-MAHDTAPAYAFDH-----LEGFAMD-----VREAQYSQLDITLRVCHYV	138
CrtM S.a	RRIMMALQHVA-QHKNIAFQSFYNL-----IDTVYKD-----QHFTMFETDALFGYCYGV	133
CrtYB X.d	AGSVAEELLVYVSWASAPSQVPATIEEREAVLVASREMGTAQLVNIARDIKGATEGRFLLPLSFFGLRDESKLAIPTDN	559
SQS homo	AGLVGIGLSRLFSASEFED-----PLVGEDTERANSMLFLQKTNIIIDYLEDQQGGREFWPQEVWSRYVKLGD----	246
SQS S.c	AGLVGDGLTRLIVIAKFAN-----ESLYSNEQLYESMGLFLQKTNIIIDYNEELVDGSSFWPKEIWSQYAPQLKD----	253
CrtB P.a	AGVVGLMMAQIMGVRD-----NATLDRACDLGLAFOLTNIAARDIVDDAHAGRCYLPASWLEHEGLNKEN----	203
CrtM S.a	AGTVGEVLTPILSDHET-----HQTYDVARRLGESLQILNILEVGEFDNERIIFSKQRLKQYEVDIAE--VY	200
CrtYB X.d	TEPRQDFDKLLSLSPSSTLPSSNASSEFRFEWKTSYSLPVAAYEDLAKHSYKGIDRLPTEVOAGMRAACASYLLIGREI	639
SQS homo	AKPENIDLAV--QC-----LNELITNALHHIPDVITYLSRLRNQSVNFCAIPQVMAIATLAAC--N--N--N	306
SQS S.c	MKPENEQLGL--DC-----INHLVLNALSHVIDVLTYLAGEHQSTFQFCAIPQVMAIATLALV--N--N--N	313
CrtB P.a	AAPENRQALS--RI-----ARRLVQEAEPYLSATAGLAGLPLRS-----AWAATAKQV--YR--K--	254
CrtM S.a	QNGVNNHYID--L-----WEYYAAIAEKDFQDQVMDQKVFISIEA-----QPIIELAARI--YI--E--	250
CrtYB X.d	KVVWKGDVGERR-TVAGWR-----RVRKVLVVMVGWEGQ-----	673
SQS homo	QQVFKGAVKIRK-GQAVTLMMDATNMPAVKAI-YQYMEEIYHRIIPDSDPSSSKTRQIIISTIR-----TONLPNCOLI	377
SQS S.c	REVLHGNVKIRK-GTTCYLILKSRTRLRGCVEIF-DYYLRDIKSLAVQDPNFKLNIQISKIEQFMEEMYQDKLPNNV--	389
CrtB P.a	IG-----VKVEQAGQQAWDQROSTTPEKLTLLLAASGQALTSRMRAHPPR-----PAHLW	305
CrtM S.a	IL-----DEVROANYTLH-----ERVFVDRKK-----KAKLF	277
CrtYB X.d	-----	673
SQS homo	SRSHYSPIYLS-----F-----VMLLAALSQYLTTLTQVTEDEVVQTGEH	417
SQS S.c	-KPNETPIPLKVKERSRYDDELFTQEEYKFNMVLSIILSVLLGFYYIYTLHRA-----	444
CrtB P.a	QRPL-----	309
CrtM S.a	HEINSKYHRI-----	287

Figure 35 : Alignement de séquence de CrtB de *X. dendrorhous* avec d'autres séquences. Alignement de séquence avec SQS humaine, de la SQS de *S. cerevisiae*, du CrtB de *P. ananatis*, et de CrtM de *S. aureus*. Les résidus conservés ayant les mêmes propriétés sont représentés dans la même couleur. Les trois motifs de régions conservés chez tous ces enzymes sont indiqués au dessus de l'alignement en rouge, vert et jaune (Blagg et al., 2002). Les résidus encadrés en violet ne sont pas importants pour la fonction de la SQS humaine (Thompson et al., 1998). Les résidus encadrés en rouge sont une séquence d'insertion au réticulum endoplasmique pour la SQS de *S. cerevisiae* (Peng et al., 2017). L'étoile indique la limite N terminale de CrtB choisie dans cette étude.



Les extrémités de CrtB de *X. dendrorhous* ne semblent donc pas être impliquées dans la fonction catalytique de l'enzyme. Les modèles de structure par homologie obtenus (figure 36 C et D) montrent les extrémités N et C terminales qui se retrouvent en surface de la protéine.

III.1.3 Modélisation de CrtY

Le domaine CrtY est une protéine avec plusieurs segments transmembranaires. De plus aucune structure d'homologue fonctionnel probable de CrtY n'a été résolue. Nous n'avons donc pas obtenu une structure par homologie de CrtY. Le linker naturel entre CrtY et CrtB est retrouvé après la dernière hélice transmembranaire de CrtY.

III-2 Stratégie d'expression

Notre objectif dans l'expression des différentes enzymes au sein de la voie de biosynthèse a été d'avoir un niveau d'expression similaire entre les enzymes. Pour cela, nous avons choisi une intégration en monocopie dans le génome de *S. cerevisiae* et une expression par un même promoteur Gal1 ou Gal10. Une même souche de départ a été créée pour accueillir les différentes constructions. Cette souche de départ a consisté en une optimisation de la voie en amont du précurseur des caroténoïdes, le GGPP. Pour cela, deux étapes limitantes de cette voie en amont ont été améliorées en intégrant une copie des enzymes limitantes dans le génome. Ces enzymes limitantes étant CrtE et HMG1. La forme tronquée appelée tHMG1 a été utilisée car elle est non régulable par rétroinhibition sous cette forme (figure 37).

Stratégie d'expression

- Intégration dans le génome de *S. cerevisiae*
- Expression sous le contrôle du double promoteur GAL1-GAL10
- Deux optimisations de la voie précurseur: **HMG1** et CRTE

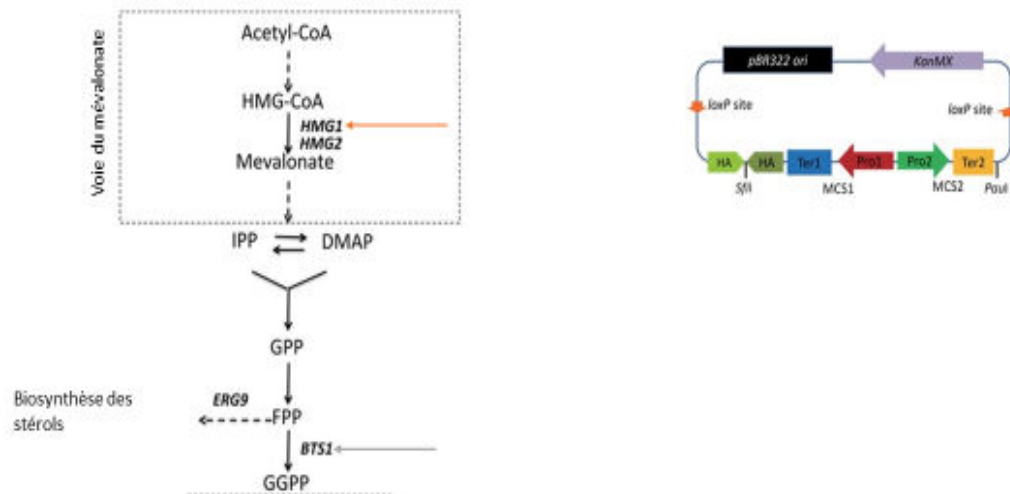


Figure 37 : Optimisation du flux de la voie en amont du GGPP et stratégie d'expression des différents gènes dans cette étude. Deux étapes limitantes sont indiquées par les flèches oranges et grises. L'organisation du plasmide d'intégration utilisé dans cette étude est représentée à droite (HA : Homologous Arm, Sfi1 : site de restriction unique Sfi1 utilisé pour linéariser le plasmide, Pro1 et Pro2 : promoteurs bidirectionnel GAL1-GAL10, Ter1 et Ter2 : terminateurs. Le promoteur bidirectionnel et les terminateurs respectifs permettent d'insérer deux gènes à un locus.

III-3 Colocalisation avec des domaines d'interaction

III.3.1 Stratégie de colocalisation avec les modules cohésine dockérine

Une stratégie de rapprochement spatiale d'enzyme peut se faire par l'utilisation de domaines d'interaction que l'on fusionne au niveau des enzymes. Dans la littérature, des travaux de colocalisation d'enzyme chez la levure par l'utilisation des domaines cohésine dockérine qui ont apporté une amélioration dans la production ont été publiés durant cette thèse (Kim and Hahn, 2014; Kim et al., 2016). Nous avons conçu un système d'enzymes dans lequel CrtYB est fusionnée avec une dockérine en C-terminale et CrtI avec une cohésine en N terminale. Les dockérines et les cohésines possèdent un linker naturel les reliant aux différents enzymes impliqués dans la dégradation du cellulosome. Dans le cellulosome, la cohésine que l'on a utilisée est trouvée en seconde position de la chaîne polypeptidique de CIPA de *Clostridium thermocellum*. Elle est flanquée en C terminale de son linker naturel. Son partenaire d'interaction est la dockérine CelS de type I de *Clostridium thermocellum* qui est flanquée en N terminale d'un linker naturel flexible. Les séquences d'ADN codant ces domaines d'interaction et leur linker ont été optimisées pour une expression chez *S. cerevisiae*. Un effet recherché par cette stratégie est une colocalisation imposée de CrtYB et de CrtI (figure 38A).

III.3.2 Construction et analyse des souches

Les constructions ont été intégrées en une copie dans le génome par l'intermédiaire d'un plasmide intégratif au niveau de deux locus différents. Les tags peptidiques n'ont pas été conservés dans les constructions pour limiter une probable interférence dans les interactions des domaines. Après 72 heures de culture, les cellules sont rassemblées et séchées pour extraction des caroténoïdes. Les souches contenant le système enzymatique YB-dockérine/cohésine-I produisent plus du double de caroténoïdes totaux en comparaison avec le taux de production du système naturel. Les productions mesurées sont de 13mg/g DCW et de 6mg/g DCW respectivement (figure 38B). L'augmentation est due essentiellement à une production élevée du dihydro- β -carotène (figure 38B et C). Le dihydro- β -carotène possède deux cycles beta comme le β -carotène mais ne possède que trois insaturations. L'ajout du cycle β présent dans le dihydro- β -carotène se fait donc avant la quatrième désaturation par CrtI.

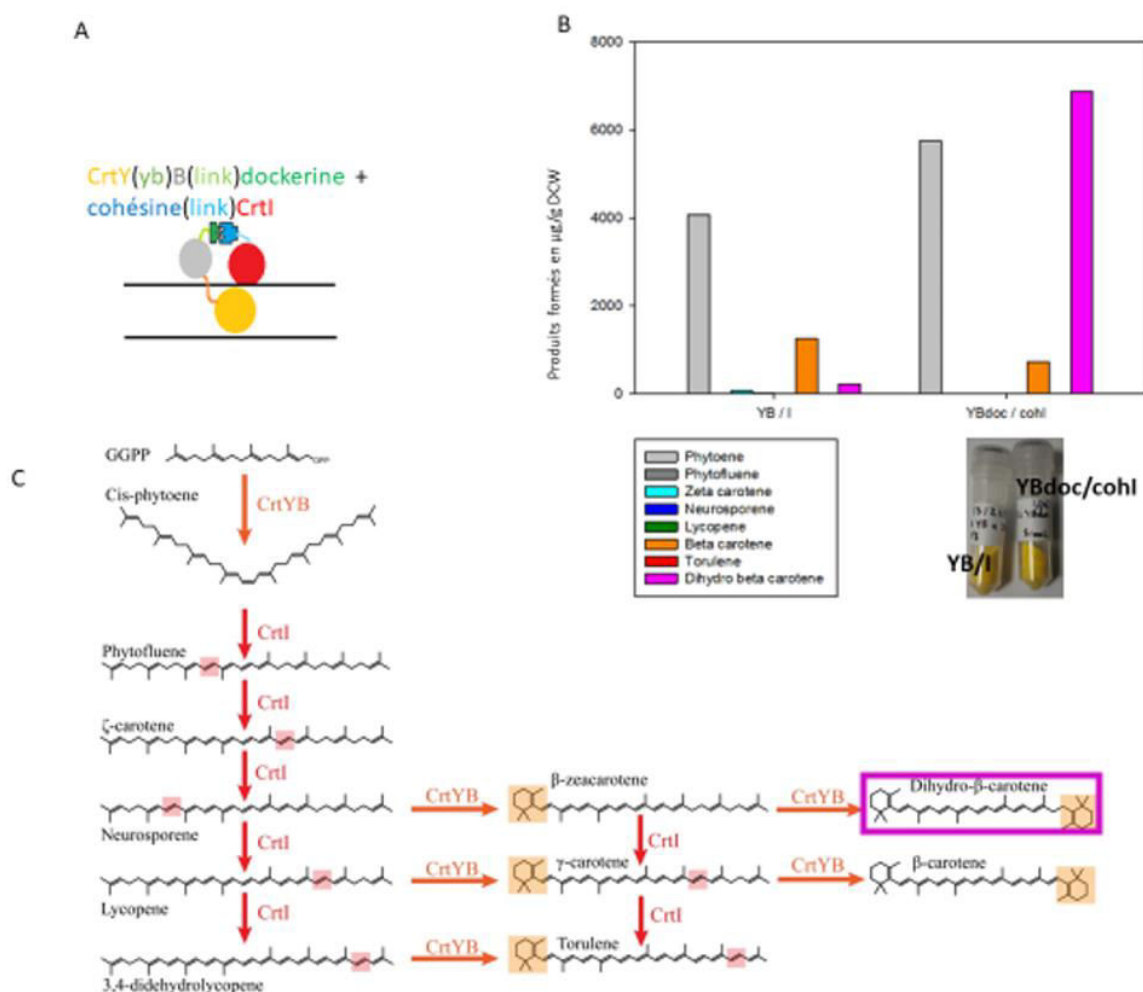


Figure 38 : Quantification des caroténoïdes produits par le système d'enzymes fusionnées aux domaines cohésine et dockérine. A. Sur le schéma représentatif des enzymes, les différentes régions de la construction sont représentées avec des couleurs respectifs. B. La quantification des caroténoïdes principaux produits par les souches est représentée. Une photo du culot cellulaire issu de 50mL de culture et après lyophilisation de la souche portant le système d'enzymes naturelles et la souche portant le système d'enzymes fusionnées avec la dockérine ou la cohésine est montrée. C. La voie de biosynthèse des caroténoïdes avec le produit majoritaire dihydro- β -carotène encadré en violet.

III.3.3 Perspectives

Un changement dans la nature des caroténoïdes produits en fusionnant des domaines d'interaction aux enzymes caroténogéniques n'a pas été attendu. Dans la littérature, des changements dans la nature des produits caroténogéniques ont été obtenus par exemple en utilisant des enzymes d'origines différentes. Par exemple, on a utilisé des désaturases pouvant catalyser 5 désaturations pour obtenir de nouvelles structures de caroténoïdes (Schmidt-Dannert et al., 2000). Dans notre cas, le changement dans la nature du caroténoïde final produit peut avoir différentes origines. L'augmentation en concentration des molécules en fin de voie, dans ce cas celui du dihydro- β -carotène, pourrait être la conséquence d'un flux plus élevé. Cette augmentation du flux de la voie s'observe aussi par le niveau total de caroténoïdes total produit qui augmente et qui se mesure aussi par une concentration plus élevée en phytoène par rapport au système d'enzymes naturelles. Une compétition entre l'étape de désaturation et de cyclisation pourrait dépendre de la concentration des intermédiaires qui sont dans notre cas substrats de deux enzymes. Un flux élevé en caroténoïdes pourrait donc favoriser la cyclisation au dépend de la désaturation. D'un autre côté, la modification des enzymes par la fusion avec les domaines d'interaction utilisées pourrait avoir des effets sur les activités catalytiques et donc modifier l'équilibre entre les niveaux d'activité phytoène désaturase et lycopène cyclase. D'autres expériences sont nécessaires pour mieux comprendre ce nouveau système enzymatique. Cela pourrait commencer par la confirmation de l'interaction in vivo des domaines.

III-4 Colocalisation par utilisation d'une protéine oligomérique

Un contrôle de l'architecture tridimensionnelle des scaffolds protéiques utilisés dans le cadre de la colocalisation d'enzyme serait idéal. Les protéines polymérisantes représentent des blocs de constructions qui peuvent s'auto-assembler de manière efficace in vivo. Les protéines formant la coque de microcompartiments bactériens pourraient constituer des blocs de construction prometteurs à réutiliser pour la conception de nouveaux scaffold protéiques in vivo (figure 39). Dans cette optique, nous avons fusionné CrtI avec CcmK2 en N terminale (figure 39).

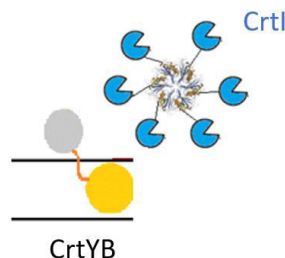
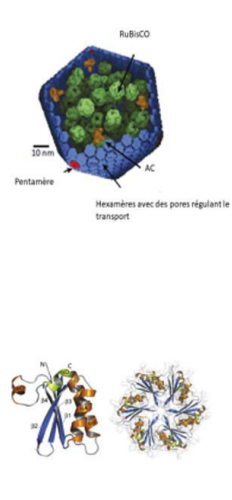


Figure 39 : Représentation schématique d'un auto assemblage de CrtI lorsque l'enzyme est fusionnée à un domaine CcmK2. Dans le panel de gauche est représenté le carboxysome (en haut) dont la cage protéique est composée, entre autres polymères, d'hexamère de la protéine CcmK2 (en bas). Dans le panel de droite est représenté une colocalisation de CrtI lorsque fusionnées à CcmK2.

III.4.1 Construction et analyse des souches

Les constructions ont été intégrées en une copie dans le génome par l'intermédiaire d'un plasmide intégratif au niveau de deux locus différents. Les tags peptidiques n'ont pas été conservés dans les constructions pour limiter une probable interférence dans les interactions des domaines. Nous avons confirmé la fonctionnalité de la voie de biosynthèse en observant la coloration des souches sur boîte de Petri. Des mesures précises des quantités de caroténoïdes produites et la confirmation du type d'interaction protéique réellement obtenu par ce système synthétique reste à réaliser.

III.4.2 Perspective

Un challenge dans l'utilisation de systèmes de protéines polymérisantes est celui de ne pas perturber les complexes quaternaires naturels des enzymes. On pourrait designer les scaffolds protéiques de telles sortes qu'ils ne perturbent pas les complexes quaternaires naturels des enzymes. CrtI par exemple pourrait fonctionner sous forme de tétramère alors que la stratégie utilisée pourrait imposer une colocalisation spatiale de six CrtI. La fonctionnalité confirmée du système enzymatique est donc intéressante. Les interactions entre les CcmK2 ne sont peut-être pas possibles dans ces conditions alors que CrtI arrive à maintenir ses modes d'interactions naturelles. Aussi, une interaction en tétramère de CrtI n'est peut-être pas finalement nécessaire pour son fonctionnement.

III-5 Colocalisation par fusion directe

III.5.1 Stratégie de colocalisation des enzymes par fusion directe

Différents paramètres ont été variés pour la construction des fusions enzymatiques. L'objectif est de sélectionner un système optimisé. Le choix de linker ou l'orientation des domaines fusionnés sont les paramètres que l'on peut varier dans cette stratégie de colocalisation. Différents linker peuvent être choisis par la réutilisation de linker naturel ou par la synthèse de linker synthétique. Les linker que nous avons utilisés ont été choisis pour être plus ou moins représentatifs d'une variation en taille et en flexibilité de linker (figure 40).

	Nombre de résidus	Caractérisations biophysiques du linker	Caractéristiques des linker
Linker naturelle reliant CrtY et CrtB dans CrtYB H G R T I Y G N K K M P S S F L I T P P V L S L F F S S R P Y S S Q P K R	38 aa	Aucun	
Linker naturel dans la NADPH Cytochrome P450 Reductase, une bidoamine enzyme C E H F G V E A T G E E S S I R Q Y	18 aa	RMN SAXS	Linker du CPR humaine Très flexible ne présentant aucune restriction de mouvement des domaines
G G G G S G G G S G G G S G G G S	20 aa	FRET SAXS	
E A A A K E A A A K E A A A K E A A A K	20 aa (30 Å)	FRET SAXS	

Figure 40 : Caractéristiques des linker utilisés dans cette étude. De haut en bas : linker yb, linker cpr, linker gs, linker ek. Le linker yb est non caractérisé, le linker cpr est très flexible, le linker synthétique gs est flexible, le linker synthétique ek forme une hélice et est donc rigide.

III.5.2 Construction des souches

Les constructions (figure 41) ont été intégrées en une copie dans le génome par l'intermédiaire d'un plasmide intégratif au niveau du même locus. Un tag en C terminale de chaque construction permet la détection par western blot.

Nom des souches	Domaine 1	Linker	Domaine 2	Linker	Domaine 3	Produits finaux attendus
B	CrtB					phytoène
B/I	CrtB		CrtI			lycopène
B/I/Y	CrtB		CrtI		CrtY	β-carotène
BI	CrtB	-	CrtI			lycopène
B(yb)I	CrtB	(yb)	CrtI			
B(cpr)I	CrtB	(cpr)	CrtI			
B(gs)I	CrtB	(gs)	CrtI			
B(ek)I	CrtB	(ek)	CrtI			
IB	CrtI	-	CrtB			
I(yb)B	CrtI	(yb)	CrtB			
I(cpr)B	CrtI	(cpr)	CrtB			
I(gs)B	CrtI	(gs)	CrtB			
I(ek)B	CrtI	(ek)	CrtB			
Y(yb)B	CrtY	(yb)	CrtB			
Y(yb)B/I	CrtY	(yb)	CrtB		CrtI	
Y(yb)B(ek)I	CrtY	(yb)	CrtB	(ek)	CrtI	β-carotène

Figure 41 : Liste des souches construites dans la partie de l'étude concernant la stratégie de fusion d'enzymes. En allant du haut vers le bas, on a les souches contenant des enzymes individuelles, les souches contenant les bifusions et enfin la fusion tripartite. Les linkers utilisés sont mis en évidence par le fond noir, le trait signifiant une fusion directe sans utilisation de linker. Le produit final susceptible d'être produit par chaque souche est indiqué.

III.5.3 Domaines séparés

En séparant les domaines dans CrtYB, on peut voir si la liaison CrtYB présente spécifiquement chez les champignons est importante ou non. Le domaine CrtB séparé avec le domaine CrtY a donc été étudié, sans CrtI, avec CrtI ou avec CrtI et CrtY. L'expression de chaque enzyme individuelle est mesurée par western blot (figure 42). L'enzyme CrtB est relativement bien exprimée dans chaque souche construite avec une expression similaire entre chaque souche. L'expression de l'enzyme CrtI est difficile à vérifier sur western blot (figure 42). Une non détection par western blot peut être due à une faible accessibilité du tag, une faible expression de l'enzyme ou une migration à plusieurs endroits sur toute la piste due par exemple au fait que la protéine soit membranaire. De même, une bande spécifique à CrtY n'est pas identifiable par western blot (figure 42). CrtY est une enzyme de faible poids moléculaire et possède plusieurs segments transmembranaires. La mesure de la quantité relative des caroténoïdes produite par chaque souche a ensuite confirmé l'expression des enzymes et a permis d'avoir une idée de leur efficacité (figure 43).

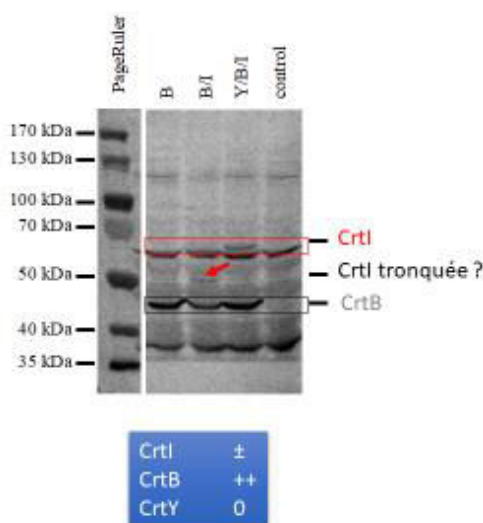


Figure 42 : Expression d'enzymes séparées mesurée par western blot. Les échantillons correspondent à des échantillons collectés après 72h de culture. La ligne contrôle est issue de la souche optimisée pour la production de GGPP mais sans enzymes caroténogéniques. La ligne de dépôt B est celle de la souche contenant CrtB, la ligne de dépôt B/I est celle de la souche contenant CrtB et CrtI, la ligne de dépôt Y/B/I est celle de la souche contenant CrtB, CrtI et CrtY. Un anticorps anti c-myc permet de détecter les enzymes d'intérêts à l'exception du domaine Y isolé qui est détecté par un anticorps anti FLAG.

La souche contenant CrtB seule produit 40 fois moins de phytoène que la souche exprimant l'enzyme naturelle CrtYB (figure 43). La séparation de CrtB avec CrtY impacte très fortement l'activité phytoène synthase. L'activité de CrtB pourrait dépendre de la présence de CrtY. Cela permettrait par exemple à CrtB d'être rapprochée à la membrane, de libérer le produit phytoène dans la membrane. La souche B/I nous informe qu'un système contenant CrtB isolée et CrtI ne permet pas de biotransformer la totalité de phytoène en lycopène (figure 43). Dans cette souche, le faible niveau de production CrtB est confirmé. CrtI ne biotransforme pas toute cette faible quantité de phytoène en lycopène. Cela peut être la conséquence d'une activité de CrtI qui est effectivement limitante. Cela pourrait aussi avoir comme origine un complexe quaternaire non efficace entre les différentes enzymes en l'absence de CrtY. En présence de CrtY, la mesure du niveau de production en caroténoïdes dans la souche Y/B/I nous montre que le niveau de production de la souche naturelle n'est pas retrouvé (figure 44). Cette souche produit 55 fois moins de caroténoïdes que la souche contenant le système naturel complet. CrtB exprimée sous forme isolée reste donc problématique. Aussi, la production du torulène par la souche Y/B/I est aussi à noter (figure 44). Le torulène possède un seul cycle β et cinq insaturations. CrtY séparée pourrait ne pas

permettre la formation de deux cycle β mais permet par contre un CrtI qui catalyse une cinquième désaturation. A l'échelle moléculaire, le mode de fonctionnement est aujourd'hui inconnu. Concernant la souche contenant les enzymes naturelles, on a retrouvé le niveau métrique et l'accumulation de phytoène comme décrite dans la littérature (figure 44). On a aussi détecté d'autres accumulations, à une échelle moins importante, d'autres intermédiaires comme le dihydro- β -carotène, le ζ -carotène ou le neurosporène.

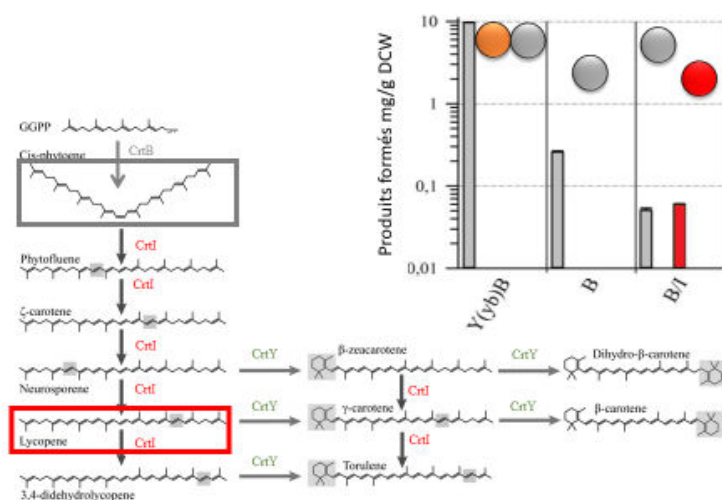


Figure 43 : Quantification des caroténoïdes produits par les souches portant des enzymes individuelles B et I. Les caroténoïdes majeurs ont été quantifiés dans les souches portant les domaines individuels (B, B/I) et ont été comparés à ceux obtenus avec la souche Y(yb)B. La voie métabolique des caroténoïdes est montrée. Les valeurs représentent la moyenne $\pm$ S.D. de la quantification de chaque métabolite de trois cultures indépendantes.

Production caroténogénique des domaines séparés

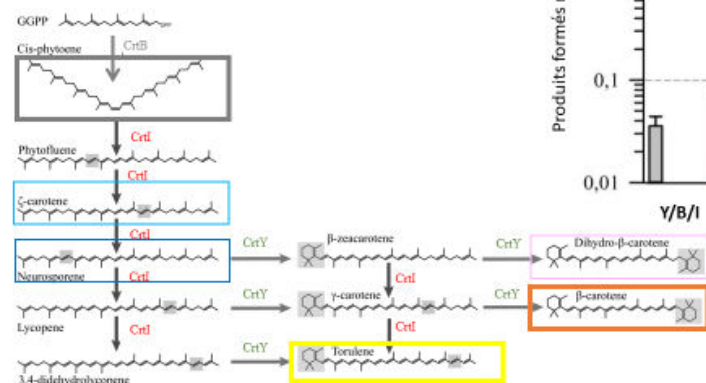


Figure 44 : Quantification des caroténoïdes produits par les souches portant des enzymes individuelles B, I et Y. Les caroténoïdes majeurs ont été quantifiés dans les souches portant les domaines individuels (B/I/Y) et ont été comparés à ceux obtenus avec la souche Y(yb)B/I. La voie métabolique des caroténoïdes est montrée. Les valeurs représentent la moyenne $\pm$ S.D. de la quantification de chaque métabolite de trois cultures indépendantes.

III.5.4 Bifusions

L'étape de biotransformation du phytoène au lycopène est limitante. En réalisant une combinatoire complète de constructions dans laquelle nous avons fusionné CrtB avec CrtI, nous avons sélectionné une conformation efficace. Les différentes souches contenant les différentes bifusions sont représentées (figure 41). Un western blot permet de connaître l'expression de chaque enzyme et de relativiser ensuite sur l'efficacité de chacune (figure 45). Les enzymes fusionnées dans l'orientation IB présente une

forte intensité de bande. Dans l'orientation BI, les bandes attendues sont plutôt faible. Une légère augmentation de l'intensité est observée pour les bandes correspondantes à B(ek)I et B(gs)I.

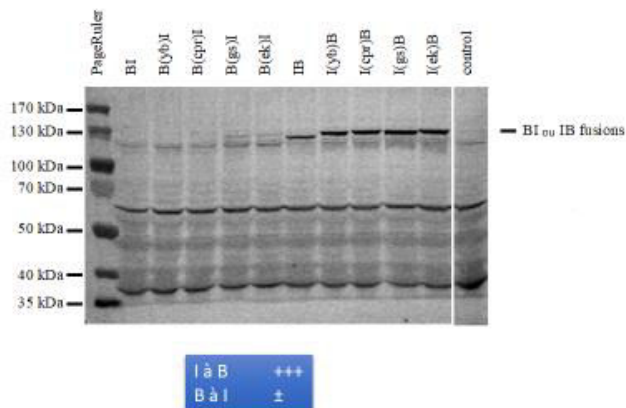
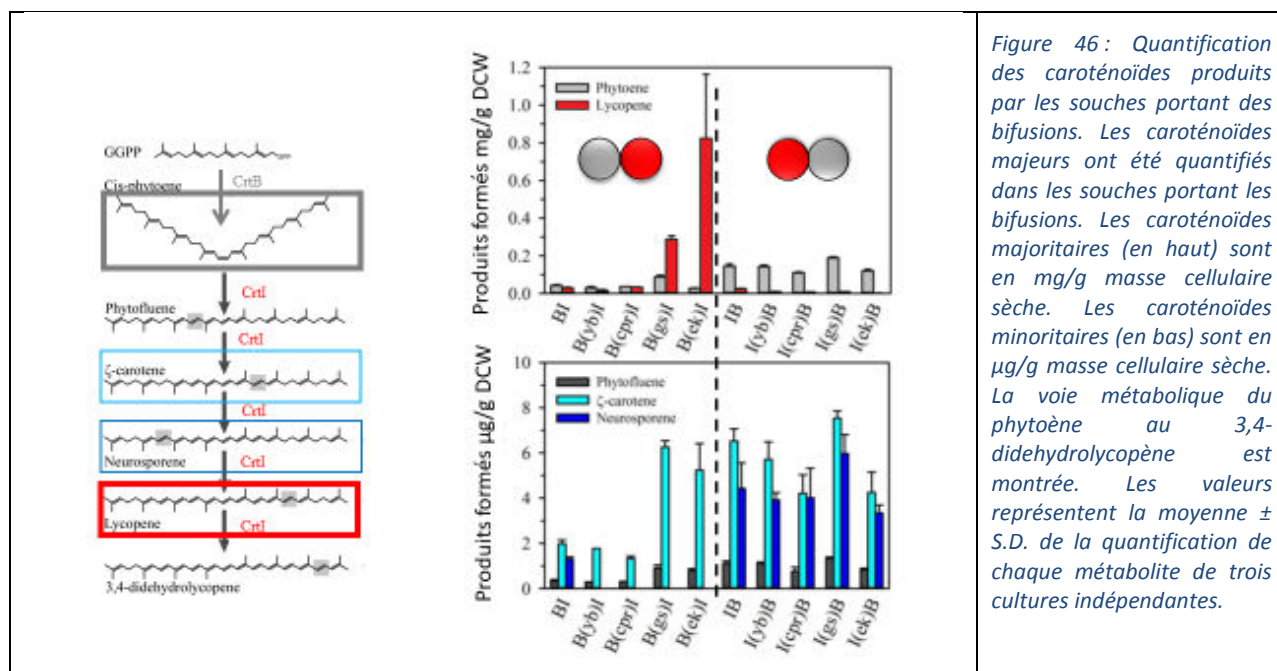


Figure 45 : Expression de bifusions détectée par western blot. Les échantillons du western blot correspondent à des échantillons collectés après 72h de culture. La ligne contrôle est issue de la souche optimisée pour la production de GGPP mais sans enzymes caroténogéniques. Un anticorps anti c-myc permet de détecter les enzymes d'intérêts.



La quantification des caroténoïdes produits mesure l'efficacité des bifusions. La capacité à produire du lycopène pour les systèmes IB est très altérée. Une fusion en C terminale de CrtI n'est donc pas adéquate (figure 46). Dans l'orientation BI, les souches B(gs)I et B(ek)I sortent du lot. En particulier, la bifusion B(ek)I permet la meilleure production en lycopène avec en plus très peu de phytoène accumulé. Ce résultat est celui que l'on a cherché à obtenir dans notre travail. Malgré apparemment une faible expression protéique, cette configuration d'enzyme est efficace pour une biotransformation du GGPP en lycopène.

Les intermédiaires phytofluène, ζ -carotène et neurosporène sont faiblement détecté par rapport au phytoène ou au lycopène (figure 46). Cependant, leur profil d'accumulation permet de comprendre l'efficacité des différentes étapes de désaturation. Dans les souches IB, les profils d'accumulations sont identiques entre les souches. Le ζ -carotène et le neurosporène sont accumulés. Dans les souches BI, le ζ -carotène est celui qui est le plus détecté alors que le neurosporène n'est plus détectable. L'amélioration de la biotransformation du neurosporène en lycopène par la fusion BI participe à l'efficacité générale de ces fusions.

III.5.5 Fusion tripartite

Nous avons donc ensuite réutilisé la fusion B(ek)I dans le contexte d'une fusion tripartite. Une fusion B(ek)IY n'a pas été construite à cause d'une probable interférence avec les mécanismes d'insertion de CrtY dans la membrane. D'autant plus que l'activité de CrtI ne tolère pas une fusion en C –terminale comme nous l'avons vu précédemment. La fusion tripartite que nous avons réalisé est une fusion CrtY(yb)B(ek)I, une fusion conservant la fusion naturelle CrtY(yb)B auquel nous avons fusionné CrtI grâce au linker (EAAAK)₄.

La quantification de l'expression de la fusion tripartite par western blot n'a pas mis en évidence une bande de migration spécifique (figure 47). Une trop faible expression peut par exemple en être l'origine. Nous avons réalisé une étude de protéomique afin de détecter la construction (figure 48). Différentes bandes issues de la migration sur gel des protéines totales issues des souches Y(yb)B(ek)I, YB/I et de la souche de départ ont été coupées et utilisées pour être analysées par spectrométrie de masse (figure 48 à gauche). Les fragments tripsiques détectés recouvrent 43% de la séquence de CrtY(yb)B(ek)I. Un seul fragment tripsique est issu de la séquence CrtY (figure 48A). Cela est dû à une faible occurrence de site tripsique dans cette partie de l'enzyme et aussi à la présence de plusieurs fragments transmembranaires donc fortement hydrophobe et donnant des peptides faiblement ionisables. Aucun fragment tripsique issue de CrtY n'a d'ailleurs été détecté dans le cas de la souche YB/I. Le reste des fragments tripsiques sont donc issus de la partie B(ek)I (figure 48A). Deux fragments chevauchent le linker (EK) avec la fin de l'enzyme CrtB et le début de CrtI respectivement. L'absence de fragment semi-tripsique détecté élimine la possibilité d'un clivage non tripsique de la fusion tripartite. Cette dernière est donc présente sous sa forme en entière in vivo. Enfin, la détection de fragment tripsique à partir de plusieurs bandes de gel (120 à 150, 70 à 80, et 40 à 50 kDa) (figure 48B) montre que la fusion tripartite migre sur plusieurs régions du gel et pourrait donc expliquer la non détection par western blot. Concernant la détection de CrtI, ce dernier est détectable par MS dans la bande de 60 à 70 kDa. CrtYB est également détectée à partir de ce fragment de gel (figure 48B).

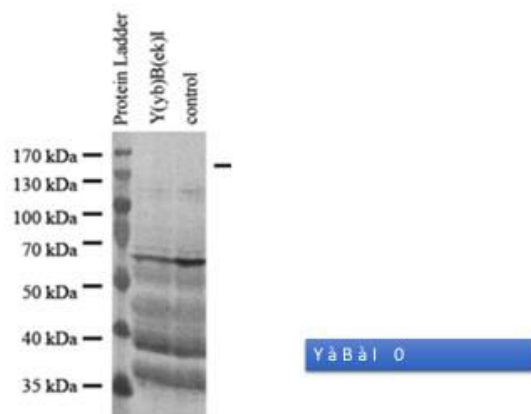


Figure 47 : Expression de la fusion tripartite détectée par western blot. Le niveau de migration attendu pour la fusion tripartite est indiqué. Les échantillons correspondent à des échantillons collectés après 72h de culture. La ligne contrôle est issue de la souche optimisée pour la production de GGPP mais sans enzymes caroténogéniques. Un anticorps anti c-myc permet de détecter les enzymes d'intérêts.

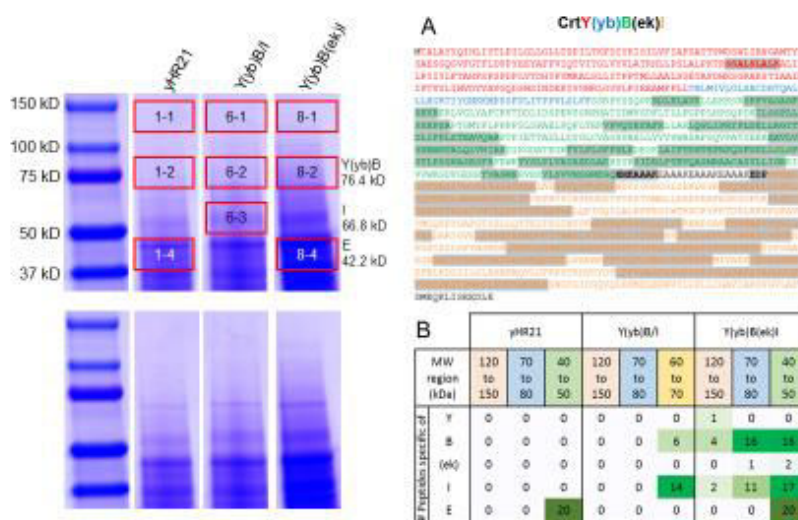


Figure 48 : Détection des peptides CrtY(yb)B(ek)I par analyse GeLC-MS. A. Séquence d'acide aminé de la construction CrtY(yb)B(ek)I. Les différents domaines de la fusion protéique sont coloriés: Y (rouge), (yb) (bleu), B (vert), (ek) (noire), I (orange). Les peptides qui sont susceptibles d'être identifiés sans équivoque sont surlignés. Les peptides identifiés recouvrent 43% de la séquence totale de CrtY(yb)B(ek)I. B. Résumé des données de GeLC-MS. Les nombres correspondent au nombre de fois où des peptides sont identifiés dans un domaine donné (ligne horizontale) et dans une portion de gel donnée (MW région) qui ont été choisies en fonction de la distance de migration suspectée de chaque construction. L'intensité de la coloration verte relate le nombre de peptides détectés.

La quantification des caroténoïdes produits par la souche contenant la fusion tripartite montre une diminution d'un facteur deux de la quantité de phytoène accumulé. La quantité de β -carotène est multipliée par deux (figure 49). La stratégie de fusion a permis d'améliorer la biotransformation du phytoène vers le β -carotène.

La détection du dihydro- β -carotène, dont la biosynthèse ne passe pas par le lycopène pourrait indiquer une voie de biosynthèse qui va du neurosporène puis au β -zeacarotène et donc ne passant pas par le lycopène. D'un autre côté, il est aussi possible que la voie qui transforme le neurosporène au β -zeacarotène puis au dihydro- β -carotène une voie concurrentielle à la voie passant par le lycopène. En effet, en absence de la cyclase, CrtI produit comme produit final majoritaire le lycopène.

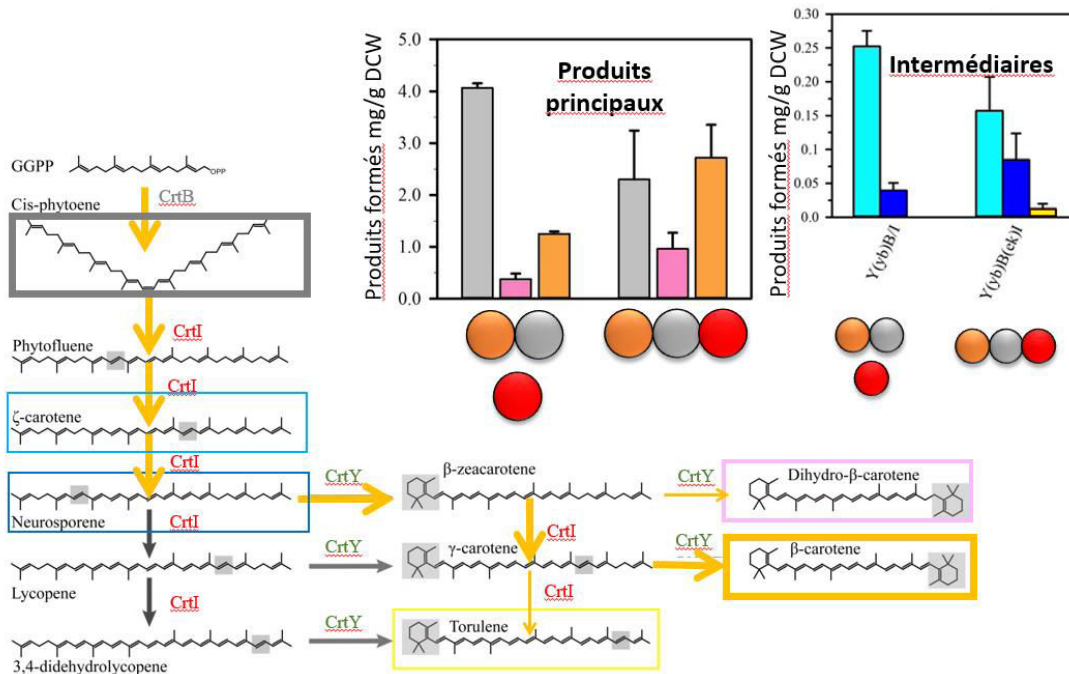


Figure 49 : Quantification des caroténoïdes produits par la souche portant l'enzyme tripartite. Les caroténoïdes majeurs ont été quantifiés dans les souches portant l'enzyme tripartite. La voie métabolique est montrée. Les valeurs représentent la moyenne $\pm$ S.D. de la quantification de chaque métabolite de trois cultures indépendantes.

III-6 Tests de production et de purification des enzymes CrtB et CrtI chez *E. coli*

Une meilleure compréhension du mode fonctionnement de ces enzymes passe par leur caractérisation biochimique et la résolution de leur structure. Une partie de ce travail de thèse a consisté en la réalisation de différents tests de production et de solubilité de CrtB et de CrtI chez *E. coli*.

Les premiers tests de production ont montré que notre forme tronquée de CrtB n'est pas exprimée chez *E. coli* ou n'est exprimée qu'à une trop faible quantité (figure 50). Des tests de production de CrtB dans deux géotypes de cellules différents que sont BL21 DE3 et C41 DE3 n'ont pas donné d'améliorations notables. La souche BL21 DE3 est déficiente pour le gène des protéases omp-t et lon et est souvent utilisée pour l'expression de protéine non toxiques pour la cellule. La souche C41 DE3 est issue de la souche BL21 DE3. Elle possède au moins une mutation qui est décrite comme une mutation diminuant la mort cellulaire due à l'expression de protéines toxiques. Elle a été utilisée pour l'expression de protéines toxiques et membranaires d'espèces d'origines différentes. Les tests d'expression de CrtB dans ces souches n'ont pas permis de détecter la production de la protéine hétérologue. Plusieurs codons rares pour une expression chez *E. coli* sont présents dans la séquence de CrtB. Une optimisation de codon de CrtB a permis la production de CrtB la production de cette dernière chez *E. coli* BL21 DE3. La majorité des protéines se retrouvent sous forme de corps d'inclusion probablement à cause d'une production trop forte ou trop rapide de la protéine. Les stratégies ont consisté à limiter ce taux de production des protéines en diminuant les températures de production ou en variant la concentration en inducteur IPTG utilisé. (figure 52). D'autres conditions de productions devraient encore être réalisées.

CrtI est bien exprimée chez *E. coli* majoritairement sous forme insoluble dans nos premiers tests de production (figure 50). L'enzyme est retenue dans la fraction membranaire ou est exprimée sous forme de corps d'inclusion. Nous avons alors utilisé différentes concentrations de détergents et d'agents chaotropiques pour solubiliser CrtI avec des détergents (figure 53) ou avec des agents chaotropiques (figure 54).

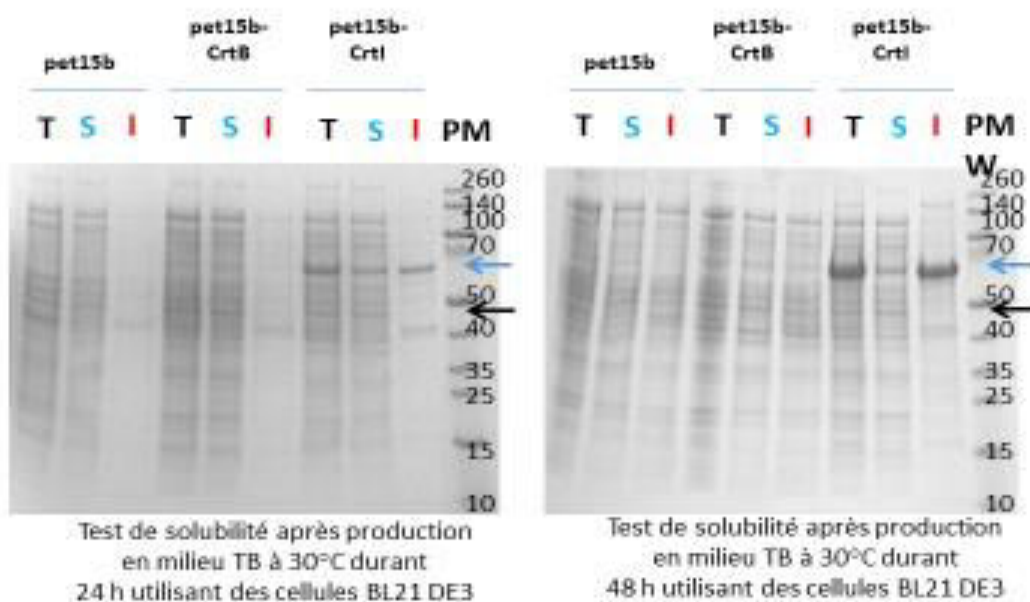


Figure 50 : Tests de production et de la solubilité initiaux des enzymes CrtB et CrtI. Les tailles attendues pour CrtB et CrtI sont 45kDa (flèche noire) et 67kDa (flèche bleue) respectivement. Une même quantité protéique d'échantillons brute total(T), de surnageant (S) et de culot cellulaire resuspendu (I) sont déposés. Les dépôts pet15b sont issus de la souche de production transformée avec le plasmide vide.

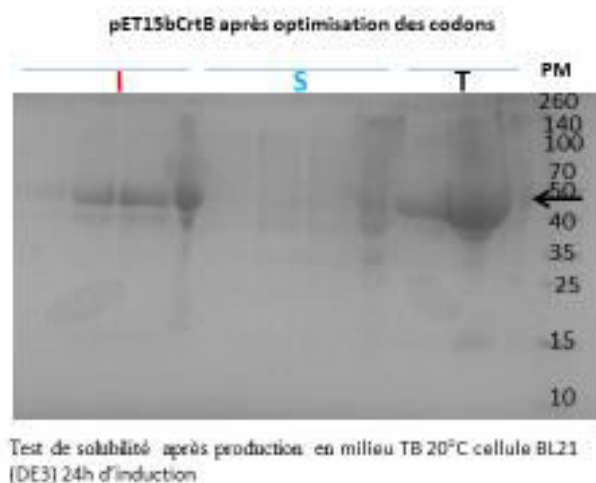


Figure 51 : Tests de production et de la solubilité de CrtB après optimisation de codon. La taille attendue pour CrtB est de 45kDa (flèche noire). Une même quantité protéique d'échantillons brute total(T), de surnageant (S) et de culot cellulaire resuspendu (I) sont déposés.

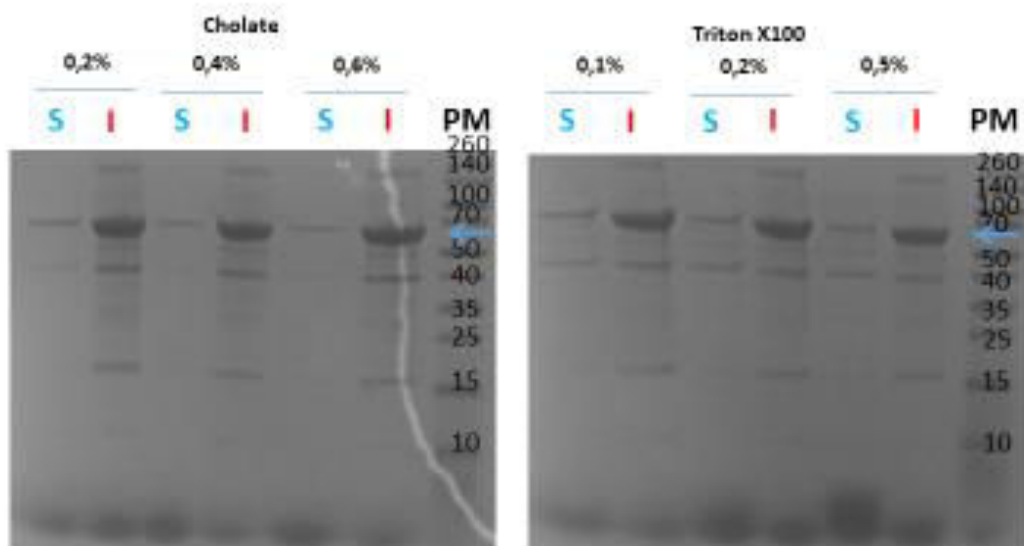


Figure 52 : Utilisation de détergents pour solubiliser la fraction insoluble de CrtI. Différentes concentrations en cholâtes et triton X100 à différentes concentrations pour une même concentration en protéine (S pour surnageant et I pour la fraction insoluble).

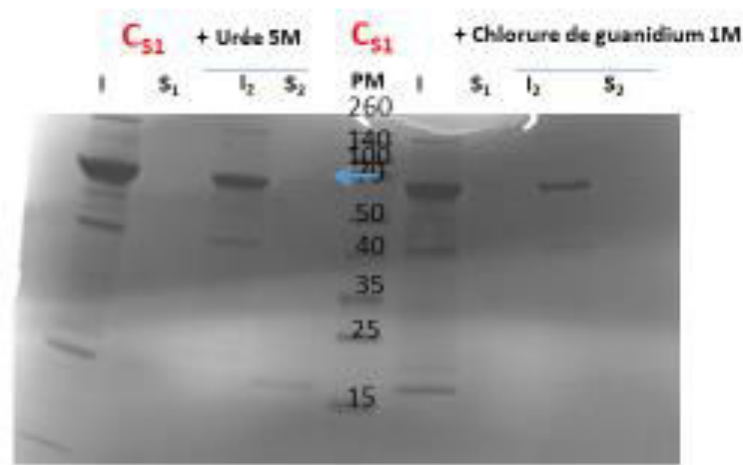


Figure 53 : Utilisation d'agents chaotropiques pour solubiliser la fraction insoluble de CrtI. Une même concentration en protéine a été utilisée et a été déposée sur gel (S pour surnageant et I pour la fraction insoluble). Cs1 est le surnageant initial après la lyse cellulaire.

Les tests de production que nous avons réalisés pour obtenir les enzymes solubles n'ont pas aboutis. D'autres conditions de production et de solubilisation peuvent encore être réalisées pour retrouver les résultats désirés.

IV DISCUSSION ET PERSPECTIVES

L'ingénierie métabolique cherche à établir des solutions pour créer des « usines cellulaires » efficaces pour la production de molécules d'intérêt. Nous avons présenté dans ce manuscrit en premier lieu la biologie de synthèse, la biologie des systèmes et l'évolution adaptative en laboratoire. Ces différentes disciplines apportent des outils dans le développement d'« usine cellulaire » efficaces. Puis nous avons fait un exposé global mais non exhaustif de différents axes de recherche possibles en ingénierie métabolique. La colocalisation spatiale d'enzymes pourrait aider à diminuer l'accumulation non désirée des intermédiaires d'une voie de biosynthèse pour ainsi favoriser la production du produit final. Les enzymes d'intérêt de notre étude sont celles de la voie de biosynthèse du β -carotène de la levure *X. dendrorhous*. Nous avons expérimenté dans ce travail de thèse différentes approches de colocalisation d'enzymes et nous avons mesuré les effets des différentes stratégies sur la voie de biosynthèse d'intérêt in vivo. Trois enzymes sont impliquées dans la biosynthèse du β -carotène, CrtB, CrtI et CrtY. Chez la levure *X. dendrorhous* et chez des champignons caroténogéniques, des enzymes impliquées dans la biosynthèse du β -carotène ont la particularité d'être fusionnées au sein de l'enzyme bifonctionnelle CrtYB. Chez les bactéries ou les plantes par exemple, ces mêmes enzymes caroténogéniques sont produites sous leur forme individuelle. Les enzymes caroténogéniques sont également décrites pour former des associations sous la forme de métabolons. Au début de ce travail de thèse, l'ingénierie métabolique pour la production du β -carotène chez *S. cerevisiae* utilisant les enzymes de *X. dendrorhous* a permis une production élevée en β -carotène par rapport à l'utilisation d'enzymes caroténogéniques bactériennes. Cependant, une accumulation importante en intermédiaire phytoène a été mesurée. Le contexte enzymatique et d'ingénierie métabolique de cette voie de biosynthèse s'apprêtaient donc particulièrement à la mise en place de différentes stratégies de colocalisation d'enzymes et à la sélection des meilleures stratégies de colocalisation. Ce modèle enzymatique, bien que largement utilisé dans la littérature, est peu caractérisé biochimiquement. L'absence de structure tridimensionnelle ou de tests biochimiques sur ces enzymes peuvent s'expliquer par des difficultés dans leur purification. Les enzymes d'intérêt ont des substrats et des produits de réaction très hydrophobes qui interagissent avec la membrane. Enfin un autre intérêt pour ce modèle enzymatique réside dans le fait que des stratégies de colocalisation d'enzyme en interaction avec la membrane sont peu étudiées alors qu'une grande majorité des voies métaboliques naturelles impliquent des enzymes membranaires et des produits non solubles.

En premier lieu, une étude in silico des enzymes a été réalisée. Une analyse de séquences comparant les bifusions naturelles CrtYB avec les séquences de CrtB individuelles et de CrtY individuelles a permis de mettre en évidence la séquence du linker CrtYB. Concernant les deux premières enzymes de la voie de biosynthèse, CrtB et CrtI, des modèles tridimensionnels par homologie ont été obtenus en utilisant les structures résolues d'homologues fonctionnels. Une structure par homologie de CrtY n'a pas été envisagée car la protéine possède plusieurs segments transmembranaires. Le modèle par homologie de CrtB peut être décrit comme étant une structure protéique compacte avec une poche au centre de la protéine contenant le site actif. Les deux extrémités N et C-terminales sont situées en surface et à priori non impliquées dans la fonction catalytique de l'enzyme. Concernant CrtI, les extrémités N et C terminales sont situées du même côté de l'enzyme. Ces extrémités sont proches de sites impliqués dans

l'interaction avec le cofacteur et/ou le canal d'entrée du substrat et du cofacteur. La comparaison de séquence de CrtI avec ses homologues met également en évidence une séquence d'acides aminés à l'extrémité C-terminale de CrtI qui lui est spécifique. Cette séquence d'acide aminés forme une hélice hydrophobe. La fusion de CrtI avec CrtB au niveau de cette extrémité impacte drastiquement la fonction enzymatique. La définition du rôle de cette hélice en C terminale de CrtI reste à envisager dans des études ultérieures. Cette hélice pourrait s'insérer dans la membrane par exemple.

La deuxième étape de l'étude a été ensuite de construire, en utilisant différentes stratégies, des enzymes colocalisées spatialement. Nous avons réutilisé en premier lieu différents domaines d'interaction naturels afin de colocaliser notre système enzymatique. Les domaines dockérine-cohésine retrouvés au niveau du cellulosome sont fusionnés à des enzymes impliquées dans la dégradation de la cellulose. Ces domaines d'interaction ont déjà été utilisés hors de leur contexte pour colocaliser des enzymes. Notre approche a été de fusionner une dockérine en C-terminale de CrtYB et de fusionner une cohésine en N terminale de CrtI. Nous conservons ainsi les configurations dans lesquelles la cohésine et la dockérine sont fusionnées dans leur système naturel. Nous avons d'ailleurs gardé leur linker naturel. Cette stratégie de colocalisation a permis une augmentation du flux total en caroténoïde. Elle a aussi résulté en un changement du type de caroténoïde majoritaire final produit, à savoir le dihydro- β -carotène à la place du β -carotène. CrtI et CrtY ayant plusieurs substrats en commun, les activités de CrtI et de CrtY sont modulées par les différents contextes enzymatiques retrouvés dans les différentes stratégies de colocalisation utilisées. L'accessibilité et la concentration des différents intermédiaires peuvent en effet varier selon les différentes réorganisations d'enzymes. Ainsi, la nature du produit final peut changer. Un avantage de la stratégie de colocalisation par l'intermédiaire de domaines d'interaction est la possibilité de créer des scaffolds protéiques composés de différents domaines d'interaction. Cela permet d'accueillir les enzymes fusionnées aux différents domaines d'interaction respectifs et de faire varier la stœchiométrie des différentes enzymes dans la voie de biosynthèse. L'idée pourrait être par exemple d'accueillir plusieurs enzymes CrtI au niveau du scaffold car elle catalyse l'étape limitante de la voie. On pourrait également imaginer jouer sur l'équilibre entre l'activité CrtI et CrtY et voir quel produit final on obtiendrait. La conception du scaffold protéique nécessitera aussi des optimisations supplémentaires.

Utilisant une autre approche, nous avons testé une stratégie de colocalisation de CrtI. La structure du carboxysome est formée, entre autres protéines, d'hexamères de CcmK2. L'idée a été de fusionner CrtI à CcmK2 et de voir l'effet sur la fonctionnalité de l'enzyme et sur la voie métabolique. L'étude concernant cette stratégie s'est arrêtée à la confirmation de la fonctionnalité *in vivo* des enzymes, cela par l'observation de la coloration des souches transformées sur boîte de petri. Dans l'avenir, la réutilisation de protéine oligomérique pour la colocalisation d'enzyme *in vivo* sera une alternative attrayante car potentiellement modulable. Les protéines sont adaptées aux conditions *in vivo*. On pourrait imaginer contrôler la structure formée par les oligomères et donc designer plus précisément la manière de colocaliser les différentes enzymes.

La troisième stratégie utilisée dans cette étude fut la fusion directe des différentes enzymes de la voie de biosynthèse entre elles. Le phytoène s'accumule lorsqu'on utilise le système d'enzyme caroténogénique naturel chez *S. cerevisiae*. Cette accumulation a été retrouvée dans notre étude. Une

accumulation moins importante d'autres substrats et produits de CrtI a également été mise en évidence dans notre étude. L'accumulation du phytoène et des autres intermédiaires confirment que l'étape catalysée par CrtI est limitante dans la voie de biosynthèse. Dans un premier temps, pour comprendre l'importance de la fusion naturelle CrtYB, notre approche a été de déconstruire le système naturel. La séparation de CrtB de CrtY donne une enzyme CrtB avec une activité quarante fois inférieure au niveau d'activité du système naturel CrtYB. On peut suggérer que la fonctionnalité de CrtB est altérée par une incapacité de la protéine à interagir seule avec la membrane en absence du domaine transmembranaire CrtY. En utilisant des séquences d'adressage au RE par exemple (Peng et al., 2017) que l'on fusionnerait à CrtB, une interaction artificielle de CrtB à la membrane pourrait être testée et l'activité enzymatique de CrtB mesurée à nouveau. En effet, une inefficacité dans l'interaction à la membrane pourrait par exemple empêcher un relargage efficace du produit de réaction phytoène et donc impacter sur l'activité enzymatique.

Nous avons ensuite cherché à colocaliser CrtB et CrtI. Une combinatoire complète de bifusions, dans les deux orientations et utilisant quatre linker différents a permis de vérifier l'effet de l'orientation et de la nature des linkers sur non seulement la fonctionnalité des fusions mais aussi sur le niveau de production des enzymes. Dans le sens IB les enzymes ont une expression élevée mais une fonctionnalité impactée négativement par la fusion. En effet, ces fusions produisent très peu du produit final attendu pour les bifusions : le lycopène. Bien que relativement peu produite in vivo, une bifusion dans le sens BI et avec un linker synthétique formant une hélice alpha a constitué la configuration la plus efficace. Cette configuration a produit le maximum de lycopène avec le minimum d'intermédiaires accumulés.

Pour produire le β -carotène, l'idée fut ensuite de réutiliser la configuration de la bifusion la plus efficace dans un système de fusion tripartite, en conservant la fusion naturelle CrtYB. La fusion tripartite CrtY(yb)B(ek)I conserve l'efficacité de la biotransformation du phytoène au lycopène. Dans cette fusion tripartite, l'accumulation en phytoène est divisée par deux et la production en β -carotène est doublée.

L'analyse des modèles par homologie permet peut-être de comprendre pourquoi certaines fusions ne sont pas fonctionnelles. CrtI possède l'hélice C-terminale très hydrophobe, présente uniquement chez l'enzyme CrtI de *X. dendrorhous*. Cette hélice, comme évoquée précédemment pourrait être située à l'interface protéine membrane ou impliquée dans une fonction importante de l'enzyme. Réaliser d'autres fusions d'enzymes en absence de cette hélice est à envisager.

Le niveau de production d'enzyme change énormément selon la configuration de la fusion. La technique du western blot n'a pas permis de détecter la présence de la fusion tripartite alors qu'elle a été l'enzyme la plus efficace de cette étude pour la production de β -carotène. Une limitation de la stratégie de fusion est la taille finale de la protéine qui est élevée et qui pourrait être problématique pour leur production. *S. cerevisiae* peut avoir du mal à produire ces protéines de grand poids moléculaire ou pourrait les dégrader rapidement par sa machinerie cellulaire. Dans notre cas, nous avons alors complété notre étude par de la protéomique en utilisant de la GeLC-MS. La technique possède un seuil de détection plus important que le western blot et a permis de confirmer la présence de la fusion tripartite. La technique a montré que certaines constructions ne migrent pas aux tailles attendues. Cette méthode a aussi confirmé que la fusion tripartite était présente sous sa forme entière dans les cellules et donc l'absence

de clivages non désirés. Les protéines membranaires, de par leur taille, leur hydrophobicité, peuvent migrer à différentes régions sur gel protéique. La GeLC-MS est en tout cas une technique à réutiliser dans des expérimentations du même genre.

Nous avons découvert dans cette étude que les fusions d'enzymes ont un effet important sur leur niveau de production protéique. D'autres stratégies de quantification d'enzyme moins directes pourraient encore compléter l'étude. Le dosage d'ARNm peut vérifier si une limite dans la production protéique a comme origine le niveau de transcription. Aussi, des techniques utilisant par exemple la luciférase NanoLuc fusionnée au niveau de l'extrémité C terminale de protéines chez la levure permettent de mesurer *in vivo* de manière assez fine la quantité protéique (Masser et al., 2016).

Pour améliorer l'efficacité de production en caroténoïde, on pourrait continuer l'étude par l'amélioration des systèmes enzymatiques. Dans le système de fusion d'enzyme, la vitesse globale dépend de l'activité spécifique de l'enzyme la plus lente au sein de la fusion protéique. Aussi parfaite que soit la conception de la fusion, une enzyme limitante en déterminera l'efficacité globale. Une stratégie pour améliorer la fusion tripartite serait d'augmenter l'activité spécifique de CrtI par ingénierie enzymatique. Ayant trouvé une configuration efficace d'enzyme tripartite, un criblage de mutant sur cette dernière pourrait donner des mutations bénéfiques que l'on n'observerait pas dans des expériences de mutagenèse réalisées sur des protéines séparées. La fusion tripartite pourrait aussi être améliorée en testant des fusions utilisant des enzymes d'origines différentes. L'absence d'interactions naturelles entre des enzymes d'origines différentes pourrait tirer profit des stratégies de colocalisation. L'idée serait d'utiliser par exemple une enzyme CrtI possédant une meilleure activité. D'autre part, une faible affinité pour le substrat peut constituer un facteur limitant dans le système. La phytoène désaturase synthétise trois désaturations successives sur trois substrats (phytoène, phytofluène, neurosporène). Les mécanismes par lesquels l'enzyme interagit avec ces substrats sont inconnus. La compétition entre ces substrats pour le site actif de CrtI peut être à l'origine de l'accumulation de phytoène. L'optimisation peut être réalisée dans cette direction par des analyses de séquences et de la mutagenèse dirigée. Par ailleurs, récemment, une enzyme tripartite naturelle produisant du β -carotène a été utilisée chez *Yarrowia lipolytica* (Gao et al., 2017). La quantité de caroténoïdes produits est relativement faible dans cette étude. Par contre, dans cette enzyme tripartite, l'organisation est IBY. Cette enzyme pourrait être étudiée pour comprendre comment la nature a sélectionné une fusion IB fonctionnelle ainsi qu'une fusion en amont de Y. Cela ouvre en tout cas sur d'autres possibilités d'organisation de fusions. Concernant l'optimisation du linker, la fusion tripartite que nous avons conçue pourrait encore être optimisée en testant d'autres caractéristiques de linker ou en variant par exemple la taille du linker EAAAK. Par exemple, d'autres types de linker ont été créés récemment dans lesquels des caractéristiques de flexibilité et de rigidité sont présentes en même temps dans le même linker (Li et al., 2016a). Ensuite, d'autres stratégies de colocalisation d'enzyme ont également été décrites. Le scaffolding par DNA Origami a été utilisé sur des enzymes membranaires (Kurokawa Tatsuki et al., 2018). Une étude décrit une stratégie de colocalisation utilisant des protéines de liaison avec la membrane (Lin et al., 2017). Cette étude et ces récents exemples démontrent qu'il est désormais possible d'étudier différentes stratégies de colocalisation sur des enzymes membranaires.

Les caractéristiques désirées d'une fusion enzymatique optimale sont une enzyme qui produit plus de produits finaux, dans une période de temps plus courte et avec une efficacité plus élevée. Nous nous sommes limités à une mesure des caroténoïdes produit à un temps fixé de la phase stationnaire. Pour caractériser plus en détail l'efficacité des enzymes *in vivo*, on pourrait commencer par mesurer les niveaux de production à différentes phases de la culture cellulaire. Si l'on arrive ensuite à définir des conditions dans lesquelles on pourrait contrôler et faire varier le niveau de flux en substrat initial GGPP, tout en fixant la concentration des enzymes d'intérêt, on pourrait mesurer *in vivo* les caractéristiques enzymatiques. On réaliserait ainsi de l'enzymologie *in vivo*.

Idéalement, les cycles de design, build, test et learn sont itératifs et doivent être accélérés. Ce ne fut pas le cas dans notre étude car le mode de culture en flasque, les méthodes d'extraction et de purification par HPLC ne l'ont pas permis. Pour accélérer le cycle de design, build, test et learn, des méthodes de quantifications des caroténoïdes, des enzymes et des conditions de cultures à haut débit doivent être imaginées (Furubayashi et al., 2014).

La purification d'enzymes permet enfin leur caractérisation biochimique. Les tests de production des enzymes en vue de leur purification n'ont malheureusement pas été concluant à l'issue de ce travail. Bien qu'il soit possible de purifier des enzymes membranaires utilisant *E. coli*, cela nécessite sûrement plusieurs tests de production et une étude beaucoup plus poussée. *E. coli* est le premier hôte de choix pour la production et la purification d'enzyme. Cet hôte pourrait ne pas posséder une machinerie cellulaire capable de produire efficacement ces enzymes sous leur forme soluble. Comme alternative, la production et la purification des enzymes directement chez *S. cerevisiae* peut être réalisées. Des enzymes sous leur forme fonctionnelle seront plus susceptibles d'être produites. La caractérisation biochimique de ces enzymes reste en tout cas un aspect important qui manque à leur compréhension.

En conclusion, ce travail de thèse a permis d'améliorer la voie de biosynthèse du beta carotène chez *S. cerevisiae* par une stratégie de fusion de gènes et une sélection d'un linker adapté. La mise en place d'une combinatoire de fusion d'enzymes a été nécessaire pour mettre en évidence une fusion efficace. Aussi d'autres stratégies de colocalisation ont permis de tester la fonctionnalité des enzymes dans ces constructions et ouvrent la voie à leur future optimisation. Un protocole de méthodes biophysiques permettant la caractérisation de l'efficacité des différentes constructions a été mis en place et pourra être utilisé pour des expériences ultérieures.

V RÉFÉRENCES BIBLIOGRAPHIQUES

- A George, R., and Heringa, J. (2002). An analysis of protein domain linkers: Their classification and role in protein folding. *Protein Eng.* 15, 871–879.
- Abernathy, M.H., He, L., and Tang, Y.J. (2017). Channeling in native microbial pathways: Implications and challenges for metabolic engineering. *Biotechnol. Adv.* 35, 805–814.
- Agapakis, C.M., Boyle, P.M., and Silver, P.A. (2012). Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat. Chem. Biol.* 8, 527–535.
- Agashe, D., Martinez-Gomez, N.C., Drummond, D.A., and Marx, C.J. (2013). Good Codons, Bad Transcript: Large Reductions in Gene Expression and Fitness Arising from Synonymous Mutations in a Key Enzyme. *Mol. Biol. Evol.* 30, 549–560.
- Ahn, J.-W., and Kim, K.-J. (2015). Crystal structure of 1'-OH-carotenoid 3,4-desaturase from *Nonlabens dokdonensis* DSW-6. *Enzyme Microb. Technol.* 77, 29–37.
- Alam, M.T., Olin-Sandoval, V., Stincone, A., Keller, M.A., Zelezniak, A., Luisi, B.F., and Ralser, M. (2017). The self-inhibitory nature of metabolic networks and its alleviation through compartmentalization. *Nat. Commun.* 8, 16018.
- Albertsen, L., Chen, Y., Bach, L.S., Rattleff, S., Maury, J., Brix, S., Nielsen, J., and Mortensen, U.H. (2011). Diversion of Flux toward Sesquiterpene Production in *Saccharomyces cerevisiae* by Fusion of Host and Heterologous Enzymes. *Appl. Environ. Microbiol.* 77, 1033–1040.
- Alcaíno, J., Baeza, M., and Cifuentes, V. (2016). Carotenoid Distribution in Nature. In *Carotenoids in Nature*, (Springer, Cham), pp. 3–33.
- Alcaíno, J., Fuentealba, M., Cabrera, R., Baeza, M., and Cifuentes, V. (2012). Modeling the Interfacial Interactions between CrtS and CrtR from *Xanthophyllomyces dendrorhous*, a P450 System Involved in Astaxanthin Production. *J. Agric. Food Chem.* 60, 8640–8647.
- Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., and Jirstrand, M. (2014). Kinetic models in industrial biotechnology – Improving cell factory performance. *Metab. Eng.* 24, 38–60.
- An, G.-H., Cho, M.-H., and Johnson, E.A. (1999). Monocyclic carotenoid biosynthetic pathway in the yeast *Phaffia rhodozyma* (*Xanthophyllomyces dendrorhous*). *J. Biosci. Bioeng.* 88, 189–193.
- An, S., Kumar, R., Sheets, E.D., and Benkovic, S.J. (2008). Reversible Compartmentalization of de Novo Purine Biosynthetic Complexes in Living Cells. *Science* 320, 103–106.
- Andrewes, A.G., and Starr, M.P. (1976). (3R,3'R)-astaxanthin from the yeast *Phaffia rhodozyma*. *Phytochemistry* 15, 1009–1011.
- Andrianantoandro, E., Basu, S., Karig, D.K., and Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.* 2.

Aragón, C.M.G., Murillo, F.J., de la GUARDIA, M.D., and Cerdæ-Olmedo, E. (1976). An Enzyme Complex for the Dehydrogenation of Phytoene in *Phycomyces*. *Eur. J. Biochem.* 63, 71–75.

Argos, P. (1990). An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J. Mol. Biol.* 211, 943–958.

Arrach, N., Fernández-Martín, R., Cerdá-Olmedo, E., and Avalos, J. (2001). A single gene for lycopene cyclase, phytoene synthase, and regulation of carotene biosynthesis in *Phycomyces*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 1687–1692.

Bailey, J.E. (1991). Toward a science of metabolic engineering. *Science* 252, 1668–1675.

Banik, U., Zhu, D.-M., Chock, P.B., and Miles, E.W. (1995). The Tryptophan Synthase .alpha.2.beta.2 Complex: Kinetic Studies with a Mutant Enzyme (.beta. K87T) To Provide Evidence for Allosteric Activation by an Aminoacrylate Intermediate. *Biochemistry* 34, 12704–12711.

Bartley, G.E., Schmidhauser, T.J., Yanofsky, C., and Scolnik, P.A. (1990). Carotenoid desaturases from *Rhodobacter capsulatus* and *Neurospora crassa* are structurally and functionally conserved and contain domains homologous to flavoprotein disulfide oxidoreductases. *J. Biol. Chem.* 265, 16020–16024.

Bartley, G.E., Scolnik, P.A., and Beyer, P. (1999). Two *Arabidopsis thaliana* carotene desaturases, phytoene desaturase and ζ -carotene desaturase, expressed in *Escherichia coli*, catalyze a poly-cis pathway to yield pro-lycopene. *Eur. J. Biochem.* 259, 396–403.

Bassard, J.-E., and Halkier, B.A. (2017). How to prove the existence of metabolons? *Phytochem. Rev.* 1–17.

Bayer, E.A., Morag, E., and Lamed, R. (1994). The cellulosome — A treasure-trove for biotechnology. *Trends Biotechnol.* 12, 379–386.

Bayer, E.A., Belaich, J.-P., Shoham, Y., and Lamed, R. (2004). The Cellulosomes: Multienzyme Machines for Degradation of Plant Cell Wall Polysaccharides. *Annu. Rev. Microbiol.* 58, 521–554.

Beurdeley, M., Bietz, F., Li, J., Thomas, S., Stoddard, T., Juillerat, A., Zhang, F., Voytas, D.F., Duchateau, P., and Silva, G.H. (2013). Compact designer TALENs for efficient genome engineering. *Nat. Commun.* 4, 1762.

Blagg, B.S.J., Jarstfer, M.B., Rogers, D.H., and Poulter, C.D. (2002). Recombinant Squalene Synthase. A Mechanism for the Rearrangement of Presqualene Diphosphate to Squalene. *J. Am. Chem. Soc.* 124, 8846–8853.

Blazeck, J., Garg, R., Reed, B., and Alper, H.S. (2012). Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnol. Bioeng.* 109, 2884–2895.

Bonk Michael, Hoffmann Benedikt, Lintig Johannes, Schledz Michael, Al-Babili Salim, Hobeika Elias, Kleinig Hans, and Beyer Peter (2004). Chloroplast Import of Four Carotenoid Biosynthetic Enzymes In

Vitro Reveals Differential Fates Prior to Membrane Binding and Oligomeric Assembly. *Eur. J. Biochem.* 247, 942–950.

Bordbar, A., Monk, J.M., King, Z.A., and Palsson, B.O. (2014). Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* 15, 107–120.

Boyle, N.R., and Gill, R.T. (2012). Tools for genome-wide strain design and construction. *Curr. Opin. Biotechnol.* 23, 666–671.

Boyle, P.M., and Silver, P.A. (2012). Parts plus pipes: synthetic biology approaches to metabolic engineering. *Metab. Eng.* 14, 223–232.

Brausemann, A., Gemmecker, S., Koschmieder, J., Ghisla, S., Beyer, P., and Einsle, O. (2017). Structure of Phytoene Desaturase Provides Insights into Herbicide Binding and Reaction Mechanisms Involved in Carotene Desaturation. *Structure* 25, 1222-1232.e3.

Britton, G. (1995). Structure and properties of carotenoids in relation to function. *FASEB J.* 9, 1551–1558.

Bro, C., and Nielsen, J. (2004). Impact of ‘ome’ analyses on inverse metabolic engineering. *Metab. Eng.* 6, 204–211.

Bu, X., Sun, L., Shang, F., and Yan, G. (2017). Comparative metabolomics profiling of engineered *Saccharomyces cerevisiae* lead to a strategy that improving β -carotene production by acetate supplementation. *PLOS ONE* 12, e0188385.

Buhaescu, I., and Izzedine, H. (2007). Mevalonate pathway: A review of clinical and therapeutical implications. *Clin. Biochem.* 40, 575–584.

Bülöw, L. (1987). Characterization of an artificial bifunctional enzyme, β -galactosidase/galactokinase, prepared by gene fusion. *Eur. J. Biochem.* 163, 443–448.

Bülöw, L., Ljungcrantz, P., and Mosbach, K. (1985). Preparation of a Soluble Bifunctional Enzyme by Gene Fusion. *Nat. Biotechnol.* 3, 821–823.

Çakar, Z.P., Seker, U.O.S., Tamerler, C., Sonderegger, M., and Sauer, U. (2005). Evolutionary engineering of multiple-stress resistant *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 5, 569–578.

Çakar Z. Petek, Turanlı-Yıldız Burcu, Alkım Ceren, Yılmaz Ülkü, and Nielsen Jens (2011). Evolutionary engineering of *Saccharomyces cerevisiae* for improved industrially important properties. *FEMS Yeast Res.* 12, 171–182.

Camara, B. (1993). [32] Plant phytoene synthase complex: Component enzymes, immunology, and biogenesis. In *Methods in Enzymology*, (Academic Press), pp. 352–365.

Cameron, D.E., Bashor, C.J., and Collins, J.J. (2014). A brief history of synthetic biology. *Nat. Rev. Microbiol.* 12, 381–390.

- Campbell, K., Xia, J., and Nielsen, J. (2017). The Impact of Systems Biology on Bioprocessing. *Trends Biotechnol.* 35, 1156–1168.
- Candau, R., Bejarano, E.R., and Cerdá-Olmedo, E. (1991). In vivo channeling of substrates in an enzyme aggregate for beta-carotene biosynthesis. *Proc. Natl. Acad. Sci.* 88, 4936–4940.
- Carlsson, H., Ljungcrantz, P., Bülow, L., and Mosbach, K. (1992). Engineering of lactose metabolism in *E. coli* by introducing β -galactosidase/galactokinase fusion enzymes. *Biotechnol. Lett.* 14, 439–444.
- Carroll, D. (2004). Using Nucleases to Stimulate Homologous Recombination. In *Genetic Recombination*, (Humana Press), pp. 195–207.
- Caspeta, L., Chen, Y., Ghiaci, P., Feizi, A., Buskov, S., Hallström, B.M., Petranovic, D., and Nielsen, J. (2014). Altered sterol composition renders yeast thermotolerant. *Science* 346, 75–78.
- Chae, T.U., Choi, S.Y., Kim, J.W., Ko, Y.-S., and Lee, S.Y. (2017). Recent advances in systems metabolic engineering tools and strategies. *Curr. Opin. Biotechnol.* 47, 67–82.
- Chao, R., Yuan, Y., and Zhao, H. (2015). Recent Advances in DNA Assembly Technologies. *FEMS Yeast Res.* 15, 1–9.
- Chen, A.H., and Silver, P.A. (2012). Designing biological compartmentalization. *Trends Cell Biol.* 22, 662–670.
- Chen, Z., and Zeng, A.-P. (2016). Protein engineering approaches to chemical biotechnology. *Curr. Opin. Biotechnol.* 42, 198–205.
- Chen, X., Zaro, J., and Shen, W.-C. (2013). Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug Deliv. Rev.* 65, 1357–1369.
- Chen Zhen, and Zeng An-Ping (2013). Protein design in systems metabolic engineering for industrial strain development. *Biotechnol. J.* 8, 523–533.
- Chiang, C.-J., Lin, L.-J., Wang, Z.W., Lee, T.-T., and Chao, Y.-P. (2014). Design of a noncovalently linked bifunctional enzyme for whole-cell biotransformation. *Process Biochem.* 49, 1122–1128.
- Compaan, D.M., and Ellington, W.R. (2003). Functional consequences of a gene duplication and fusion event in an arginine kinase. *J. Exp. Biol.* 206, 1545–1556.
- Conrado, R.J., Varner, J.D., and DeLisa, M.P. (2008). Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr. Opin. Biotechnol.* 19, 492–499.
- Copeland, W.B., Bartley, B.A., Chandran, D., Galdzicki, M., Kim, K.H., Sleight, S.C., Maranas, C.D., and Sauro, H.M. (2012). Computational tools for metabolic engineering. *Metab. Eng.* 14, 270–280.

- Cress, B.F., Trantas, E.A., Ververidis, F., Linhardt, R.J., and Koffas, M.A. (2015). Sensitive cells: enabling tools for static and dynamic control of microbial metabolic pathways. *Curr. Opin. Biotechnol.* 36, 205–214.
- Crook, N.C., Schmitz, A.C., and Alper, H.S. (2014). Optimization of a Yeast RNA Interference System for Controlling Gene Expression and Enabling Rapid Metabolic Engineering. *ACS Synth. Biol.* 3, 307–313.
- Csörgő, B., Nyerges, Á., Pósfai, G., and Fehér, T. (2016). System-level genome editing in microbes. *Curr. Opin. Microbiol.* 33, 113–122.
- Curran, K.A., and Alper, H.S. (2012). Expanding the chemical palate of cells by combining systems biology and metabolic engineering. *Metab. Eng.* 14, 289–297.
- 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.
- Curran, K.A., Crook, N.C., Karim, A.S., Gupta, A., Wagman, A.M., and Alper, H.S. (2014). Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat. Commun.* 5, 4002.
- 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.
- 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., et al. (2013). Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* 31, 1039.
- DellaPenna, D., and Pogson, B.J. (2006). VITAMIN SYNTHESIS IN PLANTS: Tocopherols and Carotenoids. *Annu. Rev. Plant Biol.* 57, 711–738.
- Dhamankar, H., and Prather, K.L. (2011). Microbial chemical factories: recent advances in pathway engineering for synthesis of value added chemicals. *Curr. Opin. Struct. Biol.* 21, 488–494.
- van Dijken, J.P., Bauer, J., Brambilla, L., Duboc, P., Francois, J.M., Gancedo, C., Giuseppin, M.L.F., Heijnen, J.J., Hoare, M., Lange, H.C., et al. (2000). An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* 26, 706–714.
- Dogbo, O., and Camara, B. (1987). Purification of isopentenyl pyrophosphate isomerase and geranylgeranyl pyrophosphate synthase from *Capsicum* chromoplasts by affinity chromatography. *Biochim. Biophys. Acta BBA - Lipids Lipid Metab.* 920, 140–148.
- 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.

Dragosits, M., and Mattanovich, D. (2013). Adaptive laboratory evolution – principles and applications for biotechnology. *Microb. Cell Factories* 12, 64.

Drubin, D.A., Way, J.C., and Silver, P.A. (2007). Designing biological systems. *Genes Dev.* 21, 242–254.

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–e142.

Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A.V., Prather, K.L.J., and Keasling, J.D. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27, 753–759.

Elleuche, S. (2015). Bringing functions together with fusion enzymes—from nature’s inventions to biotechnological applications. *Appl. Microbiol. Biotechnol.* 99, 1545–1556.

Endy, D. (2005). Foundations for engineering biology. *Nature* 438, 449–453.

Engels, B., Dahm, P., and Jennewein, S. (2008). Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. *Metab. Eng.* 10, 201–206.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLOS ONE* 4, e5553.

Eriksen, D.T., Lian, J., and Zhao, H. (2014). Protein Design for Pathway Engineering. *J. Struct. Biol.* 185, 234–242.

Eslava, A.P., and Cerdá-Olmedo, E. (1974). Genetic control of phytoene dehydrogenation in *phycomyces*. *Plant Sci. Lett.* 2, 9–14.

F. X. Cunningham, Jr., and Gantt, and E. (1998). Genes and Enzymes of Carotenoid Biosynthesis in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 557–583.

Fang, F., Salmon, K., Shen, M.W.Y., Aeling, K.A., Ito, E., Irwin, B., Tran, U.P.C., Hatfield, G.W., Da Silva, N.A., and Sandmeyer, S. (2011). A vector set for systematic metabolic engineering in *Saccharomyces cerevisiae*. *Yeast Chichester Engl.* 28, 123–136.

Fernández-Castané, A., Fehér, T., Carbonell, P., Pauthenier, C., and Faulon, J.-L. (2014). Computer-aided design for metabolic engineering. *J. Biotechnol.* 192, 302–313.

Fisher, A.K., Freedman, B.G., Bevan, D.R., and Senger, R.S. (2014). A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories. *Comput. Struct. Biotechnol. J.* 11, 91–99.

Fletcher Eugene, Krivoruchko Anastasia, and Nielsen Jens (2015). Industrial systems biology and its impact on synthetic biology of yeast cell factories. *Biotechnol. Bioeng.* 113, 1164–1170.

Fox, E.M., and Howlett, B.J. (2008). Secondary metabolism: regulation and role in fungal biology. *Curr. Opin. Microbiol.* 11, 481–487.

Fraser, P.D., and Bramley, P.M. (2004). The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res.* 43, 228–265.

Fraser, P., Misawa, N., LINDEN, H., YAMANO, S., KOBAYASHI, K., and SANDMANN, G. (1992). Expression in *escherichia-coli*, purification, and reactivation of the recombinant *erwinia-uredovora* phytoene desaturase. *J. Biol. Chem.* 267, 19891–19895.

Furubayashi, M., Ikezumi, M., Kajiwar, J., Iwasaki, M., Fujii, A., Li, L., Saito, K., and Umeno, D. (2014). A High-Throughput Colorimetric Screening Assay for Terpene Synthase Activity Based on Substrate Consumption. *PLoS ONE* 9.

Gao, S., Tong, Y., Zhu, L., Ge, M., Jiang, Y., Chen, D., and Yang, S. (2017). Production of β -carotene by expressing a heterologous multifunctional carotene synthase in *Yarrowia lipolytica*. *Biotechnol. Lett.* 39, 921–927.

Gemmecker, S., Schaub, P., Koschmieder, J., Brausemann, A., Drepper, F., Rodriguez-Franco, M., Ghisla, S., Warscheid, B., Einsle, O., and Beyer, P. (2015). Phytoene Desaturase from *Oryza sativa*: Oligomeric Assembly, Membrane Association and Preliminary 3D-Analysis. *PloS One* 10, e0131717.

Gerster, H. (1993). Anticarcinogenic effect of common carotenoids.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison Iii, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345.

Gokhale, R.S., and Khosla, C. (2000). Role of linkers in communication between protein modules. *Curr. Opin. Chem. Biol.* 4, 22–27.

Gonçalves, E., Raguz Nakic, Z., Zampieri, M., Wagih, O., Ochoa, D., Sauer, U., Beltrao, P., and Saez-Rodriguez, J. (2017). Systematic Analysis of Transcriptional and Post-transcriptional Regulation of Metabolism in Yeast. *PLoS Comput. Biol.* 13.

Grünberg, R., and Serrano, L. (2010). Strategies for protein synthetic biology. *Nucleic Acids Res.* 38, 2663–2675.

Guardia, M.D.D.L., Aragón, C.M.G., Murillo, F.J., and Cerdá-Olmedo, E. (1971). A Carotenogenic Enzyme Aggregate in *Phycomyces*: Evidence from Quantitative Complementation. *Proc. Natl. Acad. Sci.* 68, 2012–2015.

Güldener, U., Heck, S., Fiedler, T., Beinhauer, J., and Hegemann, J.H. (1996). A New Efficient Gene Disruption Cassette for Repeated Use in Budding Yeast. *Nucleic Acids Res.* 24, 2519–2524.

Guo, W., Sheng, J., and Feng, X. (2015). ^{13}C -Metabolic Flux Analysis: An Accurate Approach to Demystify Microbial Metabolism for Biochemical Production. *Bioengineering* 3.

Hammer, S.K., and Avalos, J.L. (2017). Harnessing yeast organelles for metabolic engineering. *Nat. Chem. Biol.* 13, 823–832.

Han, A., Kim, M.-J., Kwak, G.-H., Son, J., Hwang, K.Y., and Kim, H.-Y. (2016). Essential Role of the Linker Region in the Higher Catalytic Efficiency of a Bifunctional MsrA–MsrB Fusion Protein. *Biochemistry* 55, 5117–5127.

Hawkins, A.R., and Lamb, H.K. (1995). The Molecular Biology of Multidomain Proteins Selected Examples. *Eur. J. Biochem.* 232, 7–18.

Heyland, J., Blank, L.M., and Schmid, A. (2011). Quantification of metabolic limitations during recombinant protein production in *Escherichia coli*. *J. Biotechnol.* 155, 178–184.

Hiltunen, H.-M., Illarionov, B., Hedtke, B., Fischer, M., and Grimm, B. (2012). Arabidopsis RIBA Proteins: Two out of Three Isoforms Have Lost Their Bifunctional Activity in Riboflavin Biosynthesis. *Int. J. Mol. Sci.* 13, 14086–14105.

Hong, K.-K., and Nielsen, J. (2012). Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell. Mol. Life Sci.* 69, 2671–2690.

Hong, K.-K., Vongsangnak, W., Vemuri, G.N., and Nielsen, J. (2011). Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis. *Proc. Natl. Acad. Sci.* 108, 12179–12184.

Horn, A.H.C., and Sticht, H. (2015). Synthetic Protein Scaffolds Based on Peptide Motifs and Cognate Adaptor Domains for Improving Metabolic Productivity.

Hou, J., Scalcinati, G., Oldiges, M., and Vemuri, G.N. (2010). Metabolic Impact of Increased NADH Availability in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 76, 851–859.

Huttanus Herbert M., and Feng Xueyang (2017). Compartmentalized metabolic engineering for biochemical and biofuel production. *Biotechnol. J.* 12, 1700052.

Ishii, J., Kondo, T., Makino, H., Ogura, A., Matsuda, F., and Kondo, A. (2014). Three gene expression vector sets for concurrently expressing multiple genes in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 14, 399–411.

Iturriaga, E.A., Velayos, A., and Eslava, A.P. (2000). Structure and function of the genes involved in the biosynthesis of carotenoids in the mucorales. *Biotechnol. Bioprocess Eng.* 5, 263.

Jakočiūnas, T., Bonde, I., Herrgård, M., Harrison, S.J., Kristensen, M., Pedersen, L.E., Jensen, M.K., and Keasling, J.D. (2015). Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. *Metab. Eng.* 28, 213–222.

- Jensen, N.B., Strucko, T., Kildegaard, K.R., David, F., Maury, J., Mortensen, U.H., Forster, J., Nielsen, J., and Borodina, I. (2014). EasyClone: method for iterative chromosomal integration of multiple genes *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 14, 238–248.
- Jones, J.A., Toparlak, Ö.D., and Koffas, M.A. (2015). Metabolic pathway balancing and its role in the production of biofuels and chemicals. *Curr. Opin. Biotechnol.* 33, 52–59.
- Jørgensen, K., Rasmussen, A.V., Morant, M., Nielsen, A.H., Bjarnholt, N., Zagrobelny, M., Bak, S., and Møller, B.L. (2005). Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* 8, 280–291.
- Kampranis, S.C., and Makris, A.M. (2012). DEVELOPING A YEAST CELL FACTORY FOR THE PRODUCTION OF TERPENOIDS. *Comput. Struct. Biotechnol. J.* 3, e201210006.
- Keasling, J.D. (2010). Manufacturing Molecules Through Metabolic Engineering. *Science* 330, 1355–1358.
- Keasling, J.D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* 14, 189–195.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modelling, prediction and analysis. *Nat. Protoc.* 10, 845–858.
- Kelwick, R., MacDonald, J.T., Webb, A.J., and Freemont, P. (2014). Developments in the Tools and Methodologies of Synthetic Biology. *Front. Bioeng. Biotechnol.* 2.
- Kerfeld, C.A., Heinhorst, S., and Cannon, G.C. (2010). Bacterial Microcompartments. *Annu. Rev. Microbiol.* 64, 391–408.
- Kerkhoven, E.J., Lahtvee, P.-J., and Nielsen, J. (2015). Applications of computational modeling in metabolic engineering of yeast. *FEMS Yeast Res.* 15, 1–13.
- Kholodenko, B.N., and Westerhoff, H.V. (2004). *Metabolic Engineering in the Post Genomic Era* (Wymondham, Norfolk, England: Taylor & Francis).
- Kim, S., and Hahn, J.-S. (2014). Synthetic scaffold based on a cohesin–dockerin interaction for improved production of 2,3-butanediol in *Saccharomyces cerevisiae*.
- Kim, B., Du, J., Eriksen, D.T., and 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, 931–941.
- Kim, S., Bae, S.-J., and Hahn, J.-S. (2016). Redirection of pyruvate flux toward desired metabolic pathways through substrate channeling between pyruvate kinase and pyruvate-converting enzymes in *Saccharomyces cerevisiae*. *Sci. Rep.* 6.

Kimchi-Sarfaty, Jm, O., Iw, K., Ze, S., Am, C., Sv, A., Mm, G., C, K.-S., Jm, O., Iw, K., et al. (2007). A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science* 315.

Kleinig, H. (1989). The Role of Plastids in Isoprenoid Biosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 40, 39–59.

Kok, S. de, Kozak, B.U., Pronk, J.T., and Maris, A.J.A. van (2012). Energy coupling in *Saccharomyces cerevisiae*: selected opportunities for metabolic engineering. *FEMS Yeast Res.* 12, 387–397.

Kok, S. de, Stanton, L.H., Slaby, T., Durot, M., Holmes, V.F., Patel, K.G., Platt, D., Shapland, E.B., Serber, Z., Dean, J., et al. (2014). Rapid and Reliable DNA Assembly via Ligase Cycling Reaction. *ACS Synth. Biol.* 3, 97–106.

Koutinas, M., Kiparissides, A., Pistikopoulos, E.N., and Mantalaris, A. (2012). Bioprocess systems engineering: transferring traditional process engineering principles to industrial biotechnology. *Comput. Struct. Biotechnol. J.* 3, e201210022.

Krinsky, N.I., and Yeum, K.-J. (2003). Carotenoid–radical interactions. *Biochem. Biophys. Res. Commun.* 305, 754–760.

Krubasik, P., and Sandmann, G. (2000). Molecular evolution of lycopene cyclases involved in the formation of carotenoids with ionone end groups. *Biochem. Soc. Trans.* 28, 806–810.

Kuijpers, N.G., Chroumpi, S., Vos, T., Solis-Escalante, D., Bosman, L., Pronk, J.T., Daran, J.-M., and 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, 769–781.

Kurokawa Tatsuki, Kiyonaka Shigeki, Nakata Eiji, Endo Masayuki, Koyama Shohei, Mori Emiko, Tran Nam Ha, Dinh Huyen, Suzuki Yuki, Hidaka Kumi, et al. (2018). DNA Origami Scaffolds as Templates for Functional Tetrameric Kir3 K⁺ Channels. *Angew. Chem. Int. Ed.* 57, 2586–2591.

Kuznetsova, I.M., Turoverov, K.K., and Uversky, V.N. (2014). What Macromolecular Crowding Can Do to a Protein. *Int. J. Mol. Sci.* 15, 23090–23140.

Kwok, R. (2010). Five hard truths for synthetic biology. *Nature* 463, 288–290.

Lanza, A.M., Curran, K.A., Rey, L.G., and Alper, H.S. (2014). A condition-specific codon optimization approach for improved heterologous gene expression in *Saccharomyces cerevisiae*. *BMC Syst. Biol.* 8, 33.

Laursen, T., Møller, B.L., and Bassard, J.-E. (2015). Plasticity of specialized metabolism as mediated by dynamic metabolons. *Trends Plant Sci.* 20, 20–32.

Lawrence, A.D., Frank, S., Newnham, S., Lee, M.J., Brown, I.R., Xue, W.-F., Rowe, M.L., Mulvihill, D.P., Prentice, M.B., Howard, M.J., et al. (2014). Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. *ACS Synth. Biol.* 3, 454–465.

- Leavitt, J.M., and Alper, H.S. (2015). Advances and current limitations in transcript-level control of gene expression. *Curr. Opin. Biotechnol.* 34, 98–104.
- Lee, S.Y., and Kim, H.U. (2015). Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.* 33, 1061.
- Lee, D., Lloyd, N.D.R., Pretorius, I.S., and Borneman, A.R. (2016). Heterologous production of raspberry ketone in the wine yeast *Saccharomyces cerevisiae* via pathway engineering and synthetic enzyme fusion. *Microb. Cell Factories* 15.
- Lee, H., DeLoache, W.C., and Dueber, J.E. (2012). Spatial organization of enzymes for metabolic engineering. *Metab. Eng.* 14, 242–251.
- Lewis, N.E., Nagarajan, H., and Palsson, B.O. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* 10, 291–305.
- Li, M., and Borodina, I. (2015). Application of synthetic biology for production of chemicals in yeast *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 15, 1–12.
- Li, G., Huang, Z., Zhang, C., Dong, B.-J., Guo, R.-H., Yue, H.-W., Yan, L.-T., and Xing, X.-H. (2016a). Construction of a linker library with widely controllable flexibility for fusion protein design. *Appl. Microbiol. Biotechnol.* 100, 215–225.
- Li, J., Lee, E.-J., Chang, L., and Facchini, P.J. (2016b). Genes encoding norcoclaurine synthase occur as tandem fusions in the Papaveraceae. *Sci. Rep.* 6, 39256.
- Lian, J., Jin, R., and Zhao, H. (2016). Construction of plasmids with tunable copy numbers in *Saccharomyces cerevisiae* and their applications in pathway optimization and multiplex genome integration. *Biotechnol. Bioeng.* 113, 2462–2473.
- Liao, B., Hill, G.A., and Roesler, W.J. (2012). Stable expression of barley α -amylase in *S. cerevisiae* for conversion of starch into bioethanol. *Biochem. Eng. J.* 64, 8–16.
- Ligon, B.L. (2004). Penicillin: its discovery and early development. *Semin. Pediatr. Infect. Dis.* 15, 52–57.
- Lin, F.-Y., Liu, C.-I., Liu, Y.-L., Zhang, Y., Wang, K., Jeng, W.-Y., Ko, T.-P., Cao, R., Wang, A.H.-J., and Oldfield, E. (2010). Mechanism of action and inhibition of dehydrosqualene synthase. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21337–21342.
- Lin, J.-L., Zhu, J., and Wheeldon, I. (2017). Synthetic Protein Scaffolds for Biosynthetic Pathway Colocalization on Lipid Droplet Membranes. *ACS Synth. Biol.* 6, 1534–1544.
- Lindbladh, C., Persson, M., Bülow, L., and Mosbach, K. (1992). Characterization of a recombinant bifunctional enzyme, galactose dehydrogenase/bacterial luciferase, displaying an improved bioluminescence in a three-enzyme system. *Eur. J. Biochem.* 204, 241–247.

- Lindbladh, C., Brodeur, R.D., Lilius, G., Buelow, L., Mosbach, K., and Srere, P.A. (1994). Metabolic studies on *Saccharomyces cerevisiae* containing fused citrate synthase/malate dehydrogenase. *Biochemistry* 33, 11684–11691.
- Liu, C.-I., Liu, G.Y., Song, Y., Yin, F., Hensler, M.E., Jeng, W.-Y., Nizet, V., Wang, A.H.-J., and Oldfield, E. (2008). A cholesterol biosynthesis inhibitor blocks *Staphylococcus aureus* virulence. *Science* 319, 1391–1394.
- Liu, C.-I., Jeng, W.-Y., Chang, W.-J., Shih, M.-F., Ko, T.-P., and Wang, A.H.-J. (2014). Structural insights into the catalytic mechanism of human squalene synthase. *Acta Crystallogr. D Biol. Crystallogr.* 70, 231–241.
- Liu, F., Banta, S., and Chen, W. (2013a). Functional assembly of a multi-enzyme methanol oxidation cascade on a surface-displayed trifunctional scaffold for enhanced NADH production. *Chem. Commun.* 49, 3766–3768.
- Liu, G.Y., Essex, A., Buchanan, J.T., Datta, V., Hoffman, H.M., Bastian, J.F., Fierer, J., and Nizet, V. (2005). *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *J. Exp. Med.* 202, 209–215.
- Liu, L., Redden, H., and Alper, H.S. (2013b). Frontiers of yeast metabolic engineering: diversifying beyond ethanol and *Saccharomyces*. *Curr. Opin. Biotechnol.* 24, 1023–1030.
- Liu, X., Ding, W., and Jiang, H. (2017). Engineering microbial cell factories for the production of plant natural products: from design principles to industrial-scale production. *Microb. Cell Factories* 16.
- Ljungcrantz, P., Carlsson, H., Mansson, M.O., Buckel, P., Mosbach, K., and Buelow, L. (1989). Construction of an artificial bifunctional enzyme, .beta.-galactosidase/galactose dehydrogenase, exhibiting efficient galactose channeling. *Biochemistry* 28, 8786–8792.
- Long, M. (2000). A New Function Evolved from Gene Fusion. *Genome Res.* 10, 1655–1657.
- Long, M.R., Ong, W.K., and Reed, J.L. (2015). Computational methods in metabolic engineering for strain design. *Curr. Opin. Biotechnol.* 34, 135–141.
- Long, M.S., Jones, C.D., Helfrich, M.R., Mangeney-Slavin, L.K., and Keating, C.D. (2005). Dynamic microcompartmentation in synthetic cells. *Proc. Natl. Acad. Sci.* 102, 5920–5925.
- Lopez Alex B., Yang Yong, Thannhauser Theodore W., and Li Li (2008). Phytoene desaturase is present in a large protein complex in the plastid membrane. *Physiol. Plant.* 133, 190–198.
- Lu, T.K., Khalil, A.S., and Collins, J.J. (2009). Next-Generation Synthetic Gene Networks. *Nat. Biotechnol.* 27, 1139–1150.
- Lund, A.M., Kildegaard, H.F., Petersen, M.B.K., Rank, J., Hansen, B.G., Andersen, M.R., and Mortensen, U.H. (2014). A Versatile System for USER Cloning-Based Assembly of Expression Vectors for Mammalian Cell Engineering. *PLOS ONE* 9, e96693.

- Marcheschi, R.J., Gronenberg, L.S., and Liao, J.C. (2013). Protein engineering for metabolic engineering: current and next-generation tools. *Biotechnol. J.* 8, 545–555.
- Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D., and Keasling, J.D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 21, 796–802.
- März, U. (2015). The Global Market for Carotenoids: FOD025E | BCC Research.
- Masser, A.E., Kandasamy, G., Kaimal, J.M., and Andréasson, C. (2016). Luciferase NanoLuc as a reporter for gene expression and protein levels in *Saccharomyces cerevisiae*. *Yeast Chichester Engl.* 33, 191–200.
- Maudinas, B., Bucholtz, M.L., Papastephanou, C., Katiyar, S.S., Briedis, A.V., and Porter, J.W. (1977). The partial purification and properties of a phytoene synthesizing enzyme system. *Arch. Biochem. Biophys.* 180, 354–362.
- Maury, J., Asadollahi, M.A., Møller, K., Clark, A., and Nielsen, J. (2005). Microbial Isoprenoid Production: An Example of Green Chemistry through Metabolic Engineering. *Biotechnol. Future* 19–51.
- Mazzoli, R., Lamberti, C., and Pessione, E. (2012). Engineering new metabolic capabilities in bacteria: lessons from recombinant cellulolytic strategies. *Trends Biotechnol.* 30, 111–119.
- McKeague, M., Wong, R.S., and Smolke, C.D. (2016). Opportunities in the design and application of RNA for gene expression control. *Nucleic Acids Res.* 44, 2987–2999.
- Medema, M.H., and Osbourn, A. (2016). Computational genomic identification and functional reconstitution of plant natural product biosynthetic pathways. *Nat. Prod. Rep.* 33, 951–962.
- Meynial Salles, I., Forchhammer, N., Croux, C., Girbal, L., and Soucaille, P. (2007). Evolution of a *Saccharomyces cerevisiae* metabolic pathway in *Escherichia coli*. *Metab. Eng.* 9, 152–159.
- Milani, M., Pesce, A., Bolognesi, M., Bocedi, A., and Ascenzi, P. (2003). Substrate channeling: Molecular bases. *Biochem. Mol. Biol. Educ.* 31, 228–233.
- Minton, A.P. (2001). The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media. *J. Biol. Chem.* 276, 10577–10580.
- Misra, A., Conway, M.F., Johnnie, J., Qureshi, T.M., Lige, B., Derrick, A.M., Agbo, E.C., and Sriram, G. (2013). Metabolic analyses elucidate non-trivial gene targets for amplifying dihydroartemisinic acid production in yeast. *Front. Microbiol.* 4.
- Mitzka-Schnabel, U., and Rau, W. (1981). Subcellular site of carotenoid biosynthesis in *Neurospora crassa*. *Phytochemistry* 20, 63–69.
- Mizoguchi, T., Harada, J., Yamamoto, K., and Tamiaki, H. (2015). Inactivation of *bciD* and *bchU* genes in the green sulfur bacterium *Chlorobaculum limnaeum* and alteration of photosynthetic pigments in the resultant mutants. *J. Photochem. Photobiol. Chem.* 313, 52–59.

Moise, A.R., Al-Babili, S., and Wurtzel, E.T. (2014). Mechanistic aspects of carotenoid biosynthesis. *Chem. Rev.* 114, 164–193.

Moliné, M., Libkind, D., and van Broock, M. (2012). Production of Torularhodin, Torulene, and β -Carotene by *Rhodotorula* Yeasts. *Methods Mol. Biol.* Clifton NJ 898, 275–283.

Moody Suzy C., Zhao Bin, Lei Li, Nelson David R., Mullins Jonathan G. L., Waterman Michael R., Kelly Steven L., and Lamb David C. (2012). Investigating conservation of the albaflavone biosynthetic pathway and CYP170 bifunctionality in streptomycetes. *FEBS J.* 279, 1640–1649.

Moraïs, S., Morag, E., Barak, Y., Goldman, D., Hadar, Y., Lamed, R., Shoham, Y., Wilson, D.B., and Bayer, E.A. (2012). Deconstruction of Lignocellulose into Soluble Sugars by Native and Designer Cellulosomes. *MBio* 3, e00508-12.

Mourão, M.A., Hakim, J.B., and Schnell, S. (2014). Connecting the Dots: The Effects of Macromolecular Crowding on Cell Physiology. *Biophys. J.* 107, 2761–2766.

Moutsatsos, I.K., and Parker, C.N. (2016). Recent advances in quantitative high throughput and high content data analysis. *Expert Opin. Drug Discov.* 11, 415–423.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540.

Naesby, M., Nielsen, S.V., Nielsen, C.A., Green, T., Tange, T.Ø., Simón, E., Knechtle, P., Hansson, A., Schwab, M.S., Titiz, O., et al. (2009). Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in *Saccharomyces cerevisiae*. *Microb. Cell Factories* 8, 45.

Naseri, G., Balazadeh, S., Machens, F., Kamranfar, I., Messerschmidt, K., and Mueller-Roeber, B. (2017). Plant-Derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 6, 1742–1756.

Nicolaou, S.A., Gaida, S.M., and Papoutsakis, E.T. (2010). A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. *Metab. Eng.* 12, 307–331.

Nielsen, J. (2015). Yeast cell factories on the horizon. *Science* 349, 1050–1051.

O'Brien, E.J., Monk, J.M., and Palsson, B.O. (2015). Using Genome-scale Models to Predict Biological Capabilities. *Cell* 161, 971–987.

Ohto, C., Muramatsu, M., Obata, S., Sakuradani, E., and Shimizu, S. (2010). Production of geranylgeraniol on overexpression of a prenyl diphosphate synthase fusion gene in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 87, 1327–1334.

- Österlund, T., Bordel, S., and Nielsen, J. (2015). Controllability analysis of transcriptional regulatory networks reveals circular control patterns among transcription factors. *Integr. Biol.* 7, 560–568.
- Özaydin, B., Burd, H., Lee, T.S., and Keasling, J.D. (2013). Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production. *Metab. Eng.* 15, 174–183.
- Packer, L., Hiramatsu, M., and Yoshikawa, T. (1999). *Antioxidant Food Supplements in Human Health* - 1st Edition.
- Paddon, C.J., and Keasling, J.D. (2014). Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* 12, 355–367.
- Paine, M.J.I., Scrutton, N.S., Munro, A.W., Gutierrez, A., Roberts, G.C.K., and Wolf, C.R. (2005). Electron Transfer Partners of Cytochrome P450. In *Cytochrome P450*, (Springer, Boston, MA), pp. 115–148.
- Pandit, A.V., Srinivasan, S., and Mahadevan, R. (2017). Redesigning metabolism based on orthogonality principles. *Nat. Commun.* 8, 15188.
- Pandit, J., Danley, D.E., Schulte, G.K., Mazzalupo, S., Pauly, T.A., Hayward, C.M., Hamanaka, E.S., Thompson, J.F., and Harwood, H.J. (2000). Crystal Structure of Human Squalene Synthase A KEY ENZYME IN CHOLESTEROL BIOSYNTHESIS. *J. Biol. Chem.* 275, 30610–30617.
- Papagiannakis, A., de Jonge, J.J., Zhang, Z., and Heinemann, M. (2017). Quantitative characterization of the auxin-inducible degron: a guide for dynamic protein depletion in single yeast cells. *Sci. Rep.* 7.
- Parry, B.R., Surovtsev, I.V., Cabeen, M.T., O’Hern, C.S., Dufresne, E.R., and Jacobs-Wagner, C. (2014). The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. *Cell* 156, 183–194.
- Peng, B., Plan, M.R., Chrysanthopoulos, P., Hodson, M.P., Nielsen, L.K., and Vickers, C.E. (2017). A squalene synthase protein degradation method for improved sesquiterpene production in *Saccharomyces cerevisiae*. *Metab. Eng.* 39, 209–219.
- Pirie, C.M., De Mey, M., Jones Prather, K.L., and Ajikumar, P.K. (2013). Integrating the protein and metabolic engineering toolkits for next-generation chemical biosynthesis. *ACS Chem. Biol.* 8, 662–672.
- Pitera, D.J., Paddon, C.J., Newman, J.D., and Keasling, J.D. (2007). Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab. Eng.* 9, 193–207.
- Pleiss, J. (2011). Protein design in metabolic engineering and synthetic biology. *Curr. Opin. Biotechnol.* 22, 611–617.
- Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12, 32–42.
- Polakowski, T., Stahl, U., and Lang, C. (1998). Overexpression of a cytosolic hydroxymethylglutaryl-CoA reductase leads to squalene accumulation in yeast. *Appl. Microbiol. Biotechnol.* 49, 66–71.

- Polka, J.K., Hays, S.G., and Silver, P.A. (2016). Building Spatial Synthetic Biology with Compartments, Scaffolds, and Communities. *Cold Spring Harb. Perspect. Biol.* 8, a024018.
- Portela, R.M.C., Vogl, T., Kniely, C., Fischer, J.E., Oliveira, R., and Glieder, A. (2017). Synthetic Core Promoters as Universal Parts for Fine-Tuning Expression in Different Yeast Species. *ACS Synth. Biol.* 6, 471–484.
- Portnoy, V.A., Bezdan, D., and Zengler, K. (2011). Adaptive laboratory evolution—harnessing the power of biology for metabolic engineering. *Curr. Opin. Biotechnol.* 22, 590–594.
- Purdy, H.M., and Reed, J.L. (2017). Evaluating the capabilities of microbial chemical production using genome-scale metabolic models. *Curr. Opin. Syst. Biol.* 2, 91–97.
- Purnick, P.E.M., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* 10, 410–422.
- Quan, J., and Tian, J. (2009). Circular Polymerase Extension Cloning of Complex Gene Libraries and Pathways. *PLOS ONE* 4, e6441.
- Rau, W., and Mitzka-Schnabel, U. (1985). [29] Carotenoid synthesis in *Neurospora crassa*. In *Methods in Enzymology*, (Academic Press), pp. 253–267.
- Rice-Evans, C.A., Sampson, J., Bramley, P.M., and Holloway, D.E. (1997). Why Do We Expect Carotenoids to be Antioxidants in vivo? *Free Radic. Res.* 26, 381–398.
- Rico, J., Pardo, E., and Orejas, M. (2010). Enhanced Production of a Plant Monoterpene by Overexpression of the 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Catalytic Domain in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 76, 6449–6454.
- Riley, G.J.P., and Bramley, P.M. (1976). The subcellular distribution of carotenoids in *Phycomyces blakesleeanus* C115 car-42 mad-107(–). *Biochim. Biophys. Acta BBA - Lipids Lipid Metab.* 450, 429–440.
- 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., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940–943.
- Rogers, J.K., Taylor, N.D., and Church, G.M. (2016). Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol.* 42, 84–91.
- Roldão, A., Kim, I.-K., and Nielsen, J. (2012). Bridging Omics Technologies with Synthetic Biology in Yeast Industrial Biotechnology. In *Systems Metabolic Engineering*, (Springer, Dordrecht), pp. 271–327.
- Ronda, C., Maury, J., Jakociūnas, T., Baallal Jacobsen, S.A., Germann, S.M., Harrison, S.J., Borodina, I., Keasling, J.D., Jensen, M.K., and Nielsen, A.T. (2015). CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. *Microb. Cell Factories* 14, 97.

Ruban, A.V., Berera, R., Iliaia, C., Stokkum, I.H.M. van, Kennis, J.T.M., Pascal, A.A., Amerongen, H. van, Robert, B., Horton, P., and Grondelle, R. van (2007). Identification of a mechanism of photoprotective energy dissipation in higher plants. *Nature* 450, 575–578.

Sadeghi, S.J., and Gilardi, G. (2013). Chimeric P450 enzymes: Activity of artificial redox fusions driven by different reductases for biotechnological applications. *Biotechnol. Appl. Biochem.* 60, 102–110.

Samborska, B., and Kimber, M.S. (2012). A Dodecameric CcmK2 Structure Suggests β -Carboxysomal Shell Facets Have a Double-Layered Organization. *Structure* 20, 1353–1362.

Sambrook, J., and Russell, D.W. (2006). SDS-Polyacrylamide Gel Electrophoresis of Proteins. *Cold Spring Harb. Protoc.* 2006, pdb.prot4540.

Santos, C.N.S., Koffas, M., and Stephanopoulos, G. (2011). Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab. Eng.* 13, 392–400.

Sanz, C., Alvarez, M.I., Orejas, M., Velayos, A., Eslava, A.P., and Benito, E.P. (2002). Interallelic complementation provides genetic evidence for the multimeric organization of the *Phycomyces blakesleeana* phytoene dehydrogenase. *Eur. J. Biochem.* 269, 902–908.

Schaub, P., Yu, Q., Gemmecker, S., Poussin-Courmontagne, P., Mailliot, J., McEwen, A.G., Ghisla, S., Al-Babili, S., Cavarelli, J., and Beyer, P. (2012a). On the Structure and Function of the Phytoene Desaturase CRTI from *Pantoea ananatis*, a Membrane-Peripheral and FAD-Dependent Oxidase/Isomerase. *PLoS ONE* 7.

Schaub, P., Yu, Q., Gemmecker, S., Poussin-Courmontagne, P., Mailliot, J., McEwen, A.G., Ghisla, S., Al-Babili, S., Cavarelli, J., and Beyer, P. (2012b). On the Structure and Function of the Phytoene Desaturase CRTI from *Pantoea ananatis*, a Membrane-Peripheral and FAD-Dependent Oxidase/Isomerase. *PLoS ONE* 7, e39550.

Schiestl, R.H., and Gietz, R.D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr. Genet.* 16, 339–346.

Schledz, M., Al-Babili, S., Lintig, J. v, Haubruck, H., Rabbani, S., Kleinig, H., and Beyer, P. (1996). Phytoene synthase from *Narcissus pseudonarcissus*: functional expression, galactolipid requirement, topological distribution in chromoplasts and induction during flowering. *Plant J.* 10, 781–792.

Schmidhauser, T.J., Lauter, F.R., Schumacher, M., Zhou, W., Russo, V.E., and Yanofsky, C. (1994). Characterization of *al-2*, the phytoene synthase gene of *Neurospora crassa*. Cloning, sequence analysis, and photoregulation. *J. Biol. Chem.* 269, 12060–12066.

Schmidt, S.R. (2013). Fusion proteins as biopharmaceuticals – Applications and challenges. In *Fusion Protein Technologies for Biopharmaceuticals*, (Wiley-Blackwell), pp. 1–24.

- Schmidt, I., Schewe, H., Gassel, S., Jin, C., Buckingham, J., Hümbelin, M., Sandmann, G., and Schrader, J. (2011). Biotechnological production of astaxanthin with *Phaffia rhodozyma/Xanthophyllomyces dendrorhous*. Appl. Microbiol. Biotechnol. 89, 555–571.
- Schmidt-Dannert, C., Umeno, D., and Arnold, F.H. (2000). Molecular breeding of carotenoid biosynthetic pathways. Nat. Biotechnol. 18, 750.
- Schmitt, D.L., and An, S. (2017). Spatial Organization of Metabolic Enzyme Complexes in Cells. Biochemistry 56, 3184–3196.
- Schmitz Alexander C., Hartline Christopher J., and Zhang Fuzhong (2017). Engineering Microbial Metabolite Dynamics and Heterogeneity. Biotechnol. J. 12, 1700422.
- Schückel, J., Rylott, E.L., Grogan, G., and Bruce, N.C. (2012). A Gene-Fusion Approach to Enabling Plant Cytochromes P450 for Biocatalysis. ChemBioChem 13, 2758–2763.
- Schweiggert, R.M., and Carle, R. (2016). Carotenoid Production by Bacteria, Microalgae, and Fungi. In Carotenoids, (Wiley-Blackwell), pp. 217–240.
- Shao, Z., and Zhao, H. (2012). Exploring DNA assembler, a synthetic biology tool for characterizing and engineering natural product gene clusters. Methods Enzymol. 517, 203–224.
- Shao, Z., Zhao, H., and Zhao, H. (2009). DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 37, e16.
- Shi, S., Liang, Y., Zhang, M.M., Ang, E.L., and Zhao, H. (2016). A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in *Saccharomyces cerevisiae*. Metab. Eng. 33, 19–27.
- Shi, S., Choi, Y.W., Zhao, H., Tan, M.H., and Ang, E.L. (2017). Discovery and engineering of a 1-butanol biosensor in *Saccharomyces cerevisiae*. Bioresour. Technol. 245, 1343–1351.
- Shin, J.H., Kim, H.U., Kim, D.I., and Lee, S.Y. (2013). Production of bulk chemicals via novel metabolic pathways in microorganisms. Biotechnol. Adv. 31, 925–935.
- Shumskaya, M., and Wurtzel, E.T. (2013). The carotenoid biosynthetic pathway: thinking in all dimensions. Plant Sci. Int. J. Exp. Plant Biol. 208, 58–63.
- Si, T., Chao, R., Min, Y., Wu, Y., Ren, W., and Zhao, H. (2017). Automated multiplex genome-scale engineering in yeast. Nat. Commun. 8.
- Singh, R., Lanchantin, J., Robins, G., and Qi, Y. (2016). DeepChrome: deep-learning for predicting gene expression from histone modifications. Bioinformatics 32, i639–i648.
- Siu, K.-H., Chen, R.P., Sun, Q., Chen, L., Tsai, S.-L., and Chen, W. (2015). Synthetic scaffolds for pathway enhancement. Curr. Opin. Biotechnol. 36, 98–106.

Skjoedt, M.L., Snoek, T., Kildegaard, K.R., Arsovska, D., Eichenberger, M., Goedecke, T.J., Rajkumar, A.S., Zhang, J., Kristensen, M., Lehka, B.J., et al. (2016). Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nat. Chem. Biol.* 12, 951.

Sleight, S.C., Bartley, B.A., Lieviant, J.A., and Sauro, H.M. (2010). In-Fusion BioBrick assembly and re-engineering. *Nucleic Acids Res.* 38, 2624–2636.

Soma, Y., Tsuruno, K., Wada, M., Yokota, A., and Hanai, T. (2014). Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch. *Metab. Eng.* 23, 175–184.

Spitzer, J. (2011). From Water and Ions to Crowded Biomacromolecules: In Vivo Structuring of a Prokaryotic Cell. *Microbiol. Mol. Biol. Rev.* MMBR 75, 491–506.

Srere, P.A. (1987). Complexes of Sequential Metabolic Enzymes. *Annu. Rev. Biochem.* 56, 89–124.

Stahl, W., and Sies, H. (2003). Antioxidant activity of carotenoids. *Mol. Aspects Med.* 24, 345–351.

Stamellos, K.D., Shackelford, J.E., Shechter, I., Jiang, G., Conrad, D., Keller, G.A., and Krisans, S.K. (1993). Subcellular localization of squalene synthase in rat hepatic cells. Biochemical and immunochemical evidence. *J. Biol. Chem.* 268, 12825–12836.

Stephanopoulos, G. (2012). Synthetic Biology and Metabolic Engineering. *ACS Synth. Biol.* 1, 514–525.

Stephanopoulos, G., Aristidou, A.A., and Nielsen, J. (1998). *Metabolic Engineering: Principles and Methodologies* (Elsevier).

Sumegi, B., Sherry, A.D., and Malloy, C.R. (1990). Channeling of TCA cycle intermediates in cultured *Saccharomyces cerevisiae*. *Biochemistry* 29, 9106–9110.

Sun, J., and Alper, H.S. (2015). Metabolic engineering of strains: from industrial-scale to lab-scale chemical production. *J. Ind. Microbiol. Biotechnol.* 42, 423–436.

Sweetlove, L.J., and Fernie, A.R. (2013). The Spatial Organization of Metabolism Within the Plant Cell. *Annu. Rev. Plant Biol.* 64, 723–746.

Takaichi, S. (2000). Characterization of carotenes in a combination of a C18 HPLC column with isocratic elution and absorption spectra with a photodiode-array detector. *Photosynth. Res.* 65, 93–99.

Tansey, T.R., and Shechter, I. (2001). Squalene synthase: structure and regulation. *Prog. Nucleic Acid Res. Mol. Biol.* 65, 157–195.

Thodey, K., Galanie, S., and Smolke, C.D. (2014). A microbial biomanufacturing platform for natural and semi-synthetic opiates. *Nat. Chem. Biol.* 10, 837–844.

Thompson, J.F., Danley, D.E., Mazzalupo, S., Milos, P.M., Lira, M.E., and Harwood, H.J. (1998). Truncation of Human Squalene Synthase Yields Active, Crystallizable Protein. *Arch. Biochem. Biophys.* 350, 283–290.

Throop, A.L., and LaBaer, J. (2015). Recombinational Cloning Using Gateway and In-Fusion Cloning Schemes. *Curr. Protoc. Mol. Biol.* Ed. Frederick M Ausubel AI 110, 3.20.1-3.20.23.

Tippmann Stefan, Chen Yun, Siewers Verena, and Nielsen Jens (2013). From flavors and pharmaceuticals to advanced biofuels: Production of isoprenoids in *Saccharomyces cerevisiae*. *Biotechnol. J.* 8, 1435–1444.

Tsai, S.-L., Oh, J., Singh, S., Chen, R., and Chen, W. (2009). Functional Assembly of Minicellulosomes on the *Saccharomyces cerevisiae* Cell Surface for Cellulose Hydrolysis and Ethanol Production. *Appl. Environ. Microbiol.* 75, 6087–6093.

Tsai, S.-L., A Dasilva, N., and Chen, W. (2013). Functional Display of Complex Cellulosomes on the Yeast Surface via Adaptive Assembly. *ACS Synth. Biol.* 2, 14–21.

Umeno, D., Tobias, A.V., and Arnold, F.H. (2002). Evolution of the C30 carotenoid synthase CrtM for function in a C40 pathway. *J. Bacteriol.* 184, 6690–6699.

Van Dien, S. (2013). From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals. *Curr. Opin. Biotechnol.* 24, 1061–1068.

Velayos, A., Eslava, A.P., and Iturriaga, E.A. (2000). A bifunctional enzyme with lycopene cyclase and phytoene synthase activities is encoded by the carRP gene of *Mucor circinelloides*. *Eur. J. Biochem.* 267, 5509–5519.

Venayak, N., Anesiadis, N., Cluett, W.R., and Mahadevan, R. (2015). Engineering metabolism through dynamic control. *Curr. Opin. Biotechnol.* 34, 142–152.

Verdoes (1999). Cloning and characterization of the astaxanthin biosynthetic gene encoding phytoene desaturase of *Xanthophyllomyces dendrorhous* - Verdoes - 1999 - *Biotechnology and Bioengineering* - Wiley Online Library.

Verdoes, J.C., Krubasik, K.P., Sandmann, G., and van Ooyen, A.J. (1999). Isolation and functional characterisation of a novel type of carotenoid biosynthetic gene from *Xanthophyllomyces dendrorhous*. *Mol. Gen. Genet.* MGG 262, 453–461.

Verwaal, R., Wang, J., Meijnen, J.-P., Visser, H., Sandmann, G., van den Berg, J.A., and van Ooyen, A.J.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.

Wang, C., Yoon, S.-H., Jang, H.-J., Chung, Y.-R., Kim, J.-Y., Choi, E.-S., and Kim, S.-W. (2011). Metabolic engineering of *Escherichia coli* for α -farnesene production. *Metab. Eng.* 13, 648–655.

- Wang, H., Russa, M.L., and Qi, L.S. (2016). CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. Biochem.* 85, 227–264.
- Wang, L., Dash, S., Ng, C.Y., and Maranas, C.D. (2017). A review of computational tools for design and reconstruction of metabolic pathways. *Synth. Syst. Biotechnol.* 2, 243–252.
- Wang, Y., San, K.-Y., and Bennett, G.N. (2013). Cofactor engineering for advancing chemical biotechnology. *Curr. Opin. Biotechnol.* 24, 994–999.
- Wanke, M., Skorupinska-Tudek, K., and Swiezewska, E. (2001). Isoprenoid biosynthesis via 1-deoxy-D-xylulose 5-phosphate/2-C-methyl-D-erythritol 4-phosphate (DOXP/MEP) pathway. *Acta Biochim. Pol.* 48, 663–672.
- Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE* 6, e16765.
- Wheeldon, I., Minter, S.D., Banta, S., Barton, S.C., Atanassov, P., and Sigman, M. (2016). Substrate channelling as an approach to cascade reactions. *Nat. Chem.* 8, 299–309.
- Whitaker, W.R., and Dueber, J.E. (2011). Chapter nineteen - Metabolic Pathway Flux Enhancement by Synthetic Protein Scaffolding. In *Methods in Enzymology*, C. Voigt, ed. (Academic Press), pp. 447–468.
- Wilson Miles, E. (2001). Tryptophan synthase: A multienzyme complex with an intramolecular tunnel.
- Winkler, J.D., and Kao, K.C. (2014). Recent advances in the evolutionary engineering of industrial biocatalysts. *Genomics* 104, 406–411.
- Wu, F., and Minter, S. (2015). Krebs Cycle Metabolon: Structural Evidence of Substrate Channeling Revealed by Cross-Linking and Mass Spectrometry.
- Wu, G., Yan, Q., Jones, J.A., Tang, Y.J., Fong, S.S., and Koffas, M.A.G. (2016). Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol.* 34, 652–664.
- Wu, S.G., He, L., Wang, Q., and Tang, Y.J. (2015). An ancient Chinese wisdom for metabolic engineering: Yin-Yang. *Microb. Cell Factories* 14, 39.
- Xie, W., Liu, M., Lv, X., Lu, W., Gu, J., and Yu, H. (2014). Construction of a controllable β -carotene biosynthetic pathway by decentralized assembly strategy in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 111, 125–133.
- Xie, W., Ye, L., Lv, X., Xu, H., and Yu, H. (2015). Sequential control of biosynthetic pathways for balanced utilization of metabolic intermediates in *Saccharomyces cerevisiae*. *Metab. Eng.* 28, 8–18.
- Xu, J.-W., Xu, Y.-N., and Zhong, J.-J. (2012). Enhancement of Ganoderic Acid Accumulation by Overexpression of an N-Terminally Truncated 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Gene in the Basidiomycete *Ganoderma lucidum*. *Appl. Environ. Microbiol.* 78, 7968–7976.

- Xu, P., Li, L., Zhang, F., Stephanopoulos, G., and Koffas, M. (2014). Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc. Natl. Acad. Sci. U. S. A.* 111, 11299–11304.
- Yabuzaki, J. (2017). Carotenoids Database: structures, chemical fingerprints and distribution among organisms. *Database J. Biol. Databases Curation* 2017.
- Yamano, S., Ishii, T., Nakagawa, M., Ikenaga, H., and Misawa, N. (1994). Metabolic Engineering for Production of β -Carotene and Lycopene in *Saccharomyces cerevisiae*. *Biosci. Biotechnol. Biochem.* 58, 1112–1114.
- Yeates, T.O., Crowley, C.S., and Tanaka, S. (2010). Bacterial Microcompartment Organelles: Protein Shell Structure and Evolution. *Annu. Rev. Biophys.* 39, 185–205.
- You, C., and Zhang, Y.-H.P. (2013). Self-assembly of synthetic metabolons through synthetic protein scaffolds: one-step purification, co-immobilization, and substrate channeling. *ACS Synth. Biol.* 2, 102–110.
- You, C., Myung, S., and Zhang, Y.-H.P. (2012). Facilitated Substrate Channeling in a Self-Assembled Trifunctional Enzyme Complex. *Angew. Chem. Int. Ed.* 51, 8787–8790.
- Yournon, J., Kohno, T., and Roth, J.R. (1970). Enzyme Evolution: Generation of a Bifunctional Enzyme by Fusion of Adjacent Genes. *Nature* 228, 820–824.
- Yuan, J., and Ching, C.-B. (2015). Dynamic control of ERG9 expression for improved amorpha-4,11-diene production in *Saccharomyces cerevisiae*. *Microb. Cell Factories* 14.
- Zamboni, N., Fendt, S.-M., Rühl, M., and Sauer, U. (2009). ¹³C-based metabolic flux analysis. *Nat. Protoc.* 4, 878–892.
- Zhang, Y.-H.P. (2011). Substrate channeling and enzyme complexes for biotechnological applications. *Biotechnol. Adv.* 29, 715–725.
- Zhang, T., Lei, J., Yang, H., Xu, K., Wang, R., and Zhang, Z. (2011). An improved method for whole protein extraction from yeast *Saccharomyces cerevisiae*. *Yeast Chichester Engl.* 28, 795–798.
- Zhang, Y., Li, S.-Z., Li, J., Pan, X., Cahoon, R.E., Jaworski, J.G., Wang, X., Jez, J.M., Chen, F., and Yu, O. (2006). Using Unnatural Protein Fusions to Engineer Resveratrol Biosynthesis in Yeast and Mammalian Cells. *J. Am. Chem. Soc.* 128, 13030–13031.
- Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J.M., Sachs, M.S., and Liu, Y. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* 495, 111–115.
- Zi, J., Mafu, S., and Peters, R.J. (2014). To Gibberellins and Beyond! Surveying the Evolution of (Di)terpenoid Metabolism. *Annu. Rev. Plant Biol.* 65, 259–286.

VI ANNEXE: article scientifique publié

Enzyme-fusion strategies for redirecting and improving carotenoid synthesis in *S. cerevisiae*

Hery Rabeharindranto^{a§}, Sara Castaño-Cerezo^{a§}, Thomas Lautier^a, Luis F. Garcia-Alles^a, Christian Treitz^b, Andreas Tholey^b, Gilles Truan^{a*}

^a LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France.

^b Systematic Proteome Research and Bioanalytics, Institute for Experimental Medicine, Christian-Albrechts-Universität zu Kiel, Kiel, Germany.

[§] These authors contributed equally to this work

* **Corresponding author:** gilles.truan@insa-toulouse.fr; LISBP, 135 Avenue de Rangueil, 31077 Toulouse Cedex 04, FRANCE.

1 Abstract

Spatial clustering of enzymes has proven an elegant approach to optimize metabolite transfer between enzymes in synthetic metabolic pathways. Among the multiple methods used to promote colocalisation, enzyme fusion is probably the simplest. Inspired by natural systems, we have explored the metabolic consequences of spatial reorganizations of the catalytic domains of *Xanthophyllomyces dendrorhous* carotenoid enzymes produced in *Saccharomyces cerevisiae*. Synthetic genes encoding bidomain enzymes composed of CrtI and CrtB domains from the natural CrtYB fusion were connected in the two possible orientations, using natural and synthetic linkers. A tridomain enzyme (CrtB, CrtI, CrtY) harboring the full β -carotene producing pathway was also constructed. Our results demonstrate that domain order and linker properties considerably impact both the expression and/or stability of the constructed proteins and the functionality of the catalytic domains, all concurring to either diminish or boost specific enzymatic steps of the metabolic pathway. Remarkably, the yield of β -carotene production doubled with the tridomain fusion while precursor accumulation decreased, leading to an improvement of the pathway efficiency, when compared to the natural system. Our data strengthen the idea that fusion of enzymatic domains is an appropriate technique not only to achieve spatial confinement and enhance the metabolic flux but also to produce molecules not easily attainable with natural enzymatic configurations, even with membrane bound enzymes.

Keywords: Metabolic engineering, synthetic biology, metabolic flux, enzyme spatial proximity, carotenoids, multidomain enzymes.

2 Introduction

Metabolic engineering aims to design and implement metabolic pathways to produce high value compounds or novel molecules that would otherwise be more expensive or difficult to attain following chemical organic synthesis (Chen and Nielsen, 2013). However, product yields in engineered microorganisms are often low. It is therefore important to devise strategies optimizing the efficiency of both individual enzymatic steps and the global synthetic pathway (Siddiqui et al., 2012). A major drawback originates from the accumulation of metabolic intermediates. This is particularly detrimental to the cell, not only because of the loss of metabolites and non-optimized consumption of resources but also because of the possible consumption of those intermediates by other competitive metabolic pathways, potentially leading to toxicity (Dahl et al., 2013).

Spatial proximity and appropriate disposition of sequential enzymatic domains inside the metabolic pathway are assumed to contribute critically in promoting rapid and efficient consumption of pathway intermediates (Srere, 1985). In natural enzymatic systems, spatial confinement is achieved using various strategies such as the creation of enzymes combining several covalently-linked catalytic domains, the recruitment of several enzymes on protein scaffolds or the colocalization of activities within lipid- or protein-based compartments (Pompon et al., 2017; Schmitt and An, 2017). When applied to metabolic engineering, several studies have demonstrated that the efficiency of a synthetic pathway can be improved via spatial confinement. For example in *E. coli*, a 77-fold improvement in product titer was achieved by promoting artificial interactions between the three first enzymes of the mevalonate pathway, using SH3, PDZ and GBD binding domains as scaffold and their respective ligands as recruiting domains (Dueber et al., 2009). Synthetic RNAs were also used as macromolecular scaffolds to recruit and promote spatial confinement of ferredoxin and hydrogenase enzymes, leading to a 24-fold increase in the hydrogen production compared to the strain expressing free diffusing enzymes (Delebecque et al., 2011). In *Saccharomyces cerevisiae*, a 38% improvement in 2,3-butanediol was achieved by scaffolding enzymes onto cohesin and dockerin domains (Kim et al., 2016)

and a three times increase in the production of geranylgeraniol was detected upon overexpression of a synthetic bidomain enzyme (Tokuhiro et al., 2009). Similarly, by connecting farnesyl pyrophosphate (FPP) synthase with amorphaadiene and patchoulol synthases from plant origin, the biotransformation of the intermediate molecule FPP into the desired molecules of interest was improved (Albertsen et al., 2011; Baadhe et al., 2013). Other strategies promoted the use of natural compartment such as mitochondria to improve the production of branched alcohols in yeast (Avalos et al., 2013).

Carotenoids, a particular group of terpene-based molecules, have raised scientific attention due to their potential benefit for human health (Lu et al., 2014) and in several industrial applications (López et al., 2015). Biosynthesis of carotenoids begins with the production of phytoene from geranylgeranyl pyrophosphate (GGPP) by phytoene synthase (CrtB) (Figure 1). Phytoene desaturase (CrtI) then catalyzes four successive desaturation reactions to form lycopene. Finally, lycopene cyclase (CrtY) performs the cyclisation on both sides of lycopene to produce β -carotene. Eukaryotic CrtY enzymes are predicted to be transmembrane proteins. On the contrary, CrtB and possibly CrtI are supposedly cytosolic (Schaub et al., 2012). Importantly, in some fungi, the phytoene synthase and lycopene cyclase catalytic activities are regrouped within a multidomain protein named CrtYB in *Xanthophyllomyces dendrorhous* (Verdoes et al., 1999) and CarRA in *Phycomyces blakesleeanus* (Arrach et al., 2001).

The full β -carotene producing enzymatic system from *X. dendrorhous* has been expressed successfully in *S. cerevisiae* (Verwaal et al., 2007). Similarly, the natural tridomain fusion (CrtIBY) from *Schizotrium* sp. was expressed recently in *Yarrowia lipolytica* (Gao et al., 2017) but the β -carotene production yields were considerably lower than those obtained with the yeast-expressed *X. dendrorhous* configuration (Verwaal et al., 2007; Xie et al., 2015c). Although these heterologous systems, based on *X. dendrorhous* enzymes, are the most productive to date, accumulation of β -carotene precursors in the pathway, such as phytoene, was evidenced. Moreover, attempts to alleviate the presumed metabolic bottleneck by overexpression of CrtI was only partially efficient (Verwaal et al., 2007).

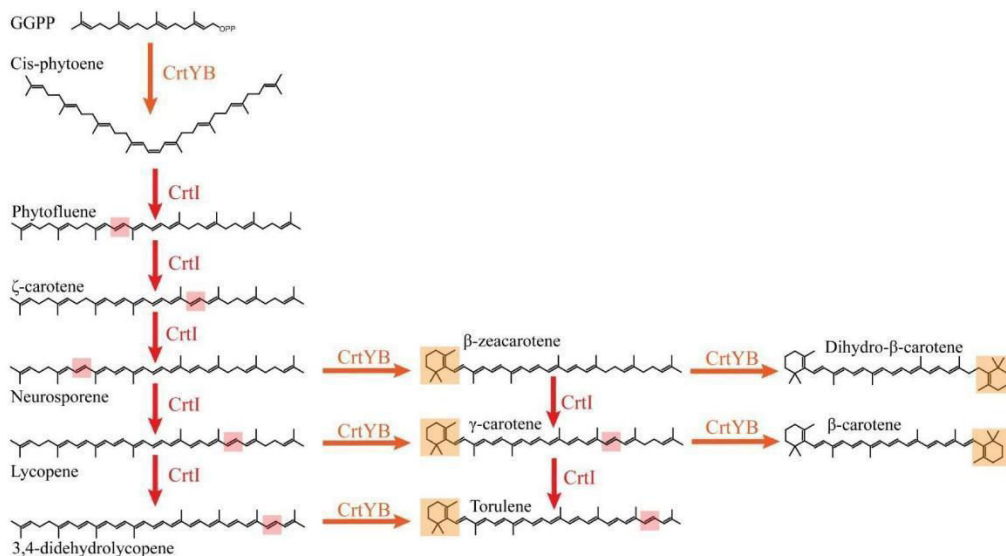


Figure 1. Presumed biosynthetic pathway for the synthesis of β -carotene in *X. dendrorhous* (Verdoes et al., 1999). The bifunctional lycopene cyclase/phytoene synthase (CrtYB) enzyme is depicted in orange, the phytoene desaturase enzyme (CrtI) is depicted in red. Orange or red boxes represent the localization of the modifications introduced by the CrtYB or CrtI enzymes respectively.

Several studies have demonstrated that the efficiency of colocalisation of metabolic reactions with membrane bound enzymes or transporters associated to cytosolic enzymes can be improved using synthetic protein scaffolds to recruit the various enzymes of the biosynthetic pathway (Lin et al., 2017; Somasundaram et al., 2017; Thomik et al., 2017). To test how spatial reorganization of the carotenoid enzymes could reduce the accumulation of intermediates and reorient the metabolic fluxes toward β -carotene production, we engineered a series of bifusion enzymes using the CrtB domain or the full CrtYB from *X. dendrorhous* connected to CrtI by means of various peptidic linkers providing variable spatial proximity and flexibility between domains. We then investigated how the relative orientation of interconnected domains and the properties of the linker (size, secondary structure, rigidity) influenced both the yield and pattern of product formation and, based on these results, we engineered a tridomain fusion that improved carotenoid production efficiency.

3 Materials and Methods

3.1 Plasmid construction

All the plasmids used in this study are presented in Supplementary Table S1. DNA Sequences for *CRTYB* and *CRTI* encoding genes were obtained by amplification of the open reading frames from the plasmid pYEplac195 *CRTYB/CRTI/CRTE* (Verwaal et al., 2007). For the construction of the different fusion proteins, the different linker sequences were obtained via commercial gene synthesis (GeneCust, Luxemburg). *CRTB* encoding gene was obtained via gene synthesis using codon optimization except for constructs using *CRTYB* where the natural sequence was used. All natural or recoded sequences used are presented in Supplementary Data S1, the linker sequences in Supplementary Data S2. The various reconstructed enzymes were expressed from the dual *GAL1-GAL10* promoters present in the pMRI31 and pMRI32 vectors (Xie et al., 2014), kindly provided by Prof. Hongwei Yu. Open reading frames encoding the various domains were cloned either at the N- or the C-terminus of one linker by classical restriction-ligation or isothermal assembly. All plasmid constructions and modified sequences were verified by sequencing. *E. coli* DH5 α strain was used for all molecular biology protocols except for In-Fusion (Takara Bio USA, Mountain View, CA, USA) experiments for which Stellar competent cells were used according to the fabricant recommendations.

3.2 Strain construction

The strain BY4741 was used to construct the various strains bearing the fusion proteins. The first step was the disruption of the *S. cerevisiae* *GAL80* gene to alleviate the repression of glucose on the *GAL1-10* promoter obtaining strain yHR22. As carotenoid synthesis in *S. cerevisiae* is highly dependent on the level of GGPP (Misawa, 2011), a supplementary of the more efficient GGPP synthase (*CRTE*) of *X. dendrorhous* and the truncated form of *HMG1* were both expressed under the control of the p*GAL1-10*, obtaining the strain yHR21. All subsequent strains are derivatives of yHR21 and are presented in the Supplementary Table S2. The pMRI integrative plasmids, bearing the same dual promoter (*GAL1-GAL10*), were used for integration (Xie et al., 2014), avoiding as much as

possible transcriptional effects due to plasmid copy number variation or transcription interference (Carquet et al., 2015). The following strain naming was also adopted for clarity: each catalytic domain is represented by a capital letter (Y, B and I), each linker by a parenthesis defining the linker type (yb, cpr, ek and gs), the absence of linker by the absence of parenthesis and a / sign when the catalytic domain is expressed separately. As an example, the *S. cerevisiae* strain bearing the natural carotenoid system from *X. dendrorhous* is notated Y(yb)B/I.

The integration of the different constructions was performed using the high efficiency Gietz transformation method (Schiestl and Gietz, 1989) with some minor modifications: 1 µg of linearized DNA fragment digested by *Sfi*I (New England BioLabs, Ipswich, MA, USA) was added in a maximum volume of 10 µl. The mixture was incubated 30 min at 30°C after which cells were heat shocked 20 min at 42°C in a water bath. The cells were recovered during two hours in 1 mL of YPD at 30°C and plated in YPD containing G418 at a concentration of 200 µg/ml.

Integrations were verified via colony PCR using two specific pairs of oligonucleotides. Each pair is composed of one oligonucleotide directed against the introduced gene or the Kan^R marker and the other one directed against the chromosomal region of insertion. In the case when the Kan^R resistance marker needed to be removed, correct colonies were then further transformed with the pSH47 vector to eliminate the Kan^R gene by the Cre-Lox P recombination system (Güldener et al., 1996). Transformed clones were grown overnight on 5 ml SD-URA Gal medium to maintain and express the Cre enzyme and plated on SG-URA plates (approximately 100 cells per plate) to allow for the loss of the *KanMX* gene. After three days growth, clones were individually tested for their Kan^R phenotype using both YPD and YPD containing 200 µg/ml of Geneticin (G418). Clones which are not growing on YPD G418 were grown overnight on YPD medium to allow for the loss of the pSH47 plasmid. Cells from the culture were plated onto YPD medium (approximately 100 cells per plate). After three days growth, clones were individually tested for their URA⁺ phenotype using both YPD and SG-URA medium.

3.3 Media composition

YPD or YPG media (20 g/L peptone, 10 g/L yeast extract, 20 g/L glucose or galactose respectively), YPD or YPG agar plates (YPD or YPG media + 20 g/L agar) were used to grow yeast. SD-URA or SG-URA media (1.7 g/L Yeast Nitrogen Base w/o amino acid, 5 g/L ammonium sulfate, 20 g/L glucose (SD) or galactose (SG), 0.77 g/L dropout mix containing all amino acids and nucleic acids without uracil (MP Biomedicals, Santa Ana, California, USA) and SD or SG -URA plates (SD or SG -URA medium with 17 g/L agar) were used to grow or select for URA⁺ colonies. Geneticin (G418) was incorporated at 200 µg/ml in YPD media for the selection of Geneticin resistant colonies.

3.4 Carotenoid production in shake flasks

Cryoconserved strains were regrown on YPG plates. Typically, one isolated clone was used to inoculate a 5 ml preculture in YPG. After one night growth, the preculture was used to inoculate several 50 ml YPG flasks at 0.05 OD_{600nm} and further cultivated at 28°C, 140 rpm during 72 hours in an incubator INFORS HT with 50 mm shaking throw. Cells were then pelleted for 3 min at 4°C at 6000 g, washed in 2 ml of distilled water, and then transferred into a pre-weighted screw capped tube, and centrifuged again at 9300 g at 4°C for 3 min. Pelleted cells were stored at -80°C until further use.

3.5 Carotenoid extraction and analysis

The frozen cell pellets were lyophilized and resuspended in 500 µL of acetone. Apocarotenal (24 mg/L in hexane) was used as internal standard (100 µL per 0.025 g of dry cells). Extraction was performed as follows: 3 cycles of 20 s at 0.05 m/s with one minute of cooling on ice using a Fast Prep FP120 (ThermoFisher, Waltham, MA, USA). The acetone phase was transferred to a new tube and the process repeated 3 times. Acetone extracts were pooled and dried under nitrogen flux and further suspended in 500 µL of hexane for HPLC analysis.

30 µl of carotenoids extract were injected on a Waters alliance HPLC system. Separation was done as follows: an isocratic gradient composed of isopropanol and acetonitrile/water (90:10, v:v) was applied for 49 min at a flow-rate of 0.3 ml/min into a 100 x 2.1 mm Brownlee RP18 column at 35°C. Absorbance spectra were recorded between 200 nm and 600 nm using a 996 Waters PDA detector. Peaks were characterized by their elution time and their spectra (Takaichi, 2000). Main

carotenoids peaks were manually integrated: at 282 nm (phytoene), at 475 nm (lycopene), at 454 nm (β -carotene), at 442 nm (neurosporene), 399 nm (ζ -carotene), 349 nm (phytofluene), 430 nm (dihydro- β -carotene), and 487nm (torulene). Peaks areas were used to calculate quantity based on a calibration curve performed with phytoene, lycopene and β -carotene standards solutions (Sigma-Aldrich, St. Louis, MO, USA) and neurosporene, zeta-carotene, phytofluene, β -zeacarotene and torulene (Carotenature, Münsingen, Switzerland) and the correction for the extraction yields were performed with the quantification of apocarotenal (Sigma-Aldrich, St. Louis, MO, USA). Attribution of dihydro- β -carotene peaks was done in reference to the elution order described in (Mizoguchi et al., 2015; Takaichi, 2000). As there is no available standard for dihydro- β -carotene, the quantification was performed with the β -zeacarotene standard curve as both products contain the same amount of conjugated double bonds. All other carotenoids were not quantified due to absence of commercial standard. Each extracted sample was injected and quantified twice (technical duplicate). A second set of batch cultures was also prepared and analyzed with the same protocol for duplication (biological duplicate). Error bars represent the standard deviation obtained from these different values.

3.6 Western blot analysis

Protein extracts were prepared following the protocol described by Zhang et al (Zhang et al., 2011). Briefly, pelleted cells from one 50 ml flask were resuspended in 1 ml PBS (Invitrogen, Carlsbad, CA, USA) by agitation. To normalize the quantity of cell before extraction, cell density was verified by OD₆₀₀ measurements. 8 μ L of the cell suspension were gently mixed on ice with 200 μ L of a 2 M lithium acetate cold solution, and left to stand for 5 min, followed by 5 min centrifugation at 5000 g, 4°C. The supernatant was discarded and 200 μ L of a 0.4 M solution of NaOH added. After gentle resuspension, and 5 min standing on ice, samples were centrifuged 5 min at 4°C. After discarding supernatants, pellets were vigorously vortexed with 100 μ L of bromophenol Blue loading dye solution supplemented with 5% β -mercaptoethanol. After denaturation for 5 min at 90°C, 15 μ L of each sample was deposited on a 10% SDS page gel. Semi-dry transfer were performed on PVDF membrane (Merck Millipore, Darmstadt, Germany) using a Trans-Blot® SD Cell BioRad apparatus (18V at 600 mA during 25 min). 5% bovine milk in TBS was used as blocking agent. Incubations with primary anti-Flag

(ThermoFischer, Waltham, MA, USA), mouse anti-c-Myc (Invitrogen, Carlsbad, CA, USA), and secondary anti-mouse IgG coupled with alkaline phosphatase (ThermoScientific, Rockford, IL, USA), diluted following instructions from the provider, lasted one hour. Proteins were detected by the incubation of BCIP/NBT AP substrate buffer (Sigma-Aldrich, St. Louis, MO, USA).

3.7 Protein analysis by GelLC-MS

Crude cell lysates were diluted with reducing Laemmli buffer (Sambrook and Russell, 2006) to a concentration of 3.3 $\mu\text{g}/\mu\text{L}$ and SDS-PAGE was performed on a 4-15 % precast gradient gel (Mini-PROTEAN TGX gel; Bio-rad, Munich, Germany). Gels were stained by Coomassie and protein bands were excised for subsequent in-gel digestion as indicated on the gel-scan (Supplementary Figure S1).

After reduction (20 mM dithiothreitol) and alkylation (55 mM chloroacetamide), proteins were digested with 0.1 μg of sequencing grade modified Trypsin (Promega, Mannheim, Germany) in 0.1 M triethylammonium bicarbonate buffer (TEAB) for 16h at 37 C. Peptides were extracted, the solvent was lyophilized to dryness and reconstituted in 20 μL of HPLC loading buffer (3% acetonitrile, 0.05% trifluoroacetic acid in water). Peptide extracts were analyzed by LC-MS on a Dionex U3000 uHPLC system coupled to a Q Exactive Plus mass spectrometer (Thermo Scientific, Dreieich, Germany). A volume of 4 μL of the samples were injected on a C18 PepMap 100 μm -precursor column (column dimensions: 300 μm i.d. x 5 mm) before separated on an Acclaim PepMap RSLC column (column dimension: 75 μm i.d. x 50 cm; Thermo Scientific). A binary gradient of eluent A (0.05% aqueous formic acid) and eluent B (80% acetonitrile, 0.04% formic acid) with a flow rate of 0.3 $\mu\text{L}/\text{min}$ was used. Peptides were eluted over a gradient from 5 % eluent B to 50 % eluent B in 70 min, followed by an increased to 95 % eluent B in 6 min. After isocratic elution at 90 % eluent B for 10 min, the column was equilibrated for 10 min with 5 % eluent B. The LC-system was directly coupled to the Q Exactive Plus mass spectrometer. Full MS scans were acquired in positive ion mode with a resolution of 35,000, with a scan range from 240 m/z to 1800 m/z. Data dependent MS/MS spectra of the ten most intense precursor ions were acquired with a resolution of 17,500; Scan parameters were set to an isolation window of 1.6 m/z, a normalized collision energy of 27. Precursors with a charge states < 2 and > 6 were excluded and precursors were excluded from subsequent isolation for 10 s. MS raw files were searched against a

database including *Saccharomyces cerevisiae* (strain CEN.PK113-7D; 5,439 UniProt entries, 10.2018) in addition to the target proteins and common contaminants using the Sequest search algorithm and the Proteome discoverer software (Thermo Scientific). Searches with a full-tryptic and semi-tryptic protease specificity, allowing for variable oxidation of methionine and N-terminal acetylation, were combined. Peptide spectrum matches and protein identifications were performed with a false discovery rate below 1%. The complete list of identified proteins is provided as a Supplementary Excel file S1.

3.8 Sequence alignments

Alignments were performed with the Muscle software and with the sequence identification numbers boxed in parentheses from the Uniprot database. The names and Uniprot codes for the different aligned proteins are the following: C.t., *Chloracidobacterium thermophilum* (G2LGW3); M.sp, *Mycobacterium sp* (A0A1A0Q6A2); N.sp., *Nanosalina sp.* (G0QG73); C.f., *Cystobacter fuscus* (S9QG19); P.a., *Pantoea ananas* (P21687); A.n., *Aspergillus niger* (A2QM49); A.o., *Aspergillus oryzae* (Q2U4X9); A.g., *Arthroderma gypseum* (E4UPP6); *Neurospora crassa* (P37295); *Phaffia rhodozyma* (Q7Z859); *Podospora anserina* (B2ATB0); *Pyrenophora tritici-repentis* (B2WAQ3); *Phaeosphaeria nodorum* (Q0V6M5); *Pantoea ananas* (P21683); *Franconibacter pulveris* (C7C5F2); *Erwinia gerundensis* (A0A0U5L0T8); *Cronobacter turicensis* (C9Y0F0); *Enterobacter agglomerans* (Q47846).

4 Results and discussion

4.1 Design and construction of the various metabolic systems

4.1.1 Defining CrtB and CrtY domains

To define the limits of the CrtB and CrtY domains within CrtYB, we compared 57 bifunctional lycopene cyclase/phytoene synthase enzymes with 35 sequences of phytoene synthases and 55 sequences of single domain lycopene cyclases (Supplementary Figure S2). This alignment evidences a large gap between the end of the various single domains CrtY enzymes and the beginning of the various CrtB enzymes and is filled, in all bifunctional enzymes, with different linker sequences (Figure

2). As seen in Figure 2 (lower panel), the C-terminal segment of the CrtY domain of monofunctional lycopene cyclases and bifunctional CrtYB enzymes seems well conserved up to residue 254 (numbering refers to CrtYB from *X. dendrorhous*) where the homology starts decreasing between homologous enzymes. Furthermore, the alignment of CrtYB with monofunctional phytoene synthases confirmed that the residues of the linker have no equivalent in single domain CrtB or CrtY enzymes. The linker region seems to possess between 20 and 25 residues depending on the species considered, leading to a start of the CrtB domain around position 290. The end of the CrtY and the beginning of the CrtB domains were chosen with extra residues on both sides to avoid possible incorrect delimitation and were thus defined at the residue G260 (end of CrtY) and at the residue S286 (beginning of CrtB). The CrtY domain, when expressed as a single domain, contained the full linker sequence to avoid, as much as possible, any impact on activity.

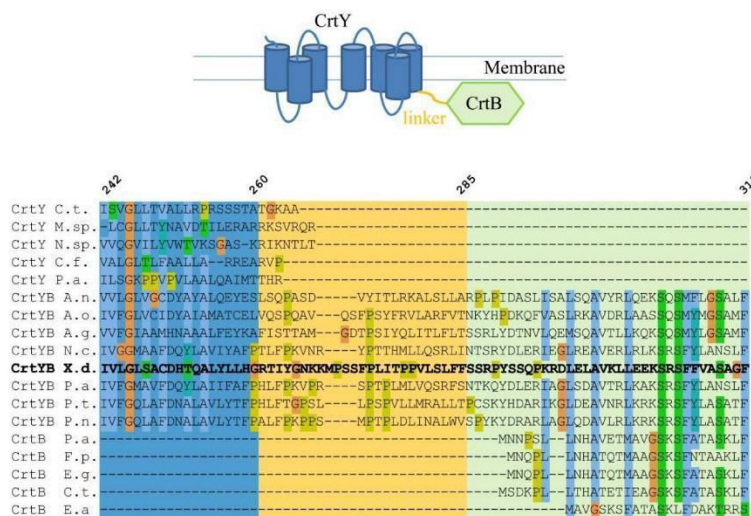


Figure 2. Localization and sequence of the linker of CrtYB from *X. dendrorhous*. Upper panel: graphical sketch of the putative structural organization of CrtYB in the membrane. The lycopene cyclase, phytoene synthase and linker domains are represented in blue, green and yellow, respectively. Eukaryotic lycopene cyclases enzymes are predicted to have seven transmembrane helices (<http://www.uniprot.org/uniprot/Q7Z859>) while phytoene synthases do not have any transmembrane segment (Schaub et al., 2012). Lower panel: partial alignment

between the CrtYB enzyme from *Xanthophyllomyces dendrorhous* (X.d.) (in bold) with other bifunctional phytoene synthases/lycopene synthases and monofunctional phytoene synthases or lycopene cyclases (See Materials and Methods).

4.1.2 Linkers properties and strains constructions

The combinatorial assembly of the domains of the 3 enzymes involved in the carotene synthesis was performed using either no linker or variable linkers. The choice of the linkers was dictated mostly to cover various physicochemical properties: size and dynamics. The longest linker used is the original linker from CrtYB, as defined previously (27 residues). Although there is no information on its dynamical properties, we expected this linker to be well tolerated when fused to the CrtB domain. We then selected a series of shorter linkers for which dynamical properties are well known. The second linker corresponds to the short segment (18 residues) connecting the FMN and FAD/NADPH domains of the human cytochrome P450 reductase enzyme (CPR) and for which high flexibility has been described by SAXS and NMR (Frances et al., 2015). The third and fourth linkers are short (20-residues long segments) and are completely synthetic sequences (Arai et al., 2004) that behave as flexible (four repetition of the GGGGS motif) or rigid α -helical connectors (four repetitions of the EAAAK motif). In the rest of the article, the linker in each construction can be inferred from the lower-case letters between the parenthesis: yb for the CrtYB linker, cpr for the connecting domain of CPR, ek and gs for the rigid or flexible small synthetic linkers corresponding to (EAAAK)₄ or (GGGGS)₄ respectively, the absence of linker by the absence of parenthesis. Furthermore, the name of the enzymatic domains was represented by a capital letter (Y for CrtY for example). A slash sign between catalytic domains indicated that domains on both sides of the sign were expressed as single domains. As an example, the *S. cerevisiae* strain bearing the natural carotenoid system from *X. dendrorhous* is notated Y(yb)B/I, corresponding to the production of CrtYB and CrtI.

	Strain name	Domain 1	Linker	Domain 2	Linker	Domain 3
Single domain	B	CrtB				
	B/I	CrtB		CrtI		

	B/I/Y	CrtB		CrtI		CrtY
Bidomain	BI	CrtB	-	CrtI		
	B(yb)I	CrtB	(yb)	CrtI		
	B(cpr)I	CrtB	(cpr)	CrtI		
	B(gs)I	CrtB	(gs)	CrtI		
	B(ek)I	CrtB	(ek)	CrtI		
	IB	CrtI	-	CrtB		
	I(yb)B	CrtI	(yb)	CrtB		
	I(cpr)B	CrtI	(cpr)	CrtB		
	I(gs)B	CrtI	(gs)	CrtB		
	I(ek)B	CrtI	(ek)	CrtB		
	Y(yb)B	CrtY	(yb)	CrtB		
	Y(yb)B/I	CrtY	(yb)	CrtB		CrtI
Tridomain	Y(yb)B(ek)I	CrtY	(yb)	CrtB	(ek)	CrtI

Table 1. Enzymatic systems used in this study. Columns after the strain names describe which domains (grey cells) and linkers (dark cells) are present in the strains. CrtY: lycopene cyclase; CrtB: phytoene synthase; CrtI: phytoene desaturase. No sign in the linker columns depicts expression of individual domains (white cell). The – sign represents a fusion without any linker present; (yb), (cpr), (gs) and (ek) represent the different linkers.

Sixteen different yeast strains containing combinations of the individual domains CrtB, CrtI and CrtY, bidomain or tridomain enzymes were constructed (Table 1). In the tridomain fusion, only the (ek) linker was tested and this choice was partly dictated by preliminary results obtained with the various bidomain enzymes (*vide infra*).

4.2 Enzyme detection by western blots in the various strains

Before analyzing the carotenoid content of the various strains, we first assessed the level of protein production of the different fusions and/or single domains. Figure 3 presents the western blots obtained for the full protein extracts of the strains in the same culture conditions used for carotenoids quantification, i.e. at 72 hours of culture. All single, double or triple domain constructs were tagged with the c-Myc peptide at the C-terminus, with the exception of the CrtY protein (when expressed alone) that carried a FLAG tag instead.

We first analyzed the production of single domain enzymes (Figure 3A). CrtB was invariably detected as intense bands, indicating a good expression in the strains B, B/I or Y/B/I. On the contrary, the CrtI protein was found at the expected size (67 kDa) only when coexpressed together with CrtY domain (strain B/I/Y). The presence of a faint band, with 50 kDa apparent molecular weight, in the B/I strain might indicate the presence of a truncated form of CrtI, reinforcing the idea of low production or stability of CrtI in the absence of CrtY. With regard to the CrtY single domain (33 kDa), its presence could not be revealed using our extractions conditions, suggesting an inadequate transfer of this integral membrane protein during blotting. Stability defects caused by the absence of the CrtB domain on the transmembrane portion are less likely, as suggested by the production of β -carotene by this construct (see below). Other constructs carrying the CrtY domain tagged with c-Myc peptide, and connected to a soluble domain (B), could be successfully detected (see below). Hence, independently of the precise reason why CrtY alone is not detected in our western blots, its expression can be inferred from the detection of the lycopene cyclase activity (*vide infra*) and also indirectly by the abovementioned effect on CrtI expression.

Expression of the natural CrtYB was well detected (Figure 3B), although with a relatively lower size (65 kDa, as compared to 75 kDa theoretical value). This sort of discrepancies are commonly observed when considering the SDS-PAGE migration pattern of integral membrane proteins (Rath et al., 2009). This also led to the non-resolution of the two individual bands when CrtI (67 kDa) was coexpressed with CrtYB. Since CrtI was never well detected, the band at 65 kDa was cautiously assigned to CrtYB. Interestingly, the intensity of the 65 kDa band decreased when CrtI and CrtYB were coexpressed together, suggesting that CrtYB level is also reduced in Y(yb)B/I as compared to Y(yb)B. Again, a possible truncation product of CrtI was faintly detected in the Y(yb)B/I strain. As will be seen further, the expression of functional CrtI in this strain will be supported by its activity.

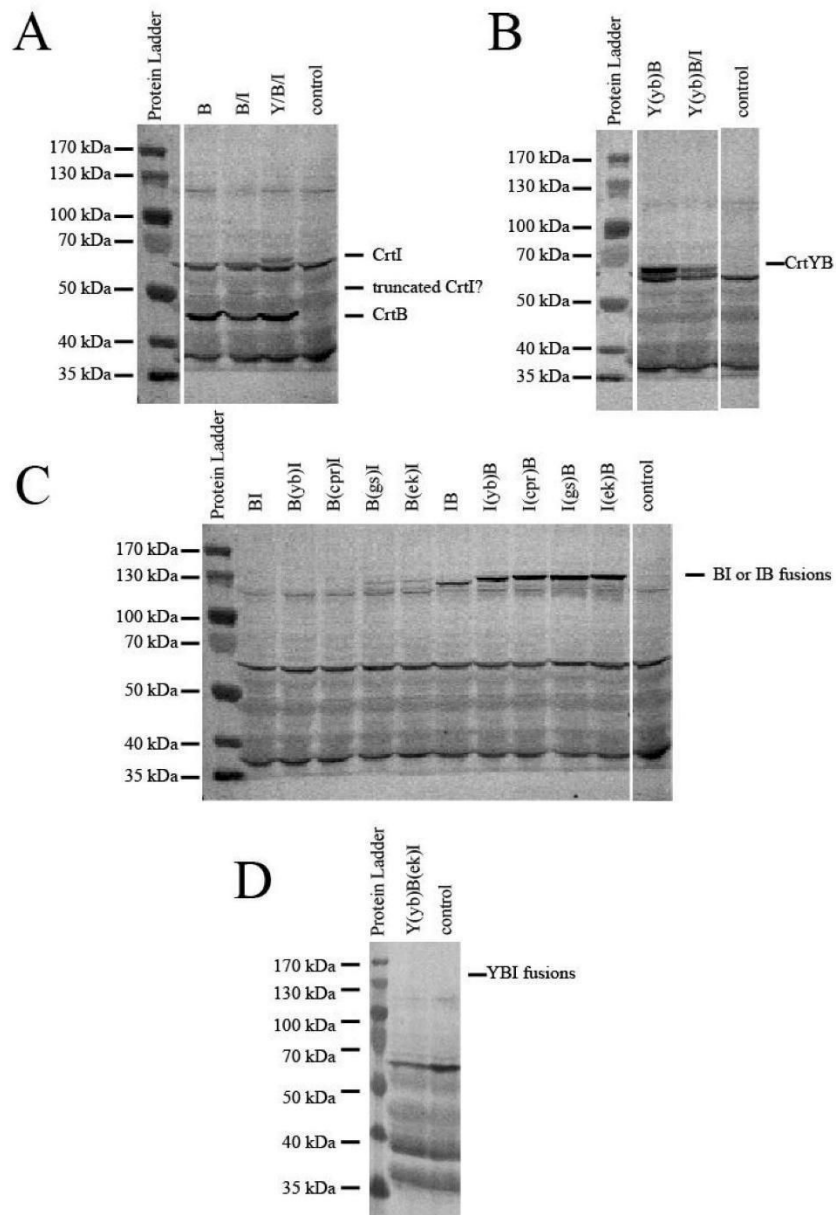


Figure 3. Expression of the different constructs in crude extracts detected in western blots. Loaded samples correspond to material withdrawn after 72 h of culture. The void space between lanes represents spatial arrangements of different lanes from the same blot. Control lanes are from extracts from the strain engineered

for GGPP overproduction but devoid of any carotenoid enzyme (yHR21). All components were blotted with antibody that detect specifically C-ter c-myc peptides, exception made of the single Y domain (third lane of panel A), which was detected by an antiFLAG antibody recognizing its C-ter FLAG peptide. **A.** Strains expressing monomeric enzymes. **B.** Strains expressing Y(yb)B based constructions. **C.** Strains with bifunctional synthetic enzymes. **D.** Strain expressing the trifunctional enzyme.

Bands for bidomain constructs with the CrtB domain connected at the N-ter of the CrtI domain (B to I orientation) were undistinguishable from those present in the control lane (compare lanes 2-6 with last one in Figure 3C). Specific protein bands only became visible for the fusions bearing the linkers (gs) and (ek) but still remained faint (lanes 5,6), suggesting that the B to I orientation is likely deleterious for expression. On the contrary, bidomain enzymes with the I to B orientation were perfectly detectable (lanes 7-11). As with the B to I constructs, the (gs) and (ek) linkers appeared most benefic for promoting higher amounts of fusions proteins (I(ek)B and I(gs)B). Remarkably, the different linkers induced apparent size shifts, in spite of their relatively small size (27 residues for the longest one). Overall, the results of the bidomain fusions expression indicate that the N- or C-terminal positioning of CrtI or CrtB catalytic domains strongly affects the protein production level. Expression of the tridomain fusion construct (strain Y(yb)B(ek)I) was also investigated without success. Different lytic and solubilization protocols were applied, but all efforts to establish its presence by western blot failed (Figure 3D).

The strong influence of domain orientation and linker composition on the protein expression level or the translocation through the periplasmic membrane of *E. coli* has already been described in the case of single chain antibodies composed of the fusion of light and heavy chains (Anand et al., 1991; Tsumoto et al., 1994). The oligomeric assembly and membrane association of the CrtI enzyme from *Pantoea ananatis* and *Oryza sativa* has been well studied, providing evidences that i), CrtI is associated to the membrane via peptidic regions that are not membrane spanning segments and, ii), that the N-terminus of CrtI likely plays a major role in directing the correct folding of the enzyme as well as its oligomerization (Gemmecker et al., 2015; Schaub et al., 2012). The fact that the I to B orientation is preferred may indicate that membrane interaction or oligomerization is indeed located

in the N-terminus of the CrtI domain and this N-terminal part, when placed between the CrtB and CrtI enzymes, might not properly promote its role in membrane interaction or oligomerization, reducing the amount of correctly folded, membrane localized CrtI and thus leading to its degradation within cells.

Overall, western blot analysis allowed us to define relative expression efficiencies for the various expression systems. Although semi-quantitative, such information is important to better understand some of the patterns of carotenoid molecules production.

4.3 Detection of the natural Y(yb)B/I and the Y(yb)B(ek)I trifusion by GeLC-MS

The unsuccessful detection of trifusion protein by western blot prompted us to perform shotgun proteomic analyses using gel electrophoresis followed by in-gel digestion and liquid chromatography coupled to mass spectrometry of the extracted peptides (GeLC-MS). After separating by SDS-PAGE the total protein extracts of yHR21, Y(yb)B/I and Y(yb)B(ek)I strains, gel regions around the MW of CrtYB(ek)I (145 kDa), CrtYB (76 kDa, but migrating as a 65 kDa protein (Figure 3B)) and CrtI (67 kDa) were treated in search for specific peptides of the Y, (yb), B, (ek) or I domains (Figure 4 and Supplementary Figure S1). A gel region around 40 kDa MW was also analyzed in both yHR21 and Y(yb)B(ek)I strains and peptides specific for CrtE were detected with high sequence coverage. A summary of all identified proteins including details for peptide evidence for the target proteins is presented in Supplementary Excel File S1.

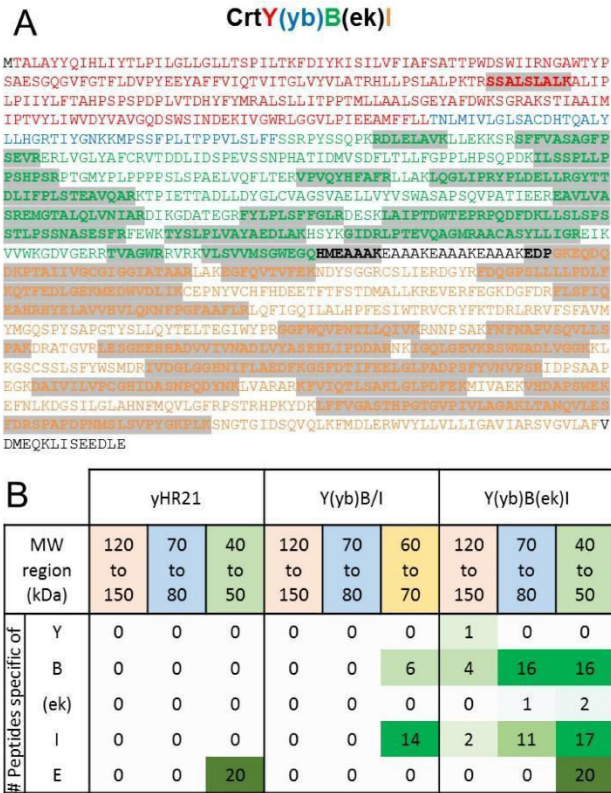


Figure 4. Detection of CrtY(yb)B(ek)I peptides by GeLC-MS analysis. **A.** Amino acid sequence of the CrtY(yb)B(ek)I construct. The different domains of the fusion protein are color coded: Y (red), (yb) (blue), B (green), (ek) (black), I (orange). Peptides that could be unequivocally identified are highlighted. Total sequence coverage by identified peptides for the fusion protein CrtY(yb)B(ek)I was 43%. **B.** Summary of GeLC-MS data. Numbers correspond to peptide hits identified for a given domain (horizontal lanes) within a given gel portion (MW region) which was selected as corresponding to the expected electrophoretic mobility for a given construct. The intensity of green coloring is related to the number of detected peptides.

Peptides identified for the trifusion protein covered about 43% of its sequence (Figure 4A), being highest for the stretch covering the B(ek)I portion. Thus, peptides from both the B and I domains could be detected in the *ca.* 145 kDa MW region. This, together with the successful identification of a peptide deriving from the Y domain (Figure 4A), strongly supported the occurrence

of the Y(yb)B(ek)I tridomain fusion. A higher coverage of the CrtYB(ek)I sequence was attained in measurements of MW gel regions from 40 to 80 kDa (Figure 4B). While most of the detected peptides correspond to either the B or I domains, three of them covered the (ek) linker, extending either towards the B or I domains (Figure 4A).

That a single peptide was detected for Y is likely explained by two parameters: i) the low ionizability of peptides generated from this hydrophobic domain, which consists of 7 membrane-spanning helices; ii) the relatively low abundance of trypsin-cleavage sites (13 potential sites over the 260 residues of the Y domain, as compared to 45 sites for the remaining 400 residues). Such difficulties were confirmed in measurements on the natural YB/I case, which failed to reveal any peptide from the Y domain in any gel region. Similarly, the analysis of the Y(yb)B/I strain shows that peptides covering both the B domain and CrtI were found in the region corresponding to the MW of CrtI but not in the region corresponding to the MW of CrtY(yb)B. This result is in agreement with the faster than expected electrophoretic migration of CrtY(yb)B observed in western blots (see above). Unfortunately, MS data did not permit to clarify whether this phenomena was due to proteolytic degradation of the natural fusion. However, semi-tryptic peptides (which would support the occurrence of non-tryptic proteolytic cleavage) were never be detected, thus indicating low occurrence of proteolysis of the Y(yb)B(ek)I protein. Overall, GeLC-MS investigations provided very good evidences of the presence of the full length Y(yb)B(ek)I protein.

4.4 Catalytic efficiency of the different natural or synthetic enzymatic systems

Quantification of all carotenoid molecules was performed using standards curves and authentic standards except for dihydro- β -carotene that does not exist as a standard but its relative elution order and spectrum are known (Mizoguchi et al., 2015; Takaichi, 2000) (Supplementary Figure S3). In the literature, there is no report on the absorption coefficient of dihydro- β -carotene. However the spectrum of dihydro- β -carotene and β -zeacarotene are relatively similar which signs relative homology between the double bonds resonances of the compounds. We then hypothesized that the two molar absorption coefficients were approximately equal and thus we could use β -zeacarotene calibration curves to quantify dihydro- β -carotene. The full quantification of carotenoids from all

strains was performed using three independent cultures and is presented in the Supplementary Table S3.

4.4.1 Functionality of single domain enzymes

The impact of splitting the CrtB domain from CrtYB on phytoene synthase activity was first evaluated (Figure 5). Phytoene production was severely decreased (40 times) in the B strain compared to the Y(yb)B strain, confirming the importance of the natural fusion between CrtY and CrtB for the phytoene synthase activity (Niklitschek et al., 2008; Verdoes et al., 1999; Xie et al., 2015a). A first hypothesis for the low activity of phytoene synthase when expressed alone could be associated with a faulty targeting of CrtB to the correct cellular environment when disconnected from CrtY, which is normally directed to the endoplasmic reticulum due to its transmembrane segments. As the low phytoene production with our CrtB construct is probably not due to defects in expression (Figure 3A), a second hypothesis is that splitting the two domains decreases the residence time of the phytoene synthase in the vicinity of the membrane, which may also hamper the attainment of the appropriate orientation required for efficient product release. Bearing in mind the difference in water solubility between the substrate (GGPP) and the product (phytoene) of CrtB, the overall effect would be a dramatic slowdown of the enzymatic process by product inhibition due to difficult water environment release.

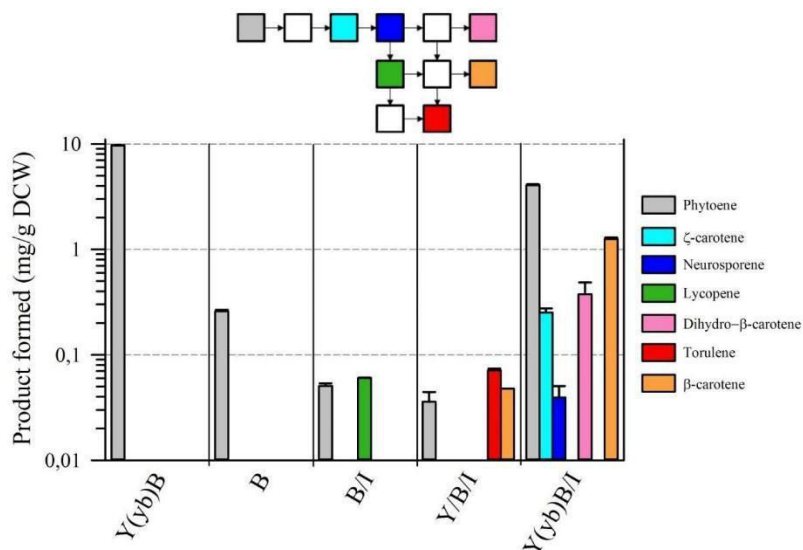


Figure 5. Quantification of carotenoids from the strains bearing single domains enzymes. Major carotenoids were quantified from strains expressing single domain enzymes (B, B/I, B/I/Y) and are compared to data obtained for strain Y(yb)B (left side) or Y(yb)B/I (right side). Metabolites are colored according to the legend in the right. The metabolic pathway from phytoene to carotene is depicted at the top with the same color code as previously, white squares correspond to intermediate metabolites that were not detected. Values represent the average $\pm$ S.D. of the quantification of each metabolite with three independent cultures.

The strain B/I produces a total amount of carotenoids around 100 $\mu\text{g/g}$ DCW, roughly comparable to the one measured in the B strain. Therefore, the coexpression of CrtI with CrtB has almost no impact on the total flux of the carotenoid pathway which again confirms that the low activity of CrtB might be attributed to the slow release of phytoene.

We then reconstituted the full β -carotene producing pathway using the three separated domains (strain Y/B/I). The natural system (strain Y(yb)B/I) produces, in parallel to two major carotenoids (phytoene and β -carotene) several intermediates in detectable quantities such as ζ -carotene, neurosporene, dihydro- β -carotene but no lycopene, confirming data already reported by Verwaal and coll. (Verwaal et al., 2007). In the strain Y/B/I, the total production of carotenoids is reduced by a factor of 55 compared to the Y(yb)B/I strain corresponding roughly to the ratio between

phytoene production when comparing YB and B strains. Interestingly, we could evidence a small, yet visible, production of torulene (71 $\mu\text{g/g}$ DCW) which, to our knowledge, has never been detected when expressing CrtYB and CrtI in *S. cerevisiae* although this molecule can be accumulated in *X. dendrorhous* (Verdoes et al., 2003). The relative levels of carotenoid production in the two metabolic systems (Y/B/I compared to Y(yb)B/I) strengthen the hypothesis of the phytoene synthase defect upon separation with the lycopene cyclase domain. As to why the Y/B/I accumulates small amounts of torulene which is absent when the metabolic flux is greater, the simplest explanation is that when low levels of γ -carotene are present, the competition between CrtI and CrtY might be in favor of the desaturation reaction, leading to an equal proportion of torulene and β -carotene. When the overall carotenoid flux is higher (higher amounts of neurosporene and lycopene due to the fusion of CrtY and CrtB), the competition between lycopene cyclase and phytoene desaturase might be in favor of CrtY, leading to the presence of dihydro- β -carotene rather than torulene.

4.4.1.1 Phytoene synthase/desaturase bidomain fusions

The major bottleneck in reconstituted β -carotene pathways is the reaction catalyzed by CrtI, leading to some accumulation of phytoene in all constructed strains. To study how the enzymatic reactions from phytoene to lycopene can be improved we constructed fusions between CrtB and CrtI domains in different orientations and using different linkers (Figure 6). From Figures 6A and 6B, there seems to be a clear distinction between the I to B and the B to I orientations: I to B fusions accumulate small quantities of intermediates up to neurosporene while the other orientation (with the exception of the BI fusion) seem to either accumulate ζ -carotene or the final product lycopene but not neurosporene. We can thus hypothesize that the geometrical constraints imposed by the order of domains as well as the linker properties influence not only the general activity of the bidomain fusions but also specific steps of the phytoene desaturase reactions.

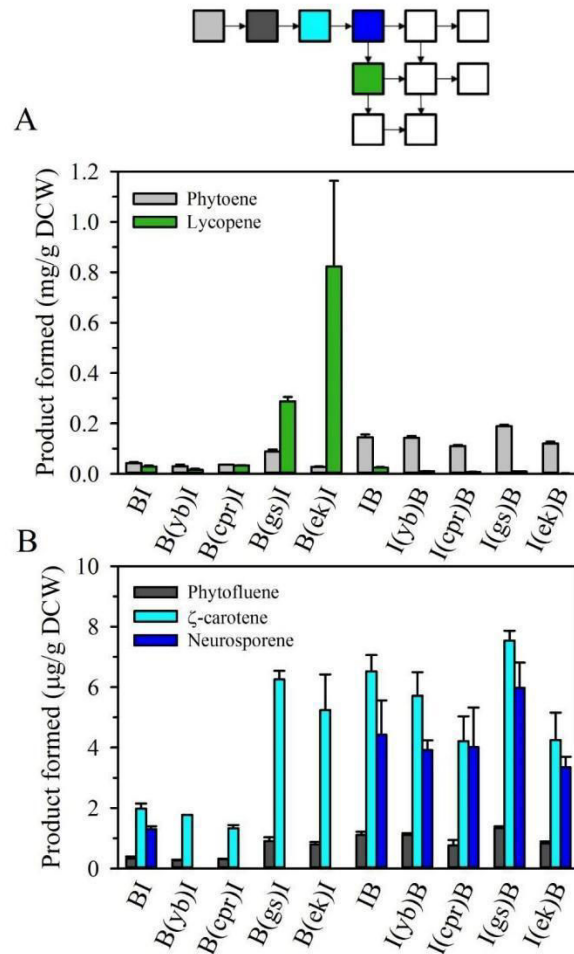


Figure 6. Quantification of carotenoids from the strains bearing bidomain fusions between phytoene synthase and phytoene desaturase enzymes. **A.** Quantification of major carotenoids: phytoene (grey bars), lycopene (green bars). **B.** Quantification of minor carotenoid molecules: phytofluene (dark grey bars), ζ-carotene (cyan bars), neurosporene (blue bars). Metabolites are colored according to the legend inside the two panels. The metabolic pathway from phytoene to lycopene is depicted at the top with the same color code as previously, white squares correspond to intermediate metabolites that were not detected. Values represent the average ± S.D. of the quantification of each metabolite with three independent cultures.

Remarkably, the two most proficient bidomain fusions do not correlate with the most expressed enzymes (Figure 3C). Figure 6A evidences a pronounced distinction between B(ek)I and B(gs)I constructs from the rest of the bidomain fusions as both produce much higher amounts of total carotenoids and with a large preference for lycopene over any other intermediate. Furthermore, the stoichiometry between these two molecules is highly dependent on the linker type present between the two domains. Overall, in the B(gs)I and B(ek)I strains, the lycopene production attained 300 and 800 µg/g DCW levels respectively, being significantly higher than phytoene production. This result demonstrates that the phytoene desaturase activity in the B(gs)I and B(ek)I strains is high enough to convert, almost completely, phytoene into lycopene. Furthermore, the total carotenoid production of the B(ek)I strain (Figure 6) is approximately 3-4 times higher than the ones measured in the B or B/I strains (Figure 5), confirming the defect of phytoene synthase activity when expressed as a single domain. However, in the I to B fusions, the overall carotene production is lower than the one seen with the CrtYB natural enzyme which implies that the CrtI domain might not locate CrtB close enough to the membrane to efficiently release phytoene, but could potentially directly transfer the latter to phytoene desaturase which would then convert it into lycopene, reducing the potential amounts of intermediate carotenoids. This hypothesis is reinforced by the fact that the four linkers tested induce differential effects on the ratio between lycopene and phytoene, possibly evidencing that both proximity and geometry between phytoene synthase and phytoene desaturase domains are important. As to why a small, rigid linker is preferable over the others is still an open question.

4.4.2 Tridomain enzyme

The improvement of lycopene production attained with the B(ek)I fusion, when compared to the combination of individual domains (B/I), prompted us to explore the possibility of optimizing the metabolic efficiency towards β-carotene production with the corresponding tridomain fusion enzymes (Figure 7).

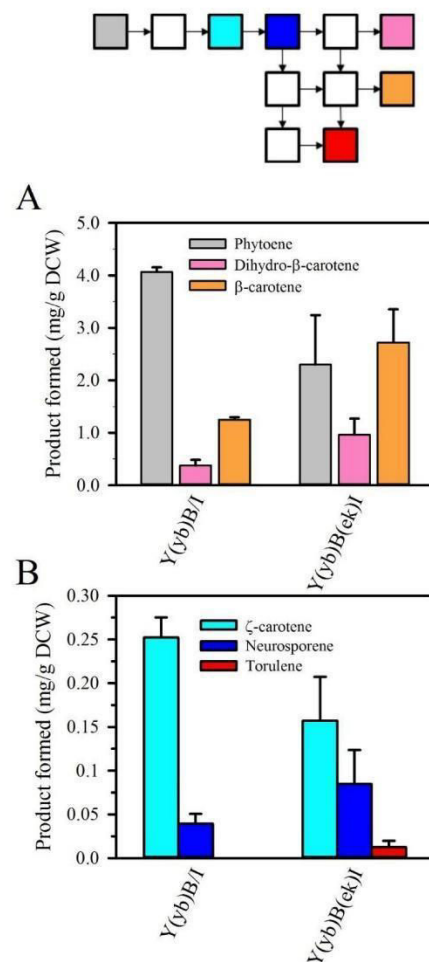


Figure 7. Quantification of carotenoids from the strain bearing tridomain fusion (*Y(yb)B(ek)I*) compared to the natural system (*Y(yb)B/I*). **A.** Quantification of major carotenoid molecules. **B.** Quantification of minor carotenoid molecules. Metabolites are colored according to the legend between the panels. The metabolic pathway from phytoene to β-carotene between the two panels with the same color code as previously, white squares correspond to intermediate metabolites that were not detected. Values represent the average ± S.D. of the quantification of each metabolite with three independent cultures.

The comparison between Y(yb)B/I and Y(yb)B(ek)I strains is striking. While the total carotenoid levels are the same in both strains (Supplementary Table S3) the relative amounts of intermediates decreased for carotenoids before ζ -carotene (phytoene and ζ -carotene) and increased for carotenoids after ζ -carotene (neurosporene, torulene, dihydro- β -carotene and β -carotene) (Figure 7A and 7B). As demonstrated above, the GeLC-MS analysis shows the presence of full-length Y(yb)B(ek)I and low evidence of specific proteolysis degradation that would lead to exact domain separation at the linker positions. Therefore, the change of quantity of late metabolites in the pathway suggests a better metabolic efficiency of the Y(yb)B(ek)I strain compared to Y(yb)B/I in pushing the overall flux toward β -carotene. Indeed, the Y(yb)B(ek)I strain was found to produce two times more β -carotene and dihydro- β -carotene than the strain carrying the natural configuration, while, on the contrary, phytoene and ζ -carotene levels dropped about two-fold. This shift in the ratio between the different carotenoid molecules, particularly those that are produced after ζ -carotene, indicates that metabolic bottlenecks have been alleviated. The absence of any accumulation of lycopene in Y(yb)B(ek)I, while the overall metabolic flux is enhanced, indicates that the last desaturation step is outcompeted by cyclisation with the CrtY enzyme. This competition is also evidenced by the accumulation of dihydro- β -carotene which should only originate through the formation of cyclisation of neurosporene followed by desaturation of β -zeacarotene. Interestingly, in the Y/B/I strain, in which the overall level of carotenoids is much smaller than in the natural system Y(yb)B/I or the trifusion Y(yb)B(ek)I, the metabolic flux is almost completely converted to torulene and β -carotene while no other intermediates could be seen. This might signify that the competition between the last desaturation step and cyclisation could depend on the amounts of these intermediates, favoring desaturation at low carotenoid flux and cyclisation when higher fluxes are present. This could also signify that the metabolic pathway to the production of β -carotene and torulene with the *X. dendrorhous* enzymatic system may favor the use of neurosporene over lycopene. Of course, as the trifusion protein should provide a exact 1:1 ratio between the cyclase (Y) and the desaturase (I) proteins, the observed “push” effect could be due to potential lower amounts of the I domain. This seems quite unlikely considering the western blot data and it is thus more probable that the assembly of the three enzymes promotes a faster cyclisation compared to the natural system.

5 Conclusion

This study shows not only the role of spatial proximity for the good performance of a branched enzymatic system but also demonstrates that such enhancements can be conducted in metabolic systems for which enzymes are present in two different subcellular compartments. The various productivities of the bidomain fusion in producing lycopene and the good efficiency of the tridomain fusion in producing β -carotene indicate that the geometry of the full assembly is a critical parameter for enzymatic activity and metabolites diffusion, both being the primary level of the efficiency of metabolic systems. As such, our system is still opened for further improvement, most notably at the characterization of the tridomain fusion but also in refining the linkers between domains, in improving individual enzymatic activities not to mention metabolic and process-based enhancements that have already shown to increase metabolic yields for carotenoids production (Xie et al., 2014, 2015b).

In summary, the system used for spatial enzymatic confinement shows two complementary advantages: it improves the pathway performance but also helps mapping branched metabolic pathways. In general, domain fusion is a fast method to test hypotheses on how spatial enzymatic proximity can improve metabolic pathway efficiency but it may be also complemented by other assembly methods in order to find the optimal combination of good expression level, high enzymatic activity and enzyme stoichiometry.

6 Conflict of Interest

The authors declare no conflict of interest.

7 Author Contributions

Experimental design: HR, TL, CT, AT, GT; experimental work: HR, SCC, TL, LFGA, CT, AT; data analysis and interpretation: all authors; writing, reviewing and editing of the manuscript: all authors; funding acquisition: GT.

8 Funding

This work was in part funded by the MICA and CEPIA department of INRA through the complete financial support of H.R. PhD salary and by the grant 11050480 “Accueil d’Equipes d’Excellence” from the Région Midi-Pyrénées.

9 Acknowledgments

We would like to thank Fayza Daboussi and Mathieu Fournié for critical reading of the manuscript and helpful discussions.

10 Abbreviations

DCW, Dry Cell Weight; HPLC, High Performance Liquid Chromatography; FPP, Farnesyl pyrophosphate; GGPP, geranyl geranyl pyrophosphate.

11 References

- Albertsen, L., Chen, Y., Bach, L. S., Rattleff, S., Maury, J., Brix, S., et al. (2011). Diversion of Flux toward Sesquiterpene Production in *Saccharomyces cerevisiae* by Fusion of Host and Heterologous Enzymes. *Appl. Environ. Microbiol.* 77, 1033–1040. doi:10.1128/AEM.01361-10.
- Anand, N. N., Mandal, S., MacKenzie, C. R., Sadowska, J., Sigurskjold, B., Young, N. M., et al. (1991). Bacterial expression and secretion of various single-chain Fv genes encoding proteins specific for a Salmonella serotype B O-antigen. *J. Biol. Chem.* 266, 21874–21879.
- Arai, R., Wriggers, W., Nishikawa, Y., Nagamune, T., and Fujisawa, T. (2004). Conformations of variably linked chimeric proteins evaluated by synchrotron X-ray small-angle scattering. *Proteins Struct. Funct. Bioinforma.* 57, 829–838. doi:10.1002/prot.20244.
- Arrach, N., Fernández-Martín, R., Cerdá-Olmedo, E., and Avalos, J. (2001). A single gene for lycopene cyclase, phytoene synthase, and regulation of carotene biosynthesis in *Phycomyces*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 1687–1692. doi:10.1073/pnas.021555298.
- Avalos, J. L., Fink, G. R., and Stephanopoulos, G. (2013). Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat. Biotechnol.* 31, 335–341. doi:10.1038/nbt.2509.

- Baadhe, R. R., Mekala, N. K., Parcha, S. R., and Prameela Devi, Y. (2013). Combination of *ERG9* Repression and Enzyme Fusion Technology for Improved Production of Amorphadiene in *Saccharomyces cerevisiae*. *J. Anal. Methods Chem.* 2013, 1–8. doi:10.1155/2013/140469.
- Carquet, M., Pompon, D., and Truan, G. (2015). Transcription interference and ORF nature strongly affect promoter strength in a reconstituted metabolic pathway. *Synth. Biol.* 3, 21. doi:10.3389/fbioe.2015.00021.
- Chen, Y., and Nielsen, J. (2013). Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks. *Curr. Opin. Biotechnol.* 24, 965–972. doi:10.1016/j.copbio.2013.03.008.
- Dahl, R. H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T. S., Redding-Johanson, A. M., et al. (2013). Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* 31, 1039–1046. doi:10.1038/nbt.2689.
- Delebecque, C. J., Lindner, A. B., Silver, P. A., and Aldaye, F. A. (2011). Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science*, 1206938. doi:10.1126/science.1206938.
- Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., et al. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27, 753–759. doi:10.1038/nbt.1557.
- Frances, O., Fatemi, F., Pompon, D., Guittet, E., Sizun, C., Pérez, J., et al. (2015). A Well-Balanced Preexisting Equilibrium Governs Electron Flux Efficiency of a Multidomain Diflavin Reductase. *Biophys. J.* 108, 1527–1536. doi:10.1016/j.bpj.2015.01.032.
- Gao, S., Tong, Y., Zhu, L., Ge, M., Jiang, Y., Chen, D., et al. (2017). Production of β -carotene by expressing a heterologous multifunctional carotene synthase in *Yarrowia lipolytica*. *Biotechnol. Lett.* 39, 921–927. doi:10.1007/s10529-017-2318-1.

- Gemmecker, S., Schaub, P., Koschmieder, J., Brausemann, A., Drepper, F., Rodriguez-Franco, M., et al. (2015). Phytoene Desaturase from *Oryza sativa*: Oligomeric Assembly, Membrane Association and Preliminary 3D-Analysis. *PLOS ONE* 10, e0131717. doi:10.1371/journal.pone.0131717.
- Güldener, U., Heck, S., Fielder, T., Beinhauer, J., and Hegemann, J. H. (1996). A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* 24, 2519–2524.
- Kim, S., Bae, S.-J., and Hahn, J.-S. (2016). Redirection of pyruvate flux toward desired metabolic pathways through substrate channeling between pyruvate kinase and pyruvate-converting enzymes in *Saccharomyces cerevisiae*. *Sci. Rep.* 6. doi:10.1038/srep24145.
- Lin, J.-L., Zhu, J., and Wheeldon, I. (2017). Synthetic Protein Scaffolds for Biosynthetic Pathway Colocalization on Lipid Droplet Membranes. *ACS Synth. Biol.* 6, 1534–1544. doi:10.1021/acssynbio.7b00041.
- López, J., Essus, K., Kim, I., Pereira, R., Herzog, J., Siewers, V., et al. (2015). Production of β -ionone by combined expression of carotenogenic and plant CCD1 genes in *Saccharomyces cerevisiae*. *Microb. Cell Factories* 14, 1–13. doi:10.1186/s12934-015-0273-x.
- Lu, M.-S., Fang, Y.-J., Chen, Y.-M., Luo, W.-P., Pan, Z.-Z., Zhong, X., et al. (2014). Higher intake of carotenoid is associated with a lower risk of colorectal cancer in Chinese adults: a case–control study. *Eur. J. Nutr.* 54, 619–628. doi:10.1007/s00394-014-0743-7.
- Misawa, N. (2011). Pathway engineering for functional isoprenoids. *Curr. Opin. Biotechnol.* 22, 627–633. doi:10.1016/j.copbio.2011.01.002.
- Mizoguchi, T., Harada, J., Yamamoto, K., and Tamiaki, H. (2015). Inactivation of *bciD* and *bchU* genes in the green sulfur bacterium *Chlorobaculum limnaeum* and alteration of photosynthetic pigments in the resultant mutants. *J. Photochem. Photobiol. Chem.* 313, 52–59. doi:10.1016/j.jphotochem.2015.06.012.
- Niklitschek, M., Alcaíno, J., Barahona, S., Sepúlveda, D., Lozano, C., Carmona, M., et al. (2008). Genomic organization of the structural genes controlling the astaxanthin biosynthesis pathway

- of *Xanthophyllomyces dendrorhous*. *Biol. Res.* 41, 93–108. doi:10.4067/S0716-97602008000100011.
- Pompon, D., F. Garcia-Alles, L., and Truan, G. (2017). “Nanotechnology for Synthetic Biology: Crossroads Throughout Spatial Confinement,” in *Nanotechnology in Agriculture and Food Science*, eds. M. A. Axelos and rcel H. V. de V. h.c (Wiley-VCH Verlag GmbH & Co. KGaA), 209–234. doi:10.1002/9783527697724.ch13.
- Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., and Deber, C. M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Natl. Acad. Sci.* 106, 1760–1765. doi:10.1073/pnas.0813167106.
- Sambrook, J., and Russell, D. W. (2006). SDS-Polyacrylamide Gel Electrophoresis of Proteins. *Cold Spring Harb. Protoc.* 2006, pdb.prot4540. doi:10.1101/pdb.prot4540.
- Schaub, P., Yu, Q., Gemmecker, S., Poussin-Courmontagne, P., Mailliot, J., McEwen, A. G., et al. (2012). On the Structure and Function of the Phytoene Desaturase CRTI from *Pantoea ananatis*, a Membrane-Peripheral and FAD-Dependent Oxidase/Isomerase. *PLoS ONE* 7, e39550. doi:10.1371/journal.pone.0039550.
- Schiestl, R. H., and Gietz, R. D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr. Genet.* 16, 339–346.
- Schmitt, D. L., and An, S. (2017). Spatial Organization of Metabolic Enzyme Complexes in Cells. *Biochemistry* 56, 3184–3196. doi:10.1021/acs.biochem.7b00249.
- Siddiqui, M. S., Thodey, K., Trenchard, I., and Smolke, C. D. (2012). Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res.* 12, 144–170. doi:10.1111/j.1567-1364.2011.00774.x.
- Somasundaram, S., Tran, K.-N. T., Ravikumar, S., and Hong, S. H. (2017). Introduction of synthetic protein complex between *Pyrococcus horikoshii* glutamate decarboxylase and *Escherichia coli* GABA transporter for the improved production of GABA. *Biochem. Eng. J.* 120, 1–6. doi:10.1016/j.bej.2016.12.020.

- Srere, P. A. (1985). The metabolon. *Trends Biochem. Sci.* 10, 109–110.
- Takaichi, S. (2000). Characterization of carotenes in a combination of a C18 HPLC column with isocratic elution and absorption spectra with a photodiode-array detector. *Photosynth. Res.* 65, 93–99.
- Thomik, T., Wittig, I., Choe, J., Boles, E., and Oreb, M. (2017). An artificial transport metabolon facilitates improved substrate utilization in yeast. *Nat. Chem. Biol.* 13, 1158–1163. doi:10.1038/nchembio.2457.
- Tokuhiro, K., Muramatsu, M., Ohto, C., Kawaguchi, T., Obata, S., Muramoto, N., et al. (2009). Overproduction of Geranylgeraniol by Metabolically Engineered *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 75, 5536–5543. doi:10.1128/AEM.00277-09.
- Tsumoto, K., Nakaoki, Y., Ueda, Y., Ogasahara, K., Yutani, K., Watanabe, K., et al. (1994). Effect of the Order of Antibody Variable Regions on the Expression of the Single-Chain HyHEL10 Fv Fragment in *Escherichia coli* and the Thermodynamic Analysis of Its Antigen-Binding Properties. *Biochem. Biophys. Res. Commun.* 201, 546–551. doi:10.1006/bbrc.1994.1736.
- Verdoes, J. C., Krubasik, P., Sandmann, G., and Van Ooyen, A. J. J. (1999). Isolation and functional characterisation of a novel type of carotenoid biosynthetic gene from *Xanthophyllomyces dendrorhous*. *Mol. Gen. Genet. MGG* 262, 453–461.
- Verdoes, J. C., Sandmann, G., Visser, H., Diaz, M., Mossel, M. van, and Ooyen, A. J. J. van (2003). Metabolic Engineering of the Carotenoid Biosynthetic Pathway in the Yeast *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*). *Appl. Environ. Microbiol.* 69, 3728–3738. doi:10.1128/AEM.69.7.3728-3738.2003.
- Verwaal, R., Wang, J., Meijnen, J.-P., Visser, H., Sandmann, G., Berg, J. A. van den, et al. (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.

- Xie, W., Liu, M., Lv, X., Lu, W., Gu, J., and Yu, H. (2014). Construction of a controllable β -carotene biosynthetic pathway by decentralized assembly strategy in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 111, 125–133. doi:10.1002/bit.25002.
- Xie, W., Lv, X., Ye, L., Zhou, P., and Yu, H. (2015a). Construction of lycopene-overproducing *Saccharomyces cerevisiae* by combining directed evolution and metabolic engineering. *Metab. Eng.* 30, 69–78. doi:10.1016/j.ymben.2015.04.009.
- Xie, W., Lv, X., Ye, L., Zhou, P., and Yu, H. (2015b). Construction of lycopene-overproducing *Saccharomyces cerevisiae* by combining directed evolution and metabolic engineering. *Metab. Eng.* 30, 69–78. doi:10.1016/j.ymben.2015.04.009.
- Xie, W., Ye, L., Lv, X., Xu, H., and Yu, H. (2015c). Sequential control of biosynthetic pathways for balanced utilization of metabolic intermediates in *Saccharomyces cerevisiae*. *Metab. Eng.* 28, 8–18. doi:10.1016/j.ymben.2014.11.007.
- Zhang, T., Lei, J., Yang, H., Xu, K., Wang, R., and Zhang, Z. (2011). An improved method for whole protein extraction from yeast *Saccharomyces cerevisiae*. *Yeast* 28, 795–798. doi:10.1002/yea.1905.