Exploring linker's sequence diversity in protein fusion for the biosynthesis of zeaxanthin.

Aurélie Bouin

Université Fédérale
Toulouse Midi-Pyrénées

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

Aurélie BOUIN

Le 6 juillet 2023

Exploration de la diversité de séquence des linkers dans les protéines de fusion pour la biosynthèse de zéaxanthine.

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

Spécialité : **Ingénieries microbienne et enzymatique**

Unité de recherche :
TBI - Toulouse Biotechnology Institute, Bio & Chemical Engineering

Thèse dirigée par
Gilles TRUAN et Nicholas LINDLEY

Jury

Mme Yanyan LI, Rapporteure
Mme Anne-Claude GAVIN, Rapporteure
M. Fabien PIERREL, Examinateur
M. Gilles TRUAN, Directeur de thèse
M. Thomas LAUTIER, Co-directeur de thèse

ABSTRACT

Title : Exploring linker's sequence diversity in protein fusion for the biosynthesis of zeaxanthin.

Spatial organisation of enzyme is a field of metabolic engineering that enables the improvement of metabolic pathways. Indeed, clustering enzymes can limit the loss of intermediates towards competitive pathways or limit the toxicity of theses intermediates within the cell. Strategies such as enzyme compartmentalization, organising of enzymes on scaffold or in metabolon, and protein fusion can bring enzymes inclose proximity. Protein fusion is the most straighforward method to bring enzymes in close proximity but its design remains empirical.

The main objective of this thesis was to investigate the parameters of linkers that impact protein fusion. Starting from a strain of *Escherichia coli* optimized for the production of lycopene, a protein fusion strategy was implemented on the lycopene cyclase CrtY and β-carotene hydroxylase CrtZ to produce zeaxanthin. A linker library was designed and cloned between the enzymes to systematically asses the role of physicochemical parameters of the linkers in the CrtY/CrtZ protein.

In addition to the common parameters such as the enzymes orientation, the length and flexibility of the linker, a fourth parameter was identified. Indeed, the use of amino acid at the extremities of the linker was shown to be more polarized than in the rest of the linker, with four amino acids more represented. We were able to verify that this rule was also observed in the 1280 linkers of natural multidomain proteins.

The fusion of CrtY and CrtZ enzymes allowed to obtain strains improved for the production of zeaxanthin in terms of final titer or specificity.

Keywords: carotenoid – linker – spatial organisation – enzyme - fusion

RÉSUMÉ

Titre: Exploration de la diversité de séquence des linkers dans les protéines de fusion pour la biosynthèse de zéaxanthine.

L'organisation spatiale des enzymes est un domaine de l'ingénierie métabolique permettant d'améliorer les voies de biosynthèses. En effet, organiser les enzymes dans un même espace permet de limiter la perte d'intermédiaire vers des voies métaboliques adverse ou limiter leur toxicité dans la cellule. Parmi les stratégies permettant de rapprocher les enzymes entre elles, on distingue la compartimentalisation des enzymes, l'organisation des enzymes sur des scaffolds ou en métabolons, ou encore les protéines de fusion. Ces méthodes existent dans la nature et ont été adaptées à l'ingénierie métabolique. La fusion des enzymes est la méthode la plus directe pour rapprocher des enzymes mais leur conception reste empirique.

L'objectif principal de cette thèse était l'étude des paramètres du linker pouvant impacter les proteins de fusion. En partant d'une souche d'*Escherischia coli* optimisée pour la production de lycopène, une stratégie de fusion des enzymes lycopene cyclase (CrtY) and β-carotene hydroxylase (CrtZ) a été mise en place pour la production de zéaxanthine. Une banque de linker a été conçue puis clonée entre les enzymes pour en étudier les paramètres physico-chimique pouvant impacter la protéine de fusion CrtY/CrtZ.

En plus des paramètres couramment connus pour impacter les protéines de fusion tels que l'orientation des enzymes, la taille ou la flexibilité du linker, nous avons identifié un quatrième paramètre important. En effet, l'usage des acides aminés aux deux extrémités du linker est beaucoup plus polarisés que dans le reste du linker, avec quatre acides aminés sur-représentés. Nous avons pu vérifier que cette régle reste vraie dans une banque des 1280 linkers également.

La fusion des enzymes CrtY et CrtZ a permi d'obtenir des souches améliorées pour la production de zeéxanthin, que ce soit en termes de spécificité de production pour la zéaxanthine, ou en termes de quantité totale de caroténoïdes produits.

Mot-clés : enzyme – linker – organisation spatiale – enzyme – fusion

ACKNOWLEDGMENT / REMERCIEMENTS

I would first like to thank Dr. Gilles Truan and Dr. Nic Lindley for generously sharing their knowledge and expertise during inspiring discussions.

This endeavor would not have been possible without Dr. Thomas Lautier. Thank you, Thomas, to have offered me the opportunity of this PhD in collaboration between A*STAR and TBI. It has been an amazing journey, both scientifically and personally. I would like to express my deepest appreciation for your kind guidance throughout this PhD, and highlight that I really appreciated our open exchanges, and your constant availability.

I would like to express my sincere gratitude to Dr. Simon Zhang for welcoming me to his team and providing helpful suggestions during my work. Many thanks to Dr. Xixian Chen as well.

Many thanks to the entire metabolic engineering team in Singapore who have provided encouragements whenever I needed it, and to the shared analytical platform for their technical support with HPLC analysis. Although there are many people I would like to thank, I would like to mention a few individuals who have made a particular impact. Sudha, you truly are the heart of the team, and I admire your incredible dedication. Rekha and Tiffany, you both are kind persons, very meticulous and organized who I look up to. Ari, thank you for your listening ear, advices and our dinners! Stella, your hard work, determination and laugher are truly an inspiration! I wish you good luck for your own PhD journey and your future career, I know you'll do great things! Naazneen, I fondly remember our friendly discussions and your reading recommendations. Nicola your enthusiasm and LOUD conversations are greatly missed! ChinChin, thank you for teaching me how to play Singaporean Mahjong, your Monday's cakes and the team B cooking sessions. Cullman thank you for always being up for a hike, and for your cheerful personality. Also thank you to Thao, Yanbin, Jieling, Xiao Hui, Jingsen, Leonard, Ching-Ning, Amy, Federica, Megan G, Deepti, Clement, Jon, Jia Xuan, KC, Shreyash, Alefiya and everyone else on the team. Please know that I consider myself extremely lucky to have known all of you and been a part of such a close-knit team.

Thanks should also go to EAD11 team who punctuated my PhD with brilliant presentations and passionate discussions! I'd like to thank Hélène for her generosity and support and Philippe for sharing news with such great storytelling! Lucie, thank you so much for our endless discussions, see you in Morocco! I would also like to express my appreciation to Sara, Luis and Denis! Outside the EAD11 team, I would like to acknowledge the people visiting the Bio3 building, Quentin, Thomas G., Denis J., and Adilya.

I am also extremely grateful to Dr. Sébastien Darras and Dr. Pierre Crozet for encouraging me to pursue a PhD in the first place.

Thank you also to the badminton group who warmly welcomed me when I barely knew how to hold a racket and count points but never made me feel as I was holding them back. Chuan, Ruphy, Min and Juliani, I will definitely let you know when I return to Singapore so that we can play again!

Je tiens également à remercier chaleureusement ma famille et mes ami·e·s qui m'ont soutenue tout au long de cette thèse. Je tiens à faire une mention spéciale pour Nico, Océ, Émilie K. ainsi que Émilie B. & Alexis qui ont fait le déplacement jusqu'à Singapour. Nos voyages ensemble restent des moments inoubliables !

Je souhaite également remercier mes parents qui m'ont soutenue durant ces longues années d'études, mes sœurs Émilie B., Fa et Emma, ainsi que mon neveu (également filleul !) et ma nièce, né·e·s pendant cette thèse.

Enfin, merci Nico pour tes encouragements et ton amour indéfectible.

Table of content

INTRODUCTION

"Architecture is what ultimately distinguishes a living cell from a soup of the chemicals of which it is composed" – (Harold, 2005)

I. Natural spatial organization of enzymes

A cell is a highly structured organization. The spatial organization is a founding principle of life and can be defined as the internal organization of an object as well as its interactions with the surrounding environment.

Enzymes are proteins that act as catalysts to accelerate specific biochemical reactions in living organisms. For an enzyme, the folding of its amino acid sequence in a tertiary structure leads to a functional unit. This internal arrangement is necessary to ensure efficient interaction with the substrate while preventing other unwanted reactions.

Enzymes can also organize themselves as multiple subunits or domains, creating more complex active sites, multifunctional enzymes or allowing regulation of enzyme activity by allosteric sites. The location of enzymes within cells and tissues also plays a key role in their function. For example, enzymes involved in metabolic pathways may be localized within specific organelles or structures to facilitate the efficient transfer of substrates and products. The organization of enzyme within the cell can help prevent toxicity of metabolites to the cell (ex: peroxisome, metabolosomes), help concentrate reactant (ex: carboxysome), or be used as storage (ferritin).

In this bibliographic section, we will discuss the different spatial arrangements of enzymes at different scales, from larger arrangements to smaller ones.

1. Compartmentation

Bacterial microcompartment

Similarly to eukaryote cells which have specialised organelles delimited by a lipidic membrane, prokaryotes have bacterial microcompartment (BMC), a self-assembling proteinaceous shell encapsulating enzyme (Kerfeld *et al.*, 2018). The protein shell is conserved among organisms and is made up of a thousand of copies of a few proteins, able to self-assemble. BMC are usually between 100 to 400 nm in diameter. Despite a very conserved shell, they have diverse functions and encapsulate various pathways.

Bacterial microcompartment can be divided into two categories, the carboxysome and metabolosomes.

Carboxysome is the only example of BMC involved in anabolism, the synthesis of molecules, known to date. The carboxysome encapsulates carbonic anhydrase (CA) and rubilose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) enzymes. Both enzymes participate in the Calvin-Benson-Bassham cycle, responsible for 90% of carbon fixation, a process by which inorganic carbon is converted to organic molecules during photosynthesis (Huffine *et al.*, 2023) (**Figure 1**). However, at similar concentration, RuBisCO has a stronger affinity to dioxygen (O_2) than carbon dioxide (CO_2) and the Earth's atmosphere is made of 18% of O_2 for less than 0.1% of CO_2. Carbon assimilation is allowed by the selective entry of CO_2 ions in the carboxysomes, effectively concentrating CO_2 around RuBisCO (Turmo, Gonzalez-Esquer and Kerfeld, 2017).

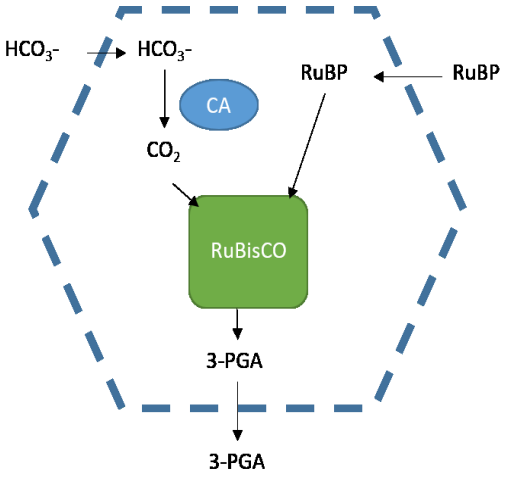


***Figure 1**: Carboxysome.*
3-PGA: 3-phosphoglycerate;
RuBP; Ribulose 1,5-biphosphate

In the case of metabolosome, catabolic microcompartment, the primary role of BMC is to protect the cell from toxic compounds. There are various pathways encapsulated within BMC, such as 1,2-propanediol utilization (pdu), coenzyme B12-dependent ethanolamine degradation (eut), choline utilization (cut) or glycyl radical propanediol (grp) (Stewart, Stewart and Bobik, 2020). These pathways have in common the production of an aldehyde intermediate during the first step of the degradation (**Figure 2**). The aldehyde being volatile and toxic to cells, the BMC allows to contain the aldehyde for detoxification and prevent the loss of carbon (Kirst and Kerfeld, 2019).

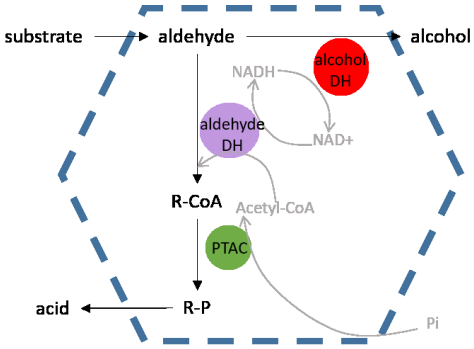


***Figure 2**: Metabolosome.*
PTAC: Phosphotransacyclase.

Encapsulins nanocompartment

Similarly to microcompartments, encapsulin are nanocompartments delimited by a self-assembling protein shell able to encapsulate specific components, called enzymatic cargo. However, the shell is made of only one type of protomer proteins which is homologous to a phage main capsid protein and similarly assemble into an icosahedral structure, usually between 24 to 42 nm in diameter

(Gabashvili *et al.*, 2020). Encapsulins form an emerging research area and their exact roles and mechanisms are still being investigated (Giessen, 2022).

It has been suggested that one encapsulin from *Thermotoga maritima* containing ferroxidase as a cargo protein could be involved in redox metabolism of anaerobic bacteria and be able to couple response to oxidative stress and to excess of Fe^{2+} iron leading to Fenton reactions (Wiryaman and Toor, 2022).

Other encapsulins contain dye-decolorizing peroxidase (DyP) cargo protein. DyP is a relatively new family of heme peroxidase catalyzing H_2O_2 dependant oxydation, widespread among prokaryotes (Xu *et al.*, 2021). Although they are known for their high efficiency on anthraquinone dyes and ability to degrade lignine substrates, encapsulin associated DyP substrates are still unknown as well as their biological function (Tang *et al.*, 2021). However, it has been shown in *Mycobacterium tuberculosis* that they directly protect cell against oxidative stress (Giessen, 2022).

Both microcompartment and nanocompartment have a self-assembling shelf capturing specific enzymes. The shell proteins assemble forming pores in the structure. Pores enable the selective entry of molecules based on their physico-chemical properties, their size, charges, hydrophobicity, etc… These exchanges with the cytosolic environment allow for the free diffusion of substrates and products of the enzymatic pathway encapsulated.

Biomolecular condensates

Biomolecular condensates, also called membraneless organelles, bodies, granules, are microcompartments without physical barriers such as a membrane or a protein shell. They concentrate biomolecules such as protein or nucleic acid. Formation and organization of biomolecular condensate is driven by liquid-liquid phase separation (LLPS) and thus happen when their essentials components reach their solubility limit (Banani *et al.*, 2017). The assembly of biomolecules in condensates depends on weak protein-nucleic acid and/or protein-protein interactions often using protein with intrinsically disordered regions (IDR) and capable of binding RNA and/or DNA (Nesterov, Ilyinsky and Uversky, 2021). Condensates are regulated through the cellular concentration of their constituent components and the modulation of the separation threshold (for exemple, post-translationnal modifications of condensate components can modify their solubility, thus the concentration at which they will form condensates) (Banani *et al*., 2017).

Biomolecular condensates have first been characterized in eukaryotes, where they regulate a wide range of cellular processes. P bodies regulate mRNA decay in *Saccharomyces cerevisiae* and Cajal bodies are sub-nuclear structures implicated in RNA-related metabolic processes as well as telomere maintenance. Functions of biomolecular condensates can be summarized as the regulation of metabolic fluxes or stress reduction (Gao *et al.*, 2021).

Direct evidence of biomolecular condensate in prokaryotes came later due to the small size of bacteria and the technical difficulty ensuing to observe these structures. The best characterized prokaryotic condensates are the BR-bodies, RNA polymerase cluster and PopZ systems (Azaldegui, Vecchiarelli and Biteen, 2021). BR-bodies control the degradation of mRNA by concentrating the mRNA decay machinery and its substrates, long and poorly translated mRNA, while excluding structured RNA (Al-Husini *et al.*, 2020). RNAP are cluster of non transcribing RNA polymerase formed when cells enter log phase in nutrient-rich media. It is proposed that these clusters improve initiation rates by increasing transcription levels. (Ladouceur *et al.*, 2020). PopZ is a regulatory protein of asymmetric cell division. By forming microdomains at the swarmer poles of *Caulobacter crescentus*, PopZ concentrates CtrA transcription factor and generates a gradient in the cell. This leads to an asymmetric cell division and different daughter cells (Lasker *et al.*, 2022).

Overall, although it was long believed that compartmentation of pathway was an exclusive feature of eukaryotic cells, it is now becoming obvious that all cells do regulate their activity through the physical partitioning of its components. While eukaryotes have diverse membrane-bound organelles and prokaryotes tend to optimise certain reactions through protein-bound compartments, both use liquid-liquid phase separation to regulate a large amount of cellular processes. The formation and abundance of the different compartments are based on environmental signals. The absence of physical barriers in LLPS-based compartments allow them to be more responsive as they do not require the synthesis of proteins and are able to directly interact with the components of the cytosol.

The formation of compartments allows to sequestrate specific biomolecules but is rarely able to control the stoichiometry of enzymes and the precise placement or enzymes between themselves. The type of enzyme that can enter the microcompartment is filtered through different means but the number and more precise location inside the compartment is not controlled. Other type of assembly can allow this level precision, the scaffolding.

2. Scaffolding: the cellulosome

The cellulosome is a protein complex responsible for the degradation of cellulose, hemicelllose and pectin, the major constituents of plant cell walls. Cellulose is a repeating polymer of glucose molecules linked by β-1,4 bonds. Despite its chemical homogeneneicity, cellulose is a recalcitrant substrate for hydrolysis because intermolecular hydrogen bonds lead to the formation of both insoluble crystalline structures called microfibrils and less ordered amorphous regions. Coupled with the association of cellulose with other polymers, it is impossible for a single enzyme to degrade cellulose (Lakhundi, Siddiqui and Khan, 2015).

The degradation of plant cell wall is performed by the synergetic action of at least three types of enzymes: endo-1,4-β-glucanase, exo-1,4-β-glucanase and β-glucosidase (Horn *et al.*, 2012). Endoglucanases cleave internal bonds in the cellulose chain while exoglucanases act on the extremities of the cellulose chain. Both endo- and exo- glucanases produce cellobiose which is then

converted to glucose by β-glucosidases. Other types of enzymes are also involved in plant cell wall degradation: chitinase, xyloglucanase, mannanases, etc… (Hirano *et al.*, 2016)

The cellulosome is an enzymatic complex able to efficiently degrade cellulose by gathering all the necessary hydrolytic components around a scaffold. The scaffold is a non-catalytic protein able to bind cellulose *via* a cellulose binding domain and containing several dockerin domains. The various catalytic domains involved in the cellulosome contain cohesin subdomains which interact specifically with dockerin domains of the scaffold (Béguin and Lemaire, 1996).

Thus, the cellulosome allow the specific recruitment of certain type of enzymes as well as their abundance and placement due to the number and positionning of the dockerin domains.

3. Metabolons

Metabolon and substrate channeling

Metabolons are transient multi-protein interactions of sequential enzymes that allow for substrate channeling,which is the direct transfer of metabolites between active sites without diffusion in the bulk phase (Zhang and Fernie, 2021). Due to the transient nature of metabolons, the demonstration of their existence is challenging. One must prove both the interaction of enzymes and the substrate channelling, as enzyme interaction does not necessarily imply that the substrate is channeled between the two active sites, and thus does not diffuse in the cell. Therefore, substrate channeling has been demonstrated in only a few cases.

In 2017, the channeling of two metabolites was demonstrated in the tricarboxylic acid (TCA) cycle of *Arabidopsis thaliana* (Zhang *et al.*, 2017). The TCA cycle metabolizes acetate to form ATP through the succession of eight reactions catalyzed by eight different enzymes. The TCA cycle occurs in most plant, animal, fungi and bacteria and plays a central role as the cell energy provider and supplier of precursors for many biomolecules. Firstly, the team demonstrated the *in vitro* interaction of enzymes catalysing the successive reactions of the pathway. The eight enzymes of the pathway interact together; however, only two of the five metabolites tested for channeling displayed a positive result.

In plant secondary metabolism, dhurrin is a defense compound which can be hydrolyzed into a toxic hydrogen cyanide upon disruption of cellular integrity. Three enzymes are involved in the synthesis of dhurrin: two membrane-anchored proteins, the CYP79A1 and CYP71E1 cytochromes P450 which convert tyrosine *via* aldoxime into cyanohydrin, and one soluble protein, the UDP-glucosyltransferase (UGT85B1) which converts the unstable cyanohydrin into dhurrin, a more stable product. A first team demonstrated protein interactions at the surface of the ER membranes between the three enzymes mentionned above, as well as with the NADPH-dependent cytochrome P450 oxydoreductase (POR2b) responsible for electron donation (Laursen *et al.*, 2016). Moreover, an early

study on dhurrin biosynthesis demonstrated substrate channeling though isotope dilution experiment (Møller and Conn, 1979).

Membrane transport metabolon

Among metabolons, the membrane transport metabolons refer to the interaction of a membrane transporter with an enzyme that acts on the transported substrate. The term metabolon here is employed broadly as it does not always include substrate channeling as per the definition of a metabolon. Membrane transport metabolon can either designate an enzyme acting on a substrate imported by the transporter or an enzyme producing a product to be exported by the transporter (Oreb, 2020) (**Figure** 3).

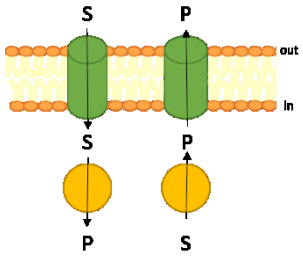


***Figure 3**: Membrane transport metabolon.*

For example, the carbonic anhydrase II of human erythrocyte, responsible for the conversion of carbon dioxide into bicarbonate ion and proton have been shown to interact with a specific cytoplasmic binding domain of the chloride/bicarbonate exchanger in plasma membrane (Vince and Reithmeier, 1998). This metabolon exists in diverse organisms from bacteria to humans. It facilitates the movement of substrate between the two proteins and increases transport efficiency by creating a pH microenvironment around the transporter (Sterling, Reithmeier and Casey, 2001).

The phosphoenolpyruvate: carbohydrate phosphotransferase system (PTS) is a bacterial system involved in the transport and phosphorylation of carbohydrates. The PTS catalyzes the transfer of a phosphoryl group from the phosphoenolpyruvate (PEP) to the carbohydrate (Postma, Lengeler and Jacobson, 1993). The PTS is composed of several actors involved in the transfer of the phosphoryl group: enzyme I (EI), Enzyme II (EII) and the histidine protein (HPr). EI and Hpr are soluble and cytosolic proteins. They are the "general" protein of the PTS, sequentially transferring the phosphoryl group from PEP to enzyme II subunit. EII is a carbohydrate specific protein complex (IIA, IIB, IIC, IID). IIC and IID are membrane transporters while IIA and IIB units transfer the phosphoryl group to the sugar (Jeckelmann and Erni, 2019). The phosphorylation of PTS actors varies according to substrate availability and thus allow the regulation of vital cellular processes such as carbon and nitrogen metabolism (Song, Wu and Beitz, 2011).

In the case of enzyme product being exported by the transporter, one membrane transport metabolon has been identified between the arsenate reductase domain and the arsenite conducting aquaglyceroporins, thus preventing toxic effect of arsenic to the cell (Song, Wu and Beitz, 2011).

Additionally, membrane transport metabolon could be involved in a wide range of metabolic pathway: amino acid, glucose but also nucleic acid, gluconate, glycerol, ammonia, etc… (Moraes and Reithmeier, 2012)

4. Multimerization and multidomain proteins

Multimerization: example of the polyketide synthases

Polyketide synthase (PKS) are a family of multifunctional enzymes complexes producing polyketides, a large group of natural products with applications in health, agriculture or medecine. Polyketides have a complex stereochemistry with multiple chiral centers. For exemple, one of the most studied PKS, the 6-deoxyerythronolide B synthase (DEBS), produces only one stereoisomer for a macrolide with ten stereogenic centers (Khosla *et al.*, 2014).

Polyketide synthases have drawn attention for their stereospecificity, but also for their unidirectional biosynthesis allowed by a complex organization. In this view, the architecture of "assembly line" or "modular" PKS is particurlarly interesting as they channel the growing polyketide chain from one module to another (Nivina *et al.*, 2019).

A typical assembly line PKS contains several modules each responsible for one round of extension of the polytketide chain with the associated functional modification. PKS can contain up to thirty modules, DEBS contains six modules. Each module is made up of at least three domains, a ketosynthase (KS), an acyltransferase (AT) and an acyl carrier protein (ACP) (Nivina *et al.*, 2019). Briefly, the KS receives the growing polyketide chain from the ACP of the previous module. Simultaneously, the AT esterifies an extender unit onto the ACP using acyl-CoA metabolite. The KS catalyzes a condensation between the polyketide intermediate and the extender unit. The intermediate can then be modified by additional domains, such as a ketoreductase, a dehydratase, an enoylreductase, a methyltransferase, or others... (Robbins *et al.*, 2016).

The assembled modules form megadalton multienzyme complexes. All domains and modules of PKS are required and organized in the linear order of enzymatic reactions for the polyketide synthesis (Katz, 2009). The six modules of the DEBS are organised in three pairs of polypeptide chains (Grininger, 2023). The polyketide synthase producing azalomycin F is composed of twenty modules organised in eight polypeptide chains. Each polypeptide chain contains one to five modules. The mechanisms for the module assembly, respecting biosynthesis steps order relies on the specific interaction docking domains at the C-terminus of the upstream module and at the N-terminus of the next downstream module (Zhai *et al.*, 2023).

While PKS are well known for their modular structure, other proteins such as the fatty acid synthases or non ribosomal peptide synthetase have a similar type of arrangement.

Multidomain proteins

70% of proteins are multidomain proteins. Gene duplication and recombination give rise to new multidomain protein. The new multidomain protein will have a function different from the one

domain protein. The multidomain protein might become bifunctional but also function in new context, lose a catalytic function, have a modified substrate binding, etc... (Bashton and Chothia, 2007)

Linkers are ubiquituously observed in natural multidomains proteins (to maintain necessary distance to prevent steric hindrance and/or allow protein-protein interactions between domains) and two studies have focused on the characteristics of linker in multidomain proteins (Argos, 1990; George and Heringa, 2002). It was found that residues in linkers were mostly polars (to be able to interact with surrounding environnement).

A few multidomain proteins have evolved the ability to channel the substrate between their two active sites. The example of the tryptophan synthase and of the thymidylate synthase are detailed below. These proteins are not called metabolon since the interaction is not transient and happens in a single (although multidomain) protein.

The tryptophan synthase is a bifunctional enzyme that catalyzes the last two steps in L-tryptophan biosynthesis. The enzyme is a dimeric unit. The alpha subunit cleaves 3-indole-D-glycerol 3'-phosphate (IGP) into indole and D-glyceraldehyde 3-phosphate (G3P). The beta subunit catalyzes the condensation of indole with serine to yield L-tryptophan and water. Between the α and β subunits, the indole is channeled through a 25 Å hydrophobic interconnecting tunnel, long enough to accommodate up to four indole molecules (Dunn *et al.*, 2008). The binding of serine at the β subunit regulates the formation of indole at the alpha subunit through conformational changes of the protein (Anderson, Miles and Johnson, 1991).

The thymidylate synthase - dihydrofolate reductase (TS-DHFR) from *Leishmania major* is another bifunctional enzyme. The TS-DHFR enzyme participate in the unique *de novo* synthesis pathway of thymidylate, one of the four DNA basis (Wang and McCammon, 2016). The TS catalyzes the reductive methylation of 2'-deoxyuridylate (dUMP) by a tetrahydrofolate and yields the 2'-deoxythymidylate (dTMP) and a dihydrofolate. DHFR catalyzes the reduction of the dihydrofolate by NADPH to yield tetrahydrofolate. The negatively charged dihydrofolate is transferred between the active sites through a tunnel by electrostatic guidance. Indeed, the TS and DHFR active sites of *L. major* DHFR-TS are connected by a tunnel with positively charged residues (Metzger *et al.*, 2014).

5. Does enzyme proximity enhance reaction kinetics?

It was long believed that the proximity of two enzymes performing successive reactions in a metabolic pathway could increase the reaction rate of the second enzyme (Zhang, 2011; Fu *et al.*, 2012; Jandt *et al.*, 2013). The reasonning was that if the product from the first enzyme was directly channeled to the active site of the second enzyme without diffusing in the bulk phase of the cell, then the reaction rate of the second enzyme would be enhanced.

This statement was challenged in the recent years. Indeed, it would imply that the diffusion rate of the substrate is the limiting factor and that when enzymes are not in close proximity, the second enzyme is thus "waiting" to encounter its substrate. However, it has now been demonstrated several times that the diffusion rate is faster than the reaction rates of enzymes, even in a crowded cellular environment. Depending on studies, the factor of metabolite diffusion and enzyme turnover differ greatly. Sweetlove and Fernie calculated that the turnover of enzymes was three orders of magnitudes slower than an average metabolite diffusion constant (Sweetlove and Fernie, 2013). In 2017, another study considered diffusion rate at least three times faster than typical enzymes turnover (Rabe *et al.*, 2017). It was mentionned in 2016 that a positive effect from proximity would only be observed if enzymes were more than 5 µm apart when separated. However, at an enzyme concentration of 1 nM, the average distance between them is usually around 1.2 µm (Wheeldon *et al.*, 2016). Although the numbers differ from one study to the other (probably due to different constant being taken), the recent consensus is that the diffusion rate is not limiting and thus enzyme proximity cannot enhance kinetics of enzymes.

The previous statement knows three exceptions, where enzyme proximity could increase the reaction rates of the enzymes, despite a diffusion rate faster than most enzyme reaction rate. The first exception concerns a system with very low concentrations of enzymes where enzymes would be so far apart (mostly in eukaryotes cells) that the diffusion rate of the substrate would indeed become limiting (Wheeldon *et al.*, 2016; Rabe *et al.*, 2017). Secondly, there is the rare case of "perfect catalytic enzymes" (Sweetlove and Fernie, 2018). "Perfect catalytic enzymes" or diffusion-limited enzymes are enzymes with a K_{cat}/K_m value so high that diffusion rate does becomes limiting. A couple of triose phosphate isomerases from extremophiles organisms displayed a K_{cat}/K_m value of 10^{10}-10^{11} $M.s^{-1}$ (Sharma and Guptasarma, 2015). In this situation also, diffusion would become limiting. Thirdly, enzyme proximity can enhance reaction rate of enzyme before they reach their steady state. The boost is significant enough to be observed but only last for an order of milliseconds (Rabe *et al.*, 2017).

An explanation for the improvement in catalytic activity of enzymes when enzymes were brought in close proximity has been explained by a change in their microenvironment rather than by substrate channeling (Zhang, Tsitkov and Hess, 2016).

Enzyme proximity can favor a reaction during steady state in presence of a competing reaction or when the intermediates are labile or toxic to the cell by preventing their diffusion in the bulk phase.

II. Spatial organisation strategies for metabolic engineering

1. Compartmentation

Targeting organelles in eukaryotes

In eukaryotes, targeting pathway to specific organelles is quite common as each compartment offers a different set of physico-chemical properties as well as a unique content of enzymes, metabolites and cofactors. Enzymatic reactions being sensible to their environment will usually be more efficient in one compartment over the others (Hammer and Avalos, 2017).

This section will not be detailed as the method always relies on the addition of a targeting peptide to the protein of interest. Several examples of targeted pathway will be given for each organelle, as well as a review describing what has been done in the organelle.

The yeast mitochondria is an organelle solicited for metabolic engineering as it can provide large supply of acetyl-CoA and redox equivalents (Duran, López and Avalos, 2020). Among the numerous pathways successfully targeted to the mitochondria, Ehlich pathway improved isobutanol production by 260% and also prevented loss of intermediates to competitive pathway (Avalos, Fink and Stephanopoulos, 2013), isoprene production was improved by 2.1-fold (Lv *et al.*, 2016) and amorphadiene by 20-fold (Farhi *et al.*, 2011).

The peroxisome is a compartment in which happens the β-oxydation of fatty acids producing acetyl-CoA. The peroxisome is thus a place of choice for the production of biomolecules derived from acetyl CoA or fatty acids (Kulagina *et al.*, 2021). Indeed, the localization of squalene pathway to the peroxisome reached a titer of 11g/L (Liu *et al.*, 2020), the production of fatty alcohol was increased by 2.7-fold (Zhou *et al.*, 2016), and the prodeoxyviolacein pigment production was 35% higher with a reduction of side-product when the pathway was localized to the peroxisome (DeLoache, Russ and Dueber, 2016).

Endoplasmic reticulum (ER) is involved, among other functions, in the synthesis and folding of proteins. By expanding the size of the endoplasmic reticulum in *S. cerevisiae* and targeting the morphine pathway to it, a 1.24-fold increase in morphine titer was obtained (Thodey, Galanie and Smolke, 2014). In another study, targeting production of squalene and protopanaxadiol to *S. cerevisiae* endoplasmic reticulum increased their titer by 71- and 8-fold, respectively (Kim *et al.*, 2019), while in *Yarrowia lypolitica*, targeting enzymes of the fatty acid ethyl ester pathway to the ER led to 19-fold increase in titer (Xu *et al.*, 2016).

Previous studies described the targeting of enzymes to the ER through the use of targeting peptides. The resulting enzymes were thus expressed in the ER along with other endogenous reticulum enzymes. A study reported the attempt of generating new ER-derived vesicules free of

endogenous enzymes/pathway in the yeast *S. cerevisiae*. This was mediated by the fusion of the enzymes of interest with a "Zera" peptide. The Zera peptide contains an ER targeting signal and a cysteine-rich region resulting in the creation of disulfide bonds between tagged proteins and thus leading to the genesis of protein bodies, a membrane-surrounded storage organelle. The Zera peptide was fused to the three enzymes of the cis,cis-muconic acid (CCM) pathway. The enzymes were successfully targeted to the ER and producing CCM. However, all three tagged enzymes showed a decreased *in vitro* activity compared to the non-tagged enzymes, and CCM titer was 3-fold lower in the ER than in the cytosol, demonstrating that the ER environment is not good for these enzymes or that compartmentation does not always improves production titer (Reifenrath *et al.*, 2020).

Similarly to organelles targeting in eukaryotes, metabolic pathway can also be adressed to protein bound compartment in prokaryotes. One advantage provided by the bacterial compartment originate from the fact that specific pathway can be isolated from all other endogenous enzymes.

Targeting to micro- and nano- compartments

In the first part of the introduction, several types of protein-bound bacterial compartments were described. Learning about these protein compartments has now allowed us to repurpose them to encapsulate various pathways. As each type of compartment has its own mechanism of assembly, they are exploited differently and thus allow the encapsulation of diverse types of enzymes or have distinct types of applications. For example, in a manner similar to the carboxysome, microcompartment have been repurposed to contain oxygen-sensitive enzymes. System based on the lumazine synthase which assemble through electrostatic interactions with cargo protein can easily encapsulate positively charged protein. Encapsulin nanocompartments have been exported to eukaryotic organisms such as yeast and mammalian cells.

Proteins are targeted to bacterial microcompartment (BMC) through encapsulation peptides (EP). EP are fifteen to twenty amino acids in length and can be located at either terminus of a protein (Lee, Palmer and Warren, 2019). Fusing encapsulation peptides to protein of interest is a straightforward way to target them to microcompartment. Repurposing microcompartment as carboxysome analogues, pathway encapsulation has been applied to the production of hydrogen and pyruvate, protecting oxygen sensitive enzymes. Indeed, hydrogen is a promising source of renewable energy but the most efficient producing enzyme, [Fe-Fe] hydrogenases are sensitive to oxygen. By expressing [FeFe]-hydrogenases and their partners in an *E. coli* strain modified to express the carboxysome and encapsulating the pathway, hydrogen production increases by 4-fold in presence of oxygen. Using microcompartment in anaerobic conditions only increases hydrogen production by 20%. Thus, the compartmentation of hydrogen producing enzymes helps to increase the hydrogen production by targeting the enzymes in a confined space free of oxygen (Li *et al.*, 2020). Similarly, a glycyl-radical enzyme, extremely oxygen-sensitive have been addressed to a microcompartment and allowed the production of 0.5 µmol of formate, when formate production was not detectable in absence of the microcompartment (Kirst *et al.*, 2022).

Another application for the repurposing of microcompartment has been the protection of metabolites from degradation. For example, intracellular polyphosphate levels depend on the polyphosphate-synthesizing enzyme, polyphosphate kinase (PPK1), and on the exopolyphosphatase enzyme responsible for polyphosphate catabolism. Targeting the PPK1 enzyme to a microcompartment resulted in an eight-fold increase of intracellular polyphosphate levels compared to the non-targeted PPK1. Sequestration of polyphosphate metabolite thus protect it from degradation (Liang *et al.*, 2017)

Based on Virus capsid:

This one has been exploited for a long time now, and several method have been set up to adress protein to virus particules.

Bacteriophage can form virus-like particle which volume and pore size can be modulated. The coat protein of the virus usually assemble through non covalent interaction with the C-termini of the scaffolding proteins. They have sometimes been modified with the SpyTag/SpyCatcher system to allow for the targeting of protein to the virus capsid. Indigo production was increased by 60% when the two enzymes of the pathway were targeted to the MS2 phage capsid (Giessen and Silver, 2016) and cellulose hydrolysis was similarly enhanced when the degrading enzymes were targeted to the hepatitis B virus capsid (Berckman and Chen, 2021).

Bacteriophage shell protein have also been expressed in yeast to compartmentalize enzymes. In a study published in 2021, the myo-inositol oxygenase enzyme involved in the rate-limiting step of the D-glucaric pathway was targeted to the murine polyomavirus virus-like particles expressed in the strain. It allowed a 20% increase in glucaric acid production compared to the strain where the enzyme was not adressed to the particule (Cheah *et al.*, 2021).

Based on the lumazine synthase: electrostatic interactions

The lumazine synthase is an enzyme involved in riboflavin biosynthesis and is present in plants and microorganisms. The lumazine synthase and the riboflavin synthase enzymes catalyze the two last steps of the pathway. The lumazine synthase forms an icosahedral capsid built by twelve identical pentameric units (Ladenstein and Morgunova, 2020). The lumazine synthase shell protein from *Aquifex aeolicus* is negatively charged and able to recruit the riboflavin synthase which contains a positively charged patch of amino acids. The supercharged GFP (+36) as well as seven different enzymes fused to the supercharged GFP were successfully encapsulated by the lumazine synthase (Azuma *et al.*, 2016). This strategy was applied to the ascorbate peroxidase enzyme and the successful polymerization of 3,3-Diaminobenzidine was demonstrated (Frey, Hayashi and Hilvert, 2016).

Based on encapsulins:

Encapsulins have been exported to eukaryotic cells, allowing to create completely orthologous protein-bound organelles. Encapsulins are self-assembling compartments and proteins are targeted to them by short targeting peptides (TP) localized at the C-termini. As proof of concept of protein targeting to encapsulin in *S. cerevisiae*, a fluorescent protein was tagged with both a proteasome degradation signal as well as an encapsulin targeting peptide. After two hours, the fluorescence signal decreased by half for the proteins tagged with the degradation tag only, while the fluorescence signal remained stable for proteins with the encapsulin targeting peptide, showing that encapsulin protected the proteins from degradation (Lau *et al.*, 2018). Similarly, encapsulins have also been expressed in mammalian cells to allow the production of melatonin. Melatonin is toxic for cells and usually produced in specialized compartments. The enzyme responsible for the conversion of tyrosine into melatonin was thus addressed to encapsulin through fusion with native protein cargo of the encapsulin. The production of melatonin in presence of the encapsulin was correlated with an increase in cell viability, confirming the sequestration of melatonin in the nanocompartment (Sigmund *et al.*, 2018).

Design of de novo compartment

De novo compartment can be formed in cell through liquid-liquid phase separation resulting in protein condensates. Understanding that liquid-liquid phase separation is driven by intrinsically disordered region (IDR) of proteins allowed to form new biomolecular condensates for enzymes with non-native IDR.

Starting from an identified minimal sequence of eight amino acids from a known disordered protein, Rec-1 resilin, artificial IDR were created in 2020. The minimal sequence was mutated, the resulting peptide was fused to proteins and the formation of condensate regions was evaluated. Using this method, intracellular droplets able to recruit catalytically active enzymes were formed. It was also demonstrated that the dynamic of phase separation is controlled by the molecular weight of the internally disordered proteins (responsible for the formation of the droplets) and by the ratio of aromatic to aliphatic residues in the minimal peptide. Using these two parameters, droplets with a controlled range of Csat (protein concentration at which phase separation will occur) were designed. These parameters were demonstrated to be valid both *in vivo* and *in vitro* (Dzuricky *et al.*, 2020).

At the same time, catalytically active membraneless compartments were created using synthetic disordered proteins in *Escherichia coli*. Fusion of the intrinsically disordered silk-like protein with the two enzymes responsible for the conversion of aspartate β-semialdehyde into 1,3-diaminopropane resulted in the production of 1,3-diaminopropane in a *de novo* compartment (Wei *et al.*, 2020). More recently, protein condensates were also used to enhance the production of the α-farnesene terpenoid in *E. coli* (Wang *et al.*, 2022).

As mentioned, liquid-liquid phase separation responds to protein concentration in the cell, which can vary based on different intra- and/or extra- cellular signals. Liquid-liquid phase separation

technology was thus coupled with optogenetic tools to create dynamic systems able to assemble or disassemble in order of seconds.

Arabidopsis cryptochrome 2 (CRY2) interacts with CIB1 under blue light stimulation to form CRY2-CIP1 heterodimers. CIB1 was fused with a phase separating protein, effectively forming liquid droplets in *E. coli* cells, while CRY2 was fused to protein of interest and distributed evenly in cells. Under blue light stimulation, CRY2 fused to the protein of interest was recruited into the LLPS-based compartment. The recruitment of the protein of interest to the liquid compartment was estimated to be around fifteen seconds and showed to be reversed in fifteen minutes. Using this system, luciferin oxidation was accelerated by 2.3-fold (Huang *et al.*, 2022).

The use of optogenetic tools coupled to the formation of molecular condensates has also been used in the yeast *S. cerevisiae* to redirect metabolic flux at a pathway node for deoxyviolacein production (DV) (**Figure 4a.**). Briefly, VioE enzyme catalyzes the formation of the protodeoxyviolacein (PTDV) intermediate. PTDV can either be catalyzed by VioC enzyme to produce DV or be spontaneously oxidized to form the prodeoxyviolacein (PDV) undesired side-product. The study is based on the hypothesis that colocalizing VioE and VioC enzymes might help direct the flux from PTDV

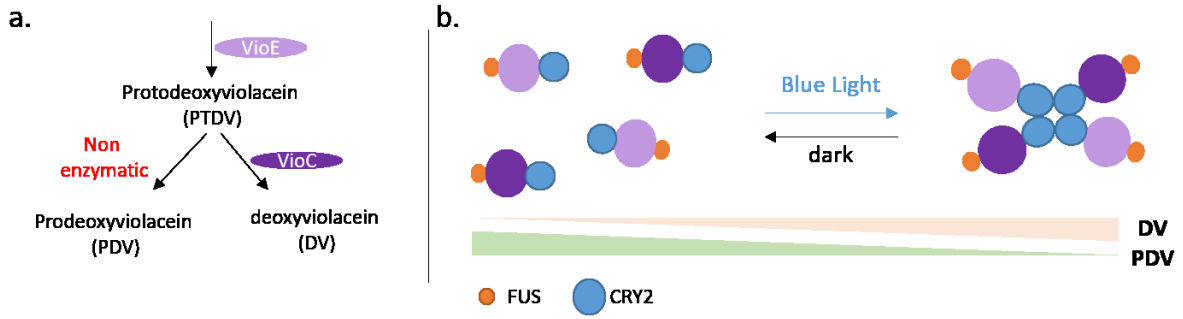


***Figure 4**: Production of proteodeoxyviolacein. a. The protodeoxyviolacein pathway. b. Optogenetic strategy for the production of protodeoxyviolacein.*

to DV. Two optogenetic systems were tested, one with CRY2 protein, which assemble under blue light stimulation, and one with PixE and PixD proteins which cluster in the dark and disassemble under similar light stimulation.

Optogenetic protein domains were fused to the intrinsically disordered N-terminal FUS protein domain. A strain with a six-fold increase in the PV product was obtained, which represented an 18-fold increase in the PV/PDV ratio (Zhao *et al.*, 2019) (**Figure 4b.**).

Mainly, the compartmentation of exogenous metabolic pathway can be done by targeting to existing organelles, repurposed bacterial micro- and nano-compartment or by the formation of biomolecular condensates driven by liquid-liquid phase separation.

2. Scaffolding

In metabolic engineering, scaffolds are used to enhance metabolic pathway by colocalizing enzymes. The scaffold can be made of protein, nucleic acid or lipid and share a similar concept; the scaffold is made of repeated elements that can specifically interact with a part of a modified enzyme. The number of repeated element can be modified to be adapted to each specific pathway and the stochiometry of the enzymes can be adjusted to balance pathway fluxes and enzyme kinetics. The proximity of the enzymes can also be tuned by changing the distance of the repeated elements on the scaffold.

Protein

Protein scaffold relies on the specific interaction between a recognition domain and a ligand.

Any natural recognition domain and its ligand, if they have a strong binding affinity and high specificity, could be used to gather enzymes. However, only a few have been typically used until now: Src Homology 2 (SH2), Src homology 3 (SH3), PSD95/DlgA/Zo-1 (PDZ), and GTPase Binding Domain (GBD) recognition domains with their ligands, as well as the cohesin-dockerin system. The cohesin-dockerin system is particular since the binding part, the cohesin, is a full protein domain whereas other ligands are small peptides of a dozen of amino acids. SH2, SH3, PDZ and GBD can thus be quite advantageous since they require only minor modifications of the enzymes due to their small size (**Table 1**). The cohesin-dockerin domains have a stronger affinity, lots of orthologues but are calcium dependant, and thus mainly used at the cell surface of yeast.

***Table 1**: Comparison of synthetic protein scaffold*

Regognition domain	Domain size (amino acids)	Ligand size (amino acids)	K_d (in M)	Ligand position
SH2	100	3-6		C-term
SH3	57	11	$1*10^{-7}$	–
PDZ	96	6	$8*10^{-6}$	C-term
GBD	80	32	$1*10^{-6}$	
Cohesin	150	70	$10^{-9} – 10^{-12}$	

The first and most emblematic reported success for enzyme scaffolding using a protein scaffold is the one from Dueber team in 2009. The scaffold, made of fused GBD, SH3 and PDZ domains was used to gather the three enzymes of the mevalonate pathway, AtoB, HMGS and HMGR (Dueber *et al.*, 2009). By gathering the enzymes together and optimizing the stochiometry of each enzyme as well as the induction level of the scaffold, the mevalonate titer was increased 77-fold. Just by using a scaffold on the bottleneck enzyme of the glucarix acid, they improve an original titer of 0.5g/L to 1.75 g/L

without any optimization. In their follow up paper (Moon et al., 2010), they manage to reach a 5-fold increase by optimizing the scaffold.

This scaffold system has now been used on a variety of pathway, both in *E. coli* and *S. cerevisiae*: electron transfer circuit, butyrate, Gamma-aminobutyric acid, resveratrol, etc… (Agapakis *et al.*, 2010; Wang and Yu, 2012; Baek *et al.*, 2013; Dung Pham *et al.*, 2016)

Cohesin-dockerin domains have also been use to colocalize successive enzyme of a pathway instead of carbohydrate active enzymes. The cohesin-dockerin system was indeed used for the cytosolic production of 2,3 butanediol in *S. cerevisiae*. The scaffold was made of seven cohesin domains and the three enzymes of the pathway, AlsS, AlsD, and Bdh1, were fused to dockerin domains. The scaffolding of the enzymes allowed a 37% increase in the 2,3-butanediol titer (Kim and Hahn, 2014).

The cohesin-dockerin domains are however mainly used with yeast at the cell surface display. They are of particular interest in the biofuel cell field where chemical energy is converted into electricity. It requires an anode and a cathode. The anode contains oxidizing enzymes able to separate electron from a donor molecule while cathode use oxygen-reducing enzymes. The cohesin-dockerin domains are used to display the electrosome at the cell surface. In 2017, the cohesin-dockerin domains were use to scaffold enzymes performing ethanol oxidizing cascade. Ethanol was oxydized by an alcohol dehydrogenase into acetaldehyde and acetaldehyde was further oxydized into acetate by a formaldehyde dehydrogenase. This system outperformed usual systems which only recover one electron from the donor molecule. Indeed, the power output for the non scaffolded enzymes was 0.25 $\mu W.cm^{-2}$, while the first oxydizing enzyme scaffolded alone displayed a power output of 1.4 $\mu W.cm^{-2}$ and both scaffolded enzymes yielded a power output of 2.7 $\mu W.cm^{-2}$. In the cathode, increasing the copy number of oxygen reducing enzymes also increase the power output and allow to compete yeast oxygen consumption (Szczupak *et al.*, 2017).

Nucleic acid

RNA scaffolds are made of RNA strands designed to fold into aptamers. Aptamers are RNA secondary structures able to bind to a protein target with high affinity and specificity. The RNA strands also contain domains that allow their polymerization into one-dimension or two-dimension scaffolds. Since each different aptamer can be specifically recognized by a binding protein domain, the binding protein domains are fused to the protein of interest to target them to the scaffold which become a docking platform.

This strategy has been applied for different pathways: the production of hydrogen (Delebecque *et al.*, 2011) and the production of succinate (Sachdeva *et al.*, 2014). For the production of hydrogen, fusion proteins were made between the PP7 binding domain and a hydrogenase and between the MS2 binding domain and a ferredoxin. Expression of the ferredoxin and hydrogenase protein in

presence of the 2D scaffold in *E. coli* increased the hydrogen production by 24-fold. It was the first nucleic acid scaffold applied *in vivo*. The succinate pathway involves three enzymes and the oxaloacetate intermediate which can be used by other intracellular enzymes to produce amino acids, fatty acids, etc… By docking the three enzymes to the RNA scaffold, the succinate production was increased by 80% in *E. coli*.

Although RNA scaffolds are easy to implement and can increase the product yield of various metabolic pathways, RNA is prone to degradation, compared to the more stable DNA (Geraldi *et al.*, 2021). DNA is a self-assembling biopolymer. Single stranded DNA can be designed to form structures such as the Holliday junction or crossovers. But more complex and multidimensional DNA nanostructures can also be designed using scaffolded DNA origami and ssDNA tiles (Fu *et al.*, 2020). DNA nanostructures can be used as a scaffold to recruit specific proteins. Several technologies can allow the recruitment of proteins on DNA scaffolds: covalent crosslinking, non-covalent binding, and fusion tags. Most of these strategies requiring DNA origami or modified nucleotides, non-natural amino acids, streptavidin tagged protein or other modifications of biomolecules are often used *in vitro* (Wilner *et al.*, 2009; Erkelenz, Kuo and Niemeyer, 2011; Liu *et al.*, 2016).

Two examples where plasmid DNA have been used as a scaffold to tether protein of interest *in vivo* in *E. coli* will be described here.

First, zinc finger binding domain sequences were incorporated into a plasmid sequence and protein of interest were fused to zinc finger protein domains. Each zinc finger protein recognized a different DNA sequence of ten nucleotides or less. The authors tested the strategy on three different pathways, the resveratrol pathway, the 1,2 propanediol pathway and the mevalonate pathway (Conrado *et al.*, 2012). Resveratrol synthesis in *E. coli* occurs in two steps. 4-coumaric acid is first converted to 4-coumaroyl-CoA by the 4-coumarate:CoA ligase (4CL) and trans-resveratrol formation is catalyzed by the stilbene synthase (STS) which condenses 4-coumaroyl-CoA and malonyl-CoA. Scaffolding of these two enzymes on a DNA scaffold generates a 5-fold increase in resveratrol production after optimization of the spacer distance between the zinc finger binding sites on the scaffold (8bp) and the number of scaffold repeat. Interestingly, the 4CL-STS direct fusion of enzymes produced 50-times less resveratrol than the scaffold system, probably due to misfolding of the large protein fusion. The similar optimization of the 1,2 propanediol pathway involving three different enzymes improved the final titer by 4.5-fold. The fold improvement for the mevalonate pathway was close to three, less than the protein scaffold (Dueber *et al.*, 2009), although no absolute quantification is given.

Later, transcription activator-like effector nuclease (TALE) was also used to create a TALE-DNA scaffolding system. TALEs are like zinc fingers as they recognize specific DNA sequences. Briefly, TALEs share a common DNA binding domain, each specific of a DNA sequence due to a highly conserved region where two amino acids confer the specificity. This gives a competitive advantage to the TALE system as the DNA sequence recognized by the protein can be modified and is thus more flexible. The

strategy was applied for 3-acetic acid production, a two-enzyme pathway with final product titer increased by almost 10-fold (Zhu et al., 2016).

Lipidic and membrane bound scaffold

The scaffold made of protein and nucleic acid described can be used *in vivo* or *in vitro* but they mostly work for soluble protein. Membrane proteins represent 20 to 25 % of all proteins in prokaryotes (Sawada et al., 2007). Membrane proteins cannot be adressed to cytosolic scaffold as it would likely alter their folding and impede their activity. A few papers have tried to specifically designed lipidic scaffold or membrane protein anchor for membrane or membrane bound proteins. Five strategies will be described here: the recruitment of proteins to the membrane using self-assembling transmembrane heteromers or a self-assembling transmembrane oligomer, a peptide tag to localize proteins to lipid droplets in yeast, generation of functional membrane microdomains in prokaryote and the formation of lipidic particules. These approaches are of particular interest for pathways involved in hydrophobic intermediates or products.

The first strategy has been developed based on the ΔpH-dependent twin-arginine translocation (tat) responsible for the active transport of folded proteins across the lipid bilayer in plant thylakoid fraction of the chloroplast and prokaryotes. Three proteins are involved in the tat complex: TatA, TatB and TatC. TatB and TatC form a tatBC heterodimer, and seven to eight heterodimers arranged together to bind substrate proteins. The tatBC/substrate is then able to recruit TatA to the complex. TatA is responsible for the translocation of the folded proteins (Rodriguez *et al.*, 2013). The self-assembly property of the TatBC dimer have been repurposed to colocalize enzymes of the dhurrin metabolon in the thylakoid membrane (**Figure 5**). The dhurrin metabolon involves the two cytochromes P450 CYP97A1 and CYP71E1 as well as a soluble UGT. In eukaryotes, P450s localize to the endoplasmic reticulum (ER) and require reductase to provide electrons. A previous study where the enzymes of the metabolon where targeted to the chloroplast instead of the ER showed that the formation of the metabolon was lost and that numerous unwanted side products were formed instead of dhurrin. Here, the author removed the two natural P450 membrane anchors of the P450s and replaced them by the TatB transmembrane domain. The soluble UGT was fused to the TatC protein with a flexible linker. This strategy resulted in a 5-fold increase in the production of the final product dhurrin and a reduction of unwanted side product in *Nicotiana*

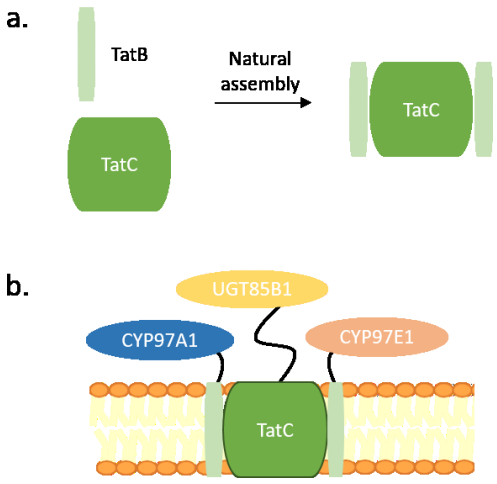


***Figure 5**: TatBC dimer assembly for enzyme colocalisation. a. natural assembly of the dimer. b. the enzyme colocalisation strategy.*

benthamiana (Henriques de Jesus *et al.*, 2017).Another team tried to designed a scaffold usable in various host. The strategy relies on the CURvature Thylakoïd1A (CURT1A) protein from the thylakoide membrane of *Arabidopsis Thaliana*. Curt1 is a small integral protein of 12.2kDa with two stroma facing amphipatic helices linked by two transmembranes helices. Curt1 self-assemble in oligomers (**Figure 6**). Curt1 was fused in one or both termini with fluorescent proteins and the different constructs were tested in *Nicotiana benthamiana* chloroplast and ER, in *S. cerevisiae* and in *E. coli*. In the chloroplast of *N. benthamiana*, the constructs were correctly inserted into the membrane and retained their oligomerization properties. However, in the ER only the C-term fusion is able to properly insert itself in the membrane and oligomerize (thus can not be applied to two different proteins and act as a scaffold?). In *S. cerevisiae*, insertion in ER cortical membrane and oligomerization (active or due to crowding) also worked. However, in *E. coli*, the constructs severely impeded growth. The activity of β-glucuronidase fusion with CURT1 fusion was improved 13-fold, and 45-fold, 200-fold in *S. cerevisiae*, in the thylakoid and in the ER of *N. benthamiana*, respectively (Behrendorff *et al.*, 2019).

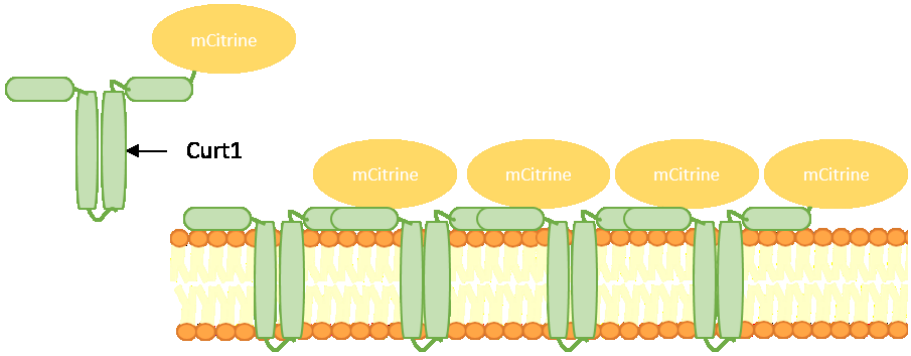


***Figure 6**: CURT1A protein assembling in oligomers. Adapted from (Behrendorff* et al.*, 2019)*

A third strategy has been to develop a protein scaffold adressed to the outer membrane of lipid droplets in *S. cerevisiae*. This strategy was applied to the ester biosynthesis pathway. Indeed, in yeast, ethyl acetate is produced by the condensation of ethanol and acetyl CoA by Atf1 enzyme that localize to the ER during the exponential phase and then into lipid droplet during the stationnary phase. The upstream enzymes of the pathway, Acs1, that catalyse the conversion of acetate to acetyl CoA, and Ald6, that catalyse the conversion of acetaldehyde to acetate, localize to the cytosol and the mitochondria. The authors hypothesize that colocolazing the upstream enzymes of the pathway, Ald6 and Acs1 to the lipid droplets would improve the pathway flux and the final conversion to ethyl acetate. The oleosin domain was identified as responsible for the lipid droplet-

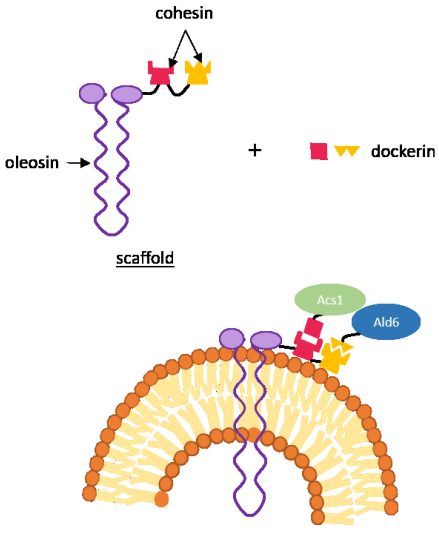


***Figure 7**: Targeting enzymes to membrane using an oleosine-based scaffold.*

targeting of various enzymes in mamalian and plant cells. Cohesin domain (from cohesin-dockering pairs of the cellulosome) were thus localized to lipid droplets by fusion to the oleosin scaffold (**Figure 7**). Dockerin domains were fused to Ald6 and Acs1 enzymes and were thus targeted to the lipid droplets as well. The scaffolded pathway had a near 2-fold improvement over the non-scaffolded pathway (Lin, Zhu and Wheeldon, 2017).

A fourth strategy was inspired by the equivalent of lipid raft in prokaryotes, functionnal membrane microdomains (FMM). FMM have high spatio and temporal stability and are involved in numerous cell process such as signal transduction, secretion and transport processes. They contain scaffolding proteins responsible for the recruitment of other proteins to lipid rafts. Two scaffolding proteins FloT and FloA have been identified in *Bacillus subtilis* as scaffolding proteins. FloT can be divided into two domains, the N-terminal SPFH domain and the C-terminal flotillin domain. The SPFH domain is responsible for protein localization to FMM. The authors used these proteins to recruit enzymes of the N-acetylglucosamine (GlcNAc) pathway to FMM and construct an enzyme cascade in *B. subtilis*. The pathway involved six enzymes; ptsGHI, pgi, glmS, GNA1 and yqaB. Among these 6 enzymes, GNA1, GlmS, Pgi and YqaB were demonstrated to be cytosolic. The four cytosolic enzymes were thus adressed to FMM through fusion with SPFH or FloA or FloT protein domains. Although the fusion caused a decrease in enzyme activity between 1.3 to 26.7 times, the final titer in GlcNAc increased by more than 3-fold. Supported by a kinetic model, the authors attributed the increase in titer to substrate channeling in the FMM (Lv *et al.*, 2020).

Lastly, a synthetic lipid-containing scaffolds (SLS) was created using two components of the bacteriophage ϕ6: the membrane protein P9 and the non structural protein P12. The expression of these two proteins alone is sufficient for the formation of lipid particules in *E. coli (***Figure 8***)*. The lipid particules are larger than 20 nm in diameter and are a mixture of lipids and proteins. Two enzymes of the indigo pathway, TnaA and FMO, were colocalized to the SLSs by fusion to the C-termini of P9 proteins. In presence of SLS, the indigo production had a 2-fold increase (Myhrvold, Polka and Silver, 2016).

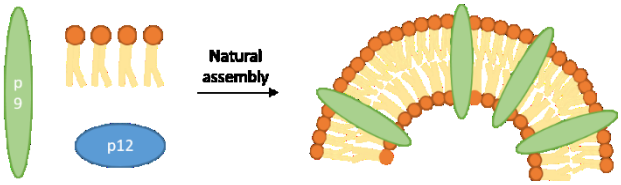


***Figure 8**: Synthetic lipid-containing scaffolds (SLS).*

Various strategies have been develop to colocalize enzymes to the membrane, and they include plant cells as well as eukaryotes such as *S. cerevisiae* and eukaryotes.

Other

Two more methods to scaffold proteins together should be mentioned as they are original and less described. One is based on affibodies while the second divert the PduA protein to form a filamentous scaffold in *E. coli* cells.

Affibodies are made of 58 amino acids shaped in triple alpha-helix, and come from the Z domain of *Staphylococcus aureus* protein A. A wide library of affibody-anti-affibody pairs with a range of binding affinity exists, and their affinity are characterized. Based on this technology, a scaffold was made by assembling anti-affibody peptides while the corresponding affibodies were fused to the protein of interests. The colocalisation of farnesyl diphosphate synthase and farnesene synthase was thereby set up in *S. cerevisiae*, yielding a 135% increase in the final product titer. The enzyme:scaffold ratio was the most impactfull parameter to optimize. The three-enzyme polyhydroxybutyrate pathway introduced in *E. coli* showed a 7-fold increase in polyhydroxybutyrate production. Thus, this method has been proven successful both in *E.coli* and *S. cerevisiae* (Tippmann *et al.*, 2017).

The PduA shell protein from the propanediol-utilization metabolosome is a self-assembling protein which form intracellular filamentous arrangement spanned along *E. coli* cytoplasm. A coiled-coiled system was used to target enzymes to the filament. The coil-coiled heterodimer is made of peptide A and peptide B. Peptide A is rather acidic while peptide B is basic and the two peptides interact together specifically. Peptide B was fused to PduA and PduA kept its filamentous property. Peptide A was fused to the ethanol producing enzymes, pyruvate decarboxylase (Pdc) and alcohol dehydrogenase (Adh). When coexpressed with the PduA scaffold, the ethanol production was increased by 221% per OD unit compared to enzymes alone, without the scaffold. It was also demonstrated that the scaffold can be directed to the cell inner membrane (Lee *et al.*, 2018).

3. Artificial membrane transport metabolon

The design of *de novo* metabolon, as the transient interaction of sequential enzyme resulting in substrate channeling is hardly achievable today since the mechanisms for such interactions are not fully understood and require precise adjustments of the protein. However, the fusion of membrane transporter with associated enzyme of the pathway has been successfully achieved. These fusions, called transport metabolons, have not always been investigated for substrate channeling.

One transport metabolon has been designed for the import and metabolization of xylose. Indeed, upon its entry in *S. cerevisiae*, xylose can either be converted to xylulose by the xylose isomerase or be side-tracked by the aldose reductase to form xylitol instead of xylose (**Figure 9**). Xylose is an early percursor of the pentose phosphate pathway and glycolysis to later form ethanol. The xylose isomerase has a low catalytic efficiency and is also inhibited by xylitol, creating a negative feedback loop. By coupling the xylose isomerase and galactose permease tranporting xylose, the consumption of xylose was accelerated, with a drop in the unwanted side product, xylitol, and an increase in ethanol production. The transport metabolon thus allowed the protection of the substrate from competing

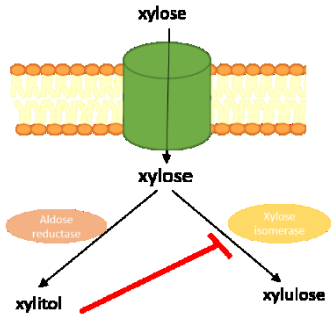


***Figure 9**: The artificial membrane transport metabolon of the xylose pathway.*

pathway (Thomik *et al.*, 2017). Similarly, the coupling of the glutamate decarboxylase, producing GABA from glutamate with GABA antiporter increased GABA production by 3.5 fold (Somasundaram *et al.*, 2017).

4. Post-translationnal protein assembly

The Dock-and-Lock method: RIAD-RIDD peptide pairs

The Dock-and-Lock method relies on the natural association of two protein domains: the regulatory subunit of protein kinase A (PKA) and the anchoring domain of A-kinase anchoring proteins (AKAP) (Figure 10a.). The regulatory subunit is made of the last 44 N-terminal residues of the PKA and form homodimers. The anchoring domain of AKAP is a helix of 14 to 18 amino acids which only binds the dimerized regulatory domain of PKA. These domains were modified with a cysteine residue to create covalent disulfide bonds between them, resulting in a site specific conjuguation method. Fusing a dimerization domain (RIDD) to a protein A allows to group them in dimers. With an additional fusion of the anchoring domain of AKAP (RIAD) with a protein B allows the formation of trimeric complexes with two proteins A and one protein B (Chang, Rossi and Goldenberg, 2007). These peptides were fused to various enzymes in order to create complexes with enhanced performance (Figure 10b.).

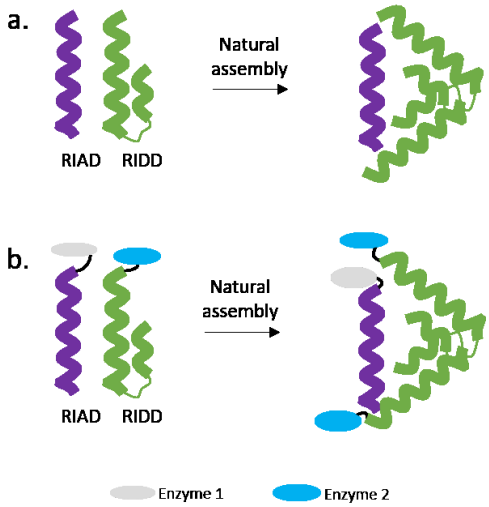


Figure 10: RIAD-RIDD peptide assembly. a. In a natural context b. For enzyme colocalisation.

These enzyme complexes were applied for the production of quinolinic acid (Zhu, Peña and Bennett, 2021) and phycocyanobilin (Y. Wang *et al.*, 2023) in *E. coli*. In both studies, the enzyme assembly was one of the strategy implemented among other optimization processes such as the selection of the best orthologous enzymes, knockout of genes negatively affecting the pathway of interest, overexpression of genes allowing the accumulation of precursors and optimization of fermentation conditions. The production of phycocianobilin was improved from 21.4 mg/L to 23.5 mg/L with the post-translationnal enzyme assembly of the two key enzymes of the pathway with the peptide tags RIDD-RIAD. The results were more spectacular for the quinolinic acid production. Quinolinic acid is produced from aspartate by two sucessive reactions. First, there is the conversion of aspartate to 2-iminosuccinate catalysed by NadB enzymes, and 2-iminosuccinate is further converted to quinolinate by NadA enzyme. 2-iminosuccinate was reported to be unstable and the authors showed that this part

of the pathway is the limiting step. Enzyme fusion of NadB and NadA enzymes led to a performance similar to the one the free enzymes. However, enzyme assembly of NadA and NadB enzymes though peptide-petide interaction of RIAD and RIDD led to a 2.5-fold increase in the quinolinate titer, going from 1g/L with the independent enzyme to 2.5g/L with the peptide tags.

The RIAD and RIDD peptide tags were also used *in vitro* for cofactor regeneration purposes. The tags were fused to the C-terminus of a phenylacetone monooxygenase, PAMO, and to a phosphite dehydrogenase, PTDH. The enzyme complexes displayed similar catalytic activity when compared with the PTDH–PAMO fusion enzyme, However, the self-assembled PAMO–PTDH complex required 4 times less PTDH for the same performance when compared with the PTDH–PAMO fusion enzyme (Purwani *et al.*, 2021).

SpyTag/SpyCatcher

The fibronectin-binding protein FbaB from *Streptococcus pyogenes* contains a domain which spontaneously forms an intramolecular isopeptide bond between Lys and Asp. The Ffab domain was engineered and split in two parts: SpyTag, a 13 amino acids peptide; and SpyCatcher, a small protein domain of 138 amino acids (15 kDa). SpyTag and SpyCatcher are able to form an irreversible amide bond together under various pH (5 to 8), temperature (4°C to 37°C) and buffer (PBS, phosphate-citrate, Hepes, Tris) conditions (Zakeri *et al.*, 2012) (Figure 11).

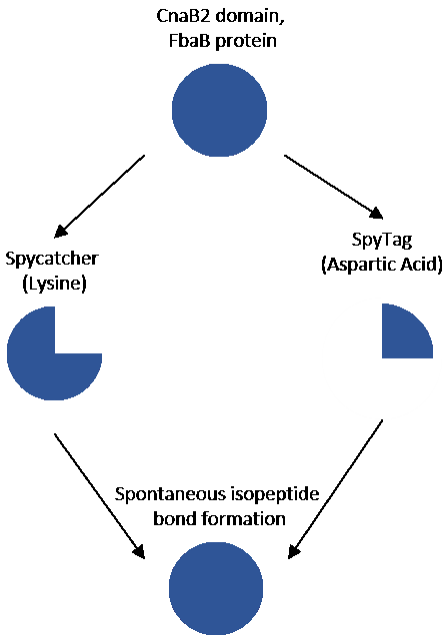


Figure 11: The SpyTag-SpyCatcher system.

This tool has been used many times to bring enzymes in close proximity. Fusion of spyTag and Spycatcher to the carbonyl reductase (CpCR) and glucose dehydrogenase (GDH) involved in ethyl(R)2-hydroxy-4-phenylbutanoate ((R)-HPBE) pathway, increases its conversion rate by 2.4-fold compared to the free enzymes (J. Wang *et al.*, 2023). Applied to the formate hydrogenase and the 2,3-butanediol hydrogenase enzymes, SpyTag/SpyCatcher system improved the catalytic rate of (R)-1-phenyl-1,2-ethanediol reduction by a 2.9-fold. The tagged enzymes also displayed a better stability in regard to pH and solvents (Peng *et al.*, 2020).

The Spycatcher system has also been polymerized into a scaffold. The SpyCatcher domain was tagged on both termini by a tyrosine residue which polymerises in presence of HRP and H_2O_2. The SpyCatcher scaffold was applied to an artificial cellulosome to hydrolyse arabinoxylan. The endoxylanase and arabinofuranosidase were both tagged with SpyTag. The reaction rate of the tagged enzymes on the SpyCatcher polymers (Km and Vmax) were similar to the one of the independent

enzymes. However, by adjusting the enzymes ratio to one endoxylanase to four arabinofuranosidase, the degradation of sugar (arabinoxylan) increase by 1.63-fold (Jia *et al.*, 2017).

All the example described above were investigated *in vivo*. One study analysed the formation of extracellular enzyme complexes with the SpyTag/SpyCatcher system. An endoxylanase enzyme was fused on both extremities with either two SpyTag peptides, two SpyCatcher domains or a Spytag and a SpyCatcher domains. All the three proteins, expressed and secreted from *Bacillus subtilis*, displayed a xylanase activity. Strains expressing the endoxylanase fused with two SpyTag or two SpyCatcher domains were grown in cocultures, and the endoxylane were found to be dimeric, proving that the conjugation is also possible in culture supernatant (Gilbert *et al.*, 2017).

Jo-In

The Jo-In system is based on the two RrgA adhesin domains of *Streptococcus pneumoniae*. It consists of two proteins, Jo and In, of similar size (10.5 and 16 kDa, respectively). The operating principle is similar to the Spytag/Spycatcher system, with the spontaneous formation of an isopeptide bond between a lysine and an aspartic acid or an asparagine of the two domains.

The Jo-In system was used to study the role of the spatial organisation for glycoside hydrolase in the cellulosome. Indeed, the effect of the cellulosome scaffold is usually explained by synergic effect of the different enzymes brought together, and not by substrate channeling as it can sometimes be described in other types of enzyme assembly. The Jo and In peptides were thus fused to two glucoside hydrolases, the endo-β-1,4-xylanase GH11A from *Neocallimastix patriciarum* (NpXyn11A) which releases oligosaccharides from xylan, and the β-1,4-xylosidase GH43 from *Bacillus halodurans* (BhXyl43) which further hydrolyses the short oligomers. The fusions did not alter the kinetic parameters of the individual enzymes, however, the enzyme complexe performance was impacted in both enzymatic efficiency and product profile, putatively due to steric hindrance. By covalently associating the two glucoside hydrolases, the authors increased the amount of hydrolytic events (Enjalbert *et al.*, 2020).

Proliferating cell nuclear antigen (PCNA)

The PCNA is a ring-shaped trimeric sliding clamp protein. It opens and closes around DNA and act as a scaffold for other DNA-related enzymes during DNA replication and repair processes. The PCNA from *Sulfolobus solfataricus* form an heterotrimer made of PCNA1, PCNA2 and PCNA3 (Vanderstraeten and Briers, 2020) (**Figure 12**a.).

The three PCNAs domains were fused to the soluble P450 of *Pseudomonas putida*, the putidaredoxin (PdX) and the putidaredoxin reductase (PdR) (**Figure 12b.**). The fusion at the C-terminus of PCNA did not affect the activity of the PdR, PdX and P450cam and generated a 50-fold increase of the monooxygenase activity in comparison to the three independent enzymes in an equimolar mixture (Hirakawa and Nagamune, 2010). The optimization of the length and flexibility of the linker between the PCNA domain and the enzymes further enhanced the activity by a 1.9-fold (Haga, Hirakawa and Nagamune, 2013). One of the major downside of the PCNA is due to its non-covalent binding nature causing dissociation of the complex at low protein concentration. By shifting to the more stable PCNA complex of *Metallosphaera sedula*, the specific monooxygenase activity increased by 2.1-fold compared to the one fused to the PCNA of *S. solfataricus* (Iwata, Hirakawa and Nagamune, 2018).

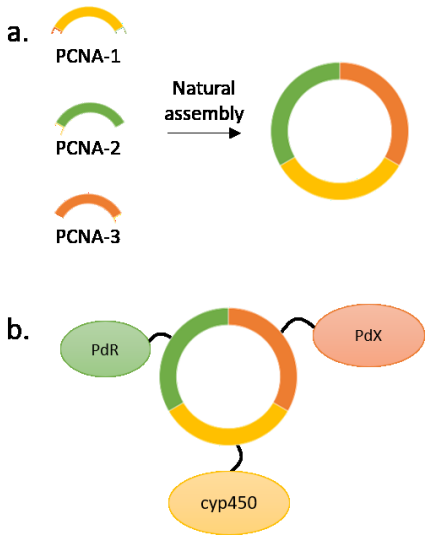


***Figure 12**: PCNA trimer. a. In natural context. b. As a scaffold for enzyme assembly.*

5. Focus on protein fusion

History

Inspired by natural multidomain protein evolved from separate enzymes, researchers started to compare bifunctional enzymes to their individual counterparts. One of the first study concerned the histidine bacterial gene operon where a mutation in the intergenic region of the mRNA led to bifunctionnal protein. The bifunctional protein was able to catalyse both reactions usually catalyzed by the individual enzymes, despite a decreased in their activity (Yourno, Kohno and Roth, 1970). A bifunctional enzyme performing two successive reactions was later designed by gene fusion of the β-galactosidase and galactokinase enzymes. Similarly, the *in vitro* assay showed a protein performing both reaction but at the cost of impaired activity (Bülow, Ljungcrantz and Mosbach, 1985).

The development of protein fusion has grown with the rise of reporter gene in research to detect spatial or temporal expression of protein, first with the β-galactosidase gene and, later, with fluorescent genes such as luciferases or fluorescent proteins (Riggs, 2013). Protein fusion have also been developed to increase the solubility of recombinant protein or to purify protein of interest. For exemple, in 2015, there were 102 reported structures of protein fused to the maltose binding protein in the PDB databank (Waugh, 2016). The maltose binding protein is often used to increase the solubility of protein expressed in *E. coli*.

Application

Application of protein fusion have now evolved from the pure research purpose and are used to design enzymes with improved performances in metabolic engineering. The improvement can relate to an increased reaction rate, to the regeneration of cofactors or to the ability to redirect fluxes at branched point. To illustrate the different applications of protein fusion in metabolic engineering, an example will be given for each application.

In 2011, a fusion protein was constructed from the endogenous farnesyl diphosphate synthase (FPPS) of yeast and a recombinant patchoulol synthase (PTS) of plant origin. Indeed, the FPPS catalyzes the production of farnesyl pyrophosphate (FPP), which is also the substrate for several other enzymes in *S. cerevisiae*. The expression of the fusion protein increased the production of patchoulol, the main sesquiterpene produced by PTS, up to 2-fold (Albertsen *et al.*, 2011).

An increased reaction rate was also observed for enzyme fusion in the case of slow catalytic enzymes. Lactose is cleaved by CelB into one molecule of galactose and one molecule of glucose. GalL catalyzed the conversion of galactose to galactose-1-phosphate and GluK the conversion of glucose to glucose-6-phosphate. Fusing the enzymes together gave a 3-fold advantage to the fused enzymes compared to independent enzymes, in the case where the first enzyme of the pathway was inhibited, and thus very slow (Patterson *et al.*, 2014).

Protein fusion also help to regenerate cofactor and this has been exemplified a lot with fusion of P450 with their redox partners. There is a natural $P450_{BM3}$ which is covalently fused to its partner enzymes. The $P450_{BM3}$ catalyzes fatty acid hydroxylation at rates at least two orders of magnitude faster than eukaryotic fatty acid hydroxylases (Munro, Girvan and McLean, 2007). Engineered P450 can also have increased reaction rate or higher electron transfer efficiency compared to separate enzymes, however, the yield still remain far from the natural P450 fusions.

Despite improvements in the design of protein fusion, substrate channeling and acceleration of the catalytic rate has rarely been demonstrated for synthetic protein fusion. Only one team was able to rationnally design a protein fusion able to substrate channel and have an increased reaction rate compared to the independent enzymes (Kummer *et al.*, 2021). An enzyme fusion was created between an alcohol dehydrogenase (ADH) and an aldehyde dehydrogenase (aldDH), allowing the conversion of acetate into ethanol, and *vice versa*. To begin the fusion protein design, the authors used a protein secondary structure prediction software to design a rigid linker forming an alpha-helix with as many cationic amino acids as possible. It resulted in a linker with a succession of lysine and arginine and one polar residue, the glutamine, to allow the formation of the helix. Next, they wanted to position the linker as close as possible to the catalytic domains without disrupting its activity. They found the best structure using Rosetta software. In the end, a final fusion protein was designed as [aldDH C-463]-KKRQKKKRK-[N-12 ADH], and the fusion displayed an improved conversion efficiency compared to the unlinked form.

It also happened spontaneously when coupling the glucose consumption pathway to glycerol production in *E. coli* (Meynial Salles *et al.*, 2007). The glucolysis pathway lead to the production of fructose-1-6-diphosphate which is cleaved into one molecule of dihydroxyacetone-P and one molecule of glyceraldehyde-3-P by endogenous enzymes. Enzymes from competitive pathways were deleted and the glycerol-3-P dehydrogenase (GPD1) and Glycerol-3-P phosphatase (GPP2) from the yeast *S. cerevisiae* were added to convert the dihydroxyacetone-P into glycerol. Hence, any improvement in glycerol production led to a higher glucose consumption and a higher growth rate. In culture, such strain naturally produced a fused enzyme between GPD1 and GPP2 enzymes with a 44 bp deletion. The protein fusion was able to partially substrate channel and had an increased catalytic reaction rate.

Linkers

Inside a protein fusion, the linker is the peptide connecting the two proteins moieties giving space for both domains to fold properly while allowing interaction between the protein domains. Based on the studies of linker in natural multidomain proteins, a few rules for the design of the linkers have been determined. Linker should not contain much hydrophobic residues and favor polar amino acids instead (Argos, 1990). Because they represent a major design step of protein fusion, several studies have focused on the impact of linker parameters on protein fusion. For example, one team studied the role of linker flexibility in protein fusion by designing a cyan fluorescent protein and a yellow fluorescent protein fusion (Van Rosmalen, Krom and Merkx, 2017). A FRET analysis of linker with various size and flexibility showed that efficiency decreased with an increasing linker length and was overall lower for linkers with less glycine. The observed efficiency was coherent with linker modeling.

In the same efforts to provide ressources to design protein fusion, several databases have been set up to help researchers find an appropriate linker for their protein fusion. In 2000, a first team set up a linker database based on loop sequences extracted from protein databank, and confirmed by DSSP program, an algorythm which can define secondary structures. The dataset was constituted after removing hairpin loop and loops of less than four amino acids. Users could then specify characteristics to get proposed a corresponding set of linkers (Crasto and Feng, 2000). SynLinker was another linker database of 2260 linkers comprised of natural linkers (2150) extracted from a non redundant set of multidomain proteins from the Protein Data Bank or artificial linkers (110) selected from publications and patents. The linkers could then be filtered by the user based on five criteria: the amino acid length, the solvent accessibility, the terminal amino acids, the composition of the linker and possible proteolytic site (Liu, Chin and Lee, 2015). Although both of these databases are not available anymore, the IBIVU database, similar to the previous ones, is still online (https://www.ibi.vu.nl/programs/linkerdbwww/). In addition to the linker parameters, the search engine also returns the PDB code of the multidomain protein from which the linker sequence has been extracted (George and Heringa, 2002).

After linkers databases, efforts have been made regarding the cloning of linkers between protein domains. Indeed, due to the repetitive sequence of artificial linkers (glycine repeats for example) and sometimes to the amount of linkers to be tested, the cloning step can be time-consuming. A first method has been elaborated, iFlinkC, the iterative functional linker cloning (Gräwe *et al.*, 2020). A collection of plasmids with linkers or protein domain delimited by type IIS restriction sites is necessary. Type IIS restriction enzymes cut the DNA outside of their recognition sites and generate DNA fragments with unique overhangs. By designing compatible overhangs for linkers and protein domains, a restriction/ligation cycle regenerates a plasmid with both domain and linker, with restriction sites similar to the starting plasmids. This allow the cycle to be repeated infinitely in order to combine several protein domain and linkers. However, the reiteration process require several cycles of digestion and ligation. A second method has been developed to generate a linker library, protaTETHER (Norris and Hughes, 2018). The variable linker sequence is encoded by a reverse set of primers. The forward primers has multiple annealing site resulting in various PCR amplicons. The PCR amplicons are then annealed to the plasmid containing the DNA of the protein domains to be fused. The plasmid can then be transformed.

The wide range of methods existing to spatially organise enzyme is quite benificial to allow us to choose the best method for different system. Interestingly, the pathways of the resveratrol and of the hydrogen have been engineered to enhance proximity of enzymes by two different methods, either by scaffolding or by direct protein fusion. For the resveratrol pathway, the scaffold strategy (Wang and Yu, 2012) led to a 5-fold increase in the resveratrol yield, whereas the protein fusion only help increase the final yield by a 1.8 times (Guo *et al.*, 2017). However, for the production of hydrogen, a DNA scaffold only increased the final yield by 3-fold (Agapakis *et al.*, 2010) when the protein fusions increased it by 4.4-fold. Probably due to more favorable interactions or more favorable enzyme ratio in the protein fusion.

III. The carotenoid pathway as a model

Isoprenoids are the largest group of natural molecules produced in living organisms. They are a highly diverse secondary metabolites group with endless applications in food, cosmetic and health industry (Perveen, 2018). They are made up of the 5-carbon (C5) isoprene units – dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP). The C5 building unit is then assembled to form terpenes of various sizes, hemiterpenes (C5), monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20), triterpenes (C30) and tetraterpenes (C40). While terpenes are simple hydrocarbures, they are often functionnalized with oxygen-containing groups to form terpenoids.

Carotenoids are a large part of the tetraterpenes family and widely found in photosynthetic systems such as plants, bacteria or algae. Tetraterpenes carotenoids represent 95% of carotenoids but the diverse family also include C30 and C50 molecules. Carotenoids are divided in two groups, xanthophylls and carotenes. The latter are hydrocarbons made only of carbons and hydrogens, such

as lycopene and β-carotene. The former (oxygenated derivatives) also include hydroxyl groups to form carotenoids such as zeaxanthin or astaxanthin (Thomas and Johnson, 2018).

Due to their long unsaturated aliphatic chain, carotenoids are mainly lipophilic and found in the membrane bilayer. The presence of conjugated double bounds confer them a visible color from yellow to red. The color varies based on the number of conjugated double bonds. On top of their pigment characteristic, they also play a huge role as antioxidant.

1. Metabolic pathway of carotenoids

The isoprenoids precusors IPP and DMAPP are produced by two independent pathways, the mevalonate (MVA) pathway, and the non-mevalonate pathway also known as the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway or the 1-deoxy-D-xylulose 5-phosphate (DXP) pathway (**Figure 13**). *E. coli* and most bacteria have the non-mevalonate pathway which uses glycerol-3-phosphate and pyruvate as precursors while eukaryote have the mevalonate pathway (MVA) which uses acetyl-CoA as precursor (Wang, Quan and Xiao, 2019). The production of carotenoids in *E. Coli* often require the addition of an exogenous MVA pathway to complement the endogenous MEP pathway.

The isoprene C5 units are successively condensed together by IspA enzyme to form the C15 farnesyl pyrophosphate (FPP). CrtE catalyzes the addition of another C5 unit onto FPP to form the geranylgeranyl-pyrophosphate (GGPP) diterpene. Lastly, CrtB enzyme condenses two GGPP molecules into one molecule of phytoene. The phytoene desaturase CrtI catalyses the four desaturations leading to the production of lycopene (**Figure 13**). Due to its eleven conjugated double bonds, lycopene displays a bright red color and is the first visible carotenoid of the pathway.

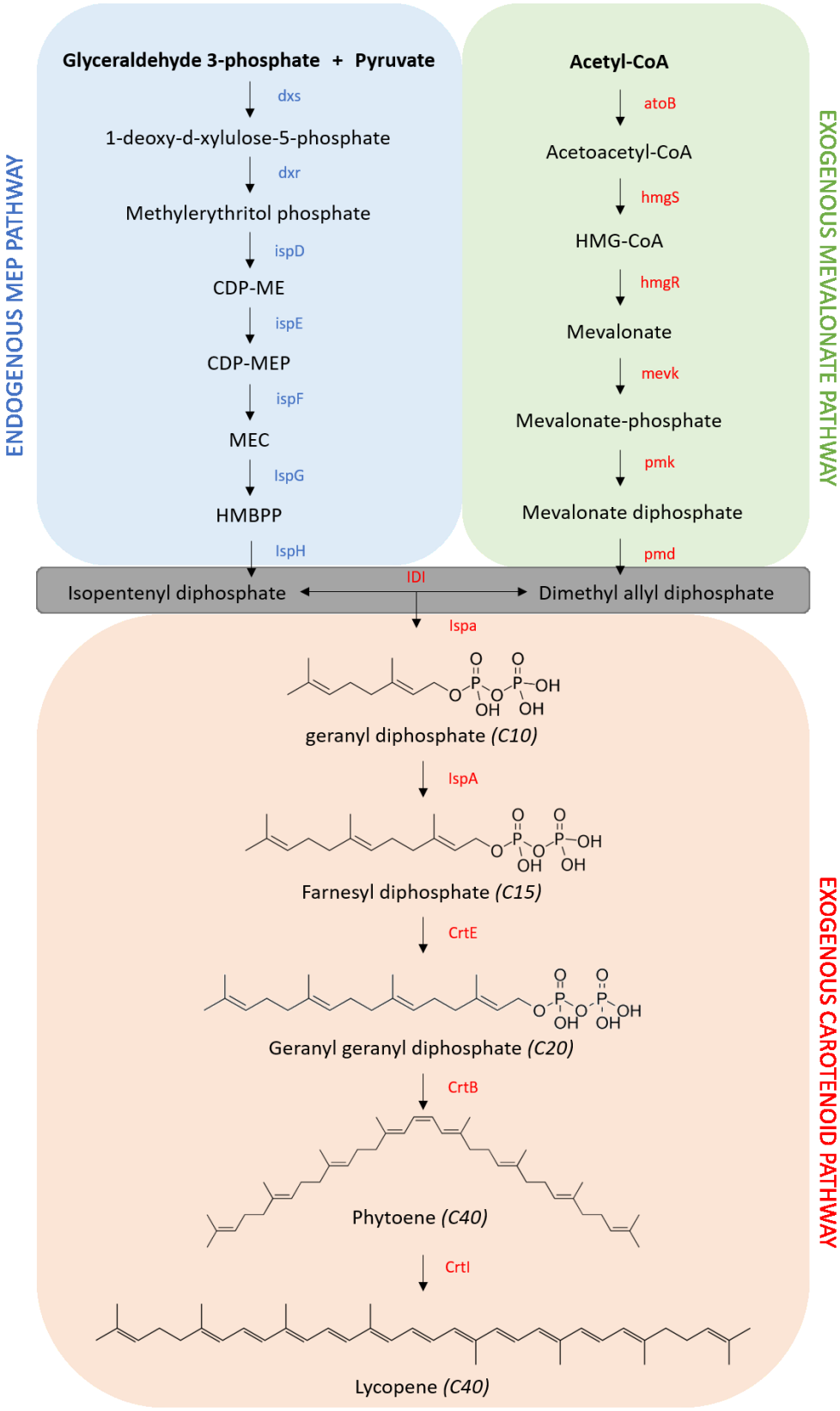


Figure 13: isoprenoid pathway for the biosynthesis of carotenoids in E.coli. CDP-ME: 4-diphosphocytidyl-2C-methyl-d-erythritol; CDP-MEP: 4-diphosphocytidyl-2C-methyl-d- erythritol-2-phosphate; MEC: 2C-methyl-d-erythritol-2,4-cyclo-diphosphate; HMBPP: 4-hydroxy-3-methyl-2-(E)-butenyl-4-diphosphate; HMG-CoA: 3-hydroxy-3-methyl-glutaryl-coenzyme.

Lycopene is cyclised by lycopene cyclases, adding a ring to each extremity of the lycopene. If both cyclisation are performed by a lycopene β-cyclase, β-carotene will be formed, whereas if one of the cyclisation is performed by a lycopene ε-cyclase, α-carotene will be formed (**Figure 14**). α-carotene is the precursor of lutein, an important carotenoid known for its anti-inflammatory properties and its ability to prevent age-related macular disease (Buscemi *et al.*, 2018). β-carotene rings can also be functionalized by differents enzymes to produce different carotenoids. For exemple, a ketolation by a β-carotene ketolase on position 4 of one of the β-carotene ring produces echinenone, while two ketolations on position 4 and 4' of the β-carotene rings produced canthaxanthin.

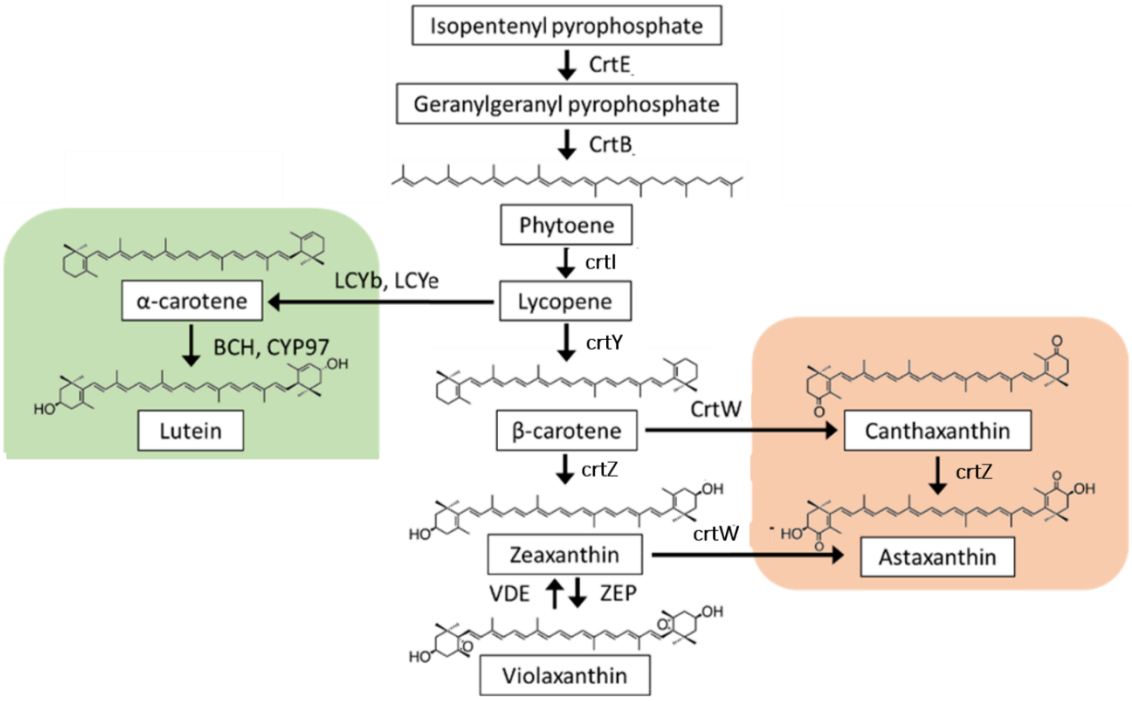


***Figure 14**: Carotenoids structure and biosynthesis. CrtE, geranylgeranyl pyrophosphate synthase; CrtB, phytoene synthase; CrtI, phytoene desaturase; LCYb and CrtY, lycopene β-cyclase; LCYe, lycopene ε-cyclase; CrtZ carotene hydroxylase; ZEP, zeaxanthin epoxidase; VDE, violaxanthin de-epoxidase; CrtW, β-carotene ketolase. Adapted from (Tamaki, Mochida and Suzuki, 2021).*

2. Spatial organisation of the carotenoid pathway in metabolic engineering

Carotenoids such as lycopene, β-carotene, canthaxanthin, zeaxanthin, lutein and astaxanthin are industrially relevant (Barreiro and Barredo, 2018) for use as nutritional supplement and feed supplementation to enhance skin and flesh colors of animals (Martínez-Cámara *et al.*, 2021). Because

of their high market value, the production of carotenoids by metabolic engineering of microorganisms has been extensively studied. Indeed, metabolic engineering offers a sustainable alternative to common chemical synthesis, and have usually higher biological property because of a more favorable cis / trans isomers mixture. Several reviews described the efforts made in the production of astaxanthin (Wan *et al.*, 2020), canthaxanthin (Rebelo *et al.*, 2020) and other carotenoids or in the general strategies applied for the production of carotenoids

The optimization method of the carotenoid pathway relying on a modification of the spatial organization can be categorized in four main strategies: targeting of enzymes to organelles, expand membrane volume to increase storage space available for the carotenoids produced, assembling enzymes together and lastly, organize the export of carotenoid out of the cells.

Targeting to organelles

In 2009, crtE, crtB and crtI were successfully targeted to the peroxisome of *Pichia pastoris* by fusion with peroxisomal targeting sequence. However, the location of enzymes did not impact carotenoid production as both strains with peroxisomal or cytosolic enzymes produced between 12 to 14 mg/L of lycopene (Bhataya, Schmidt-Dannert and Lee, 2009).

However, expression of astaxanthin pathway to subcellular organelles of the oleaginous yeast *Y. lipolytica* increased astaxanthin titer and decreased the accumulation of the ketocarotenoids intermediates (Ma *et al.*, 2021). Indeed, when the pathway was expressed to the lipid body, the ER or the peroxisome, a 1.62-fold, 1.84-fold and 2.03-fold increase in astaxanthin production were respectively observed when compared to the cytosolic enzymes. When the three different targeting strategies were combined, a synergetic effect was observed and astaxanthin titer increased by 4-fold.

Expansion of membrane volume

Increasing membrane volume to enhance the production of carotenoids is a strategy that has been tested in *E. coli*, *S. cerevisiae*, and *Y. lipolytica* for the production of various carotenoids. As carotenoids are lypophilic molecules, the increase of membrane volume provides more storage space for the carotenoids produced. This can in turn lower the toxicity of the carotenoids towards the cell.

In *E. coli*, this strategy has been implemented for the production of β-carotene (Wu *et al.*, 2017) and lycopene (Wu *et al.*, 2018). The surface area of membrane was increased by overexpression of membrane bending protein as well as overexpression of the glycerophospholipids pathway. A synergetic effect was observed when both overexpression were combined and yielded a 2.9-fold increase in β-carotene titer and 1.32-fold increase in lycopene titer. When implemented in a high producing strain, these changes increased the final carotenoid titer by 39% in both cases.

In *S. cerevisiae*, lycopene , β-carotene and astaxanthin were produced in strains with modified expression of its lipidic compartments. Overexpression of INO2 transcription factor involved in the synthesis of sterol and phospholipids in *S. cerevisiae* resulted in a 10% improvement of lycopene production (Chen *et al.*, 2016). Through the regulation of the synthesis, size and degradation of lipid droplets, the yield of β-carotene increased by 34% (Bu *et al.*, 2022). This effect was boosted with the addition of external oleic acid which promotes production of triacylglycerol involved in the formation of lipid droplets and thus allow to direct more acetyl-CoA for β-carotene production, as acetyl-CoA is a common precursor of triacylglycerols and β-carotene. Moderate upregulation of lipid synthesis and expension of lipid droplets yielded 9.79 mg/g DCW of astaxanthin (Li *et al.*, 2022).

In *Y. lipolytica*, the strain that overproduce lipid has a 1.93 higher expression of β-carotene than the wild-type strain (Larroude *et al.*, 2018).

Enzyme assembly

On top of relocalizing enzymes in the cell, enzymes can also be assembled together to increase pathway efficiency. As described in the previous section, the assembly of enzymes can be mediated by a scaffold, by interacting peptides allowing post-translational assembly or by protein fusion.

Two types of scaffold were used for the production of carotenoids in *E. coli* in 2022 and 2023.

First, lutein was produced using CipA and CipB scaffold protein (Park *et al.*, 2022). Lutein pathway from lycopene require four enzymes, a lycopene β-cyclase (LCYB) as well as a lycopene ε-cylase (LUT2) that forms an asymetric α-carotene molecule. The α-carotene is then functionalysed by a β-carotene hydroxylase (LUT5) and a carotene ε- monooxygenase (LUT1) to form lutein. The β-carotene hydroxylase adds a hydroxyl group to the β-ring of the α-carotene while the carotene monooxygenase adds a hydroxyl group to the ε-ring of the α-carotene. The expression of the four enzymes responsible for the production of lutein in a lycopene producing strain did not yield any lutein. It was hypothesized that the metabolic flux where diverted from α-carotene to β-carotene by the LCBY promiscuous enzyme. By fusing LUT2 and LCYB enzymes to CypB scaffold protein, enzymes were able to cluster and produce 0.84mg/L of lutein. Moreover, the LUT5 and LUT1 enzymes are cytoplasmic P450 requiring the p450 reductase (ATR2) localised in the membrane. Thus, a second enzymatic cluster was formed using the CipB scaffold protein, which is totally independent of the CipA scaffold protein. The strain containing the clustered LUT5, LUT1 and ATR2 enzymes had a 3.41-fold increase in the production of lutein.

Next, in 2023, rare carotenoids capsanthin, capsorubin, cucurbitaxanthin A, and capsanthin 3,6-epoxide were produced using a scaffold system based on the S-tag and S-binding protein (Hattan *et al.*, 2023). From the zeaxanthin carotenoid, zeaxanthin epoxidase (ZEP) catalyses the formation of violaxanthin in two steps and the capsorubin synthase (CCS) can catalyses the formation of the four different carotenoids. Previously, the expression of CCS with ZEP in cells, only produced low amount

of capsanthin. The S-protein can oligomerized to form a protein scaffold. Both enzymes individually fused to the S-tag increased interaction between ZEP and CCS enzymes and led to the formation of the four products for the first time in *E. coli*.

Less complex than scaffold, peptide interactions were used to assemble CrtE and Idi enzymes. Post-translationnal fusion between CrtE and Idi enzymes allowed an increase in final astaxanthin titer by 2.7-fold and a 5.7-fold increase in total carotenoids production (Kang *et al.*, 2019). Indeed, CrtE is a membrane bound enzyme catalysing the last reaction of the mevalonate pathway, which is the starting point of numerous other pathway, whereas Idi is a cytosolic enzyme catalysing the first reaction of the β-carotene pathway. The protein fusion allows the re-localisation of CrtE to the membrane and prevents the loss of IPP and DMAPP precursor toward competitive pathways. When the system was transfered to *S.cerevisiae*, a 58% increase in lycopene production was observed.

Protein fusion are also largely used to enhance the production of different carotenoids molecules. Fusion have been designed between enzymes of the downstream mevalonate pathway in order to guide the carbon flux towards the production of carotenoids instead of other endogenous pathway, but also for the production of β-carotene, zeaxanthin and astaxanthin, which production involves numerous intermediates.

Similarly to the CrtE and Idi enzyme assembly, the GGPP synthase was fuse to the phytoene synthase to avoid the consumption of GGPP by other terpenoid pathways and concentrate the flux of GGPP toward the production of phytoene in plants (Camagna *et al.*, 2019). The synthetic fusion of GGPS and PSY enzyme was tested both *in vitro* and *in vivo*. *In vitro*, all the GGPP was converted into phytoene, and the orange colored cotyledons in *Arabidopsis thaliana* in presence of the enzyme fusion indicated an increase in the production of carotenoids, compared to the independently expressed enzymes.

A tridomain protein fusion was designed for the production of β-carotene in the yeast *S. cerevisiae* (Rabeharindranto *et al.*, 2019). Indeed, in *Xanthophyllomyces dendrorhous*, the phytoene synthase activity and the lycopene cyclase activity are gathered in a single CrtYB protein. The two catalytic domains of the CrtYB protein as well as the phytoene desaturase CrtI were fused together in different succession to form a tridomain protein fusion. The yield of β-carotene doubled with the protein fusion while the precursor accumulation decreased, leading to an improved pathway efficiency as compared with the natural system.

For the production of zeaxanthin in *E.coli*, the authors started from a strain producing 33.43 mg/gDCW of lycopene in *E.coli* obtained by chromosomal modifications only. By expressing CrtY and CrtZ genes from *Pantoea ananatis*, the strain accumulated β-carotene only. To verify if the expression of CrtY and CrtZ genes was balanced, a CrtY-CrtZ chimera was designed: CrtY-(GGGGS)3-CrtZ. The enzyme fusion was able to produce 1.54 mg/g DCW of zeaxanthin confirming that balancing CrtZ and CrtY expression was key for the production of zeaxanthin (Li *et al.*, 2015).

The largest example of enzyme assembly for the production of carotenoids has been enzyme fusions for the production of astaxanthin. Astaxanthin is formed from β-carotene by four successive reactions catalysed by two enzymes, the β-carotene hydrolase (CrtZ) and the β-carotene ketolase (CrtW). These two enzymes yield ten possible intermediates between the β-carotene and astaxanthin.

The fusion attemtps of the different CrtW and CrtZ enzymes for the production of astaxanthin carotenoid are summarized in **Table 2**. The fusion of both CrtZ and CrtW usually yield an increase in the production of astaxanthin (between 1.4-fold to 5-fold) as well as a reduction of intermediate and/or precursor accumulation. As illustrated in the table, the enzyme orientation in the fusion is always critical to obtain a catalytically active enzyme. When tested, the linker flexibility did not seem to impact the protein fusion, however, the results seem contradictory for the impact of the size of the linker. It is interesting to note that in all fusion the enzymes used come from different organisms.

***Table 2**: comparison of CrtZ and CrtW enzyme fusion for the production of astaxanthin.*

Host organism	crtW	crtZ	Astaxanthin yield	Enzyme orientation	Linker size	Linker flexibility	Reference
E. coli	*Brevundimonas sp.*	*Pantoea agglomerans*	*2.15	Yes	-	-	(Ye *et al.* , 2018)
E. coli	*Brevundimonas sp.*	*Alcaligenes sp.*	*2.28	Yes	Yes	No	(Y. Wu *et al.* , 2019)
E. coli	*Brevundimonas sp*	*Brevundimonas sp*	*1.4	Yes	No	No	(Nogueira *et al.* , 2019)
C. glutamicum	*Fulvimarina pelagi*	*Fulvimarina pelagi*	*5	Yes	-	-	(Henke and Wendisch, 2019)
Y. lypolitica	*Paracoccus sp*	*Haematococcus pluvialis*	*2.8	Yes	-	-	(Ma *et al.* , 2021)
S. cerevisiae	*Brevundimonas vesicularis*	*Agrobacterium aurantiacum*	*1.6	Yes	Yes	-	(Ding *et al.* , 2022)

Yes: the enzyme orientation or linker parameter impact the production of astaxanthin.
No: the enzyme orientation or linker parameter does not impact the production of astaxanthin.
- : not tested

Secretion and export of carotenoids

Increase in carotenoid can be observed when the volume of membrane is expended. This is usually understood as membranes have a limited space and that high production of carotenoids can saturate membrane. Well, some strategies, instead of focusing onto the increase of membrane space (which will also display a limit, although higher), have focused on the extraction of the carotenoids from the membrane to transfer them into the extracellular space and thus free the space from carotenoids, thus removing the limit.

Extraction of the carotenoids from the membrane can be done either through transporter or by excretion of the carotenoids in vesicles. Due to the lipophilic nature of the carotenoids, the strain

expressing the transporter should be cultured in a media under a layer of dodecane for example, which will allow for the capture of carotenoids without being toxic to the cell.

Two transporters have been identified in yeast as able to transfer carotenoids from the membrane to the extracellular space. In *S. cerevisiae*, an ABC transporter was used to increase the secretion of B-carotene. The addition of the transporter in the strain was coupled with engineering strategies in order to increase supply of ATP as ABC transporters require ATP to export molecules as well as strain engineering to promote membrane flexibility as overexpression of membrane protein in inner membrane tend to rigidify membrane. The combination of these three methods yielded a 5.80-fold increase of β-carotene secretion and a 1.71-fold increase of intracellular β-carotene production compared to the starting strain. The transporter was identified by proteomics because it was being upregulated during expression of B-carotene (Bu *et al.*, 2020). In the oleaginous yeast *Rhodosporidium toruloides*, the expression of the ptdr10 transporter increased the export of carotenoids and enhanced the production of carotenoids from 1.9 to 2.9 ug/ug and 1.8 ug/mg of carotenoids were efficiently exported (Lee *et al.*, 2016).

In *E. coli*, several transporters can export different carotenoids. The MsbA transporter from *Salmonella enterica* increased zeaxanthin export by 2.4-fold and lycopene export by 4.3-fold. The MsbA transporter from *E. coli* increases both canthaxanthin and β-carotene export by 4.4-fold. The production of carotenoids was not affected by the expression of the transporter. The identified MsbA transporter is also an ABC transporter and is responsible for the transfer of lipid A from inner membrane to outer membrane of *E.coli* (Doshi, Nguyen and Chang, 2013).

On top of the MsbA transporter, *E. coli* has also been shown to be able to excrete β-carotene by secretion by an artificial membrane vesicle trafficking system (AMVTS). Indeed, gram negative bacteria shed membrane components through outer membrane vesicules. However, this process is very limited and has been increased for the AMTVS purpose by knocking out or surexpressing genes involed in these membrane vesicles. To compensate the loss of membrane components, the pathways involved in the biosynthesis of membrane components was overexpressed. Using this strategy, the amount of β-carotene exported increased by 71.5-fold and the production of β-carotene was improved by 3-fold (T. Wu *et al.*, 2019).

3. Zeaxanthin

Generalities

Zeaxanthin is a terpenoid molecule of forty carbons. The two ends are rings hydroxylated on their 3 and 3' positions (**Figure 15**).

***Figure 15**: Molecule of zeaxanthin.*

It is used as a nutraceutical supplemented for human eye health and also as food additive for fish and poultry as it gives a healthy yellow colour to flesh and egg yolk for exemple. Actual production of zeaxanthin mainly relies on solvent extraction from plants producing carotenoids or chemical production. However, there is an alternative way to these methods which is the microbial production of zeaxanthin, either by microorganisms naturally producing zeaxanthin (Y. Zhang *et al.*, 2018) or by metabolic engineering of microorganisms such as *E. coli*, S. *cerevisiae*, *Y. lypolitica* or X *dendrorhous*.

The production of zeaxanthin from lycopene involves two enzymes, CrtY and CrtZ.

CrtY enzyme

The lycopene cyclase (CrtY) from *Pantoea ananatis* catalyzed an FADred-dependent non-redox reaction (Yu *et al.*, 2010). It is responsible for the cyclisation of both extremities of the lycopene. Only

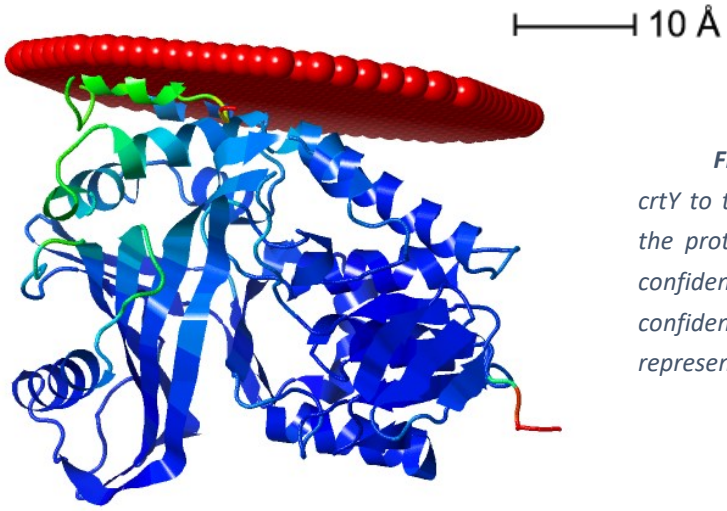


***Figure 16**: Predicted anchoring of crtY to the membrane. Color gradient of the protein from red to blue represents confidence score of Alphafold from low confidence to high confidence. Red dots represent the membrane.*

little is know about the catalytic reaction performed by CrtY (for example, its active site has not been elucidated yet), however, it is predicted to be membrane bound (**Figure 16**) (Yu *et al.*, 2010).

CrtZ enzyme

The enzyme responsible of the hydroxylation of both β-carotene rings in *Pantoea ananatis* is a β-carotene hydroxylase (CrtZ). Two main groups of β-carotene hydroxylases exist, the cytochrome p450 monooxygenases and the non-heme di-iron hydroxylases that are related to the fatty acid desaturases (Martín, Gudiña and Barredo, 2008). The β-carotene hydroxylase from *Pantoea ananatis* belongs to the latter. This family is characterized by the presence of four histidine motifs (HXXXXH or HXXHH) and the HDGLVHXRXP amino acid sequence, overlapping with one of the histidine motifs and called motif 1. The spacing between the four histidine motifs is conserved among the β-carotene hydroxylases of this family (**Figure 17**). There are involved in the binding of iron atoms and activation of dioxygen, making them essential for the enzyme activity. On top of the iron atoms, enzymes from this family require molecular oxygen, ferredoxin and ferredoxin reductase (Bouvier and Keller, 1998).

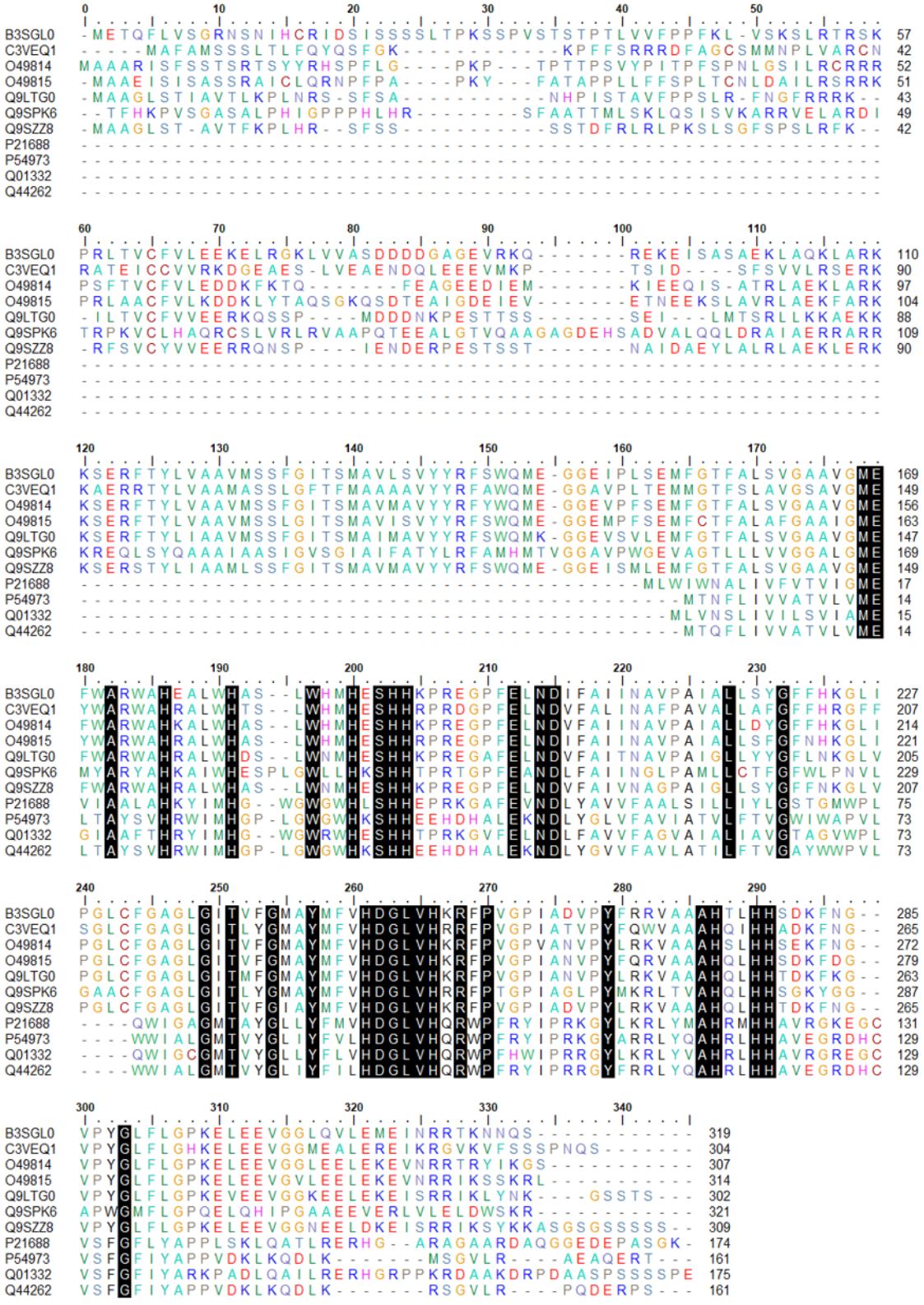


***Figure 17**: Alignment of ß-carotene hydroxylases. The sequences were retrieved from the PFAM fatty acid desaturases (PF04116) protein family. A clustal W multiple alignment was performed. Amino acids are colour coded based on their properties and conserved amino acids are highlighted in black.*

The non-heme di-iron hydroxylases / monooxygenases contains putative transmembrane

helices suggesting that they are integral parts of the membrane bilayer (**Figure 18**).

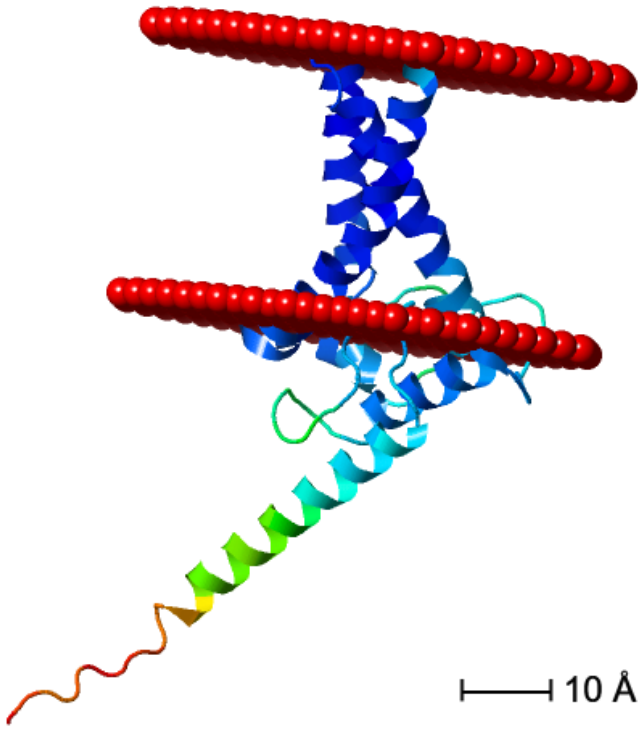


Figure 18: *Predicted position of crtZ in the membrane bilayer. Color gradient of the protein from red to blue represents confidence score of Alphafold from low confidence to high confidence. Red dot represent the membrane.*

The β-carotene hydroxylases CrtZ are promiscuous enzymes able to take different substrate as it was recently demonstrated by two groups (**Figure 19).** CrtZ from *Pantoea agglomerans* can synthesize rhodoxanthin using zeaxanthin as a substrate and this activity is increased when substituting F51I and A53P (Royer *et al.*, 2020). CrtZ from both *Pantoea agglomerans* and *Pantoea ananatis* can use violaxanthin or antheraxanthin epoxycarotenoids as substrate to form carotenoids with 6-hydroxy-3-keto-e-ends. The F52I/ A53P substitutions have been introduced in *Pantoea ananatis* and the mutant had a higher activity on 6-hydroxy-3-keto-e-ends (Furubayashi, Maoka and Mitani, 2022). In both studies, the biosynthesis of rhodoxanthin or 6-hydroxy-3-keto-e-ends is proposed to happen through ketolation, desaturation and double bond rearrangement. F52I and A53P substitutions could help increase the efficiency of the last step.

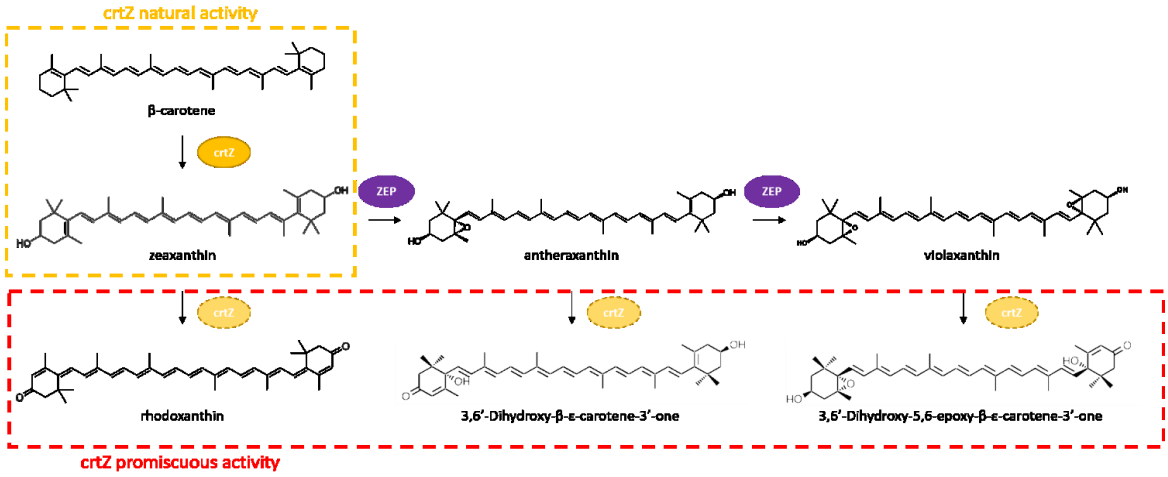


***Figure 19**: crtZ promiscuous activity, with the identification of natural and promiscuous substrates and products.*

SCIENTIFIC QUESTIONS AND OBJECTIVES

Spatial organisation of enzyme is a developping field of metabolic engineering. Spatial organisation refers to the localization of the enzyme inside of the microbial cell or to the position of enzymes toward each other. The underlying concept is that enzyme proximity allows for a more efficient flux by (i) protecting the cell from intermediate toxicity, or intermediate from degradation or diversion to competitive pathway or by (ii) confining enzymes in specialized structures thus providing an environment free of oxygen or enriched in cofactor for example. However, enzyme proximity does not enhance the kinetics of the enzymes but in very specific cases often not relevant in biological conditions.

Among the strategies existing to bring enzymes in proximity, targeting to organelles, targeting to protein-bound microbial compartment, scaffold (made of proteins, nucleic acids or lipids), molecular condensate, protein fusions are some of them. Protein fusion have been used for a long time, first in research as reporter genes or for purification purposes, to create bifunctionnal enzymes, regenerate cofactors, etc… They also have therapeutic application as Fc-based fusion proteins.

A few parameters well known to adjust for enzyme fusion are the orientation of the enzymes as well as the linker length and flexibility. However, design of protein fusions remain very empiric, with trial and error needed to successfully achieve a protein fusion.

Carotenoids are terpenoids with large diversity. Carotenoids are lipophilic molecules and carotenoids producing enzymes are thus at the interface between membranes and the cytosol. Moreover, promiscuity of enzymes can lead to variety of intermediates accumulated, even with a few enzymes. For example, from β-carotene, you can obtain ten different intermediates by only expressing two different enzymes, CrtW and CrtZ.

This work is the continuation of a previous project carried out in TBI lab where the carotenogenous enzymes CrtI and CrtYB from *Xanthophyllomyces dendrorhous* were fused as bi- or tri- fusion in *S. cerevisiae*. This led to an increase in the final product of β-carotene, while the accumulation of intermediates decreased. The objectives of my thesis were to investigate which parameters of the linker would impact a protein fusion and how protein fusion would impact the production of carotenoids. For this, we used the bacterial chassis *E.coli* since the singaporian team in SIFBI, A*STAR had just developed a lycopene producing strain with modules allowing the expression of various downstream carotenoids.

The result part is divided in two sections. The first section relate the fusion of two enzymes of the carotenoid pathway, CrtY and CrtZ. This enzymatic model has been used to try to establish rules of assembly for enzymes in a fusion. The second part is dedicated to the study of carotenoids repartition inside the cell.

RESULTS

I. Impact of linkers characteristics in zeaxanthin producing protein fusion

1. Introduction

Article summary

The main objective of my project was to understand how each characteristics of the linker in a protein fusion would impact the protein fusion. After choosing the CrtY and CrtZ protein from the carotenoid pathway as a model for this study, we first determined the optimal orientation of the crtY and CrtZ protein in a protein fusion, and confirmed that CrtZ should be placed at the N-terminus of the protein fusion to yield CrtZ-CrtY constructs. We then tested a wide range of linker between the enzymes to study the impact of the linker parameters. The size and the flexibility of the linker were found to impact the efficiency of the protein fusion, as well as the amino acid at the extremities of the linker, which was never reported before. The optimization of the linker parameter in the protein fusion leading to the production of zeaxanthin resulted in strains improved for the production of carotenoids or for zeaxanthin specificity.

Contribution

The lycopene producing strain used as chassis in this study was provided by C. Zhang. The linker sequences were retrieved from an online database. The method for high performance liquid chromatography (HPLC) were previously optimized. I realized all necessary molecular cloning, cultures and carotenoid extractions, as well as data analysis.

Reference

This chapter refers to the following publication: Aurélie Bouin, Congqiang Zhang, Nic D. Lindley, Gilles Truan, Thomas Lautier, Exploring linker's sequence diversity to fuse carotene cyclase and hydroxylase for zeaxanthin biosynthesis, Metabolic Engineering Communications, 2023, e00222, ISSN 2214-0301, https://doi.org/10.1016/j.mec.2023.e00222.

2. Article

Abstract

Fusion of catalytic domains can accelerate cascade reactions by bringing enzymes in close proximity. However, the design of a protein fusion and the choice of a linker are often challenging and lack in guidance. To determine the impact of linker parameters on fusion proteins, a library of linkers featuring various length, secondary structure, extension and hydrophobicity was designed. Linkers were used to fuse the lycopene cyclase (crtY) and β-carotene hydroxylase (crtZ) from *Pantoea ananatis* and to create fusion proteins to produce zeaxanthin. The fusion efficiency was assessed by comparing the carotenoids content in a carotenoid-production *Escherichia coli* strain. It was shown that in addition to the orientation of the enzymes and the size of the linker, the first amino acid of the linker is also a key factor in determining the efficiency of a protein fusion. The wide-range of sequence diversity in our linker library enables the fine tuning of protein fusion and this approach can be easily transferred to other enzyme couples.

Introduction

Carotenoids are tetraterpenoid pigments commonly found in bacteria, fungi, algae and plants. Carotenoids act as antioxidants and are responsible for light absorption in photosynthetic organisms (Zakynthinos and Varzakas, 2016). Among carotenoids, zeaxanthin is a xanthophyll giving corn or egg yolk their characteristic yellow colour (Sajilata, Singhal and Kamat, 2008). It is also known as colourant E161h in the food industry and used as a feed additive for fish and poultry (Sajilata et al., 2008). Moreover, zeaxanthin is also used as a dietary supplement for human eye health. Indeed, zeaxanthin and lutein are the only two carotenoids found in the vicinity of the human retina where they have a putative preventive effect against macular degeneration (Arunkumar, Gorusupudi and Bernstein, 2020). By 2030, the market demand for zeaxanthin is expected to reach US$ 210 Million (Zafar *et al.*, 2021).

Zeaxanthin can be produced by chemical synthesis. However, organic synthesis typically produces racemic mixtures, and only specific isomers present a biological activity, such as the 3R,3'R-zeaxanthin and 3R,3'S-RS- zeaxanthin in the retina (Mares, 2016). The environmental impact of the chemical synthesis is also incompatible with the required sustainable production, which needs to adopt an ecological integrative workflow as described the One Health paradigm from the World Health Organisation (Mackenzie and Jeggo, 2019). Zeaxanthin can also be extracted from natural producers such as the marigold flowers (Barreiro and Barredo, 2018) or maize (OLIVEIRA *et al.*, 2013) but this method is limited by its low yield. An alternative way to produce biomolecules is by microbial fermentation. Zeaxanthin can be produced in both natural and genetically engineered microorganisms (Y. Zhang *et al.*, 2018). Production of high-value compounds such as zeaxanthin via metabolic engineering is a serious alternative to the other two methods (extraction from natural producers and chemical synthesis) because microbial fermentation processes use renewable feedstocks and is safer

and environment-friendly (Rinaldi, Ferraz and Scrutton, 2022). Engineered microbial approach offers tools to optimise the production and to diversify the type of products. However, challenge remains to design a robust microbial strain, involving protein and metabolic engineering, and to establish an efficient bioprocess including fermentation and downstream product purification.

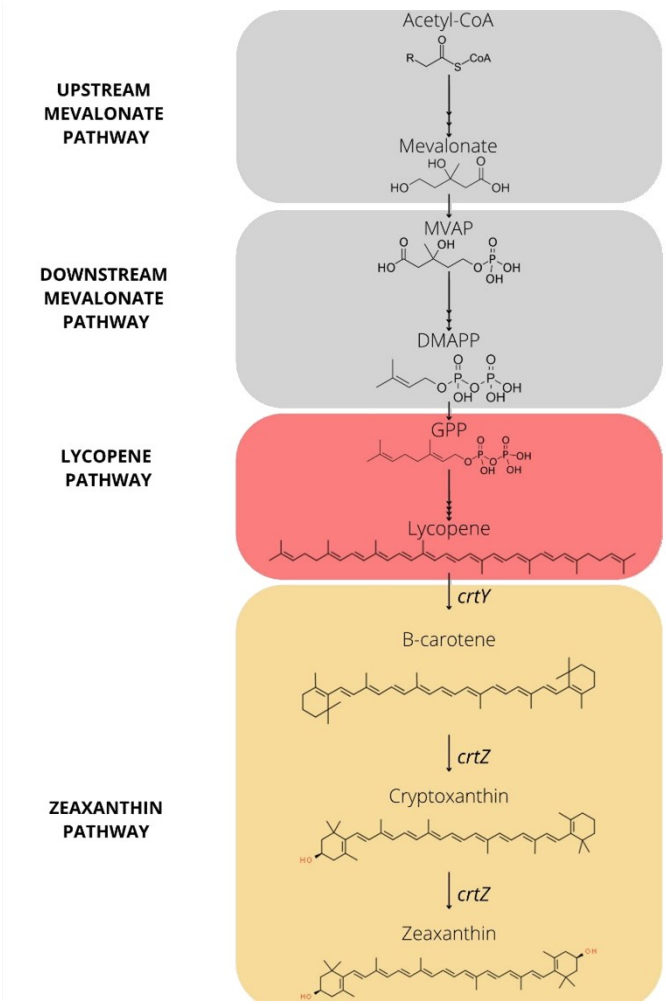


***Figure 20**: Metabolic pathway of zeaxanthin. Each block represents a module of a set of genes clustered on one plasmid (C. Zhang et al., 2018). MVAP: phosphomevalonate; DMAPP: dimethylallyl pyrophosphate; GPP: geranyl pyrophosphate.*

In the engineered *Escherichia coli* strain producing zeaxanthin, the two terpenoid precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are produced by implementing the heterologous mevalonate pathway to supplement the prokaryotic endogenous non-mevalonate pathway (**Figure 20**). The terpenoid building blocks are then successively assembled into lycopene and both extremities of lycopene are cyclised by the β-carotene cyclase (crtY) to form β-carotene. Zeaxanthin is produced from β-carotene by a hydroxylation step on each of the β-carotene rings, on position 3 and 3'. The enzyme responsible for the reaction is the β-carotene hydroxylase, crtZ.

Previous efforts for the production of zeaxanthin in *E. coli* include a study on the most efficient gene arrangement of zeaxanthin gene pathway on an operon (Nishizaki *et al.*, 2007), the use of

tunable intergenic region (TIGR) to adjust individually the expression of crtY and crtZ genes (Li *et al.*, 2015), the regulation of the mevalonate pathway using dynamically control TIGR approach (Shen *et al.*, 2016), the optimization of the initial codon in zeaxanthin pathway genes (Z. Wu *et al.*, 2019) and the multidimensional regulation of genes grouped into modules (Chen *et al.*, 2021). *Saccharomyces cerevisiae* (Carquet, Pompon and Truan, 2015), the red yeast *Xanthophyllomyces dendrorrhous* (Pollmann, Breitenbach and Sandmann, 2017; Breitenbach, Pollmann and Sandmann, 2019) and *Yarrowia lypolitica* (Xie, Chen and Xiong, 2021) have also been engineered for the production of zeaxanthin. These studies focus on the modulation of gene expression level to achieve pathway efficiency. However, β-carotene is never fully converted into zeaxanthin and the concentrations of zeaxanthin obtained are not economically viable yet.

A complementary approach to the transcriptional regulation is to improve the pathway efficiency through enzymatic spatial organisation, by bringing enzymes of the same metabolic pathway in close proximity. Indeed, spatial proximity of enzymes is thought to improve reaction velocity by reducing diffusion of intermediates and increasing local concentration of enzymes and substrates (Qiu *et al.*, 2018). Spatial optimization can be achieved at different scales. Microcompartments allow for the sequestration of enzymes and hydrogenases have previously been targeted in a repurposed carboxysome (Li *et al.*, 2020). Synthetic scaffolds made of proteins or nucleic acids are widely used to anchor enzymes (Geraldi *et al.*, 2021; Park *et al.*, 2022). Finally, at a scale limited to two or three enzymes, protein fusions allow to bring enzymes together in a one-to-one ratio (Elleuche, 2015).

Protein fusions are made of at least two protein domains. A linker joining both domains is often required to maintain a proper protein folding while allowing domain interactions (Wriggers, Chakravarty and Jennings, 2005). Linker characteristics were studied in natural multidomain proteins in terms of length, conformation and amino acid composition (Argos, 1990; George and Heringa, 2002). More recently, the role of the linker flexibility (Li *et al.*, 2016; Van Rosmalen, Krom and Merkx, 2017), cloning strategies for a wide selection of linkers between proteins (Norris and Hughes, 2018; Gräwe *et al.*, 2020) and linkers for membrane proteins (Sadaf *et al.*, 2016) were studied on artificial protein fusion, however, the impact of the linker's sequence on the fusion efficiency was not analysed and there is a lack of knowledge to predict which linker sequence will lead to a functional enzymatic fusion.

In this study, a protein-fusion approach was conducted to increase the production of zeaxanthin in an engineered *Escherichia coli* strain. To reduce the accumulation of the intermediate, β-carotene, the last two enzymes of the zeaxanthin pathway, crtY and crtZ were fused. The enzyme fusion was optimised by testing a collection of linkers, shortlisted according to idiosyncratic properties. In the end, we provided insight regarding the design of protein fusion and also discussed the limitations of the approach.

Materials and methods

Strain and plasmid

E. coli Bl21-Gold DE3 strain (Stratagene) was used in this study. The plasmids p15A-spec-hmgS-atoBhmgR, p15A-cam-mevK-pmk-pmd-idi, p15A-kan-crtEBI-ispA, p15A-crtYZ and p15a-crtY were obtained from a previous study (C. Zhang *et al.*, 2018) and the resulting lycopene producing strain was used as a platform for the production of zeaxanthin (**Figure 20**). p15Aamp-crtZY plasmid was obtained by insertion of CrtZ gene in p15A-crtY plasmid.

Construction of the linker library

The linkers were cloned based on a generic plasmid using a method inspired from golden gate cloning. To insert the linkers sequences, a SapI restriction site was inserted between the sequences of crtY and crtZ in p15Aamp-crtYZ and p15Aamp-crtZY plasmids. The digestion of the vectors by SapI enzyme from New England Biolabs (NEB) resulted in opened plasmids with 3 bases overhang on each side.

The creation of the linker library was achieved by designing primers combining the sequence of the linker plus three nucleic acids complementary to the one of the SapI overhang sequences. Primer couples were mixed into water, heated at 70°C for five minutes and slowly cooled down at room temperature to allow annealing of matching sequences and obtain double stranded DNA. The double stranded DNA of the linker, with 3 bp overhang, was then ligated with T4 DNA ligase in the SapI digested plasmids to obtain p15A-amp-crtY-linker-crtZ and p15Aamp-crtZ-linker-Y plasmids. Correct plasmids identified by sequencing were transformed in the *E. coli* K12 MG1655 producing strain with the three plasmids harbouring the mevalonate and lycopene pathway genes.

Culture conditions

After an overnight preculture in 2XPY medium (20 g/L peptone, 10 g/L yeast extract and 10 g/L NaCl), cells were inoculated at OD_{600} = 0.1 in 2XPY medium supplemented with 20 g/L glycerol, 50 mM 4-(2-hydroxyethyl)- 1-piperazineethanesulfonic acid (HEPES) and Tween 80 0.5%, as previously described (C. Zhang et al., 2018a). The cells were grown at 37°C and 250 rpm until OD_{600} reached ~ 0.8 when they were induced by 0.05 mM IPTG, and were then grown at 30°C for 24 hours. Antibiotics (34 µg/ml chloramphenicol, 50 µg/ml kanamycin, 50 µg/ ml spectinomycin and 100 µg/ml ampicillin) were added to the culture to maintain the four plasmids.

Extraction and quantification of carotenoids

Total intracellular carotenoids were extracted from cellular pellets according to the acetone extraction method (C. Zhang et al., 2018a). Briefly, 20 µL of bacterial culture were centrifuged for 10 minutes at 14000 g. The cell pellets were resuspended in 200 µL of acetone. After a 20 minutes

incubation at 50°C and 1500rpm, the mixture was centrifuged for 10 minutes at 14000 g. The supernatant was filtered using a PTFE, 0.45 µm filter and subjected to HPLC analysis carried out by an Agilent 1260 Infinity LC system equipped with a ZORBAX, Eclipse Plus C18, 4.6 × 250 mm, 5 µm column and diode array detector (DAD). Isocratic conditions (80% acetonitrile, 18% methanol, and 2% water) were maintained at 1 mL/min for 10 minutes for all experiments except for linker mutation analysis. The run for the mutation of linker mutation was performed using a more separative gradient starting at 48% methanol, 12% water and 40% acetonitrile and lasting one minute. For the next three minutes, a gradient was applied to reach 16% methanol, 4% water and 80% acetonitrile, which was applied until the end of the run. The entire run lasted 10 minutes. The carotenoids were detected and quantified through absorbance at 450 nm. Standard curves were generated using chemically synthesised lycopene, β-carotene, β-cryptoxanthin and zeaxanthin (CaroteNature, Switzerland).

For the time-course profile experiment, the usual ratio 1:10 of cell to solvent was adjusted between 2:1 to 1:10 along the experiment to allow the extraction and detection of smaller amounts of carotenoids.

SDS-PAGE

SDS-PAGE analysis was carried out on BL21 strain only carrying the last plasmid module (p15a-Amp plasmid either empty, or with crtY gene or crtZ-crtY gene fusion). 10 mL of culture were centrifuged at 4000g for 10 minutes at 4°C. Cell pellets were concentrated 10 times in 1 mL of buffer I (200 mM Tris Hcl, 50 mM NaCl, 1 mg/ml lysozyme and 1x protease inhibitor) and incubated for 30 minutes at 4°C. Cells were then lysed by 3 cycles of freeze-thawing. DNAse at 1.5µl/ml and 2 mM MgCl2 were then added to the lysate mixture. Protein concentration was determined by BCA assay. 6 µL of sample containing 10 µg of protein were mixed with 1 µl of 10X reducing agent (Invitrogen™ NuPAGE™) and 2.5 µl of 4x loading buffer (Invitrogen™ NuPAGE™). Samples were loaded on SDS-PAGE gel with 4 to 12% polyacrylamide gradient and run at 180V until the migration front reached the end on the gel. Proteins were detected by Coomassie blue staining.

Statistics

Rstudio software (version Rstudio/2022.12.0+353) was used for statistical analysis. One-way ANOVA was used to compare one independent factor in three independent groups. Two-way ANOVA was used to compare two independent factors in four independent groups. When the ANOVA was significant, the Tukey post Hoc test or t-test were used to make pairwise comparisons between groups. The normality of variance (shapiro test) and homogeneity of variance (Levene's test) were verified for all ANOVA analysis. When the aforementioned hypotheses were not verified, a logarithmic transformation was applied to the variable. P-values were calculated and represented as follow: $P < 0.05$, $P < 0.01$, $P < 0.001$ and $P < 0.0001$ were indicated by *, **, *** and **** respectively. Only significant differences were indicated.

The orientation of the fused enzymes crtY and crtZ is crucial for their activities.

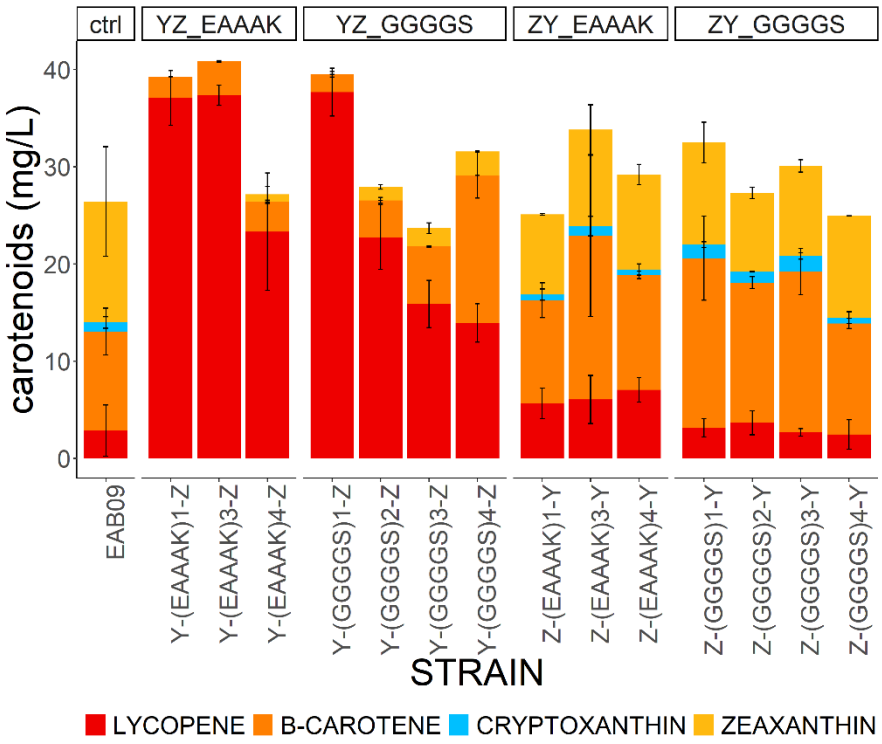


***Figure 21**: Carotenoids content in mg/L of the strains harbouring the fusion constructs Y-linker-Z and Z-linker-Y. EAB09 is the control (ctrl) strain expressing independent enzymes. Errors bars represents the standard deviation of two independent experiments. The control strain experiment was repeated four times.*

To establish the importance of the enzyme order in the fused assembly crtY/crtZ, the enzymes were fused in both orientations with synthetic linkers typically used in literature (Li *et al.*, 2016). The linkers were constituted of one to four repeats of either the flexible motif (GGGGS) or the rigid spacer (EAAAK). The fifteen resulting plasmids expressing the independent or fused crtY/crtZ couples were transformed in the lycopene producing strain to reconstitute a full zeaxanthin pathway. The control strain with independent enzymes (EAB09) accumulates around forty percent of the β-carotene precursor while the final product zeaxanthin represents fifty percent of the total carotenoids.

The first orientation consisted to fused the C-terminal of crtY to the linker, leading to a set of crtY-linker-crtZ fusions. In the case of the crtY-(EAAAK)x-crtZ fusions, lycopene is accumulated and a small amount of β-carotene is detected (**Figure 21**). Fusing the C-terminal of crtY with GGGGS linkers restore the β-carotene production to a level comparable to the control strain EAB09. The amount of β-carotene accumulated gradually increased with the size of the linker. Independently of the linker type, these results indicated that crtY activity is affected by its order in the fusion. The amino acid sequences of crtY and crtZ were analysed to identify the presence of a transmembrane domain using

the web-server Phobius (Madeira *et al.*, 2019). AlphaFold predicted structure of both crtY and crtZ were retrieved from AlphaFoldDB (**Figure 28**). Phobius web-server does not predict any transmembrane domain for crtY, and the predicted Alphafold structure is rather globular. We thus hypothesize that prokaryotic crtY from *Pantoea ananatis* is more likely a membrane-associated protein rather than anchored to the membrane unlike eukaryotic isoforms of crtY (Krubasik and Sandmann, 2000; Rabeharindranto *et al.*, 2019). Indeed, the expression of crtY from *P. ananatis* was previously optimised by fusion to a N-terminal Maltose Binding Protein (MBP) tag suggesting that membrane location for crtY in *E. coli* is favourable (Yu *et al.*, 2010). Regarding the zeaxanthin production, even if more β-carotene is accumulated, very little zeaxanthin is produced, giving the impression that crtZ activity is also impeded in this orientation. This would be consistent with previous studies (Henke and Wendisch, 2019; Nogueira *et al.*, 2019; Y. Wu *et al.*, 2019; Ding *et al.*, 2022) where fusion proteins with crtZ placed at the C-terminal of the construct were not producing zeaxanthin. The Phobius prediction shows three transmembrane helices at the N-terminal of crtZ, with the N-terminal orientated toward the extracellular space. The fusion of crtY enzyme at the N-terminal of crtZ could thus impeed the correct orientation of crtY towards the membrane and its access to the substrate, leading to the accumulation of the substrate of crtY, lycopene in the crtY-crtZ fusions. We hypothesized that crtY require a long and flexible linker (crtY-(GGGGS)4-crtZ) to compensate an unfavourable first location in the crtY-crtZ fusion. Surprisingly, strains with unfunctional crtY protein (crtY-(EAAAK)x-crtZ), had a significantly higher production of lycopene (or total carotenoids) than the strains producing the downstream products β-carotene and zeaxanthin. The decrease in total carotenoid production in presence of functional crtY and crtZ enzymes might be due to the consumption of cofactor in the cell. Indeed, crtY has been shown to consume NADH and NADPH (Yu et al., 2010) and crtZ, NADPH (Bouvier and Keller, 1998).

The second orientation consisted of the C-terminal of crtZ fused to the linker, leading to a set of crtZ-linker-crtY fusions. Independently of the linker motif, all crtZ-crtY fusions were able to produce β-carotene and zeaxanthin to a level similar to the one of the non-fused enzymes (**Figure 21**). According to these first results, the crtZ-crtY orientation was chosen to further characterise and improve the crtZ-crtY protein fusion. As differences were observed based on the size and motifs of the linkers, exploring the sequences of the linkers could lead to optimize the fusion efficiency and was assay by building a linker library.

Conception of a linker library.

SIZE	Helical	C-alpha extent category	Number of linker	Proportion
LARGE			19	21%
	Helical		8	9%
		<2 Å	7	8%
		>2 Å	1	1%
	non helical		11	12%
		<2 Å	9	10%
		>2 Å	2	2%
MEDIUM			52	57%
	Helical		28	31%
		<2 Å	17	19%
		>2 Å	11	12%
	non helical		24	26%
		<2 Å	13	14%
		>2 Å	11	12%
SMALL			20	22%
	Helical		7	8%
		<2 Å	2	2%
		>2 Å	5	5%
	non helical		13	14%
		<2 Å	2	2%
		>2 Å	11	12%
TOTAL			**91**	**100%**

***Figure 22**: Classification of the linker properties and their repartition in the library. Heading in dark grey give the category of the linker. Sub-headings in light grey represents a class within the category and numbers in these lines represent the sub total of linkers in each class.*

To optimize the crtZ-crtY protein fusion, a library of 91 linkers was designed out of the 1280 present in the online linker database IBIVU, based on natural linkers from natural multidomain protein (George and Heringa, 2002). The 91 linkers were selected to conserve a proportion of linkers by category similar to the one in the online database. Each linker of the pool was annotated using four criteria: size, structure, C-alpha extent and hydrophobicity (**Figure 22**). For the size, 3 classes were designed. Small linkers contain from two to five amino acids (notated S), medium linkers are between six and thirteen amino acids (notated M) and large linkers between fourteen to fifty-eight amino acids (notated L). The linker structure can either be helical (notated H) or non-helical (notated N). The extension of the linker, C-alpha extent, is defined by the average distance between its amino acids divided by the number of amino acids minus one. Two classes of C-alpha extent were designed: those with an average inferior to 2Å (notated I) and those with an average superior to 2Å (notated S). Lastly, linkers are ranged by their average hydrophobicity according to the Eisenhower scale. For example, the MHI05 linker is a Medium size linker, Helical, with a C-alpha extent Inferior to 2Å. It is more hydrophobic than MHI04 but less hydrophobic than MHI06.

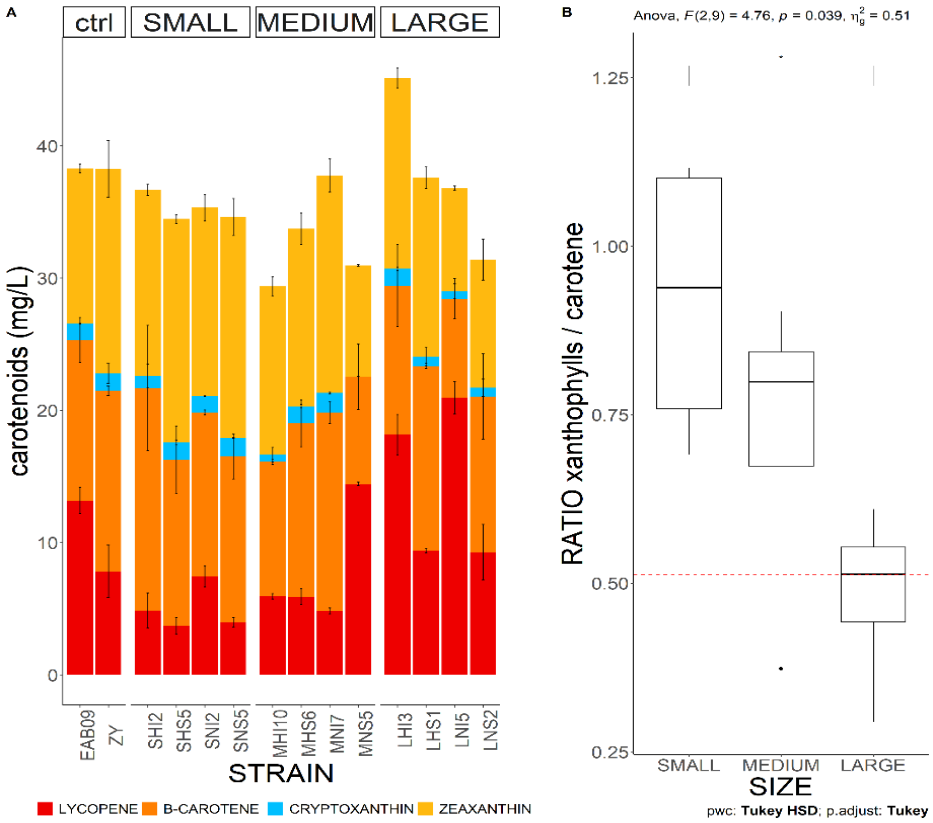


***Figure 23**: **A:** Carotenoid content in mg/L of the strains harbouring the fusion constructs Z-linker-Y with small, medium or large linkers. EAB09 is the control strain with non-fused enzymes. ZY is the strain with enzymes fused directly, without linker. Errors bars rep resents the standard deviation of two independent experiments. **B:** Ratio of xanthophylls to carotenes with the strains clustered according to the size of the linker. Data is presented as standard boxplots. Dark bars represent median values, boxes are the range from first to third quartile, whiskers represent minima and maxima. The dashed red line represents the ratio of xanthophylls to carotenes in the strain with independent enzymes (strain EAB09).*

Small linkers improve pathway efficiency.

To optimize the size of the linker in the crtZ-crtY fusion protein, four linkers of each size's class (small, medium and large) were cloned between crtZ and crtY and the carotenes content was quantified in the 12 strains. All strains displaying a crtZ-crtY fusion construct with a linker produced a similar amount of carotenoid to the strain EAB09 expressing the non-fused enzymes. This indicates that all linkers lead to a functional fusion, although the overall carotenoid production was not improved (**Figure 23**, A). To refine the impact of the linkers, the ratio of xanthophylls and carotenes, reflecting the efficiency of the pathway to produce final metabolites, was compared between strains expressing different sizes of linker in the fusion protein (**Figure 23**, B). Strains expressing a small or medium linker between crtZ and crtY had a 1.8-fold increase in their ratio of xanthophylls compare to the control strain and the strain with large linkers. Strains harbouring large linkers accumulate more precursors (lycopene or β-carotene) than the non-fused construct. The balance of pathway intermediates is then modified based on the linker size, with the more favourable ratio involving the small linkers. The size effect between crtY and crtZ could be due to folding issues or linker proteolysis

effects, favoured in the longer class. Another hypothesis could be that a small linker will create a compact and more stable enzymatic complex, closer to the membrane, where lycopene, the hydrophobic substrate, could be more accessible for crtY. Thus, small linkers were chosen to further optimize the crtZ-crtY fusion.

crtZ-crtY fusions with small linkers show diverse carotenoids profiles

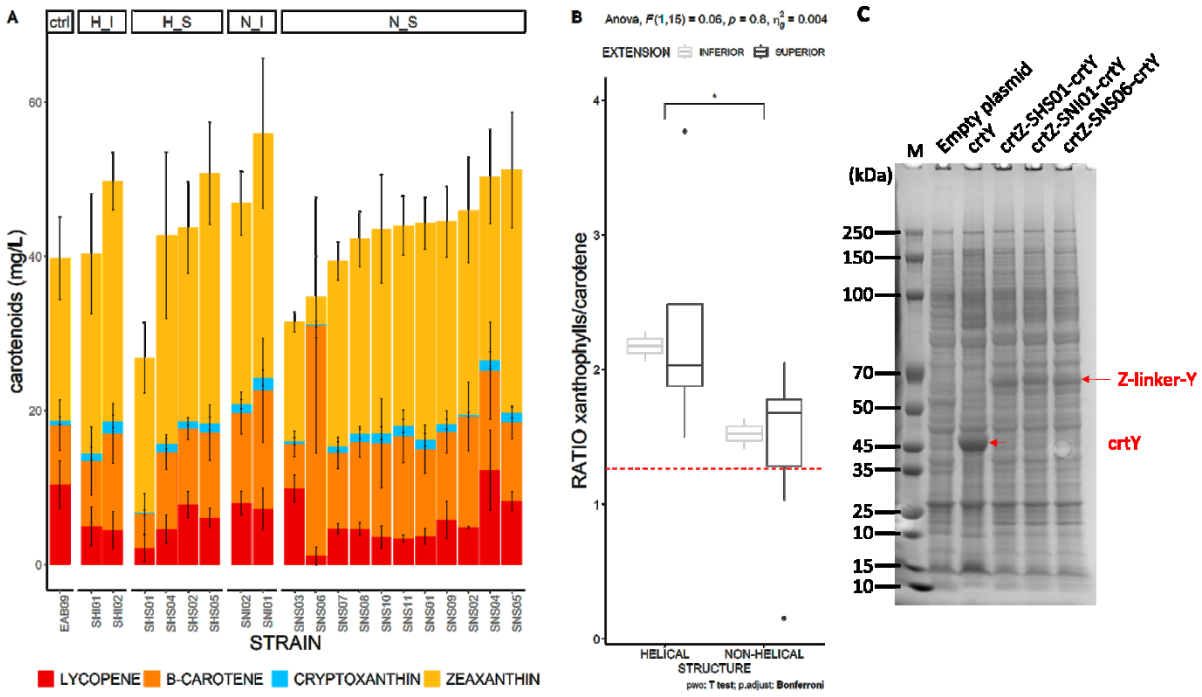


Figure 24: A: *Carotenoid content in mg/L of the strains harbouring the fusion constructs crtZ-linker-crtY with small linkers. EAB09 is the control strain with non-fused enzymes. Errors bars represents the standard deviation of two independent experiments. N_I: NON-HELICAL_INFERIOR; H_I: HELICAL_INFERIOR; H_S: HELICAL_SUPERIOR; N_S: NON-HELICAL_SUPERIOR. Non-helical and helical refers to the presumed secondary structure of the linker while inferior and superior refers to the extension of the linker. Among each category, strains are ranged by ascending order of total amount carotenoids.* ***B:*** *Ratio of xanthophylls to carotenes with the strains clustered according to the structure and extension of the linker. Data is presented as standard boxplots. Dark bars represent median values, boxes are the range from first to third quartile, whiskers represent minima and maxima. The dashed red line represents the ratio of xanthophylls to carotenes in the strain with independent enzymes (strain EAB09).* ***C:*** *SDS-PAGE of BL21 strain containing the final plasmid of the pathway with: an empty vector, crtY enzyme, fusion enzymes crtZ-SHS01-crtY, crtZ-SNI01-crtY, crtZ-SNS06-crtY. M indicates the line with protein molecular weight marker. Red arrow indicates the proteins of interest.*

To assess the role of linkers physical parameters in crtZ-linker-crtY fusions, nineteen small linkers from the linker library were selected and cloned between both enzymes. Although composed of only three to five amino acids, linkers were described in the database presenting a range of flexibility, extension or hydrophobicity different from one another. The carotenes content was quantified in the nineteen resulting strains as well as in the strain with the non-fused enzyme, EAB09.

The amounts of carotenoids produced in each strain is shown in **Figure 24**, strains are grouped by the category of the linkers in the enzyme fusion, and ranged by ascending order of total carotenoids amounts (**Figure 24**, A). Among the nineteen protein fusions with small linkers, eighteen were functional, as zeaxanthin is produced in the correlated strain. Although most strains displayed a similar profile of carotenoid accumulated, a few stood out from the others. First, the strain having the crZ and crtY enzyme fused together with the SNS06 linker accumulated only β-carotene, with lycopene precursor being almost fully converted to β-carotene and little to no zeaxanthin produced, suggesting that this linker is not compatible with crtZ enzymatic activity. This result is unlikely to be due to a full cleavage of the crtZ-crtY protein since the electrophoresis analysis by SDS-PAGE (**Figure 24**, C) showed an accumulation of signal migrating at the protein fusion full-length. However, this result needs to be taken with caution. Indeed, the proteins were only detectable on SDS-PAGE in the BL21 strain only expressing the plasmid with the protein fusion and not in the strain expressing the four plasmids, probably due to a lower expression level of the enzymes, all under the control of the T7 promoter. Secondly, the strain with SNI01 linker produced a total of 56 mg/L of carotenoids which represent a 40% increase compared to the production in the control strain with independent enzymes. Finally, the strain with the SHS01 linker displays a two-fold improvement in the xanthophylls over carotenes ratio compared to the control strain with independent enzymes. This result demonstrates a better conversion of precursors toward the final products, despite a lower amount in total carotenoids. To refine which properties of the linker could explain the different productions observed, the sub criteria of the linker library, i.e. the supposed structure (helical/non-helical) and C-alpha extent were analysed (**Figure 24**, C). On average, the strains with fused enzymes have a higher ratio of xanthophylls than the strain with independent enzymes. The strain displaying a linker annotated as helical have a significantly higher xanthophylls ratio than the one having a non-helical (flexible) linker. However, variability in the C-alpha extent of the linker did not yield any significant differences in the ratio of carotenoids produced in the strains. So far, we have demonstrated that both the size and the flexibility of the linker impact the behaviour of the fused enzymes.

Time-course profile of carotenoids accumulations in strains of interest

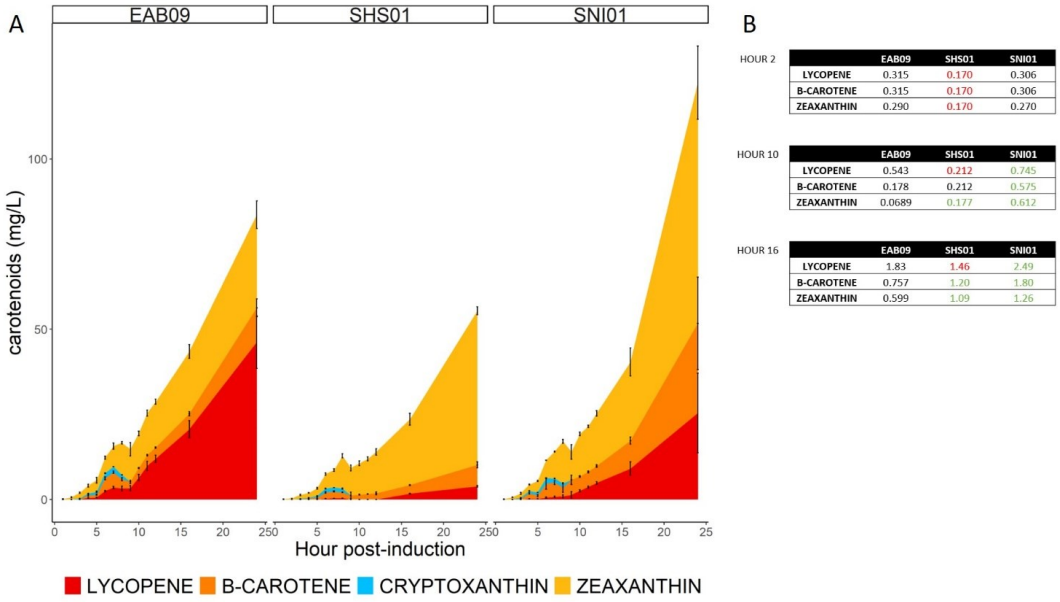


Figure 25: A: *Carotenoid accumulation profile in mg/L of the strains EAB09, SHS01 and SNI01 over time. Errors bars represents the standard deviation of three independent experiments.* ***B:*** *Flux of carotenoids at selected time expressed in mg/L/hour/OD600. Numbers in green are higher than their control equivalent (EAB09). Numbers in red are lower than their control equivalent (EAB09).*

To decipher if the improvement of xanthophylls ratio was due to a better enzymatic rate from the crtZ-crtY fusion or to a slower pathway flux, a time-course profile of the accumulation of carotenoid was performed over 24 hours. The experiment was carried out for the control strain with independent enzymes, EAB09, the strain with the highest xanthophylls ratio, SHS01, and the strain with the highest amount of total carotenoid, SNI01 (**Figure 25**, A). In these conditions, the EAB09 strain with independent enzymes accumulated 55% and 12% of lycopene and β-carotene precursors for a final amount of zeaxanthin of 32%. On the contrary, the strain with crtZ-SHS01-crtY enzyme fusion produced up 81% of zeaxanthin with only 7% and 11% of lycopene and β-carotene remaining at the end of the reaction. This corresponds to a 9-fold increase in the xanthophylls to carotene ratio. The strain with crtZ-SNI01-crtY fusion had an intermediary profile. Zeaxanthin represented up to 57% of the carotenoids produced, a 1.35-fold improvement compared to the control strain with independent enzymes. The quantity of 70 mg/L obtained here is similar or higher to previous studies in *E. coli* where zeaxanthin was produced up to 51.8mg/L, 60mg/L, or 6.33mg/L (Li *et al.*, 2015; Z. Wu *et al.*, 2019; Chen *et al.*, 2021). It could however be improved by fermentation processes as fermentation has allowed the production of zeaxanthin titter to reach 722 mg/L (Shen *et al.*, 2016). The total amount of carotenoids reached 121.7 mg/L, a 1.46-fold increase to the strain with independent enzymes. The metabolites fluxes were then calculated for each step of the pathway with $Fl = \frac{\sum A_{m,\ h} - \sum A_{m,\ h-1}}{OD_{600,\ h}}$, where Fl is the flux for a metabolite m, $\sum A_{m,\ h}$, the sum of the amounts of metabolite m and metabolites derived from m at hour h , $\sum A_{m,\ h-1}$, the sum of the amounts of metabolite m and metabolites derived from m at hour h-1 and $OD_{600,\ h}$, the OD_{600} nm at hour h (**Figure 25**, B). All three

strains displayed an overall pathway flux increasing overtime (lycopene reaction rate for the strain EAB09 went from 0.315 mg/L/hour/OD at hour 2 post-induction to 1.81 mg/L/hour/OD at hour 16 post-induction), probably due to a progressive accumulation of the crtI enzyme in the cell. The SHS01 strain lycopene flux is twice slower than the control strain with independent enzymes at the tenth hour post induction and after 10 hours, β-carotene is produced faster than in the control strain. These fluxes pointed out to an acceleration of the step involving crtY, which could be due to its location near the membrane when fused to crtZ, a location known to be favourable to its activity (Yu *et al.*, 2010). Both the decrease in the upper pathway flux and increase in the reaction rate of crtY explain the better zeaxanthin selectivity in SHS01 strain. Zeaxanthin production rate is 2.5-fold higher in the SHS01 strain than in the control strain. This could be explained by the increase in the pool of β-carotene, indicating that crtZ enzyme is not limiting. On the contrary, SNI01 strain shows a lycopene flux increased by a 1.4-fold compared to the control strain with independent enzymes. Even with a higher flux, lycopene is not accumulated (**Figure 25**, A). Indeed, in the SNI01 strain, lycopene is converted to β-carotene more than 3 times faster than in the control strain and zeaxanthin in converted almost 9 times faster in SNI01 than in the control strain (**Figure 25**, B). The crtZ-crtY fusion with SNI01 linker helps to increase the overall flux in the zeaxanthin pathway, and thus increases the total amount of zeaxanthin.

Alanine is mainly found at the 1st position of the linker.

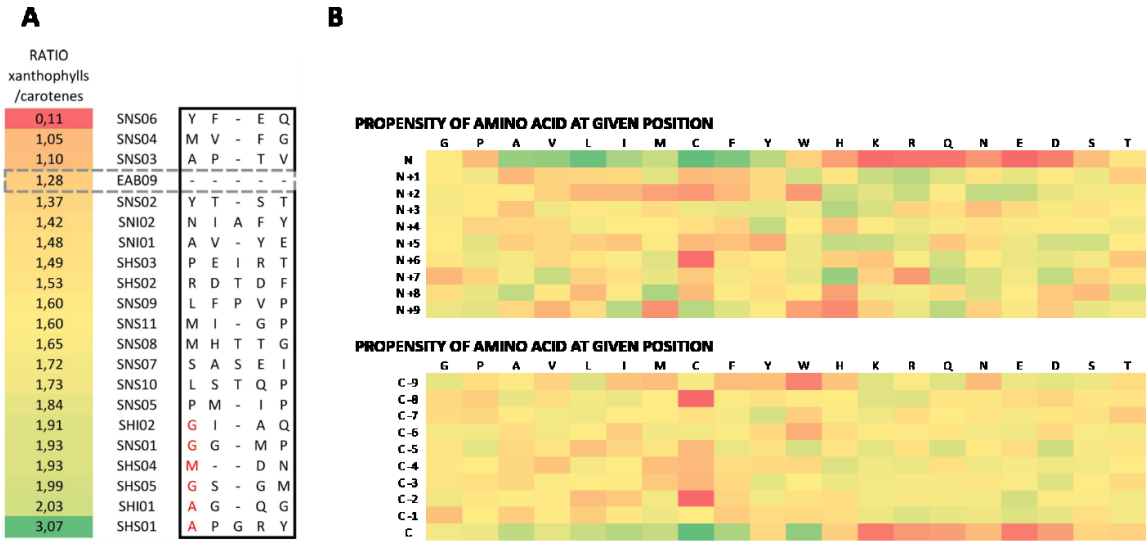


Figure 26: A: *Sequence alignment of small linkers. The dashed frame surrounds the independent enzymes. The linker sequences are framed by a plain line box. Constructions are ranged based on the ratio of xanthophylls to carotenes accumulated in the strains. Alignment performed with clustalW.* ***B:*** *Amino acid propensity by position in the linker based on the 1280 linkers from the online database. Propensity values are represented as a red to green colour gradient, from smaller to higher values respectively. Top panel represent position 1 to 10 of linkers from the N-terminal. Bottom panel represent last ten residues of linkers. First raw of table is the one letter code for amino acid.*

While the enzyme orientation and the size of the linker were shown to have a crucial effect on the pathway efficiency, no other criteria among the small linkers were shown to impact the

performance of the tested fusions. This raised the question of how a variation of such a small peptide (3-5 amino acids), far from the catalytic pockets of the 564 amino acid structure can cause different profiles in carotenoids accumulation. The role of the amino acid sequences in small linkers was analysed. Thus, an alignment of the linker sequences was performed using clustalW. The strains with all the small linkers were then ranged by ascending orders of the xanthophylls to carotenes ratio. Among the crtZ-crtY enzyme fusion strains, the strains with the higher ratio of xanthophylls to carotenes were the ones displaying an alanine or a glycine at their first position (**Figure 26**, A). This result was also compared to the set of 1280 linkers found in natural multidomain proteins from the online IBIVU database. Linkers were aligned either by their C-terminal or N-terminal end and the propensity of each amino acid was determined for different positions in the linker. The propensity was determined as $Pa = \frac{Nr_{i,p}/\,Nl_p}{\Sigma_i\, Nr_{i,s}}$, where Pa is the propensity of the amino acid i, $Nr_{i,p}$, the occurrence of the amino acid i at position p, Nl_p, the number of linker and $\sum_i Nr_{i,s}$ the occurrence of the amino acid i in the whole linker set. The results are shown in **Figure 26** (B). Firstly, it is interesting to note that the first position of the linker, whether at the N-terminal or the C-terminal side, has the most polarized use of amino acid. Indeed, amino acids such as cysteine, tryptophan, phenylalanine, leucine, alanine and valine are overrepresented at the first position whereas amino acids such as the glutamic acid, lysine and arginine are underrepresented. This polarization does not exist at other positions where a more balanced propensity of the amino acids is observed. This result is predictable, as the amino acid in direct contact with the enzyme is expected to bear the highest constraint. On the contrary of the bulky and charged glutamic acid, lysine and arginine, the residues leucine, alanine and valine are rather small, hydrophobic and neutral amino acids, which could be more generic as a first residue out of the globular shape of the first enzyme in the natural fusions. Based on the results obtained from the entire database, we could expect that alanine would be more represented than other amino acids in the best linkers from the small linker set, although not more than leucine, phenylalanine or cysteine.

Alanine is a better amino acid at first position of most linkers.

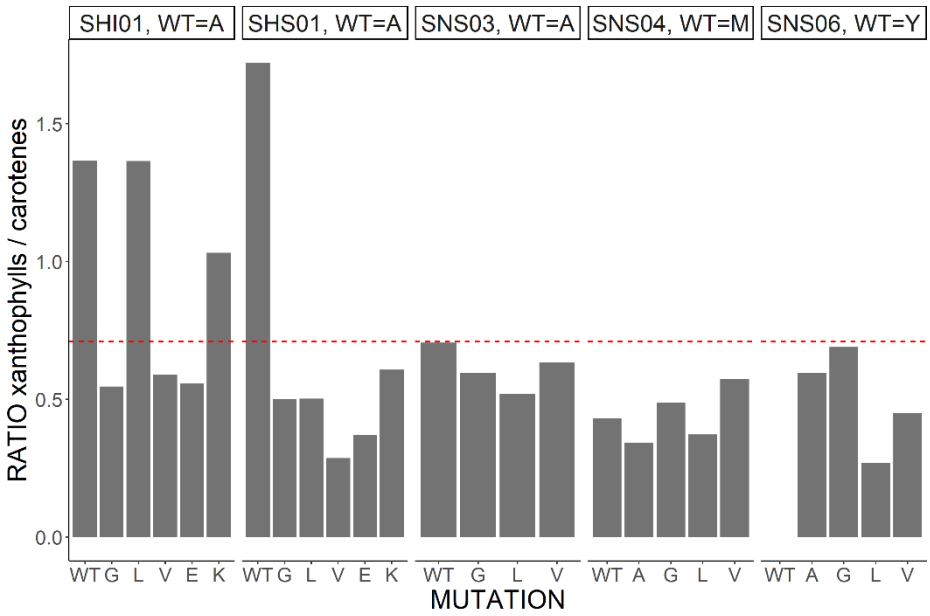


Figure 27: *Ratio of xanthophylls to carotenes produced by crtZ-linker-crtY protein fusion with the five linkers SHI01, SHS01, SNS03, SNS04 and SNS06. The original first amino acid of the linker is indicated at the top of each panel as the Wild-Type (WT) and the ratio of carotenoids in the corresponding strain is represented by the left bar of each panel. Ratio of mutation with amino acids alanine (A), glycine (G), leucine (L), valine (V), glutamic acid (E) and lysine (K) are displayed after the WT left to right. The dashed red line represents the ratio of xanthophylls to carotenes in the strain with independent enzymes.*

Since the first amino acid of the linker seemed to have an important impact in both the crtZ-crtY fusion construct and natural enzyme fusions, the hypothesis was verified by performing single point mutation of this position. Strains displaying the lowest and highest ratio of xanthophylls over carotenes were selected and the first amino acid of the linker was mutated. The mutations were made to alanine or glycine, which were shown to be the best in the crtZ-crtY enzyme fusion. Alanine with leucine or valine were also part of the most represented amino acids at the first position of linkers in the database, whereas acid glutamic and lysine were the least represented. Acid glutamic and lysine mutations were only performed on strains SHS01 and SHI01, having a good ratio of xanthophylls in the wild-type (WT) linker.

Mutants to SHS01 and SHI01 strains, which have an alanine as WT amino acid, lose their advantage and revert to a ratio similar to the control strain. SNS03 and SNS04 strains had a similar ratio independently of the nature of their first amino acid, indicating that amino acids in other positions also impact the protein fusion. Surprisingly, the profile of the SNS06 strain was completely reverted with the single mutation. The strain which was not producing xanthophylls with the tyrosine at its first position had a ratio close to the control strain with independent enzymes with an alanine or glycine. The reversion also happened with leucine and valine mutants, although to a lower extent. The tyrosine at the first position of SNS06 linker cannot be entirely responsible of the lack of

xanthophylls in the strain. Indeed, the enzyme fusion in the SNS02 strain also has a tyrosine at the first position of the linker, and this strain does produce xanthophylls. Overall, looking at the carotenoid ratio in the crtZ-crtY enzyme fusion, in certain cases, alanine seems to be the best amino acid for the first position of the linker, and in other cases, glycine or valine are equally good or even better. However, adjustment to the first amino acid of the linker is not enough to change the crtZ-crtY fusion into one accumulating xanthophylls.

Discussion and conclusion

Expressing the zeaxanthin pathway in *E. coli* leads to a partial conversion of β-carotene into zeaxanthin. To improve the overall bioprocess, the two enzymes crtY and crtZ catalysing the bottleneck metabolic steps were fused using a linker library. By successively adjusting different linker criteria between crtZ and crtY enzymes, the impact of the enzyme orientation and linker size were shown to be crucial for the performance of the crtZ-crtY enzyme. CrtY was only efficient when placed at the C-terminal of the enzyme fusion and crtZ at the N-terminal of the enzyme fusion as it was demonstrated in previous studies (Z. Wu *et al.*, 2019; Ding *et al.*, 2022). We speculated that it could be explained by the orientation of both enzymes around the membrane. CrtZ is predicted to be membrane bound and crtY membrane-associated and this could require more or less flexibility to achieve an optimal configuration allowing the enzyme to reach the hydrophobic substrate embedded in the cell membrane. By testing a library of linkers with a large diversity, small linkers of three to four amino acids were found to give strains more efficient than those with larger linkers, and more efficient than the independent enzymes in 85% of the cases. Among the linkers tested, some linkers sequences led to specific phenotypes. In particular, the crtZ-crtY enzyme fusion with SHS01 linker gave a strain with a 9-fold improvement of the xanthophylls over carotene ratio when compared to the strain with independent enzymes. The enzyme fusion with SNI01 linker allowed a 1.46-fold increase of the pathway flux. Some improvements of the enzyme fusion might be due to a higher expression of crtZ when fused to crtY. Unfortunately, efforts to quantify the crtY, crtZ or crtZ-crtY protein using western-blot remained unsuccessful, probably due to the membrane bound/associated behaviour of the enzyme complexes. Gathering enzyme domains using a linker enforces a 1:1 ratio of the enzymes that might not be the same when the enzymes are independent from each other and which affects the overall metabolites distribution, as seen in the strains SHS01, SNS06 and SNI01. On top of the physical parameters of the linker involved in the efficiency of the fusion, the sequence of the linker was also a determinant criterion, with the first amino acid of the sequence being crucial, both in natural enzyme and crtZ-crtY fusion. Exploring the idiosyncrasy of linkers can allow to fine-tune an enzyme fusion and this approach can be used on any model due to the versatility of the method.

While the fusion strategy is simple to implement, the linker sequences to explore leads to a time-consuming process. To narrow the sequences to explore, our study and others (Li *et al.*, 2016; Guo *et al.*, 2017) suggest that a good approach to fuse enzymes is to assay the two enzyme orientations using the flexible motif GGGGS or the rigid spacer EAAAK and tune the first amino acid of the most efficient assembly.

However, the improvement obtained by enzyme fusion is often limited, with an improvement range often between 1 to 5-fold (Bakkes *et al.*, 2015; Baklouti *et al.*, 2020; Wang *et al.*, 2021). In the recent years, several studies (Wheeldon *et al.*, 2016; Poshyvailo, Lieres and Kondrat, 2017; Rabe *et al.*, 2017; Kuzmak *et al.*, 2019) showed that enzyme proximity could only enhanced cascade reaction temporarily (order of millisecond before steady state) as diffusion is fast compared with usual catalytic reaction rates and benefits can only be seen in presence of a competing reaction. Moreover, the imposed stoichiometry of active sites in enzymes fusion implies that the slowest reaction will limit the overall activity. To overcome this limitation, enzyme clusters using post-translational assembly or scaffold assembly can remove this constraint by compensating the different reaction rate and adjusting enzyme stoichiometry. Overall, enzymatic assembly is just one of the tools of metabolic engineering and should be combined with others, such as enzymatic engineering. Obtaining more efficient individual enzymes could then present a synergic efficiency when embedded into enzymatic complexes.

In this study, the β-carotene cyclase crtY and the β-carotene hydroxylase crtZ were fused. A collection of linkers was used to fine-tune the enzyme fusion and the role of linker size and structure was demonstrated in this specific set of protein. Moreover, we showed for the first time that the amino acids at the extremities of the linker in contact with the enzymes have a higher selection constraint than amino acid in other position. In the end, we obtained the SNI01 strain with a 1.46-fold increase in the pathway flux and the SHS01 strain with a 9-fold increase in the ratio of xanthophylls produced. Our linker library approach can be implemented for any protein fusion.

Author contributions

Aurélie Bouin: Conceptualization, Investigation, Writing - Original Draft;

Congqiang Zhang: Conceptualization, Writing (review and editing), Funding acquisition:

Nic D Lindley: Conceptualization, Writing (review and editing), Funding acquisition;

Gilles Truan: Conceptualization, Writing (review and editing), Funding acquisition;

Thomas Lautier: Conceptualization, Writing (review and editing), supervision, Funding acquisition

Supplementary Data

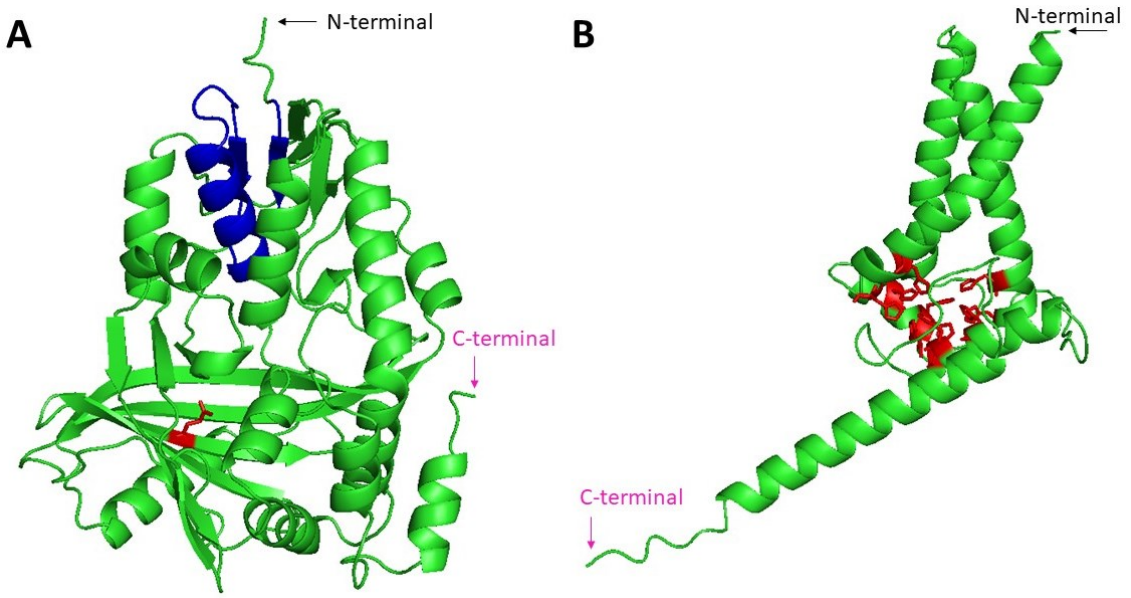


Figure 28: A: *CrtY structure prediction by AlphaFold Monomer v2.0. In blue, a NAD+ binding site and in red, an amino acid required for CrtY activity.* ***B:*** *CrtZ structure prediction by AlphaFold Monomer v2.0. In red, histidine involved in iron binding.*

3. Detection of protein

Introduction

One of the limits we had when exploring the protein fusions was the lack of understanding we had from the results obtained. Indeed, the accumulation of carotenoids allowed us to confirm the presence of the enzyme in the cell, as well as its activity. However, if the enzyme was active, any changes in carotenoid accumulation could have been due to a change in either the expression level of the enzyme, a change in its structure or its location, etc.… To understand our system better, we thus tried to quantify the expression of proteins in the cell, both to compare relative expression of the different protein fusions as well as to compare the expression level of both independent enzymes, CrtY and CrtZ with their level in the protein fusion. Indeed, as the protein fusion enforces a 1:1 ratio of both enzymes, controlling the expression level of each enzyme when they are independent could help assess if an improvement obtain with protein fusion is only due to the change in the protein ratio for example.

This section describes the effort made to quantify CrtY, CrtZ and the protein fusion by Western-Blot.

Material and methods

Cell culture

After an overnight preculture in 2XPY medium (20 g/L peptone, 10 g/L yeast extract and 10 g/L NaCl), cells were inoculated at OD_{600} = 0.1 in 2XPY medium supplemented with 20 g/L glycerol, 50 mM 4-(2-hydroxyethyl)- 1-piperazineethanesulfonic acid (HEPES) and Tween 80 0.5%, as previously described (C. Zhang et al., 2018a). The cells were grown at 37°C and 250 rpm until OD_{600} reached ~ 0.8 when they were induced by 0.05 mM IPTG, and were then grown at 30°C for 16 hours. Antibiotic(s) were added to the culture to maintain the plasmid(s).

Cell lysis

The cells were harvested by a centrifugation at 4000g for 10 minutes at 4°C. The supernatant with culture media was discarded and the pellets were resuspended in lysis buffer (200 mM Tris HCl pH8, 50 mM NaCl, 1 mg/mL of lysozyme and 1X protease inhibitor) in a volume allowing 10x concentration. The mixture was incubated on ice for 30 minutes and the cell were then lysed by three cycles of freeze-thawing, from liquid nitrogen to 37°C. 2mM of MgCl2 and 1.5µl/mL of DNase I were then added to the lysis and incubated for five minutes in order to digest the DNA and have a less viscous solution. The mixture obtained was considered as the total cell fraction. 100µl of total cell fraction was aliquoted and the remaining solution was centrifugated at 14000g for 30 minutes and 4°C in order to separate the membrane and cytosol fraction. The cytosolic fraction corresponds to the supernatant and the pellet corresponds to the membrane fraction. Membrane fraction was then resuspended with detergent.

Histidine-tag purification

For a histidine-tag purification, the 50 ml of cells were harvested by a centrifugation at 4000g for 10 minutes at 4°C and the pellet was resuspended in 10 ml lysis buffer (100 mM Tris HCl, 5 mM $CaCl_2$, 300mM NaCl, 10% glycerol, 20 mM imidazole, 10 mg/mL of lysozyme, 1.5µl/ml of DNAse I). After three cycles of freeze-thawing, the total cell fraction was obtained. 1 mL of resin was transferred to a 15 mL Falcon tube and the supernatant was removed. 2.5 mL of binding buffer (100 mM Tris HCl, 5 mM $CaCl_2$, 300mM NaCl, 10% glycerol, 20 mM imidazole) were added to the resin (equilibration) and removed after the resin had settled. The total fraction lysate was added to the tube with the resin and incubated for 1h at 4°C with agitation (binding). 500 µl of the solution was aliquoted as the flowthrough. The resin was then transferred to a gravity flow column and washed three times with 5 mL of washing buffer. The elution step was performed five times with 500 µl of elution buffer (100 mM Tris HCl, 5 mM $CaCl_2$, 300mM NaCl, 10% glycerol, 450 mM imidazole).

SDS-PAGE

Protein concentration of the different samples was sometimes determined by BCA assay. Samples were mixed with reducing agent (Invitrogen™ NuPAGE™) and loading buffer (Invitrogen™ NuPAGE™) and then loaded on gel with 4 to 12% polyacrylamide gradient. The run was performed at 180V until the migration front reached the end on the gel. Proteins were detected by Coomassie blue staining.

Protein transfer

First the proteins were run on a SDS-PAGE gel as described above. Then the proteins are transferred on a membrane either by wet or dry transfer. For the dry transfer, an iBLOT 2 Dry Bloting system from Life Technologies was used. Briefly, the transfer is done on a PVDF membrane at 20 Volt for one minute, 23 V for four minutes and 25 V for two minutes for a total time of seven minutes. The wet transfer is performed on the nitrocellulose membrane within the transfer buffer (10% of transfer buffer, 20% of methanol and 70% of distilled water) and the transfer is performed at 120V for one hour. The membrane was rinsed three times for five minutes with PBST.

Ponceau S staining

The membrane was submerged in Ponceau red stain for five minutes and rinsed with distilled H_2O. The image was captured and the Ponceau was removed by four washing steps of five minutes each.

Western-Blotting

For the blotting steps, the membrane was first blocked with a 5% milk TBST buffer (Tris-Buffered Saline with 1% tween 20). The membrane then incubated at room temperature for one hour with agitation. The blocking solution was removed and replaced by a 1:2000 His conjugated antibody in 1% milk in TBST. After two hours of incubation at room temperature with agitation, the membrane was washed three times for ten minutes with TBST. The membrane was then revealed with 2 mL of 1:1 chemiluminescence substrate and imaged.

Dot-blot

From cell lysis samples, 2µl are dropped onto a nitrocellulose membrane. The following steps are then the same as the western blotting, from the saturation step and until the revelation step.

Results

The first attempt to quantify the protein in *E. coli* in the producing strain was infructuous as no signal were observed for CrtY, CrtZ or the protein fusion. Because the T7 promoter used to expressed CrtY and CrtZ gene is also shared with all the other eleven enzymes of the zeaxanthin pathway, we assumed that the absence of signal was due to a low expression of the enzymes. The enzymes were

then transformed into the strain without the other enzymes of the pathway. This led to the detection of both CrtY enzymes and the protein fusions by SDS-PAGE (**Figure 29**). However, the protein fusion does not appear on the Ponceau staining and none of the proteins of interest were detectable by Western blot. Since the cytosolic protein was detected well, we assumed that the lack of detection in the Western-Blot was specific to our protein of interest and could be due to a problem of (i) antibody accessibility to the histidine tag epitope, (ii) an *in vivo* cleavage of the epitope which was expressed at the N-terminus of our protein of interest, (iii) a poor transfer of the proteins from the polyacrylamide gel to the western-blot membrane (at least for the protein fusion).

A dot-blot was performed to verify the presence and accessibility of the histidine tag by the

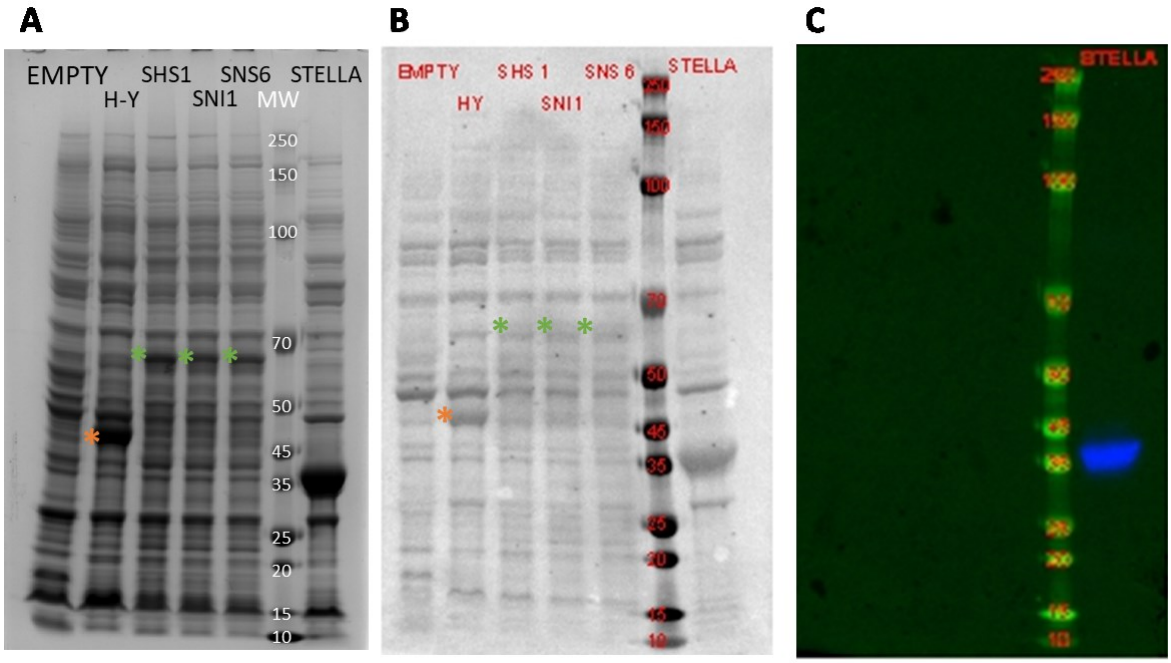


***Figure 29**: Protein detection in the one plasmid strain. **A:** SDS-PAGE with Coomassie blue staining, **B:** Ponceau red staining of PVDF membrane; **C:** Western-Blot. EMPTY: E. coli BL21 strain transformed with p15a plasmid; H-Y: E. coli strain transformed with p15a CrtY plasmid; SHS1: E. coli strain transformed with p15a CrtZ-SHS1-CrtY fusion protein; SNI1: E. coli strain transformed with p15a CrtZ-SNI1-CrtY fusion protein: SNS6: E. coli strain transformed with p15a CrtZ-SNS6-CrtY fusion protein; MW: Molecular Weight marker; STELLA: cytosolic control protein of 37 kDa. Orange asterisk indicate the presence of CrtY and green asterisk indicate the presence of a fusion protein (61 kDa).*

antibody. Indeed, a dot-blot and a western-blot share the same detection method, however, for a dot-blot the protein sample is directly loaded onto the blotting membrane, getting rid of both the SDS-PAGE run and the transfer from the polyacrylamide gel to the membrane. To also address the possible cleavage of the histidine tag, we also tested the detection of CrtY protein with a C-terminal histidine tag. In these conditions, only the cytosolic protein was detected while CrtY and CrtZ were not detected (**Figure 30**). We thus concluded that the absence of detection was most likely due to the lack of accessibility of the antibody to the epitope.

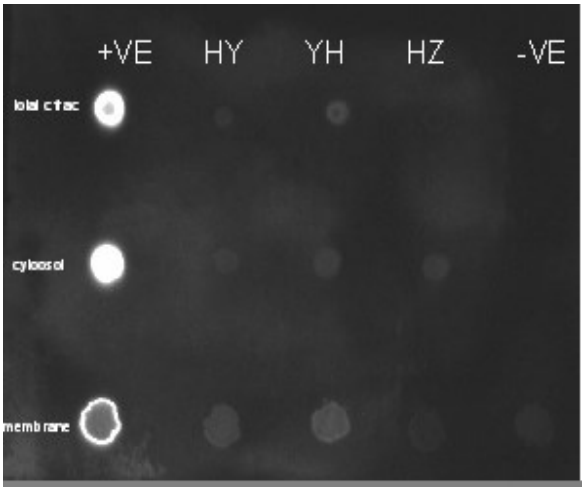


***Figure 30**: Dot blot of CrtY and CrtZ proteins. +ve: histidine tagged cytosolic protein; HY: CrtY with N-terminal histidine tag; HY: CrtY with C-terminal histidine tag; HZ: CrtZ with N-terminal histidine tag; -ve: empty plasmid. Top row: total cell fraction; middle row: cytosolic fraction; bottom row: membrane fraction.*

A spacer was thus added between the protein and the histidine tag (**Figure 31**). With the spacer, CrtY is detectable by SDS-PAGE but not CrtZ, probably due to low quantities being produced when expressed individually. It seems that the CrtY enzyme poorly transfer to the western-blot membrane (although it could also be due to a weaker detection by Ponceau red staining). When the histidine tag was added at the C-terminus of CrtY enzyme with a spacer, the enzyme was detectable by western-blot. This is an interesting as this might allow to detect protein fusion similarly. However, I did not have the time to test it.

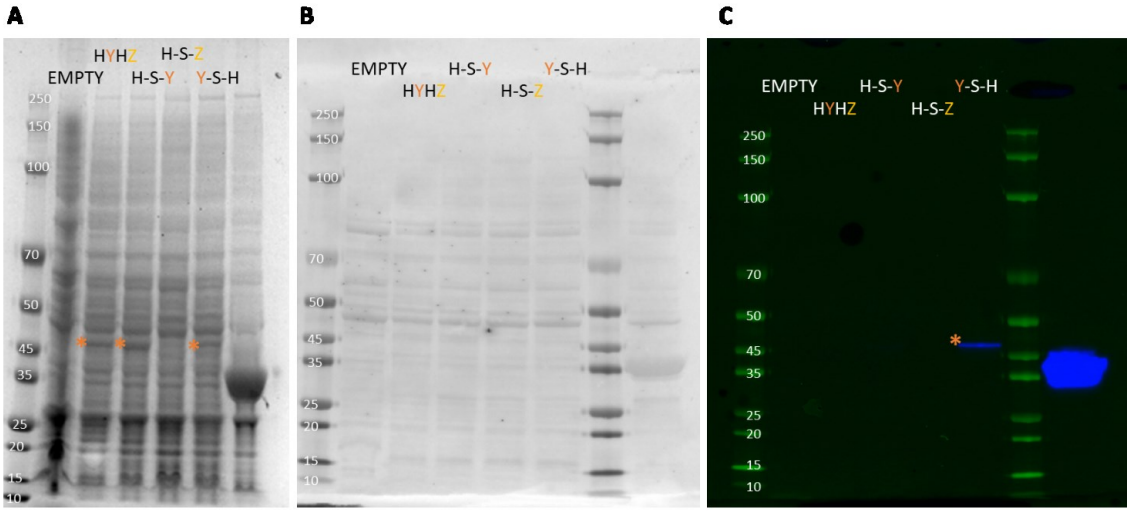


***Figure 31:** Detection of CrtY and CrtZ enzyme with a spacer added before the Histidine tag. **A:** SDS-PAGE with Coomassie blue staining, **B:** Ponceau red staining of PVDF membrane; **C:** Western-Blot.*

Conclusion

Although we have been able to detect our protein of interest We have not been able to optimize the method in order to quantify even in a relative manner the quantity of protein fusion produced in our strain of interest.

4. Prediction of the protein fusion 3D structure

Introduction

Similarly, to the protein detection section, the modeling section aims to understand better the results obtain during the study “Exploring linker's sequence diversity to fuse carotene cyclase and hydroxylase for zeaxanthin biosynthesis”. Ideally, we were looking for relation between the structure of the protein fusion and the outcome obtained in terms of carotenoids production.

Material and methods

The 3D structure predictions of protein fusions were obtained on AlphaFold Colab (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb). Pymol was used to visualize the predicted structures. For each run, five models were generated and ranked according to their binding energy prediction according to the PRODIGY webserver. The structure of protein fusion with lowest binding energy were kept for further analysis. All amino acids interactions between CrtZ and CrtY protein domain were also retrieved from the server.

Results

The starting hypothesis of this section was that the differences in carotenoid accumulation obtained among the strains with different linkers were due to a change in the orientation of both enzymes induced by the linker. Protein fusion resulting in the strains with either the best and worst total carotenoids amount or the best and worst zeaxanthin ratio were thus selected (**Table 3**) and the prediction of their structure was obtained from AlphaFold Colab.

Zeaxanthin ratio	LINKER	LINKER SEQUENCE
1,37	SHS01	APGRY
1,18	SHI01	AGQG
1,17	SHS05	GSGM
0,9	SNS03	APTV
0,88	SNS04	MVFG
1,02	SNI02	NIAFY
carotenoids (mg/L)	LINKER	LINKER SEQUENCE
69,5	SNI01	AVYE
71,4	SHS03	PEIRT
37,5	SNS03	APTV
33	SHS01	APGRY

***Table 3**: List of linkers for which the protein fusion structure has been predicted.*

The structure of the protein fusion with linkers resulting in the strain with the most striking difference in carotenoid profile, namely, SHS01, for its high zeaxanthin ratio, SNI01 for its high level of carotenoids accumulated and SNS06 for the absence of zeaxanthin accumulation, were aligned in Pymol (**Figure 32**). Although the structures are similar and the enzymes orientated in the same direction, which was expected as only four or five amino acids are changing in the whole protein sequence, differences between the three structures can be observed.

***Figure 32:** Alignment in Pymol of predicted fusion structure of: CrtZ-SHS01-CrtY (green), CrtZ-SNI01-CrtY (blue) and CrtZ-SNS06-CrtY (magenta). crtY is on the left side and CrtZ is on the right side of the picture.*

In order to quantify these differences, all the amino acids interactions between CrtY and CrtZ domain were retrieved from the PRODIGY webserver. Interestingly, the energy prediction structure

of the independent CrtZ and CrtY enzymes was around -8 kcal.mol^{-1} which is lower than any of the protein fusion. In total, CrtY and CrtZ domains of protein fusion had between 29 and 39 amino acids interacting together. To simplify the interpretation of the data, we only considered the amino acid interactions happening within 3 Angstroms. The amino acids interaction between CrtZ and CrtY domains in protein fusion resulting in the strains with the highest or lowest zeaxanthin ratio are summarized in **Figure 33** while the amino acids interaction between CrtZ and CrtY domains in protein fusion resulting in the strains with the highest or lowest total amount of carotenoids are summarized in **Figure 34**.

crtZ	crtY	SHS01	SHI01	SHS05	SNI02	SNS03	SNS04
PRO49	ARG305				X		
ARG50	GLN309				X		
LYS51	TRP308				X		
LYS51	GLN309				X		
ARG111	ARG312						X
PHE110	LEU355	X	X			X	
PHE110	ARG358	X	X			X	
ARG111	GLN309	X	X			X	
ARG111	GLN310		X		X		
ARG111	ARG312	X		X		X	X
ARG111	ARG358			X			
ILE113	GLN309						X
ILE113	GLN310	X				X	
ARG115	GLN310			X			X
ARG127	ALA287				X		
GLY186	MET01		X	X		X	X
GLY186	GLN02		X	X		X	X
TYR187	MET1	X			X		
TYR187	GLN2				X		

***Figure 33**: Comparison of amino acids interacting at less than a 3 Å distance between crtZ and crtY in protein fusion with high (green) or low (red) ratio of xanthophylls to carotenoids.*

As shown in both **Figure 33** and **Figure 34**, no common points were found between the protein fusions exhibiting a similar carotenoids accumulation outcome, neither in terms of total number of interactions, global binding energy or similitudes in the amino acids interacting together. There is no common point shared only by all protein fusion of a category of linkers (either high or low carotenoid producer for example) and not by the other.

crtZ	crtY	SHS03	SNI01	SHS05	SHI02	SNS05	SNS06	SNS03	SHS01
PRO49	ARG305	X							
LYS51	GLU249		X						
LYS51	GLN253		X						
LYS51	ARG284						X		
LYS51	HIS302						X		
LYS51	GLU306						X		
GLY52	GLU249		X						
GLY52	GLN253		X						
ALA53	GLU249		X						
GLN106	ARG312					X			
PRO109	LEU355	X							
PRO109	ARG358	X							
PHE110	LEU355	X						X	X
PHE110	ARG358							X	X
ARG111	GLN298		X						
ARG111	GLN309							X	X
ARG111	GLN310	X							
ARG111	ARG312			X	X	X		X	X
ARG111	ARG358			X	X				
ILE113	GLN309					X			
ILE113	GLN310							X	X
ARG115	GLN310			X	X	X			
MET186	MET01	X	X	X	X	X	X	X	
MET186	GLN02			X	X		X	X	
TYR187	MET1								X

Figure 34: *Comparison of amino acids interacting at less than a 3 Å distance between CrtZ and CrtY in protein fusion with high (green) or low (red) total amount of carotenoids.*

Conclusion

The comparison of the 3-dimensional predicted structures of the different protein fusion did not lead to a clear connection between the structure of the protein and the outcome in carotenoid profile observed in the corresponding strain. Several limitations could explain this result. First, the position of the linker in the protein structure. The linker is indeed placed at the N-terminus of CrtZ, which end in an unstructured region (**Figure 18**). This could result in an overall quite flexible region allowing movement between both enzymes. Since the models obtained only represent a snapshot of a possible conformation of the protein at one point in time, and some information could be missing. A molecular dynamic would allow to get more information from the models. This leads us to the second limitation which is the nature of the protein structure used in this study. Indeed, we used a protein structure without its iron cofactor, bound substrate or membrane environment! All three of these points could impact the models and should be considered to deepen our comprehension of the model.

II. Glycosylation

1. Introduction

Article summary

This publication aimed to produced glycosylated carotenoids (notably glycosylated zeaxanthin and glycosylated astaxanthin) which increase the solubility of the carotenoids. The project was carried out in *E. coli* and we report the highest production of glycosylated carotenoids produced by microbial engineering. As we suggested in the previous study, we believed that the position of CrtY near the membrane help increase the conversion of lycopene to downstream product. I was thus interested on the role of glycosylation on the carotenoids and their possible change of localization in the cell based on their properties.

Contribution

In this work, I participated in the elucidation of glycosylated and non-glycosylated carotenoids localization in *E. coli*.

Reference

This chapter refers to the following publication: Chen, X., Lim, X., **Bouin, A**. et al. High-level de novo biosynthesis of glycosylated zeaxanthin and astaxanthin in *Escherichia coli*. Bioresour. Bioprocess. 8, 67 (2021). https://doi.org/10.1186/s40643-021-00415-0

2. Article

Abstract

Because of wide applications in food, feed, pharmaceutical and cosmetic industries, the carotenoid market is growing rapidly. Most carotenoids are hydrophobic, which limits their bioavailability. Glycosylation is a natural route that substantially increases the water solubility, as well as the bioavailability, photostability and biological activities of carotenoids. Here, we report metabolic engineering efforts (e.g., promoter and RBS engineering, optimization of carbon sources and supplementation of bottleneck genes) to produce glycosylated carotenoids in *Escherichia coli*. By finetuning the carotenoid-biosynthetic genes (crtX, crtZ and crtY), our strain produced up to 47.2 mg/L (~ 11,670 ppm) of zeaxanthin glucosides, ~ 78% of the total carotenoids produced. In another construct with mevalonate, astaxanthin pathway and crtX genes, the strain produced a mixture of carotenoid glucosides including astaxanthin and adonixanthin glucosides with a total yield of 8.1 mg/L

(1774 ppm). Our work demonstrated a proof-of-concept study for the microbial biosynthesis of glycosylated carotenoids.

Introduction

Carotenoids (> 1100) are natural pigments widely distributed in plants, animals, algae and microbes (Yabuzaki, 2017; Zhang, 2018). The structures of carotenoids typically consist of an electron-rich polyene chain with nine or more conjugated double bonds. This unique feature contributes primarily to their photoprotection and light-harvesting properties, antioxidant activities to quench free radicals and singlet oxygen, and vivid colors (Sandmann, 2019). Carotenoids function as photosynthesis and photoprotection agents in photosynthetic organisms (e.g., plants and algae) and protect non-photosynthetic organisms (e.g., bacteria, archaea and fungi) from photooxidative damages (Hashimoto, Uragami and Cogdell, 2016). Carotenoids also serve as structural molecules by integrating in lipid membranes, hence, modulating membrane fluidity (Richter, Hughes and Moore, 2015). Because of these properties, especially for the pigment and health benefits, carotenoids have various applications in food, feed, nutraceutical and pharmaceutical industries, and the industrial demand is growing rapidly. For example, the global market of astaxanthin is projected to reach $2.57 billion worldwide by 2025 (Zhang, Chen and Too, 2020).

However, most natural carotenoids are lipophilic and hardly soluble in water. The hydrophobicity of carotenoids limits their application in medicine and food where enhanced water dispensability is required to facilitate their effective uptake or use (Dembitsky, 2005; Háda *et al.*, 2012). Therefore, several attempts, mainly chemical approaches (e.g., converting carotenoids to salts of carotenoid esters, or forming carotenoid–cyclodextrin complex), have been made to increase the carotenoid hydrophilicity (Háda *et al.*, 2012). Alternatively, glycosylation is an excellent natural way to increase carotenoid solubility. In nature, a large number of hydrophobic natural products (e.g., lipids and terpenes) are glycosylated into more water-soluble products by glycosyltransferases (Elshahawi *et al.*, 2015). In fact, water-soluble carotenoids, although rare, are present in nature, such as crocins (or glycosyl polyene esters) in saffron (Dembitsky, 2005). In addition, several other glycosylated carotenoids are uncovered in various microbes, such as zeaxanthin glucoside (Misawa *et al.*, 1990), astaxanthin glucoside (Yokoyama, Shizuri and Misawa, 1998), adonixanthin-β-D-glucoside (Yokoyama, Adachi and Shizuri, 1995), sioxanthin (Richter, Hughes and Moore, 2015) and a C50 decaprenoxanthin diglucoside (Krubasik *et al.*, 2001).

Natural metabolites are typically produced meaningfully with biological functions for host living organisms. Primary metabolites are synthesized to support their growth and development. Secondary metabolites typically increase the competitiveness of the organism within its environment. Likewise, glycosylated carotenoids should have meaningful functions for their hosts. It is reported that glycosylated carotenoids play important roles in maintaining cell wall structure and their localization stabilizes the thylakoid membrane in cyanobacteria where the glycosyl moiety serves as a binding motif that enables the proper folding and stacking of the thylakoid membrane (Mohamed *et al.*, 2005).

The first bacterial gene that encodes the enzyme to catalyze carotenoid glycosylation was identified in *Pantoea ananatis* (previously as *Erwinia uredovora*) (Misawa *et al.*, 1990) and it was reported that glycosylation can alter carotenoid deposition in plants (Wurtzel, 2019). As a phytopathogen, this might contribute to the virulence of *P. ananatis* with host plant cells. Moreover, carotenoid glucosides contribute to the heat resistance of the Thermus species, and hence, are also named thermoxanthins (Háda *et al.*, 2012). As for commercial applications, apart from improved water solubility (e.g., the solubility of zeaxanthin, zeaxanthin mono- and diglucosides are 12.6, 100 and 800 ppm in water, respectively (Hundle *et al.*, 1992), glycosylation of carotenoids also leads to structural diversity and several other benefits, such as increased bioavailability and efficacy as food supplements and medicines, and improved photostability (Polyakov *et al.*, 2009) and biological activities (e.g., antioxidant activity) of carotenoids (Matsushita *et al.*, 2000). It is proposed that the increase in antioxidant activities is not from their intrinsic ability of additional glucosides to scavenge free radicals, but arises from the enhanced affinity with singlet oxygen, the location and orientation in cells (Matsushita *et al.*, 2000; Choi *et al.*, 2013).

Carotenoids are glycosylated by glycosyltransferases (GTs), which is a large enzyme family. GTs typically catalyze a hydroxyl or carboxyl group of lipophilic substrates as the substituent moiety for glycosylation. For carotenoid glycosylation, the hydroxyl group is the commonest substituent moiety, and the carotenoid GTs belong to GT family 1 or GT1. Uridine diphosphate-αD-glucose (UDP-glucose) is the most abundant sugar donor to carotenoid glycosylation. In addition, other sugars such as L-rhamnose, L-fucose, D-xylose and L-quinovose can also be recruited especially in cyanobacteria (Choi *et al.*, 2013).

To date, only a couple of studies have demonstrated the biosynthesis of carotenoid glucosides in Escherichia coli and in several natural microbial producers (Misawa *et al.*, 1990; Yokoyama, Adachi and Shizuri, 1995; Yokoyama, Shizuri and Misawa, 1998; Choi *et al.*, 2013). However, these studies only produced detectable amount of carotenoid glucosides and were far from the minimal requirement for industrial applications. Here, using the zeaxanthin glucosyltransferase (ZGT, the gene crtX, UniProt ID D4GFK6) from *P. ananatis*, we have constructed a 14- and 15-gene pathway in *E. coli* to synthesize various carotenoid glucosides, such as zeaxanthin D-glucoside (yellow) and astaxanthin D-glucoside (red). The carotenoid yields have been improved by rational metabolic engineering approaches and bioprocess optimization.

Materials and methods

Strain and plasmid construction

E. coli Bl21-Gold DE3 strain (Stratagene) was used in this study. The plasmids p15A-spec-hmgS-atoB-hmgR (L28), p15A-spec-crtY-hmgS-atoB-hmgR (L2-8) p15A-cammevK-pmk-pmd-idi (L2-5), p15A-kan-crtEBI-ispA were designed as previously described (C. Zhang *et al.*, 2018). The zeaxanthin GT gene (crtX) from *Pantoea ananatis* was codon optimized (DNA sequence was provided in Additional file 1:

Supplementary note) and synthesized by Integrated DNA Technologies, Singapore. Subsequently, crtX was cloned with the primers (Additional file 1: Table S2) into the operon of the plasmids p15Aamp-crtYZ (L2-9) and p15A-amp-crtYZW (L2-9) (C. Zhang *et al.*, 2018) to obtain p15A-amp-crtYZX and p15A-ampcrtYZWX, respectively.

Construction of RBS library

CrtZ RBS library was created using the degenerate primer and followed by screening and sequencing validations, using the same cloning method as previously described (Zhang et al. 2018). RBS strengths or translation efficiencies were predicted by RBS Calculator, version 2.0 (Farasat *et al.*, 2014)

Tube culture of the *E. coli* strains

The medium used was TB medium (20 g/L tryptone, 24 g/L yeast extract, 17 mM KH2PO4, and 72 mM K2HPO4) and 2XPY medium (20 g/L peptone, 10 g/L yeast extract and 10g/L NaCl), supplemented with 10 g/L glycerol or 10–20 g/L glucose or their mixture (5 g/L glucose + 5 g/L glycerol), 50 mM 4-(2-hydroxyethyl)1-piperazineethanesulfonic acid (HEPES), as previously described (Zhang et al. 2018). For strain optimization, the cells were grown in 1mL of TB or 2XPY medium in 14ml BD Falcon™ tube at 28°C/250 rpm for 2–3days. The cells were also grown in 50 mL culture in shaking flasks for validation of the carotenoid production. The cells were initially grown at 37°C/250 rpm until OD_{600} reached ~ 0.8, induced by 0.03–0.1mM IPTG, and were subsequently grown at 28°C for 2 days. The antibiotics (34 μg/ml chloramphenicol, 50 μg/ml kanamycin, 50μg/ ml spectinomycin and 100 μg/ml ampicillin) were supplemented in the culture to maintain the four plasmids.

Microscope imaging of *E. coli* cells

For microscopy assay, *E. coli* cells were directly sampled from cell cultures. Cell amount was normalized by OD600 and directly observed at 1000 magnitude using a Leica DM6000B microscope. Neither centrifuge nor washing steps were introduced to avoid perturbation of the cell morphologies.

Extraction and quantification of carotenoids

Total intracellular carotenoids were extracted from cellular pellets according to the acetone extraction method (Zhang et al. 2018). Briefly, 10–50 μL bacterial culture (depending on the content of carotenoids in the cells) was collected and centrifuged. Cell pellets were washed with PBS and were resuspended in 20 μL of water, followed by addition of 180 μL of acetone and vigorous homogenization for 20 min. After 10 min of centrifugation at 14,000 g, the supernatant was collected and filtered using a PTFE, 0.45 μm filter. The separation of carotenoids from cytosol and cell membranes was done by differential centrifugation. Briefly, cell pellets collected from 1 mL of culture were resuspended in 1 ml lysis buffer (50 mM Tris HCl of pH 7.5, 200 mM NaCl, 1 mg/ml lysozyme of pH 8) before 3 × 30 s sonication at 4 °C (75% amplitude). The cell lysate was subsequently

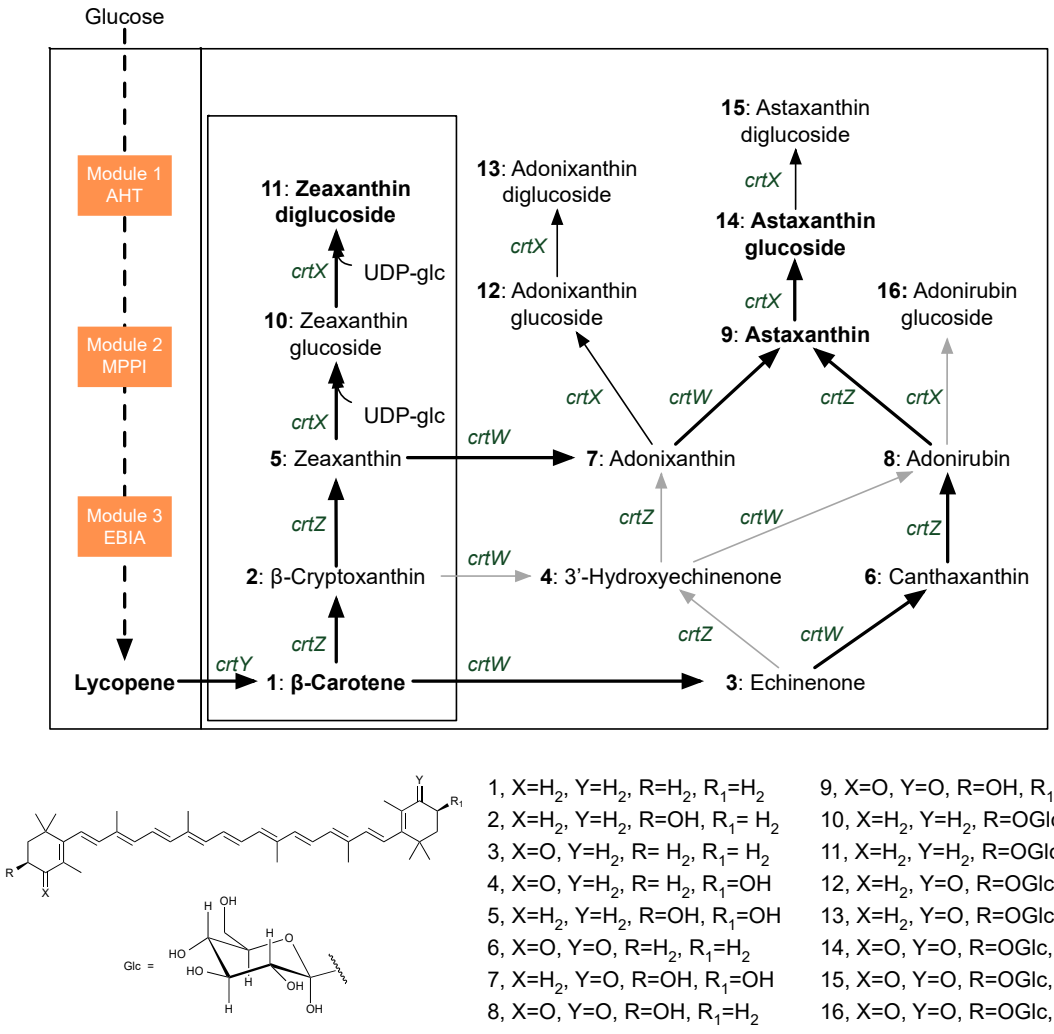
centrifugated for 10 min at 14,000 g. The supernatant containing the cytosol fraction of carotenoids and the pellet debris containing the membrane fraction were extracted separately with by 1 mL of extraction buffer (hexane: acetone: ethanol at 2:1:1 volumetric ratio).

Quantification of carotenoids

All the carotenoids were analyzed by Agilent 1290 Infinity II UHPLC System coupled with Diode Array Detector (DAD) detector and 6230B TOF-MS platform. The LC/MS method was similar to previously described (Zhang et al. 2018). Briefly, 1 μL of purified carotenoids in acetone was injected into the Agilent ZORBAX RRHD Eclipse Plus C18 2.1X50 mm, 1.8 um. Separation was carried out at a flow rate of 0.5 mL/min. The mobile phase and gradient used were as follows. The analysis started from 10% water (0.1% formic acid), 10% methanol (0.1% formic acid) and 80% acetonitrile (0.1% formic acid) and this condition was maintained for 2 min, followed by the increase in methanol from 10 to 90% and the decrease in water from 10% to 0 and acetonitrile from 80 to 10% within 0.1 min. The condition (90% methanol and 10% acetonitrile) was continued for 7 min. The whole analysis finished at 10 min. Mass spectrometry was operated to scan 100–1100 m/z in ESI-positive mode with 4000 V capillary voltage. Nebulizer gas was supplied at 35 psig and dry gas flow was 10 L/min. Gas temperature was set at 325°C. Shealth gas was set at 350°C and 12 L/min. Retention time was determined with chemical standards or calculated based on chromatography profile for those carotenoids without standards.

Carotenoid concentrations were calculated based on the peak area of each compound extracted by their corresponding m/z value (Table 1) or UV absorbance at 450 nm (Additional file 1: Figure S2). Standard curves were generated for the five chemical standards with extracted-ion chromatogram (EIC) peak areas (Additional file 1: Figure S3): lycopene, β-carotene, astaxanthin, canthaxanthin (Sigma-Aldrich, St. Luis, MO, USA), and zeaxanthin (Santa Cruz Biotechnology, Dallas, TX, USA). For those carotenoids without standards, the concentration was calculated based on the relative peak area to its close compartment. For example, the concentrations of zeaxanthin glucoside and zeaxanthin diglucoside were calculated based on that of zeaxanthin; the concentrations of astaxanthin glucosides, adonixanthin and its diglucosides were calculated based on that of astaxanthin. Carotenoid contents were calculated by normalizing the titres with dry cell weight (μg carotenoids per gram DCW, or ppm) (Zhang et al. 2018).

The pathway design for glycosylated carotenoids



1, X=H_2, Y=H_2, R=H_2, R_1=H_2
2, X=H_2, Y=H_2, R=OH, R_1= H_2
3, X=O, Y=H_2, R= H_2, R_1= H_2
4, X=O, Y=H_2, R= H_2, R_1=OH
5, X=H_2, Y=H_2, R=OH, R_1=OH
6, X=O, Y=O, R=H_2, R_1=H_2
7, X=H_2, Y=O, R=OH, R_1=OH
8, X=O, Y=O, R=OH, R_1=H_2
9, X=O, Y=O, R=OH, R_1=OH
10, X=H_2, Y=H_2, R=OGlc, R_1=OH
11, X=H_2, Y=H_2, R=OGlc, R_1=OGlc
12, X=H_2, Y=O, R=OGlc, R_1=OH
13, X=H_2, Y=O, R=OGlc, R_1=OGlc
14, X=O, Y=O, R=OGlc, R_1=OH
15, X=O, Y=O, R=OGlc, R_1=OGlc
16, X=O, Y=O, R=OGlc, R_1=H_2

Figure 35: *Biosynthetic pathway of carotenoid glucosides. The biosynthetic pathway: module 1 AHT, including atoB, hmgS and thmgR; module 2 MPPI, including mevk, pmk, pmd and idi; module 3 EBIA, including crtEBI and ispA (Zhang et al. 2018); and module 4 YZX or YZWX, including crtYZX or crtYZWX. Dashed arrow indicates multiple enzymatic steps. The glycosylation of all carotenoids required UDP-glucose (UDP-glc), here we only used zeaxanthin glucosides as representatives. The genes expressed encode the following enzymes: crtY, lycopene beta-cyclase; crtW, β-carotene ketolase; crtZ, β-carotene hydroxylase; crtX, zeaxanthin glucosyltransferase (ZGT). Thicker and thinner arrows represent the higher and lower carbon flux, respectively; gray arrows represent that the metabolites (e.g., β-cryptoxanthin-β-D-glucoside and 3'-hydroxyechinenone-β-D-glucoside) were not detected in our strains.*

The metabolic pathway for glycosylated carotenoids was designed on top of our previous optimized astaxanthin strain (Zhang et al. 2018). Briefly, the mevalonate pathway genes were cloned into the modules 1 (AHT, the genes atoB, hmgB and truncated hmgR) and 2 (MPPI, the genes mevk,

pmk, pmd and idi) and the lycopene pathway genes (crtEBI and ispA) were located in module 3 (EBIA). The last module (module 4, YZX or YZWX) consists of the genes to produce zeaxanthin glucosides (crtY, crtZ, and crtX) or to produce astaxanthin glucosides (crtY, crtZ, crtW, and crtX) (**Figure 35**). All the modules were controlled by T7 and its variants (e.g., TM1, TM2 and TM3) and induced by isopropyl β-d-1thiogalactopyranoside (IPTG) (Zhang *et al.*, 2015). This modular arrangement provides the flexibility to balance the global pathways (14–15 genes) and to fine tune the local pathways (e.g., module 4). In addition, as the module 4 controls the cyclization (crtY), hydroxylation (crtZ), ketolation (crtW), and glycosylation (crtX) of carotenoids, it is relatively simple to switch from one carotenoid (e.g., using crtYZ to produce zeaxanthin) to another one (e.g., using crtYZWX to astaxanthin glucoside) without modifying the upstream pathways genes.

The production of glycosylated zeaxanthin

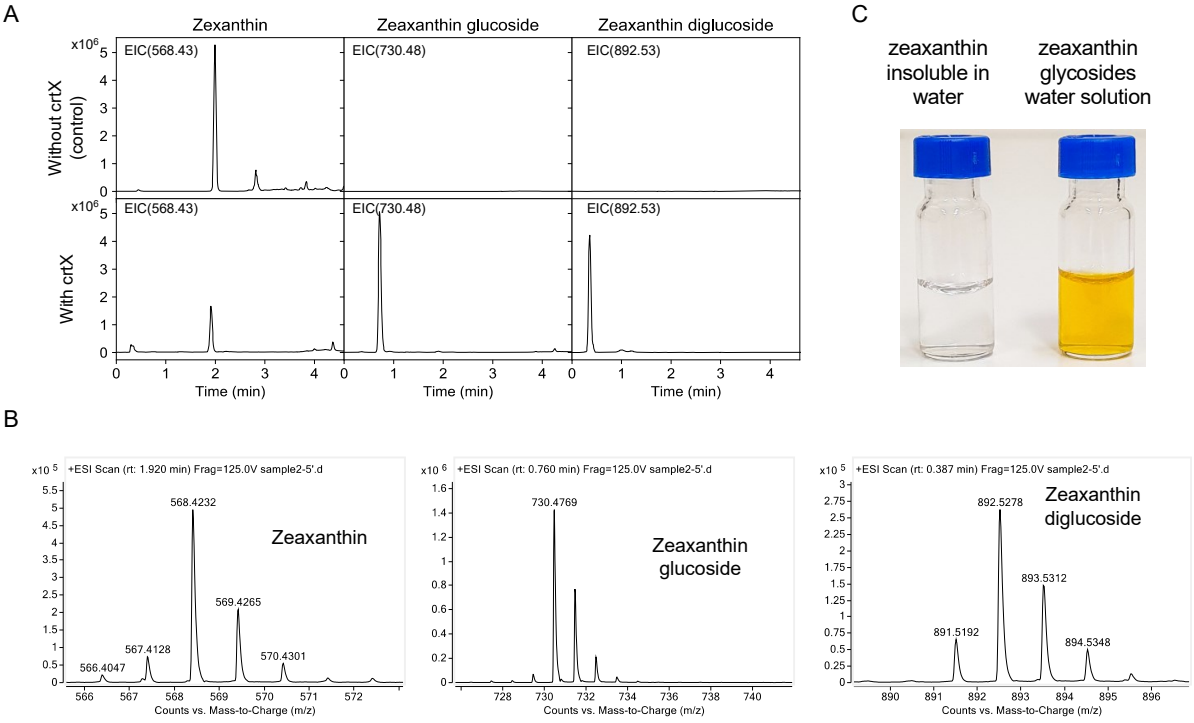
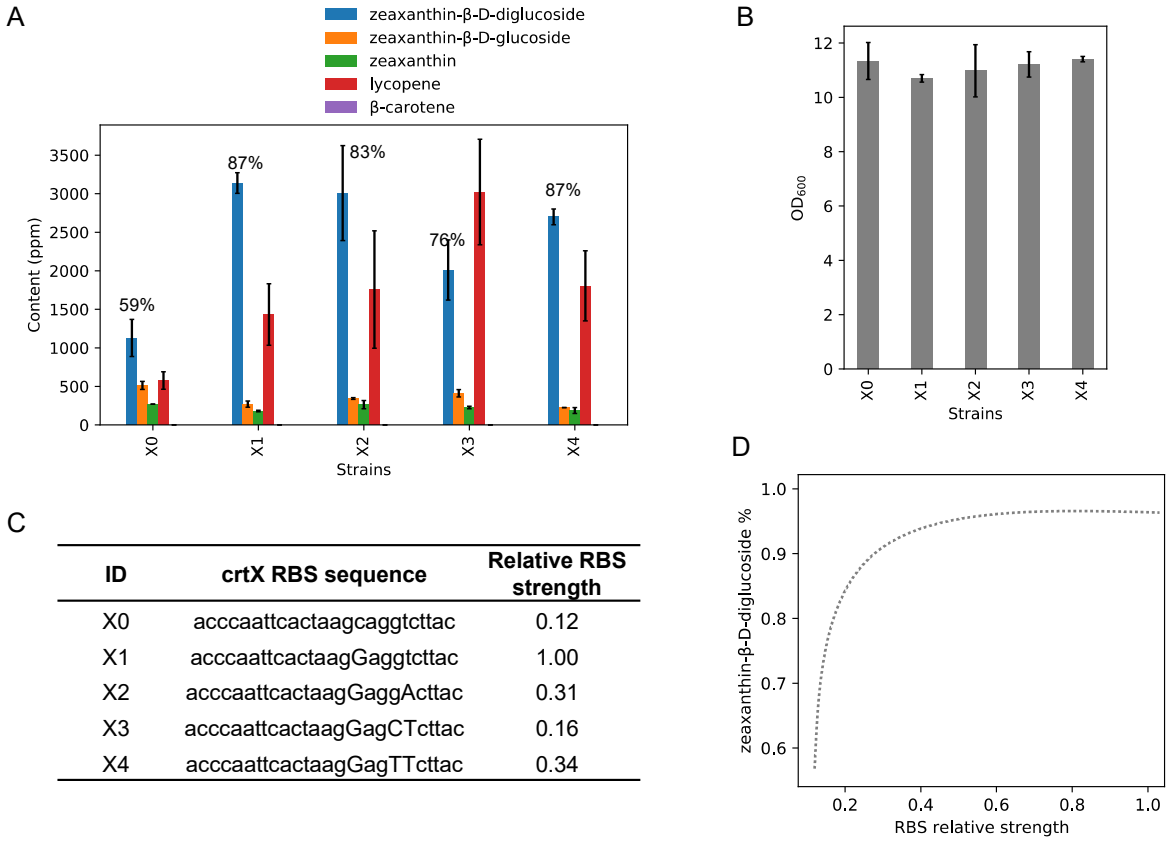


***Figure 36**: Production of zeaxanthin glucosides. A LC/MS chromatograms of zeaxanthin strains with and without the expression of crtX. B Mass spectra of zeaxanthin and its glucosides. C The water solutions of zeaxanthin and zeaxanthin glucosides*

Before we produce glycosylated zeaxanthin, we first optimized a strain that produces zeaxanthin (the last module contains crtY and crtZ, or module YZ), the combination of TM3-AHT, TM2-MPPI, TM2-EBIA and T7-YZ resulted in the best production of zeaxanthin (~ 12,000 ppm or 51.8 mg/L). On top of this strain with the same modules 1–3, we introduced module 4 (YZX) to demonstrate the capability to produce zeaxanthin glucoside. We developed a LC–TOF-MS method to detect the carotenoids and their glucosides (summary in Table 1). In the constructed strain with crtX, we managed to detect five carotenoids: lycopene, β-carotene, zeaxanthin, zeaxanthin-β-D-glucoside and zeaxanthinβ-D-diglucoside (Additional file 1: Figure S1); whereas, the control strain without crtX did

not produce either glycosylated zeaxanthin (**Figure 36**, A). The intermediate β-cryptoxanthin was not detected in either strain. The LC chromatograms and mass spectra for zeaxanthin (m/z 568.428, Table 1), zeaxanthin-β-D-glucoside (m/z 730.481) and zeaxanthin-β-D-diglucoside (m/z 892.534) are shown in **Figure 36** (A, B). In addition, we also purified some zeaxanthin glucosides from the strain with crtX and obtained a yellow aqueous solution (~ 30 mg/L). In contrast, zeaxanthin barely dissolves in water leading to a transparent water solution (**Figure 36**, C).

Optimization of glycosylation of zeaxanthin



C

ID	crtX RBS sequence	Relative RBS strength
X0	acccaattcactaagcaggtcttac	0.12
X1	acccaattcactaagGaggtcttac	1.00
X2	acccaattcactaagGaggActtac	0.31
X3	acccaattcactaagGagCTcttac	0.16
X4	acccaattcactaagGagTTcttac	0.34

***Figure 37**: Tuning the translation of zeaxanthin glucosyltransferase. A Carotenoid contents of zeaxanthin glucoside strains. 'zea' strain is the parental zeaxanthin strain without expressing crtX. B OD600 of different strains. Error bars, mean ± s.d., n = 2 or 3. C Different RBSs used for crtX and their relative strengths. D Correlation between the glycosylation efficiency of zeaxanthin and the RBS strength of crtX. The glycosylation efficiency is defined as the percentage of zeaxanthin diglucoside yield to the total yield of zeaxanthin and its two glucosides.*

In our first design strain X0, the glycosylation of zeaxanthin was incomplete: ~ 26.8% of monoglycosylated and 59.0% of diglycosylated (here the percentage was calculated by normalizing to the total yield of zeaxanthin and its two glucosides) and 14.2% of zeaxanthin remained unglycosylated (**Figure 37**, A, B). We hypothesized that glycosylation of zeaxanthin could be limited by insufficient activity of ZGT. To test it, we re-designed another four ribosomal-binding sites (RBSs) of crtX which have relatively higher translational efficiencies than the initial RBS in strain X0 (**Figure 37**, C). Indeed,

we observed that using stronger RBS for ZGT (crtX) led to higher glycosylation of zeaxanthin (**Figure 37**, A, D). Strain X1 had the strongest RBS and produced the highest amount of zeaxanthin-βD-diglucoside (~ 3139 ppm and ~ 87.4% of total zeaxanthin and its glucosides). We attempted to correlate RBS strengths to zeaxanthin-β-D-diglucoside production. Zeaxanthin-β-D-diglucoside produced appears to reach a saturated percentage when RBS relative strength was higher than 0.3 (**Figure 37**, D). It was noteworthy that the total yield of carotenoids in zeaxanthin glucoside strains (X0X4) was about 50–80% lower than that of parental zeaxanthin strain (zea).

Next, we evaluated the effect of different carbon sources on the biosynthesis of zeaxanthin glucosides. As an abundant and inexpensive carbon source, we chose glucose and hypothesized that glucose might be advantageous to supply additional UDP-glucose, which is the key cofactor for carotenoid glycosylation. UDP-glucose can be produced from glucose with three enzymes: glk: glucokinase, pgm: phosphoglucomutase, galU: UDP-glucose pyrophosphorylase (Mao, Shin and Chen, 2006; Shrestha *et al.*, 2019). In addition, we also chose glycerol as it is inexpensive and was reported to favor carotenoid production (Zhang *et al.*, 2013). For X1 strain, the glucose supplementation (10 g/L) led to higher production of zeaxanthin glucosides (~ 3650 ppm) than the supplementation of 10 g/L of glycerol or the mixture of glucose (5 g/L) and glycerol (5 g/L) (**Figure 38**, A). Subsequently, we increased the amount of supplemented glucose from 10 to 20 g/L, the yield of zeaxanthin diglucoside was further increased from ~ 3400 (or 15.1 mg/L) to~ 4690 ppm (or 25.3 mg/L). At the same time, OD600 was also increased from 10.8 to 13.1 (**Figure 38**, B). Of the total carotenoids produced including lycopene and β-carotene, zeaxanthin glucosides reached about 64% in X1 strain.

In addition, we also observed that lycopene was accumulated as the main intermediate carotenoid for all the strains and conditions in **Figure 37** (A), **Figure 38** (A). We hypothesized that the accumulation of lycopene could arise from the insufficient activity of lycopene cyclase (or crtY, **Figure 35**). Indeed, the introduction of extra copies of crtY (" + crtY" strain) significantly boosted zeaxanthin diglucoside yield from 3400 to 7150 ppm (or 23.1 mg/L) and zeaxanthin glucoside yield from 350 to 4520 ppm (14.6 mg/L) in the medium supplemented with 10 g/L glucose (**Figure 38**, B). Furthermore, for the " + crtY" strain, the titres of zeaxanthin diglucoside and glucoside were further increased to 31.0 and 16.3 mg/L, respectively, as the supplemented glucose was increased from 10 to 20 g/L (**Figure 38**, B). Lastly, the yields of zeaxanthin glucosides of " + crtY" strain were about 78% of that of total carotenoids produced.

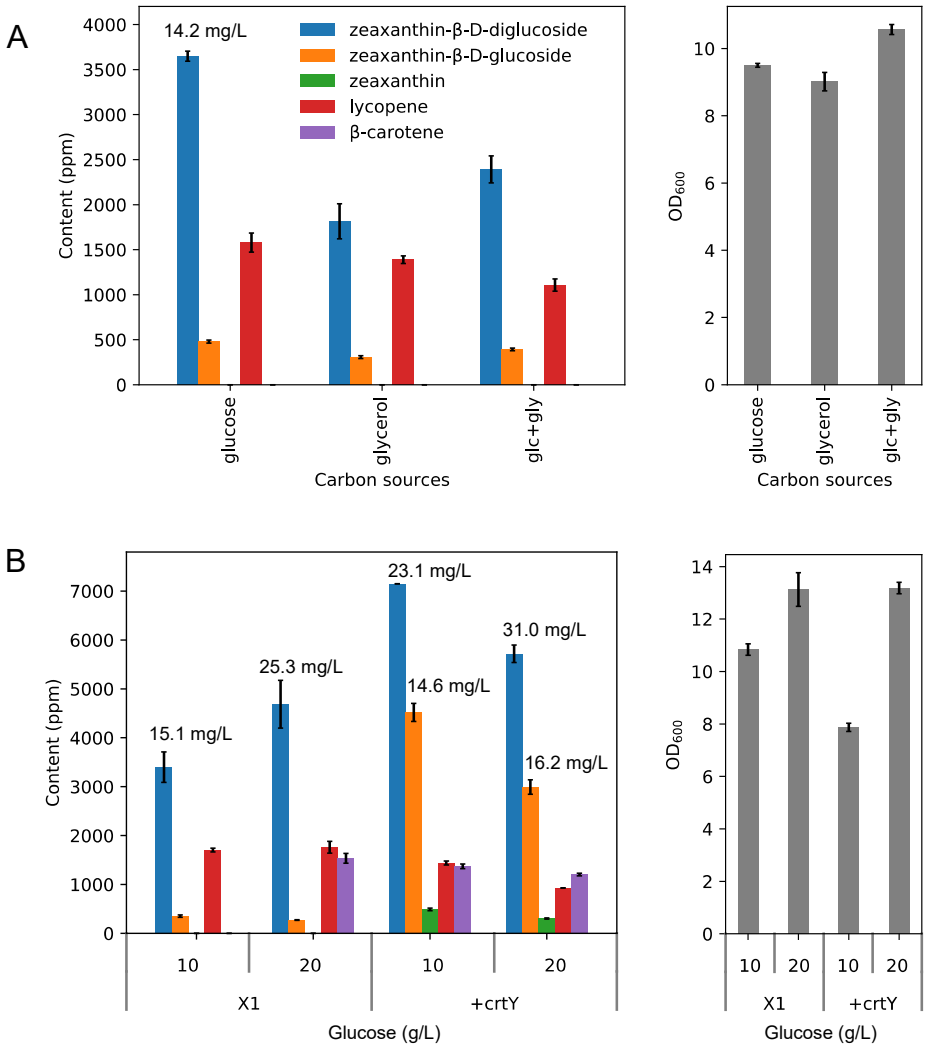


Figure 38: *The effect of carbon sources on the production of zeaxanthin glucosides.* ***A:*** *Carotenoid contents and OD600 of strain X1 by comparison of different carbon sources: 10 g/L glucose, 10 g/L glycerol and their mixture, 5 g/L glucose + 5 g/L glycerol (glc + gly).* ***B:*** *Carotenoid contents and OD600 of strains X1 and " + crtY" by optimizing the concentrations of glucose and introduction of additional copies of crtY. Error bars, mean ± s.d., n=2*

Distribution of carotenoids in *E. coli* cells

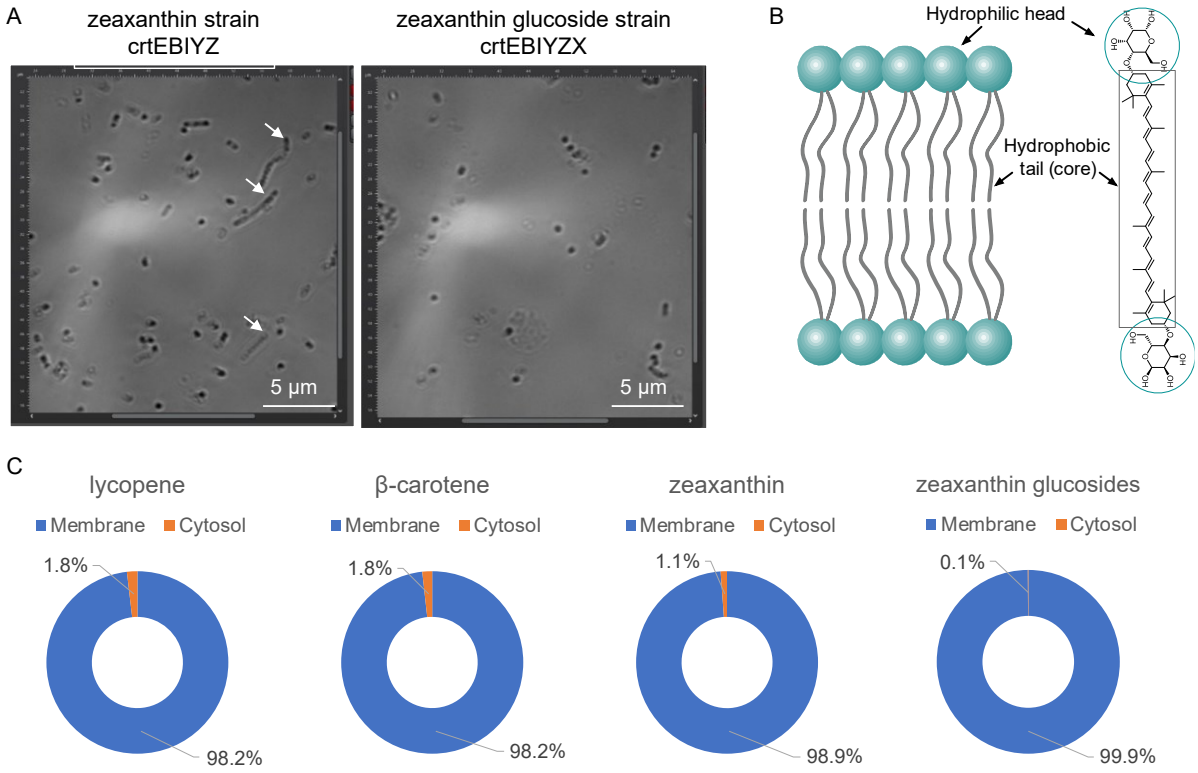


Figure 39: *Structural similarity between membrane and carotenoid diglucosides and its biological benefits.* ***A****: Comparison between zeaxanthin and zeaxanthin glucosides strains.* ***B:*** *Carotenoid distribution between cytosol and membrane.* ***C:*** *Structural similarity between phospholipid bilayers and zeaxanthin diglucoside and their dimensions.*

While studying the zeaxanthin glucoside strain, we observed that some cells of zeaxanthin production strain were longer than others in microscopes (**Figure 39**, A). In comparison, there were no elongated cells for zeaxanthin glucoside production strain. We wondered if the cell shape difference was attributed to the higher hydrophilicity of zeaxanthin glucosides so that most zeaxanthin glucosides may be distributed in cytosol. To test the hypothesis, we analyzed the distribution of carotenoids between cytosol and membrane. Unexpectedly, it was found that all the four carotenoids (lycopene, β-carotene, zeaxanthin and zeaxanthin glucosides) were predominantly localized in membrane (**Figure 39**, B). Less than 2% of them were present in cytosol. In addition, less zeaxanthin glucosides (0.08%) was distributed in cytosol as compared to zeaxanthin (1.13%). Our data supported the notion that zeaxanthin and its glucosides might have higher affinity with membrane than cytosol. Structurally, the glucoside and carotene of carotenoid glucosides resemble the hydrophilic head and the hydrophobic tail of phospholipid bilayers, respectively; also, the dimension of bilayer inner membrane (37.5 ± 0.5 Å) (Mitra *et al.*, 2004) is close to that of zeaxanthin diglucoside (~ 30 Å) (**Figure 39**, B). Carotenoid glucosides are reported to be clustered in rigid patches and such local rigidity can protect the membrane integrity under internal or external stress (e.g., oxidative and extreme temperature) (Mohamed *et al.*, 2005). This might attribute to cell shape difference between

zeaxanthin and zeaxanthin glucoside-producing cells, and further study is warranted to explore the mechanism.

The production of glycosylated astaxanthin

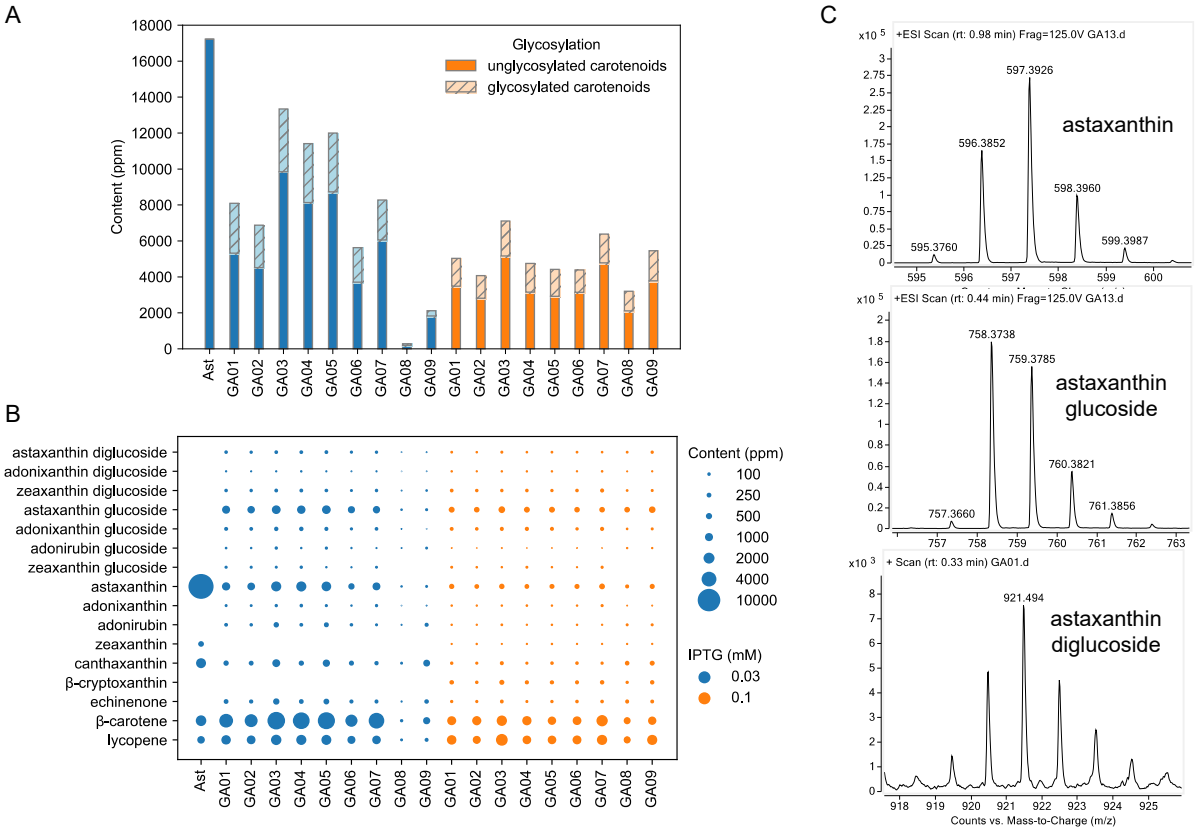


Figure 40: *Production of astaxanthin glucosides and other carotenoids.* ***A:*** *The content sums of glycosylated and unglycosylated carotenoids in different strains. The arrows refer to the two control strains ('Ast' and 'GA01').* ***B:*** *Carotenoid contents produced in different strains. Blue: 0.03 mM IPTG; orange: 0.1 mM IPTG. 'Ast' strain is the parental astaxanthin strain without expressing crtX. 'GA01' is the control strain with the highest RBS strength of crtZ.* ***C:*** *Mass spectra of astaxanthin and its glucosides.*

After demonstrating our design was working for zeaxanthin glycosylation, we further tested the other design with module YZWX to produce astaxanthin glucosides. With the addition of the gene crtX in one of our best astaxanthin producer strains (Ast strain, **Figure 40**, A, B) (Zhang et al. 2018), we tested the astaxanthin glycosylation capability (the resulting strain was named GA01). Overall, seven carotenoid glucosides are detected in GA01: zeaxanthin-β-D-glucoside, adonirubin-β-Dglucoside (m/z 742.444), adonixanthin-β-D-glucoside (m/z 744.460), astaxanthin-β-D-glucoside (m/z 758.439), zeaxanthin-β-D-diglucoside, adonixanthinβ-D-diglucoside (m/z 906.513) and astaxanthin-β-Ddiglucoside (m/z 920.492, **Figure 40**, A, B, Table 1, mass spectra **Figure 40**, C and Additional file 1: Figure S2, and LC chromatograms in Additional file 1: Figure S3 and S4). Among them, astaxanthin-β-D-glucoside was the main glycosylated product with a yield of 4.51 mg/L (968 ppm), about 68% of

total carotenoid glucosides. In addition, about 4.82 mg/L astaxanthin (1035 ppm) was not glycosylated and larger amount of β-carotene (16.0 mg/L, 3426 ppm) remained in GA01 strain. Furthermore, we observed that the introduction of crtX resulted in a 54% decrease of the total carotenoid yields in GA01 strain, as compared to its parental Ast strain (**Figure 40**, A), which might be due to the overall perturbation to the mevalonate and carotenoid pathway carbon fluxes or feedback regulations.

Here, we would like to highlight that in our conditions (ESI mode, with water, methanol and acetonitrile as mobile phase), the detected molecular-related ions of carotenoids had two kinds: [M]+. and [M + H]+. Different carotenoids can have different ratios of these two ion species. For zeaxanthin and its glucosides, [M]+. was predominant (**Figure 36**, B), while for astaxanthin and its glucosides, both [M]+. and [M + H]+ co-existed (**Figure 39**, C). This phenomenon was observed previously (Rivera, Christou and Canela-Garayoa, 2014), and it was found that the high polyene conjugation, the presence of oxygen in carotenoids and solvent system have a strong impact on the formation and stability of molecular ion species (Rivera, Christou and Canela-Garayoa, 2014).

Optimization of glycosylation of astaxanthin

Moreover, the higher IPTG concentration reduced the total yield of glycosylated carotenoids from 6.61 to 3.60 mg/L (1418 to 799 ppm) and non-glycosylated (or aglycones) carotenoids from 24.8 to 15.7 mg/L (5320 to 3485 ppm, **Figure 40**, A), possibly because IPTG perturbed the whole biosynthetic pathway where all the genes were controlled by T7 promoter variants and/or it promoted a competition between CrtZ and CrtW with intermediate accumulation (**Figure 35**). It has been observed that the translational efficiency of the β-carotene hydroxylase (crtZ) is more crucial than that of β-carotene ketolase (crtW) on astaxanthin production (Zhang et al. 2018). Therefore, we used nine different ribosomal-binding sites (RBSs, Additional file 1: Table S1) covering from 1 to 100% of translational efficiencies (the strains were named G0109, translational efficiencies were normalized to that of strain GA01, the strongest among them) to optimize the production of glycosylated carotenoids, especially glycosylated astaxanthin.

Essentially, GA01–09 were strains with the same design except for the different RBSs of crtZ (Additional file 1: Table S1). Indeed, the RBS had marked effects on the carotenoid production and distribution (**Figure 40**, A, B, Additional file 1: Figure S5). For GA08 and GA09, the total carotenoid yields were very low, below 10 mg/L (< 2000 ppm), and the carotenoid glucosides were also very low, below 0.4 mg/L (< 100 ppm). GA01 and GA02 had the highest glycosylation efficiency (~ 21%, **Figure 40**, A), but with relatively lower total carotenoid yields as compared to GA03, GA04 and GA05. Surprisingly, GA03, with a relatively weaker RBS (Additional file 1: Table S1), had the highest yield of total carotenoids (11,623 ppm) and total glycosylated carotenoids (1774 ppm). Similar to GA01, strains GA02–07 had lower yields of carotenoids (including glycosylated carotenoids) when IPTG concentrations increased from 0.03 to 0.1 mM. In contrast, strains GA08–09 had higher yields when IPTG dosage increased, likely due to the relatively weaker RBSs of crtZ. RBS engineering of crtZ has enhanced the production of glycosylated and total carotenoids by 25% and 72%, respectively, as

compared to that of GA01. However, unlike the obvious positive effect of RBS of crtX on zeaxanthin glucosides (**Figure 37**, D), the data in Additional file 1: Figure S5 indicated the lack of correlation between the RBS strength of crtZ and carotenoid production. The lack of correlation was not surprising as the top two producers, GA03 and GA05, had relatively weaker RBSs.

Discussion

Here, we successfully engineered *E. coli* to produce carotenoid glucosides in high amounts. Particularly, our zeaxanthin glucoside strain produced 11,670 ppm of two zeaxanthin glucosides (~ 7150 ppm of zeaxanthin diglucose, ~ 4520 ppm of zeaxanthin glucoside) in 2-day batch fermentation (**Figure 38**, B). In contrast, the astaxanthin glucoside strains (GA01–09) produced lower amount of total carotenoid glucosides (1774 ppm) but with high diversity where 7 carotenoid glucosides were detected. To the best of our knowledge, our study is the first to produce these carotenoid glucosides (up to 7 varieties) in recombinant microbes.

Glycosylation plays a crucial role in secondary metabolite biosynthesis such as carotenoids and flavonoids. Similar to carotenoids, the glycosylation improves their solubility, stability, and biological activities of flavonoids. The GTs for both flavonoids and carotenoids are GT1 family, for example, flavonol 3-O-glucosyltransferase (EC 2.4.1.91); anthocyanidin 3-O-glucosyltransferase (EC 2.4.1.115); ZGT (EC 2.4.1.-). The GT1 family comprises a highly divergent, polyphyletic genes/enzymes, with GTs identified from animals, plants, fungi, bacteria, and viruses. Flavonoid GTs are relatively well studied and characterized. To date, 35 flavonoid GTs are reviewed in UniProt database, and most of them show broad activities to a large range of structurally similar flavonoids and sugar donors, e.g., Anthocyanidin 3-O-glucosyltransferase UFGT from Vitis vinifera (UniProt ID, P51094) can accept cyanidin, delphinidin, kaempferol, malvidin, quercitin, etc. as substrates and use UDP-glucose, UDP-galactose, guanosine 5'-diphosphoglucose (GDP-glucose), dTDP-glucose, etc. as sugar donors. In contrast, none of the carotenoid GTs has been well studied. Our results here supported that ZGT was able to glycosylate various other carotenoids (e.g., adonirubin, adonixanthin), in addition to the reported zeaxanthin and astaxanthin (Hundle *et al.*, 1992; Yokoyama, Shizuri and Misawa, 1998). Furthermore, if xanthophylls have two hydroxyl groups (e.g., astaxanthin), diglycosylated products can also be produced by ZGT. Considering the complexity of the carotenoid pathway and the promiscuity of ZGT, the product diversity was not surprising as the glycosylation reaction competed with other reactions (hydroxylation or ketolation, **Figure 35**). The presence of bulky glycoside moiety may prevent the glycosylated intermediates (e.g., zeaxanthin and adonixanthin) from further ketolation to astaxanthin glucosides by the β-carotene ketolase (crtW); hence, all the carotenoid glucosides became the end products (**Figure 35**).

This study was largely built on our previous astaxanthin platform. By removing the crtW gene, we obtained a high-yield zeaxanthin strain, and furthermore, the high-yield production of zeaxanthin and astaxanthin glucosides. The success indicated that our carotenoid platform is highly expandable for the production of various carotenoids and serves a good starting point for further optimization.

Yet even with such a good platform, it is still not trivial to further tune the pathways for the production of carotenoid glucosides, much efforts are still required to enhance the yields toward industrial viability. To improve the glycosylation of zeaxanthin, we have employed RBS engineering (strong RBS for ZGT), media optimization and supplementation of additional lycopene cyclase (crtY). All the strategies were very effective, collectively, they enhanced the yields of the two zeaxanthin glucosides from 1640 ppm to 11,670 ppm, or by 7.1 fold.

However, it was not straightforward for astaxanthin glycosylation. A possible reason is that the ZGT from *P. ananatis* might have relatively lower activity for astaxanthin than zeaxanthin. The keto group may also stabilize the hydroxyl group or introduces steric hindrance and, thus, reduces accessibility by ZGT. Also, the competitions for carotenoid intermediates by ketolases (CrtW), hydroxylases (CrtZ) and ZGT increase the ramification of the metabolic pathway. To further improve the production of astaxanthin glucosides, four strategies can be employed in the future: (1) to explore the natural diversity of ZGTs for more suitable enzymes; (2) to balance the expression of Module 4 (**Figure 35**); (3) to further manipulate the intracellular UDP-glucose supply; and 4) to implement a dynamic regulation to trigger glycosylation after the formation of astaxanthin. A search in UniProt database resulted in 254 zeaxanthin GT homologues from 69 microbial genera, particularly in *Pseudomonas*, *Pantoea* and *Massilia*, which have 88, 22, 12 of homologues identified, respectively. Experimental screening may lead to identifying some candidates with higher activities and/or specificities for astaxanthin. Furthermore, the data in **Figure 37** and **Figure 40** indicated that the perturbation of crtZ and crtX expression had strong effects on both yields of total carotenoids and glycosylated carotenoids. The parental strain (Ast) had produced astaxanthin as the main product; however, all the GA01–09 strains had β-carotene accumulated intracellularly (**Figure 40**, B). This indicated that previously balanced pathway was perturbed by the introduction of ZGT. A solution is to refine the module 4 by RBS/promoter engineering or organization shuffling of operon genes to minimizing the accumulation of intermediates (e.g., lycopene and β-carotene, **Figure 40**, B). Lastly, unlike zeaxanthin glycosylation strain with high glycosylation efficiency (> 90%), the astaxanthin glycosylation was relatively low (40–50%) indicating they might be still limited by the accessible intracellular UDP-glucose, whose supply can be enhanced by overexpressing UDP-glucose biosynthetic pathway genes (e.g., glk: glucokinase, pgm: phosphoglucomutase, galU: UDPglucose pyrophosphorylase) and by utilizing other types of UDP-sugars with glycosyltransferases. The strategy has been successfully applied to increase the production of flavonoids such as anthocyanins (Shrestha *et al.*, 2019; Zha, Wu and Koffas, 2020) and is worth exploring on carotenoid glycosylation.

Conclusion

We have developed microbial strains to overproduce various carotenoid glucosides. The metabolic engineering and bioprocess strategies are proven to be effective and have synergic effects in improving the yields of carotenoid glucosides by balancing the metabolic pathways and supplying carbon precursors and important cofactors. Our study here demonstrated a proof-of-concept study

for microbial production of glycosylated carotenoids and might inspire the production for other high-value metabolites, especially other glycosylated metabolites.

Acknowledgements

We acknowledge Prof Heng-Phon Too, National University of Singapore, Dr Nicholas David Lindley, Dr Ee Lui Ang and Dr Hazel Khoo, Singapore Institute of Food and Biotechnology Innovation, for the support and insightful suggestions of the project.

Author contributions

CZ conceived the project, analyzed the results, and wrote the manuscript. XC, XL, and CZ designed and did the experiments of strain engineering. TL and AB designed and did the microscope imaging of *E. coli* and quantification of carotenoid distributed in membrane and cytosol.

All authors have read and approved the final manuscript.

CONCLUSION AND PERSPECTIVES

Starting from a strain optimized for the production of lycopene by combinatorial modules, the introduction of CrtY and CrtZ enzymes led to the incomplete conversion of β-carotene into zeaxanthin. We thus implemented a protein fusion strategy to try to increase the conversion of β-carotene into zeaxanthin. A linker library was designed to systematically asses the role of physicochemical parameters of linkers in the CrtY/CrtZ protein. On top of the common parameters such as the length and flexibility of the linker, a new criterion was found and should be taken into consideration when designing fusion protein. Indeed, the use of amino acid at the extremities of the linker were shown to be more polarized than in the rest of the linker, probably due higher constraint due do the direct proximity of the enzymes.

The understanding of the linker parameters impacting the protein fusion could be applied to other enzymes. Although the theory of the importance of amino acid as the extremities of the linker was confirmed with the 1280 linkers of the database, it would have been interesting to test it on other enzymes. Indeed, one of the limits of the database is that all membrane proteins have been removed from the set, making extrapolation of our conclusion to membrane protein quite difficult. It is interesting to note that the first amino acid parameter applied to the CrtZ-CrtY protein as well as the dataset. It could now be interesting to see if different amino acids are favored in linker of cytosolic protein fusion and membrane protein fusion.

The fusion of CrtY and CrtZ enzymes allowed to obtain improved strain for the production of zeaxanthin. The SHS01 strain displayed a 9-fold increase in zeaxanthin specificity while the SNI01 strain had a 1.46-fold increase in the total production of carotenoids.

The increase in zeaxanthin production could be due to the relocalization of CrtY to the membrane when fused to CrtZ, as it is predicted to be a membrane bound protein. This hypothesis could be verified by fusing CrtY to a membrane anchor such as MBP and see if the improvement remains. The location of CrtY enzyme could also be verified by fusion with a fluorescent protein. An improvement in zeaxanthin production due to relocation of CrtY to the membrane agrees with previous studies that demonstrated that enzyme proximity does not enhance the reaction rate of enzymes but rather provides a situational advantage such as a possible change in local microenvironment, protection of intermediates from competing pathway, protection of enzyme from O_2, or its location in the cell as hypothesized in this work.

SDS-PAGE results showed that CrtZ protein is less expressed than CrtY in *E. coli* cells. The fusion of CrtZ to CrtY protein could enhance the production of CrtZ in the strain with enzyme fusion and thus explained changes in the profiles of carotenoids accumulation, notably the increase in CrtZ specificity. Any changes in the production of carotenoids and more specifically in the rate of the enzymes could

be better understood with an enzymatic study of the proteins. It would indeed be interesting to compare the reaction rate of the independent enzymes to the enzymes fused together with different linkers, notably the fusion with SHS01, SNI01 and SNS06 linkers which yielded completely different carotenoids profiles in *E. coli*.

One limit of this work is the lack of understanding of the protein structures. Indeed, no correlation were found between the structure of the protein and the efficiency of carotenoid production. We also lack structural information, for example, although the structure of both enzymes was retrieved from Alphafold, the active site of the CrtY enzyme is still unknow rendering quite difficult the interpretation of protein fusion models developed in this study.

It could also be interesting to verify if the carotenoids profile accumulated in the *E. coli* strain with different protein fusion would remain similar in another organism such as *S. cerevisiae*. Although we expect the carotenoid yield to be impacted, since the ratio of CrtY and CrtZ enzyme would remain a 1:1 ratio, we would expect a similar hierarchy in the carotenoid accumulation than the one observed in *E. coli*.

Protein fusion strategy is limited by the number of protein domain able to be folded together. Protein fusion with more than three catalytic domains have not been reported. Moreover, the protein fusion will always be limited by the slowest enzyme of the fusion, and enzyme proximity does not enhance the reaction rates of the enzymes. These limits of protein fusion should encourage the use of more complex enzyme assembly technology such as the use of protein scaffold which have recently been reported for the production of lutein or the rare capsanthin carotenoids, or peptide-ligand interaction.

REFERENCES

Agapakis, C.M. *et al.* (2010) 'Insulation of a synthetic hydrogen metabolism circuit in bacteria', *Journal of Biological Engineering*, 4(1), p. 3. Available at: https://doi.org/10.1186/1754-1611-4-3.

Albertsen, L. *et al.* (2011) 'Diversion of Flux toward Sesquiterpene Production in Saccharomyces cerevisiae by Fusion of Host and Heterologous Enzymes', *Applied and Environmental Microbiology*, 77(3), p. 1033. Available at: https://doi.org/10.1128/AEM.01361-10.

Al-Husini, N. *et al.* (2020) 'BR-bodies provide selectively permeable condensates that stimulate mRNA decay and prevent release of decay intermediates', *Molecular cell*, 78(4), pp. 670-682.e8. Available at: https://doi.org/10.1016/j.molcel.2020.04.001.

Anderson, K.S., Miles, E.W. and Johnson, K.A. (1991) 'Serine modulates substrate channeling in tryptophan synthase. A novel intersubunit triggering mechanism', *Journal of Biological Chemistry*, 266(13), pp. 8020–8033. Available at: https://doi.org/10.1016/S0021-9258(18)92934-0.

Argos, P. (1990) 'An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion', *Journal of Molecular Biology*, 211(4), pp. 943–958. Available at: https://doi.org/10.1016/0022-2836(90)90085-Z.

Arunkumar, R., Gorusupudi, A. and Bernstein, P.S. (2020) 'The macular carotenoids: A biochemical overview', *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1865(11), p. 158617. Available at: https://doi.org/10.1016/j.bbalip.2020.158617.

Avalos, J.L., Fink, G.R. and Stephanopoulos, G. (2013) 'Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols', *Nature Biotechnology*, 31(4), pp. 335–341. Available at: https://doi.org/10.1038/nbt.2509.

Azaldegui, C.A., Vecchiarelli, A.G. and Biteen, J.S. (2021) 'The emergence of phase separation as an organizing principle in bacteria', *Biophysical Journal*, 120(7), pp. 1123–1138. Available at: https://doi.org/10.1016/j.bpj.2020.09.023.

Azuma, Y. *et al.* (2016) 'Quantitative Packaging of Active Enzymes into a Protein Cage', *Angewandte Chemie International Edition*, 55(4), pp. 1531–1534. Available at: https://doi.org/10.1002/anie.201508414.

Baek, J.-M. *et al.* (2013) 'Butyrate production in engineered *Escherichia coli* with synthetic scaffolds: Butyrate Production in Engineered *Escherichia coli*', *Biotechnology and Bioengineering*, 110(10), pp. 2790–2794. Available at: https://doi.org/10.1002/bit.24925.

Bakkes, P.J. *et al.* (2015) 'Design and improvement of artificial redox modules by molecular fusion of flavodoxin and flavodoxin reductase from Escherichia coli', *Scientific Reports*, 5(1), p. 12158. Available at: https://doi.org/10.1038/srep12158.

Baklouti, Z. *et al.* (2020) 'Biochemical Characterization of a Bifunctional Enzyme Constructed by the Fusion of a Glucuronan Lyase and a Chitinase from Trichoderma sp.', *Life*, 10(10), p. 234. Available at: https://doi.org/10.3390/life10100234.

Banani, S.F. *et al.* (2017) 'Biomolecular condensates: Organizers of cellular biochemistry', *Nature reviews. Molecular cell biology*, 18(5), pp. 285–298. Available at: https://doi.org/10.1038/nrm.2017.7.

Barreiro, C. and Barredo, J.-L. (2018) 'Carotenoids Production: A Healthy and Profitable Industry', in C. Barreiro and J.-L. Barredo (eds) *Microbial Carotenoids*. New York, NY: Springer New York (Methods in Molecular Biology), pp. 45–55. Available at: https://doi.org/10.1007/978-1-4939-8742-9_2.

Bashton, M. and Chothia, C. (2007) 'The Generation of New Protein Functions by the Combination of Domains', *Structure*, 15(1), pp. 85–99. Available at: https://doi.org/10.1016/j.str.2006.11.009.

Béguin, P. and Lemaire, M. (1996) 'The Cellulosome: An Exocellular, Multiprotein Complex Specialized in Cellulose Degradation', *Critical Reviews in Biochemistry and Molecular Biology*, 31(3), pp. 201–236. Available at: https://doi.org/10.3109/10409239609106584.

Behrendorff, J.B.Y.H. *et al.* (2019) 'Membrane-Bound Protein Scaffolding in Diverse Hosts Using Thylakoid Protein CURT1A', *ACS Synthetic Biology*, 8(4), pp. 611–620. Available at: https://doi.org/10.1021/acssynbio.8b00418.

Berckman, E.A. and Chen, W. (2021) 'Self-assembling protein nanocages for modular enzyme assembly by orthogonal bioconjugation', *Biotechnology Progress*, 37(5), p. e3190. Available at: https://doi.org/10.1002/btpr.3190.

Bhataya, A., Schmidt-Dannert, C. and Lee, P.C. (2009) 'Metabolic engineering of Pichia pastoris X-33 for lycopene production', *Process Biochemistry*, 44(10), pp. 1095–1102. Available at: https://doi.org/10.1016/j.procbio.2009.05.012.

Bouvier, F. and Keller, Y. (1998) 'Xanthophyll biosynthesis: molecular and functional characterization of carotenoid hydroxylases from pepper fruits žCapsicum annuum L./', p. 9.

Breitenbach, J., Pollmann, H. and Sandmann, G. (2019) 'Genetic modification of the carotenoid pathway in the red yeast Xanthophyllomyces dendrorhous: Engineering of a high-yield zeaxanthin strain', *Journal of Biotechnology*, 289, pp. 112–117. Available at: https://doi.org/10.1016/j.jbiotec.2018.11.019.

Bu, X. *et al.* (2020) 'Engineering endogenous ABC transporter with improving ATP supply and membrane flexibility enhances the secretion of β-carotene in Saccharomyces cerevisiae', *Biotechnology for Biofuels*, 13(1), p. 168. Available at: https://doi.org/10.1186/s13068-020-01809-6.

Bu, X. *et al.* (2022) 'Dual regulation of lipid droplet-triacylglycerol metabolism and ERG9 expression for improved β-carotene production in Saccharomyces cerevisiae', *Microbial Cell Factories*, 21(1), p. 3. Available at: https://doi.org/10.1186/s12934-021-01723-y.

Bülow, leif, Ljungcrantz, P. and Mosbach, K. (1985) 'Preparation of a bifunctional enzyme by gene fusion', *Nature Biotechnology*, 3.

Buscemi, S. *et al.* (2018) 'The Effect of Lutein on Eye and Extra-Eye Health', *Nutrients*, 10(9), p. 1321. Available at: https://doi.org/10.3390/nu10091321.

Camagna, M. *et al.* (2019) 'Enzyme Fusion Removes Competition for Geranylgeranyl Diphosphate in Carotenogenesis1[OPEN]', *Plant Physiology*, 179(3), pp. 1013–1027. Available at: https://doi.org/10.1104/pp.18.01026.

Carquet, M., Pompon, D. and Truan, G. (2015) 'Transcription Interference and ORF Nature Strongly Affect Promoter Strength in a Reconstituted Metabolic Pathway', *Frontiers in Bioengineering and Biotechnology*, 3. Available at: https://www.frontiersin.org/articles/10.3389/fbioe.2015.00021 (Accessed: 24 January 2023).

Chang, C.-H., Rossi, E.A. and Goldenberg, D.M. (2007) 'The Dock and Lock Method: A Novel Platform Technology for Building Multivalent, Multifunctional Structures of Defined Composition with Retained Bioactivity', *Clinical Cancer Research*, 13(18), pp. 5586s–5591s. Available at: https://doi.org/10.1158/1078-0432.CCR-07-1217.

Cheah, L.C. *et al.* (2021) 'Artificial Self-assembling Nanocompartment for Organizing Metabolic Pathways in Yeast', *ACS Synthetic Biology*, 10(12), pp. 3251–3263. Available at: https://doi.org/10.1021/acssynbio.1c00045.

Chen, X. *et al.* (2021) 'High-level de novo biosynthesis of glycosylated zeaxanthin and astaxanthin in Escherichia coli', *Bioresources and Bioprocessing*, 8(1), p. 67. Available at: https://doi.org/10.1186/s40643-021-00415-0.

Chen, Y. *et al.* (2016) 'Lycopene overproduction in Saccharomyces cerevisiae through combining pathway engineering with host engineering', *Microbial Cell Factories*, 15, p. 113. Available at: https://doi.org/10.1186/s12934-016-0509-4.

Choi, S.-K. *et al.* (2013) '3-β-Glucosyl-3'-β-quinovosyl zeaxanthin, a novel carotenoid glycoside synthesized by Escherichia coli cells expressing the Pantoea ananatis carotenoid biosynthesis gene cluster', *Applied Microbiology and Biotechnology*, 97(19), pp. 8479–8486. Available at: https://doi.org/10.1007/s00253-013-5101-9.

Conrado, R.J. *et al.* (2012) 'DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency', *Nucleic Acids Research*, 40(4), pp. 1879–1889. Available at: https://doi.org/10.1093/nar/gkr888.

Crasto, C.J. and Feng, J.A. (2000) 'LINKER: A program to generate linker sequences for fusion proteins', *Protein Engineering*, 13(5), pp. 309–312. Available at: https://doi.org/10.1093/protein/13.5.309.

Delebecque, C.J. *et al.* (2011) 'Organization of intracellular reactions with rationally designed RNA assemblies', *Science*, 333(6041), pp. 470–474. Available at: https://doi.org/10.1126/science.1206938.

DeLoache, W.C., Russ, Z.N. and Dueber, J.E. (2016) 'Towards repurposing the yeast peroxisome for compartmentalizing heterologous metabolic pathways', *Nature Communications*, 7(1), p. 11152. Available at: https://doi.org/10.1038/ncomms11152.

Dembitsky, V.M. (2005) 'Astonishing diversity of natural surfactants. 3. Carotenoid glycosides and isoprenoid glycolipids', *Lipids* [Preprint]. Available at: https://scholar.google.com/scholar_lookup?title=Astonishing+diversity+of+natural+surfactants.+3.+Carotenoid+glycosides+and+isoprenoid+glycolipids&author=Dembitsky%2C+V.M.&publication_year=2005 (Accessed: 27 April 2023).

Ding, Y.-W. *et al.* (2022) 'Directed evolution of the fusion enzyme for improving astaxanthin biosynthesis in Saccharomyces cerevisiae', *Synthetic and Systems Biotechnology*, 8(1), pp. 46–53. Available at: https://doi.org/10.1016/j.synbio.2022.10.005.

Doshi, R., Nguyen, T. and Chang, G. (2013) 'Transporter-mediated biofuel secretion', *Proceedings of the National Academy of Sciences of the United States of America*, 110(19), pp. 7642–7647. Available at: https://doi.org/10.1073/pnas.1301358110.

Dueber, J.E. *et al.* (2009) 'Synthetic protein scaffolds provide modular control over metabolic flux', *Nature Biotechnology*, 27(8), pp. 753–759. Available at: https://doi.org/10.1038/nbt.1557.

Dung Pham, V. *et al.* (2016) 'Efficient production of gamma-aminobutyric acid using Escherichia coli by co-localization of glutamate synthase, glutamate decarboxylase, and GABA transporter', *Journal of Industrial Microbiology and Biotechnology*, 43(1), pp. 79–86. Available at: https://doi.org/10.1007/s10295-015-1712-8.

Dunn, M.F. *et al.* (2008) 'Tryptophan synthase: the workings of a channeling nanomachine', *Trends in Biochemical Sciences*, 33(6), pp. 254–264. Available at: https://doi.org/10.1016/j.tibs.2008.04.008.

Duran, L., López, J.M. and Avalos, J.L. (2020) '¡Viva la mitochondria!: harnessing yeast mitochondria for chemical production', *FEMS Yeast Research*, 20(6), p. foaa037. Available at: https://doi.org/10.1093/femsyr/foaa037.

Dzuricky, M. *et al.* (2020) 'De novo engineering of intracellular condensates using artificial disordered proteins', *Nature Chemistry*, 12(9), pp. 814–825. Available at: https://doi.org/10.1038/s41557-020-0511-7.

Elleuche, S. (2015) 'Bringing functions together with fusion enzymes—from nature's inventions to biotechnological applications', *Applied Microbiology and Biotechnology*, 99(4), pp. 1545–1556. Available at: https://doi.org/10.1007/s00253-014-6315-1.

Elshahawi, S.I. *et al.* (2015) 'A comprehensive review of glycosylated bacterial natural products', *Chemical Society reviews*, 44(21), pp. 7591–7697. Available at: https://doi.org/10.1039/c4cs00426d.

Enjalbert, T. *et al.* (2020) 'Characterisation of the Effect of the Spatial Organisation of Hemicellulases on the Hydrolysis of Plant Biomass Polymer', *International Journal of Molecular Sciences*, 21(12), p. 4360. Available at: https://doi.org/10.3390/ijms21124360.

Erkelenz, M., Kuo, C.-H. and Niemeyer, C.M. (2011) 'DNA-Mediated Assembly of Cytochrome P450 BM3 Subdomains', *Journal of the American Chemical Society*, 133(40), pp. 16111–16118. Available at: https://doi.org/10.1021/ja204993s.

Farasat, I. *et al.* (2014) 'Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria', *Molecular Systems Biology*, 10(6), p. 731. Available at: https://doi.org/10.15252/msb.20134955.

Farhi, M. *et al.* (2011) 'Harnessing yeast subcellular compartments for the production of plant terpenoids', *Metabolic Engineering*, 13(5), pp. 474–481. Available at: https://doi.org/10.1016/j.ymben.2011.05.001.

Frey, R., Hayashi, T. and Hilvert, D. (2016) 'Enzyme-mediated polymerization inside engineered protein cages', *Chemical Communications*, 52(68), pp. 10423–10426. Available at: https://doi.org/10.1039/C6CC05301G.

Fu, J. *et al.* (2012) 'Inter-enzyme substrate diffusion for an enzyme cascade organized on spatially addressable DNA nanostructures', *Journal of the American Chemical Society*, 134(12), pp. 5516–5519. Available at: https://doi.org/10.1021/ja300897h.

Fu, J. *et al.* (2020) 'DNA-Scaffolded Proximity Assembly and Confinement of Multienzyme Reactions', *Topics in Current Chemistry (Cham)*, 378(3), p. 38. Available at: https://doi.org/10.1007/s41061-020-0299-3.

Furubayashi, M., Maoka, T. and Mitani, Y. (2022) 'Promiscuous activity of β-carotene hydroxylase CrtZ on epoxycarotenoids leads to the formation of rare carotenoids with 6-hydroxy-3-keto-ε-ends', *FEBS Letters*, pp. 1873-3468.14342. Available at: https://doi.org/10.1002/1873-3468.14342.

Gabashvili, A.N. *et al.* (2020) 'Encapsulins—Bacterial Protein Nanocompartments: Structure, Properties, and Application', *Biomolecules*, 10(6), p. 966. Available at: https://doi.org/10.3390/biom10060966.

Gao, Z. *et al.* (2021) 'Liquid-Liquid Phase Separation: Unraveling the Enigma of Biomolecular Condensates in Microbial Cells', *Frontiers in Microbiology*, 12, p. 751880. Available at: https://doi.org/10.3389/fmicb.2021.751880.

George, R.A. and Heringa, J. (2002) 'An analysis of protein domain linkers: Their classification and role in protein folding', *Protein Engineering*, 15(11), pp. 871–879. Available at: https://doi.org/10.1093/protein/15.11.871.

Geraldi, A. *et al.* (2021) 'Synthetic Scaffold Systems for Increasing the Efficiency of Metabolic Pathways in Microorganisms', *Biology*, 10(3), p. 216. Available at: https://doi.org/10.3390/biology10030216.

Giessen, T.W. (2022) 'Encapsulins', *Annual review of biochemistry*, 91, pp. 353–380. Available at: https://doi.org/10.1146/annurev-biochem-040320-102858.

Giessen, T.W. and Silver, P.A. (2016) 'A Catalytic Nanoreactor Based on in Vivo Encapsulation of Multiple Enzymes in an Engineered Protein Nanocompartment', *ChemBioChem*, 17(20), pp. 1931–1935. Available at: https://doi.org/10.1002/cbic.201600431.

Gilbert, C. *et al.* (2017) 'Extracellular Self-Assembly of Functional and Tunable Protein Conjugates from *Bacillus subtilis*', *ACS Synthetic Biology*, 6(6), pp. 957–967. Available at: https://doi.org/10.1021/acssynbio.6b00292.

Gräwe, A. *et al.* (2020) 'iFLinkC: an iterative functional linker cloning strategy for the combinatorial assembly and recombination of linker peptides with functional domains', *Nucleic Acids Research*, 48(4), pp. e24–e24. Available at: https://doi.org/10.1093/nar/gkz1210.

Grininger, M. (2023) 'Enzymology of assembly line synthesis by modular polyketide synthases', *Nature Chemical Biology*, pp. 1–15. Available at: https://doi.org/10.1038/s41589-023-01277-7.

Guo, H. *et al.* (2017) 'Effect of flexible linker length on the activity of fusion protein 4-coumaroyl-CoA ligase::stilbene synthase', *Molecular BioSystems*, 13(3), pp. 598–606. Available at: https://doi.org/10.1039/c6mb00563b.

Háda, M. *et al.* (2012) 'Hydrophilic Carotenoids: Recent Progress', *Molecules*, 17(5), pp. 5003–5012. Available at: https://doi.org/10.3390/molecules17055003.

Haga, T., Hirakawa, H. and Nagamune, T. (2013) 'Fine Tuning of Spatial Arrangement of Enzymes in a PCNA-Mediated Multienzyme Complex Using a Rigid Poly-L-Proline Linker', *PLoS ONE*, 8(9), p. e75114. Available at: https://doi.org/10.1371/journal.pone.0075114.

Hammer, S.K. and Avalos, J.L. (2017) 'Harnessing yeast organelles for metabolic engineering', *Nature Chemical Biology*, 13(8), pp. 823–832. Available at: https://doi.org/10.1038/nchembio.2429.

Harold, F.M. (2005) 'Molecules into Cells: Specifying Spatial Architecture', *Microbiology and Molecular Biology Reviews*, 69(4), pp. 544–564. Available at: https://doi.org/10.1128/MMBR.69.4.544-564.2005.

Hashimoto, H., Uragami, C. and Cogdell, R.J. (2016) 'Carotenoids and Photosynthesis', *Sub-cellular biochemistry*, 79, pp. 111–139. Available at: https://doi.org/10.1007/978-3-319-39126-7_4.

Hattan, J. *et al.* (2023) 'Reconstruction of the Native Biosynthetic System of Carotenoids in *E. coli* –Biosynthesis of a Series of Carotenoids Specific to Paprika Fruit', *ACS Synthetic Biology*, 12(4), pp. 1072–1080. Available at: https://doi.org/10.1021/acssynbio.2c00578.

Henke, N.A. and Wendisch, V.F. (2019) 'Improved astaxanthin production with corynebacterium glutamicum by application of a membrane fusion protein', *Marine Drugs*, 17(11). Available at: https://doi.org/10.3390/md17110621.

Henriques de Jesus, M.P.R. *et al.* (2017) 'Tat proteins as novel thylakoid membrane anchors organize a biosynthetic pathway in chloroplasts and increase product yield 5-fold', *Metabolic Engineering*, 44, pp. 108–116. Available at: https://doi.org/10.1016/j.ymben.2017.09.014.

Hirakawa, H. and Nagamune, T. (2010) 'Molecular Assembly of P450 with Ferredoxin and Ferredoxin Reductase by Fusion to PCNA', *ChemBioChem*, 11(11), pp. 1517–1520. Available at: https://doi.org/10.1002/cbic.201000226.

Hirano, K. *et al.* (2016) 'Enzymatic diversity of the Clostridium thermocellum cellulosome is crucial for the degradation of crystalline cellulose and plant biomass', *Scientific Reports*, 6(1), p. 35709. Available at: https://doi.org/10.1038/srep35709.

Horn, S.J. *et al.* (2012) 'Novel enzymes for the degradation of cellulose', *Biotechnology for Biofuels*, 5(1), p. 45. Available at: https://doi.org/10.1186/1754-6834-5-45.

Huang, Z. *et al.* (2022) 'Rapid regulations of metabolic reactions in Escherichia coli via light-responsive enzyme redistribution', *Biotechnology Journal*, 17(9), p. 2200129. Available at: https://doi.org/10.1002/biot.202200129.

Huffine, C.A. *et al.* (2023) 'Role of carboxysomes in cyanobacterial CO_2 assimilation: CO_2 concentrating mechanisms and metabolon implications', *Environmental Microbiology*, 25(2), pp. 219–228. Available at: https://doi.org/10.1111/1462-2920.16283.

Hundle, B.S. *et al.* (1992) 'Functional expression of zeaxanthin glucosyltransferase from Erwinia herbicola and a proposed uridine diphosphate binding site', *Proceedings of the National Academy of Sciences of the United States of America*, 89(19), pp. 9321–9325. Available at: https://doi.org/10.1073/pnas.89.19.9321.

Iwata, F., Hirakawa, H. and Nagamune, T. (2018) 'A Stable Artificial Multienzymatic Complex Using a Heterotrimeric Protein From Metallosphaera sedula', *Biotechnology Journal*, 13(11), p. 1700662. Available at: https://doi.org/10.1002/biot.201700662.

Jandt, U. *et al.* (2013) 'Compartmentalization and Metabolic Channeling for Multienzymatic Biosynthesis: Practical Strategies and Modeling Approaches', in An-Ping Zeng (ed.) *Fundamentals and Application of New Bioproduction Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg (Advances in Biochemical Engineering/Biotechnology), pp. 41–65. Available at: https://doi.org/10.1007/10_2013_221.

Jeckelmann, J.-M. and Erni, B. (2019) 'Carbohydrate Transport by Group Translocation: The Bacterial Phosphoenolpyruvate: Sugar Phosphotransferase System', in A. Kuhn (ed.) *Bacterial Cell Walls and Membranes*. Cham: Springer International Publishing (Subcellular Biochemistry), pp. 223–274. Available at: https://doi.org/10.1007/978-3-030-18768-2_8.

Jia, L. *et al.* (2017) 'Polymeric SpyCatcher Scaffold Enables Bioconjugation in a Ratio-Controllable Manner', *Biotechnology Journal*, 12(12), p. 1700195. Available at: https://doi.org/10.1002/biot.201700195.

Kang, W. *et al.* (2019) 'Modular enzyme assembly for enhanced cascade biocatalysis and metabolic flux', *Nature Communications*, 10(1). Available at: https://doi.org/10.1038/s41467-019-12247-w.

Katz, L. (2009) 'Chapter 6 The DEBS Paradigm for Type I Modular Polyketide Synthases and Beyond', in *Methods in Enzymology*. Elsevier, pp. 113–142. Available at: https://doi.org/10.1016/S0076-6879(09)04606-0.

Kerfeld, C.A. *et al.* (2018) 'Bacterial microcompartments', *Nature reviews. Microbiology*, 16(5), pp. 277–290. Available at: https://doi.org/10.1038/nrmicro.2018.10.

Khosla, C. *et al.* (2014) 'Assembly Line Polyketide Synthases: Mechanistic Insights and Unsolved Problems', *Biochemistry*, 53(18), pp. 2875–2883. Available at: https://doi.org/10.1021/bi500290t.

Kim, J.-E. *et al.* (2019) 'Tailoring the Saccharomyces cerevisiae endoplasmic reticulum for functional assembly of terpene synthesis pathway', *Metabolic Engineering*, 56, pp. 50–59. Available at: https://doi.org/10.1016/j.ymben.2019.08.013.

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', *Journal of Biotechnology*, 192, pp. 192–196. Available at: https://doi.org/10.1016/j.jbiotec.2014.10.015.

Kirst, H. *et al.* (2022) 'Toward a glycyl radical enzyme containing synthetic bacterial microcompartment to produce pyruvate from formate and acetate', *Proceedings of the National Academy of Sciences of the United States of America*, 119(8), p. e2116871119. Available at: https://doi.org/10.1073/pnas.2116871119.

Kirst, H. and Kerfeld, C.A. (2019) 'Bacterial microcompartments: catalysis-enhancing metabolic modules for next generation metabolic and biomedical engineering', *BMC Biology*, 17(1), p. 79. Available at: https://doi.org/10.1186/s12915-019-0691-z.

Krubasik, P. *et al.* (2001) 'Detailed biosynthetic pathway to decaprenoxanthin diglucoside in Corynebacterium glutamicum and identification of novel intermediates', *Archives of Microbiology*, 176(3), pp. 217–223. Available at: https://doi.org/10.1007/s002030100315.

Krubasik, P. and Sandmann, G. (2000) 'Molecular evolution of lycopene cyclases involved in the formation of carotenoids with ionone end groups', *Biochemical Society Transactions*, 28.

Kulagina, N. *et al.* (2021) 'Peroxisomes: A New Hub for Metabolic Engineering in Yeast', *Frontiers in Bioengineering and Biotechnology*, 9, p. 659431. Available at: https://doi.org/10.3389/fbioe.2021.659431.

Kummer, M.J. *et al.* (2021) 'Substrate Channeling by a Rationally Designed Fusion Protein in a Biocatalytic Cascade', *JACS Au* [Preprint]. Available at: https://doi.org/10.1021/jacsau.1c00180.

Kuzmak, A. *et al.* (2019) 'Can enzyme proximity accelerate cascade reactions?', *Scientific Reports*, 9. Available at: https://doi.org/10.1038/s41598-018-37034-3.

Ladenstein, R. and Morgunova, E. (2020) 'Second career of a biosynthetic enzyme: Lumazine synthase as a virus-like nanoparticle in vaccine development', *Biotechnology Reports*, 27, p. e00494. Available at: https://doi.org/10.1016/j.btre.2020.e00494.

Ladouceur, A.-M. *et al.* (2020) 'Clusters of bacterial RNA polymerase are biomolecular condensates that assemble through liquid–liquid phase separation', *Proceedings of the National Academy of Sciences*, 117(31), pp. 18540–18549. Available at: https://doi.org/10.1073/pnas.2005019117.

Lakhundi, S., Siddiqui, R. and Khan, N.A. (2015) 'Cellulose degradation: a therapeutic strategy in the improved treatment of Acanthamoeba infections', *Parasites & Vectors*, 8(1), p. 23. Available at: https://doi.org/10.1186/s13071-015-0642-7.

Larroude, M. *et al.* (2018) 'A synthetic biology approach to transform Yarrowia lipolytica into a competitive biotechnological producer of β-carotene', *Biotechnology and Bioengineering*, 115(2), pp. 464–472. Available at: https://doi.org/10.1002/bit.26473.

Lasker, K. *et al.* (2022) 'The material properties of a bacterial-derived biomolecular condensate tune biological function in natural and synthetic systems', *Nature Communications*, 13(1), p. 5643. Available at: https://doi.org/10.1038/s41467-022-33221-z.

Lau, Y.H. *et al.* (2018) 'Prokaryotic nanocompartments form synthetic organelles in a eukaryote', *Nature Communications*, 9(1), p. 1311. Available at: https://doi.org/10.1038/s41467-018-03768-x.

Laursen, T. *et al.* (2016) 'Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum', *Science*, 354(6314), pp. 890–893. Available at: https://doi.org/10.1126/science.aag2347.

Lee, J.J.L. *et al.* (2016) 'Engineering Rhodosporidium toruloides with a membrane transporter facilitates production and separation of carotenoids and lipids in a bi-phasic culture', *Applied Microbiology and Biotechnology*, 100(2), pp. 869–877. Available at: https://doi.org/10.1007/s00253-015-7102-3.

Lee, M.J. *et al.* (2018) 'Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm', *Nature Chemical Biology*, 14(2), pp. 142–147. Available at: https://doi.org/10.1038/nchembio.2535.

Lee, M.J., Palmer, D.J. and Warren, M.J. (2019) 'Biotechnological Advances in Bacterial Microcompartment Technology', *Trends in Biotechnology*, 37(3), pp. 325–336. Available at: https://doi.org/10.1016/j.tibtech.2018.08.006.

Li, G. *et al.* (2016) 'Construction of a linker library with widely controllable flexibility for fusion protein design', *Applied Microbiology and Biotechnology*, 100(1), pp. 215–225. Available at: https://doi.org/10.1007/s00253-015-6985-3.

Li, M. *et al.* (2022) 'Spatiotemporal Regulation of Astaxanthin Synthesis in S. cerevisiae', *ACS Synthetic Biology*, 11(8), pp. 2636–2649. Available at: https://doi.org/10.1021/acssynbio.2c00044.

Li, T. *et al.* (2020) 'Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production', *Nature Communications*, 11(1), p. 5448. Available at: https://doi.org/10.1038/s41467-020-19280-0.

Li, X.R. *et al.* (2015) 'Metabolic engineering of Escherichia coli to produce zeaxanthin', *Journal of Industrial Microbiology and Biotechnology*, 42(4), pp. 627–636. Available at: https://doi.org/10.1007/s10295-014-1565-6.

Liang, M. *et al.* (2017) 'Bacterial microcompartment-directed polyphosphate kinase promotes stable polyphosphate accumulation in E. coli', *Biotechnology Journal*, 12(3), p. 1600415. Available at: https://doi.org/10.1002/biot.201600415.

Lin, J.-L., Zhu, J. and Wheeldon, I. (2017) 'Synthetic Protein Scaffolds for Biosynthetic Pathway Colocalization on Lipid Droplet Membranes', *ACS Synthetic Biology*, 6(8), pp. 1534–1544. Available at: https://doi.org/10.1021/acssynbio.7b00041.

Liu, C., Chin, J.X. and Lee, D.-Y. (2015) 'SynLinker: an integrated system for designing linkers and synthetic fusion proteins', *Bioinformatics*, 31(22), pp. 3700–3702. Available at: https://doi.org/10.1093/bioinformatics/btv447.

Liu, G.-S. *et al.* (2020) 'The yeast peroxisome: A dynamic storage depot and subcellular factory for squalene overproduction', *Metabolic Engineering*, 57, pp. 151–161. Available at: https://doi.org/10.1016/j.ymben.2019.11.001.

Liu, M. *et al.* (2016) 'A Three-Enzyme Pathway with an Optimised Geometric Arrangement to Facilitate Substrate Transfer', *ChemBioChem*, 17(12), pp. 1097–1101. Available at: https://doi.org/10.1002/cbic.201600103.

Lv, X. *et al.* (2016) 'Dual regulation of cytoplasmic and mitochondrial acetyl-CoA utilization for improved isoprene production in Saccharomyces cerevisiae', *Nature Communications*, 7(1), p. 12851. Available at: https://doi.org/10.1038/ncomms12851.

Lv, X. *et al.* (2020) 'Synthetic metabolic channel by functional membrane microdomains for compartmentalized flux control', *Metabolic Engineering*, 59, pp. 106–118. Available at: https://doi.org/10.1016/j.ymben.2020.02.003.

Ma, Y. *et al.* (2021) 'Targeting pathway expression to subcellular organelles improves astaxanthin synthesis in Yarrowia lipolytica', *Metabolic Engineering*, 68, pp. 152–161. Available at: https://doi.org/10.1016/j.ymben.2021.10.004.

Mackenzie, J.S. and Jeggo, M. (2019) 'The One Health Approach—Why Is It So Important?', *Tropical Medicine and Infectious Disease*, 4(2), p. 88. Available at: https://doi.org/10.3390/tropicalmed4020088.

Madeira, F. *et al.* (2019) 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic Acids Research*, 47(W1), pp. W636–W641. Available at: https://doi.org/10.1093/nar/gkz268.

Mao, Z., Shin, H.-D. and Chen, R.R. (2006) 'Engineering the E. coli UDP-Glucose Synthesis Pathway for Oligosaccharide Synthesis', *Biotechnology Progress*, 22(2), pp. 369–374. Available at: https://doi.org/10.1021/bp0503181.

Mares, J. (2016) 'Lutein and Zeaxanthin Isomers in Eye Health and Disease', *Annual review of nutrition*, 36, pp. 571–602. Available at: https://doi.org/10.1146/annurev-nutr-071715-051110.

Martín, J.F., Gudiña, E. and Barredo, J.L. (2008) 'Conversion of β-carotene into astaxanthin: Two separate enzymes or a bifunctional hydroxylase-ketolase protein?', *Microbial Cell Factories*, 7, pp. 1–10. Available at: https://doi.org/10.1186/1475-2859-7-3.

Martínez-Cámara, S. *et al.* (2021) 'Main Carotenoids Produced by Microorganisms', *Encyclopedia*, 1(4), pp. 1223–1245. Available at: https://doi.org/10.3390/encyclopedia1040093.

Matsushita, Y. *et al.* (2000) 'Antioxidant activity of polar carotenoids including astaxanthin-β-glucoside from marine bacterium on PC liposomes', *Fisheries science*, 66(5), pp. 980–985. Available at: https://doi.org/10.1046/j.1444-2906.2000.00155.x.

Metzger, V.T. *et al.* (2014) 'Electrostatic Channeling in P. falciparum DHFR-TS: Brownian Dynamics and Smoluchowski Modeling', *Biophysical Journal*, 107(10), pp. 2394–2402. Available at: https://doi.org/10.1016/j.bpj.2014.09.039.

Meynial Salles, I. *et al.* (2007) 'Evolution of a Saccharomyces cerevisiae metabolic pathway in Escherichia coli', *Metabolic Engineering*, 9(2), pp. 152–159. Available at: https://doi.org/10.1016/j.ymben.2006.09.002.

Misawa, N. *et al.* (1990) 'Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functional analysis of gene products expressed in Escherichia coli.', *Journal of Bacteriology*, 172(12), pp. 6704–6712.

Mitra, K. *et al.* (2004) 'Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol', *Proceedings of the National Academy of Sciences of the United States of America*, 101(12), pp. 4083–4088. Available at: https://doi.org/10.1073/pnas.0307332101.

Mohamed, H.E. *et al.* (2005) 'Myxoxanthophyll is required for normal cell wall structure and thylakoid organization in the cyanobacterium Synechocystis sp. strain PCC 6803', *Journal of Bacteriology*, 187(20), pp. 6883–6892. Available at: https://doi.org/10.1128/JB.187.20.6883-6892.2005.

Møller, B.L. and Conn, E.E. (1979) 'The biosynthesis of cyanogenic glucosides in higher plants. N-Hydroxytyrosine as an intermediate in the biosynthesis of dhurrin by Sorghum bicolor (Linn) Moench.', *Journal of Biological Chemistry*, 254(17), pp. 8575–8583. Available at: https://doi.org/10.1016/S0021-9258(19)86931-4.

Moraes, T.F. and Reithmeier, R.A.F. (2012) 'Membrane transport metabolons', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1818(11), pp. 2687–2706. Available at: https://doi.org/10.1016/j.bbamem.2012.06.007.

Munro, A.W., Girvan, H.M. and McLean, K.J. (2007) 'Cytochrome P450–redox partner fusion enzymes', *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1770(3), pp. 345–359. Available at: https://doi.org/10.1016/j.bbagen.2006.08.018.

Myhrvold, C., Polka, J.K. and Silver, P.A. (2016) 'Synthetic Lipid-Containing Scaffolds Enhance Production by Colocalizing Enzymes', *ACS Synthetic Biology*, 5(12), pp. 1396–1403. Available at: https://doi.org/10.1021/acssynbio.6b00141.

Nesterov, S.V., Ilyinsky, N.S. and Uversky, V.N. (2021) 'Liquid-liquid phase separation as a common organizing principle of intracellular space and biomembranes providing dynamic adaptive responses', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1868(11), p. 119102. Available at: https://doi.org/10.1016/j.bbamcr.2021.119102.

Nishizaki, T. *et al.* (2007) 'Metabolic Engineering of Carotenoid Biosynthesis in Escherichia coli by Ordered Gene Assembly in Bacillus subtilis', *Applied and Environmental Microbiology*, 73(4), pp. 1355–1361. Available at: https://doi.org/10.1128/AEM.02268-06.

Nivina, A. *et al.* (2019) 'Evolution and Diversity of Assembly-Line Polyketide Synthases', *Chemical Reviews*, 119(24), pp. 12524–12547. Available at: https://doi.org/10.1021/acs.chemrev.9b00525.

Nogueira, M. *et al.* (2019) 'Construction of a fusion enzyme for astaxanthin formation and its characterisation in microbial and plant hosts: A new tool for engineering ketocarotenoids', *Metabolic Engineering*, 52, pp. 243–252. Available at: https://doi.org/10.1016/j.ymben.2018.12.006.

Norris, J.L. and Hughes, R.M. (2018) 'protaTETHER – a method for the incorporation of variable linkers in protein fusions reveals impacts of linker flexibility in a PKAc-GFP fusion protein', *FEBS Open Bio*, 8(6), pp. 1029–1042. Available at: https://doi.org/10.1002/2211-5463.12414.

OLIVEIRA, S. *et al.* (2013) *DEVELOPMENT OF EXTRACTION PROCESS OF THE XANTHOPHYLLS LUTEIN AND ZEAXANTHIN FROM BRAZILLIAN MAIZE (ZEA MAYS L.) LANDRACES, p610*. Available at: https://doi.org/10.5301/EJO.2013.11169.

Oreb, M. (2020) 'Construction of artificial membrane transport metabolons – an emerging strategy in metabolic engineering', *FEMS Microbiology Letters*, 367(1), p. fnaa027. Available at: https://doi.org/10.1093/femsle/fnaa027.

Park, S.Y. *et al.* (2022) 'Metabolic engineering of Escherichia coli with electron channelling for the production of natural products', *Nature Catalysis*, 5(8), pp. 726–737. Available at: https://doi.org/10.1038/s41929-022-00820-4.

Patterson, D.P. *et al.* (2014) 'Encapsulation of an Enzyme Cascade within the Bacteriophage P22 Virus-Like Particle', *ACS Chemical Biology*, 9(2), pp. 359–365. Available at: https://doi.org/10.1021/cb4006529.

Peng, F. *et al.* (2020) 'Co-immobilization of multiple enzymes by self-assembly and chemical crosslinking for cofactor regeneration and robust biocatalysis', *International Journal of Biological Macromolecules*, 162, pp. 445–453. Available at: https://doi.org/10.1016/j.ijbiomac.2020.06.141.

Perveen, S. (2018) 'Introductory Chapter', in *Terpenes and Terpenoids*. IntechOpen. Available at: https://doi.org/10.5772/intechopen.79683.

Pollmann, H., Breitenbach, J. and Sandmann, G. (2017) 'Engineering of the carotenoid pathway in Xanthophyllomyces dendrorhous leading to the synthesis of zeaxanthin', *Applied Microbiology and Biotechnology*, 101(1), pp. 103–111. Available at: https://doi.org/10.1007/s00253-016-7769-0.

Polyakov, N.E. *et al.* (2009) 'Water soluble complexes of carotenoids with arabinogalactan', *The Journal of Physical Chemistry. B*, 113(1), pp. 275–282. Available at: https://doi.org/10.1021/jp805531q.

Poshyvailo, L., Lieres, E. von and Kondrat, S. (2017) 'Does metabolite channeling accelerate enzyme-catalyzed cascade reactions?', *PLOS ONE*, 12(2), p. e0172673. Available at: https://doi.org/10.1371/journal.pone.0172673.

Postma, P.W., Lengeler, J.W. and Jacobson, G.R. (1993) 'Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria.', *Microbiological Reviews*, 57(3), pp. 543–594.

Purwani, N.N. *et al.* (2021) 'Modular Assembly of Phosphite Dehydrogenase and Phenylacetone Monooxygenase for Tuning Cofactor Regeneration', *Biomolecules*, 11(6), p. 905. Available at: https://doi.org/10.3390/biom11060905.

Qiu, X. *et al.* (2018) 'Spatial organization of enzymes to enhance synthetic pathways in microbial chassis: a systematic review', *Microbial Cell Factories*, 17(1), p. 120. Available at: https://doi.org/10.1186/s12934-018-0965-0.

Rabe, K.S. *et al.* (2017) 'Cascades in Compartments: En Route to Machine-Assisted Biotechnology', *Angewandte Chemie International Edition*, 56(44), pp. 13574–13589. Available at: https://doi.org/10.1002/anie.201703806.

Rabeharindranto, H. *et al.* (2019) 'Enzyme-fusion strategies for redirecting and improving carotenoid synthesis in S. cerevisiae', *Metabolic Engineering Communications*, 8(December 2018), pp. 1–11. Available at: https://doi.org/10.1016/j.mec.2019.e00086.

Rebelo, B.A. *et al.* (2020) 'Canthaxanthin, a Red-Hot Carotenoid: Applications, Synthesis, and Biosynthetic Evolution', *Plants*, 9(8), p. 1039. Available at: https://doi.org/10.3390/plants9081039.

Reifenrath, M. *et al.* (2020) 'Artificial ER-Derived Vesicles as Synthetic Organelles for *in Vivo* Compartmentalization of Biochemical Pathways', *ACS Synthetic Biology*, 9(11), pp. 2909–2916. Available at: https://doi.org/10.1021/acssynbio.0c00241.

Richter, T.K.S., Hughes, C.C. and Moore, B.S. (2015) 'Sioxanthin, a novel glycosylated carotenoid, reveals an unusual subclustered biosynthetic pathway', *Environmental Microbiology*, 17(6), pp. 2158–2171. Available at: https://doi.org/10.1111/1462-2920.12669.

Riggs, P. (2013) 'Fusion Protein', in *Brenner's Encyclopedia of Genetics*. Elsevier, pp. 134–135. Available at: https://doi.org/10.1016/B978-0-12-374984-0.00565-9.

Rinaldi, M.A., Ferraz, C.A. and Scrutton, N.S. (2022) 'Alternative metabolic pathways and strategies to high-titre terpenoid production in *Escherichia coli*', *Natural Product Reports*, 39(1), pp. 90–118. Available at: https://doi.org/10.1039/D1NP00025J.

Rivera, S.M., Christou, P. and Canela-Garayoa, R. (2014) 'Identification of carotenoids using mass spectrometry', *Mass Spectrometry Reviews*, 33(5), pp. 353–372. Available at: https://doi.org/10.1002/mas.21390.

Robbins, T. *et al.* (2016) 'Structure and Mechanism of Assembly Line Polyketide Synthases', *Current opinion in structural biology*, 41, pp. 10–18. Available at: https://doi.org/10.1016/j.sbi.2016.05.009.

Rodriguez, F. *et al.* (2013) 'Structural model for the protein-translocating element of the twin-arginine transport system', *Proceedings of the National Academy of Sciences*, 110(12), pp. E1092–E1101. Available at: https://doi.org/10.1073/pnas.1219486110.

Royer, J. *et al.* (2020) 'Rhodoxanthin synthase from honeysuckle; a membrane diiron enzyme catalyzes the multistep conversion of β-carotene to rhodoxanthin', *Science Advances*, 6(17), p. eaay9226. Available at: https://doi.org/10.1126/sciadv.aay9226.

Sachdeva, G. *et al.* (2014) 'In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner', *Nucleic Acids Research*, 42(14), pp. 9493–9503. Available at: https://doi.org/10.1093/nar/gku617.

Sadaf, A. *et al.* (2016) 'A class of rigid linker-bearing glucosides for membrane protein structural study', *Chemical Science*, 7(3), pp. 1933–1939. Available at: https://doi.org/10.1039/c5sc02900g.

Sajilata, M.G., Singhal, R.S. and Kamat, M.Y. (2008) 'The Carotenoid Pigment Zeaxanthin—A Review', *Comprehensive Reviews in Food Science and Food Safety*, 7(1), pp. 29–49. Available at: https://doi.org/10.1111/j.1541-4337.2007.00028.x.

Sandmann, G. (2019) 'Antioxidant Protection from UV- and Light-Stress Related to Carotenoid Structures', *Antioxidants*, 8(7), p. 219. Available at: https://doi.org/10.3390/antiox8070219.

Sharma, P. and Guptasarma, P. (2015) '"Super-perfect" enzymes: Structural stabilities and activities of recombinant triose phosphate isomerases from Pyrococcus furiosus and Thermococcus onnurineus produced in Escherichia coli', *Biochemical and Biophysical Research Communications*, 460(3), pp. 753–758. Available at: https://doi.org/10.1016/j.bbrc.2015.03.102.

Shen, H.J. *et al.* (2016) 'Dynamic control of the mevalonate pathway expression for improved zeaxanthin production in Escherichia coli and comparative proteome analysis', *Metabolic Engineering*, 38, pp. 180–190. Available at: https://doi.org/10.1016/j.ymben.2016.07.012.

Shrestha, B. *et al.* (2019) 'Combinatorial approach for improved cyanidin 3-O-glucoside production in Escherichia coli', *Microbial Cell Factories*, 18, p. 7. Available at: https://doi.org/10.1186/s12934-019-1056-6.

Sigmund, F. *et al.* (2018) 'Bacterial encapsulins as orthogonal compartments for mammalian cell engineering', *Nature Communications*, 9(1), p. 1990. Available at: https://doi.org/10.1038/s41467-018-04227-3.

Somasundaram, S. *et al.* (2017) 'Introduction of synthetic protein complex between Pyrococcus horikoshii glutamate decarboxylase and Escherichia coli GABA transporter for the improved production of GABA', *Biochemical Engineering Journal*, 120, pp. 1–6. Available at: https://doi.org/10.1016/j.bej.2016.12.020.

Song, J., Wu, B. and Beitz, E. (2011) 'Functional and evolutional implications of natural channel-enzyme fusion proteins', 2(5), pp. 439–444. Available at: https://doi.org/10.1515/BMC.2011.037.

Sterling, D., Reithmeier, R.A.F. and Casey, J.R. (2001) 'A Transport Metabolon: FUNCTIONAL INTERACTION OF CARBONIC ANHYDRASE II AND CHLORIDE/BICARBONATE EXCHANGERS*', *Journal of Biological Chemistry*, 276(51), pp. 47886–47894. Available at: https://doi.org/10.1074/jbc.M105959200.

Stewart, K.L., Stewart, A.M. and Bobik, T.A. (2020) 'Prokaryotic Organelles: Bacterial Microcompartments in E. coli and Salmonella', *EcoSal Plus*, 9(1), p. 10.1128/ecosalplus.ESP-0025–2019. Available at: https://doi.org/10.1128/ecosalplus.ESP-0025-2019.

Sweetlove, L.J. and Fernie, A.R. (2013) 'The Spatial Organization of Metabolism Within the Plant Cell', *Annual Review of Plant Biology*, 64(1), pp. 723–746. Available at: https://doi.org/10.1146/annurev-arplant-050312-120233.

Sweetlove, L.J. and Fernie, A.R. (2018) 'The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation', *Nature Communications*, 9(1), p. 2136. Available at: https://doi.org/10.1038/s41467-018-04543-8.

Szczupak, A. *et al.* (2017) 'The Electrosome: A Surface-Displayed Enzymatic Cascade in a Biofuel Cell's Anode and a High-Density Surface-Displayed Biocathodic Enzyme', *Nanomaterials*, 7(7), p. 153. Available at: https://doi.org/10.3390/nano7070153.

Tang, Y. *et al.* (2021) 'Cryo-EM structure of *Mycobacterium smegmatis* DyP-loaded encapsulin', *Proceedings of the National Academy of Sciences*, 118(16), p. e2025658118. Available at: https://doi.org/10.1073/pnas.2025658118.

Thodey, K., Galanie, S. and Smolke, C.D. (2014) 'A microbial biomanufacturing platform for natural and semi-synthetic opiates', *Nature chemical biology*, 10(10), pp. 837–844. Available at: https://doi.org/10.1038/nchembio.1613.

Thomas, S.E. and Johnson, E.J. (2018) 'Xanthophylls', *Advances in Nutrition*, 9(2), pp. 160–162. Available at: https://doi.org/10.1093/advances/nmx005.

Thomik, T. *et al.* (2017) 'An artificial transport metabolon facilitates improved substrate utilization in yeast', *Nature Chemical Biology*, 13(11), pp. 1158–1163. Available at: https://doi.org/10.1038/nchembio.2457.

Tippmann, S. *et al.* (2017) 'Affibody Scaffolds Improve Sesquiterpene Production in *Saccharomyces cerevisiae*', *ACS Synthetic Biology*, 6(1), pp. 19–28. Available at: https://doi.org/10.1021/acssynbio.6b00109.

Turmo, A., Gonzalez-Esquer, C.R. and Kerfeld, C.A. (2017) 'Carboxysomes: metabolic modules for CO2 fixation', *FEMS Microbiology Letters*, 364(18). Available at: https://doi.org/10.1093/femsle/fnx176.

Van Rosmalen, M., Krom, M. and Merkx, M. (2017) 'Tuning the Flexibility of Glycine-Serine Linkers to Allow Rational Design of Multidomain Proteins', *Biochemistry*, 56(50), pp. 6565–6574. Available at: https://doi.org/10.1021/acs.biochem.7b00902.

Vanderstraeten, J. and Briers, Y. (2020) 'Synthetic protein scaffolds for the colocalisation of co-acting enzymes', *Biotechnology Advances*, 44, p. 107627. Available at: https://doi.org/10.1016/j.biotechadv.2020.107627.

Vince, J.W. and Reithmeier, R.A.F. (1998) 'Carbonic Anhydrase II Binds to the Carboxyl Terminus of Human Band 3, the Erythrocyte Cl−/HCO3−Exchanger*', *Journal of Biological Chemistry*, 273(43), pp. 28430–28437. Available at: https://doi.org/10.1074/jbc.273.43.28430.

Wan, X. *et al.* (2020) 'Reprogramming microorganisms for the biosynthesis of astaxanthin via metabolic engineering', *Progress in Lipid Research*, p. 101083. Available at: https://doi.org/10.1016/j.plipres.2020.101083.

Wang, J. *et al.* (2023) 'Construction of Bi-Enzyme Self-Assembly Clusters Based on SpyCatcher/SpyTag for the Efficient Biosynthesis of (R)-Ethyl 2-hydroxy-4-phenylbutyrate', *Biomolecules*, 13(1), p. 91. Available at: https://doi.org/10.3390/biom13010091.

Wang, N. and McCammon, J.A. (2016) 'Substrate channeling between the human dihydrofolate reductase and thymidylate synthase', *Protein Science : A Publication of the Protein Society*, 25(1), pp. 79–86. Available at: https://doi.org/10.1002/pro.2720.

Wang, Q., Quan, S. and Xiao, H. (2019) 'Towards efficient terpenoid biosynthesis: manipulating IPP and DMAPP supply', *Bioresources and Bioprocessing*, 6(1), p. 6. Available at: https://doi.org/10.1186/s40643-019-0242-z.

Wang, X. *et al.* (2021) 'Efficient production of oxidized terpenoids via engineering fusion proteins of terpene synthase and cytochrome P450', *Metabolic Engineering*, 64, pp. 41–51. Available at: https://doi.org/10.1016/j.ymben.2021.01.004.

Wang, Y. *et al.* (2022) 'Phase-Separated Multienzyme Compartmentalization for Terpene Biosynthesis in a Prokaryote', *Angewandte Chemie International Edition*, 61(29), p. e202203909. Available at: https://doi.org/10.1002/anie.202203909.

Wang, Y. *et al.* (2023) 'Enhancement of phycocyanobilin biosynthesis in Escherichia coli by strengthening the supply of precursor and artificially self-assembly complex', *Synthetic and Systems Biotechnology*, 8(2), pp. 227–234. Available at: https://doi.org/10.1016/j.synbio.2023.02.005.

Wang, Y. and Yu, O. (2012) 'Synthetic scaffolds increased resveratrol biosynthesis in engineered yeast cells', *Journal of Biotechnology*, 157(1), pp. 258–260. Available at: https://doi.org/10.1016/j.jbiotec.2011.11.003.

Waugh, D.S. (2016) 'Crystal structures of MBP fusion proteins', *Protein Science*, 25(3), pp. 559–571. Available at: https://doi.org/10.1002/pro.2863.

Wei, S.-P. *et al.* (2020) 'Formation and functionalization of membraneless compartments in Escherichia coli', *Nature Chemical Biology*, 16(10), pp. 1143–1148. Available at: https://doi.org/10.1038/s41589-020-0579-9.

Wheeldon, I. *et al.* (2016) 'Substrate channelling as an approach to cascade reactions', *Nature Chemistry*, 8(4), pp. 299–309. Available at: https://doi.org/10.1038/nchem.2459.

Wilner, O.I. *et al.* (2009) 'Enzyme cascades activated on topologically programmed DNA scaffolds', *Nature Nanotechnology*, 4(4), pp. 249–254. Available at: https://doi.org/10.1038/nnano.2009.50.

Wiryaman, T. and Toor, N. (2022) 'Recent advances in the structural biology of encapsulin bacterial nanocompartments', *Journal of Structural Biology: X*, 6, p. 100062. Available at: https://doi.org/10.1016/j.yjsbx.2022.100062.

Wriggers, W., Chakravarty, S. and Jennings, P.A. (2005) 'Control of protein functional dynamics by peptide linkers', *Biopolymers - Peptide Science Section*, 80(6), pp. 736–746. Available at: https://doi.org/10.1002/bip.20291.

Wu, T. *et al.* (2017) 'Membrane engineering - A novel strategy to enhance the production and accumulation of β-carotene in Escherichia coli', *Metabolic Engineering*, 43, pp. 85–91. Available at: https://doi.org/10.1016/j.ymben.2017.07.001.

Wu, T. *et al.* (2018) 'Engineering membrane morphology and manipulating synthesis for increased lycopene accumulation in Escherichia coli cell factories', *3 Biotech*, 8(6), p. 269. Available at: https://doi.org/10.1007/s13205-018-1298-8.

Wu, T. *et al.* (2019) 'Engineering an Artificial Membrane Vesicle Trafficking System (AMVTS) for the Excretion of β-Carotene in Escherichia coli', *ACS Synthetic Biology*, 8(5), pp. 1037–1046. Available at: https://doi.org/10.1021/acssynbio.8b00472.

Wu, Y. *et al.* (2019) 'Combinatorial expression of different β-carotene hydroxylases and ketolases in Escherichia coli for increased astaxanthin production', *Journal of Industrial Microbiology and Biotechnology*, 46(11), pp. 1505–1516. Available at: https://doi.org/10.1007/s10295-019-02214-1.

Wu, Z. *et al.* (2019) 'Combinatorial modulation of initial codons for improved zeaxanthin synthetic pathway efficiency in Escherichia coli', *MicrobiologyOpen*, (August), pp. 1–22. Available at: https://doi.org/10.1002/mbo3.930.

Wurtzel, E.T. (2019) 'Changing Form and Function through Carotenoids and Synthetic Biology', *Plant Physiology*, 179(3), pp. 830–843. Available at: https://doi.org/10.1104/pp.18.01122.

Xie, Y., Chen, S. and Xiong, X. (2021) 'Metabolic Engineering of Non-carotenoid-Producing Yeast Yarrowia lipolytica for the Biosynthesis of Zeaxanthin', *Frontiers in Microbiology*, 12, p. 699235. Available at: https://doi.org/10.3389/fmicb.2021.699235.

Xu, L. *et al.* (2021) 'Dye Decoloring Peroxidase Structure, Catalytic Properties and Applications: Current Advancement and Futurity', *Catalysts*, 11(8), p. 955. Available at: https://doi.org/10.3390/catal11080955.

Xu, P. *et al.* (2016) 'Engineering Yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals', *Proceedings of the National Academy of Sciences of the United States of America*, 113(39), pp. 10848–10853. Available at: https://doi.org/10.1073/pnas.1607295113.

Yabuzaki, J. (2017) 'Carotenoids Database: structures, chemical fingerprints and distribution among organisms', *Database: The Journal of Biological Databases and Curation*, 2017, p. bax004. Available at: https://doi.org/10.1093/database/bax004.

Yokoyama, A., Adachi, K. and Shizuri, Y. (1995) 'New Carotenoid Glucosides, Astaxanthin Glucoside and Adonixanthin Glucoside, Isolated from the Astaxanthin-Producing Marine Bacterium, Agrobacterium aurantiacum', *Journal of Natural Products*, 58(12), pp. 1929–1933. Available at: https://doi.org/10.1021/np50126a022.

Yokoyama, A., Shizuri, Y. and Misawa, N. (1998) 'Production of new carotenoids, astaxanthin glucosides, by Escherichia coli transformants carrying carotenoid biosynthesis genes', *Tetrahedron Letters*, 39(22), pp. 3709–3712. Available at: https://doi.org/10.1016/S0040-4039(98)00542-5.

Yourno, J., Kohno, T. and Roth, J.R. (1970) 'Enzyme Evolution: Generation of a Bifunctional Enzyme by Fusion of Adjacent Genes', *Nature*, 228(5274), pp. 820–824. Available at: https://doi.org/10.1038/228820a0.

Yu, Q. *et al.* (2010) 'The Lycopene Cyclase CrtY from Pantoea ananatis (Formerly Erwinia uredovora) Catalyzes an FADred-dependent Non-redox Reaction', *The Journal of Biological Chemistry*, 285(16), pp. 12109–12120. Available at: https://doi.org/10.1074/jbc.M109.091843.

Zafar, J. *et al.* (2021) 'Biochemical and Immunological implications of Lutein and Zeaxanthin', *International Journal of Molecular Sciences*, 22(20), p. 10910. Available at: https://doi.org/10.3390/ijms222010910.

Zakeri, B. *et al.* (2012) 'Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin', *Proceedings of the National Academy of Sciences of the United States of America*, 109(12), pp. E690–E697. Available at: https://doi.org/10.1073/pnas.1115485109.

Zakynthinos, G. and Varzakas, T. (2016) 'Carotenoids: From Plants to Food Industry', *Current Research in Nutrition and Food Science Journal*, 4(1), pp. 38–51. Available at: https://doi.org/10.12944/CRNFSJ.4.Special-Issue1.04.

Zha, J., Wu, X. and Koffas, M.A. (2020) 'Making brilliant colors by microorganisms', *Current Opinion in Biotechnology*, 61, pp. 135–141. Available at: https://doi.org/10.1016/j.copbio.2019.12.020.

Zhai, G. *et al.* (2023) 'Insights into azalomycin F assembly-line contribute to evolution-guided polyketide synthase engineering and identification of intermodular recognition', *Nature Communications*, 14(1), p. 612. Available at: https://doi.org/10.1038/s41467-023-36213-9.

Zhang, C. *et al.* (2013) 'Combining Genotype Improvement and Statistical Media Optimization for Isoprenoid Production in E. coli', *PLoS ONE*, 8(10), p. e75164. Available at: https://doi.org/10.1371/journal.pone.0075164.

Zhang, C. *et al.* (2015) 'Experimental design-aided systematic pathway optimization of glucose uptake and deoxyxylulose phosphate pathway for improved amorphadiene production', *Applied Microbiology and Biotechnology*, 99(9), pp. 3825–3837. Available at: https://doi.org/10.1007/s00253-015-6463-y.

Zhang, C. (2018) 'Biosynthesis of Carotenoids and Apocarotenoids by Microorganisms and Their Industrial Potential'. Edited by L.Q. Zepka, E. Jacob-Lopes, and V.V.D. Rosso. Available at: https://doi.org/10.5772/intechopen.79061.

Zhang, C. *et al.* (2018) 'Multidimensional heuristic process for high-yield production of astaxanthin and fragrance molecules in Escherichia coli', *Nature Communications*, 9(1), pp. 1–12. Available at: https://doi.org/10.1038/s41467-018-04211-x.

Zhang, C., Chen, X. and Too, H.-P. (2020) 'Microbial astaxanthin biosynthesis: recent achievements, challenges, and commercialization outlook', *Applied Microbiology and Biotechnology*, 104(13), pp. 5725–5737. Available at: https://doi.org/10.1007/s00253-020-10648-2.

Zhang, Y. *et al.* (2017) 'Protein-protein interactions and metabolite channelling in the plant tricarboxylic acid cycle', *Nature Communications*, 8(1), p. 15212. Available at: https://doi.org/10.1038/ncomms15212.

Zhang, Y. *et al.* (2018) 'Biotechnological production of zeaxanthin by microorganisms', *Trends in Food Science & Technology*, 71, pp. 225–234. Available at: https://doi.org/10.1016/j.tifs.2017.11.006.

Zhang, Y. and Fernie, A.R. (2021) 'Metabolons, enzyme–enzyme assemblies that mediate substrate channeling, and their roles in plant metabolism', *Plant Communications*, 2(1), p. 100081. Available at: https://doi.org/10.1016/j.xplc.2020.100081.

Zhang, Y., Tsitkov, S. and Hess, H. (2016) 'Proximity does not contribute to activity enhancement in the glucose oxidase–horseradish peroxidase cascade', *Nature Communications*, 7(1), p. 13982. Available at: https://doi.org/10.1038/ncomms13982.

Zhang, Y.-H.P. (2011) 'Substrate channeling and enzyme complexes for biotechnological applications', *Biotechnology Advances*, 29(6), pp. 715–725. Available at: https://doi.org/10.1016/j.biotechadv.2011.05.020.

Zhao, E.M. *et al.* (2019) 'Light-based control of metabolic flux through assembly of synthetic organelles', *Nature chemical biology*, 15(6), pp. 589–597. Available at: https://doi.org/10.1038/s41589-019-0284-8.

Zhou, Y.J. *et al.* (2016) 'Harnessing Yeast Peroxisomes for Biosynthesis of Fatty-Acid-Derived Biofuels and Chemicals with Relieved Side-Pathway Competition', *Journal of the American Chemical Society*, 138(47), pp. 15368–15377. Available at: https://doi.org/10.1021/jacs.6b07394.

Zhu, F., Peña, M. and Bennett, G.N. (2021) 'Metabolic engineering of Escherichia coli for quinolinic acid production by assembling L-aspartate oxidase and quinolinate synthase as an enzyme complex', *Metabolic Engineering*, 67, pp. 164–172. Available at: https://doi.org/10.1016/j.ymben.2021.06.007.