Bacterial microcompartments : from study of shell subunit assembly to the development of tools for the nanotechnologies

Lucie Barthe

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

Université de Toulouse

Doctorat de l'Université de Toulouse

préparé à l'INSA Toulouse

Les microcompartiments bactériens : étude de l'assemblage des protéines hexamériques des coques et développement d'outils pour les nanotechnologies

Thèse présentée et soutenue, le 28 février 2024 par

Lucie BARTHE

École 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
Philippe URBAN et Luis F. GARCIA ALLES

Composition du jury
M. Patrick BRON, Président, INSERM Occitanie-Méditerrannée
M. Juan REGUERA, Rapporteur, INSERM PACA et Corse
M. Nicolas BARNICH, Rapporteur, Université Clermont Auvergne
Mme Stéphanie CABANTOUS, Examinatrice, INSERM Occitanie-Pyrénées
M. Philippe URBAN, Directeur de thèse, CNRS Occitanie-Ouest
M. Luis GARCIA-ALLES, Co-directeur de thèse, CNRS Occitanie-Ouest

Bacterial microcompartments: from the study of shell subunit assembly to the development of tools for the nanotechnologies

by Lucie BARTHE

Declaration of authorship

I, Lucie Barthe, declare that this thesis manuscript, entitled “Bacterial microcompartments: from the study of shell subunit assembly to the development of tools for the nanotechnologies” and the work presented in it are my own. I confirm that the research included within this thesis is my own work or was carried out in collaboration with others, duly acknowledged below. For previously published material or illustrations used in this manuscript, the original authors are always clearly stated.

A portion of the vector constructs used in the first chapter were made by Luis Garcia-Alles before the beginning of my thesis. The whole section dealing with the sumoylation of RMM was carried out in collaboration with Dorian Grégnanin. All the samples for transmission electron microscopy analysis were prepared (included, stained and sliced) by Vanessa Soldan and Stéphanie Balor from the Microscopie électronique integrative (METi) platform of the Centre de Biologie Intégrative of Toulouse. The construction of *Klebsiella* BMC-H pair library was partially performed by Paola Randazzo, Mickael Dinclaux and Sandra Serres from the strain engineering platform of Toulouse White Biotechnology. The design of BMC-H variants was made by 2 collaborator teams: one from the Mathématiques et Informatique Appliquées de Toulouse unit (composed by Thomas Schiex, Marianne Defresne, Samuel Buchet and Simon de Givry) and the other from Toulouse Biotechnology Institute (composed by Sophie Barbe, Delphine Desseaux and Jérémy Esqué).

Acknowledgments

L'espace d'un instant, j'aimerais revenir au français afin de mieux exprimer aux différentes personnes qui ont importé durant cette thèse et, de manière globale, durant mon parcours, toute ma gratitude.

Je voudrais remercier en premier lieu Monsieur Nicolas Barnich, Professeur et Directeur de Recherche dans l'unité Microbes Intestin Inflammation et Susceptibilité de l'Hôte de Clermont Ferrand, et Monsieur Juan Reguera, Directeur de recherche au laboratoire Architecture et Fonction des Macromolécules Biologiques de Marseille, pour l'honneur qu'ils m'ont fait en acceptant d'être rapporteurs de cette thèse. J'adresse également tous mes remerciements à Madame Stéphanie Cabantous, Chercheuse au Centre de Recherches en Cancérologie de Toulouse, ainsi qu'à Monsieur Patrick Bron, Directeur de recherche au Centre de Biologie Structurale de Montpellier pour avoir accepté d'être les examinateurs de ma thèse.

Ensuite, je voudrais remercier mon encadrant de thèse, Luis Garcia-Alles, pour m'avoir donné la chance de réaliser cette thèse ainsi que d'acquérir et d'étendre mes connaissances en biologie et biotechnologies durant ces 3 années et quelques mois de thèse. Je tiens aussi à remercier les personnes de mon équipe, Sara, Philippe, Hélène, Denis, Christophe et Laura pour leur accueil, leur bonne humeur et leurs conseils.

Egalement à tous ceux qui sont venus et sont repartis en laissant une trace ; pour l'aide qu'il m'a apporté dans les expériences, un grand merci à mon stagiaire, Dorian Grégnanin ! Pour le partage des déboires de la thèse et les instants Littérature et Psychologie, merci Aurélie Bouin. Je ne peux terminer ce tour de laboratoire sans penser à vous 4, Adilya Dagkesamanskaya, Thomas Gosselin-Monplaisir, Denis Jallet et Hanna Kulyk. Merci pour tout votre soutien, vos conseils et pour tous ces instants partagés. Cette thèse ne serait sûrement pas arrivée à son terme si vous n'aviez pas été là. Encore merci et ne changez pas !

Avant de passer aux personnes qui me sont le plus chères, j'aimerais remercier toutes les personnes qui m'ont marquée durant mon parcours de formation et qui ont contribué à faire de moi la scientifique que je suis. Je pense, entre autres, à Madame Hennebois, à Monsieur Bonnemaison qui m'a fait aimer les maths, à Madame Bousquet-Dreux qui m'a appris l'anglais et m'a donné le goût des langues vivantes, à Monsieur Dumas qui m'a suivie durant tout mon parcours à l'Université Paul Sabatier et a su satisfaire mon irrépressible envie d'apprendre, à Madame Nieto et Monsieur Ecochard qui m'ont appris toutes les ficelles de la biologie moléculaire.

Je voudrais également dire un grand merci à mon précédent encadrant, Eric Agius, qui, malgré les nombreuses années qui ont passé, est toujours disponible et prêt à m'aider. A Valérie Lobjois, ma précédente cheffe, pour sa gentillesse, sa disponibilité et l'opportunité qu'elle m'a offerte d'apprendre davantage et de m'épanouir en tant qu'ingénieure d'études, merci. A Brice Enjalbert pour m'avoir permis d'encadrer les étudiants au concours iGEM et de transmettre à mon tour mon savoir, merci beaucoup !

Pour finir, j'aimerais remercier toute ma famille. A ma mère, mon père et mes deux sœurs. Je ne serais pas arriver là où je suis sans vous, sans votre soutien et votre amour. Je ne serais pas la personne que je suis. Mille mercis ! A mon compagnon qui m'a soutenue durant toute ma thèse et particulièrement durant ces derniers mois de rédaction. Mille mercis à toi aussi ! A ma nièce aux câlins réconfortants, j'en voudrais encore plein !

Statements

This PhD thesis was funded by the French National Research Agency: ANR-19-CE09-0032-01. It has also been labelled by the EUR BioEco and it benefited from a grant managed by the same agency, under the "Investissements d'Avenir" program: ANR-18-EURE-0021.

During the course of this thesis, efforts on plastic utilization have been done. Disposable plastic petri dishes, tubes and culture tubes were replaced by reusable glass equipment. Plastic 96-well plates and pipet tips were washed, sterilized and reused when possible.

This allowed to save at least 81kg of plastic or, in other words, to prevent the production as well as the collection and destruction or storage of these 81kg and thus the utilization of energy and water resources that come along. Of note, the 81kg figure is an underestimated value which does not take into account all the bottles that were used for solution and medium preparation instead of plastic tubes.

This was set in a global view of ecology and awareness that scientists should be the examples, making effort towards energetic sobriety and sustainability, for a greener planet, for longer…

Table of contents

Table of contents

Introduction

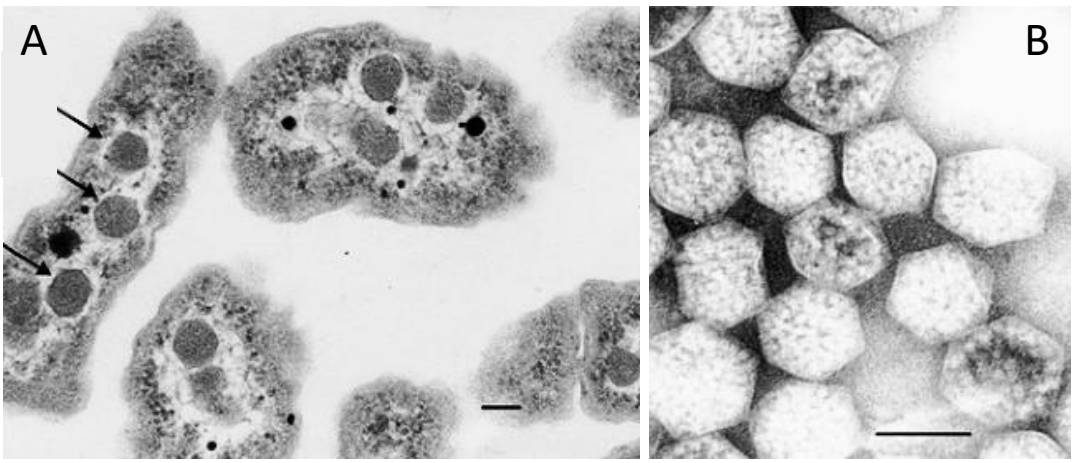


Figure 1. Morphology of the bacterial microcompartment (BMC).
TEM observations of *Halothiobacillus neapolitanus* carboxysomes, *in vivo* (**A**) and *in vitro,* after purification (**B**). Black arrows indicate BMC structures. Scale bar = 100nm. TEM acquisitions from (Tsai *et al*, 2007).

Part 1. Introduction

1. The bacterial microcompartment, a natural factory

1.1. General features of the bacterial microcompartment

Bacterial microcompartments (BMC) are organelles found naturally in several bacteria (figure 1). Discovered for the first time on electro-microscopy micrographs of cyanobacteria (Drews & Niklowitz, 1956), it was shown recently that BMCs were present in at least 45 different bacterial phyla. This includes the firmicutes, deltaproteobacteria, gammaproteobacteria and the actinobacteria which have the biggest BMC diversity (Sutter *et al*, 2021).

BMCs do not possess a lipidic membrane like eukaryotic organelles but are instead proteinaceous structures of 40 to 600nm in diameter, that encapsulate specific metabolic pathways. BMCs are composed of a semi-permeable shell made up of multiple protein subunits that most often adopt a polyhedral geometry. This shell encloses an enzymatic core and segregates it from the rest of the cell cytosol (figure 2). Both the shell proteins and the enzymatic set are generally encoded within a single *locus* and form an operon (Rae *et al*, 2013; Herring *et al*, 2018; Chowdhury *et al*, 2014).

1.2. BMCs are optimized metabolic factories present within bacteria

BMCs play an important role in the optimization of bacterial metabolic pathways. Indeed, they accelerate substrate catalysis by concentrating defined enzymes, in a restricted place, the BMC lumen (Jakobson *et al*, 2017; Tcherkez *et al*, 2006). Their semi-permeable shell represents a physical barrier that selectively sequesters the intermediates of reaction that would otherwise diffuse in the cytosol. Thus, indirectly, the shell prevent cytosolic enzymes to compete for substrate with luminal enzymes. Moreover, it can impede the escape of volatile intermediates of the metabolic pathway out of the cell, which would represent a loss of valuable carbon and would decrease the catalysis efficiency (Penrod & Roth, 2006; Cai *et al*, 2009). Finally, BMC shell can retain toxic intermediates such as aldehydes that

could damage the cell or impact its growth (Sampson & Bobik, 2008; Havemann *et al*, 2002; Chowdhury *et al*, 2015).

Many bacteria of the mammal *flora* present BMC operons (Prentice, 2021; Asija *et al*, 2021; Sutter *et al*, 2021). Some studies have proposed that BMCs could be involved in bacterial pathogenesis. Indeed, many pathogenic bacteria such as *Salmonella*, *Clostridium*, *Streptococcus*, *Citrobacter*, *Shigella* or *Klebsiella* are endowed with BMCs. The BMC types that are more frequently associated with pathogens are the ethanolamine utilization BMC (EUT), propanediol utilization BMC (PDU), glycyl radical enzyme-associated BMC (GRM) or the sugar phosphate utilization BMC (SPU). These BMCs use different products resulting from the cell degradation (membrane phospholipid or deoxyribonucleic acid (DNA)) as substrates. Their processing can furnish a valuable nutrient source to the host cell.

Some of them, like the EUT, were shown to confer a selective growth advantage to the bacteria which carry them. Pathogenic *E. coli LF82* had an increased growth rate compared to the MG1655 strain (deprived of EUT), when cultured in presence of the EUT substrate, ethanolamine, and were more prone to infect the mouse gut than an engineered EUT-deficient LF82 strain (Delmas *et al*, 2019). In the same extent, in *Salmonella enterica*, the PDU was also highlighted as a factor a virulence (Faber *et al*, 2017) and granted the bacteria with the ability to thrive and grow within macrophages (Prentice, 2021).

Yet, some commensal bacteria have also BMC *loci* such as some *E. coli* strains that code for the EUT or PDU or *Lactobacillus reuteri* and *Enterobacterium hallii* that contain a PDU (Sutter *et al*, 2021). The Nissle and HS commensal strains of *E. coli* have a *eut locus* and were shown to be able to grow on ethanolamine and even to outcompete pathogenic enterohaemorrhagic *E. coli* (Rowley *et al*, 2018). Besides, inactivation of the *eut* operon in pathogenic *Enterococcus faecium* was shown to increase its competitiveness for the mice intestine colonisation and the same inactivation of *Clostridium difficile eut* operon conferred a higher lethality in the hamster upon infection (Kaval *et al*, 2018; Nawrocki *et al*, 2018).

Then, although evidences contradict on whether BMCs are associated with pathogenesis, one thing appears clear: BMCs represent a selective advantage that allow bacteria to grow on niche substrates (ethanolamine, propanediol,…).

1.3. Different structural proteins compose the BMC shell

BMCs are complex macrostructures forming through the spontaneous self-assembly of hundreds to thousands of proteins (Sun *et al*, 2022). This fact plus their natural bioreactor functions within bacterial cells makes them of big interest for bioengineering and synthetic biology.

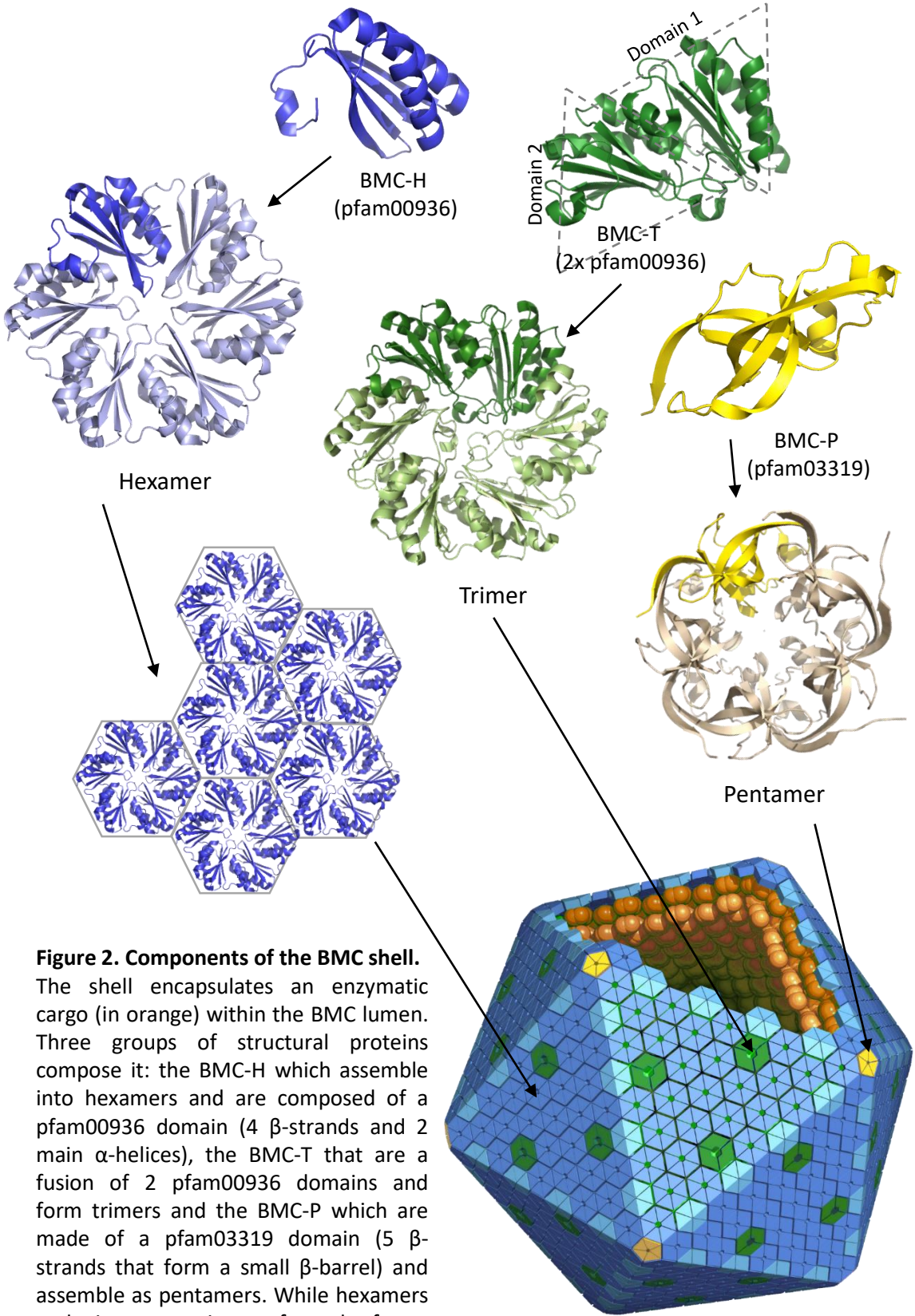


Figure 2. Components of the BMC shell.
The shell encapsulates an enzymatic cargo (in orange) within the BMC lumen. Three groups of structural proteins compose it: the BMC-H which assemble into hexamers and are composed of a pfam00936 domain (4 β-strands and 2 main α-helices), the BMC-T that are a fusion of 2 pfam00936 domains and form trimers and the BMC-P which are made of a pfam03319 domain (5 β-strands that form a small β-barrel) and assemble as pentamers. While hexamers and trimers associate to form the facets and edges of the BMC, the pentamers cap the vertices.
The BMC illustration is adapted from Todd Yeates drawing.

The BMC shell is composed of 3 groups of proteins (figure 2). The main group are protomers which adopt a pfam00936 fold (4 antiparallel β-strands flanked by 2 α-helices), oligomerize per 6 and take the form of an hexagon. As such, they are called BMC-H for BMC hexamer-forming protomer.

The protomers belonging to the second group are constituted by a fusion of 2 pfam00936 domains. Thus, they share the same hexagonal symmetry as the hexamers but actually associate as trimers. Generally, the BMC trimer-forming protomer (BMC-T) fused domains have a low sequence identity (less than 16%), suggesting that, contrary to hypotheses that BMC-T arose from gene duplication, they are more probably resulting from the fusion of 2 contiguous BMC-H homologs (Sagermann *et al*, 2009).

BMC-H and BMC-T compose the facets of the BMC shell (Kerfeld *et al*, 2005; Tanaka *et al*, 2010). They play a role in shell semi-permeability (Chowdhury *et al*, 2015; Slininger Lee *et al*, 2017). Indeed, these subunits have a central pore whose size and residue content dictate the selectivity of molecules allowed to penetrate in and out the BMC. The trimer pore is generally larger (8 to 13Å) than the hexamer one (4 to 7Å) (Cai *et al*, 2013; Kerfeld *et al*, 2005; Tanaka *et al*, 2010) and it was proposed to allow the passage of large molecules such as cofactors (NAD^+/NADH, coenzyme A, ATP, for instance). However, in order to prevent the escape of all the metabolites out of the BMC, the aperture of the trimer pore seems to be tightly controlled. Upon crystallisation, trimers like EutL or CsoS1D were observed to adopt alternatively a closed or an open pore configuration (Tanaka *et al*, 2010; Klein *et al*, 2009).

The third and last group of shell subunits are the less abundant proteins of the BMC, making them difficult to detect in purified-BMC SDS-PAGE (Parsons *et al*, 2010a). These protomers adopt a distinct fold than the BMC-H and BMC-T. They contain a pfam03319 domain (5 antiparallel β-strands organizing as a β-barrel) and associate as pentamers, hence their name of BMC-P. Their pentagonal geometry fits the 5-fold symmetric gap left by the BMC facets (Tanaka *et al*, 2008; Wheatley *et al*, 2013). Thus, the BMC-P functions are thought to be limited to capping the BMC polyhedron vertices.

Incorporation of pentamers into the BMC shell appears to be one of the final step of the BMC biogenesis, allowing BMC closure and stopping subsequent nucleation (see section 3.1) (Cameron *et al*, 2013; Parsons *et al*, 2010a).

2. Specialization of BMC in a wide range of substrate metabolism

BMCs are specialized metabolic structures which functions vary according to the enzymatic set they encapsulate. Sutter *et al* described the existence of a minimum of 11 BMC types, some of them

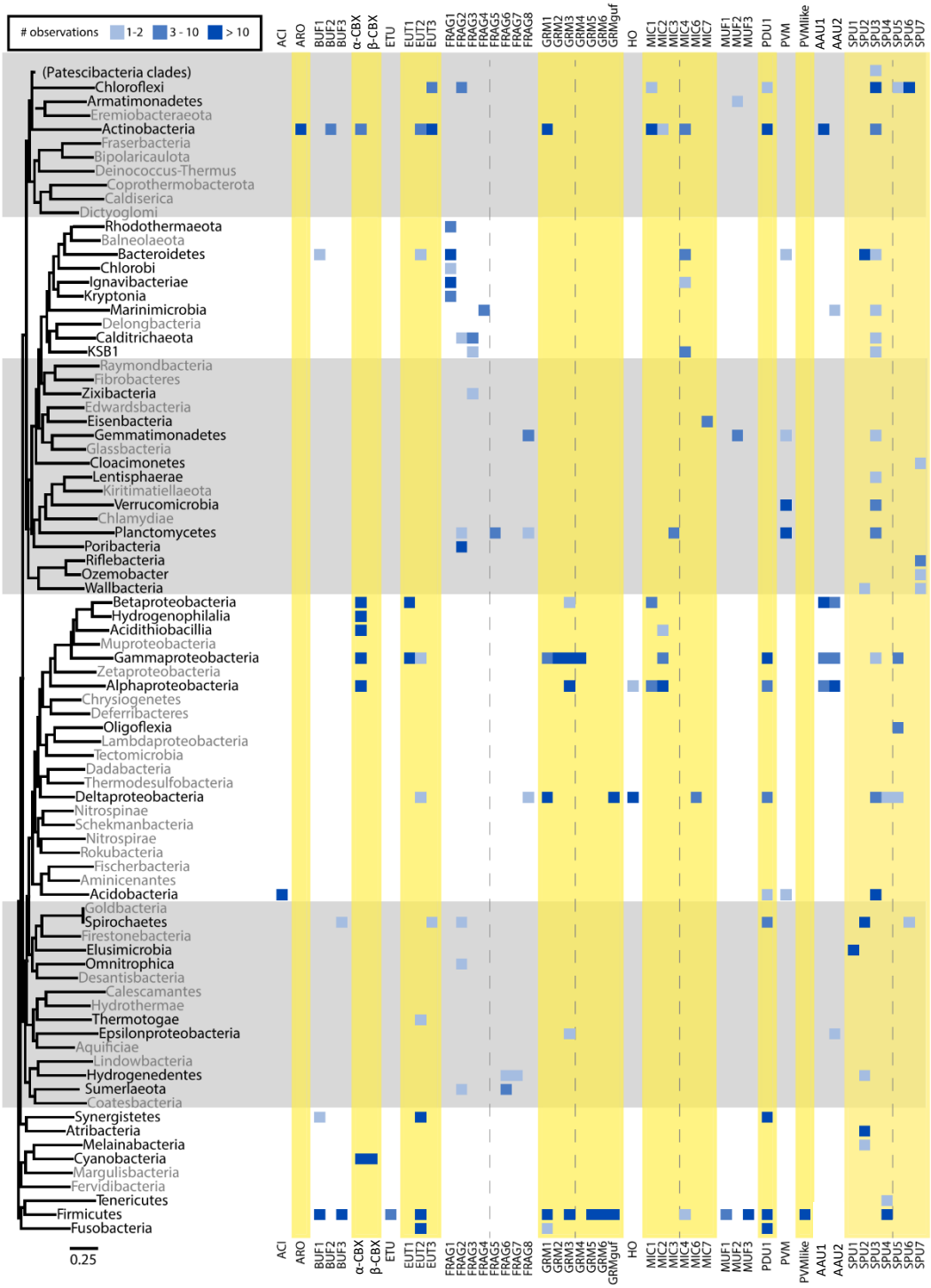


Figure 3. Distribution of the different BMC types across bacterial *phyla*.
Over 83 *phyla*, 45 bear one or multiple BMC genetic *loci*. The most BMC-populated *phyla* are the actinobacteria, proteobacteria and the firmicutes. Shades of blue represent the number of bacterial species in the *phyla* containing a given BMC type. *Phyla* written in grey are devoid of BMC *locus*. **BMC of known functions: ARO** for aromatic substrate BMC, **CBX** for carboxysome (carbon fixation), **ETU** for ethanol utilization BMC, **EUT** for ethanolamine utilization BMC, **GRM** for glycyl radical enzyme-associated microcompartment (choline, 1,2-propanediol, fuculose-phosphate or rhamnulose-phosphate degradation, depending on the subtype), **PDU** for 1,2-propanediol utilization BMC, **PVM** for *Planctomycete* and *Verrucomicrobia* microcompartment (fucose, rhamnose, fucoidan degradation), **AAU** for aminoacetone utilization BMC (amino-2-propanol degradation) and **SPU** for sugar-phosphate utilization BMC. **BMCs with unknown functions: ACI** for BMC from acidobacteria, **BUF** for BMC of unknown functions, **FRAG** for fragmented-type BMC in multiple genetic *loci*, **HO** for *Haliangium ochraceum* BMC, **MIC** for metabolosome *locus* with incomplete core (lacks one of the "signature enzymes", alcohol dehydrogenase or phosphotransacylase) and **MUF** for metabolosome of unknown functions. Illustration from (Sutter *et al*, 2021).

divided in several subtypes (figure 3), which differ in gene order or in shell subunit number coded within the BMC operon (Sutter *et al*, 2021). However, all the BMC operons uncovered to date do not have a defined metabolic function yet. In this study, some incomplete BMC *loci* could also be detected thanks to an homology search directed against shell subunits and analysis of juxtaposed genes belonging to the same operon.

The most well-known BMCs are the carboxysome (CBX), the EUT and the PDU involved in atmospheric carbon fixation, ethanolamine or 1,2-propanediol catabolism, respectively. In the literature, there are 2 other model BMCs: the AAU for aminoacetone utilization BMC (previously referred to as RMM for *Rhodococcus* and *Mycobacterium* microcompartment) and the HO BMC, called after the organism in which it was first identified, *Haliangium ochraceum* (Mallette & Kimber, 2017; Lassila *et al*, 2014). The shell of these BMCs has been extensively studied and engineered (Mallette & Kimber, 2017; Sutter *et al*, 2017; Hagen *et al*, 2018b, 2018a; Greber *et al*, 2019). However, the metabolic functions associated with the HO BMC are still unclear and require more scrutiny.

BMCs are involved in diverse other metabolic pathways such as sugar metabolism (degradation of rhamnose, fucose, fucoidan: in the *Planctomycete* and *Verrucomicrobia* BMC also called PVM), sugar phosphate derived from nucleic acids degradation (in the SPU), ethanol catabolism (in the ETU) or amino acid and derivatives catabolism (degradation of choline in the GRM for example) (Sutter *et al*, 2021).

2.1. The carboxysome

The CBX is found in practically all cyanobacteria and some chemoautotrophs from the fresh waters and oceans. Although constitutively expressed, the CBX expression can be modulated by different parameters. As demonstrated by several studies, bacteria responded to a CO_2 shortage or a high-light stimulus by overexpressing the CBX (McKay *et al*, 1993; Sun *et al*, 2016).

The CBXs were the first BMCs to be identified in electron microscopy thanks to their very regular and tight icosahedral shape (Drews & Niklowitz, 1956). These BMCs are specialized in atmospheric carbon fixation (figure 4A), a process catalysed by the ribulose-1,5-biphosphate carboxylase/oxygenase (RuBisCO). According to some estimations, 25% of carbon fixation on Earth would be dependent on the CBX activity (Behrenfeld *et al*, 2001).

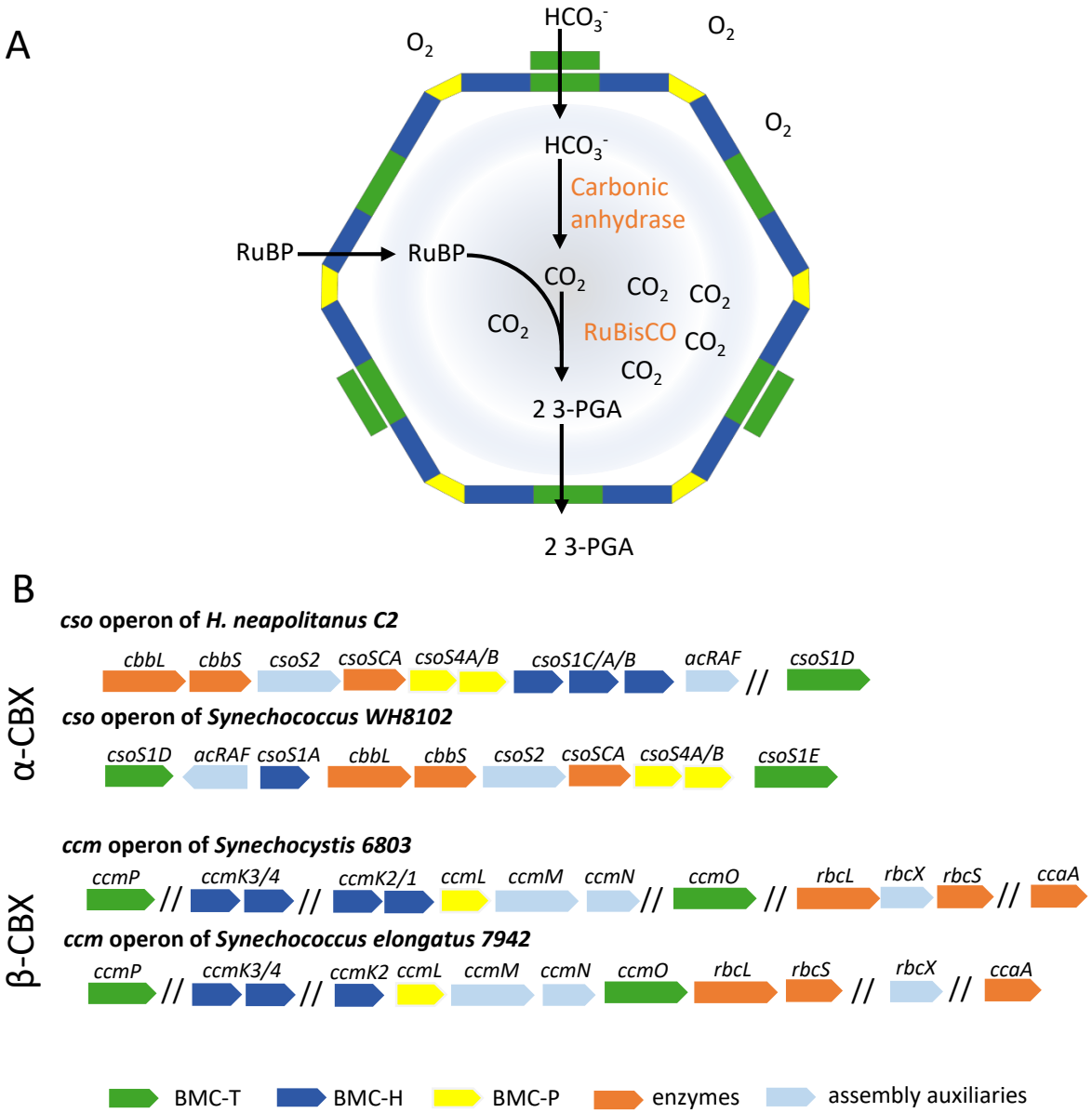


Figure 4. The carboxysome.

A. Carbon fixation pathway encapsulated in the carboxysome (CBX). RuBP: ribulose-1,5-biphosphate. 3-PGA: 3-phosphoglycerate. **B.** Genetic organization of operons coding for the α-CBX (*cso*) or β-CBX (*ccm*) in different CBX bearing organisms. The cbbL/rbcL or the cbbS/rbcS sequences code for the RuBisCO subunits. They require the action of acRAF/rbcX gene product in order to be assembled. The carbonic anhydrase is coded by the *csoSCA* or *ccaA*. The double slash signifies independent *loci*. Illustrations adapted from (Bobik *et al*, 2015; Rae *et al*, 2013).

The CBX is specialized in carbon fixation

Carbon fixation begins with concentration of bicarbonate in the bacterium cytosol by active system uptakes (Rae *et al*, 2013). Bicarbonate and ribulose-1,5-bisphosphate (RuBP) then penetrate the CBX where the bicarbonate is converted to CO_2 by the carbonic anhydrase (CA) (figure 4A). Molecular dynamic simulations on BMC-H homologs from the CBX indicated that their pores would have a lower permeability to CO_2 than to anionic bicarbonate, implying that CO_2 could not diffuse back to the cytosol and would accumulate around the RuBisCO which would favour the carboxylation of RuBP (Mahinthichaichan *et al*, 2018). This reaction gives rise to 2 molecules of 3-phosphoglycerate (3-PGA) that finally exits the CBX to join the central metabolism.

CBX has a double role: while preventing CO_2 escape from the cell, it also improves the RuBisCO catalytic activity by increasing CO_2 concentration around the enzyme (up to 1000-fold the atmospheric concentration), thus favouring the carbon fixation reaction over the photorespiration (Cai *et al*, 2009; Badger, 2003).

Indeed, the RuBisCO has a low affinity for CO_2. In presence of oxygen or of low CO_2 concentration, it would preferentially catalyse the first step of photorespiration (production of 2-phosphoglycolate plus 3-PGA from RuBP and O_2) which would lead to the release of CO_2 and carbon loss. Plants do not possess an encapsulated RuBisCO. Instead, the enzyme localises within the chloroplasts where CO_2 is free to diffuse across the membrane. If one particular group of plants (named the C4 plants) has a carbon concentrating mechanism (CCM) that provides the RuBisCO with an environment enriched in CO_2, most plants do not and undergo a loss of 30 to 60% in carbon due to photorespiration (Zhu *et al*, 2010).

The main CBX subtypes

It exists different CBX subtypes that differ on the RuBisCO form they enclose. The α-CBX contains the form IA (CbbL/S subunits) whereas the β-CBX has the form IB (RbcL/S subunits), the same form found in plants (Badger, 2003). These 2 subtypes of CBXs are encoded by different operons (figure 4B). The α-CBX is coded by a single genetic *locus*, the *cso* operon, found in marine α-cyanobacteria and some chemoautotrophs (proteobacteria and actinobacteria; figure 3) (Rae *et al*, 2013). On the other hand, the β-CBX is expressed from the main *ccm* operon and distinct satellite *loci* coding for the shell subunits CcmK3 and CcmK4 or CcmP, or the carbonic anhydrase CcaA. The β-CBX exclusively clusters in β-cyanobacteria from fresh waters (figure 3) (Rae *et al*, 2013). Finally, they also diverge from each other by the composition of their shell and the mechanism of its assembly. While α-CBX shell and cargo enzymes assemble concomitantly, with the enzymes paving the inner shell, the β-CBX shell is formed

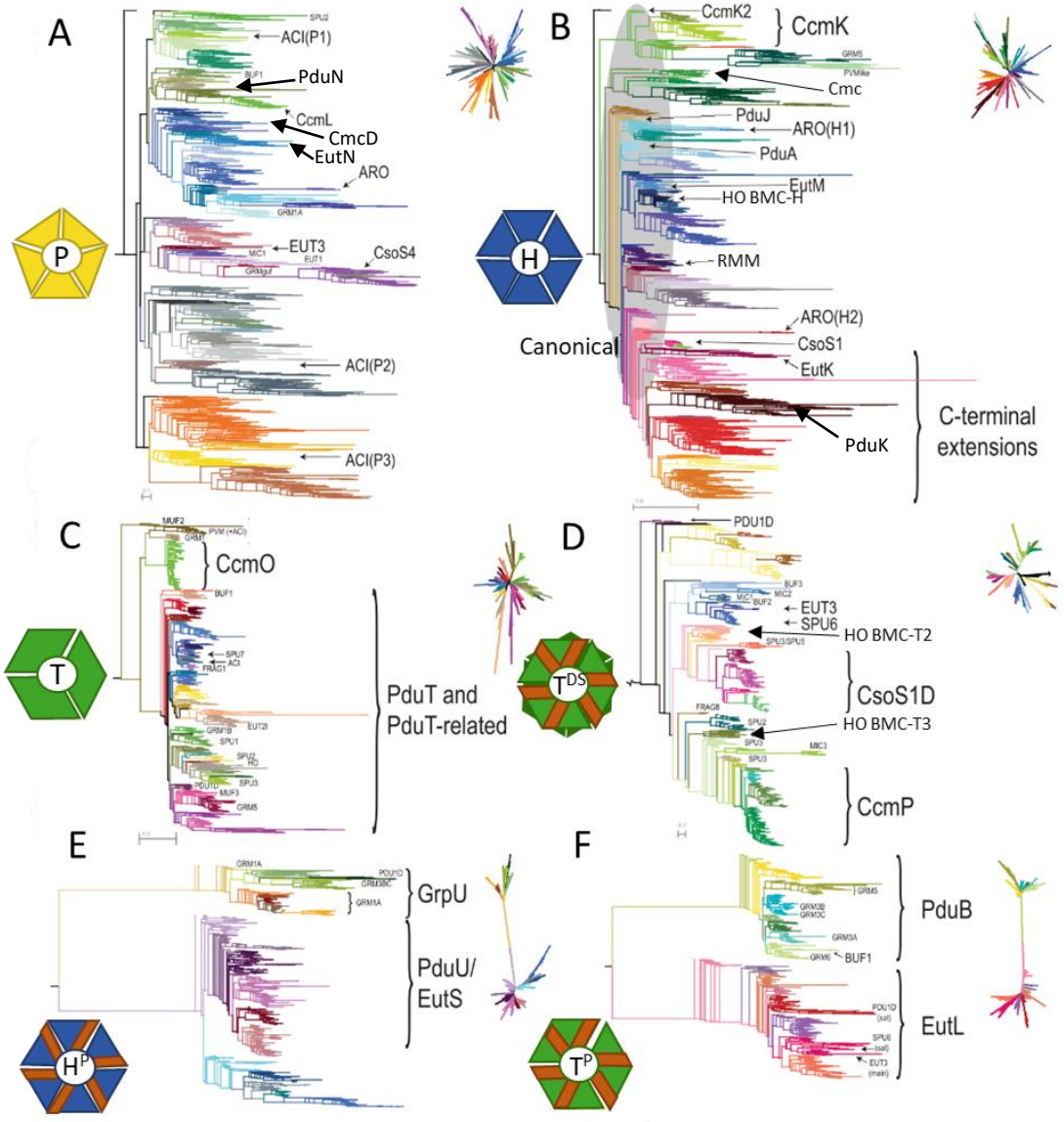


Figure 5. Phylogenetic trees of the BMC shell subunits.
The different trees were built on the basis of sequence alignments between BMC-P (**A**), BMC-H (**B**), BMC-T (**C**) or subunit variants: double-stacked circularly-permuted BMC-T (**D**), circularly-permuted BMC-H (**E**) or simple circularly-permuted BMC-T (**F**).
Illustration adapted from (Sutter *et al*, 2021).

around a dense proto-CBX composed of RuBisCO, CcmM, a scaffolding protein, and CcaA (Long *et al*, 2007; Dai *et al*, 2018).

Phylogenetic analysis suggested that these operons appeared separately. For instance, the β-CBX shell subunits (CcmK, CcmP or CcmL) are more related to other BMC type subunits than to their homologs from the α-CBX (CsoS1, CsoS1D or CsoS4; figure 5) (Sutter *et al*, 2021). Then, it is possible that the β-CBX was the result of horizontal gene transfer and evolution from another BMC type operon rather than evolution from a common *cbx* ancestor.

The BMC shell as the nucleation centre of the α-CBX

Generally, the α-CBX counts 10 different subunits, a number which may vary between organisms. It is formed through the shell-first assembly (see section 3.1.1). In this scheme, a highly connected protein network forms between the shell and the cargo enzymes.

CsoS1A, CsoS1B and CsoS1C are shell subunit homologs that form hexamers. Their 3D structures were determined by X-ray crystallography, except for CsoS1B but as it shares approximately 90% of sequence identity with its counterparts (the main difference residing in an extra C-terminal 12-residue long extension for CsoS1B), one could presume that it would also share the same behaviour (Tsai *et al*, 2007, 2009). CsoS1A and CsoS1C hexamers self-assemble to form the facets of the BMC.

CsoS1D is a BMC-T that was shown to assemble as a dimer of trimers (Klein *et al*, 2009). In this configuration, the 2 CsoS1D trimers superimpose with their face bearing the N and C-termini oriented at the opposite, creating a large tunnel connecting their central pore (figure 6). Klein *et al* have shown that CsoS1D pore, which size is 14Å, could adopt 2 configurations: an open or closed conformations where the open state could allow the entry of large substrates as the RuBP (Klein *et al*, 2009).

Finally, CsoS4A and CsoS4B homologs are pentamers that serve the role of BMC vertices.

CsoS2 is a scaffolding protein, highly disordered on its own (Ni *et al*, 2023), and specific to the α-CBX. It is the third most abundant protein of the α-CBX and its deletion was shown to be deleterious for BMC assembly (Cai *et al*, 2015a). Indeed, upon interactions with the RuBisCO, it adopts a more compact structure and connects the enzymes to the shell subunits, allowing their encapsulation (figure 7) (Cai *et al*, 2015a; Ni *et al*, 2023).

Contrary to the RuBisCO, CsoSCA, a CA that works jointly with the RuBisCO to fixate CO_2, and which is also called CsoS3, attaches directly to the luminal side of the shell (Rae *et al*, 2013).

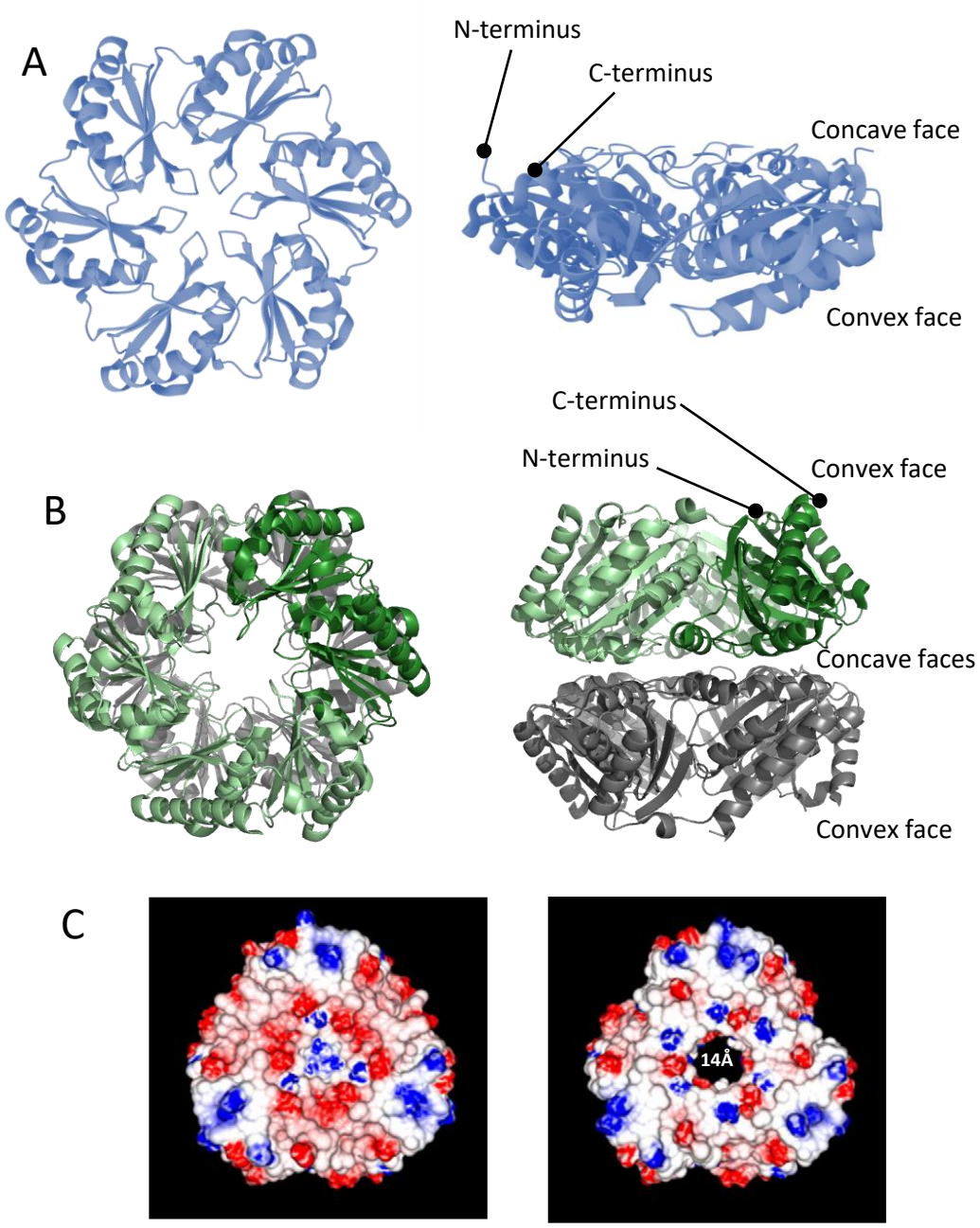


Figure 6. Variety of structures for the shell subunits of the α-carboxysome.
Elucidated 3D structures of the BMC-H CsoS1C (3H8Y) in panel **A** and the circularly-permuted and double-stacked BMC-T CsoS1D (3F56) in panel **B**. **C.** Electrostatic surface representation of CsoS1D convex face (exposed to the bacterial cytosol) upon pore opening. Positively charged regions are in blue while negatively charged regions are in red. Illustration from (Klein *et al,* 2009).

The β-CBX subunits are densely packed

In the same extent than in the α-CBX, a great deal of interplays are in action to ensure the β-CBX formation. Here, a proto-BMC composed of the RuBisCO (4 RbcS plus 4 RbcL subunits), CcaA, CcmM and CcmN, is formed prior to enzyme encapsulation (Long *et al*, 2010).

CcmM has both a CA domain and 3 RbcS-like domains (figure 7). It interacts with the RuBisCO thanks to the RbcS-like domains and was shown to be the trigger of RuBisCO RbcL subunit nucleation (Cameron *et al*, 2013). CcmM can have 2 forms: the full-length CcmM which has a molecular weight of 58kDa and a smaller variant of 35kDa, lacking the N-terminal CA domain and which is more abundant than the 58kDa form (Long *et al*, 2007).

CcmN interacts directly with the CA domain of CcmM via its N-terminus while its C-terminus bears an encapsulation peptide (EP) (see section 3.1.2) that is able to recruit shell subunits such as CcmK2 hexamers to the proto-BMC (Cameron *et al*, 2013; Kinney *et al*, 2012). Incorporation of CcmN is crucial for shell formation and BMC budding out of the polar aggregated materials where the proto-BMC initially forms. Indeed, deletion of *ccmN* was identical to *ccmK2* or *ccmO* (trimeric shell protein) deletion that is stalling BMC biogenesis at the proto-BMC stage. CcmM and CcmN are scaffolding proteins in charge of linking the enzymes to the shell.

Depending on the organism, the *ccm* operon can code for up to 4 different BMC-H homologs (CcmK1 to 4; figure 4B). CcmK1 and 2 are very similar with the unique exception of a C-terminal 8-residue long extension for CcmK1 (Samborska & Kimber, 2012). CcmK3 and CcmK4 are not included in the main *ccm* operon but are part of a distant satellite *locus*. Although individual deletions of *ccmK3* or *ccmK4* did not produce a different phenotype than wild type CBX, combined impairment of their expression was shown to impact the β-CBX dispersion in the cell cytosol, leading to aggregated CBX (Rae *et al*, 2012). Thus, it was proposed that CcmK3/4 could be involved in the spatial arrangement of the BMC, maybe by interacting with the bacterial cytoskeleton.

Likewise, CcmP is coded in another satellite *locus*. CcmP is a BMC-T which, like CsoS1D, associates as a dimer of trimer (Cai *et al*, 2013). The trimer concave sides face each other creating an inner pocket where large molecules can accommodate. Besides, CcmP has a central pore of 13Å which can adopt a closed or open conformations, allowing the passage of large molecules in and out from the CBX. It was shown that closing of the canal could be triggered by the fixation of 3-PGA, the final product of the CBX pathway (Cai *et al*, 2013). Hence, triggered pore opening addresses the paradox of how large molecules might enter the shell without allowing the escape of smaller and/or volatile molecules.

Finally, CcmL is the unique pentameric shell protein encoded by the *ccm* operon. It caps the β-CBX vertices and allows its final closure, ending the CBX biogenesis. This event induces CBX budding.

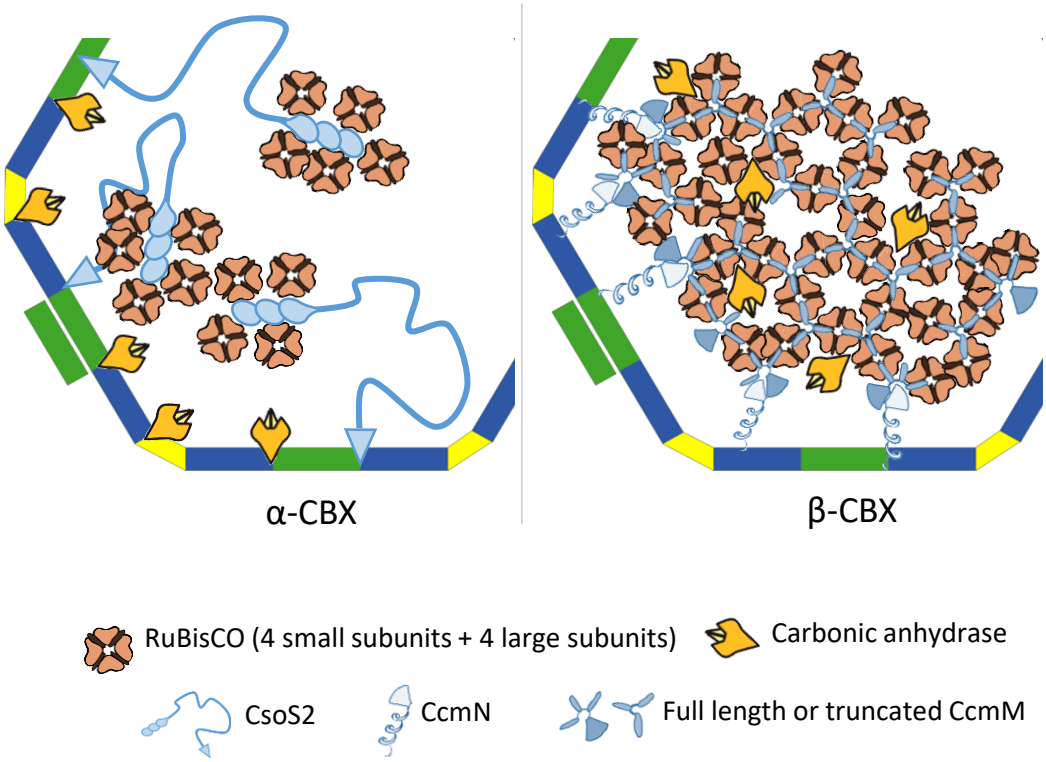


Figure 7. The carboxysome inner organization.
The α- and β-carboxysome (CBX) do not share the same mode of assembly. For the first one, the shell subunits (CsoS1 homologs, CsoS4 and CsoS1D) coalesce. Subsequently, CsoS2, a scaffolding protein, associates with the shell and recruits the RuBisCO enzymatic complex. In parallel, the CsoS3 carbonic anhydrase (CA) binds directly to the inner shell to be encapsulated. In the β-CBX, the cargo proteins aggregate first and form a proto-BMC, mediated by the scaffolding protein CcmM. In particular, CcmM interacts with the RuBisCO by substituting one of the RuBisCO small subunits (RbcS) by its own RbcS domain. Also, CcmM interacts with the CA CcaA. While the truncated form of CcmM induces the proto-BMC formation, full-length CcmM has an extra N-terminal domain that binds to CcmN, another scaffolding protein. The latter contains an encapsulation peptide that protrudes from the proto-BMC and recruits the shell subunits (CcmK homologs, CcmO, CcmP and CcmL). Illustration adapted from (Kerfeld *et al*, 2016).

Of note, capping by CcmL is probably one of the processes controlling CBX shape as elongated proto-BMCs were observed in absence of CcmL (Cameron *et al*, 2013).

2.2. The propanediol utilization BMC

PDU prevalence

The PDU is specialized in 1,2-propanediol (1,2-PD) metabolism, a by-product of rhamnose or fucose sugar fermentation (figure 8A). The PDU primary function is to sequester an intermediate of reaction, the propionaldehyde, which is volatile and can be toxic when accumulated within cells (Sampson & Bobik, 2008).

By transforming the 1,2-PD into 1-propanol or propionate, the PDU provides the cells with an alternative source of carbon for bacterial growth in diverse anaerobic environments such as intestines, sediments or soil depth (Bobik *et al*, 2015). The *pdu* operons are widespread in bacteria, found in both soil-dwelling bacteria and enterobacteria (*Salmonella, Klebsiella, Shigella, Yersinia, Listeria, Lactobacillus, Clostridium* and *Escherichia*; figure 3) (Axen *et al*, 2014; Sutter *et al*, 2021).

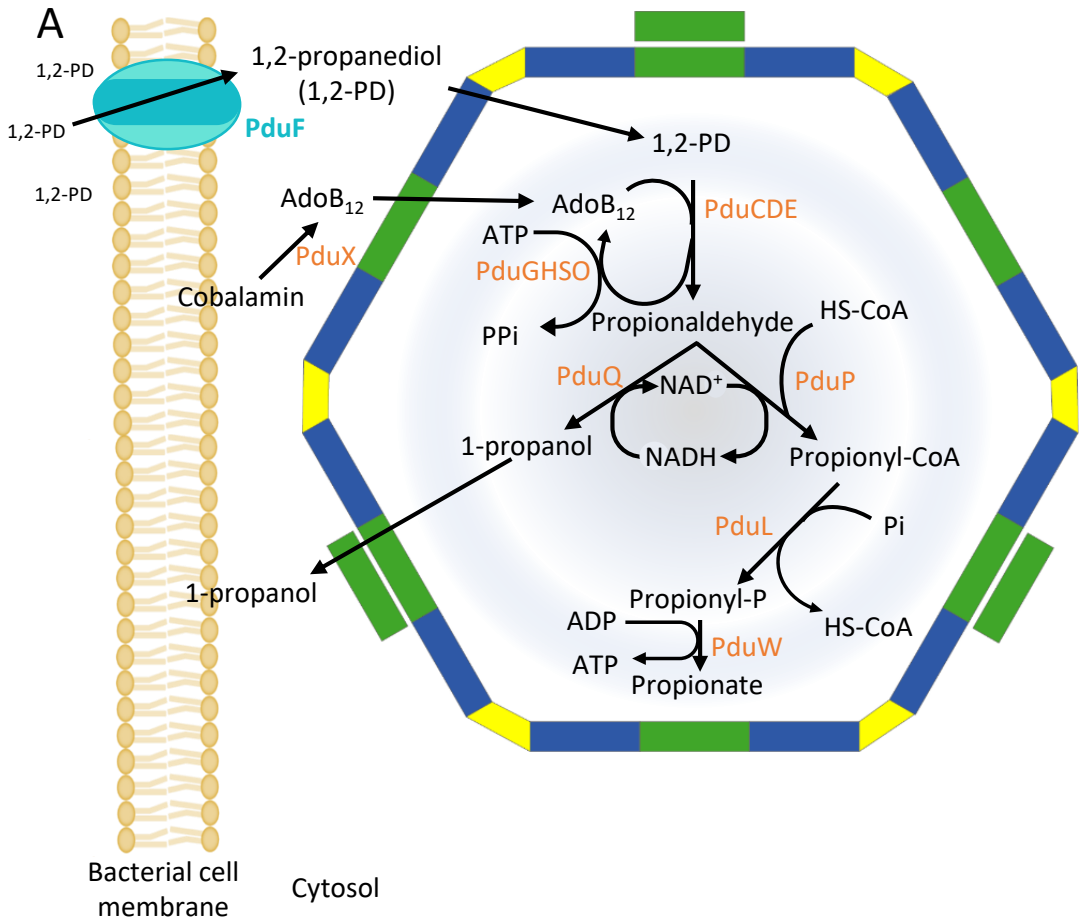
It was sometimes shown to grant a growth competitive advantage to enterobacteria over the BMC-free bacteria although no prevalence for pathogenesis could be drawn. Indeed, both some pathogenic and commensal bacteria code for the *pdu* (Dank *et al*, 2021; Bobik *et al*, 1999).

Metabolism of 1,2-propanediol

PduF is a 1-2-PD diffusion facilitator that is believed to be membrane-bound. Probably PduF is responsible for 1,2-PD uptake from the bacterium microenvironment.

Once the 1,2-PD enters the PDU, it is converted to propionaldehyde by the diol dehydratase complex PduCDE which uses cobalamin as a cofactor (figure 8A). Of note, this step damages the cofactor thus, to ensure a continuity in the PDU functions, the cobalamin is recycled within the BMC in multiple steps by the PduGH, PduS and PduO enzymes. Alternatively, new cytosolic vitamin B_{12} can be synthetized by the cytosolic enzymes from the *cob* operon which is adjacent to the *pdu* and then be processed into cobalamin by PduX before entering the PDU (Chowdhury *et al*, 2014).

Then, propionaldehyde is processed, either through reduction, into 1-propanol by the deshydrogenase PduQ or oxidation and coenzyme A (CoA) transfer by the deshydrogenase PduP to form propionyl-CoA. Both steps require opposite redox potential of NAD+/NADH, creating an equilibrium between NADH consumption and reduction inside the PDU. Finally, propionyl-CoA is phosphorylated by PduL and becomes propionyl-phosphate that is able to leave the BMC.



B

pdu* operon of *Salmonella enterica LT2

pocR pduF pduA pduB pduC/D/E pduG/H pduJ/K pduL pduM pduN pduO/P/Q/S pduT pduU pduV pduW/X

pdu* operon of *Lactobacillus reuteri DSM 20016

pduF eutJ pocR pduA pduB pduC/D/E pduG/H pduK/J pduL pduM pduN pduO/P/Q/W pduU // pduV

pdu* operon of *Listeria monocytogenes EGD-e

pduS pduT pduU pduV // pocR pduA pduB pduC/D/E pduG/H pduK/J pduL pduM pduN // pduO/P/Q pduF pduW // pduX

BMC-T BMC-H BMC-P enzymes assembly auxiliaries
transporters transcription regulators unknown functions

Figure 8. The propanediol utilization BMC.
A. Propanediol degradation in the PDU. CoA: coenzyme A, propionyl-P: propionyl-phosphate. **B.** Genetic organization of operons coding for the PDU (*pdu*) in different PDU bearing organisms. The double slash signifies independent *loci*. Illustrations adapted from (Chowdhury, 2014; Chen, 2017).

A possible last step involves PduW which catalyses the transfer of the phosphate contained in the propionyl-phosphate to an ADP molecule, releasing ATP and propionate (Palacios *et al*, 2003). Propionate, propionyl-CoA and 1-propanol subsequently join the bacterium central metabolism and can serve as carbon sources.

Interconnectivity of the PDU subunits

The *pdu* operon contains close orthologs to CBX shell proteins (figure 8B). Four BMC-H (PduA, J, K and U) are found along with one BMC-P (PduN) and two BMC-T (PduT and PduB).

By expressing recombinantly the shell subunits PduA, B, J, K, N, T and U in *E. coli*, Parsons *et al* described the possible formation of PDU empty shells (Parsons *et al*, 2010a), suggesting that the PDU would follow the α-CBX assembly fashion. Contrasting with these data, another team showed that PDU assembly could be bimodal with equal proportion of cells following the α-CBX or the β-CBX mode (Yang *et al*, 2022)

In this context, PduB and PduM were shown to be crucial for cargo loading (Yang *et al*, 2022; Kennedy *et al*, 2022). Indeed, PduB trimers associate with PduM scaffolding protein that, in turn, interacts with the cargo enzymes (PduD, G, L, O, P and W; figure 9). Of note, PduB has 2 variants: a full-length form and a truncated form, PduB', which lacks the 37-first N-terminal residues due to an alternative translation initiation on the *pdu* polycistronic messenger ribonucleic acid (mRNA). No cargo loading is happening when only PduB' form is present, supporting the idea that PduB N-terminal region iss crucial for the cargo encapsulation.

Alike CsoS1D and CcmP, PduB trimers have 2 possible conformations, open or closed (Pang *et al*, 2012), however they do not form double-stacks. Intriguingly, the closed conformation demonstrated 3 small pores or pockets in which glycerol molecules could accommodate. Glycerol is neither a substrate nor a product of the PDU. Besides, it is very close to 1-propanol. Then, it might be that the real molecule that fixates upon PduB pockets is the 1-propanol. In the same fashion as CcmP central pore opening is controlled through CBX end-product fixation, 1-propanol would induce PduB closed state.

PduM recruits notably PduD that is in complex with PduCE (figure 9). Direct interaction data are still lacking for PduP and PduL. But one could suppose that PduM is also recruiting them to the PDU lumen or, as both cargo proteins bear a N-terminal EP, that they interact directly with shell proteins (Fan *et al*, 2012a; Bradley-Clarke *et al*, 2022). On the contrary, PduG and PduQ encapsulation might go through PduD as in PduD absence, these enzymes were not associated with the PDU but rather cytosolic (Yang *et al*, 2022).

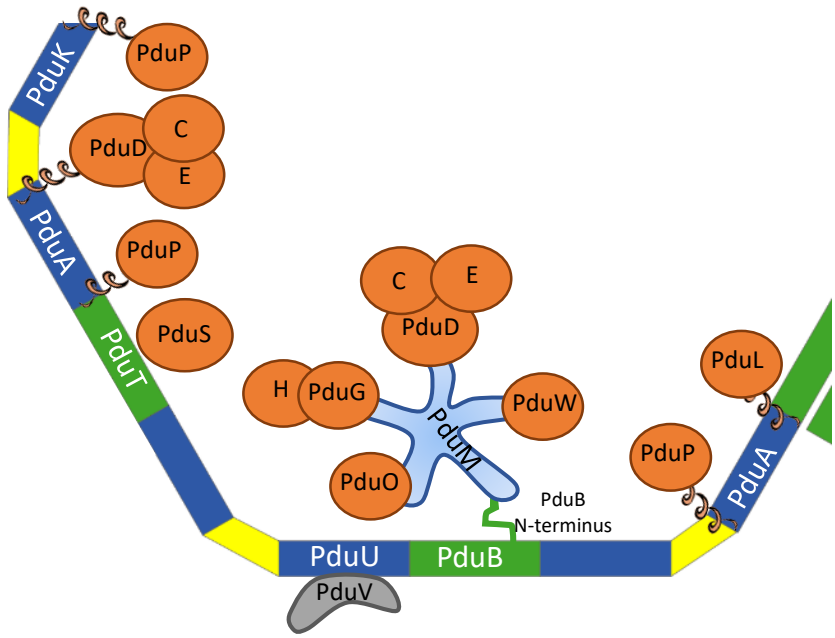


Figure 9. Proposed model for the propanediol utilization BMC inner organization. In the propanediol utilization BMC (PDU), a concomitant shell and cargo nucleation seems to occur. Cargo enzymes (in orange) can either be encapsulated through interaction of their encapsulation peptide with the shell subunits, in particular the PduA BMC-H; through complex formation with their enzymatic partners like PduC and E that are loaded by the intermediary of PduD; or be loaded thanks to the scaffolding protein PduM. Of note, PduM binds specifically to PduB N-terminal domain (37 residues).

PduA, PduJ and PduB are the most abundant proteins of the PDU. PduA was predicted to be an hub protein, totalizing 11 protein-protein interactions (PPI) with cargo enzymes as well as with other shell proteins (Jorda *et al*, 2015; Trettel *et al*, 2022). As we just saw, PduB is also implicated in cargo loading (Kennedy *et al*, 2022). PduA and PduB always occur as the first translated proteins from the main *pdu* operon. Recently, it has been shown that protein order within operon is crucial for complex assembly (Bertolini *et al*, 2021; Shieh *et al*, 2015a). Furthermore, Chowdhury *et al* demonstrated that, when *pdua* is deleted from *Salmonella enterica* genome, aberrant PDUs formed, although PduJ, its closest homolog (80% sequence identity), was present (Chowdhury *et al*, 2016). Surprisingly, this phenotype could be rescued by expressing *pduj* from *pdua* chromosomal *locus*. Thus, it seemed very likely that protein coding order in BMC operon depicts their importance in BMC biogenesis, *i.e.* the proteins in operon pole position would be the centre of BMC nucleation.

PduT forms a trimer whose pore is blocked by a characteristic [4Fe-4S] cluster, bound on the 3 Cys38 of the trimer (Pang *et al*, 2011). Although its exact functions are still to be uncovered, PduT was proposed to be involved in redox reactions inside the PDU, mediated by its [4Fe-4S] cluster. Also, PduT was shown to co-purify with PduS (Parsons *et al*, 2010b), pointing at potential extra contacts between the shell and the cargo enzymes.

In addition to PduT interplays, PduK is able to interact with the PduP enzyme, especially its N-terminal 18 residues which constitute an EP.

PduU is a circularly-permuted BMC-H which means that the order of the secondary structure elements in the protein is modified. This leads notably to the permutation of its N- and C-termini orientation from the concave to the convex face. Besides, it contains a N-terminal extension which coalesces with contiguous PduU N-termini to form a central β-barrel within the hexamer (Crowley *et al*, 2008). Of note, PduU central pore is occluded by this particular quaternary structure element. Recently, an interaction was evidenced by yeast two-hybrid (Y2H) between PduV and PduU (Jorda *et al*, 2015). Despite having a N-terminal EP supposed to promote its encapsulation within the BMC lumen like cargo enzymes, PduV would associate to the outer shell (Parsons *et al*, 2010a). Only few data are available on PduV exact functions. Notably, PduV has a faint GTPase activity and was shown to associate with filamentous structures resembling the cell cytoskeleton.

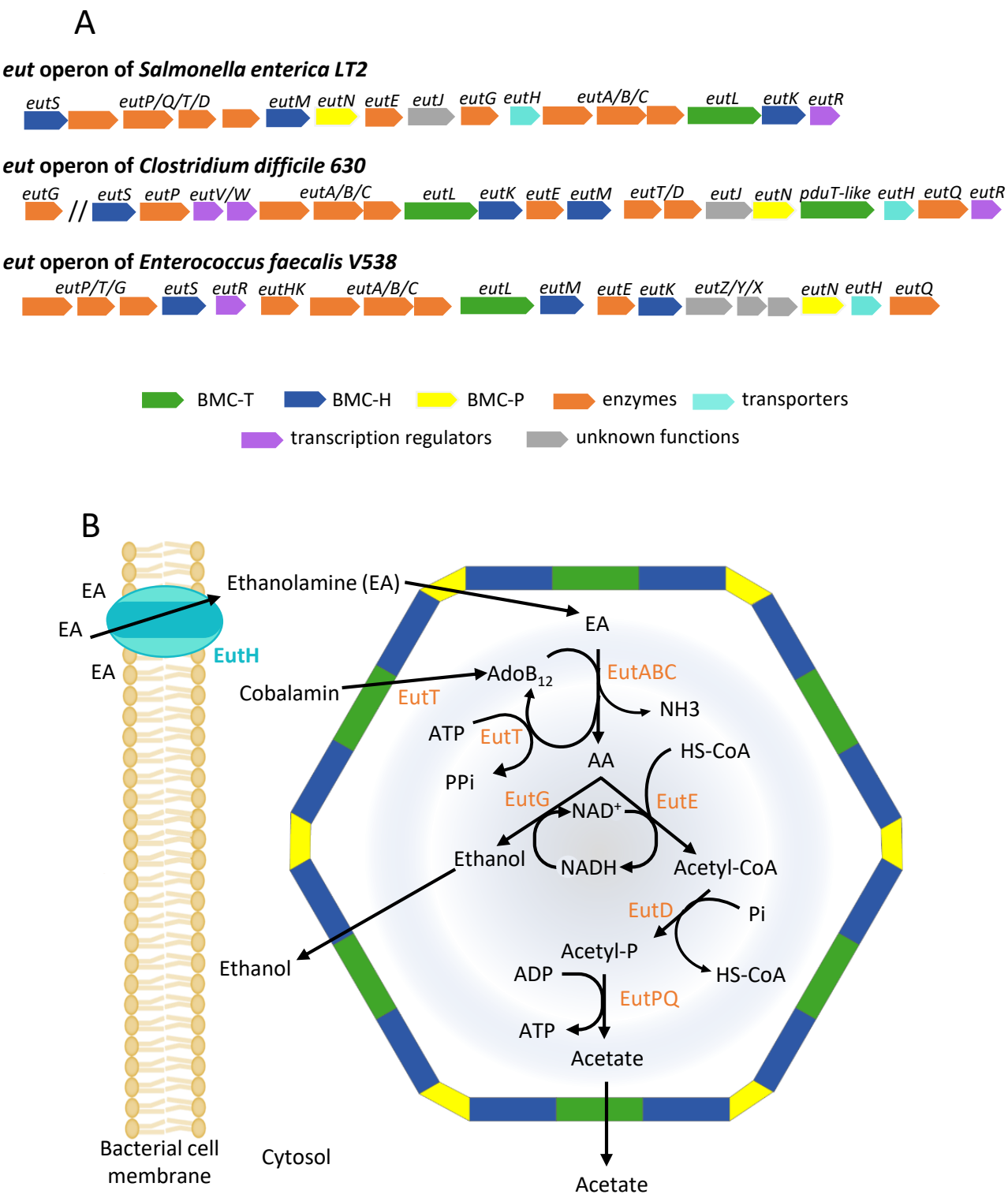


Figure 10. The ethanolamine utilization BMC.
A. Genetic organization of operons coding for the EUT (*eut*) in different EUT-bearing organisms. The double slash signifies independent *loci*. (Pitts, 2012; Del Papa, 2008; Chowdhury, 2014) **B.** Ethanolamine degradation in the EUT. $adoB_{12}$: 5'-deoxyadenosyl B_{12}, CoA: coenzyme A, acetyl-P: acetyl-phosphate.

2.3. The Ethanolamine utilization BMC

EUT prevalence

The EUT is involved in ethanolamine (EA) degradation into ethanol or acetate and to do so, it requires cobalamin as cofactor. EA originates from phosphatidylethanolamine breakdown which is the main components of cell membranes. For instance, EA-rich environments include mammalian guts where EA is released from dead epithelial or microbial cells or derives from the host diet. EA utilization constitutes a growth advantage as it can be both carbon and nitrogen sole sources (Chang & Chang, 1975). The *eut* operon is mainly found in actinobacteria, proteobacteria and firmicutes (*Mycobacterium, Klebsiella, Enterococcus, Salmonella, Clostridum, E. coli*; figure 3 & 10A) (Axen *et al*, 2014; Sutter *et al*, 2021).

During EA processing, an acetaldehyde (AA) intermediate is produced (figure 10B). Penrod *et al* showed that mutations in shell proteins affected shell integrity that resulted in a major AA leakage (Penrod & Roth, 2006). This loss of carbon negatively impacted bacterial growth. Besides, another team observed *Salmonella* cells that were unable to grow on EA or 1,2-PD when mutated for the DNA polymerase I and suggested that DNA polymerase I repair functions were needed in order to counteract aldehyde toxicity (Rondon *et al*, 1995). Then, the EUT shell is a protective barrier that sequesters the reaction and prevents carbon loss and cellular damages that might be induced by AA.

Co-occurrence with the pdu operon

Among the organisms that bear several BMC operons in their genome, the *eut* and the *pdu* operons are the ones that most often co-occur (Sutter *et al*, 2021). For instance, *Salmonella, Klebsiella* and some strains of *E. coli* possess both operons. However, expression of the 2 BMC types are finely tuned. While the *eut* is under the control of the transcription factor EutR (Roof & Roth, 1992), the *pdu* expression is positively controlled by PocR (Bobik *et al*, 1992).

PocR is generally coded upstream the *pdu* operon, in a reverse transcription orientation (figure 8B). Its expression is induced by both the 1,2-PD and vitamin B_{12}. Besides inducing the *pdu* transcription, PocR enhances its own expression in a feedback loop (Bobik *et al*, 1992). On the contrary, EutR is most often the last protein encoded by the *eut* (figure 10A). It is expressed at a week basal level from its selfish constitutive promoter (Roof & Roth, 1992). In presence of EA and vitamin B_{12}, EutR is activated and promotes the *eut* main promoter expression which also promotes its own expression.

In *Salmonella enterica*, the *eut* and *pdu* operon preclude one another (Sturms *et al*, 2015). The *eut* is repressed in presence of 1,2-PD which shows a preference for growth on 1,2-PD. This repression, which is mediated by PocR regulation factor, is critical because when the PDU was produced with a

concomitant expression of shell proteins EutL or EutS, hybrid BMCs were subsequently assembled with disrupted metabolic functions (Sturms *et al*, 2015).

All *eut*-bearing organisms do not share the same EUT regulation. In addition to EutR control, *Enterococcus faecalis* as well as *Clostridium* and *Listeria* species have a double-regulation mechanism controlled by EA and vitamin B_{12} presence (Fox *et al*, 2009). Indeed, EA was shown to induce EutW auto-phosphorylation which in turn phosphorylates EutV. EutV is thought to be retrieved from *eut* mRNA hairpin structures upon phosphorylation, preventing premature transcription termination. In the second mechanism, the vitamin B_{12} binds specific 3D structures on the *eut* mRNA, upstream *eutG* sequence notably. This fixation would induce a conformation change of the mRNA which also prevents premature transcription termination.

In the same extent, PDU-endowed bacteria can have alternative regulation such as in *Listeria monocytogenes* where a RNA antisens of *pocR* sequence was shown to be produced and to repress PocR translation (Mellin *et al*, 2013). However, this antisens RNA transcription is reduced upon vitamin B_{12} addition. Indeed, vitamin B_{12} would bind to a riboswitch present at the beginning of the antisens RNA and induce its premature termination.

Contrasting with the *pdu* domination over the *eut*, another team showed, also in *Salmonella enterica*, that the *eut* or *pdu* main promoters could be induced in the concomitant presence of EA and 1,2-PD along with vitamin B_{12} (Jakobson *et al*, 2015). Furthermore, Delmas *et al* recently evidenced that *eut* and *pdu* polycistronic mRNA were simultaneously expressed in *E. coli LF82* strain cultured in a medium containing bile salts (source of both EA and 1,2-PD) (Delmas *et al*, 2019). Then it might be possible that *eut* expression prevails over the *pdu* or that both BMCs co-exist, depending on the organism.

Metabolism of ethanolamine

EA entry within the cell is thought to be mediated by EutH which possesses 11 transmembrane domains (Kofoid *et al*, 1999). It is then addressed to the EUT where the EutBC complex, also named EA ammonia lyase, transforms it to AA. This step requires cobalamin as cofactor and releases ammonia (figure 10B). Likewise in the PDU, the cobalamin can be recycled *in situ* by EutA (reactivates EutBC by evicting inactive B_{12}) and EutT (transfers an ATP on B_{12} to reactivate it into cobalamin) or be provided by *de novo* synthesis in anaerobic conditions exclusively, by proteins encoded in the *cob* operon.

Subsequently, EutE processes AA to acetyl-CoA, producing a molecule of NADH. Acetyl-CoA can join the cytosol and the tricarboxylic acid cycle or the glyoxalate shunt or either be phosphorylated by

EutD to give an acetyl-phosphate. In the cytosol, housekeeping acetate kinase AckA can further transform it to acetate that will be excreted. In this process, a molecule of ATP is produced.

Although EutP and EutQ functions have not been precisely determined yet, some data point to the fact that they might act together to play a role similar to AckA (Moore & Escalante-Semerena, 2016). Also, data on EutP and EutQ encapsulation within the EUT are still missing.

In parallel, NAD^+/NADH balance is maintained by AA reduction into ethanol which consumes NADH and produces NAD^+.

The EUT subunit connections are overlooked

EutM is the main shell component of the EUT. Its 3D structure has recently been determined by X-ray crystallography with other EUT shell proteins (Tanaka *et al*, 2010). It is a 97-residue long protein with a canonical BMC-H domain pfam00936. It assembles as a flat hexamer and bears a central positively charged pore of approximately 8Å, suggesting that the pore allows the passage of small negatively charged molecules inside the EUT.

Unlike the PDU and CBX, little is known about the PPIs that leads to EUT biogenesis. However, as EutM shares a high homology to PduA/J or CcmK1/2, and is the more abundant protein of the EUT, one can supposed that it also shares their role as critical hub protein for the EUT biogenesis. However, a crystallographic attempt showed that no mixed EutM/L crystal could be obtained (Takenoya *et al*, 2010). But as crystal organization does not mimic natural cytosol constraints and that some shell proteins were shown to have a bending angle rather than being flat, maybe observations of the association between bent and flat protein oligomers is made impossible in crystal.

EutS has also a single pfam00936 domain but surprisingly, this domain is circularly-permuted compared to other BMC-H. The 6 protomers form a central β-barrel, similarly to PduU. Moreover, contrary to flat EutM, one EutS homolog has a distorted hexagonal shape and a 40° bending (Tanaka *et al*, 2010). This angle of curvature was dictated by a particular residues, the Gly39. Indeed, the Gly39Val mutation completely abrogated EutS bending and flat hexamers were obtained. This residues is conserved in many EutS homologs but it differs in other BMC-H among which PduU, the closest structural homolog.

Surprisingly, while EutK contains a canonical BMC-H domain, it remained monomeric in solution (Tanaka *et al*, 2010). To date the full 3D structure of EutK has not been determined; only the C-terminal 60-residue long extension was elucidated as helix-turn-helix motif, typical of DNA binding proteins. Furthermore, EutK extra domain bears a large patch of positively charged residues supporting a potential interaction with negatively charged DNA molecules. Lack of EutK self-assembly might indicates that EutK is incorporating itself within mixed hexamers.

In former studies, EutN quaternary structure was determined as hexameric although EutN is a BMC-P, suggesting it would be pentameric and act as the EUT vertices (Forouhar *et al*, 2007). In this hexamer, a large pore of 24Å was present which is considerable in size and would allow the passage of very large molecules. However, a more recent report determined that it was rather pentameric in solution (Wheatley *et al*, 2013). The hexameric state reported for EutN could either be an artefact due to crystallization conditions, a minor EutN form or indicate an atypical function for this BMC-P homolog.

There is only one BMC-T coded in the *eut* operon, EutL (Kofoid *et al*, 1999). Its 3D structure has been elucidated and demonstrated 2 possible conformations which mainly differ in pore opening (Tanaka *et al*, 2010). While the open conformation showed a 10-12Å pore, ordered loops from each EutL subunit occluded the pore in the closed form. Besides, the closed form had 3 small pockets of 1,3Å, one on each EutL, to which EA could bind specifically (Thompson *et al*, 2015). Low packing density around the pockets suggested that they provided space for conformational rearrangement that led to EutL open form. Thus, upon binding, EA would prevent EutL pore opening by steric hindrance, maybe to impede small and large molecules to escape the EUT while high EA concentration, *i.e.* high EUT metabolism is reached.

2.4. The glycyl radical enzyme-associated BMC

GRM prevalence

The GRMs are a wide BMC family which catabolise diverse metabolites thanks to glycyl radical enzymes (GRE). To be functional, these enzymes require a post-translational modification that generates an enzyme-bound glycyl free radical (Gly$^{\cdot}$) which allow them to carry out radical-based chemistry in anoxic environments. Such modification is performed by the GRE-activating enzyme which utilises a [4Fe-4S] cluster to create radical species on its substrate proteins.

Up to date, 6 different GRM subtypes were identified (Ferlez *et al*, 2019a). The GRM1 and 2 are involved in choline degradation while the GRM3, 4 and 6 have a function analogous to the PDU (but B_{12}-independent). The last subtype is the GRM5 which shares the same enzyme set as the GRM3, 4 and 6 but possesses additional enzymes, *i.e*. a fuculose-phosphate aldolase and a lactaldehyde reductase. This would enable it to process fuculose- or rhamnulose-phosphate, 2 by-products of complex polysaccharide degradation (Ferlez *et al*, 2019a). Here, we will focus on the choline-degradative GRM subtypes.

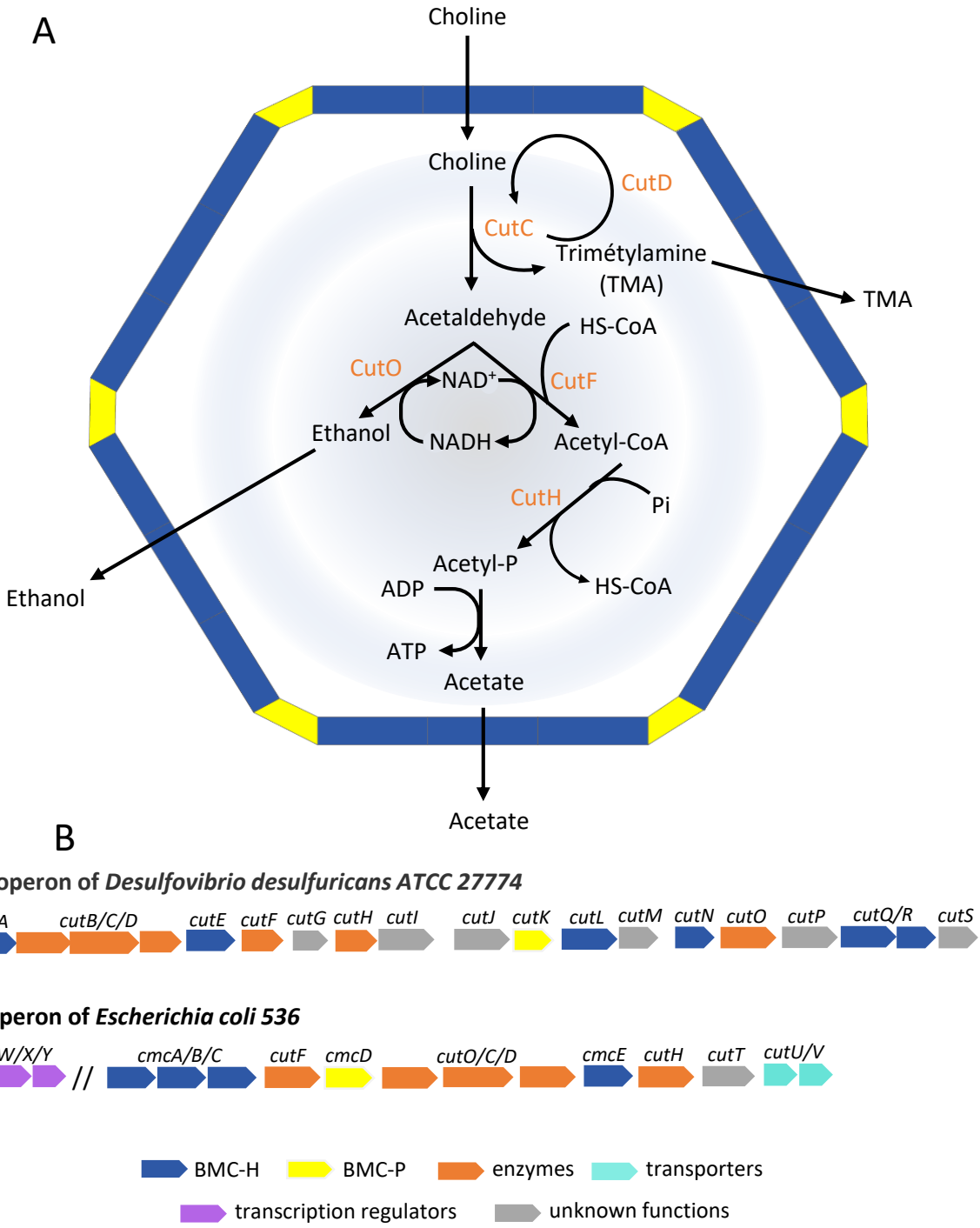


Figure 11. The glycyl-radical enzyme containing BMC.
A. Choline degradation in the GRM1 or 2. CoA: coenzyme A, acetyl-P: acetyl-phosphate. **B.** Genetic organization of operons coding for the GRM (*cut*) in different GRM-bearing organisms. The double slash signifies independent *loci*. (Martinez-del Campo, 2015; Herring, 2018)

According to genomic survey, the GRM1 (*choline utilization* type 1 operon) is found in various bacterial phyla: actinobacteria, proteobacteria, firmicutes, fusobacterial (figure 3) (Sutter *et al*, 2021; Zarzycki *et al*, 2015). On the contrary, the GRM2 (*cut2*) is restricted to pathogenic gammaproteobacteria such as *E. coli, Klebsiella* or *Raoultella ornithinolytica* (Zarzycki *et al*, 2015). Organisms able to process choline mostly live in anaerobic niches such as the human gut and urinary tract. Indeed, the *cut* operons are overexpressed only in anoxic environments, in presence of choline, suggesting a possible inhibition by oxygen (Herring *et al*, 2018). Moreover, GREs were shown to be very sensitive to oxygen as exposure to oxygen induced a polypeptidic chain cleavage on the residue on which the radical was located (Wagner *et al*, 1992). Then, besides sequestering AA intermediate to avoid toxicity and carbon loss, the GRM1/2 shell might play a role in protecting the GREs from oxygen inactivation.

Metabolism of choline

Choline is released from membrane phospholipids following mammal or bacterial cell breakdown. Upon entry within the GRM, choline is cleaved into trimethylamine (TMA) and AA by the TMA lyase CutC, which has been preliminary activated by CutD (figure 11A) (Craciun & Balskus, 2012). TMA is not used as a nitrogen source but is excreted by the cells thanks to CutUV efflux pumps. Then, AA is either process by CutF into acetyl-CoA or by CutO into ethanol. These steps require NAD^+ or NADH, respectively, which balances the luminal NAD^+ consumption and recycling. While ethanol is egressed to the cytosol, the acetyl-CoA is phosphorylated by the phosphotransacylase CutH, giving rise to acetyl-phosphate that can exit the GRM and serves as a carbon source.

First inner-organization details of the GRM

While extended studies were performed on the CBX and PDU subunit inner organization, only sparse data are available for the choline-degrading GRMs. This is mainly due to the fact that interests in GRM enzymatic functions and structure are only recently emerging (Craciun & Balskus, 2012; Zarzycki *et al*, 2015; Kalnins *et al*, 2020).

Operon analyses showed that GRMs were coding for several BMC-H (from 4 to 6) and only 1 BMC-P (figure 11B) (Zarzycki *et al*, 2015; Martínez-del Campo *et al*, 2015). Surprisingly, the *cut loci* are practically deprived of BMC-T coding sequence. Indeed, this subunit is absent from the GRM2 while only few GRM1-endowed organisms encode for BMC-T such as *Clostridium* and *Streptococcus* (Zarzycki *et al*, 2015; Kalnins *et al*, 2020).

Structures have already been resolved for some of the BMC-H shell proteins. CmcA, CmcB, CmcC from the GRM2 and CutN from the GRM1 are canonical BMC-H assembling as relatively flat hexamers

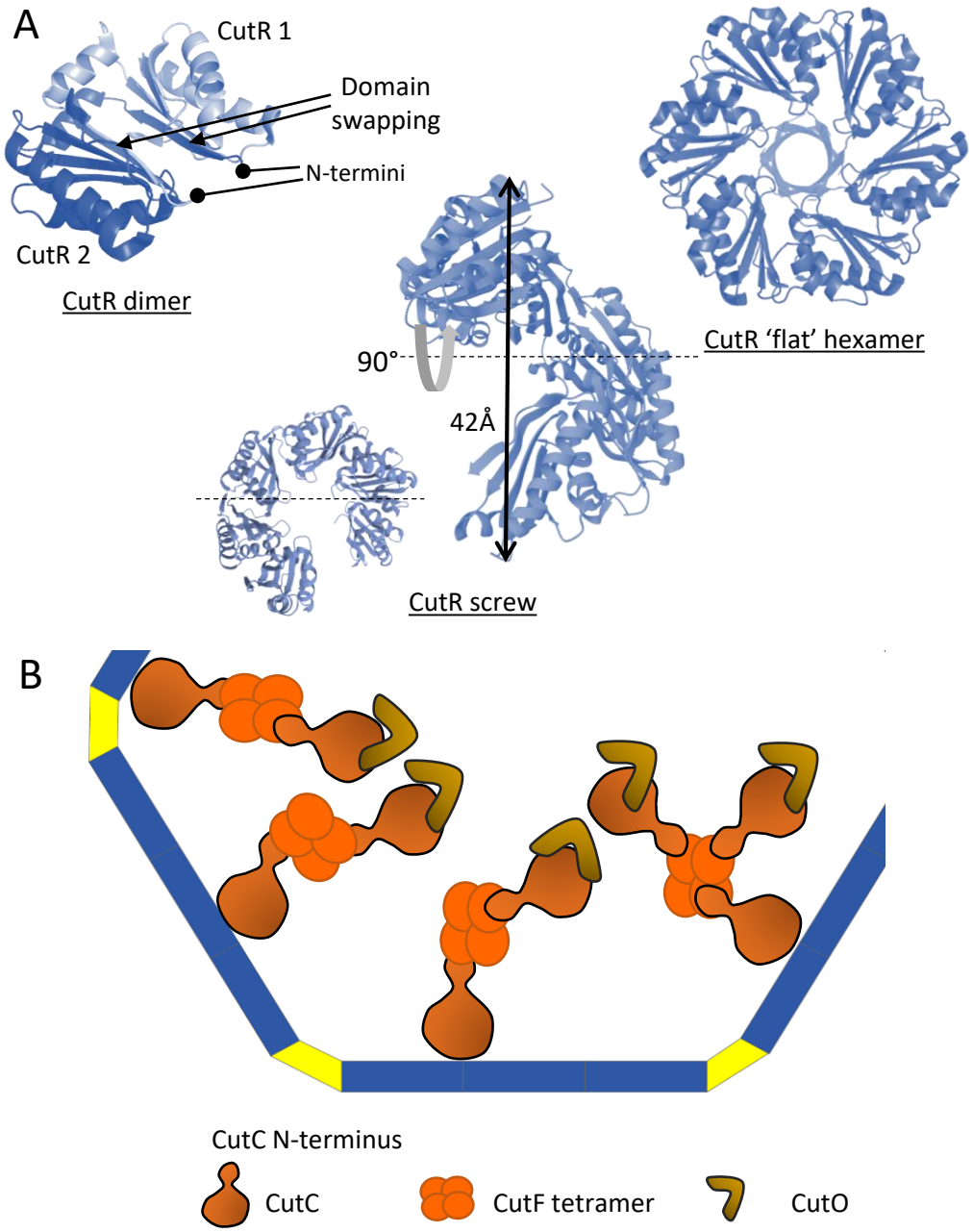


Figure 12. Structure and organization of the GRM subunits.
A. Alternative 3D structures of the symmetry-breaker BMC-H CutR from the GRM1. CutR is a circularly-permuted BMC-H that can form different oligomers: a dimer (6XPH), a relatively 'flat' hexamer (6XPI) with a central β-barrel, resembling PduU and EutS hexamer or a helical hexamer (6XPK) with a screw pitch of 42Å. **B.** Proposed model of the GRM2 inner organization. Besides its enzymatic functions, CutC would serve as an adapter for other enzyme encapsulation, in particular CutF tetramer with which it interacts through its N-terminus or CutO through its C-terminal region. Adapted from (Kalnins *et al*, 2020).

(Ochoa *et al*, 2021). CmcA, B and C have 78% sequence identity whereas CutN only share 52% of identity with them. This highlights the divergence that exists between both choline-degrading GRMs.

On the contrary, CutR from *Streptococcus intermedius,* is a circularly-permuted BMC-H which had an enigmatic behaviour upon crystallization (Ochoa *et al*, 2020). CutR demonstrated several conformations, notably a CutR dimer which performed domain swapping between N-termini or a screw-shaped hexamer with a screw pitch of 42Å (figure 12A). CutR was shown to undergo a disulphide bond between Cys37 and Cys73. When the Cys37 was mutated to Ala, the screw shape was abolished and CutR changed its conformation for a planar hexamer with a central β-barrel, similar to PduU and EutS. While CutR dimeric form seemed artifactual (crystallized from purified fraction corresponding to an hexamer weight), the screwed conformation was relevant and might indicate a particular function in shell architecture.

The only elucidated 3D structure for a BMC-P is the one of CmcD from the *Klebsiella pneumonia* GRM2 which shows a classical pentamer with un unusual hydrophobic central pore (Kalnins *et al*, 2020). In the same study, Kalnins *et al* have resolved the GRM2 shell architecture (see section 4.1) along with interactions governing cargo protein encapsulation. CmcE is a BMC-H that has a C-terminal extension of 40 residues but which structure has not been resolved yet. Inclusion of CmcE in the set of shell proteins recombinantly expressed to produced minimal GRM2 resulted in larger BMCs (Kalnins *et al*, 2020). However, CmcE was not necessary for cargo protein loading, suggesting that CmcE role is restricted to controlling shell architecture.

Although CutF and CutH contain an EP-like sequence (see section 3.1.2), cargo protein loading was proposed to occur through interaction with CutC (Kalnins *et al*, 2020). CutC is a large enzyme of approximately 1500 residues. It is partially disordered up until choline binding (Kalnins *et al*, 2015). It has a 340 residue-long N-terminal extension, homologous to the subsequent 340 residues. CutC was the unique cargo enzyme capable of being encapsulated within the shell by its own and it did so independently of the presence of its N-terminal extension (Kalnins *et al*, 2020). Rather, this extension was shown to interact with CutF and mediate its encapsulation (figure 12B). As CutF is a tetramer, CutF could represent a centre for other CutC to nucleate, increasing the shell capacity of CutC loading. While CutC also mediated CutO loading, no CutH could be observed in the recovered BMCs. Possible explanation is that CutH was encapsulated in a too small amount to be observable or that CutH could not be taken in charge in these minimal GRM2 because it normally interacts with CmcE through its EP (CmcE was not present) or that CutH is normally cytosolic.

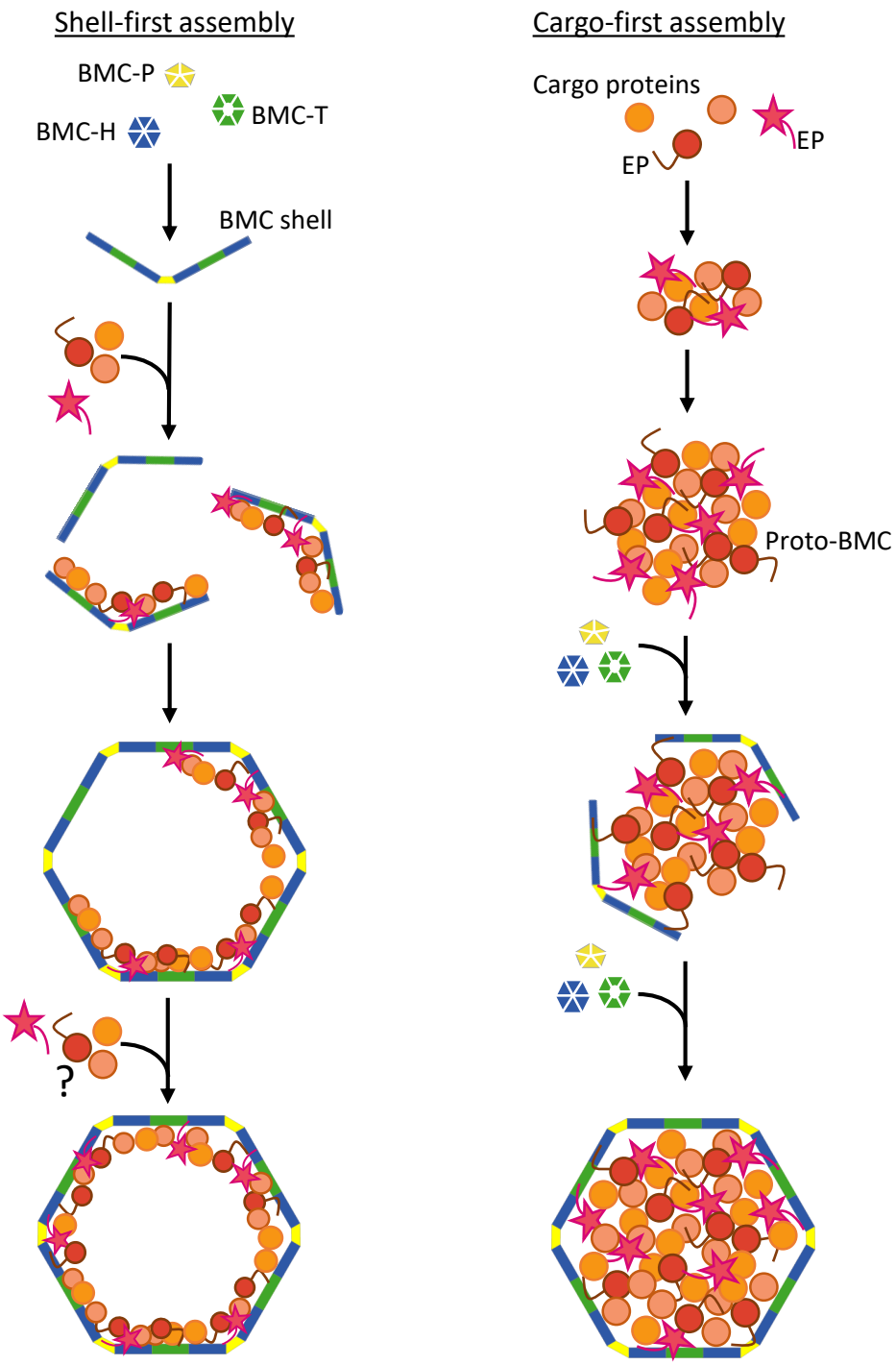


Figure 13. Proposed BMC assembly modes.
In the shell-first mode, the different subunits composing the BMC shell nucleate first to form shell fragments onto which cargo proteins bind prior to shell closing. Alternatively, it was proposed that a complete and closed shell assembles and that cargo proteins penetrate this empty shell, in a second step, although no mechanism permitting to explain how large proteins could enter the BMC had been elucidated yet. **In the cargo-first mode,** cargo proteins form a dense core called a proto-BMC, presumably thanks to the aggregating properties of the encapsulation peptide (EP) they bear and direct protein-protein interactions. The shell subunits are recruited to this proto-BMC through interaction with the EPs of the cargo and encapsulate it before budding from the cell poles where the material aggregated. Illustration adapted from (Kerfeld *et al*, 2015).

3. From biogenesis to BMC end

3.1. BMC assembly

3.1.1. Two modes of BMC assembly

BMC biogenesis is a critical step for BMC functions as it was demonstrated that disrupted BMCs failed to fixate CO_2 or to metabolize EA or 1,2-PD, impairing bacterial growth. Also, overproduction of individual shell proteins led to altered BMC shape (Parsons *et al*, 2008). Conserving a certain ratio of each protein present in the BMC is very important to ensure a functional BMC.

Yet, the mode of BMC assembly has not been completely unravelled. Still, some studies brought some clues and two distinct mechanisms seem to exist : the α-CBX-like and the β-CBX-like assemblies.

The BMC shell as a scaffold for cargo loading

In *Halothiobacillus neapolitanus*, a chemoautotrophic model organims for the α-CBX study, electron cryo-tomographies revealed partially assembled shells along with RubisCO complexes (Iancu *et al*, 2010). This suggested that a co-assembly was occurring for shell and cargo proteins in the α-CBX. In this mode of assembly, the cargo proteins were lining up against the inner shell, leaving α-CBX lumen partially empty (Shively *et al*, 1973; Iancu *et al*, 2010) contrary to β-CBX that depicted a dense core (Kaneko *et al*, 2006). Furthermore, empty α-CBX could form in absence of cargo proteins (Shively *et al*, 1973). This mode of BMC biogenesis will be referred to as the shell-first assembly (figure 13).

Priority to the core coalescence

Regarding the second mode of BMC assembly, it follows the β-CBX scheme where the enzymatic set forms a proto-BMC at the cell pole, highly packed and ordered, before shell subunits encapsulate them (Cameron *et al*, 2013). In this assembly mode, the shell proteins are recruited to the proto-BMC thanks to a small peptide on cargo proteins, the encapsulation peptide. Subsequent budding from aggregated material gives rise to a complete BMC. Then, this is a cargo-first BMC assembly (figure13). Of note, this peptide is absent in the subunits of the α-CBX, corroborating that another mechanism is in action. Indeed, a proto-β-CBX was clearly visible in transmission electron microscopy (TEM) observations of *Synechococcus elongatus 7942* (Kinney *et al*, 2012). This proto-CBX was exclusively composed of the RuBisCO, the carbonic anhydrase CcaA and CcmM. It packed with a polyhedral geometry which might dictate BMC shape.

3.1.2. The encapsulation peptide

A short helicoidal sequence to mediate cargo encapsulation

In order to recruit BMC shell subunits to the enzymatic core, bacteria have evolved a particular signal sequence called an encapsulation peptide (EP). This peptide is generally 18-residue long and can be found either on the N- or C-terminus of cargo proteins. For instance, EP were observed on EutC and EutG N-termini (Fan *et al*, 2010) or on CcmN C-terminus (Aussignargues *et al*, 2015). Surprisingly, no EP could be detected on any of the α-CBX cargo proteins, hinting at an assembly mechanism that diverges from other BMCs.

Although EP does not show any residue sequence conservation, it maintains a peculiar alternation between hydrophobic and polar residues (typically 2 hydrophobic residues followed by 2 polar, repeated at least 2 times). Its 3D structure was elucidated by nuclear magnetic resonance as an α-helix on which polar residues are distributed on one side while hydrophobic residues localize to the opposite side (Lawrence *et al*, 2014).

The EP is connected to the cargo protein through a linker whose length might range from 1 residue for some members of the phosphotransacylase family to up to 277 residues for CcmN (Aussignargues *et al*, 2015). The different linkers found in encapsulated proteins do not share a sequence homology nor any conserved residues. The only characteristic they have in common is a high content in hydrophilic residues.

The PDU has emerged as the BMC model for the study of EP. Indeed, multiple cargo enzymes were shown to have an EP like PduP, PduD and PduL (Fan *et al*, 2012a; Lawrence *et al*, 2014; Bradley-Clarke *et al*, 2022). PduE was predicted to bear a N-terminal EP, yet the enzyme could not be encapsulated in PDU shell on its own. Rather, its encapsulation was mediated by the formation of the diol dehydratase complex along with PduC and PduD, the latter having an EP that targeted PduE and PduC to the PDU (Fan & Bobik, 2011). While it is still unclear whether or not PduS bears an EP, it was shown to interact via its N-terminus with the shell protein PduT (Parsons *et al*, 2010b). Besides PduT, cargo enzymes rely on PduB for BMC encapsulation (Kennedy *et al*, 2022). Indeed, empty shells were seen in a PduB deleted strain.

EP binding onto the shell hexameric subunits

In *Salmonella enterica*, PduP was shown to associate with PduA or J through interactions between its EP and the C-terminal small α-helix of the shell proteins (Fan *et al*, 2012). To decipher such association, Jorda *et al* modelled diverse PDU cargo protein EPs binding onto PduA hexamer (Jorda *et*

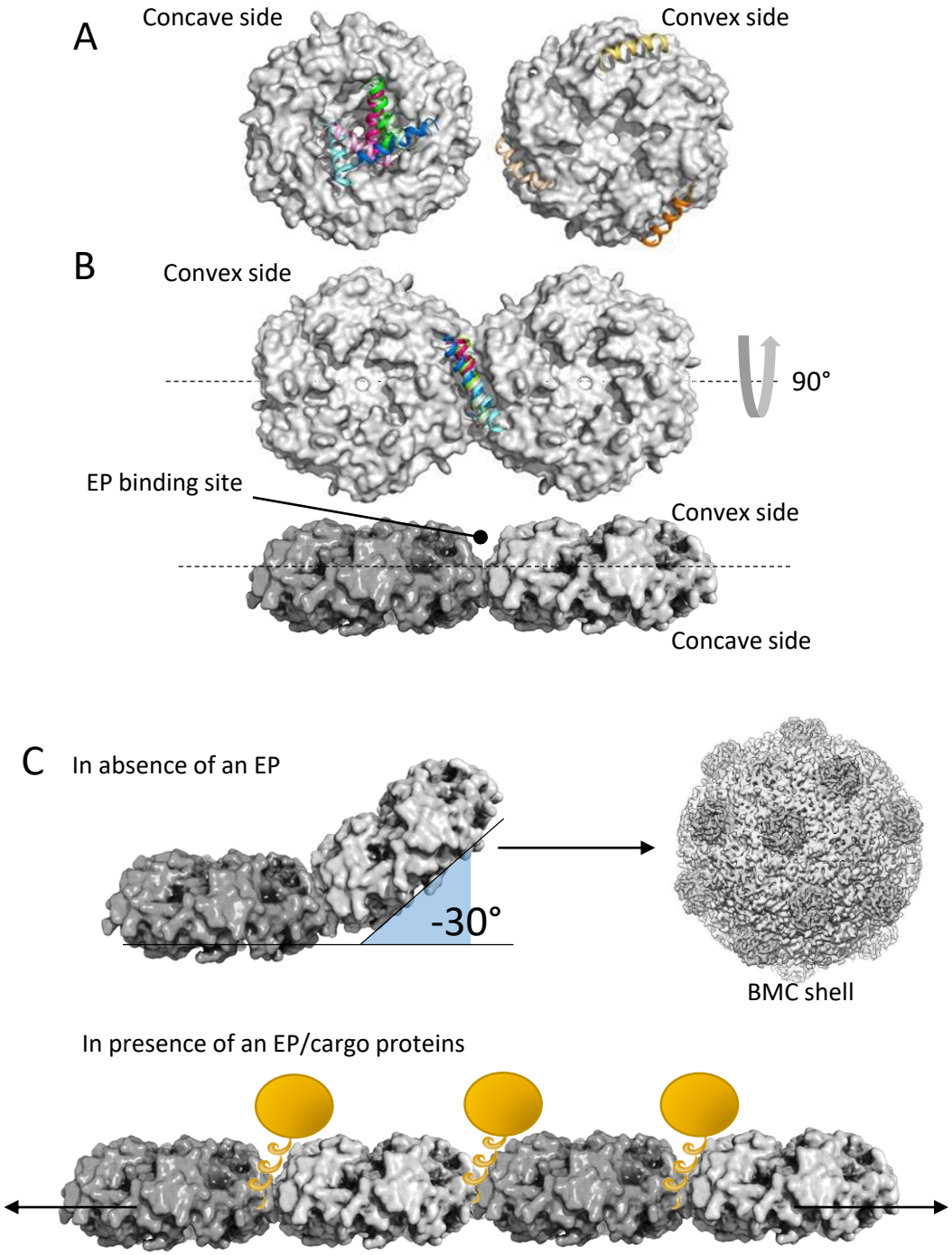


Figure 14. Binding of the encapsulation peptide (EP) onto tessellating hexamers. Modelling of PduL (in pink), PduP (in green) and PduD (in blue) EPs docking onto individual PduA hexamer concave or convex faces (**A**) or onto tessellating hexamers (**B**). All EPs (in orange) coalesce on the same groove present in between hexamer convex sides. Illustration adapted from (Bradley-Clarke *et al*, 2022). **C.** Proposed mechanism of EP control over shell curvature. In absence of an EP (thus, of extra cargo proteins to encapsulate), the groove in between shell subunits remains free, allowing bending of the shell and subsequent closure. On the contrary, EP binding onto tessellating hexamers sterically hinders curvature which induces shell expansion and loading of more cargo proteins. The BMC shell 3D map was taken from (Sutter *et al*, 2017) and the adjacent hexamers from (Uddin *et al*, 2018).

al, 2015). Their model seemed to indicate that the EP was interacting with a particular cleft of the hexamer concave face (figure 14A).

Multiple whole-BMC 3D shell structures have been newly elucidated and these structures all point at the fact that shell subunit concave faces are oriented toward the exterior of BMCs (see section 4.1) (Tan *et al*, 2021; Kalnins *et al*, 2020; Sutter *et al*, 2017; Greber *et al*, 2019; Ni *et al*, 2023). If the binding model was correct, the cargo proteins would be associated with the BMC outer shell and not within the lumen. Yet, when the enhanced green fluorescent protein (eGFP) was fused to EutC EP to be targeted to EUT, it was protected from anti-GFP immunoblotting whereas it could be tagged in disrupted shells (Choudhary *et al*, 2012), demonstrating that the eGFP was encapsulated within the BMC and not associated with the outer shell.

Another study might provide us with a better understanding on EP association with the BMC shell (Bradley-Clarke *et al*, 2022). Both modelled and experimental data highlighted a greater propensity of different EPs from the PDU to bind to the convex face of tessellated PduA hexamers (figure 14B). In this context, EPs localized to the groove formed by adjacent hexamers, blocking bending of the hexamer interface. Indeed, upon EP binding, PduA which normally formed nanotubes, were assembling as flat sheets. Together, this demonstrated that EPs play a crucial role in shell formation: while recruiting shell subunits, they also dictate shell size and shape. Hence, a low EP-tagged cargo filling of BMC would induce premature bending and closure of the shell, resulting in smaller BMC what can be seen in recombinant empty shell (Juodeikis *et al*, 2020; Kennedy *et al*, 2022). On the contrary, high EP-tagged content would promote extended shell facet formation and bigger BMCs (figure 14C).

Of note, EPs were also shown to induce cargo protein coalescence in absence of the shell subunits (Juodeikis *et al*, 2020). This phenomenon, also referred to as liquid-liquid phase separation (LLPS), was proposed to be at the origin of BMC following the cargo-first assembly mode (*i.e*. formation of a densely packed proto-BMC) (Zang *et al*, 2021).

Knowledge of the EP biology and role is of foremost importance for synthetic biology as it could enable us to target heterologous proteins to the BMC lumen. Some attempts have already been made in this direction. Thanks to PduP and PduD N-terminal EPs, the pathway for ethanol production (pyruvate decarboxylase and alcohol dehydrogenase) was successfully addressed to recombinant PDU (Lawrence *et al*, 2014). Encapsulated within the PDU, the catalytic efficiency was up to 10-fold higher than those of cytosolic enzymes.

3.2. BMC spatial control and final degradation

BMCs are large structures that can have a diameter of 40 to 600nm and weight as much as a gigadalton. They comprise several hundreds to thousands of proteins. For instance, in the α-CBX, an estimation of around 5000 CcmK protomers and 250 RuBisCO were present (Iancu *et al*, 2007) and there is an average of 3,7 CBXs per cell (Savage *et al*, 2010). In comparison, around 7600 BMC-H were recorded per PDU shell, along with 2000 cargo proteins (Yang *et al*, 2020).

These structures grant a selective advantage to bacteria (Rowley *et al*, 2018; Delmas *et al*, 2019) or sustain their whole metabolism, like the CBX which furnishes the fundamental brick (CO_2) for carbon compound biosynthesis. Indeed, the cyanobacteria *Synechococcus elongatus 7942* cannot grow if depleted in β-CBX (Savage *et al*, 2010), highlighting the importance of BMC maintenance within the bacterium.

3.2.1. Cytoskeleton and nucleoid involvment in BMC maintenance

During their lifetime, BMCs appeared to be taken in charge by the bacterial cytoskeleton which tightly controls BMC localization and homogeneous repartition inside the cell (Savage *et al*, 2010; Parsons *et al*, 2010a). Indeed, proteins normally associated to the cytoskeleton like ParA or MreB were shown to ensure that BMCs are equally passed on to the daughter cells during division as evidenced by *parA* or *mreB* deletions that disrupted CBX distribution (Savage *et al*, 2010). Besides, a *parA* deletion led to random segregation at cell division. Some daughter cells did not receive any CBX and temporarily lost their ability to fixate CO_2 and perform photosynthesis (Savage *et al*, 2010; Hill *et al*, 2020).

Recently, MacCready *et al* determined that CBX localization within cells was controlled by a ParA-like ATPase protein, McdA (Maintenance of CBX Distribution A), in collaboration with McdB (MacCready *et al*, 2018, 2020, 2021). Briefly, McdA binds to double-stranded DNA, presumably the nucleoid, non-specifically while McdB would interact with the shell proteins CcmK2-4, CcmO or CcmL (MacCready, 2018). When McdA and McdB enter in contact, McdB induces McdA ATPase activity which releases McdA from the bacterium nucleoid. Then, it binds DNA back in regions with a lower McdB concentration. On the contrary, McdB follows McdA gradient across the cell, drawing with it the CBX it is anchored to. In this way, CBXs are taken in charge by the McdA/B system from the cell poles where they bud to be distributed evenly along the cell longitudinal axis.

Thus, equal repartition of the BMCs ensures equal partitioning between daughter cell upon cell division. The *mcdA/B* genes generally cluster near *cso* or *ccm* operons or satellite *loci*. But the McdA/B system is not exclusive of CBX-coding organisms. These proteins were also found in close proximity to

pdu, *eut* or *cut* operons (MacCready *et al*, 2021). In *mcdA/B* depleted strains, CBXs remain as polar aggregates showing that McdA/B system is also very important in CBX and more generally BMC budding (MacCready *et al*, 2018).

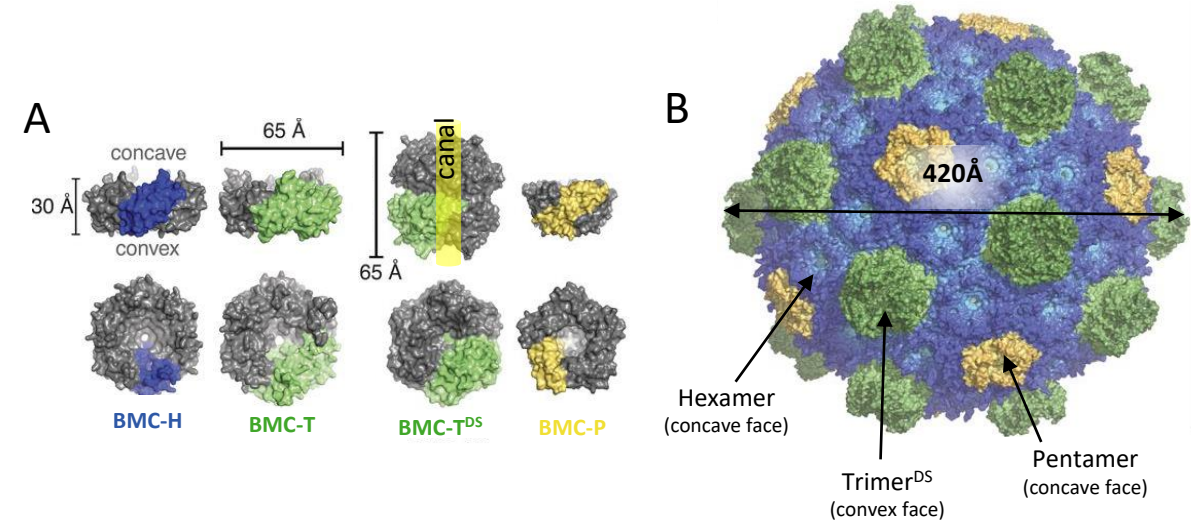
To recall, EutK, a BMC-H from the EUT, has an C-terminal extension, rich in positively-charged residues and which adopts a fold similar to nucleic acid-binding proteins (Tanaka *et al*, 2010). In light of the activity of McdA/B system, one could suppose that EutK extension would bind to the bacterial nucleoid while its C-terminal part (the BMC-H domain) would incorporate the shell. Of note, EutK is not oligomerizing with itself in solution (Tanaka *et al*, 2010) but it could be plausible that EutK is involved in hetero-hexamers with other BMC-H from the EUT. EutK would mimic the McdA/B system and act in synergy to control the EUT spatial organization within the cytosol

In the PDU, PduV, a protein with unknown functions, was shown to be associated to the outer shell and transiently associated with filamentous structures resembling the cytoskeleton (Parsons *et al*, 2010a). Also, PduV presence in the PDU is linked to BMC dynamics within the cell. PduV shares a high sequence similarity to the Ras-like GTPase protein family and it was determined that PduV could have a small GTPase activity (Parsons *et al*, 2010a). Data suggest that PduV might interact with the bacterium cytoskeleton to control BMC dynamics and that energy to do so might be provided by its GTPase activity.

Moreover, it was recently shown that PduK depleted cells had PDU distribution issues (Yang *et al*, 2022). Indeed, PDUs were not budding from the cell poles and remained with aggregated materials. Yet, it was not crucial for PDU assembly. PduK is one of the PDU shell proteins. It has a C-terminal extension compared to other canonical BMC-H such as CcmK2 or EutM, which functions have not been determined. One could assume that such extension is homologous to EutK C-terminal extension and that it might be involved in PDU distribution and segregation. Yet, no PduK structure is available impeding to go further on the assumptions.

3.2.2. BMC degradation

In the last stage of BMC life, BMCs gradually lose their metabolic functions upon multiple daughter cell generations. Indeed, thanks to an engineered *Synechococcus 7002* strain harbouring a single β-CBX, Hill *et al* tracked the same CBX over time and found that CO_2 fixation, *i.e.* cyanobacteria growth, stopped after several cell divisions (Hill *et al*, 2020). In this aging CBX population, GFP-fused RbcL was shown to return to the cell poles. Subsequently, GFP fluorescence disappeared indicating that both shell and RuBisCO had been degraded.



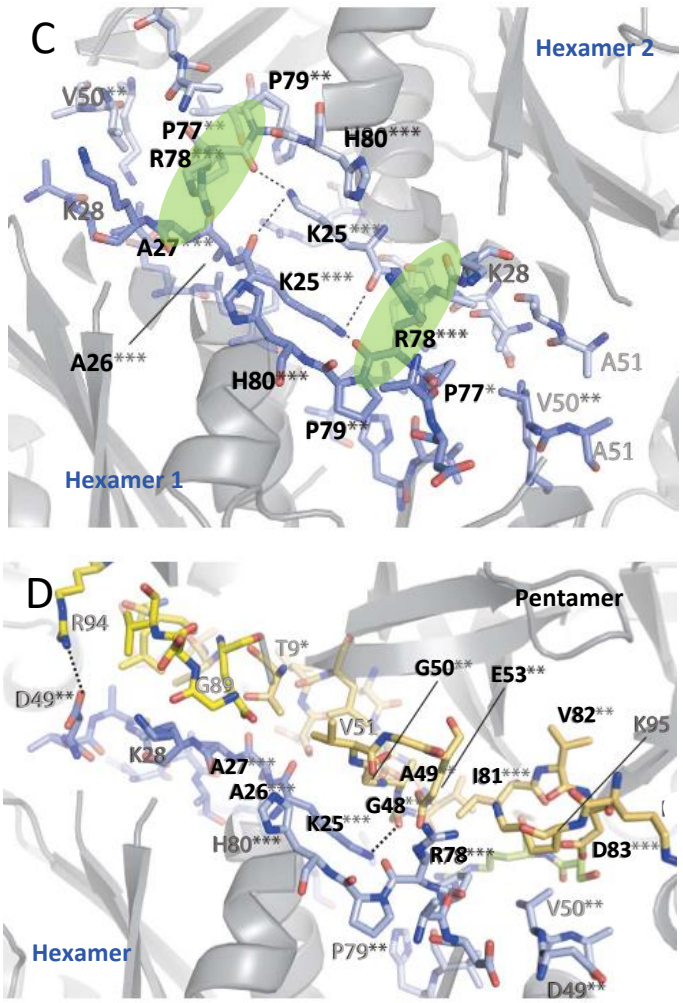


Figure 15. BMC shell architecture.
A. Shell subunit general topology. DS: double stacked. **B.** Whole-BMC shell 3D structure resolved by cryo-EM and X-ray crystallography. **C.** Residues involved in the stability of hexamer-hexamer interface. The KAA (residues 25-27 and PRPH motifs (residues 77-80) make hydrogen bonds while the R78 protrudes and contacts a negatively charged residue present on a pocket formed by the adjacent hexamer, which clamp hexamers together. A better view of this clamping is provided in the panel A of the figure 16. **D.** Residues involved in the hexamer-pentamer interface stabilization. The pentamer is pictured in yellow and hexamer in blue. The pentamer IVD motif stabilizes the R78 while the GAGxGE motif makes hydrogen bonds with the KAA motif. Illustrations were taken from (Sutter *et al*, 2017).

Then, BMCs are not long-lasting structures but rather dynamic protein structures that are built in the cell poles and bud from protein aggregates. Recently, it was proposed that BMC assembly could occur through a LLPS scheme where BMC subunits condensate through EP coalescence and separate from the rest of the cytosol (Oltrogge *et al*, 2020; Zang *et al*, 2021; Kumar & Sinha, 2022). In this scheme, aggregated proteins would remain well-folded and active if they were enzymes.

Specific protein systems are involved in its dynamic across the cell, systems which identities might depend on BMC type and might implicate different cellular entities such as the cytoskeleton or the nucleoid. Globally, these entities take in charge BMCs as early as their birth, control their even distribution along the cytosol throughout their life and control their return to budding site for the final degradation.

4. Interactions governing the shell assembly

4.1. The BMC shell architecture

Although BMC assembly modes are relatively well studied for diverse BMC types, very little was known about how the different structural subunits organize in the shell. Basically, up to now, only hexamer/hexamer interactions had been extensively examined (Sutter *et al*, 2016; Faulkner *et al*, 2019). But, very recently, thanks to the combination of individual shell subunit crystallographic structures and cryo-electron microscopy (cryo-EM) resolution enhancement, solving of whole-BMC 3D structures was made possible (Sutter *et al*, 2017, 2019; Greber *et al*, 2019; Kalnins *et al*, 2020; Ni *et al*, 2023). In an attempt to design BMCs with a restricted set of shell subunits, diverse minimal BMCs were studied and the structures of a minimal HO BMC, an α- and β-CBX and a GRM2 were determined.

General features of the shell architecture

All BMC structural subunits (BMC-P/T/H) have 2 distinct faces (figures 15A & 6A): one concave (hollow face) and one convex (domed face). The subunit C- and N-termini are always present on the same face and usually localize on the concave face with some exceptions like circularly-permuted BMC-T and -H. Indeed, in circular permutants, the protein termini are switched due to translocation of two secondary structure elements of the pfam00936 domain from the C- to the N-terminus. Then, termini are localized to the convex face. BMC-T circular permutants include notably EutL, PduB, CsoS1D, HO BMC-T2, -T3 and CcmP (figure 5). CsoS1D and CcmP as well as HO BMC-T2 and T3 associate as double-stacked trimers that form an inner chamber or tunnel, linking superimposed trimer pores (figures 15A & 5). BMC-H permutants like PduU, CutR and EutS (figure 5) share a peculiar topology. On the convex side, their N-termini form a protruding β-barrel at the 6-fold symmetry axis (Crowley *et al*, 2008; Ochoa

et al, 2020; Pitts *et al*, 2012). While PduV was proposed to interact with PduU hexamer through binding onto this protruding β-barrel, the functions of such protrusion remain unclear and more data are required (Jorda *et al*, 2015).

BMC-H are the main subunits of the shell with a ratio of 129 BMC-H:30 BMC-T:1 BMC-P per PDU (Yang *et al*, 2020), depicting their importance for shell architecture. Despite some evidences that BMC-H such as CcmK2 could self-assemble as a double layer (concave-to-concave face stacking) (Samborska & Kimber, 2012), whole-BMC cryo-EM structures revealed that the shell was made of a single layer of structural proteins that included double-stacked trimers occasionally (figure 15B) (Sutter *et al*, 2017; Greber *et al*, 2019). In these configurations, all subunits concave faces were oriented toward the exterior and this feature was conserved among the four BMC types studied. Again, the only exception was for double-stacked trimers for which concave faces were facing up and one convex face was oriented toward the lumen while the second was outward (Greber *et al*, 2019).

Pentamers which are the less abundant subunits, have a truncated pyramidal shape (figure 15A). At the shell vertices, pentamers were always surrounded exclusively by hexamers. Although BMC-T/BMC-P interactions are not impossible in principle, such assemblies were not observed in cryo-EM suggesting that interactions with the hexamers are preferred in both cases.

Rather than icosahedral with clear facets and edges as observed in TEM (Iancu *et al*, 2010; Shively *et al*, 1973), the shell had a round shape. This might be explained by the absence of an enzymatic cargo which seemed to be important for shell size and geometry. Indeed, loading the α-CBX with a cargo GFP increased its size from 217 to 247Å and switched its T = 3 symmetry to a T = 4, *i.e.* respectively 3 or 4 proteins in an asymmetric unit that repeats x times in space to create a full shell (Tan *et al*, 2021). While 12 pentamers and 30 hexamers were found in a T = 4 symmetric shell, only 12 pentamers and 20 hexamers were present for the T = 3 symmetry. Thus, cargo loading also changed BMC-H/BMC-P ratio and would probably affect shell subunit repartition and interactions.

BMC-H are the privileged interactants of shell subunits

Canonical BMC-H, like CsoS1 or CcmK homologs and HO BMC-H (figure 5), were found to be the vital link between all shell subunits. They are able to make contacts with every subunits, including themselves. The hexamer peripheral interface contains many patches of hydrophobic residues, as do the pentamer and trimer interfaces. Interactions between the hexamers and the other subunits were proposed to be mediated by shape complementarity of these patches (Sutter *et al*, 2017; Tan *et al*, 2021) rather than specific residues interactions.

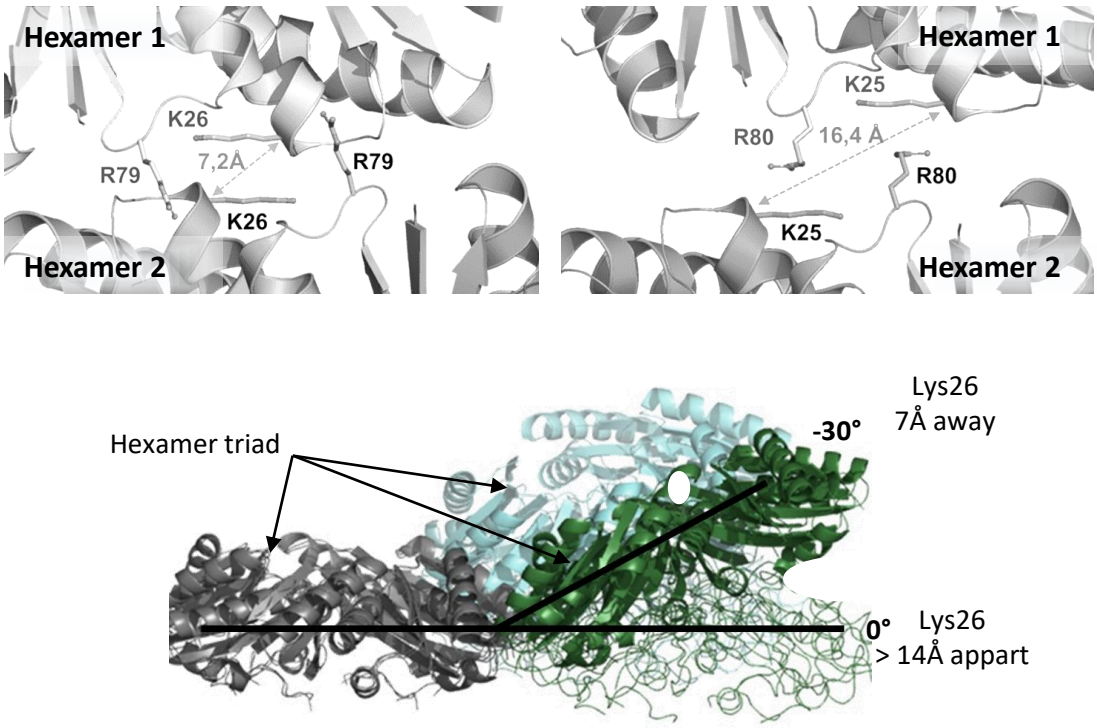


Figure 16. Hexamer facet binding dictated by the distance between Lys25.
Depending on the BMC-H homolog, conserved antiparallel K25 adopt different behaviour in the hexamer-hexamer interface. While the K26 α carbons are very close in PduA (7Å away), allowing for proper clamping of the interface by the R79 which performs electrostatic bonds with the E83 of juxtaposed hexamer, the K25 of CcmK1 or 2 are further away. The proximity of conserved antiparallel Lys (7-8Å) induces a -30° bending angle between hexamers. More distant Lys have no curvature preference. Illustration from (Garcia-Alles *et al*, 2023).

Shape complementary would allow interfaces with different peripheral residue content but sharing the same shape to interact with the same hexamer interface or in other words, this would make the hexamer peripheral interface promiscuous to any shell subunit providing it has a specific shape. This was notably highlighted in HO BMC where the three trimers (composed of BMC-T1, T2 or T3), despite considerable sequence divergence, could occupy a similar position alongside the unique BMC-H of the HO operon (Sutter *et al*, 2017).

However, it seems that some residues played a role in clamping subunit interfaces together, thus stabilizing their assembly. Among these residues are the KAA (residues 25-27) and PRPH (residues 77-80) motifs on BMC-H (figure 15C). These motifs are highly conserved among BMC-H paralogs hinting at their crucial role.

The Lys25 makes hydrogen bonds with the conserved GAGxGE motif on BMC-P while the Arg78 and the BMC-P IVD motif are involved in a salt bridge (figure 15D) (Sutter *et al*, 2017). In the hexamer/hexamer interface, the Lys25 (Lys29 in CsoS1A) of each subunit are facing each other in antiparallel while the Arg78 (Arg83 in CsoS1A, Arg80 in CcmK) localize on both sides of the Lys and clamp the interface through multiple hydrogen bonds (figure 15C) (Sutter *et al*, 2017; Kalnins *et al*, 2020; Tan *et al*, 2021).

Control of the shell curvature by conserved antiparallel Lys

Recently, our team has shown that distance between juxtaposed and highly conserved Lys25 (Lys26 in the study) was dictating hexamer interface curvature (Garcia-Alles *et al*, 2023). When the α-carbons of the Lys25 were very close (7 to 8Å), hexamer triad had a -30° bending angle (figure 16). Of note, this negative angle predicted that concave faces of the hexamers would point outward which is in agreement with resolved shell 3D structures. In this configuration, the Arg78 (Arg79 in the study) could localize in the small pocket present on the opposite hexamer and interact with Glu83 to stabilize the interface. On the contrary, when the Lys25 were more distant (14 to 17Å), hexamer triad had no specific curvature trend and the Arg78 loosed their interacting partners, making them more mobile. Thus, data suggested that besides having a role in shell subunit interface stabilization, Lys25 and Arg78 were involved in shell curvature.

The hexamer/trimer interfaces

No BMC-T is coded in the *cut2 locus*, hence no information could be drawn from the GRM2 shell structure. Unfortunately, in both the α- and β-CBX BMC, CcmO or double-stacked CsoS1D could not be detected in the crystallographic unit, suggesting that they had not been integrated into the shells

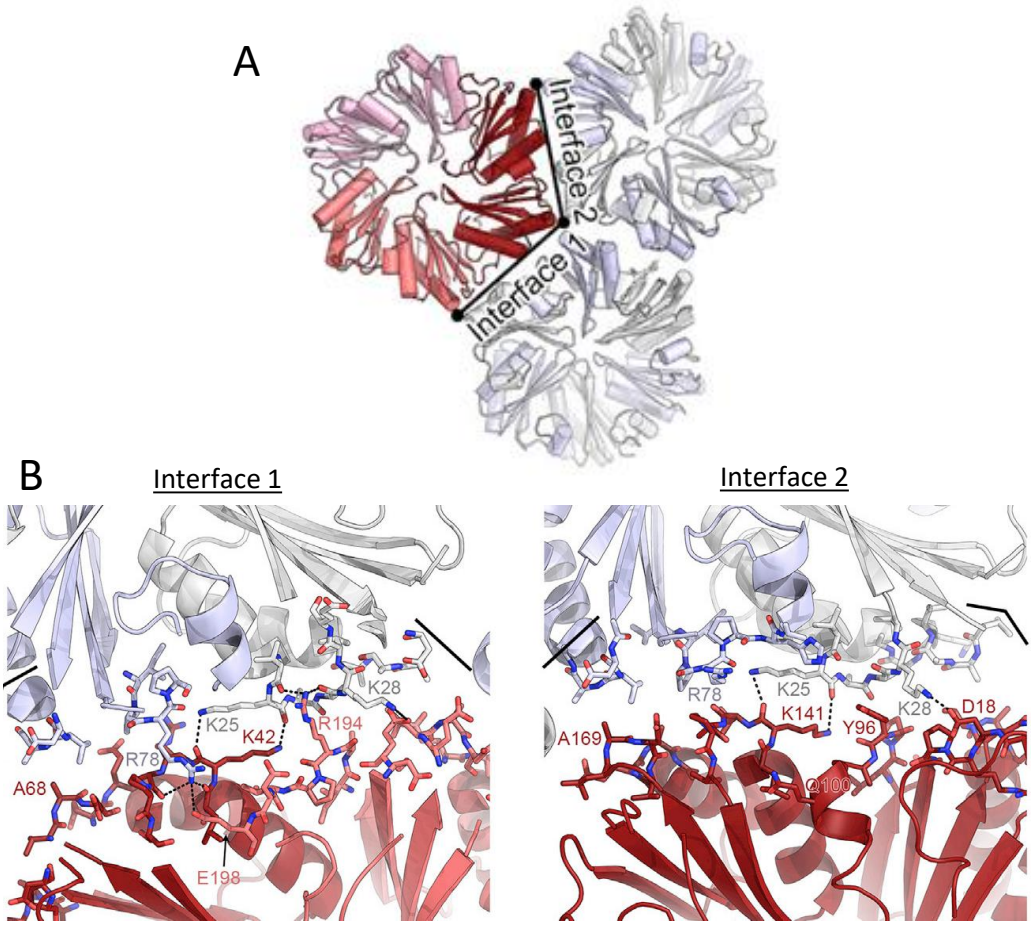


Figure 17. The trimer-hexamer interface.

A. The hexamer, in grey, contacts 2 pfam00936 domains either localized on 2 different BMC-T (interface 1), in red, or on the same BMC-T (interface 2). **B.** Antiparallel K25 and 42 or 141 are involved in hydrogen bonding, stabilizing the interface. In the first interface, the R78 of the hexamer protrudes towards the negatively charged E198 or equivalent R194 of the trimer contacts the opposite E83. By contrast, in the interface 2, the hexamer R78 does not stand out to reach opposite D100. Illustrations from (Greber *et al*, 2019).

(Sutter *et al*, 2019; Tan *et al*, 2021). Yet, trimers were successfully observed in the HO BMC shell, allowing to determine the residues or motifs involved in hexamer/trimer interactions.

As BMC-T are a fusion of 2 pfam00936 domains, their interactions with an hexamer could occur through 2 different interfaces (figure 17A). In these interfaces, the Lys of the hexamer and trimer (Lys25 and Lys42 or Lys141, respectively) orient in antiparallel, stabilizing the interaction (figure 17B) (Greber *et al*, 2019). Besides, in the first interface, the trimer Glu198 is performing hydrogen bonds and a salt bridge with the hexamer Arg78, notably, while in the second interface, the side chain of the Arg78 is further away from the trimer which impairs any possible bond with the trimer Asp100 that occupies the position equivalent to Glu198 on the second pfam00936 domain.

Small-chained residue at the subunit 3-fold axis to ensure assembly

Another conserved feature in the shell subunit interfaces is the obligated presence of a residue with a small lateral chain at the 3-fold axis where subunits meet. The Ala68 and Ala169 were found at the trimer corners (Greber *et al*, 2019) or Ala51 for the hexamer in HO BMC shell (Sutter *et al*, 2017). This place is occupied by the Ser51 in CcmK1 and 2 and Gly51 for Cmc homologs (Sutter *et al*, 2019; Kalnins *et al*, 2020). These small residues are crucial for assembly as a bulkier residue would create a steric hindrance and impede shell assembly.

Although numerous structures are available and allowed to determine the global architecture of the BMC shell, more data are still necessary, notably, high resolution studies to examine the exact role of each subunit homolog in the shell architecture. Indeed, as nowadays techniques rely on averaging acquisition to increase overall image resolution, there is an information loss for very similar proteins like HO BMC-T2 and T3 (Greber *et al*, 2019) or the Cmc homologs (Kalnins *et al*, 2020).

4.2. BMC-H property to self-assemble

When overexpressed in *E. coli*, BMC-H of different BMC types were observed to form higher-order macrostructures. This characteristic makes them of great interest as potential scaffolds to be engineered for synthetic biology. PduA from *Citrobacter freundii* typically assembles as nanotubes that extend within the bacterial cytosol and sometimes impair cell septation (Pang *et al*, 2014). These structures were also observed for RMM, the unique BMC-H of the AAU of *Mycolicibacterium smegmatis* (Noël *et al*, 2016).

In TEM, nanotubes appear as densely packed, long and hollow filaments of 18-20nm in diameter in cell longitudinal view and honeycomb structures in transversal view (figure 18A). They are the result

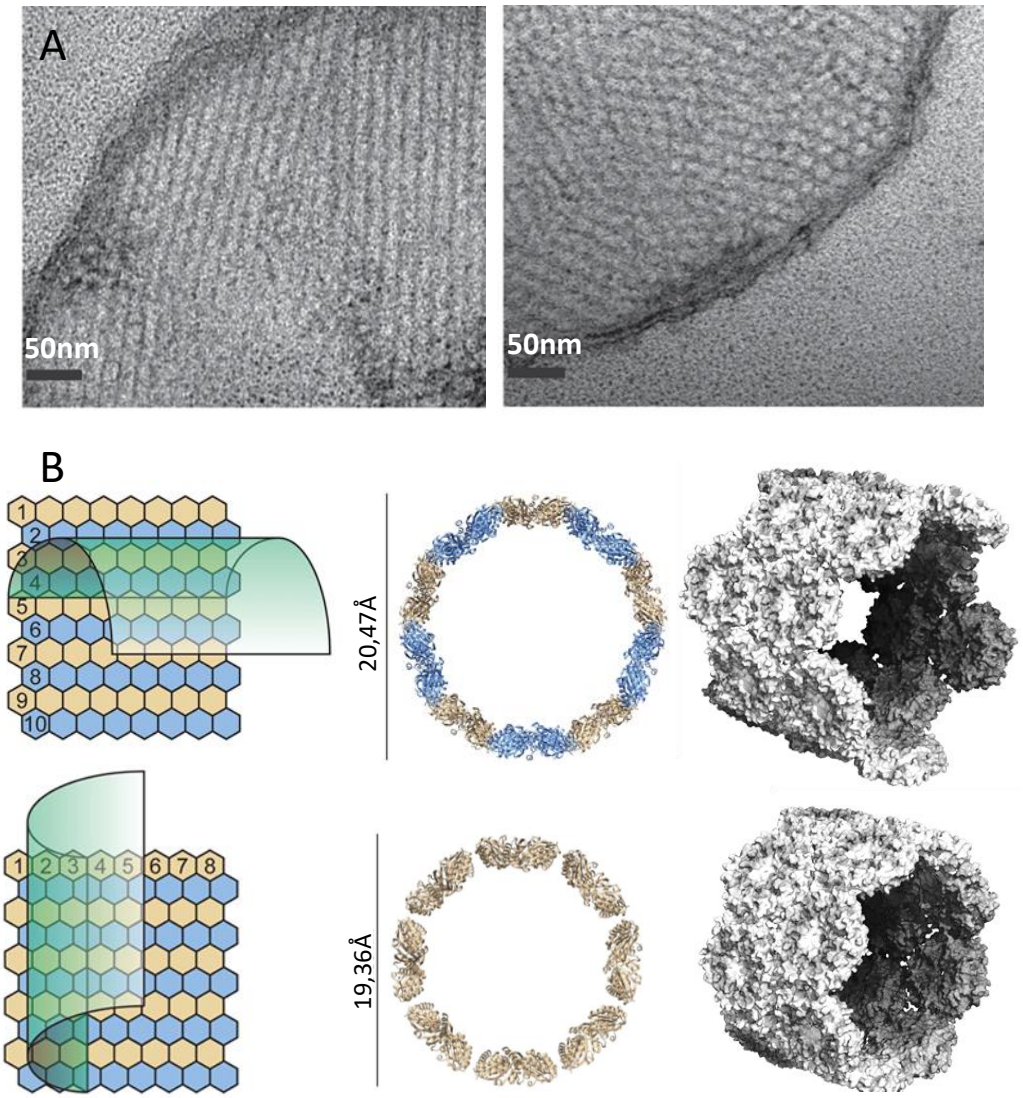


Figure 18. Nanotubes formed by overexpressing the BMC-H RMM in *E. coli*.
A. Nanotubes formation by RMM from *Mycolicibacterium smegmatis*, imaged by TEM. **B.** Models to explain nanotube formation. RMM hexamer would coalesce as a planar sheet before rolling up and circularizing as a nanotube. Topology of the nanotube formed would vary according to hexamer tiling, leading to a 20Å-wide nanotube when 10 hexamers are present per turn or to smaller nanotubes for 8 hexamers per turn. Illustrations from (Noël *et al*, 2016).

of hexamer sheets that circularized (with 8 to 10 tiling hexamers per turn), leaving a hollow luminal space, clear to electron (figure 18B).

Another group of BMC-H produced different macrostructures with laminar features resembling Swiss-rolls (typically a hexamer sheet that is rolled up on itself). These macrostructures were denoted in cells expressing CD1918, from the EUT of *Clostridium difficile*, an homolog of EutM from *Salmonella enterica* (figure 19A) (Pitts *et al*, 2012). They were also present in *E. coli* overexpressing HO BMC-H, (also called MicH), the sole BMC-H from *Haliangium ochraceum* BMC (figure 19B) (Young *et al*, 2017). These rolled sheets were clearly visible in high-speed atomic force microscopy (HS-AFM) with a 3,7nm spacing which corresponded to the thickness of a hexamer (figure 19C) (Faulkner *et al*, 2019).

Surprisingly, CcmK2 and CcmK4 from *Synechococcus elongatus 7942* were not forming prominent macrostructures in the bacterium cytosol (figure 19D) (Young *et al*, 2017). Yet, our team identified their homologs from *Synechocystis 6803* as prone to self-assembly into flat sheets in HS-AFM (figure 19E) (Garcia-Alles *et al*, 2017).

Finally, a BMC-H group with notably EutS from *E. coli K-12*, CutR from *Streptococcus intermedius*, PduU from *Salmonella enterica* or CD1908 (shares 65% of sequence identity with PduU) from *Clostridium difficile* EUT (Tanaka *et al*, 2010; Ochoa *et al*, 2020; Crowley *et al*, 2008; Pitts *et al*, 2012) is completely unable to self-assemble. They were proposed to be symmetry breaker and create diversity and dynamics within the formed BMC.

As a matter of fact, hexamer facet assembly is a quite dynamic process which can be finely monitored by HS-AFM (Sutter *et al*, 2016). Both association and dissociation of individual hexamer can be observed from hexamer patches. In that way, it was also observed that hexamers embedded within the patch could be excised but to a lower frequency than peripheral hexamers. This phenomenon was proposed to happen in the BMC and make possible the exchange of damaged shell subunits or, to a further extent, grant the BMC with the ability to adapt according to its environment (heat, pH, salinity) by replacing an hexamer by another homologs with different characteristics.

4.3. Intra-hexamer associations

Besides their ability to self-assemble at the inter-subunit level, BMC-H also self-assemble at the intra-subunit level, as do BMC-T and BMC-P, which will be referred to here as the intra-hexamer associations or interactions. In that matter, plenty of BMC-H 3D structures have already been resolved

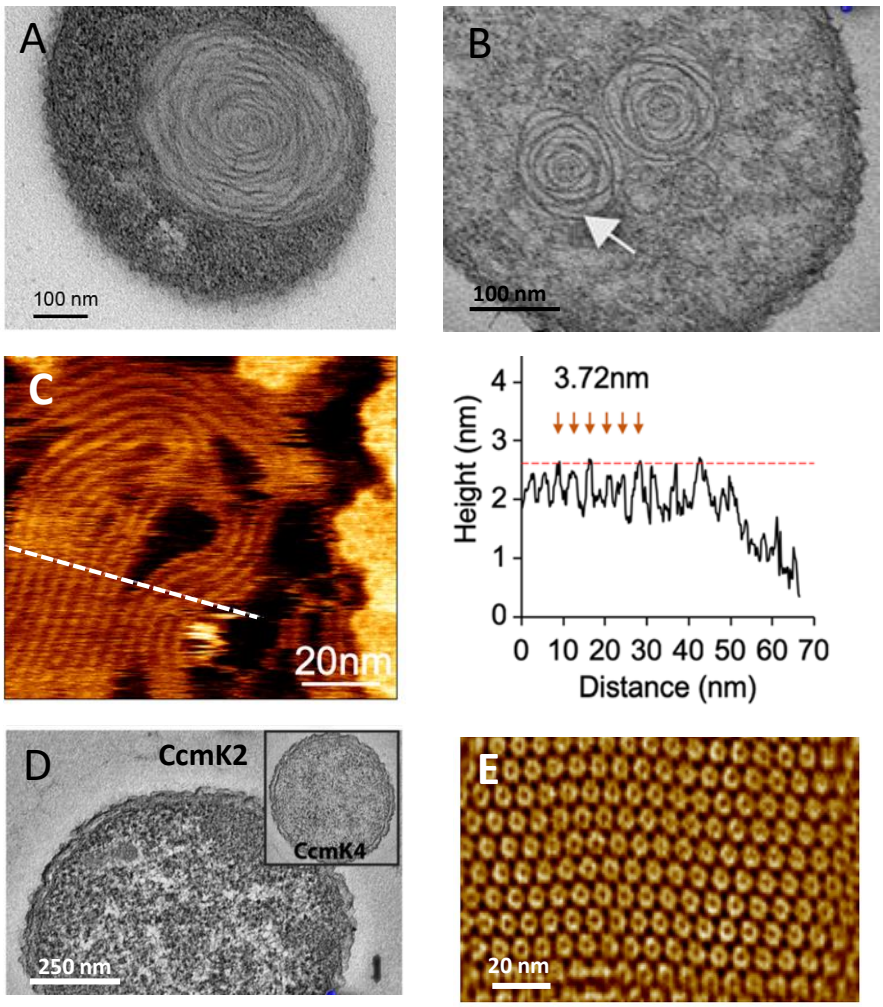


Figure 19. Variety in macrostructure formation by BMC-H homologs.
TEM observation of recombinantly expressed BMC-H: Swiss-rolls formed by CD1918 from *Clostridium difficile* in panel **A** (Pitts *et al*, 2012) or by HO BMC-H from *Haliangium ochraceum* in panel **B** (Young *et al*, 2017) or absence of visible macrostructure for CcmK2 and CcmK4 from *Synechococcus elongatus* in panel **D** (Young *et al*, 2017). **C.** HO BMC-H Swiss-roll formation observed in high-speed atomic force microscopy (Faulkner *et al*, 2019). The different sheets were 3,72nm-distant from each other as depicted by distance measures along the white dashed line. This distance is consistent with the thickness of a hexamer (around 30Å). **E.** *Synechocystis* CcmK2 hexamer tiling as flat sheet in HS-AFM (Garcia-Alles *et al*, 2017).

by X-ray crystallography. A website compiling all these structures was created recently: MCPdb. It also gathers the 3D structures of BMC-T, BMC-P and BMC shell along with BMC-associated cargo enzyme structures. All BMC-H structures deposited to date are homo-hexamers, *i.e.* the same BMC-H repeating 6 times.

Although no study has focus yet on the exact residues governing intra-hexamer interactions, we can mention that the BMC-H internal interfaces are populated with extended hydrophobic patches. One could suppose that theses patches serve the same purpose as in peripheral interfaces, and would be complemented by clamping hydrophilic residues to stabilize the interaction.

More data are necessary as to which residues are involved in BMC-H association. However such study seems difficult to undertake because even a point mutation on the intra-hexamer interface that would normally have a small effect is repeated on 6 interfaces which could act in synergy and produce deleterious effect on hexamer assembly while the residues were not of the greatest importance.

4.4. Hetero-hexamers, anomaly or physiologically relevant ?

If homo-hexamers were long thought to be the only possible oligomerization state, this dogma was recently shattered by the observation that hetero-hexamers could also form. Indeed, two teams, including ours, have shown that, in β-CBX, BMC-H homologs could interact together to form hybrid hexamers (Garcia-Alles *et al*, 2019; Sommer *et al*, 2019). Thanks to protein co-purification and western blot analyses, our team demonstrated that hetero-hexamers could form between CcmK1 and CcmK2 and also between CcmK3 and CcmK4 from the *Synechocystis 6803* β-CBX (Garcia-Alles *et al*, 2019). This result was verified by mass spectrometry which also allowed to determined that BMC-H homolog ratio varied among the hetero-hexamer population.

Study of these hetero-hexamers by AFM depicted a decrease in 2D-sheet assembly compared to the CcmK4 homo-hexamers that formed extended homogeneous patches. This suggested that hetero-hexamers might break the hexamer symmetry. Furthermore, the absence of CcmK3 in CcmK3/CcmK4 crystals corroborated this hypothesis as crystallization conditions favour very symmetric and packed hexamers. Of note, CcmK3 crystals could not be observed because CcmK3 of *Synechocystis* is insoluble when recombinantly expressed on its own, in *E. coli.* Thus, if CcmK3 is not committed in hetero-hexamer formation with CcmK4, it is likely to be aggregated.

In parallel, Sommer *et al* obtained the same hetero-hexamers with CcmK3 and CcmK4 homologs from *Halothece 7418* and *Synechococcus elongatus 7942* β-CBX (Sommer *et al*, 2019). *Halothece* CcmK3 has a bulky Glu residue in the pore region which would hinder stable CcmK3 homo-hexamer

formation. On the contrary, CcmK4 has a Gly residue, conserved between the other CcmK, that would limit the steric clashes around the pore and allow the formation of a stable CcmK3/K4 hetero-hexamer. CcmK3 Glu38 interaction with the Arg38 of CcmK4 was also proposed to stabilize the complex through hydrogen bonding.

In their study, an average ratio of 4 CcmK4 for 2 CcmK3 per hexamer was estimated (Sommer *et al*, 2019). Surprisingly, they also evidenced that these hetero-hexamers were able to superimpose and form double stack, with concave faces facing each other, as do some BMC-T and CcmK2, (Klein *et al*, 2009; Cai *et al*, 2013; Samborska & Kimber, 2012). Modelling of CcmK3/K4 dodecamer allow them to propose that this particular conformation was mediated by the C-terminal helix of the CcmK proteins. Yet, surface area involved in the double-stacking was far less than in trimer double stack (2200 Å² *vs*. 6500 Å² respectively) which may indicate a smaller stability in solution (Sommer *et al*, 2019; Klein *et al*, 2009) and a preponderance for simple hetero-hexamer associations.

5. Questions and objectives of my PhD thesis

5.1. Searching for hetero-hexameric associations beyond the β-CBX

BMC-H are the main and the most diverse shell subunits, in terms of number of homologs within a single operon. Genomic surveys indicate an average of 3,5 BMC-H homologs per operon, with some organisms like *Clostridium saccharolyticum WM1* coding for up to 15 BMC-H, split between 3 BMC types (Axen *et al*, 2014). Recently, hetero-hexamer formation was evidenced between BMC-H homologs, in 2 different β-CBX-expressing bacteria (Garcia-Alles *et al*, 2019; Sommer *et al*, 2019). Indeed, numerous BMC-H homologs share a high sequence identity, notably at the intra-hexamer interfaces (Sutter *et al*, 2017).

When considering the presence of multiple BMC-H per operon, one could wonder if hetero-hexamer formation is a β-CBX-restricted phenomenon or whether it also happens in other BMC types such as the PDU or the EUT. What is the prevalence of these structures? Do the hetero-hexamers comply with particular functions within the BMC?

Besides paving the way for possible hetero-hexamer formation beyond the β-CBX, inside organisms equipped with one BMC type, these 2 recent studies raise the question of possible cross-interactions between BMC-H coming from multiple BMC types.

While some organisms have multiple BMC operons in their genome (Sutter *et al*, 2021) and are able, in theory, to express them simultaneously, there is still a lack of information on the possibility for BMC-H from different BMC subtypes to cross-interact together. Yet, if such hetero-hexamer could

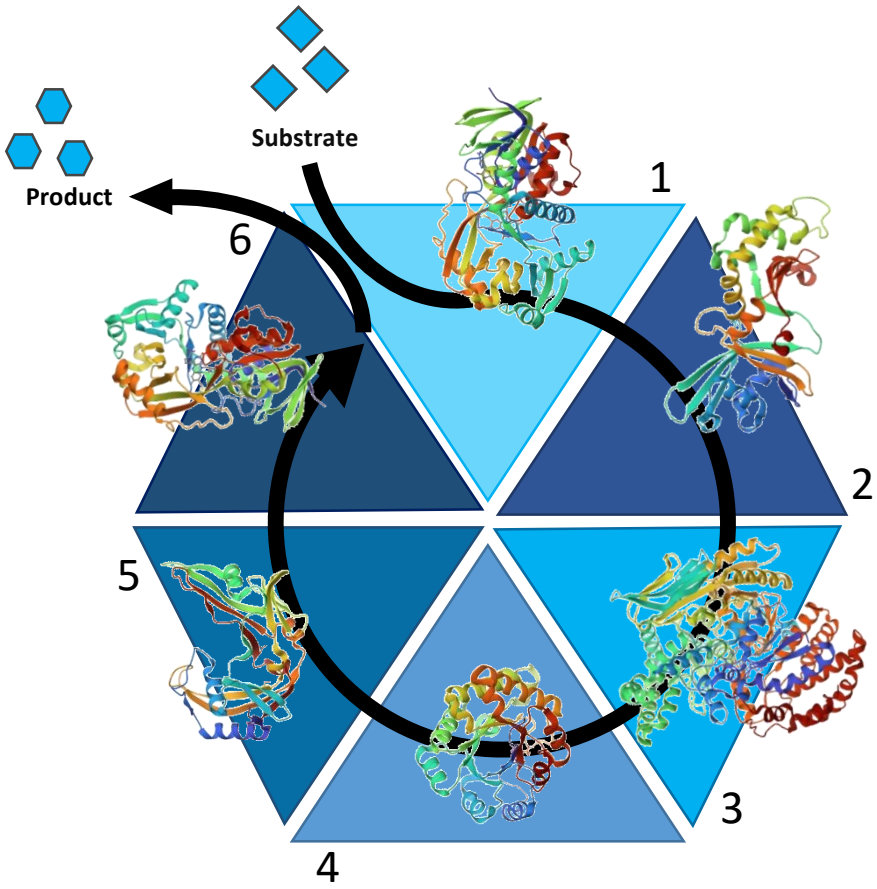


Figure 20. An hetero-hexameric protein platform for synthetic biology.

formed, what would be the impact on BMC functions? Of note, when EutL or EutS were co-expressed with the PDU, the BMC shell integrity was affected and impaired BMC metabolic functions (Sturms *et al*, 2015).

One objective during my PhD thesis was to examine the occurrence of hetero-hexamers in nature. To this end, I first had to find a technology to study protein-protein interactions (PPI) and to adapt it to the particular case of the BMC-H. In that matter, the tripartite GFP was selected. Different parameters to express the BMC-H pairs for the interaction assay (vector strategy, genetic organization, expression control…*et cetera*) were tested. Each one will be detailed during the first chapter of the results.

The best suited parameters were validated with known PPI-status BMC-H pairs before being implemented on the case study of *Klebsiella pneumonia 342* BMC-H, which will be the subject of the second chapter. Of note, this organism is very interesting because it has in its genome 3 BMC *loci,* comprising a total of 11 BMC-H homologs. Indeed, it is capable of expressing the EUT, the PDU and the GRM2. Then, besides allowing to determine whether hetero-hexamers do form aside from the β-CBX, in 3 other BMC types, the study of its BMC-H homologs would also bring some answer elements to the question of the cross-interactions between BMC-H arising from different BMC types.

5.2. Elaboration of a protein platform on the basis of a hetero-hexamer

A novel method to enhance a pathway catalytic efficiency (other than by classical enzymatic engineering) is gaining more and more interests nowadays: enzyme spatial organization. The idea is that, by putting in close proximity or in an arranged fashion the enzymes from a metabolic pathway, one could increase the efficiency of the pathway, through substrate channelling between the different enzymes, for instance, or enzyme clusterisation.

As we saw earlier, the majority of hexamers have the intrinsic property to self-assemble and form higher-ordered macrostructures (nanotubes, fibres, Swiss-rolls, 2D sheets) when recombinantly expressed alone in *E. coli*. This peculiarity has already been exploited in multiple studies to create a protein scaffold for the immobilization of enzymes (Lee *et al*, 2018b; Zhang *et al*, 2018; Liu *et al*, 2022). In these proof-of-concepts, a sole BMC-H was used to build the scaffold, which would only permit to immobilized different enzymes in a random fashion.

Here, we propose to go further with the idea of spatial organization and aimed to elaborate a protein platform starting from an hetero-hexamer. This hetero-hexamer would be composed by 2 up

to 6 different BMC-H, with each BMC-H constituting an anchoring point for a future enzymatic domain (figure 20). With such platform, the spatial organization of the enzymes would be more finely controlled which would further enhance the catalysis efficiency of a metabolic pathway.

To meet this goal, *de novo* designed BMC-H were created by 2 collaborator teams of computational design. In the third chapter of my PhD thesis, I studied them and searched for BMC-H couples that would depict orthogonal intra-hexamer interfaces. Indeed, to be able to control precisely the organization onto the platform, this would require to ensure a specific BMC-H order within the hetero-hexamer and thus, tightly control which BMC-H is adjacent to which one and prevent any other association.

Results

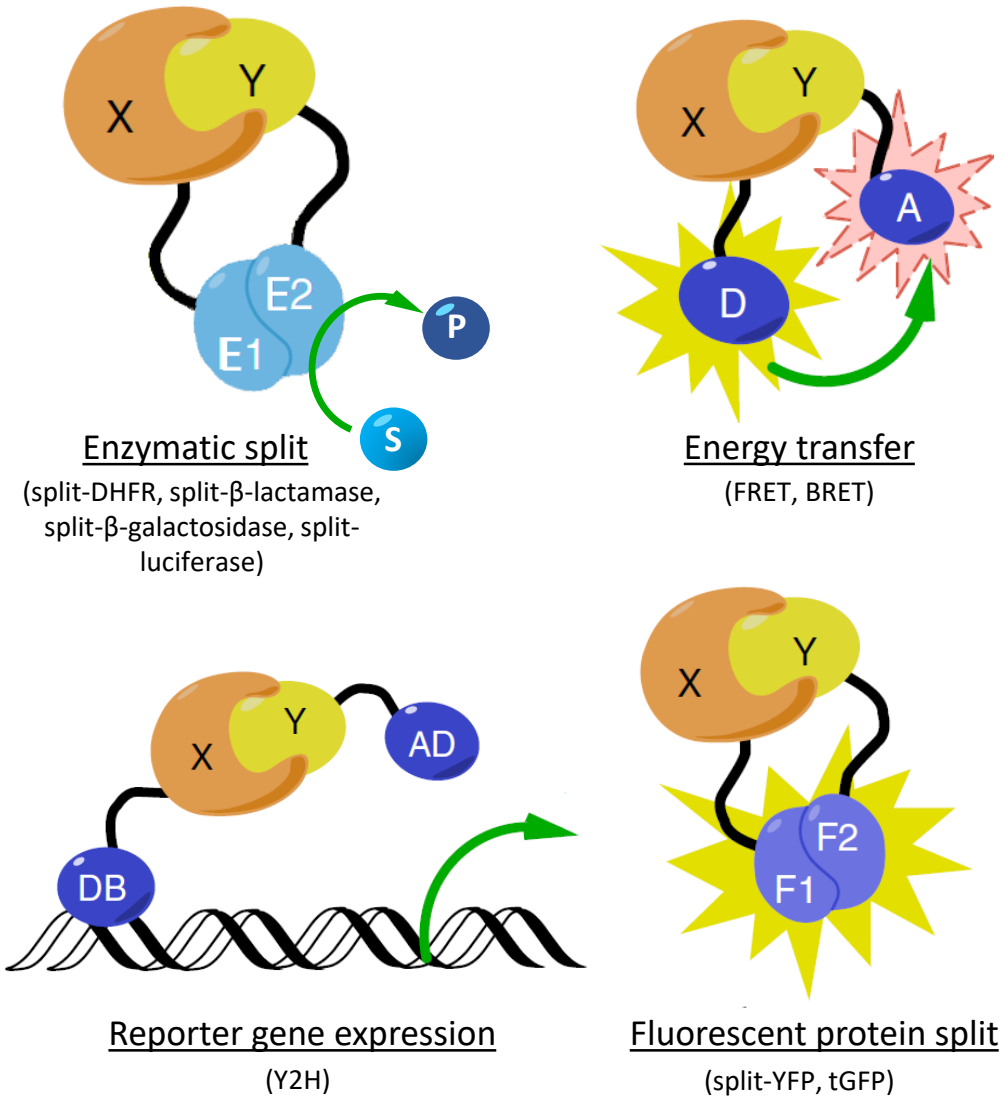


Figure 21. Protein-protein interaction (PPI) study tools.
The main PPI study tools rely on the rapprochement of 2 protein domains which is mediated by the interaction between the partners X and Y. They can be divided in 4 classes. The first class is based on an enzyme (E) reconstitution which is followed up by the apparition of its product (P). In the second class, interacting partners ensure a spatial proximity that allow energy transfer from a donor (D) to an acceptor protein (A) that would subsequently emit light. In FRET experiment, both donor and acceptor are fluorophores while in BRET, the donor is a luciferase. Indeed, the energy that luciferase normally releases, upon ATP and luciferin addition, in the form of photons, is transferred to the fluorophore. In the third class, a transcription factor is split into 2: its DNA binding domain (DB) and its activating domain (AD). The DB is still binding promoter sequences but only the interaction between X and Y can recruit the AD and induce the reporter gene expression. Finally, the last class is based on fluorescent protein (F) reconstitution through partner interaction, leading to fluorescence emission. Illustration adapted from (Choi *et al*, 2019).

Part 2. Results

Chapter 1

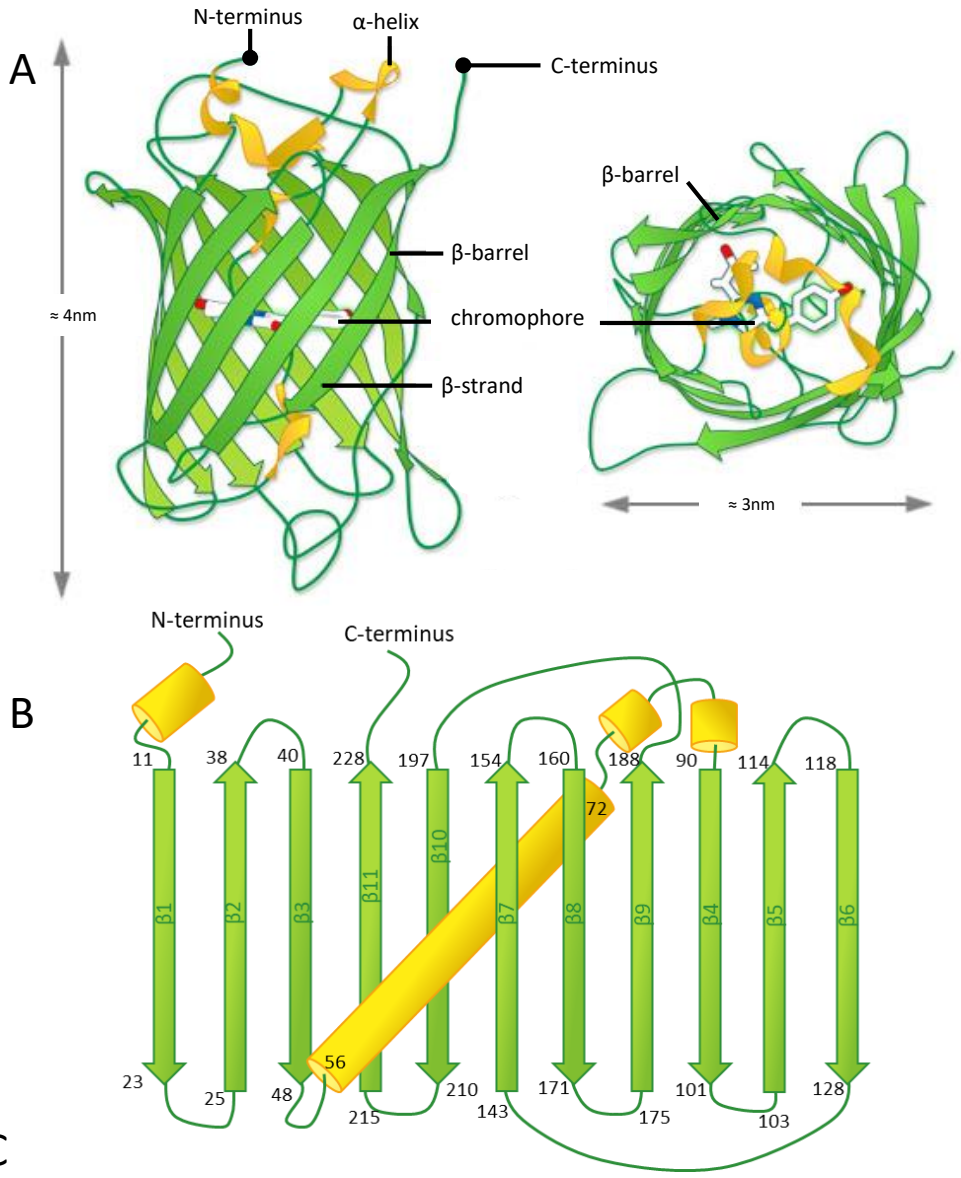
Adaptataion of the tGFP technology for the study of BMC-H interactions

1.1. Introduction to the GFP as a PPI study tool

1.1.1. Protein-protein interaction study tools

Within cells, proteins usually work in complex networks or signalling cascades. Then, to fulfil their biological role, transient or perennial interactions may be established with other protein partners. By determining the whole set of PPIs they might have, their functions and mechanism of action could be better understood. This also applies to BMC proteins, where the existence of, sometimes, multiple BMC-H homologs has been postulated to provide flexibility to the immense BMC structure in response to environmental variations.

Although diverse technologies exist to study PPIs *in cellulo* (figure 21), none of them offers a perfect coverage of PPIs and combining multiple approaches is often required (Choi *et al*, 2019). One of the most used screens is the Y2H which relies on protein complementation assay (PCA) (Fields & Song, 1989). Basically, a transcription factor is split in two parts, its DNA-binding domain and its activating domain which are fused to bait and prey proteins. In the context of a positive PPI, interactions between the bait and prey proteins bring together the 2 portions of the transcription factor, thus restoring its function and inducing the production of a reporter protein. The major drawback is that false negatives happen when the proteins of interest (POI) to be tested are part of the transcription machinery.



C

		β1	β2			β3				β4	β5	β7			β8					β10			β11
	2	16	28	30	39	43	64	65	80	99	105	145	149	153	158	163	166	171	198	205	206	211	223
avGFP		V	S	S	Y	T	F	S	Q	F	N	Y	N	M	K	V	K	I	D	S	A	P	F
frGFP							L	T	R	S				T		A							
sfGFP				R	N		L	T	R	S	T	F		T		A		V			V		
tGFP	R	E	F	R	I	S	L	T	R	Y	T	F	K	T	N	A	T	V	N	T	I	L	Y

Figure 22. The green fluorescent protein (GFP) and its engineering.
A. Schematic representation of the GFP 3D structure. **B.** Decomposed 3D structure of the GFP. βX indicates the number X of the β-strand. The numbers at the beginning and end of each secondary structure element are the numbers of the first and last residues composing the element. **C.** Residue differences between the original *Aequorea victoria* GFP (avGFP) and its engineered forms: fragment reporter GFP (frGFP), super folder GFP (sfGFP) and the tripartite GFP (tGFP). Illustration adapted from (Cabantous *et al*, 2013; Pédelacq *et al*, 2019).

Other technologies based on the same principle of reconstituting a split enzyme in order to monitor PPI were designed: the split dihydrofolate reductase, split luciferase, split β-lactamase, split β-galactosidase, *et cetera* (Blaszczak *et al*, 2021). These technologies all share the same disadvantage that the products of the enzyme reconstitution diffuse away from the PPI site, hampering its localization. Another kind of PPI assay exploits the FRET (Förster resonance energy transfer) phenomenon where POIs to be tested are fused to an energy donor and acceptor fluorophores (Lin *et al*, 2018) or the closely related BRET where the energy donor is a luciferase and the acceptor a fluorophore (Machleidt *et al*, 2015).

Fluorescent PCAs can also be used in that matter. They are easily implementable, do not require the addition of external substrate/components, allow high-throughput PPI studies and the assay can be performed on classical fluorescence microplate readers. Then, subsequent BMC-H interaction studies were performed with a PCA technology based on a split GFP.

1.1.2. Discovery of the GFP

The GFP was first observed in *Aequorea victoria* jellyfish by Shimomura *et al,* in 1962 (Shimomura *et al*, 1962). Observations were made that, upon stimulation (mechanical stress on the jellyfish or addition of Ca^{2+} in crude extracts), green light was emitted at 510-515nm. This resulted from radiative energy transfer between the aequorin luciferase, activated by Ca^{2+} in presence of luciferine that emits blue light, and the GFP (Morise *et al*, 1974). Afterwards, many fluorescent proteins were discovered in other organisms such as the DsRed from *Discosoma,* a cyan fluorescent protein from *Clavularia,* and a yellow one from *Zoanthus* with a wide range of fluorescence emission spectrum (Matz *et al*, 1999).

The GFP is a protein of 238 residues (around 27kDa) that organizes into 11 β-strands around a central α-helix composing the chromophore, a 3D fold called β-barrel (figures 22A & B) (Ormö *et al*, 1996). Three residues are necessary to give rise to the chromophore: the Ser65, Tyr66 and Gly67. In order to become fluorescent, the GFP chromophore undergoes different auto-catalysed steps of maturation: cyclisation between NH of the Gly and CO of Ser, dehydration and oxidation.

1.1.3. GFP engineering

Classically, in fluorescence microscopy or in fluorescence-activated cell sorting based on the GFP utilisation, samples were excited via an argon laser lamp that produced light at a wavelength of 488nm and green light emitted could be observed with a fluorescein isothiocyanate filter that allowed fluorescence passage at 510nm. The *Aequorea victoria* GFP (avGFP) had a maximal excitation peak at

396-398nm and a secondary peak at 476-478nm. Of note, the second peak of excitation induced less fluorescence photobleaching than the first peak (Heim *et al*, 1995).

First engineering of the avGFP was performed with the aim of increasing the excitation efficiency of the second wavelength to preserve GFP fluorescence. Heim *et al* performed point mutations on the avGFP and found that when the Ser65 was changed for a Thr, only one excitation peak was present for the GFP, at 490nm, with no change in maximal emission wavelength (Heim *et al*, 1995). This excitation peak resulted in a fluorescence 5 times brighter than in the wild type GFP. Besides, chromophore maturation happened in approximately 1h30 compared to 6h for the avGFP.

When recombinantly expressed in *E. coli*, a large portion of the avGFP is detected in inclusion bodies in a non-fluorescent form (Cormack *et al*, 1996). In the same study, random mutagenesis was performed on residues encompassing the chromophore (residues 55 to 74) to improve the GFP solubility. In that manner, combinations of mutations on the Ser65 and on Phe64, Val68 or Ser72 increased the GFP brightness 10 to 100-fold. Above all, the combination of Ser65Thr and Phe64Leu mutations was shown to promote folding, solubility (90% of the GFP was now soluble) and to decrease the maturation delay. Furthermore, in less than 8min, fluorescence started to appear and reached a maximum in 1h. When excited at 488nm, this particular mutant was 35-fold more fluorescent than the avGFP. Later, it gave rise to the eGFP, after codon optimization for human cells, with a high expression and brightness in eukaryotic cells.

With its brightness and its ability to be recombinantly expressed, the GFP gained tremendous interest in the biology field, especially for protein localization within eukaryotic cells (Nikles *et al*, 2008; Böhm *et al*, 2017). It was also used as a folding-reporter (frGFP) after the observation that the GFP solubility and fluorescence was correlated to the solubility of the POI it was fused to (Waldo *et al*, 1999). Indeed, fluorescence screens permitted to report on POI solubility improvement along evolution rounds. To avoid this major pitfall when analysing POI expression and localization, Pédelacq *et al* engineered a superfolder GFP (sfGFP) out of the frGFP, able to fold properly, independently of the solubility properties of the fused protein (Pédelacq *et al*, 2006). This sfGFP version included 6 new mutations (figure 22C), was more stable to urea denaturation and less prone to dimerization.

As the GFP is a relatively bulky protein (27kDa) and that it could impact subcellular localization (Cui *et al*, 2016), another strategy was designed to track proteins *in vivo*: the split GFP where the POI was fused to the β-strand 11 of the GFP (GFP11) and reconstitution of the full GFP was possible through concomitant expression of the remaining GFP1-10 portion (Cabantous *et al*, 2005). The major advantage of this strategy is that the small size of the GFP tag (only 20 residues) would presumably

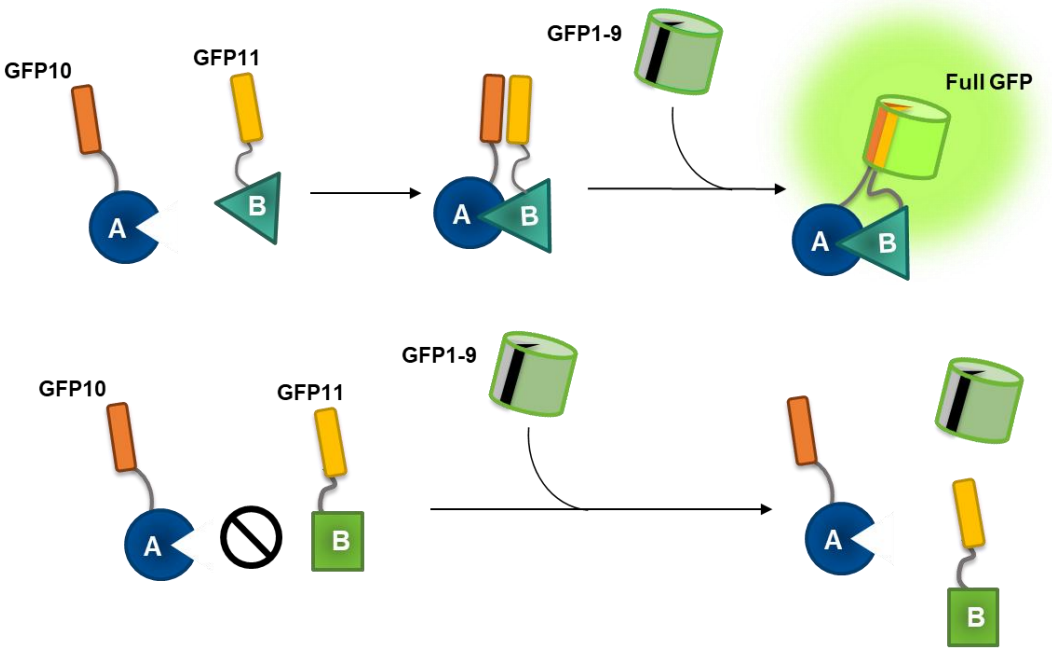


Figure 23. The tripartite GFP technology.
The GFP is composed of 11 β-strands that can be split into 3 parts (the GFP1-9, GFP10 and GFP11) and be used to report on protein-protein interactions (PPI). In a positive PPI between the proteins A and B, the interacting partners bring in close proximity the GFP10 and 11 tags to which they are fused. Upon the GFP1-9 arrival, this favours the reconstitution of a full GFP emitting green fluorescence. On the contrary, if proteins A and B do not interact together, no GFP reconstitution happens, hence no fluorescence.

reduce the perturbations on POI expression or folding, subcellular localization and interactions. Yet, the GFP1-10 was insoluble when expressed on its own in *E. coli* and the GFP11 fusion reduced POI solubility. Then, several rounds of directed evolution were performed. An optimized GFP1-10 (GFP1-10 OPT) was designed with 11 mutations compared to the frGFP, resulting in improved solubility (50% of the GFP1-10) and *in vitro* complementation with POI-GFP11 by 80-fold. Also, a variant of the GFP11 tag, called GFP11 M3, was created with mutations Leu221His, Phe223Tyr and Thr225Asn to improve POI-GFP11 fusion solubility.

1.1.4. A GFP-based technology to study PPIs

Later on, the same team developed a PPI-sensing system starting from the split GFP1-10/GFP11 with better solubility, folding and maturation kinetics: the tripartite GFP (tGFP) (Cabantous *et al*, 2013). This technology is composed of the β-strands 10 (residues 194-212; GFP10) and 11 (residues 213-233; GFP11) of the GFP which are fused to bait and prey proteins (figure 23). Briefly, interacting partners bring together the GFP10 and GFP11 tags which, in the presence of co-expressed GFP1-9 fragment (residues 1-193), favors the reconstitution of an entire and functional GFP emitting fluorescence. On the contrary, in absence of PPI, the probability of reconstitution is diminished due to distant GFP10 and GFP11 tags.

The main advantages here are that very small tags are affected to both POIs and this would have fewer impacts on POI expression, folding or interactions than fusing the GFP1-10 to one of the POI to be tested. Also, fluorescence allows direct visualization of PPI localization if need be. Finally, the GFP reconstitution is irreversible, allowing the detection of transient and low-affinity PPIs. While this characteristic, mixed with POI cytosolic accumulation would cause an increase in unspecific GFP reconstitution due to fortuitous encounters in bipartite GFP assays, dividing the GFP into 3 fragments would decrease the frequency of random encounters, leading to fewer unspecific signals.

One of the aim of my thesis was to characterize the interactions within the oligomeric subunits of the BMC shell with a special focus on those occurring inside hexamers. Since these proteins belong to the same structural family, I opted for adapting the tGFP technology (which seemed the best suited) to this specific case rather than merging different techniques.

In this first chapter, we will go through the different parameters that were optimized to be able to explore BMC-H PPIs, namely the choice of vector strategy, genetic organization, promoter control and linker length. Some problems arose during this study and will be tackled here such as the poor GFP1-9 solubility or the macrostructure formation by BMC-H hexamers.

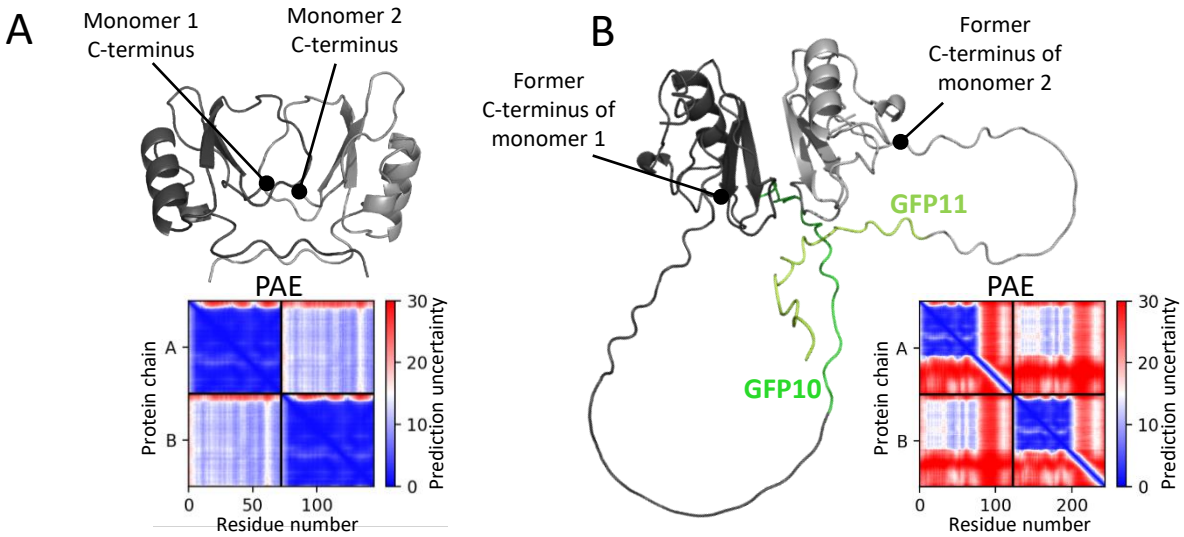


Figure 24. Structure predictions for BWI negative control.
BWI was crystallized and resolved as a monomeric protein (3RDY). Though Alphafold2 predicted a dimer with both C-termini buried in the interacting interface (**A**). When run on BWI tagged with the GFP10 or the GFP11, a dimeric association was no longer proposed but distinct monomers (**B**). Predicted aligned errors (PAE) are provided for each AF2 prediction. Note that the flexible linkers and tags appear as unstructured and mobile loops which increases the prediction uncertainty of these segments while BWI core remains well predicted.

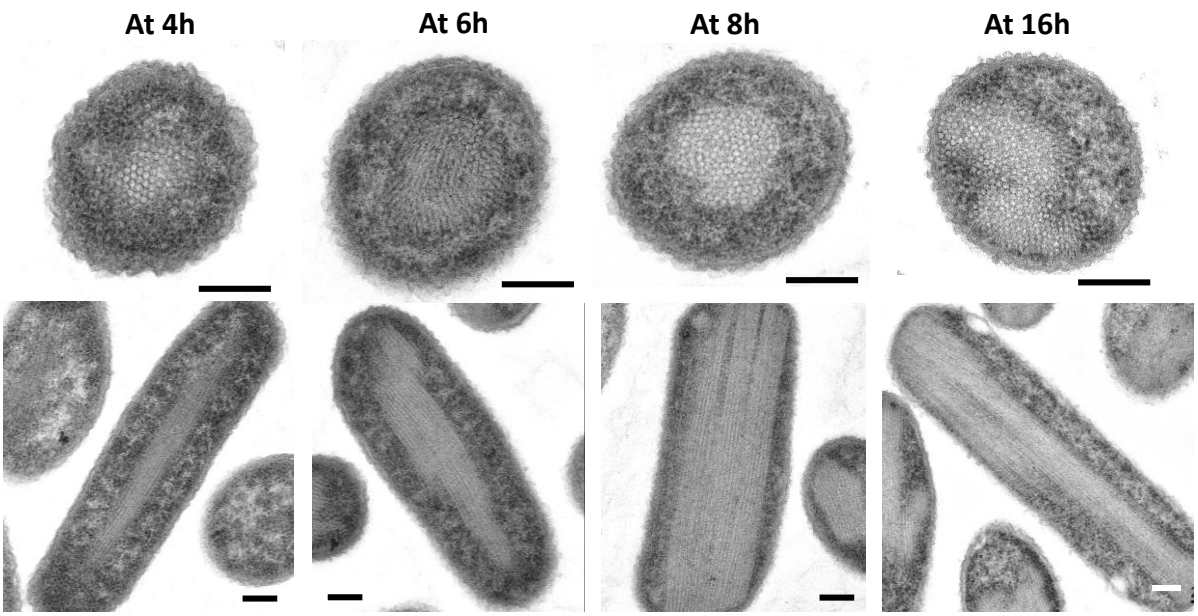


Figure 25. Kinetics of nanotube formation in *E. coli*.
TEM observations of cells overexpressing His_6-tagged RMM at different culture times after IPTG induction. Longitudinal and transverse views are provided for each time point. Scale bar = 200nm.

1.2. Pursue of the best parameters to study BMC-H interactions

1.2.1. Starting vector construct and design of controls

The general organization of the starting tGFP vector was the following. The same vector coded for the 3 tGFP partners (POI1-GFP10, POI2-GFP11 and GFP1-9). Each open reading frame (ORF) was under the control of a T7 promoter (T7p) and terminator, leading to 3 independent transcripts. The POIs were connected to GFP tags by flexible linkers of 30 and 27 residues for the GFP10 and 11, respectively. Finally, the GFP1-9 bears a His_6 tag on its C-terminus. The tGFP assays were performed *in vivo*, in *E. coli*.

Multiple phenomena can contribute and create nonspecific signals in the reporter technology. For instance, random encounters between the 3 tGFP partners could occur with increasing intracellular POI concentration. The buckwheat trypsin inhibitor (BWI) was selected to account for it. Though Alphafold2 (AF2) predictions pointed to a potential dimerization of BWI (figure 24A), a monomeric state was revealed in crystal structure (Wang *et al*, 2011). Furthermore, since the last C-terminal residue of BWI was embedded in the AF2-predicted dimer, no dimerization interface was found when AF2 was run with the C-terminally-tagged BWI-GFP10/BWI-GFP11 pair, thus reinforcing monomer prevalence in the tGFP assay (figure 24B).

Another phenomenon which contribution had to be probed was the formation of aggregated material. CcmK3 is a BMC-H from *Synechocystis 6803* which is highly expressed in *E. coli* but in insoluble form, regardless of tag identity or position (Garcia-Alles *et al*, 2017), and was chosen for this matter.

Finally, BMC-H have the ability to assemble as macrostructures when expressed alone *in vivo* (see part 1, section 4.2). Then, it is possible that a portion of the GFP signal emanates from inter-hexamer interactions in addition to BMC-H oligomerization. Multiple controls would be elaborated to determine whether inter-hexamer assembly participated in the GFP reconstitution and signal in the next section.

1.2.2. BMC-H macrostructure contribution to the tGFP signal

BMC-H often coalesce to form higher-ordered macrostructures like nanotubes or sheets or Swiss-rolls when overexpressed inside cells or when observed *in vitro*, from purified proteins (Pang *et al*, 2014; Young *et al*, 2017; Pitts *et al*, 2012; Garcia-Alles *et al*, 2023). During the course of this thesis, I used as model protein the *Rhodococcus* and *Mycobacterium* Microcompartment BMC-H, referred to as RMM, from the AAU of *Mycolicibacterium smegmatis MC2 155* (Mallette & Kimber, 2017). RMM is well expressed in *E. coli* and assembles into nanotubes (Noël *et al*, 2016).

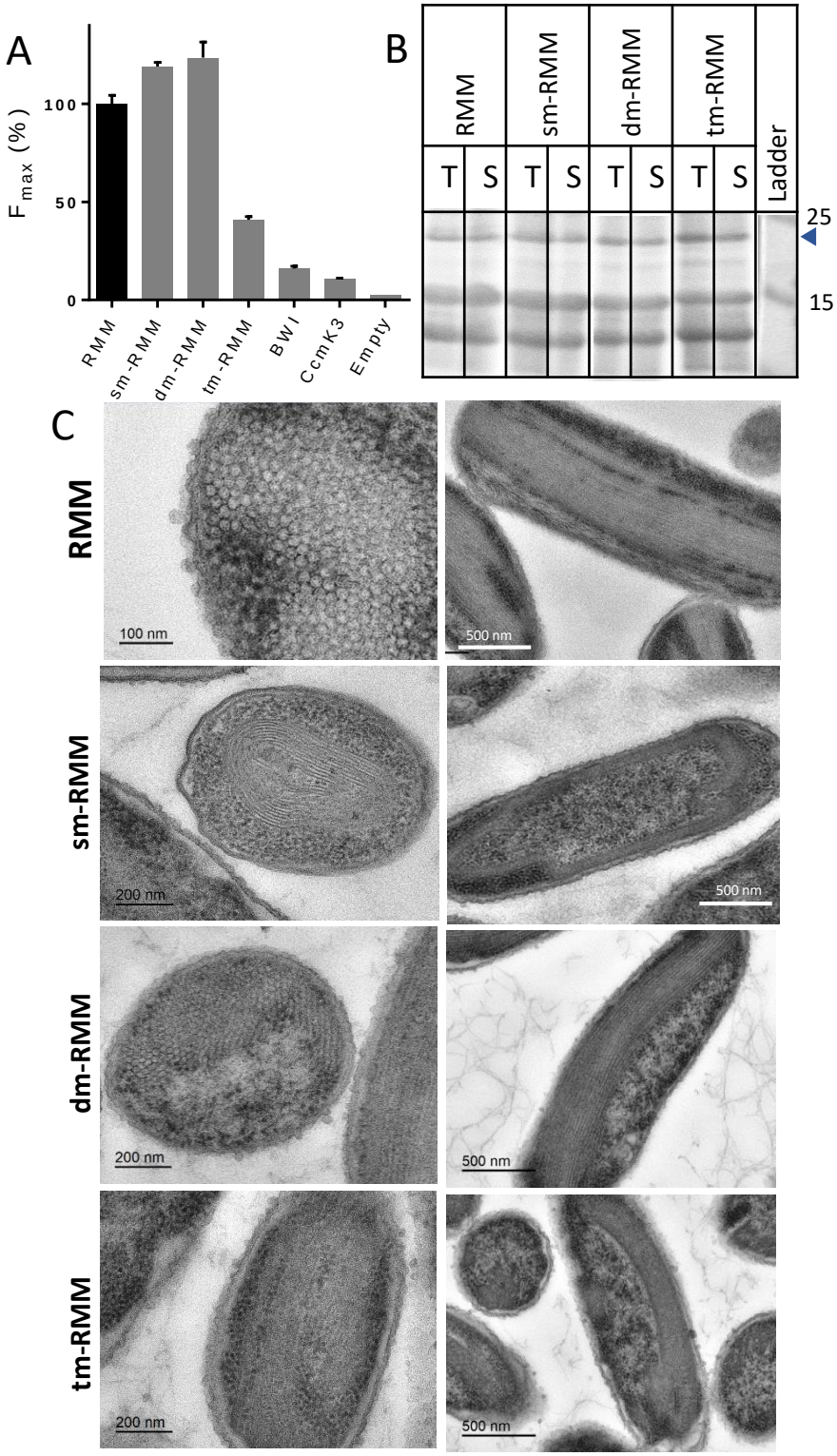


Figure 26. Mutations on RMM peripheral residues to impede inter-hexamer assembly. **A.** tGFP interaction assay of RMM or RMM^{K26D} (sm-RMM), RMM^{N29D,A53D} (dm-RMM), RMM^{K26D,N29D,A53D} (tm-RMM) homo-pairs. Plotted values are the maximal fluorescence values (F_{max}), given as a percentage of the RMM reference case (black bar). **B.** Analysis of total (T) or soluble (S) protein fractions by SDS-PAGE. Molecular weights are in kDa and the GFP1-9 migration area is depicted by the blue arrow. **C.** TEM observations of *E. coli* cells overexpressing His_6-tagged wild-type RMM or its mutants after 16h of induction. Longitudinal and transverse views are provided for each case.

Under our standard induction conditions (10µM IPTG from the beginning of the culture), inter-hexamer assembly was already visible in TEM after 4h of culture, possibly even before as earlier times were not inspected (figure 25). Nanotubes were nucleating in *E. coli* cytoplasm. They had a diameter of 21 ± 2nm, consistent with 12 hexamers per turn and a -30° bending angle between 2 adjacent hexamers (Pang *et al*, 2014). They appeared as honeycomb structures in cell transversal view and as bundles in longitudinal view. By the time of 6h of culture, nanotube pool had considerably increased and by 8h, *E. coli* cytosol was filled with macrostructures. Nanotubes expanded throughout the cells, sometimes interfering with septation. These macrostructures were observable up until the typical end of the culture (16h).

Mutations on hexamer peripheral residues

In the context of 2 assembled hexamers, RMM C-termini are very close to each other. If these adjacent BMC-H were tagged with the GFP10 and GFP11, connected by the long flexible linkers Lk30/27, the GFP could in principle be reconstituted. Then, it is possible that GFP signal emanates from inter-hexamer assembly.

To rule out nanotube participation in the tGFP signal, I attempted to prevent inter-hexamer assembly. Mutations were introduced on RMM peripheral residues that were shown to be involved in hexamer-hexamer interactions in other BMC-H (Garcia-Alles *et al*, 2017; Pang *et al*, 2014; Sutter *et al*, 2019; Garcia-Alles *et al*, 2023): RMM^{K26D} (sm-RMM), $RMM^{N29D,A53D}$ (dm-RMM), $RMM^{K26D,N29D,A53D}$ (tm-RMM). The Lys26Asp mutation was selected based on published TEM observations of an absence of assembly for the equivalent PduA mutant (Pang *et al*, 2014). The choice of Asn29Asp mutation was motivated by a study where PduA mutated on the Asn29 led to impaired shell integrity (Sinha *et al*, 2014). Finally, the Ala53 was chosen for its special localization on the hexamer edges. Indeed, as we saw earlier, the residues present at the centre of a hexamer triad are generally short-chained residues (see part 1, section 4.1) (Sutter *et al*, 2017, 2019; Kalnins *et al*, 2020) that do not perturb inter-hexamer assembly. Then, the Ala53 was changed for a bulky Asp.

Homo-pairs of RMM or its mutant forms were assayed in tGFP. Optical density (OD) at 600nm and GFP fluorescence were monitored during 16h of induction with 10µM of IPTG. The OD_{600nm} and fluorescence curves were fitted to a sigmoidal function (see Material and methods). Maximal fluorescence (F_{max}) values were extracted and normalized by the wild-type RMM (wt-RMM) F_{max} value (figure 26A).

Both sm-RMM and dm-RMM resulted in F_{max} values slightly higher than the wt-RMM. A significant drop was noticed for tm-RMM (approximately to 40% of the wt-RMM value). Delays between midpoints of cellular growth and fluorescence curves were calculated. They were similar for the wt-RMM (3,4h), sm-RMM (3,5h) and dm-RMM (3,4h) but increased considerably for the tm-RMM (5,6h).

To explain such drop in the tm-RMM pair fluorescence, protein expression was verified in SDS-PAGE (figure 26B). No difference in neither protein expression nor solubility could be evidenced between all cases.

To determine whether introduced mutations disrupted macrostructure assembly, wt-RMM and its mutants, each carrying a C-terminal His_6 tag, were overexpressed in *E. coli* before TEM observation (figure 26C).

Well-defined bundles of nanotubes were visible for the wt-RMM (measured nanotube diameter of 21 ± 2nm). Surprisingly, nanotubes still formed with the dm-RMM (21 ± 2nm). Nanotube formation was not evidenced for any of the 2 mutants incorporating the Lys26Asp mutation (sm-RMM or tm-RMM). Yet, compact assemblies continued to form with the sm-RMM, which appeared like Swiss-rolls (11 ± 2nm inter-spacing). Data were more contrasted for the tm-RMM which depicted signs of assembly albeit in a lesser extent than its counterparts and with a morphology not resembling neither nanotubes nor Swiss-rolls. Occasionally, tm-RMM even showed a propensity to self-aggregate at the cell pole, a characteristic that was more pronounced when the GFP version was visualized by TEM (supp figure 1). Of note, macrostructure formation was less evident in GFP-tagged RMM (*i.e*. nanotubes appeared loosely packed in transversal views). This was mostly true for the mutant forms. Indeed, although expressed proteins seemed to collapsed together, no clear repetitive patterns reminiscent of nanotubes or Swiss-rolls could be seen.

Collectively, these data suggested that inter-hexamer assembly is a very robust phenomenon which could not be prevented by a single point mutation of conserved Lys26 nor by combinations of mutations on peripheral residues. Thus, this method did not permit to draw any conclusion on the participation of inter-hexamer assembly in the tGFP assay. Or, at the very least, one could assume that, if a part of the GFP signal was owed to macrostructure formation, this part was affected by the type of structures formed.

Playing on linker length to monitor different associative phenomena

Linker length is an impacting factor for the detection of PPIs. This was shown for instance in a large-scale study using another PCA based on the dihydrofolate reductase where longer linkers allowed to capture a higher number of PPIs (Chrétien *et al*, 2018). Besides, within the original GFP, β-strands 10 and 11 are aligned in antiparallel (figure 22B). Thus, for the tGFP assay, linkers should be long enough to enable the good orientation of GFP tags and proper reconstitution of the GFP. Though a long linker could lead to the detection of inter-hexamer assembly in addition to BMC-H oligomerization.

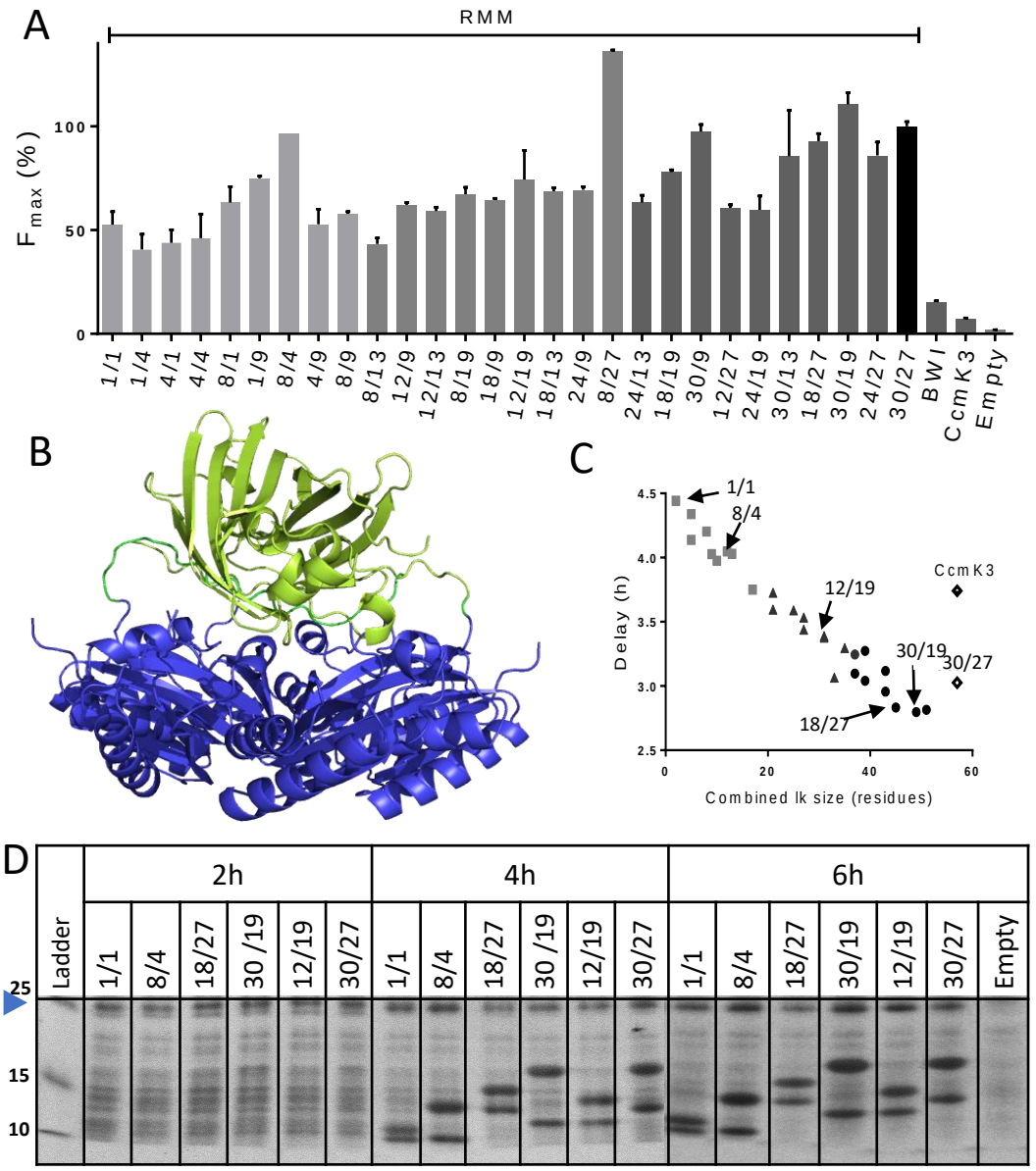


Figure 27. Influence of the linker length on the GFP reconstitution.
RMM was connected to the GFP10 or GFP11 tags by flexible linkers (Lk, Gly/Ser-rich) of varied sizes. The name of each case is defined as follow: residue number connecting the POI1 to the GFP10/residue number connecting the POI2 to the GFP11. **A.** tGFP assay on RMM homo-pair with different Lk length. F_{max} results are given as a percentage of the RMM Lk30/27 reference (black bar). Of note, BWI and CcmK3 had a Lk30/27 combination. **B.** Improper GFP reconstitution prediction by Alphafold2 when the linkers connecting RMM to the GFP tags are Lk1/1. **C.** Delay in fluorescence apparition, calculated as the time interval between reaching half F_{max} and half maximal cellular growth. BWI delay value which exceeded 6h, is not shown here. **D.** Expression kinetics of the tGFP partners according to the linker length. The cases indicated by arrows in (C) were collected after 2, 4 or 6h of induction for protein soluble fraction analysis in SDS-PAGE. The empty case is cells transformed with an empty pET26b. The blue arrow points at the GFP1-9 migration position. Molecular weights are in kDa.

Next attempted strategy to restrict inter-hexamer assembly contribution to the tGFP signal was to shorten the length of the linker between the POIs and GFP tags. Different sizes for the 2 linkers connecting the RMM pair to the GFP10 or 11 tags were tested in tGFP: from 30 or 27 residues, respectively (Lk30/27, original linker size) down to 1 residue each (Lk1/1).

The global trend was a decrease in fluorescence when reducing linker length, with approximately a 3-fold difference between F_{max} values of extreme cases (figure 27A). Surprisingly, the combination of the shortest linkers still fluoresced although a single residue length was initially considered incompatible with a correct GFP reconstitution by 2 BMC-H belonging to the same hexamer. Such structural incompatibilities were highlighted by AF2 predictions, which indicated incomplete anchoring of GFP10 and GFP11 tags during reconstitution of the tGFP for the Lk1/1 (figure 27B).

Fluorescence not only decayed for the shortest linkers but was also delayed in time (figure 27C). Indeed, the time lapse measured between the half maximal cell growth and half F_{max} times increased from less than 3h with the longest linker combinations to about 4,5h with the Lk1/1. Of note, this delay reached 6h for the BWI control (with Lk30/27). This was in agreement with BWI being a negative control that informed on the participation of random encounters in the GFP signal. Indeed, random encounter frequency would increase belatedly, when proteins had accumulated and reached a high cytosolic concentration.

To certify that these differences in fluorescence apparition were not due to different kinetics of protein accumulation, protein expression was analysed in SDS-PAGE at 3 different moments of the culture, for several linker combinations that exhibited extreme and intermediate delays: RMM pair with either the Lk1/1, Lk8/4, Lk12/19, Lk18/27, Lk30/19 or Lk30/27 (figure 27D).

Comparable protein expression was observed at each time.

Globally, these data suggested that according to the linker length, the signal obtained in tGFP was emanating from different associative phenomena. BMC-H oligomerization probably happened as early as the end of BMC-H synthesis. Some studies even showed that translation and association of operon-coded protein complexes are occurring concomitantly (Shieh *et al*, 2015b; Bertolini *et al*, 2021). Inter-hexamer assembly and macrostructure formation would only take place afterwards and would require a consistent pool of hexamers to start nucleating. In our hands, macrostructure formation was evidenced at least 4h after the beginning of induction. BMC-H oligomerization would mainly drive GFP reconstitution with a combination of long linkers while the signal would arise from random encounters of freely diffusing hexamers or from the interactions between 2 BMC-H belonging to 2 assembly-committed adjacent hexamers with the short linkers.

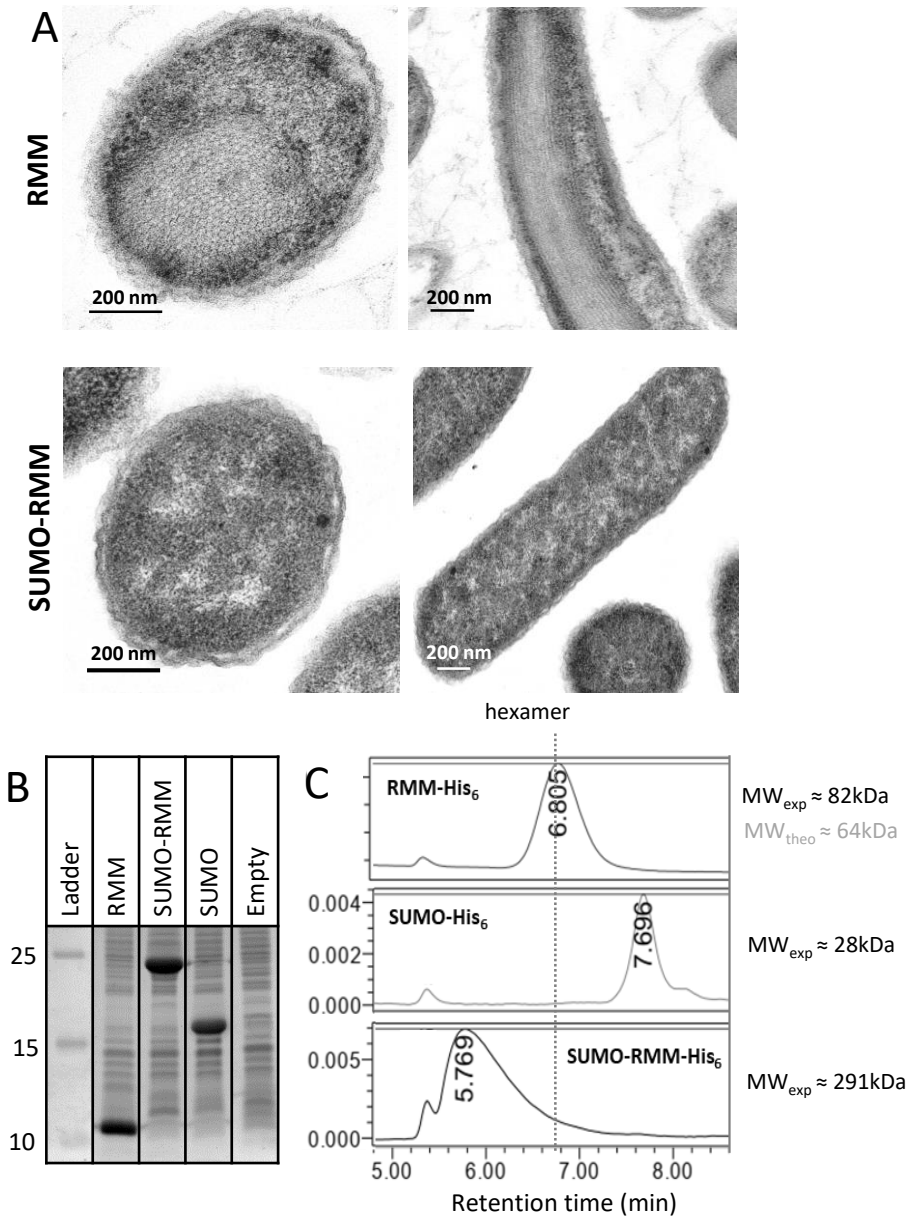


Figure 28. Impediment of inter-hexamer assembly by sumoylating RMM.
A. TEM observations of *E. coli* cells overexpressing RMM-His_6 N-terminally fused or not to a SUMO domain. Longitudinal and transverse views are provided for each case. **B.** Total protein fractions of the same constructs were analysed by SDS-PAGE. The empty case is cells transformed with an empty pET26b while SUMO case is the His_6-tagged SUMO domain alone. Molecular weights are given in kDa. **C.** His_6-tagged versions of RMM, SUMO or SUMO-RMM were purified by cobalt affinity chromatography before injection in size-exclusion high-pressure liquid chromatography to determine their oligomerization state. Aggregated species (>2MDa) elute within the 5,4min peak. Experimental molecular weights (MW) were calculated thanks to standard proteins and are indicated on the right of the panel.

Sumoylation of RMM to preclude inter-hexamer assembly

Efforts to gauge inter-hexamer assembly participation in the tGFP signal were carried with another strategy. In order to prevent macrostructure formation, sterically-hindering small ubiquitin-related modifier (SUMO) domain from *Saccharomyces cerevisiae* was fused to RMM. This protein is exclusively found in eukaryotes where it is added post-translationally onto proteins to modify their functions. SUMO tagging is widely used in recombinant protein production to improve protein expression and solubility (Malakhov *et al*, 2004). Besides, SUMO fusion has already been implemented on BMC-H which allowed production and purification of unassembled, yet highly concentrated hexamers (Hagen *et al*, 2018b).

In preliminary experiments, the SUMO domain was inserted at the N-terminus of RMM-His_6 and cells expressing this construct were observed in TEM.

While RMM-His_6 was forming nanotubes extending throughout *E. coli* cytosol, no structure was visible for SUMO-RMM-His_6 (figure 28A). Importantly, RMM-His_6 and SUMO-RMM-His_6 were expressed in equivalent quantities (figure 28B).

Then, SUMO fusion succeeded in interrupting inter-hexamer assembly, corroborating previously published data (Hagen *et al*, 2018b).

Within a hexamer, the SUMO domain would be in 6 exemplars. To determine whether repetition of this bulky domain was impacting intra-hexamer interactions, RMM and SUMO-RMM oligomerization states were analysed by size-exclusion chromatography (SEC) after protein purification. Elution of different standards (see Material and methods) was also monitored to establish a standard curve, permitting to calculate an experimental molecular weight (MW) for RMM and SUMO-RMM.

RMM eluted at 6,8min (MW_{exp} ≈ 82kDa), the expected retention time for a hexamer (figure 28C). The retention time of SUMO-RMM was of 5,7min (MW_{exp} ≈ 291kDa), consistent with an oligomer, demonstrating that RMM was still able to self-oligomerize when fused to the SUMO domain.

SUMO domains were fused on a GFP10/11-tagged RMM pair (on a bicistronic vector leading to POI transcription on the same messenger ribonucleic acid (mRNA) and independent GFP1-9) to probe macrostructure contribution to the GFP signal. The C- and N-termini of RMM are protruding on the same hexamer face (the concave face). In the same fashion that 6 SUMO domains were shown to preclude inter-hexamer assembly, these bulky domains might also impact the GFP1-9 approach and the GFP reconstitution. Then, partial SUMO tagging was also implemented on the RMM pair (either on RMM-GFP10 or RMM-GFP11 or both partners). Besides, to unveil the phenomenon behind

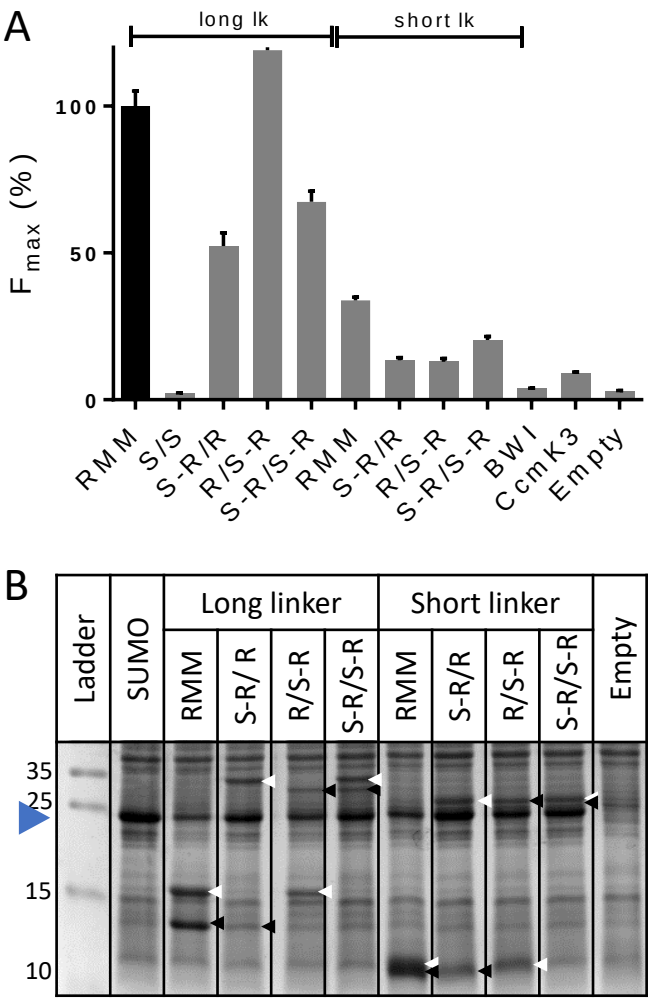


Figure 29. Fluorescence reconstitution with assembly-impaired SUMO-RMM.
A. tGFP assay on cells expressing the RMM pair or different combinations of SUMO-RMM (S-R) with RMM (R), connected to the GFP10/11 with either the Lk30/27 (long linker) or Lk1/1 (short linker). Of note, the SUMO domain is fused to RMM N-terminus. F_{max} values are given as the percentage of the value of RMM pair reference with a Lk30/27 (black bar). BWI and CcmK3 controls had the Lk30/27. **B.** Expression of the constructs assayed in panel A were verified by analysis of the total protein fractions on SDS-PAGE. White and black arrows indicate the POI-GFP10 and POI-GFP11, respectively, while the blue arrow depicts the GFP1-9 migration zone. The empty case is cells transformed with an empty pET26b. Molecular weights are indicated in kDa.

fluorescence apparition with short linker combination, the same constructs were tested with a Lk1/1 (figure 29A).

F_{max} monitored for the RMM/SUMO-RMM pair corresponded to 115% of the reference RMM F_{max} value, whereas the values decreased to 52-67% for the 2 other constructs. These data, which were obtained with the Lk30/27, contrasted with the more severe decrease in GFP signal with RMM/SUMO-RMM carrying the Lk1/1 (50% of RMM pair with Lk1/1 signal). Other sumoylated RMM combinations with the Lk1/1 exhibited 38-60% of the fluorescence level of the RMM pair with Lk1/1.

In order to conclude on these results, expression of the different GFP-tagged constructs was analysed by SDS-PAGE (figure 29B).

Several points were to be noted. First, expression of sumoylated RMM cases was lower compared to the non-sumoylated RMM pair, with long linkers as well as with the shortest. Second, there was an imbalance between the GFP10- and the GFP11-tagged POIs, except for the reference. Indeed, the POI-GFP10 was always predominant.

Unfortunately, with these data, no final conclusion could be given because if lower protein quantities could explain a lower F_{max} for SUMO-RMM/RMM and SUMO-RMM/SUMO-RMM, the increased signal of RMM/SUMO-RMM, in parallel to a lower protein expression, remained enigmatic. As for the GFP10/11 specie imbalance, this may be explained by the POI order within the operon (here, experiments were performed with constructs involving a bicistronic mRNA encoding both POIs). Indeed, it was shown that gene order within an operon influences protein expression with top position gene becoming the more translated protein (Gerngross *et al*, 2022; Lim *et al*, 2011).

Cells expressing the GFP-tagged and SUMO-fused RMM pairs were inspected in TEM to ascertain that macrostructure assembly was also interrupted in these cases (figure 30). Assembly were clearly impeded in all combinations involving SUMO domains compared to the non-sumoylated RMM pair. Surprisingly, a polar aggregate was observed inside cells overexpressing SUMO-RMM/RMM and SUMO-RMM/SUMO-RMM pairs and occasionally with RMM/SUMO-RMM.

Overall, data confirmed that SUMO domain fusion prevented macrostructure formation. In that manner, probably the entirety of the GFP signal with sumoylated RMM is arising from intra-hexamer interactions. While GFP-tagged RMM homo-pair was forming nanotubes in *E. coli* cytosol, its fluorescence was relatively similar to RMM/SUMO-RMM. This fact, along with a greater expression level for the RMM pair, hinted at a pool of BMC-H not participating in the GFP signal, most probably the pool of hexamers involved in nanotube formation. Indeed, it seemed likely that the BMC-H embedded in these structures could not reconstitute the GFP because inaccessible. Thus, in the

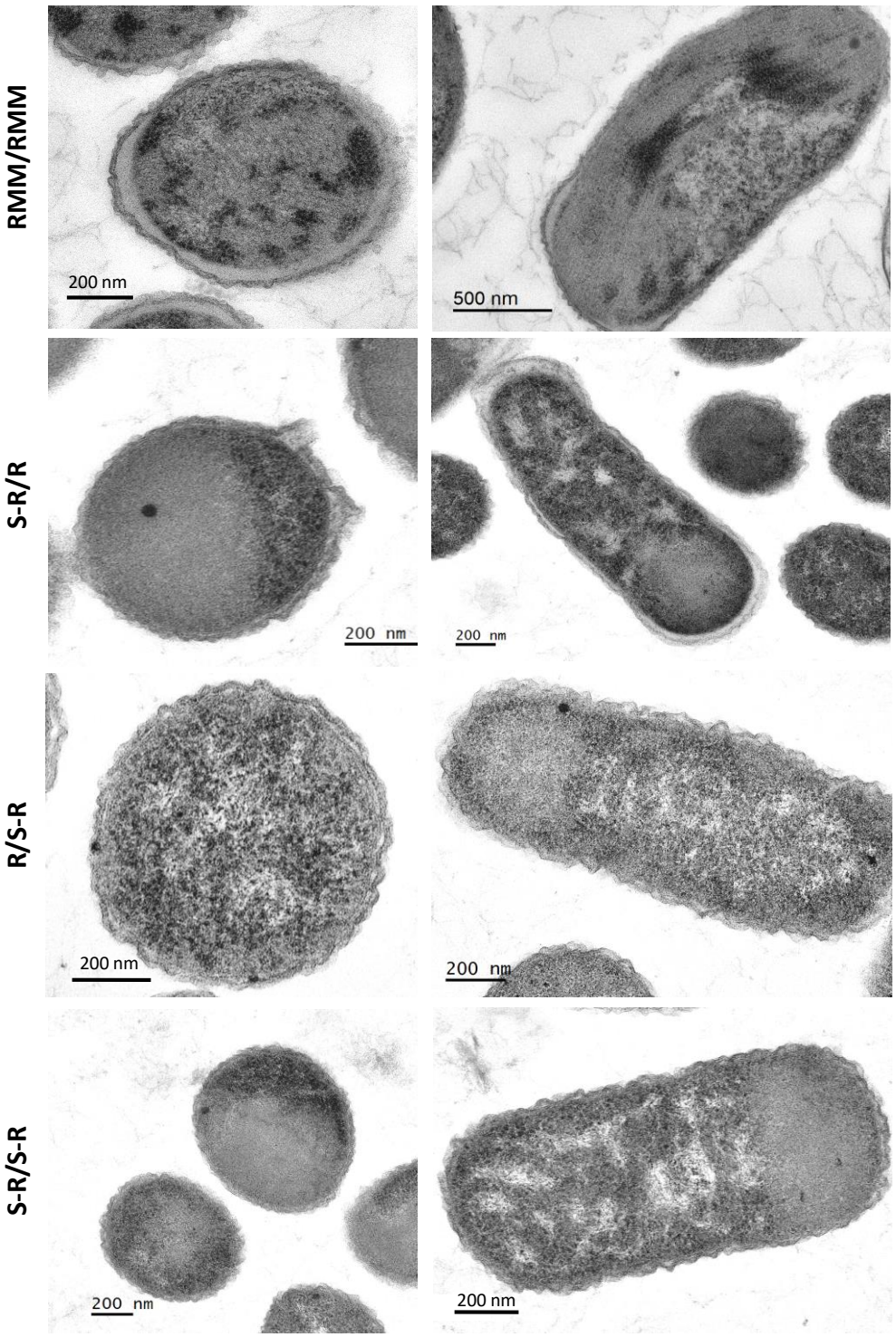


Figure 30. Prevention of macrostructure formation in tGFP by the SUMO domain.

TEM observations of cells over-expressing the RMM pair or different combinations of SUMO-RMM (S-R) with RMM (R), connected to the GFP10/11 with the Lk30/27. Of note, the SUMO domain is fused to RMM N-terminus. The GFP1-9 partner is also expressed in all cases.

context of long linkers, the inter-hexamer assembly leading to macrostructure formation was not a major contributor to the GFP signal. On the contrary, for Lk1/1, the significant decrease in fluorescence for sumoylated RMM showed that inter-hexamer assembly partially participated in the GFP reconstitution. This explained the delayed fluorescence profile of the Lk1/1 case (figure 27C) for which sufficient protein accumulation was necessary prior to fluorescence development.

1.2.3. tGFP assay with partners encoded on 2 independent vectors

POIs coded on 2 compatible vectors

Molecular biology efforts to construct a PPI library can be considerably reduced by expressing the POIs to be tested on 2 independent vectors, compared to using a single vector. Thus, to determine whether this strategy would be adequate to study BMC-H interactions, *E. coli* BL21(DE3) were transformed with different combinations of compatible vectors: (1) a pACYC coding for the POI-GFP10 and the GFP1-9 plus a pET15b carrying the POI-GFP11 or (2) a pACYC with the POI-GFP11 alongside the GFP1-9 plus a pET26b coding for the POI-GFP10. Combinations were compared in fluorescence to a pET26b coding the 3 partners of the tGFP on independent transcripts (figure 31A).

Surprisingly, the RMM pair remained at the same fluorescence level as the negative controls (BWI and CcmK3) when expressed from both 2-vector combinations. On the contrary, a strong signal occurred when RMM expression was carried out from a single vector.

BMC-H have extended patches of hydrophobic residues in their intra-hexamer interfaces. Besides, it was demonstrated that plasmids cluster in bacterial cytoplasm according to their replication origin (ORI) (Ho, 2002). Indeed, λ-P1, pOX38 and RK2 plasmids which bear different ORIs, formed independent *foci* in fluorescence *in situ* hybridization experiments. Then, one possibility for this lack of fluorescence in the 2-vector strategy would be that BMC-H cannot travel the distance between plasmid clusters due to their low solubility as monomers and prefer to form homo-tagged hexamers.

To test this hypothesis, I included soluble interacting proteins in the tGFP assay: the Im9/E9 couple (*E. coli* immunity protein 9 and colicin endonuclease 9) (Garinot-Schneider *et al*, 1996) and leucine zipper domain K1coil/E1coil pair (Tripet *et al*, 1997) (figure 31A). Of note, E9 was inactivated thanks to His575Ala mutation on the active site to prevent any toxicity resulting from its endonuclease activity in case of E9/Im9 stoichiometric imbalance.

Surprisingly, fluorescence of these positive PPI pairs remained also at the negative control level except when they were expressed from a single vector.

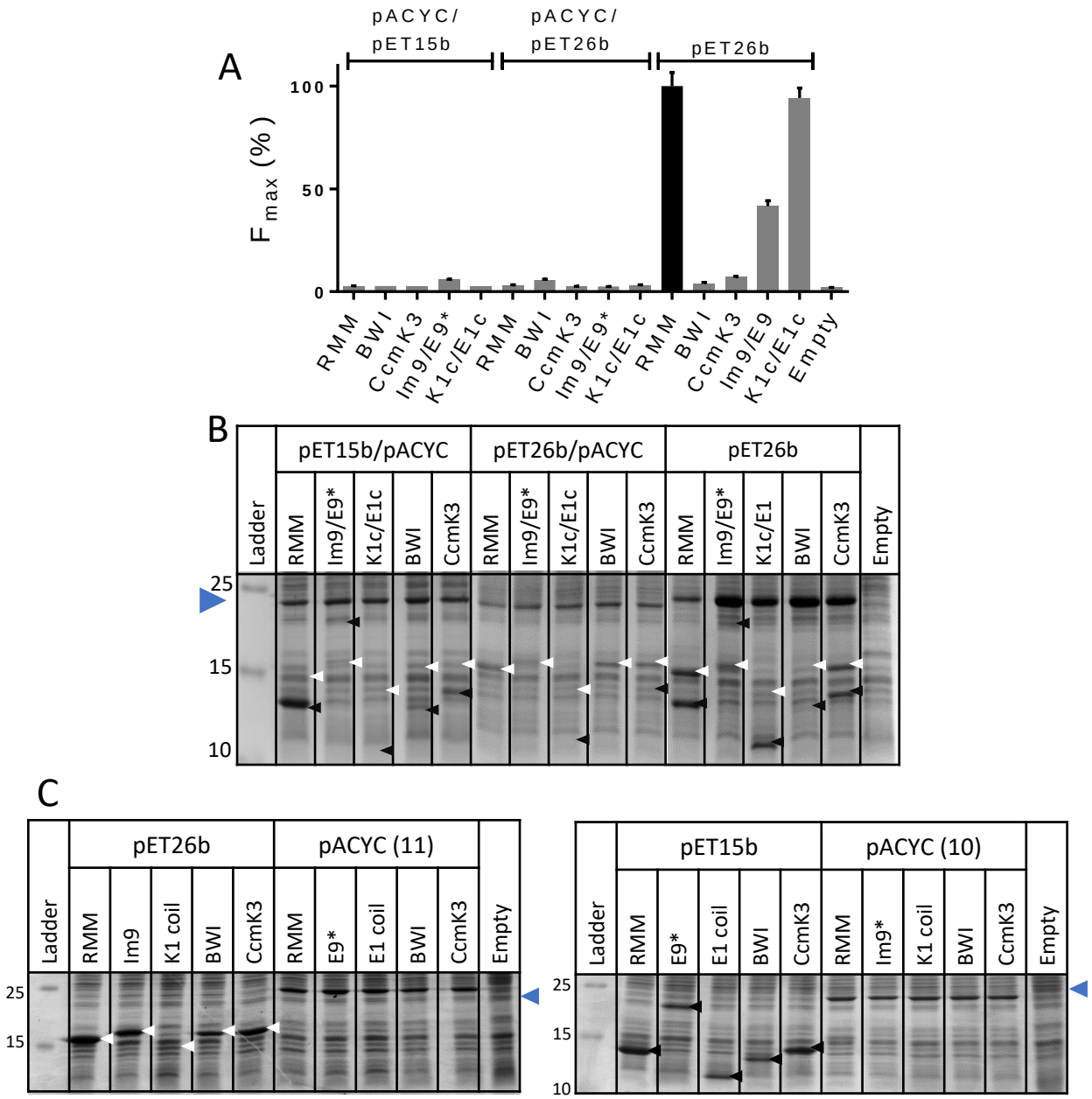


Figure 31. tGFP signal and partner expression decrease when using the 2-vector strategy. POIs were coded on 1 single vector bearing also the GFP1-9 (pET26b) or on 2 separate vectors: either a combination of a pACYC carrying the GFP1-9 and the POI1-GFP10 plus a pET15b for the POI2-GFP11 (pET15b/pACYC), or a combination of pET26b for the POI1-GFP10 together with a pACYC coding for the GFP1-9 and the POI2-GFP11 (pET26b/pACYC). **A.** tGFP assay using the 1- or 2-vector strategy. A single name is given when the same POI is fused to GFP10 and GFP11 tags, otherwise the 2 POIs are indicated, separated by a slash. F_{max} values are given as percentages of the value measured for the RMM homo-pair coded in a single vector (black bar). **B.** Protein expression verification with total protein extracts for the cases studied in panel A on SDS-PAGE. **C.** The vectors used in the 2-vector strategy were transformed individually in *E. coli* and total protein fractions were analysed on SDS-PAGE. For a more comprehensive analysis, profiles from the vectors that were combined in the 2-vector strategy are juxtaposed in gels. White and black arrows indicate the approximate migration of the POI-GFP10 and POI-GFP11, respectively, whereas the blue arrow is for the GFP1-9 position. The molecular weights are given in kDa.

Pull-down effect of the GFP1-9

One striking phenomenon when comparing protein expression from the 2-vector different combinations was that the POI coded alongside the GFP1-9 was virtually absent on the SDS-PAGE gels whereas both POIs were visible in the 1-vector strategy (figure 31B).

To sustain such observation, cells transformed independently with each plasmid from the 2-vector strategy (pACYC(10) or pACYC(11) or pET15b or pET26b) were analysed by SDS-PAGE (figure 31C).

A clearer-cut view was obtained: only POIs from the pET15b or pET26b (*i.e*. plasmids without the GFP1-9) showed high expression patterns. Although the differences in protein expression might partially originate from the lower copy number of pACYC compared to pET vectors (10 *vs.* 15-20 copies, respectively), all the data collected pointed to a deleterious effect of the GFP1-9 on adjacent POI viability.

Indeed, the GFP1-9 is known to be poorly soluble on its own (Park *et al*, 2022) despite protein engineering attempts to improve its folding and solubility. Then, it is possible that the GFP1-9 interferes with the GFP tag of adjacent POI (POI encoded on the same plasmid). In absence of the POI interacting partner, that would cluster in another cytosolic sub-localization due to a different plasmid ORI, the partial reconstitution of the GFP would not be stabilized. This would cause the POI to precipitate along with the GFP1-9 and to be subsequently degraded.

This hypothesis was further explored by analysing a new set of protein couples. Additional control cases were constructed using the 1-vector strategy: besides Im9/E9, close homolog Im2 was assayed with E9 that bore a His575Ala mutation on its catalytic site to prevent its endonuclease activity (E9*). The CutA (cutinase A from *T. thermophiles*) and CobT (cobalamin adenosyl-transferase from *P. horikoshii*) positive pairs were also included. As for the negative cases, PIH1D1 N-terminal domain (PIH1D1-N from *H. sapiens*), nanobody VHH (nanobody from *C. dromedaries*), Smt3 (SUMO domain from *S. cerevisiae*) and K1coil homo-pairs were constructed.

The tGFP assay validated their PPI status (figure 32A).

In parallel, expression of each case was assessed in presence of the GFP1-9 or in absence (figure 32B). Indeed, the same couples were constructed in a 1-vector pET26b lacking the GFP1-9 coding sequence.

Unexpectedly, while all positive couples were visible in SDS-PAGE, no band could be noticed for the negative cases when the GFP1-9 was also expressed. This result completely changed when the GFP1-9 was retrieved from the tGFP vectors. In this context, clear bands could be visualised for almost every case. Furthermore, positive control expression was also increased in absence of the GFP1-9. This

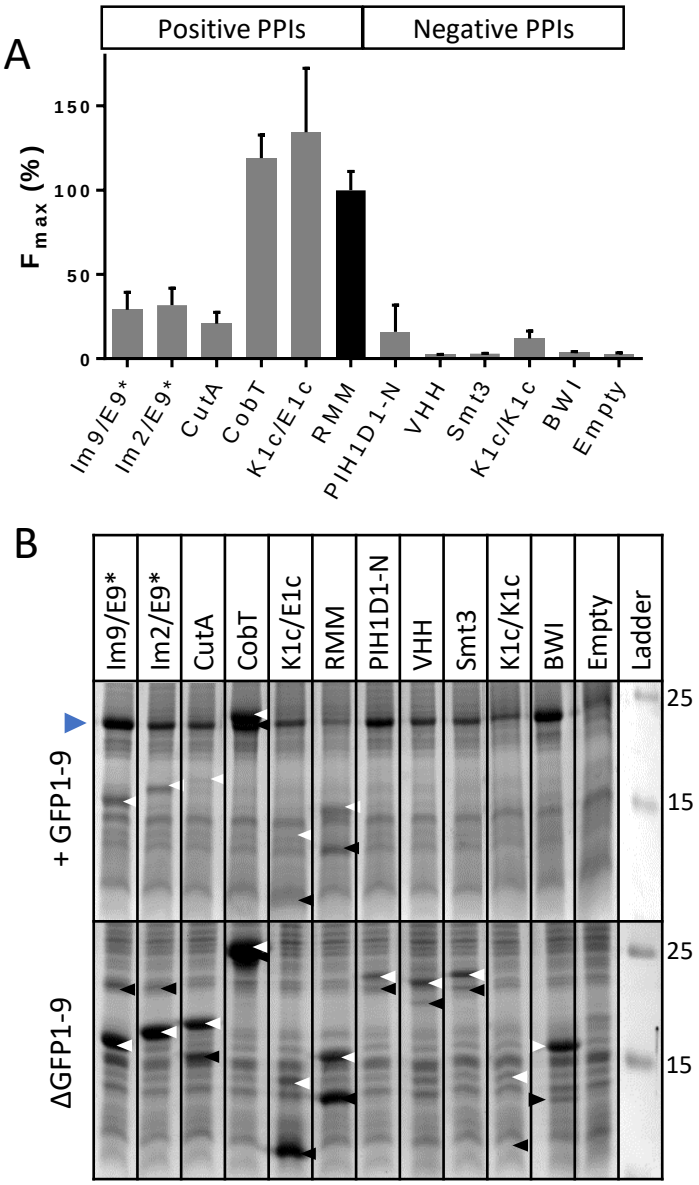


Figure 32. Evaluation of the tGFP robustness in detecting PPIs.
A. tGFP assay on positive (Im9/E9* or Im2/E9*, CutA, CobT, K1coil/E1coil and RMM) or negative PPI controls (PIH1D1-N, VHH, Smt3, K1coil/K1coil or BWI) coded on a single pET26b. F_{max} values are provided as percentages of the RMM reference pair (black bar). A single POI name is provided for homo-pairs, otherwise the 2 component names are given, separated by a slash. **B.** Pull-down effect of the GFP1-9 on POIs. Total protein fractions of cells expressing all 3 tGFP partners (top) or only both GFP-tagged POIs (bottom) from a single vector were collected after an 8h IPTG induction and analysed by SDS-PAGE. White arrows point at the POI-GFP10, whereas black arrows identify the POI-GFP11. The blue arrow notifies the GFP1-9 band. The molecular weights are given in kDa.

withdrawn the possibility that the lack of negative pair expression was due to a problem of expression and/or folding of one of the POI in the couple.

Globally, this demonstrated again the detrimental impact that the GFP1-9 can have on POI partners, especially on negative couples for which the GFP1-9 reduced to practically zero the expression levels. This effect might raise difficulties when it comes to verify if tGFP assay results were negative because of a lack of protein expression or because of a negative PPI. On the other hand, this fact could be viewed positively as the GFP1-9 pull-down effect would clear out non-interacting partners, thus avoiding unspecific GFP reconstitution due to POI accumulation.

Pull-down rescue with vectors sharing the same ORI

To further verify the hypothesis of the GFP1-9 pull-down effect, rescue experiments were undertaken in the 2-vector strategy. As a reminder, it was shown that plasmids might cluster according to their ORI in the cytosol (Ho, 2002). Then, in order to force the spatial proximity of the 2 vectors, plasmids sharing the same ORI were selected for the tGFP assay: a pET26b coding for the GFP1-9 and POI-GFP10 and the pET15b used previously which coded the POI-GFP11. In principle, these plasmids are considered incompatible because they are genetically instable. Indeed, as they have the same ORI, the cell machinery recognizes both plasmids without distinction and during subsequent cell division, one of them can be lost due to uneven partition. Nevertheless, as each vector was bearing a different resistance cassette (kanamycin^R and ampicillin^R), concomitant use of both antibiotics favored maintenance of both vectors during the tGFP assay.

The RMM, BWI, CcmK3, Im9/E9 and K1coil/E1coil pairs were assayed with this set-up (figure 33A). Fluorescence signals were recorded for all positive pairs in the 2-incompatible-vector strategy.

In this configuration, RMM homo-pair and K1coil/E1coil reached nearly 30% of RMM F_{max} value in the 1-vector mode and 14% for Im9/E9*.

To verify whether these discrepancies in GFP signals were due to a difference in protein expression, total protein fractions were collected after a 16h IPTG induction and analysed by SDS-PAGE (figure 33B).

The same POIs were more expressed when coded from a single vector than from the 2 incompatible vectors.

Thus, bringing together the plasmids used in 2-vector tGFP test succeeded in rescuing the GFP1-9 pull-down effect. However, resulting protein expression was lower than in the 1-vector strategy and affected the F_{max} value.

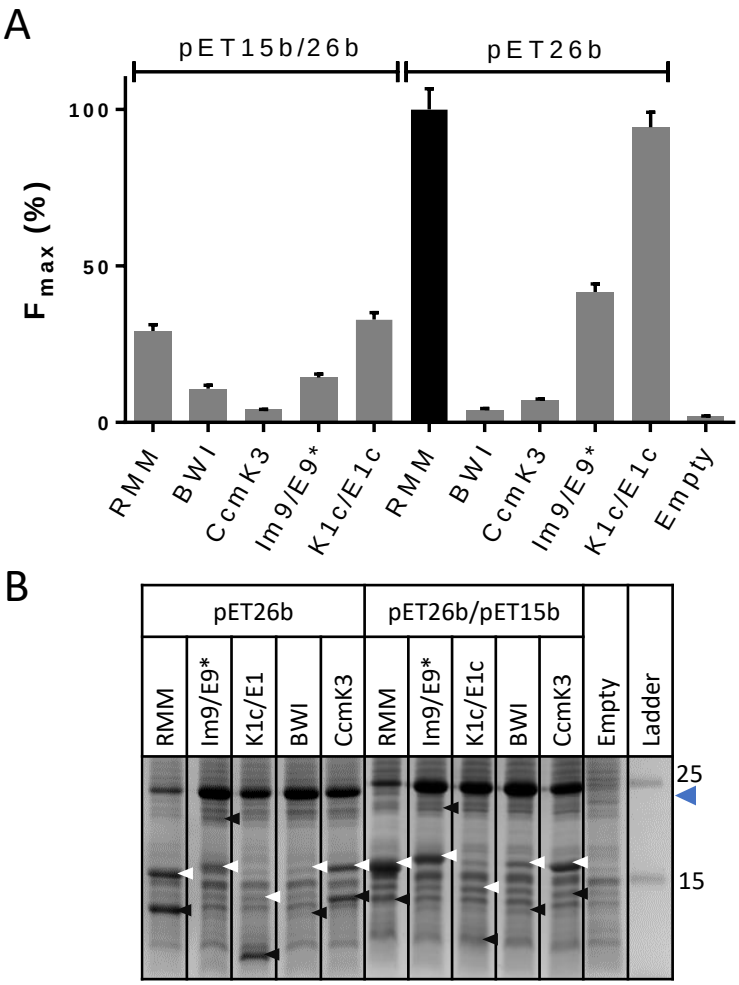


Figure 33. Forcing vector proximity to circumvent the GFP1-9 pull-down effect.
A. tGFP assay performed on cell overexpressing a combination of a pET26b carrying the GFP1-9 and POI1-GFP10 and a pET15b with the POI2-GFP11 (pET15b/pET26b) or a single pET26b vector coding all 3 tGFP partners. F_{max} values are provided as percentages of the RMM reference pair coded by a single vector (black bar). A single POI name is given when the same POI is fused to the GFP10 and GFP11 tags, otherwise the 2 components are separated by a slash. **B.** Verification of POI expression. Total protein fractions of cells expressing the same constructs as in panel A were analysed on SDS-PAGE. White and black arrows point at the POI-GFP10 or POI-GFP11, respectively, while the blue arrow notifies the GFP1-9 band. The molecular weights are given in kDa.

Someone willing to use the 2-vector strategy should consider the possibility to code both POIs on the same vector while the GFP1-9 transcription is maintained on an independent vector as it was performed in the original study (Cabantous *et al*, 2013). Unfortunately, this would not decrease molecular biology efforts required to construct future BMC-H pair library, which was my initial objective. Besides, as the 1-vector strategy showed great success in tGFP assay, I decided to pursue this path and to optimize it further to the study of BMC-H interactions.

1.2.4. Impact of the genetic organization of the tGFP partners

BMC proteins are generally coded within one single operon (Rae *et al*, 2013; Chowdhury *et al*, 2014), with some exceptions as CcmK3 and CcmK4 from most β-cyanobacteria, which are present in a genetic *locus* remote from the rest of the β-CBX main *locus*. Furthermore, genetic organization was shown to improve the efficiency of protein complex assembly (Wells *et al*, 2016). Indeed, oligomerization efficiency is enhanced if proteins are translated from the same mRNA (Shieh *et al*, 2015; Bertolini *et al*, 2021).

In light of these 2 facts, the optimal genetic organization scheme for studying BMC-H oligomerization was sought. Three different constructs were built, leading to (1) all 3 tGFP partners independently transcribed, (2) POIs transcribed on a bicistronic mRNA while the GFP1-9 remained independent or (3) all 3 partners on a single tricistronic mRNA (figure 34A).

When these constructs were assayed in tGFP, a strong fluorescence was recorded with all organizations involving the GFP10/11-tagged RMM pair (figure 34B). Signal resulting from the CcmK3 pair remained weak with the 3 organizations, slightly above cellular auto-fluorescence. On the contrary, while BWI pair fluorescence was low for both the independent and bicistronic transcripts, a significant signal increase occurred when the pair was transcribed on a tricistronic mRNA (approximately 55% of tricistronic RMM).

These data suggested that transcription of the 3 tGFP partners on the same mRNA and subsequent proximate translation boosted unspecific GFP reconstitution, *i.e.* random encounters between the GFP1-9, GFP10 and GFP11. Hence, a biosynthetic *locus* coding for all the 3 tGFP partners should be avoided. As for selecting between genetic organizations giving rise to independent transcripts or to a bicistronic mRNA, the bicistron was preferred for the rest of my thesis on the basis that it resembles more BMC-H natural operon organization and transcription fashion.

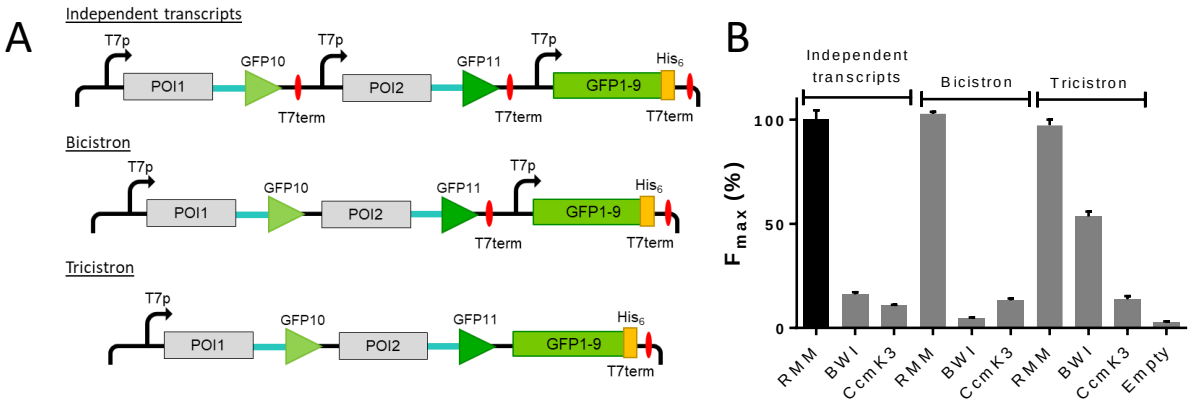


Figure 34. Effect of the genetic organization of the tGFP partners on tGFP reconstitution.
A. tGFP vectors were designed to give rise either to 3 independent mRNAs coding for the POI1-GFP10, POI2-GFP11 and the GFP1-9 (independent transcripts), or to 2 distinct mRNAs, one coding for both POIs and the second for the GFP1-9 (bicistron), or to a single mRNA encoding the 3 tGFP partners (tricistron). T7p: promoter T7; T7term: terminater T7. **B.** tGFP assay on cell expressing one of the 3 constructs presented in panel A. F_{max} are given as percentages of the value obtained for the independently transcribed RMM pair (black bar).

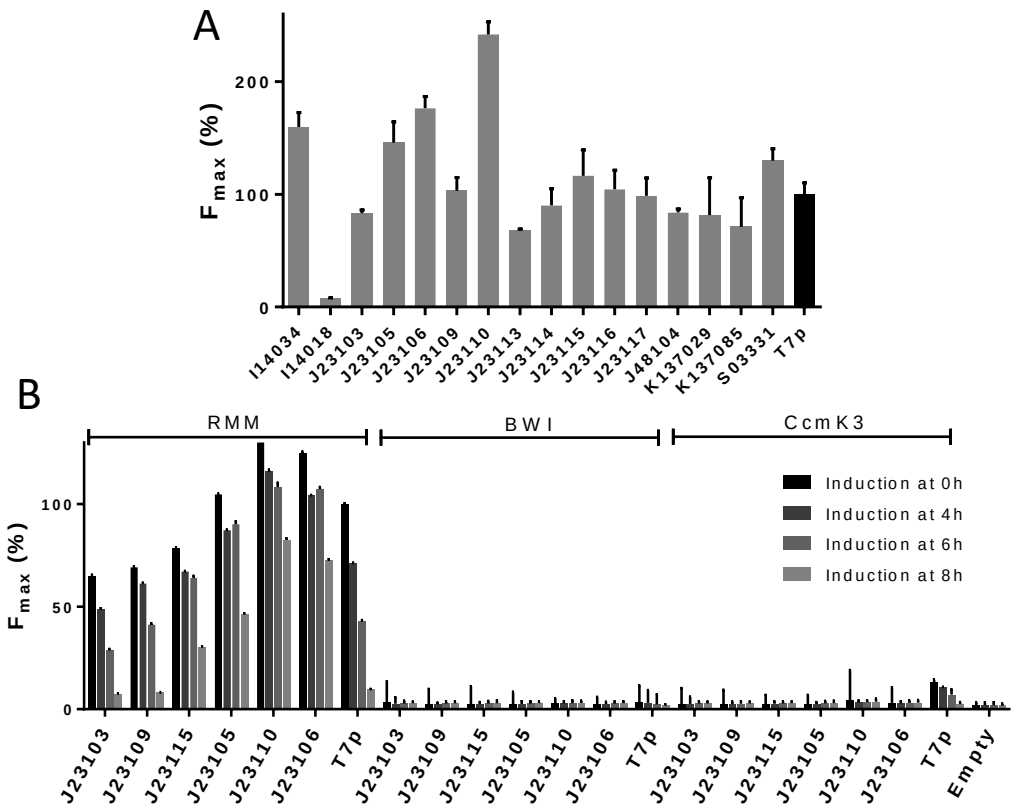


Figure 35. Promoter control over the transcription of the tGFP partners.
A. Evaluation of the expression strength of the constitutive promoters (CP) with RMM homo-pairs compared to the T7 promoter (T7p). In all cases, the GFP1-9 transcription remained under the control of a T7p and IPTG induction was performed from the beginning of the tGFP assay. **B.** Selected CPs were assayed in tGFP with RMM or BWI and CcmK3 negative controls. IPTG was added at 0, 4, 6 or 8h of culture to induce an uncoupled POIs/GFP1-9 expression. In panels A and B, F_{max} are given as percentages of the T7p-controlled RMM pair value induced at 0h.

1.2.5. Control over the expression of the tGFP partners

Uncoupling the production of the 2 GFP-tagged POIs from the GFP1-9 expression might permit to reduce the frequency of random encounter events between the GFP1-9 and the GFP10/11 tags. Besides, this could have another beneficial effect by limiting the suspected GFP1-9 pull-down of individual POI prematurely bound to it.

In the bicistronic constructs, the T7p controlling the POI expression was changed for a constitutive promoter (CP) while the GFP1-9 remained under the control of T7p. A set of 16 different CPs were selected from the iGEM promoter repertoire. Their expression strength was first evaluated in tGFP with the RMM homo-pair and a GFP1-9 induction from the beginning of the screen (figure 35A).

Six CPs which expression strengths were distributed between the T7p and the strongest CP were chosen and implemented with BWI and CcmK3 pairs to examine the effect of postponing the GFP1-9 expression on unspecific signal apparition (figure 35B). In that matter, expression of the GFP1-9 was induced either from the beginning of the assay or after 4, 6 or 8h of fluorescence monitoring.

J23105, J23106 and J23110 CPs led to higher signals than the one of the T7p-controlled RMM pair, whereas lower signals occurred with J23115, J23109 and J23103. Besides, RMM F_{max} values decreased progressively when increasing the delay of the GFP1-9 induction and this was more prominent for the T7p RMM reference. For an induction after 8h, the F_{max} was approximately 10% of the same condition induced from the beginning. Surprisingly, a strong drop also manifested with the weakest CPs when postponing the GFP1-9 expression. Indeed, when the GFP1-9 was induced after 8h of culture, the J23109 and J23103 promoters only reached around 16% of their fluorescence when induced from the beginning (7% of T7p-controlled RMM reference induced at 0h).

Insufficient GFP1-9 production or decline in the cellular resources were ruled out as arguments to explain such signal drop because significantly higher fluorescence could be reconstituted with stronger CPs when inducing after 8h of culture (73 to 83% of T7p RMM reference induced at 0h). Regarding the negative controls, when the GFP1-9 induction was performed from the beginning of the culture or after 4, 6 or 8h of culture, the BWI pair fluorescence remained unchanged with CPs compared to the inducible T7p. On the contrary, there was a slight decrease for the CcmK3 pair signal, showing a possible advantage in uncoupling the GFP1-9 expression from POI interactions.

If the signal drop obtained when postponing cell induction was justified for the T7p-controlled RMM as both the GFP1-9 and the POI expression were delayed, this was unexpected for the weakest CPs. Such observation could be interpreted as indicative of inter-hexamer assembly which would act as a molecular sink, absorbing freely-diffusing RMM hexamers. As we saw when RMM was sumoylated (see section 1.2.2), comparison of fluorescence signals to the expression levels indicated that a

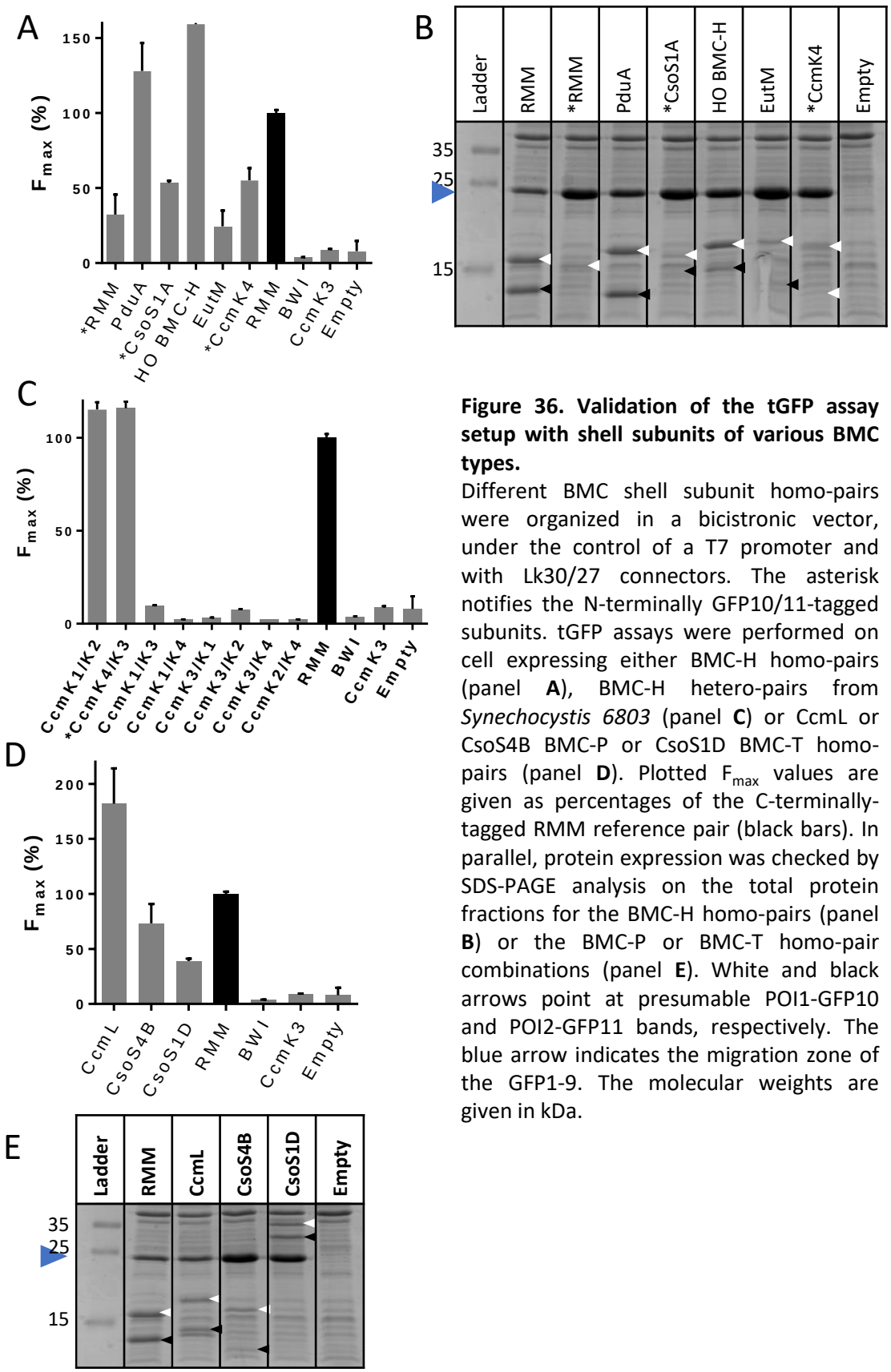


Figure 36. Validation of the tGFP assay setup with shell subunits of various BMC types.

Different BMC shell subunit homo-pairs were organized in a bicistronic vector, under the control of a T7 promoter and with Lk30/27 connectors. The asterisk notifies the N-terminally GFP10/11-tagged subunits. tGFP assays were performed on cell expressing either BMC-H homo-pairs (panel **A**), BMC-H hetero-pairs from *Synechocystis 6803* (panel **C**) or CcmL or CsoS4B BMC-P or CsoS1D BMC-T homo-pairs (panel **D**). Plotted F_{max} values are given as percentages of the C-terminally-tagged RMM reference pair (black bars). In parallel, protein expression was checked by SDS-PAGE analysis on the total protein fractions for the BMC-H homo-pairs (panel **B**) or the BMC-P or BMC-T homo-pair combinations (panel **E**). White and black arrows point at presumable POI1-GFP10 and POI2-GFP11 bands, respectively. The blue arrow indicates the migration zone of the GFP1-9. The molecular weights are given in kDa.

substantial portion of GFP-tagged RMM might not participate in the tGFP signal (figure 29) and this portion was proposed to be the hexamers committed to nanotube formation. Then, for all the CPs, when inducing the GFP1-9 expression concomitantly to POIs, the GFP1-9 would get to label intra-hexamer interactions before hexamers commit to nanotube formation. On the contrary, when uncoupling both things, the hexamers that were produced before the GFP1-9 expression might already be embedded within nanotubes, impossible to be labelled thereafter. This phenomenon might be overcome in strong CPs that overproduced BMC-H. Indeed, quantity of new hexamers might still be produced when the GFP1-9 is induced. As for the weak CPs, the majority of the hexamers would be embedded in nanotubes (then unlabelable) and new production would be too low to observe a fluorescence increase in the same timeframe.

Uncoupling POI interactions from GFP1-9 production did not achieve the significant improvements I expected, probably because the GFP1-9 pull-down effect already maintained unspecific GFP reconstitution events at a very low apparition frequency. Then, for the rest of my thesis, I decided to keep the POIs under the control of a T7p which granted me with a better control over the beginning of the tGFP assay.

1.3. Validation of the assay parameters with BMC shell components

1.3.1. Homo-hexamer associations

In order to validate the use of the tGFP for BMC-H interaction study, well-described BMC-H homo-pairs were put under the test with our adapted tGFP set-up: PduA from *S. enterica*, CsoS1A from *H. neapolitanus*, HO BMC-H from *H. ochraceum*, EutM from *E. coli* and CcmK4 from *Synechocystis 6803*. These cases were constructed in the 1-vector mode, with a genetic organization giving rise to a bicistronic mRNA coding both POIs connected to GFP10 or GFP11 tags by a Lk30 or Lk27, respectively. Of note, preliminary results pointed out a deleterious effect of a C-terminal tag orientation for CsoS1A (data not published) and CcmK4 (Garcia-Alles *et al*, 2019). Therefore, GFP tags were placed on the N-terminus for these 2 cases. In parallel, RMM was also constructed with N-terminal tags (*RMM where the asterisk depicts the tag orientation) to benchmark the impact of tag orientation on BMC-H interactions and tGFP assay (figure 36A).

All homo-pairs revealed to be positive in tGFP. Signals were especially strong with PduA and HO BMC-H. On the contrary, fluorescence deriving from EutM pair overexpression remained low, albeit above the threshold level established by the BWI and CcmK3 negative controls or cellular auto-

fluorescence measured for empty vector-transformed cells. Finally, switching tag orientation in RMM caused a 4-fold drop in fluorescence.

Discrepancies in POI expression evidenced in SDS-PAGE partially explained tGFP signal variations (figure 36B). However, while more than 10-fold protein expression differences were noticed between CsoS1A or CcmK4 and PduA or RMM, this only reverberated in 2- to 2,5-fold differences in fluorescence signal. A similar observation could be made for *RMM which underwent a 4-fold drop compared to C-terminally-tagged RMM although its expression pattern was almost invisible in SDS-PAGE.

Once more, this suggested that a portion of the hexamer pool escaped labelling by the GFP1-9 when BMC-H are overexpressed.

1.3.2. Hetero-hexamer associations

A genomic survey estimated the existence of an average of about 3,5, 1,4 and 1,2 gene copies coding for BMC-H, BMC-T and BMC-P, respectively, per organism (Axen *et al*, 2014). Besides, previous studies reported that a simultaneous expression of CcmK homologs could result in the formation of CcmK1/CcmK2 and CcmK3/CcmK4 hetero-hexamers (Sommer *et al*, 2019; Garcia-Alles *et al*, 2019). Of note, BMC-H partners were co-purified by affinity tag-mediated purification of specific BMC-H and identified through western blot analysis.

To further validate the tGFP set-up for the study of hetero-hexamer formation, I applied the technology to explore already characterized *Synechocystis 6803* CcmK1, CcmK2, CcmK3 and CcmK4 cross-interactions. Using the same construct organization as in the previous section, combinations of CcmK1, CcmK2, CcmK3 and CcmK4 homologs were created and assayed in tGFP (figure 36C).

Remarkably, the tGFP assay succeeded in reproducing published data (Garcia-Alles *et al*, 2019). Indeed, high fluorescence signals were measured exclusively for the *CcmK4/CcmK3 and CcmK1/CcmK2 pairs, with calculated F_{max} that were even higher than that of the RMM reference pair. Moreover, as reported, CcmK4/CcmK3 signal was dependent on tagging orientation: the couple F_{max} value dropped to background levels when the GFP10 tag was transferred from CcmK4 N-terminus to its C-terminal side.

Altogether, these data demonstrated that the tGFP technology is well-suited for the identification of BMC-H interactions. In particular, the tGFP set-up permitted to assess the formation of hetero-hexamers.

Over 8 different BMC-H combinations, only 2 produced a fluorescence signal. This indicated that POI involved in the positive pairs were well expressed (CcmK1-GFP10, CcmK2-GFP11, GFP10-CcmK4 and CcmK3-GFP11). The same POIs gave negative results in other combinations. Albeit it was not

possible to verify the protein expression in tGFP due to the GFP1-9 pull-down effect, it seemed reasonable to think that they were also expressed in these negative combinations prior to be pulled down and degraded. Thus, the absence of fluorescence would be linked to an absence of interactions, demonstrating that the tGFP assay highlighted only positive PPI. Then, the method results were trustworthy.

β-CBX interacting pairs CcmK1 and CcmK2 are part of the main *ccm* operon while CcmK3 and CcmK4 are found in a satellite *locus* (Rae *et al*, 2013). As mentioned earlier, BMC-H are unlikely to diffuse alone, as monomers; they need to oligomerize in order to shield their hydrophobic interfaces and become stable in solution. Thus, presumably, in the tGFP assay where genetic *locus* distances are abolished, every paralog could interact with its counterparts. Yet, only the BMC-H originally included within the same *loci* were able to cross-interact.

This could suggest the existence of different co-evolutionary constraints imposed on the main *ccm* operon or on satellite *loci*, pointing to the possibility that CcmK3 and CcmK4 could play auxiliary functions that would apply only under certain conditions.

Taken together, these data suggested that BMC organization into operons is of foremost importance for BMC-H interactions. Furthermore, 2 BMC-H genetic proximity could hint at more probable interacting partners as demonstrated by means of statistical analyses (Wells *et al*, 2016).

1.3.3. The tGFP is amenable to study all shell component interactions

Three different classes of proteins constitute the BMC shell: BMC-H, BMC-T and BMC-P. Although, this was not part of my thesis objectives, the tGFP validation tests were extended to other BMC shell components.

To determine whether the tGFP could also be fit for the study of these subunit interactions, BMC-P homo-pairs of CsoS4B from *H. neapolitanus* and CcmL from *Snechocystis 6803* along with a BMC-T homo-pair composed of CsoS1D, also from *H. neapolitanus*, were monitored in tGFP and compared to the RMM reference pair (figure 36D).

Pentameric CcmL from the β-CBX resulted in the highest F_{max} value (approximately twice as high as the RMM value). Substantial fluorescence also emerged with CsoS4B from the α-CBX. However, trimeric CsoS1D signal was lower (38% of the RMM value) although still superior to threshold values established by negative controls.

To explain such variations in the GFP signal, protein expression was analysed by SDS-PAGE (figure 36E).

While the CsoS1D pair was expressed in quantity equivalent to CcmL, an approximate 4,6-fold difference was observed in fluorescence. More surprisingly, CcmL band intensities were very slightly marked compared to RMM and yet, its F_{max} value was practically doubled.

In this context, 2 explanations could be valid: (1) in the same extent that RMM fluorescence level contrasted with its expression level, hinting at inter-hexamer assembly, the decrease of CsoS1D fluorescence might indicate inter-trimer assembly, similarly to what had been described for the nanotube-forming trimer PduB (Uddin *et al*, 2018) or (2) unequally distributed GFP10- and GFP11-tagged POI within a trimer might lead to homo-labelled trimers lost for the tGFP assay.

For the latter, probabilities indicated that up to 25% of the trimers would be GFP10 or GFP11 homo-labelled. Contrasting with this number, only 6% of the pentamers and 3% of the hexamers could not participate in the tGFP assay by such phenomenon (when considering at least one GFP reconstituted by oligomer). Homo-tagged trimers might also be enriched by the association of monomers emerging from adjacent ribosomes acting in *cis* on the same mRNA (Bertolini *et al*, 2021), a phenomenon that could be enhanced by the bi-domain nature of BMC-T.

To recall, CsoS1D forms a double stack with concave faces oriented towards the interior of the stack (Klein *et al*, 2009). However, contrary to BMC-T such as PduT, CsoS1D is circularly permuted which provokes a switch in protein terminus orientation from the concave to the convex BMC-T side. Then, GFP reconstitution hampering due to sandwiched CsoS1D trimers (hiding GFP tags from the GFP1-9) seemed unlikely.

Chapter 2

Exploration of the cross-interactions between *Klebsiella pneumoniae* BMC-H

2.1. Introduction to *Klebsiella pneumoniae*, a 3-BMC-coding bacterium

2.1.1. A mammal pathogen and a plant mutualist

Klebsiella pneumoniae (*Kpe*), formerly called *Klebsiella variicola*, is a rod-shaped Gram-negative bacterium and facultative anaerobe belonging to the gammaproteobacterium *phylum* and more precisely the enterobacterium family. It has been found in very diverse environments: watercourses, soils, plants and mammals (Bagley, 1985).

Different strains exist, with preferred habitats for each one but the majority were isolated from clinical patient samples. Indeed, *Kpe* is an opportunistic pathogenic bacterium that can provoke severe conditions such as urinary tract infection (from the bladder to the kidney), pneumonia, septicaemia, meningitis or liver abscess (Navon-Venezia *et al*, 2017). Generally, infection involves contaminated food and the gastrointestinal tract as an entryway.

Some *Kpe* strains are multidrug resistant, mainly falling back on efflux pumps and β-lactamase activity to evade antibiotic toxicity. As such, *Kpe* has raised worldwide medical concerns and has been the object of many studies (Navon-Venezia *et al*, 2017).

On the contrary, other strains showed a preference for plant colonization. For instance, *Kpe 342* was originally collected on maize (Chelius & Triplett, 2000). However, it can also be a mammal pathogen. Genome sequencing evidenced that it bears virulence factor-coding genes and mouse model infection assays demonstrated that it was able to cause urinary tract infection and pneumonia although not as virulent as the obligate pathogen *Kpe C3091* (Fouts *et al*, 2008).

Plant colonization by *Kpe 342* occurs without inducing the plant defence systems nor the creation of a symbiotic structure. Rather it spreads homogeneously from roots to shoots (Iniguez *et al*, 2004). *Kpe 342* is a mutualistic endophyte which is able to fix nitrogen and to pass it through to the plant in exchange for shelter and probably a carbon source (Mahl *et al*, 1965; Iniguez *et al*, 2004). Endophytic relationship between wheat and *Kpe 342* was notably shown to provide fitness to the plant, displayed

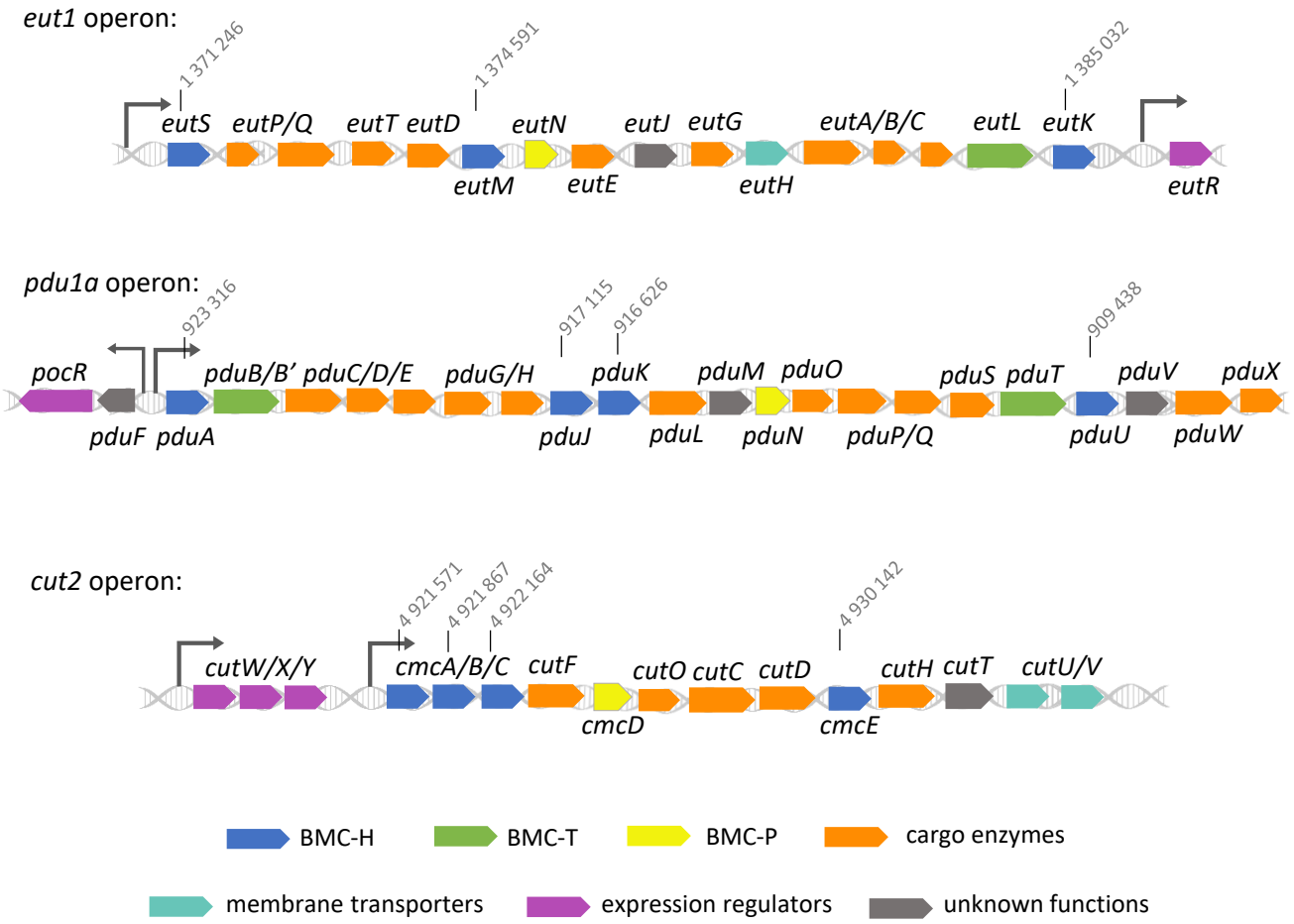


Figure 37. *Klebsiella pneumoniae 342* BMC operons.
Klebsiella pneumoniae 342 codes for 3 different BMCs: the EUT1, PDU1A and GRM2 (*cut2*). Here are presented the genetic organization of each operon. In the *eut1*, 3 BMC-H are encoded: EutM, EutK and EutS. The *pdu1a* bears 4 BMC-H: PduA, PduJ, PduK and PduU. As for the *cut2*, 4 different BMC-H are present: CmcA, CmcB, CmcC and CmcE. Although the main promoter positions are clearly defined, the presence of operon-interned promoters or alternative expression regulation mechanisms such as riboswitches is not excluded. The number indicated above each BMC-H coding sequence represents the nucleotide of *Kpe 342* genome at which their sequence starts.

by taller, greener and stronger seedlings (Iniguez *et al*, 2004). Hence, *Kpe* could be of great interest for future worldwide agricultural use because it could reduce nitrogen fertilizers requirement for efficient crop growth.

2.1.2. Three BMC *loci* are present in *Kpe 342* genome

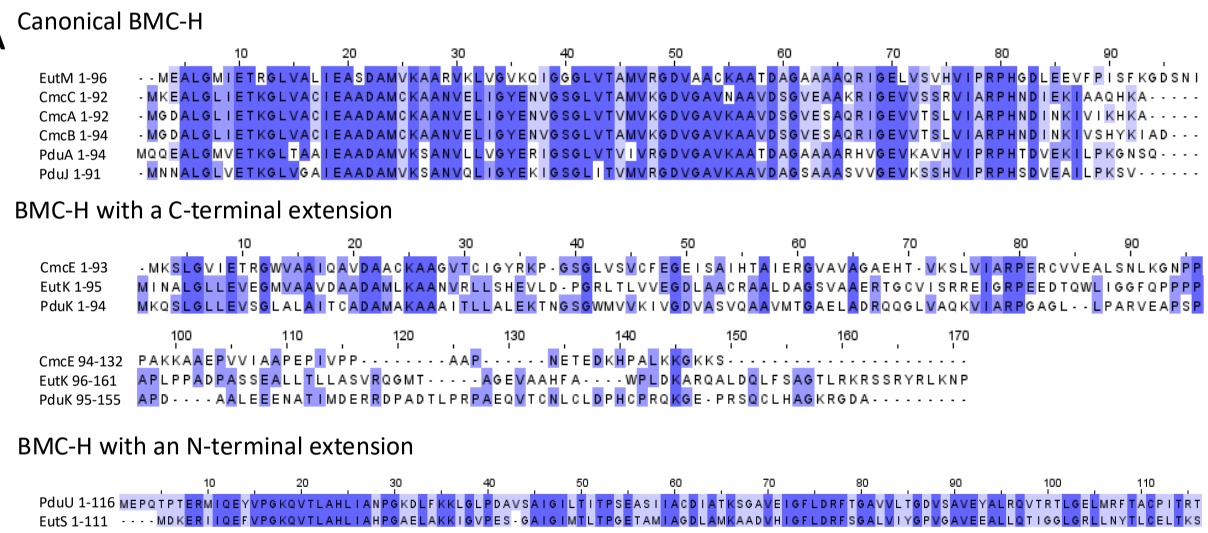
Thanks to whole-genome sequencing and genomic surveys, it was highlighted that *Kpe 342* was endowed with 3 different BMC-coding *loci* (figure 37) (Fouts *et al*, 2008; Axen *et al*, 2014). Indeed, *Kpe 342* codes for the EUT1, PDU1A and GRM2 which, as a reminder, catabolize either EA, 1,2-PD or choline, respectively. These metabolites can result from both mammal and plant cell membrane breakdown. Thus, the BMCs present in *Kpe 342* could grant it with the ability to colonize mammals and plants and live on their by-products. Some close relatives like *Klebsiella oxytoca* also codes for the 3 identical BMCs while other *Klebsiella pneumoniae* subspecies like the clinical strain MGH78578 lack the *cut2* operon coding for the GRM2 (Axen *et al*, 2014).

Despite abundant data on *Salmonella*, *Clostridium* or *Escherichia* EUT and PDU metabolism or shell structure, little is known about the ones of *Klebsiella*. Back in 1976, a study evidenced the utilization of EA in a subgroup of the *Klebsiella* genus (Scarlett & Turner, 1976). EA was shown to trigger the activity of an EA ammonia lyase. Its catabolism was dependent on vitamin B_{12} (cobalamin) presence in the culture medium and led to AA production, similarly as in the EUT. Yet, no further study was undertaken to determine whether EA degradation was BMC-bound and data are still missing.

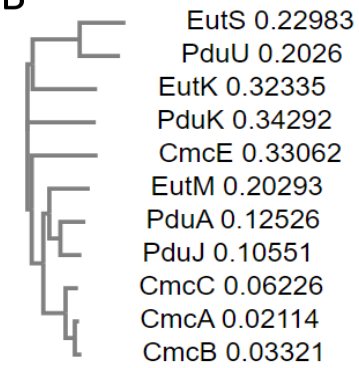
Studies have mainly focused on *Kpe* GRM2. Two teams showed that the *cut2* operon was functional: *Klebsiella* could metabolize choline into TMA through the activity of CutC (Martínez-del Campo *et al*, 2015; Kalnins *et al*, 2015). Of note, the TMA is subsequently transformed into TMA N-oxide (TMAO) which was shown to be involved in inflammation and cardiovascular diseases (Liu & Dai, 2020). Also, simplified versions of the GRM2 BMC were characterized recently (Kalnins *et al*, 2020; Cesle *et al*, 2021).

The *eut1* operon of *Kpe 342* shares the same genetic organization as *Salmonella enterica* and *E. coli*, including the presence of the EutR transcription factor, downstream the operon (figure 37), but differs from the organization in *Clostridium difficile* which contains additional EutV/W regulators (Axen *et al*, 2014; Pitts *et al*, 2012). This indicates that *Kpe 342 eut1* is rather regulated through a EutR-dependent mechanism. Of note, *E. coli* strains classically used in labs for protein expression (K12 and BL21(DE3)) also possess an *eut1* operon.

A



B



C

	CmcA	CmcB	CmcC	CmcE	EutK	EutM	EutS	PduA	PduJ	PduK	PduU
CmcA		95	86	56	42	56	22	63	66	41	26
CmcB			86	44	40	57	22	59	64	38	24
CmcC				42	44	59	22	61	66	40	24
CmcE					27	38	21	37	40	28	26
EutK						42	21	40	45	32	20
EutM							19	63	61	34	22
EutS								20	22	18	56
PduA									77	35	25
PduJ										39	23
PduK											20
PduU											

Figure 38. ***Klebsiella pneumoniae (Kpe) 342*** **BMC-H diversity.**
A. Sequence alignment of *Kpe* BMC-H. Note that 3 groups of BMC-H cluster: canonical BMC-H only constituted by the pfam0936 domain, N-terminally extended BMC-H also called circular permutants or C-terminally extended BMC-H. Both the N- and C-terminally extended BMC-H possess extensions beyond the pfam0936 domain. **B.** Phylogenetic tree of *Kpe* BMC-H. **C.** Sequence homology among *Kpe* BMC-H. Figures in the table are percentages of identity calculated only by taking into account the common pfam0936 domain. Note that if extensions were included in the calculus, the percentage of identity between canonical and extended BMC-H would decrease.

On the other hand, *Kpe 342 pdu1a* operon is homolog to the ones of *Salmonella enterica* and *Citrobacter genus* (figure 37) (Axen *et al*, 2014). Thus, we could rely on such model organisms to make deductions on *Kpe* BMC biology.

Besides, structures of the *E. coli 536* GRM2 BMC-H were elucidated recently which could also give us insight into shell subunit topology (Ochoa *et al*, 2021).

2.1.3. Diversity of BMC-H subunits in *Kpe 342*

The 3 BMC *loci* of *Kpe 342* totalize 11 different BMC-H sequences, namely EutK, EutM and EutS for the *eut1*, PduA, PduJ, PduK and PduU for the *pdu1a* and finally CmcA, CmcB, CmcC and CmcE for the *cut2*. Upon comparison, BMC-H can be divided into 3 groups, on the basis of their sequence and predicted structure but independently of their BMC origins (figures 38A, B & 39).

Canonical BMC-H

The first group is composed of canonical BMC-H (1 structural domain pfam00936 made of 4 β-strands and 2 α-helices). PduA/J, EutM and CmcA/B/C belong to this group (figure 39). They share between 56% and 95% of sequence identity for extreme cases or, an average residue conservation of 64% (figure 38C). Alignment of their 3D structure predicted by AF2 depicted a practically perfect structure alignment with a root mean square deviation (RMSD) of 0,460 to 1,373Å.

C-terminally-extended BMC-H

The second group is constituted of canonical BMC-H which have an additional long C-terminal extension (figures 38A & B). Based on structural considerations, these extensions might confer specific functionalities to each protein.

While CmcE (39-residue long) extension was predicted to be fully disordered by AF2, PduK (59-residue long) extension seemed to be only partially disordered and displayed a quite ordered Cys-rich domain at the extremity of the flexible extension (figure 39). Of note, when compared to the protein 3D structure data bank with Dali server, no match could be found with any known protein domain.

EutK (66-residue long) extension adopted a particular conformation (3 α-helices followed by 2 small β-strands, resembling *E. coli* EutK C-terminal domain structure that was resolved from crystals) (Tanaka *et al*, 2010). The closest structural fold to this peculiar domain was an helix-turn-helix motif found in many nucleic acid binding proteins. Unlike *E. coli* EutK which remained monomeric in solution (Tanaka *et al*, 2010), *Kpe 342* EutK was predicted to associate as a hexamer. In this context, the C-terminal extensions formed independent domains.

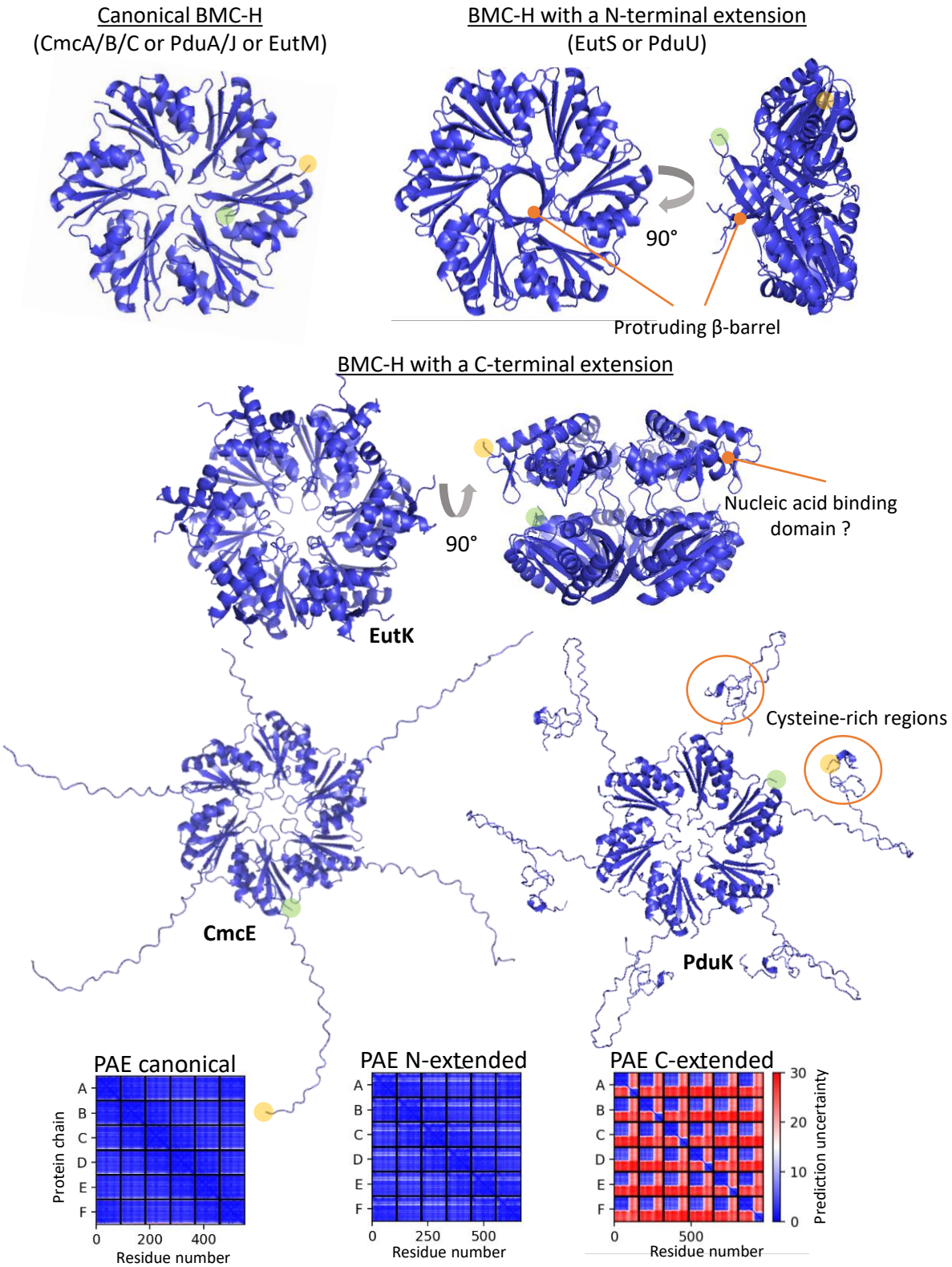


Figure 39. Alphafold2 3D-structure predictions of *Klebsiella pneumoniae 342* BMC-H.
Canonical BMC-H are only composed of the pfam0936 domain. Extension of the N-terminally extended BMC-H would fold as a β-strand that associates with homologous extensions within the hexamer to form a β-barrel that protrudes on the convex face. The C-terminally extended BMC-H would not share a common behaviour: a well-folded extra domain was predicted for EutK while CmcE and PduK extensions would be more flexible and unstructured. The yellow and green circles indicate the localization of the C- or N-terminus of one BMC-H, respectively. Representative predicted aligned errors (PAE) are provided for each BMC-H class structure.

Of note, all extensions were predicted to protrude from the BMC-H concave face which, according to consensual BMC structural models, would orient such elements towards the cytosolic side of the shell. Besides, sequence alignment of C-terminally-extended BMC-H indicated a lower resemblance within the group (approximately 30% sequence identity) than when comparing each one with canonical BMC-H of their origin BMC (35% for PduK/PduA to 56% for CmcE/CmcA; figure 38C).

Globally, predicted structures and sequence alignment suggested that, although CmcE, PduK and EutK all have a C-terminal extension, these extensions probably do not adopt a similar conformation nor hold similar functions in the BMC. Indeed, CmcE was shown to impact GRM2 shell size as CmcE absence in minimal BMCs recombinantly expressed in *E. coli* generated smaller particles (Kalnins *et al*, 2020). Nevertheless, it is unclear whether this CmcE extension was involved in this phenomenon.

On the other hand, deletion of PduK led to impediment of PDU budding from the cell poles, but without any impact on shell integrity (Yang *et al*, 2022). Then, PduK controls PDU dynamics and localization and it might do so through extension-mediated interactions with the McdA/B system which bind to the nucleoid and induce BMC movement along the cell axis (MacCready *et al*, 2018).

Finally, EutK extension was proposed to be a DNA-binding domain (Tanaka *et al*, 2010) therefore EutK could play an equivalent role to PduK but through direct interaction with the cell nucleoid.

Circularly-permuted BMC-H

The third and last group is composed of circularly-permuted BMC-H (figures 38A & B). Due to their circular permutation, PduU and EutS secondary structural element order is modified: the normally final β-strand and small α-helix are moved to the protein N-terminus. This also provokes a switch in N- and C-termini orientation from the concave to the convex face as described earlier in the BMC shell architecture (see part 1, section 4.1).

Furthermore, the N-terminal extension of PduU from *Salmonella enterica*, or EutS from *Clostridium difficile* and CutR from *Streptococcus intermedius* was shown to form a β-strand that protrudes on the convex face and associates with other intra-hexamer extensions into a β-barrel, occluding the central pore (Crowley *et al*, 2008; Pitts *et al*, 2012; Ochoa *et al*, 2020), a structure that was also predicted by AF2 for *Kpe 342* homolog BMC-H (figure 39). Surprisingly, in *Kpe* EutS, the Gly39 which was observed in *E. coli* EutS to induce the formation of a bent hexamer (Tanaka *et al*, 2010), is also present however AF2 proposed a flat version.

No information has been collected yet concerning the circularly-permuted BMC-H face orientation within the shell. However, a preliminary response element could be retrieved from PduU/PduV interaction analysis (Jorda *et al*, 2015). Co-evolution study depicted that, among all other

PDU components, PduV would interact exclusively with PduU. Y2H assay supported the predicted interaction and docking studies suggested that PduU protruding β-barrel would serve as an anchoring base for PduV binding. Of note, PduV appeared to localize to the exterior of the PDU and be linked to PDU dynamics within the cell (Parsons *et al*, 2010a). These data would indicate that, contrary to other BMC-H, circularly-permuted BMC-H convex face would probably point outward the BMC, allowing external PduV or an equivalent protein binding and subsequent BMC movements. *Kpe* EutS extension has 66% of sequence identity with PduU and their N-terminal extensions were predicted to adopt a similar conformation. One could assume that EutS and PduU extensions share the same function in BMC dynamics.

Considering all the diversity of *Kpe 342* BMC-H, and that hetero-hexamer formation was evidenced previously (Garcia-Alles *et al*, 2019; Sommer *et al*, 2019) and in the first chapter of this thesis with β-CBX BMC-H, we could wonder whether BMC-H cross-associations also happen within the different BMCs of *Kpe*. Besides, sequence alignment depicted a high homology, not only between BMC-H originating from the same BMC but also between homologs from different BMC types. Thus, cross-interactions could also arise with BMC-H deriving from 2 distinct BMC types. In this second chapter, I explored the extent of cross-interactions among *Kpe 342* BMC-H. To this end, a library of BMC-H pairs was constructed and assayed in tGFP in *E. coli*.

2.2. Construction of the BMC-H pair library

As seen in a previous study and in the chapter 1, tag orientation can be deleterious on PPI (Garcia-Alles *et al*, 2019). Here, I aimed to screen 11 different BMC-H as combinations of pairs. Yet, no data were available on the preferred tag orientation for GFP10 or 11 labelling of each BMC-H. AF2 predictions were scrutinized to try to determinate such parameter (figure 39). Several BMC-H had their N- and C-termini clearly accessible, protruding on the hexamer surface: CmcA, CmcB, CmcC, CmcE, PduA, PduJ, PduK and EutK although its N-terminus could potentially be hidden by its structured C-terminal domain. On the contrary, EutM, EutS and PduU C-termini oriented towards the hexamer interior. Of note, while EutM C-terminus was predicted as a flexible loop, EutS and PduU C-termini were predicted to fold as a β-strand, closely intertwined in the intra-hexamer interface. Thus, a C-terminal tagging could result in hexamer destabilization for both cases.

As no clear argument could be drawn from AF2 predictions in favour of a C- or N-terminal tagging, except maybe for EutS and PduU, all the possible tag attributions and orientations were constructed on individual vectors: POI-GFP10, GFP10-POI, POI-GFP11 and GFP11-POI. These vectors were then

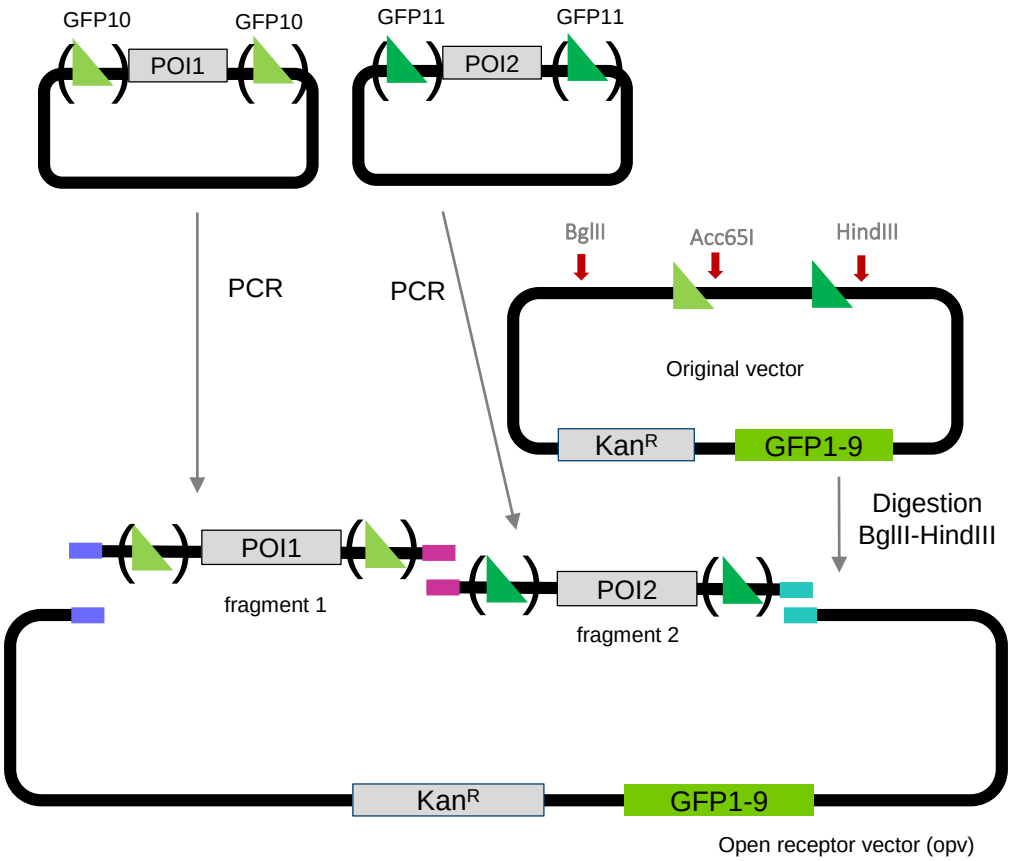


Figure 40. Gibson assembly strategy for the construction of the BMC-H pair library. C- or N-terminus alternatively-tagged POIs (represented by the brackets) are coded individually on pET26b vectors from which they are amplified by PCR. Primers used in this step add regions at the extremities of the POIX-GFPX fragment, homolog to the open vector or to the future adjacent fragment. In parallel, the original receptor vector is opened by BglII-HindIII digestion (in later phase, the protocol was adapted to include an additional digestion with Acc65I, which cleaves inside the fragment excised by BglII-HindIII, thus reducing the chance of original vector reconstitution during the assembly). Finally, both fragments are assembled with the open vector by Gibson assembly reaction, in one step for preliminary construction test, in two steps for subsequent optimized reactions (first step with opv plus fragment 2 and the second step with the addition of fragment 1 to the reaction mix).

used as template to amplify tagged-POI fragments. Pairs of POIs were subsequently assembled by Gibson into a bicistronic pET26b receptor vector (opened by enzymatic digestion) to create the final tGFP vectors (figure 40). In that manner, a total of 484 BMC-H pairs were constructed with the help of the strain engineering platform of Toulouse White Biotechnology.

Utilisation of an unique strain for both cloning and expression

In an attempt to decrease handling efforts to create such library, the possibility to perform all the cloning work directly in a strain normally intended for protein expression was evaluated. Home-made TOP10 cloning strain, BL21(DE3) or T7 express (a BL21 derivative) expression strains were transformed with Gibson assembly reactions and compared. The tGFP vectors were positive-PPI CcmK1/CcmK2, negative-PPI CcmK1/CcmK3 and *Kpe* BMC-H pairs of unknown-PPI-status: CmcA*/CmcC*, *CmcB/CmcC*, EutM*/EutS*, *PduA/*PduK, PduJ*/EutS* and *PduU/*PduK (where the asterisk indicates the orientation of the GFP tag).

While no clone was visible after BL21(DE3) transformations, between 60 to 100 clones were obtained for TOP10 bacteria. By contrast, approximately 300 clones were present in T7 express transformations which depicted a greater transformation efficiency for home-made T7 express over the TOP10 and BL21(DE3) (10^8 against 10^7 and 10^6 clones per µg of pUC19 plasmid, respectively).

Surprisingly, after a 2- to 3-day storage at 4°C, a portion of the T7 express transformants became brightly fluorescent in all cases, including the negative couple CcmK1/CcmK3 (figure 41A). Of note, some cases were already displaying fluorescence after a 1-day storage at 4°C and fluorescent clones were also visible in the Gibson negative control composed of the open receptor vector (opv) alone, probably resulting from the original receptor vector religation (the opv used was not purified thus, excised original fragment coding for CcmO-GFP10/GFP11-CcmP was still present). Also, the majority of the clones were non-fluorescent.

In order to explain such fluorescent clone apparition in all T7 express cases and to determine whether fluorescence could be used to screen properly-assembled tGFP vectors, plasmids from fluorescent and non-fluorescent clones were purified and sequenced.

Non-fluorescent clones were all misassembled vectors lacking either the fragment 1 (GFP10-tagged POI1) and 2 (GFP11-tagged POI2) or only the fragment 2. Fluorescent clones were mostly correctly-assembled tGFP vectors bearing both POI-coding fragments although some exceptions seemed to derive from plasmid recombination.

Indeed, the T7 express strain is coding for the recombinase which does not preclude plasmid recombination as in TOP10 bacteria in which its gene was deleted. Then, recombination events between highly repetitive GFP10/11 linker sequences (GFP10-Lk-GFP11) or between BMC-H homo-

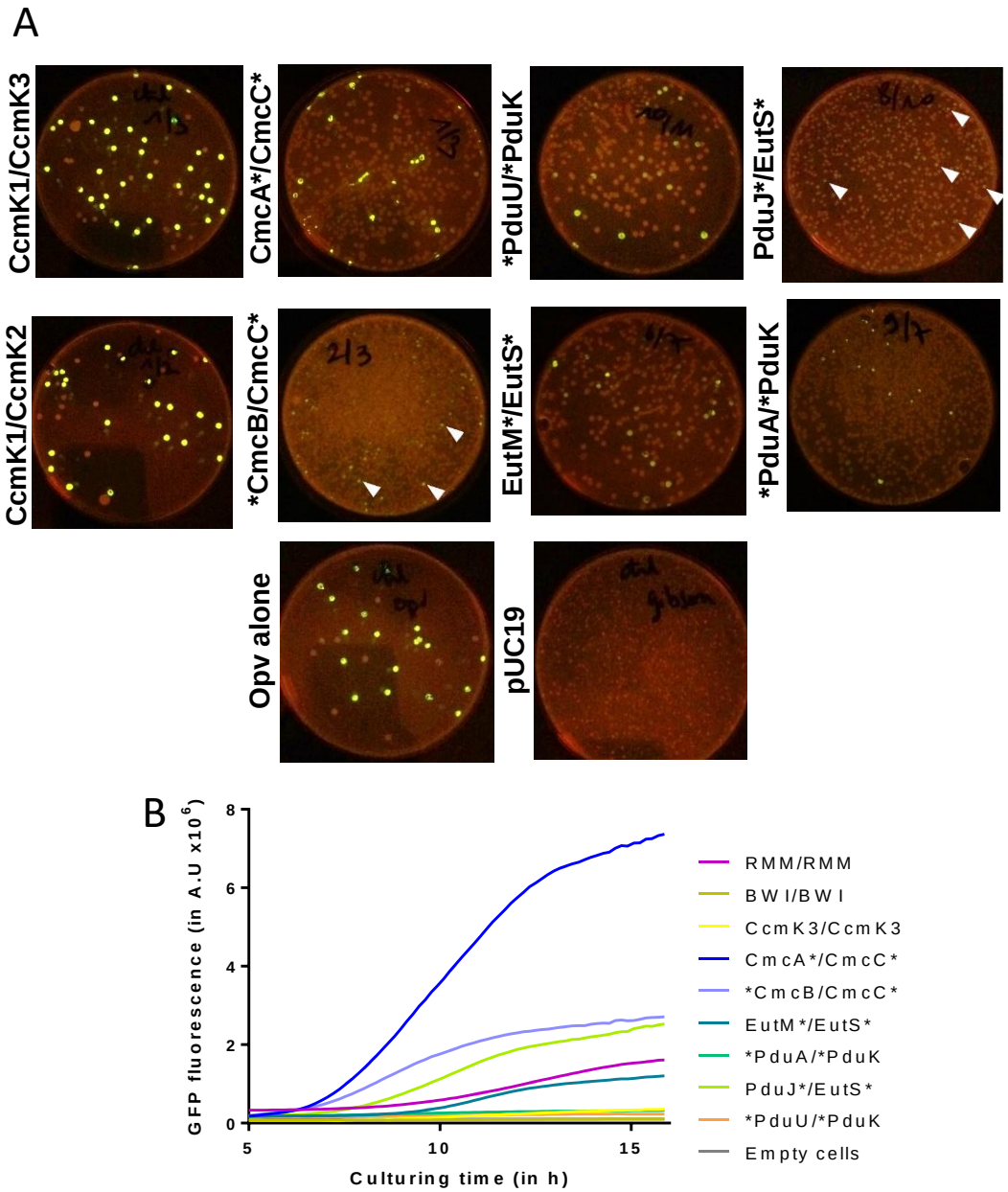


Figure 41. T7 express strain as both a cloning and an expression strain.
A. Fluorescent clones observed (under blue light and an orange filter) after T7 express cell transformation with Gibson assembly products and a 3-day storage at 4°C. Opv alone case is the receptor vector submitted to Gibson assembly but without fragment. The pUC19 case is standardized fragments of the pUC19 plasmid that were reassembled by Gibson assembly (control to ascertain that the Gibson reaction worked). White arrows point at fluorescent clones. **B.** Preliminary tGFP assay with sequence-verified fluorescent clones. T7 express cells were induced from the beginning of the culture with IPTG and fluorescence apparition was followed up during 16h, at 37°C. The asterisk position depicts the N- or C-terminal orientation of the GFP10 or 11 tags.

pairs (GFP10-Lk-POI-Lk-GFP11) were observed. However, these events were in a minority and could be easily discriminated by the size of the ORF, through a colony PCR or an enzymatic digestion.

As for the fluorescent clones present in the opv condition, they were indeed arising from the original vector religation which could also be evidenced by PCR or digestion (ORF of approximately 2000 base pairs against 1200 at most for *Kpe* BMC-H pairs).

To shed some light on why all correct constructs produced fluorescent clones, including the negative couple CcmK1/CcmK3, and in absence of an IPTG induction, a preliminary tGFP assay was performed on the 6 constructs transformed in T7 express and compared to the RMM homo-pair reference (figure 41B).

While *CmcB/CmcC*, PduJ*/EutS*, EutM*/EutS* and CmcA*/CmcC* had a GFP signal equivalent or superior to RMM, the fluorescence of *PduA/*PduK and *PduU/*PduK was below the CcmK3 negative control threshold. Thus, although the last 2 BMC-H couples produced fluorescent clones upon storage at 4°C in absence of IPTG induction, the same clones, isolated and assayed in tGFP, happened to be negative PPI. It is important to note here that final conclusions on BMC-H cross-interactions could not be drawn at these point because the tag orientation tested might not be the best suited.

Firstly, the apparition of fluorescent clones suggested that the T7p in the T7 express strain was leaky and that the 3 tGFP partners were expressed, even in absence of induction. Secondly, as this fluorescence was not necessarily linked to a positive PPI, this pointed to an increase in unspecific tGFP reconstitution when the cells are kept at 4°C.

Protein synthesis is negatively affected by low temperatures. Under 8°C, translation initiation in *E. coli* is inhibited (Friedman *et al*, 1971). Only the elongation of proteins whose translation has already begun continues until completion, but with a slower rate than at 37°C (Farewell & Neidhardt, 1998). These proteins synthetized at low temperatures would add up to the proteins already produced during the overnight incubation at 37°C, after transformation, and prior to the 4°C storage (due to the leakiness of the T7p).

In theory, at low temperatures, the newly synthetized proteins would benefit from a better folding. Then, aggregation-prone proteins such as the GFP1-9 would be more stable and last longer within the cytosol. Besides, the activity of most proteases is decreased by a temperature downshift (Francis & Page, 2010).

Taken together, this could explain fluorescent clones apparition. Indeed, they could be the result of accumulated POIs within the cells (due to a reduced proteolytic activity at low temperatures) which would randomly collision and induced unspecific reconstitution of the GFP. On the other hand, the

GFP1-9 would be more stable which would decrease the pull-down effect it held on GFP-tagged POIs and thus the clearance of non-interacting partners.

The T7 express strain proved to be very interesting for the library construction as it could be used as both a cloning and expression strain, thus significantly decreasing the amount of work required for plasmid construction and tGFP assays. Unexpectedly, the T7p leakiness observed in T7 express upon a 4°C storage led to the apparition of fluorescent clones. Importantly, fluorescence reflected correctly-assembled tGFP vectors for the majority of the clones, independent of BMC-H pair PPI status. Thus, for the whole library construction, I took advantage of this characteristic and implemented it as the screening strategy to select the correct clones.

Increasing the Gibson assembly efficiency for robotized library construction

The BMC-H pair library was intended to be built, transformed into T7 express cells and screened on the basis of fluorescence apparition, by a pipetting automaton. This implied reduced volumes in assembly mix and in competent cells. Yet, 2 issues existed for the miniaturization of the library construction: the low percentage of fluorescent clones among the different cases and the unspecific fluorescent clones observed in the negative control of Gibson assembly (opv alone). Indeed, by reducing the volumes of assembly mix and of competent cells, there was a risk that transformation of constructed vectors gives rise to fewer clones with potentially no fluorescent ones. Besides, to increase the fitness of the screen, I needed to ascertain that each fluorescent clones were correct clones and not the religated original vector. These problems were tackled independently.

To decrease the frequency of the original vector recircularization, and as purification of the opv would lead to a great loss in material, another strategy was put in place. Besides BglII and HindIII that were normally used to prepare the opv, an extra enzyme, Acc65I, was added to the digestion mix (figure 40). This enzyme had a restriction site localized in the middle of the CcmO-GFP10/GFP11-CcmP fragment removed from the original vector. A negative control composed of the unpurified opv alone was constructed with the new opv-preparation strategy or with the former double-digestion strategy. After 3 days at 4°C, the number of fluorescent clones was assessed.

While both strategies gave rise to a comparable number of clones (around 200), fluorescent clones were exclusively present, for the negative control, in the double-digestion strategy (11% of the clones compared to 0% for the triple digestion).

The triple digestion succeeded in eliminating the unspecific fluorescent clones. Thus, for the library construction by robotics, the opv preparation was performed with a triple digestion.

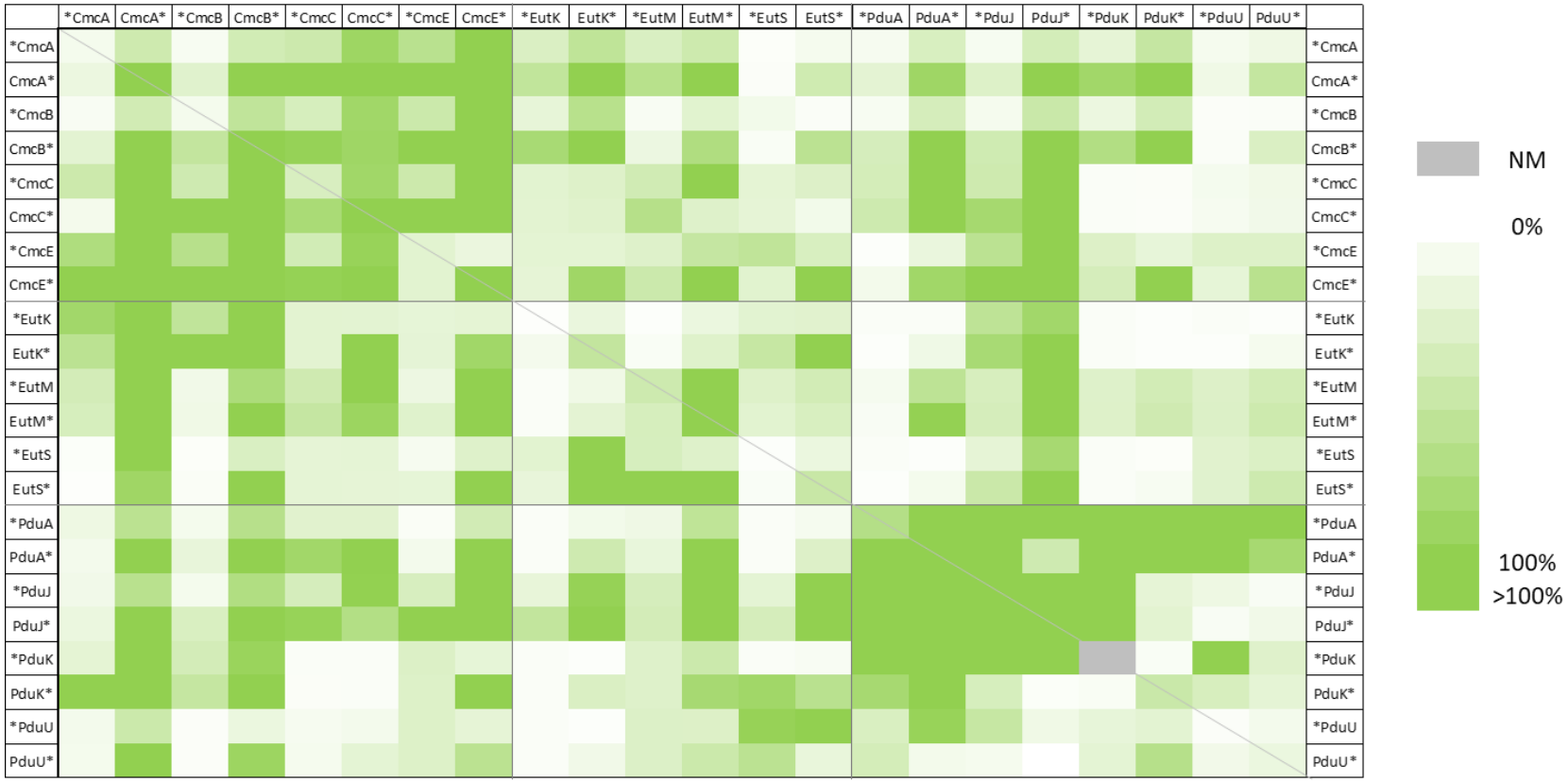


Figure 42. Protein-protein interaction matrix of *Klebsiella pneumonia 342* BMC-H pairs.

tGFP assay on cells expressing the different combinations of the 11 BMC-H coded by the 3 BMC *loci* present in *Klebsiella pneumonia 342* genome. *E*ach BMC-H was tested with either a GFP10 (horizontal repartition in the table) or GFP11 tag (vertical), localized on its N- or C-terminus (depicted by the place of the asterisk; left for N-terminus and right for the C-terminus). Green shades represent the value of the F_{max} obtained for each BMC-H pair, expressed as percentages of the RMM reference pair value. NM: not measured (vector construction failed).

In parallel, different strategies were tested to decrease the percentage of non-fluorescent clones. As the non-fluorescent clones were more prominent in the negative control composed of the opv plus fragment 1, the fragment 1 was suspected to be the main cause of such clone apparition. Then, I tempted to shift the homology regions between the fragments 1 and 2 (8 base pairs upstream present regions) or to vary the ratio between the fragment 1, fragment 2 and the opv (decreasing the fragment 1 or increasing the fragment 2 quantity) but did not get any improvement.

As the issue arose mainly from the fragment 1 unspecific recombination with the right opv extremity, normally allocated to the fragment 2 ligation, two 2-step assembly strategies were designed to favour the fragment 2 ligation. Basically, in the strategy (1), fragments 1 and 2 were assembled before subsequent assembly with the opv while in the (2), the fragment 2 was assembled with the opv prior to fragment 1 addition.

Whereas 300 to 500 clones were counted in the 1-step assembly with 30 to 60% of fluorescent clones, the clone number was significantly decreased in the 2-step assembly (1): 100 to 200 clones with approximately the same percentage of fluorescent clones. By contrast, in the 2-step assembly (2), the same clone number as in the 1-step assembly was obtained but with an increase in fluorescent clone proportion (40 to 70%).

Thus, the 2-step assembly with the fragment 2 and opv ligation prior to fragment 1 addition was selected for the library construction which was carried out by robotic means on the strain engineering platform of Toulouse White Biotechnology (see *Material and methods* for the detailed protocol).

2.3. Interaction assay within the library; homomer formation

After construction and sequencing of the whole BMC-H pair library, T7 express correct clones were assayed in tGFP as before. Briefly, OD_{600nm} and GFP fluorescence were monitored during a 16h culture, induced from the beginning with 10µM of IPTG. General F_{max} results are summarized in the PPI matrix (figure 42). Of note, the *PduK/*PduK couple, coloured in grey in the matrix, was the sole case which construction failed and which interaction could not be tested.

For more clarity, the different associations, *i.e.* formation of homo- or hetero-hexamers with BMC-H from the same or from different BMC types, will be presented separately. In this section, only the homo-pairs will be analysed and commented.

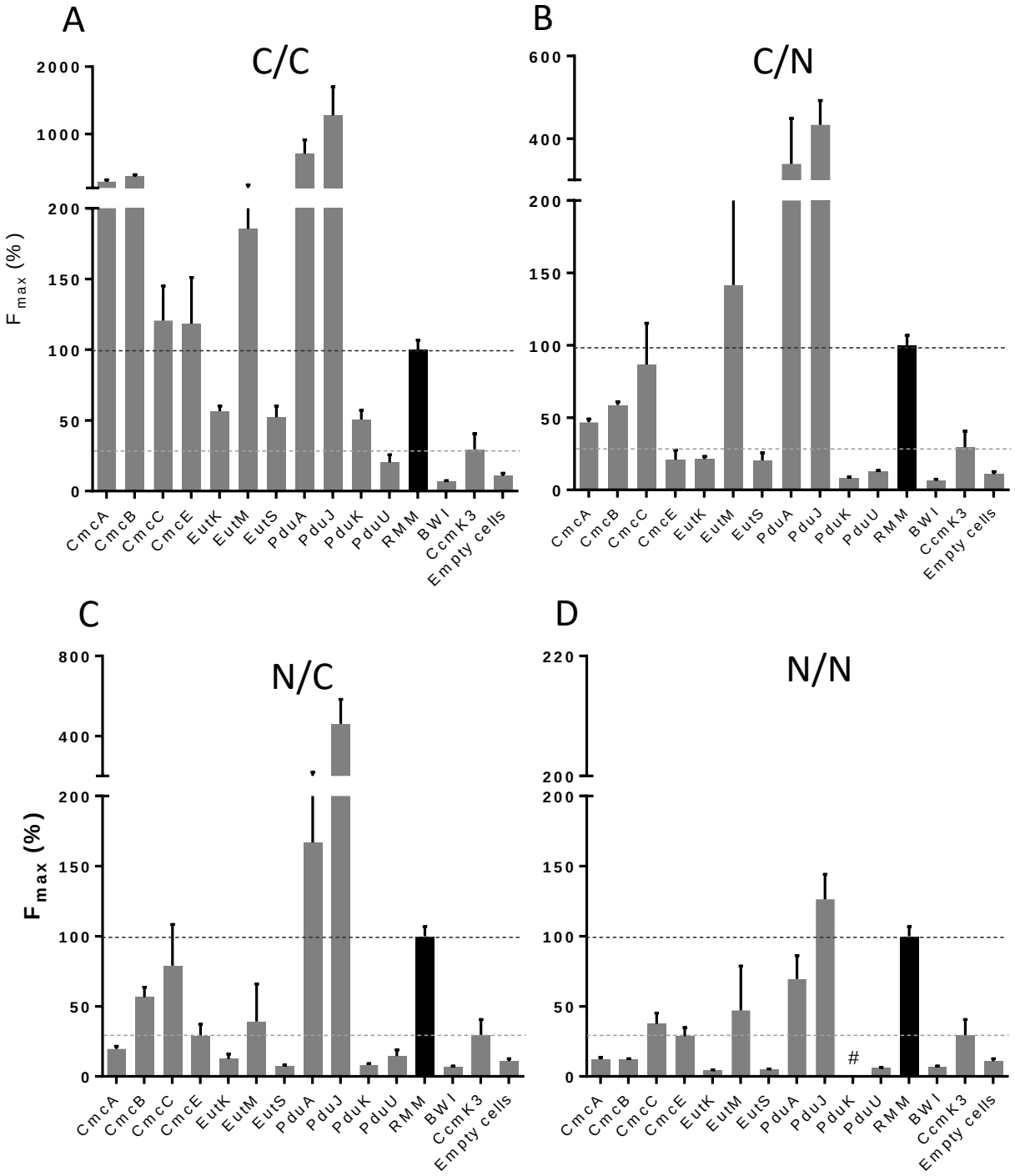


Figure 43. Homo-hexamer formation according to tag orientation.
Kpe 342 different BMC-H were constructed as homo-pairs with either both GFP10 and 11 tags in C-terminal (**A**), the GFP10 in C- and GFP11 in N-terminal (**B**), the GFP10 in N- and GFP11 in C-terminal (**C**) or both GFP tags in N-terminal (**D**). They were then assayed in tGFP and their maximal fluorescence values (F_{max}) were obtained by a sigmoidal fit on the fluorescence apparition curve and reported as a percentage of the RMM homo-pair (C/C version) value. Of note, the PduK homo-pair with the N/N tags could not be constructed and thus, not assayed.

General preference for C-terminus tagging

Data presented in the first chapter highlighted the importance of tag orientation on protein expression and interaction. Here, as we ignored the preference of presently tested *Kpe* BMC-H, all combinations were constructed and homo-pairs were analysed to determine such property (figure 43).

PduA and PduJ had a F_{max} value 7 to 12 times greater than the RMM reference for a C-terminal GFP10 and GFP11 (C/C) while CmcA, CmcB and EutM fluorescence were almost 2 to 4-fold the one of RMM. With the same tag orientation, CmcC, CmcE and RMM had comparable F_{max} values. On the other hand, EutK, EutS, PduK and PduU fluorescence were below the positive threshold (50% of RMM value for the first 3 BMC-H homo-pairs and 20% for PduU). A lower fluorescence, under the negative CcmK3 pair value, was even obtained with the 3 other tag orientation combinations.

Globally, all GFP signals decreased, for homo-pairs that seemed negative PPIs as well as for homo-pairs that were positive in C/C, in the N/C, C/N and N/N combinations. The N/N combination appeared to be the worst combination as only PduJ pair remained above the RMM F_{max} value.

These data showed a clear preference for C-terminal tagging of the different BMC-H along with a deleterious effect of N-terminal tagging on PPI study.

In order to be able to conclude on homomer formation by the different BMC-H, protein expression was monitored. Indeed, a low fluorescence signal for the EutK, EutS, PduK or PduU pairs could be the result of a poor POI expression rather than an absence of PPI. Yet, the GFP-1-9 was previously shown to provoke the pull-down of non-interacting partners, making POI expression verification from the tGFP vector inadequate. Thus, I opted for evaluating the level of expression of each POIs individually (with the 4 different tag configuration). To this end, T7 express cells were transformed with the plasmids that served as templates for the amplification of Gibson assembly fragments. Afterwards, they were induced with 10µM of IPTG for 16h before total protein collection and analysis in SDS-PAGE (figure 44).

A strong protein overexpression was evidenced for all tag orientations with CmcC, EutK, EutM, EutS, PduA and PduJ. CmcA, CmcB, PduK and PduU were well expressed with N- or C-terminal GFP10 and C-terminal GFP11 but no band could be clearly observed for the N-terminally-GFP11-tagged form. These data confirmed that C-terminal tagging was more tolerated for both the GFP10 and GFP11 as depicted by greater band intensities. The only exceptions were PduU and PduK which were more expressed as N-terminally-GFP10-tagged forms than the C-terminal forms and CmcE which was surprisingly more prominent with N-terminal GFP tags. Indeed, only slight bands were visible for CmcE C-terminal GFP10 or GFP11-tagged forms. Of note, lower molecular weight bands were noticed for PduK (C-terminally-GFP10/11-tagged and N-terminally-GFP10-tagged) which could be indicative of a tendency for proteolysis.

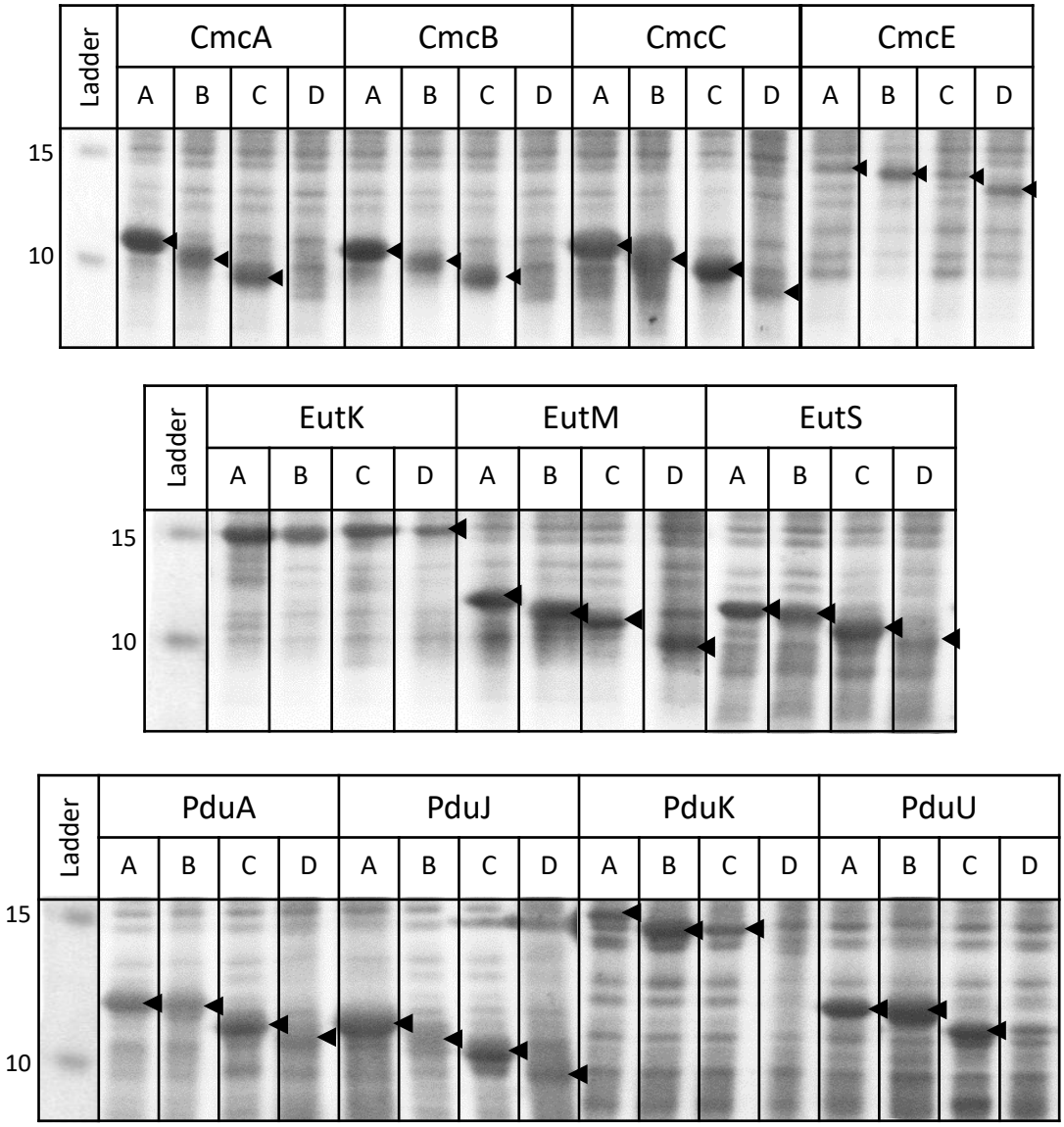


Figure 44. Individual BMC-H expression according to tag attribution and orientation.

Total protein fractions of cells expressing each BMC-H individually, tagged either with the GFP10 in C- (A) or N-terminal (B) or with the GFP11 in C- (C) or N-terminal (D) were analysed on SDS-PAGE. The molecular weights are indicated in kDa. Black arrows identifies the different POI bands.

Overall, greater fluorescence signals were obtained for combinations carrying C/C tags, most likely because C-terminal tagging favoured higher expression levels whereas N-terminal tagging, especially with the GFP11, seemed to impact protein production. However, protein expression did not fully explain tGFP results. For instance, CmcE was more expressed in the N-terminal GFP10 or 11 version and yet, the C/C CmcE resulted in a higher fluorescence. Also, EutK, EutS, PduK and PduU homo-pairs were well-expressed, for at least the C-terminally-tagged forms, which contrasted with the absence of signals observed in the tGFP assay. This could suggest 2 things: as total protein fractions (and not soluble fractions) were analysed, POIs, although expressed, could be aggregated and unable to interact, or POIs were soluble but could not associate as homo-hexamers.

In light of these data, first conclusions could be drawn. C/C combinations were generally the fittest for the tGFP assay of BMC-H interactions and this was due to a better POI expression. Yet, for some rare cases like CmcE or GFP10-tagged PduK and PduU, N-terminal tagging was not to be excluded for the test. Although they did not produce a GFP signal in tGFP, POI expression were enhanced and I could not rule out that with the appropriate partner, these forms would be the fittest to demonstrate an interaction.

Homomer formation

Protein structures have already been resolved by X-ray crystallography for some BMC-H homologs coming from other organisms (PduA, J, U, CmcA, B, C, EutS and EutM) and showed a hexameric form in the asymmetric unit (Ochoa *et al*, 2021; Tanaka *et al*, 2010; Pang *et al*, 2014; Chowdhury *et al*, 2016).

Here, high fluorescence signals in tGFP evidenced the formation of homomers for CmcA, CmcB, CmcC, EutM, PduA and PduJ (figure 43) which was in accordance with the data from the bibliography. However, as the oligomerization state of each homo-pair was not inspected, positive pairs could only be denominated as homomers, involving a minimum of 2 identical BMC-H. An extra experiment such as SEC would have be needed to be able to refer to the positive pairs as homo-hexamers.

In the tGFP assay, CmcE homo-pair was positive, indicating that CmcE was able to form a homomer. However, its precise oligomeric state could not be ascertained by the screen. While CmcE is a BMC-H, and as no CmcE homolog structure has been previously determined, nothing excluded that CmcE was associating as a dimer such as a portion of the crystallised circular-permutant CutR (Ochoa *et al*, 2020) or into a higher-order oligomeric state. Further studies would be necessary such as a SEC or a native PAGE to determine the molecular weight of CmcE in solution and thus its oligomeric state.

No interaction could be observed for EutK, EutS, PduK or PduU homo-pairs, indicating that these BMC-H were not prone to form homomers (figure 43).

E. coli EutK crystal structure could not be determined except for its C-terminal extension domain which resembled a DNA binding domain (Tanaka *et al*, 2010). Besides, EutK was previously shown to be monomeric in solution thanks to ultracentrifugation sedimentation equilibrium measure and was proposed to oligomerize only with other BMC-H from the EUT, assembling as mixed hexamers (Tanaka *et al*, 2010). To the best of our knowledge, no study has been performed yet to corroborate this theory. Only one potential interacting partner was identified by large-scale PPI study in *E. coli K12* through protein co-purification: EutL, a BMC-T (Arifuzzaman *et al*, 2006).

The next section on hetero-hexamer formation with BMC-H from the EUT might provide some element of response on the matter.

Similarly, PduK structure has not been resolved yet. AF2 predicted the oligomerization of PduK monomers into a hexamer.

Besides a lack of interaction with itself, different phenomena might also explain why PduK appeared negative in the tGFP assay (figure 43). The first hypothesis is that, when expressed on its own in *E. coli*, PduK was insoluble, as did CcmK3, or subject to proteolysis. Indeed, several lower molecular weight bands were observed on PduK SDS-PAGE profile which might indicate a partial degradation (figure 44). Crowley *et al* obtained the same profile with PduT, which central pore is composed of 3 Cys sheltering a [4Fe-4S] cluster (Crowley *et al*, 2010). They proposed that these lower bands were the result of reactive oxygen species degradative action on the metal centre. To recall, PduK has a Cys-rich region on its C-terminal extension which could be the binding site of a [Fe-S] cluster. If so, PduK could have follow the same fate as PduT which would imply the loss of the GFP tag, when C-terminally-tagged.

Another explanation would be that, while the C-terminal GFP tagging was generally preferred among *Kpe* BMC-H, including PduK, such orientation would keep the GFP10 and 11 tags away from each other if implemented on PduK. Indeed, by tagging on the C-terminus, the length of PduK extension (59-residue long) would add up to the linker length. Furthermore, as PduK extension was predicted to be almost completely disordered, it would be highly mobile, decreasing even more the probability of adjacent GFP10 and 11 and thus the probability of the GFP reconstitution. The N/N combination could have solved this issue as the linker length would not be incremented by an extension. However, the N-terminal GFP11-tagged PduK was not observed in SDS-PAGE (figure 44), making unlikely the apparition of fluorescence for the N/N combination.

Another possibility would be that PduK extension has an equivalent role to EutK C-terminal domain. By doing so, PduK and EutK would share a conserved mechanism of assembly as exclusive hetero-hexamer.

Despite a high expression pattern (with the exception of the N-terminally-GFP11-tagged forms), EutS and PduU homo-pairs gave low signals in tGFP, pointing at an absence of interaction (figures 43 & 44).

On the contrary, in the literature, 2 EutS homolog structures have already been elucidated : EutS from *E. coli* or from *Clostridium difficile* (Tanaka *et al*, 2010; Pitts *et al*, 2012). To recall, the first one was resolved as a distorted and bent hexamer while the latter was a regular hexagonal hexamer. Sequence alignment revealed that the closest homolog to *Kpe* EutS was the one coded in *E. coli* (sequence identity of 92% compared to 52% for *Clostriodium* EutS). The Gly39 which is responsible for EutS distortion in *E. coli* is also present in *Kpe* EutS. Thus, EutS homo-hexamer conformation would follow *E. coli* EutS swirl-shaped structure.

As for PduU, a PduU homolog from *Salmonella enterica* was shown to form homo-hexamers (Crowley *et al*, 2008). One could wonder why the present data and data from the bibliography are in contradiction.

According to AF2 predicted structures, the C-termini of EutS and PduU were not buried inside the hexamer but largely embedded (figure 39). They barely emerged from the hexamer surface and appeared to be involved in intra-hexamer interactions. Thus, C-terminal tagging could potentially destabilize BMC-H association.

Another possibility to explain the lack of signal for EutS and PduU homo-pairs would be that the protrusion of their central β-barrel which is on the same face than their C-termini could have created a steric hindrance, impeding the approach and reconstitution of the GFP. More data would be required in order to conclude on EutS and PduU absence of signal in tGFP.

2.4. Compatibilities between BMC-H arising from the same BMC type

This section will focus on the description of the results obtained in the tGFP assay performed on BMC-H pairs coded within the same BMC operon in *Kpe 342*. Thus, we will go through PPI within the EUT1 or the PDU1A or the GRM2 independently (figure 45). Of note, the F_{max} values presented in the different graphs corresponded to the tag combination depicting the highest fluorescence (the C/C

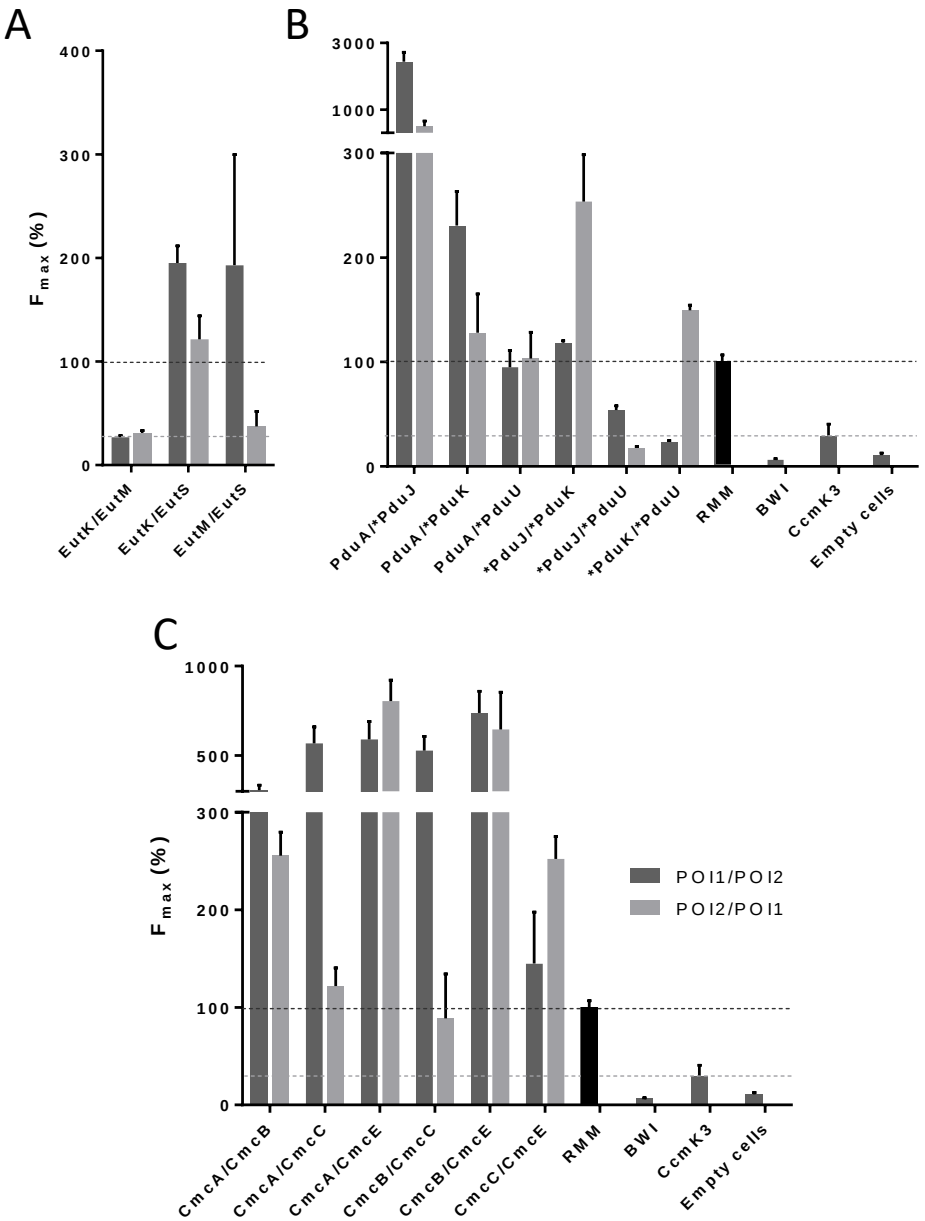


Figure 45. Hetero-hexamer formation with BMC-H pairs coming from the same BMC type.

tGFP assays were performed on BMC-H pairs with all GFP tag attribution and orientation combinations. The results shown here are the F_{max} values calculated on the POI1-GFP10/POI2-GFP11 (dark grey) or POI2-GFP10/POI1-GFP11 (light grey) combinations, with GFP tags in C-terminal, except when noted otherwise by the asterisk position. They are expressed as a percentage of the C/C RMM homo-pair fluorescence value. Of note, the name of each couple is given on the graphs as POI1/POI2.

combination in general, unless otherwise stipulated by an asterisk which indicated a N-terminally-tagged POI).

EutS cross-interacts with the other BMC-H from the EUT

Whereas EutS was found unable to associate as an homo-hexamer, it interacted surprisingly with both EutK and EutM (figure 45A). The same observation could be made about *E. coli* EutK which was monomeric on its own (Tanaka *et al*, 2010). However, EutK could form hetero-hexamers but exclusively with EutS as depicted by a F_{max} value for EutK/EutM couple under the CcmK3 negative threshold.

According to AF2 predictions and in agreement with published structures for other EutS homologs, a central β-barrel is formed within EutS hexamer (figure 39). Whether or not this structural element holds a particular function in the BMC is still unclear. Yet, a proper question to ask would be what happens to this structure in an hetero-hexamer.

EutM is a canonical BMC-H. It has no long terminal extension. Thus, in theory, no steric hindrance would be applied on neither BMC-H faces and this would allow interactions of the EutM/EutS pair.

On the contrary, EutK has a C-terminal extension that was expected to impact the interaction. Strikingly, when EutK/EutS hetero-hexamer was modelled by AF2 with a ratio 1/5, respectively, EutK extension and EutS β-barrel were protruding on opposite faces (figure 46A), making EutK/EutS cross-interaction compatible. However when EutK/EutS ratio increased, predicted hexamers seemed unstable as revealed by improperly-closed hexamers, making hetero-hexamer formation less probable (figure 46A).

As it was proposed previously (Tanaka *et al*, 2010), EutK could only be viable in the context of mixed hexamers. Data indicated here that its only interacting partner among the EUT shell subunits would be EutS. A possibility is that EutK might integrate EutS hexamers in a low stoichiometry to ensure the stability of the hetero-hexamers. Besides, thanks to both BMC-H extensions, EutK DNA binding-resembling domain and EutS protruding barrel, such hetero-hexamers could be bi-functional.

It is worth noting here that, in *Kpe 342* genome, *eutK* is approximately 13 700-base pairs away from *eutS* (figure 37), implying the same distance between both sequences in the polycistronic mRNA. Thus, in the natural host, and as typical BMC-H cannot exist as monomers because their interfaces bear several hydrophobic patches, the mRNA should adopt a 3D fold allowing the rapprochement of *eutK* and *eutS* sequence so that they be translated in close proximity and be able to associate. Alternatively, this natural genetic organization might indicate that EutK/EutS homo-hexamer formation is unlikely in *Kpe 342* or would represent a small minority of the EUT hexamers.

PduA, the central node for hetero-hexamer formation

The tGFP assay evidenced a lot of cross-interactions happening between all PDU BMC-H (figure 45B). For instance, PduA was shown to interact with every BMC-H. Its closest homolog, PduJ, followed the same trend except for PduU with which hetero-hexamer formation was unlikely. Remarkably, PduK which could not associate as a homomer, resulted in a strong GFP signal with PduA, PduJ and PduU.

PduA heteromer formation with PduJ was something expected due to their high sequence identity (77%; figure 38). By contrast, such cross-interactions were more startling with PduK and PduU as PduA only shares 35 to 37% of sequence identity with them, respectively.

Through sequence coevolution study, Jorda *et al* had predicted that PduA could interact with PduJ, PduK or PduU shell subunits (Jorda *et al*, 2015) although, at that time, they did not consider the possibility that their data might hint at intra-hexamer interactions rather than inter-hexamer ones. Possibly in favour of the importance of PduA as a hub BMC-H, the corresponding gene is found in pole position in the *Kpe 342 pdu1a* operon (figure 37). Yet, in the context of a polycistronic mRNA, gene order might be important for coded protein assembly (Parsons *et al*, 2010a; Chowdhury *et al*, 2016). As PduA translation would happen first, PduA BMC-H would probably be the nucleating centre of the majority of hetero-hexamers formed with upcoming PDU BMC-H.

Of note, in *Kpe 342* genome, *pdua* is 6000 to 13000-base pairs away from its homolog BMC-H coding sequences while in the tGFP vector, the POI ORFs were 44-base pairs apart from each other, which might have favoured interactions between naturally distant BMC-H. Thus, as well as in the EUT, unless the mRNA coding for the PDU subunits adopts a 3D fold that put in close proximity distant BMC-H ORFs and translation sites, PduA and PduU, for instance, would be unlikely to interact in the natural context.

Here, contrary to Jorda *et al* study which only predicted an interaction between PduK and the shell protein PduA (also with PduG, PduM and PduW enzymes) (Jorda *et al*, 2015), heteromer formation was noticed in combination with all other PDU BMC-H (figure 45B). This contrasted with the homomer formation results and might indicate that PduK would prefer to form mixed oligomers, as it was proposed for EutK, rather than homomers.

Similarly to the circularly-permuted EutS, the tGFP assay with hetero-pairs revealed that PduU, which also carries a domain permutation, could interact with PduA and PduK whereas it did not form homo-hexamers in spite of already crystallised 3D structure from *Salmonella* PduU (figure 45B) (Crowley *et al*, 2008). PduU and EutS shared a significant sequence homology (56% of identity) and their predicted 3D structures practically superimposed. It was then very likely that impediment of PduU homo-hexamer formation was due to the same phenomenon that prevented EutS homo-hexamer.

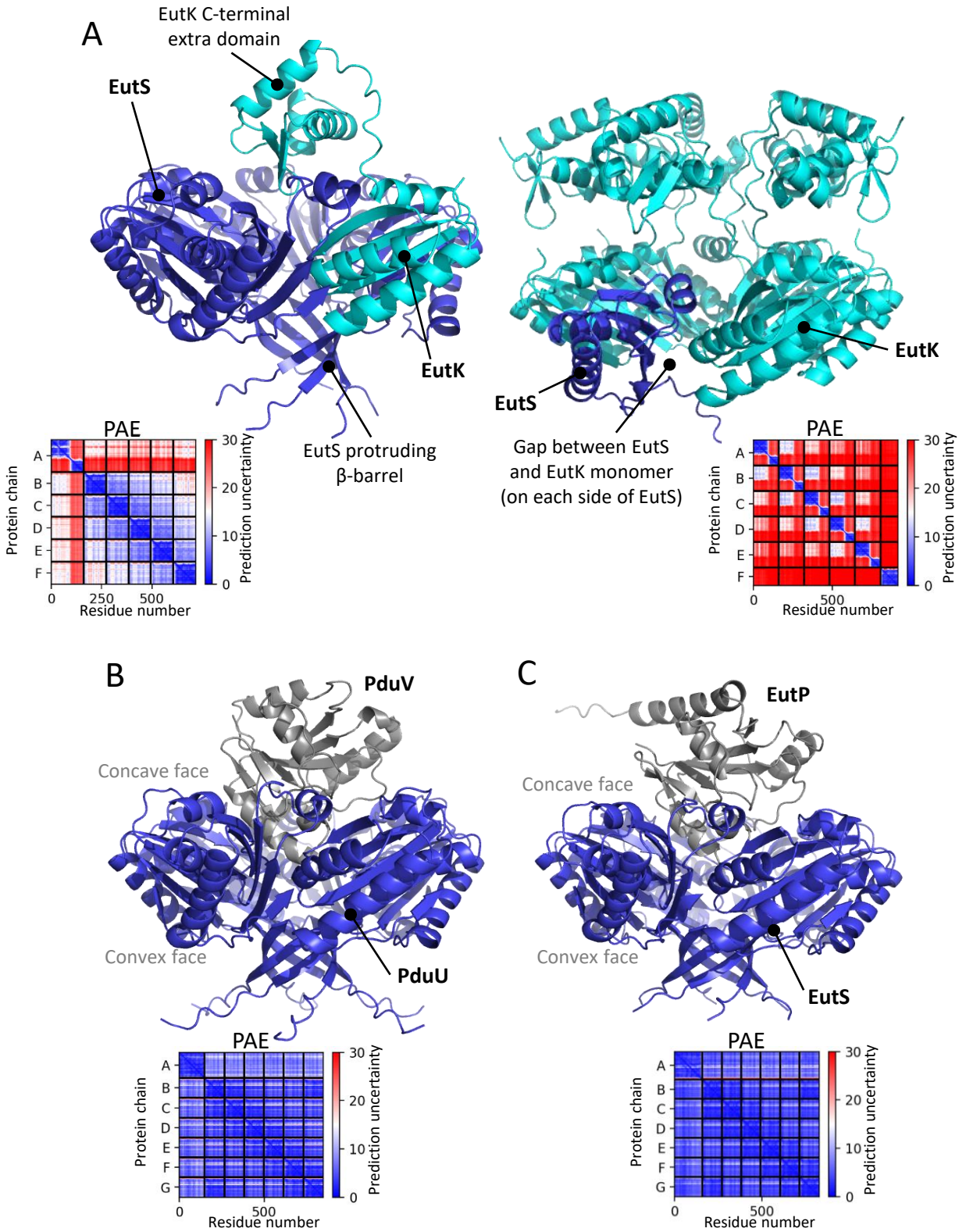


Figure 46. Alphafold2 predictions of different hexamer structures.
A. 3D structures predicted by Alphafold2 for an hetero-hexamer EutK/EutS with a 1/5 ratio or a 5/1 ratio. Comparison of the predicted 3D structures of a PduU hexamer interacting with PduV (**B**) or a EutS hexamer interacting with EutP (**C**). Predicted aligned error (PAE) plots are provided for each structure.

Possibly, PduU and EutS require another protein to help stabilizing them in the homo-hexamer form. PduV was shown to bind PduU and this interaction was proposed to be mediated by the protruding β-barrel of PduU (Jorda *et al*, 2015). Then, PduV binding could be the stabilizer that permits PduU homo-hexamer formation, an hypothesis that is supported by the genetic coding order of both partners which are adjacent (figure 37), thus translated in close proximity. However, when AF2 was launched with a PduU hexamer and PduV as query sequences, I obtained contrasting results (figure 46B). PduV was predicted to interact with PduU but on its concave face where PduV docked on the PduU central cavity.

As PduV-GFP fusion was observed to be addressed to the outer surface of the BMC shell (Parsons *et al*, 2010a), this would imply that PduU concave side is pointing outward the BMC, corroborating all whole-BMC structures published up to now in which all the shell subunits were oriented with their concave faces out (Sutter *et al*, 2017; Greber *et al*, 2019; Kalnins *et al*, 2020; Tan *et al*, 2021).

A protein homologous to PduV might exist in the EUT for EutS stabilization. EutP which was shown to have a GTPase activity like PduV along with an ATPase activity (Moore & Escalante-Semerena, 2016) shares a mild sequence identity with PduV (34%). Furthermore, like PduU and PduV, EutP is coded downstream EutS (figure 37) and when an EutS hexamer and EutP were submitted to AF2, EutP was predicted to dock onto the central cavity of EutS concave face, in the same fashion as PduV with PduU hexamer (figure 46C). Then, PduV and EutP could have homologous stabilizing functions.

Presence of highly promiscuous BMC-H in the GRM2

When the tGFP assay was performed with BMC-H coming from *Kpe 342* GRM2, F_{max} values 3 to 8-fold greater than the RMM fluorescence were monitored for each hetero-pair (figure 45C). Indeed, every BMC-H was positive with all its counterparts, suggesting that hetero-hexamers could form with every BMC-H combinations. This depicted a high promiscuity between GRM2 BMC-H interfaces, in agreement with a high sequence identity (86 to 95% between canonical BMC-H and 56% between canonical and C-terminally-extended CmcE; figure 38).

In published studies, when each BMC-H was deleted individually from *E. coli 536 cut2* operon, the mutants were still able to utilize the choline (Herring *et al*, 2018). Besides, no impact was observed on cell growth, suggesting an absence of AA toxicity which would be due to impaired shell integrity. Taken together, this would imply that the GRM2 BMC-H are interchangeable: all homologs play the same role. Cross-interactions discovered here indicated that, in addition to the redundant roles that CmcA, CmcB, CmcC and CmcE seemed to hold, different hetero-associations could form, potentially still playing the same role.

GRM2 BMC-H cross-interaction analysis raised an important question on whether a BMC-H trio or quartet could associate to form a hexamer (*i.e*. a hexamer composed of a combination of 3 or 4 of

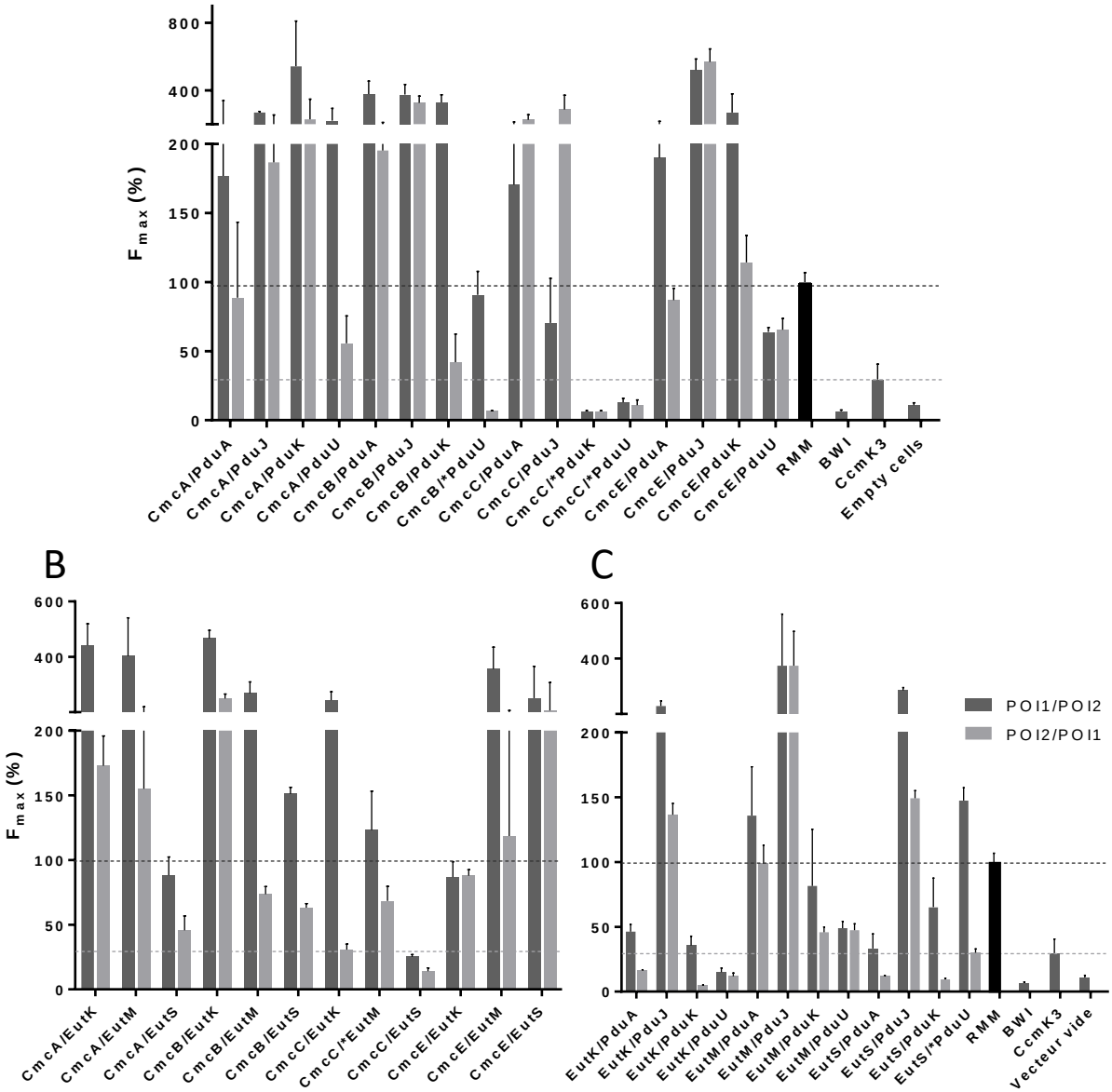


Figure 47. Hetero-hexamer formation with BMC-H pairs arising from different BMC types.

tGFP assays were performed on BMC-H pairs with all GFP tag attribution and orientation combinations. The results shown here are the F_{max} values calculated on the POI1-GFP10/POI2-GFP11 (dark grey) or POI2-GFP10/POI1-GFP11 (light grey) combinations, with GFP tags in C-terminal, except when noted otherwise by the asterisk position. They are expressed as a percentage of the C/C RMM homo-pair fluorescence value. Of note, the name of each couple is given on the graphs as POI1/POI2.

the GRM2 BMC-H). In theory, the genetic order of CmcA, CmcB and CmcC which are all adjacent in *Kpe 342* could favour such hetero-hexamers (figure 37). However, CmcE, which is remote in the *cut2* operon (practically 8000-base pairs away from CmcC sequence), would be less prone to hetero-hexamer formation with its counterparts *in vivo*, unless, as proposed for EutK/EutS or PduA/PduU, the mRNA coding all the GRM2 subunits adopts a particular fold that would ensure spatial proximity of newly translated BMC-H. As the tGFP assay showed that every BMC-H could interact with each other, there would be no structural incompatibility for the assembly of such hetero-hexamers (figure 45C).

These hetero-hexamers formed by a BMC-H trio or quartet would be very interesting for the elaboration of a hexameric platform for synthetic biology, provided that we could control each BMC-H presence and stoichiometry. Also, compared to an homo-hexamer on which only one enzymatic domain could be fused, the hetero-hexamer could accommodate up to 4 different domains.

Until today, hetero-hexamer formation had only be shown with CcmK1/CcmK2 and CcmK3/CcmK4 couples coming from the β-CBX (Garcia-Alles *et al*, 2019; Sommer *et al*, 2019). Data were missing for the other types of BMC. In this section, heteromer formation was monitored by tGFP between BMC-H from the same BMC type. Several BMC-H interacting couples were evidenced, and this, in the 3 BMCs of *Kpe 342*: the GRM2, PDU1A and EUT1. Thus, hetero-hexamer formation extends beyond the β-CBX and might even be a general phenomenon concerning all BMCs endowed with multiple BMC-H homologs.

2.5. Heteromer formation with BMC-H from different BMC types

General results of the tGFP assay with BMC-H from different BMC types

Finally, in this section, I will present in details the results of the tGFP assays on couples of BMC-H arising from different BMC types (figure 47).

For the GRM2 BMC-H that were combined with those from the PDU1A (figure 47A), the group of canonical BMC-H (CmcA, CmcB and CmcC) was observed to interact with their PDU1A canonical homologs (PduA and PduJ). CmcE also followed that trend, depicting that its C-terminal extension was not hampering the establishment of contacts for association with other BMC-H. Of all BMC-H, CmcA, PduA and PduJ were the ones totalizing the biggest interaction number as they could interact with any BMC-H from the other BMC type. On the contrary, PduU was only able to interact with CmcA for sure and also with CmcB and CmcE but the couple F_{max} values were slightly lower than the RMM reference

(90 or 65%, respectively). Surprisingly, PduK was found positive when combined with CmcA, CmcB and CmcE.

When BMC-H of the GRM2 were crossed with BMC-H from the EUT1 (figure 47B), similar results were obtained for canonical BMC-H, *i.e.* they all interacted with each other, including the CmcE/EutM pair. EutK was found to be able to associate with all BMC-H from the GRM2. This contrasted with data from tGFP assay performed on the EUT BMC-H in which EutK was unable to interact with the EUT canonical BMC-H, EutM (figure 45A). The analysis of the PPI matrix also showed that the only couple for which there was no interaction was EutS/CmcC. It is worthy to note that CmcB and EutM cross-interacted with every other BMC-H (canonical, C- or N-terminally-extended BMC-H).

Analysis of the combinations of the EUT and PDU BMC-H (figure 47C) revealed that canonical BMC-H were also observed to interact all together (PduA and PduJ with EutM). EutS/PduU couple was positive, as depicted by a F_{max} value of 147% compared to the RMM reference fluorescence. This was very surprising as both EutS and PduU, which share 56% of sequence identity, failed to form homo-hexamers separately. EutK was only able to associate with PduJ and, in a general manner, PduJ seemed to be able to form heteromers with all 3 BMC-H from the EUT.

Canonical BMC-H have a higher intra-hexamer interface plasticity

A general finding of this section was that the canonical BMC-H were highly promiscuous, interacting with a wide range of homologs, canonical as well as C- or N-terminally-extended. This was particularly the case for PduA, PduJ, CmcA, CmcB and EutM which could associate with a large majority of the BMC-H arising from other BMC types. Although bearing a C-terminal extension, CmcE was noticed to behave more as a canonical BMC-H.

Thus, canonical BMC-H might offer more malleable intra-hexamer interfaces that can adapt to a high diversity in residue composition. Their quite unspecific interactions would make an ideal nucleating centre for hetero-hexamers out of them, which might reveal of great interest for the elaboration of a synthetic platform on the basis of a hetero-hexamer.

On the other hand, non-canonical BMC-H interactions were more restricted. For instance, PduU was only able to associate with CmcA, CmcB, CmcE or EutS. Also, EutK interacted mainly with PduJ or all the BMC-H from the GRM2. Globally, non-canonical BMC-H appeared to be more specific in their interactions. Probably their extensions, and especially the circular permutations (for EutS and PduU), were the reason of this specificity.

With the platform design in mind, this fact was also very interesting as it could allow to control the localization of given BMC-H within the hybrid hexamer through implementation of canonical /non-canonical interfaces.

Biological relevance of the cross-interaction of BMC-H from different BMC types

In this study, BMC-H from different BMCs were shown to be able to cross-interact. The 3 BMC types are present in one single organism, *Kpe 342*. Choline, 1,2-PD and EA can all originate from cell membrane degradation. Thus, in theory, *Kpe* could find itself subjected to the 3 substrates at the same time and be expressing the 3 BMCs. Yet, as we just saw, several BMC-H of these BMCs could associate to form heteromers. Then, if all BMCs were to be expressed concomitantly, this would lead to hybrid BMC shells and very probably to a mixing in cargo enzymes inside the BMC lumen.

In that respect, it was previously shown that when *Salmonella pdu* operon was engineered to permit a concomitant expression of EutL or EutS, hybrid BMCs with impaired metabolic functions were produced (Sturms *et al*, 2015). In the strain of origin, a concomitant expression of multiple BMC types seems unlikely because the functions of these BMCs would be affected. Besides, BMCs are megastructures that require a lot of cellular resources to be built and production of non-functional structures would constitute a huge loss.

Generally, BMC production is a finely tuned event. For instance, for the PDU, 1,2-PD was shown to induce PocR expression (Bobik *et al*, 1992). PocR is a regulation factor which is coded upstream the *pdu* operon, in a reverse transcription orientation (figure 37). In a feedback loop, PocR enhances its own expression besides inducing the transcription of the *pdu* operon.

By contrast, the *pdu* operon is catabolically repressed by glucose, probably because glucose utilization is preferred by the cells (Staib & Fuchs, 2015). Also, in *Listeria monocytogenes*, a RNA antisens of *pocR* sequence was shown to be produced and to repress PocR translation (Mellin *et al*, 2013). However, this antisens RNA transcription is reduced upon vitamin B_{12} addition. Indeed, vitamin B_{12} would bind to a riboswitch present at the beginning of the antisens RNA and induce its premature termination.

The *eut* operon encodes a positive transcription factor *eutR*, the most distal sequence of the operon (figure 37). In *Salmonella*, EutR expression is under the control of the main operon promoter but is also constitutively transcribed at a basal level, by a weak proximal and exclusive promoter (Roof & Roth, 1992). EutR would switch to an active form in presence of both EA and vitamin B_{12} and trigger EUT protein expression.

Of note, another system composed of EutV/W regulates the *eut* expression in organisms such as *Enterococcus faecalis* (Fox *et al*, 2009), a system which is absent in *Kpe 342*.

As for the *cut* operon, it exists a small 3-protein coding operon upstream the main operon (figure 37). Among these 3 proteins, there is a positive transcription factor, CutX, which induced the transcription of the *cut* operon in presence of choline (Herring *et al*, 2018). Besides, there is CutY that does not share any similarity with the transcription factor family but was shown to play a role in the *cut* induction. Of note, it was proposed to act in concert with CutX.

The *cut* operon was found repressed in aerobiosis, potentially to avoid the destruction of oxygen-sensitive GRE enzymes before encapsulation inside the BMC shell.

Despite relatively abundant data on individual BMC type regulation, very little is known about the regulation in organisms encoding multiple BMC types. Few, yet, crucial answering elements were provided by Sturms *et al* with *pdu*- and *eut*-encoding *Salmonella enterica* (Sturms *et al*, 2015). They showed that when both EA and 1,2-PD were added in the culture medium, only the PDU was active.

In the presence of 1,2-PD, PocR repressed the *eut* operon which hinted at some kind of hierarchy between BMC types. Indeed, when both BMC substrates were present in the medium, PDU expression prevailed over EUT expression. The hierarchy order was quite surprising because EA can be a source of carbon as well as nitrogen and energy while 1,2-PD is mainly a source of carbon. It would have been expected that the EUT be favoured.

Contrasting with these data, Jakobson *et al* showed, thanks to a GFP reporter gene controlled either by the main *eut* or *pdu* promoter, that when *Salmonella enterica* was treated with EA and 1,2-PD along with vitamin B_{12}, both plasmid-encoded promoters could be activated (Jakobson *et al*, 2015). Besides, another team demonstrated that the *eut* and *pdu* operon could be concomitantly expressed (Delmas *et al*, 2019). Indeed, by culturing pathogenic *E. coli LF82* in presence of bile salts, whole mRNA sequencing revealed that both *eut* and *pdu* transcipts were overexpressed compared to the same strain cultured in M9 minimal medium.

While *Salmonella enterica, E.coli* and *Kpe* share the same *eut* and *pdu* operon organization, we could wonder whether *Kpe 342* would have also conserved their mechanisms of regulation.

The next question is what about the GRM2? Is choline utilization preferred over 1,2-PD and/or EA? In *E. coli 536*, the GRM2 expression is restricted to anaerobic conditions (Herring *et al*, 2018) and its metabolic functions do not require vitamin B_{12}. The PDU and EUT are active in aerobiosis as well as anaerobiosis and, by contrast, need vitamin B_{12} to ensure the catabolism of their respective substrate.

Then, the GRM2 could prevail in anaerobiosis, in absence of vitamin B_{12} but BMC hierarchy would be more tricky to decipher in presence of the cofactor.

To determine BMC hierarchy and co-regulation mechanisms, operon concomitant inductions with due substrates (choline ± EA ± 1,2-PD) in aerobic or anaerobic conditions should be performed on *Kpe 342*. Currently, *in vivo* tests are still undergoing by a collaborator of the Laboratoire Microorganismes: Génome Environnement of Clermont-Ferrand, Damien Balestrino. They consist in monitoring BMC operon induction by RTqPCR according to substrate addition.

Chapter 3

Development of hetero-hexameric platforms with tailored BMC-H positioning

3.1. Introduction to the engineering of BMCs and BMC shell components

3.1.1. Enzyme spatial organization to increase the catalysis efficiency

In the synthetic biology field, in order to increase the bio-production of a molecule, engineering efforts most often focus on the enzyme itself and the improvement of its substrate affinity or catalytic turnover. However, an innovative alternative or additional manner is gaining more interests recently: enzyme spatial organization. Basically, the new methods developed rely on the spatial proximity of the enzymes of a certain metabolic pathway to increase its catalytic efficiency, through substrate channelling, for instance, or enzyme clusterisation or sequestration of the substrate or of reaction intermediates (Sweetlove & Fernie, 2018; Castellana *et al*, 2014; Jakobson *et al*, 2017).

Many options exist and have already proved successful (Li *et al*, 2020b; Aalbers & Fraaije, 2017; You & Zhang, 2013; Delebecque *et al*, 2011; Küffner *et al*, 2020). To chose the most suitable approach, 2 important criteria should be considered: the number of enzymes involved in the metabolic pathway of interest and the production system (*invitro* or *in cellulo*).

Forcing spatial proximity by protein fusion

The first and simplest method to force enzyme proximity is by fusing enzymes together via a polypeptide linker (figure 48A) (Yu *et al*, 2015). As an example, Aalbers *et al* fused an alcohol dehydrogenase with a cyclohexanone monooxygenase in order to improve ε-caprolactone production from cyclohexanol (Aalbers & Fraaije, 2017). By doing so, they succeeded in converting more than 99% of cyclohexanol *in vitro* while free enzymes only reached 42% of conversion.

Enzyme connection via a linker mimics natural multi-domain enzymes for which the substrate passes from one domain to the next to be transformed more efficiently such as in the Fatty acid synthase type I (Smith, 1994).

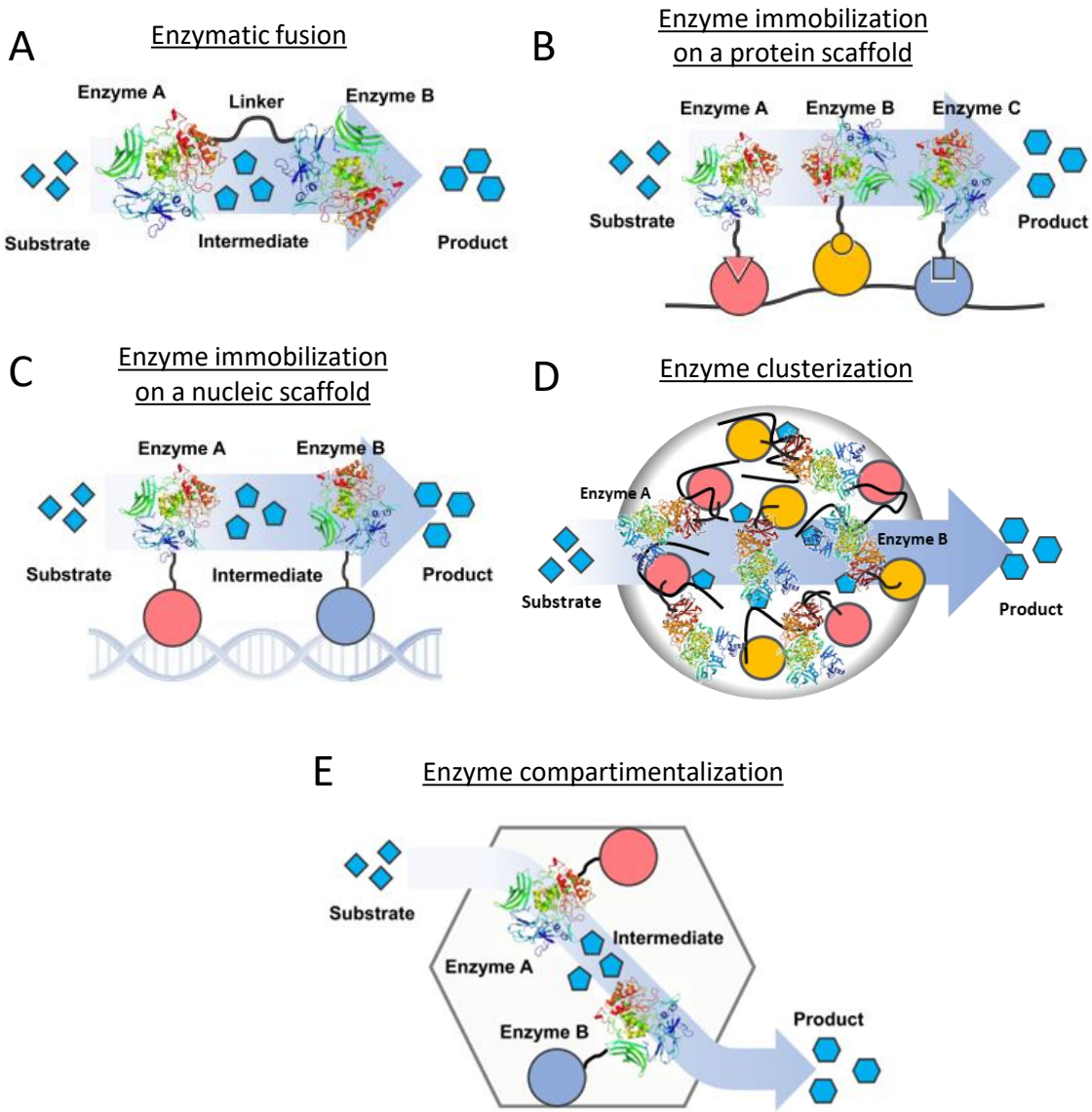


Figure 48. Enzyme spatial organization as an emerging tool to improve catalysis efficiency.

The method to organize enzymes spatially can go through enzyme translational fusion (by the intermediate of a polypeptidic linker of variable residue content, length and flexibility) (**A**) or enzyme immobilization onto a scaffold. Of note, this scaffold can be made of protein (**B**) such as in the cohesin/dockerin system where dockerin-fused enzymes associate on a cohesin scaffold. Alternatively, the scaffold can be a nucleic acid (**C**) with which DNA-binding domain-fused enzymes, for instance, can interact specifically. Enzymes clusterisation can also be induced through the formation of granules by liquid-liquid phase separation *in vivo* or *in vitro* thanks to the addition of polymers such as polyethylene glycol or by fusing the enzymes with intrinsically disordered proteins (**D**). Finally, enzymes can be addressed to a cellular compartment like a virus capsid or a bacterial microcompartment or, alternatively, a liposome. Illustration adapted from (Li *et al*, 2020).

In the context of protein fusion, both the linker size and flexibility were shown to be important (Ruiz *et al*, 2016; Bouin *et al*, 2023). Indeed, linker properties may dictate enzyme relative positioning and the substrate transit route from one enzyme to the next. Furthermore, enzyme orientation (enzyme A-B or enzyme B-A) is crucial as a N- or C-terminal linkage was shown to impact the enzyme activity, probably by perturbing their 3D structure or catalytic site (Bouin *et al*, 2023).

If protein fusion is classically used for a binary enzymatic complex, it is often superseded by other methods when the complex expands in terms of components. Moreover, no predictive tool exist for now in order to predict the best linker for a given enzyme couple, requiring to test multiple linker parameters before finding the most suited.

Enzyme immobilization onto scaffolds

Nucleic acid or protein-based scaffolds pursue to imitate enzyme natural organization enzyme within the cells. For instance, ribosomes coalesce around RNA scaffold (Kornprobst *et al*, 2016). The enzymes involved in the glycolysis associate to the F-actin cytoskeleton (Araiza-Olivera *et al*, 2013). Another example is the cellulosome, a large protein complex localized on the outer membrane of the bacteria able to degrade plant cell wall (Lamed *et al*, 1983).

The cellulosome is composed of a cohesin scaffold onto which enzymatic subunits bind through dockerin domains (figure 48B). The cohesin/dockerin system has been widely hijacked in synthetic biology to immobilize different metabolic pathways: for the production of fructose-6-phosphate or 1,3-propanediol, for instance (You & Zhang, 2013; Xu *et al*, 2021). The original system allowed to assemble up to 9 different enzymes onto one scaffold thanks to 9 existing cohesin/dockerin specific couples (Lamed *et al*, 1983). It has the advantages to be very malleable. Indeed, a new enzymatic set is implementable on the same system by simply switching enzymes fused to the dockerin modules. Also, it is secreted by the cell so the catalysis products can be easily collected.

Concerning nucleic acid scaffolds, studies are still sparse and are mainly based on aptamers (synthetic single stranded RNA which adopt a 3D structure and bind to a specific protein) or plasmids bearing sequences recognized by DNA-binding domains of either the zinc finger proteins (ZFP) or the transcription activator-like effector proteins (TALE) or the Cas9, for example (figure 48C) (Siu *et al*, 2015).

Delebecque *et al* created aptamers that self-assembled as sheets and, after fusing a hydrogenase and a ferredoxin to specific proteins recognized by the designed aptamers, they enhanced dihydrogen production by 48-fold (Delebecque *et al*, 2011).

On the other hand, a 5-fold improvement of the 1,2-PD production from dihydroxyacetone phosphate was achieved by immobilizing the 3 necessary enzymes onto a plasmid scaffold through ZFP

(Conrado *et al*, 2012). Besides, the same team showed that by tightly controlling enzyme ratio and spatial repartition, resveratrol and mevalonate production could be increased 2,5- to 5-fold compared to a random scaffolding.

Enzyme condensation by liquid-liquid phase separation

LLPS is a natural organizing phenomenon that leads to the formation of dense granules that remain relatively dynamic and mobile within cells. Many cellular processes are presently being revised because, despite previous descriptions, they would happen via a LLPS such as plasmid partitioning or transcription in both prokaryotes and eukaryotes and even BMC biogenesis (see part 1, section 3.2) (Azaldegui *et al*, 2021; Wang *et al*, 2021; Kumar & Sinha, 2022).

LLPS can be recreated thanks to intrinsically disordered proteins, polymers or RNA (figure 48D) (Küffner *et al*, 2020; Guo *et al*, 2022). Addressing the adenylate kinase to LLPS granules was shown to increase its activity by 5 times (Küffner *et al*, 2020) while the sumoylation of substrate proteins was accelerated by 36-fold when the enzymatic pathway, composed of the SAE1/2 heterodimer and Ubc9, was clustered within a granule compared to free enzymes (Peeples & Rosen, 2020).

Sequestration of a metabolic pathway

Finally, another method to organize the enzymes from a metabolic pathway is by separating them from the cell cytosol by a physical barrier through compartmentalization (figure 48E). Compartmentalization is one of the criteria of the definition of Life and, as such, it is an organizing process common to all kingdoms, from phages to superior eukaryotes. Two types of compartmentalization coexist in nature, either protein-based such as in virus or phage capsids, in encapsulins or in BMCs (Wiryaman & Toor, 2022; Chowdhury *et al*, 2014), or lipid-based like for the magnetosomes or anammoxosomes of bacteria, for eukaryotic organelles or the cells themselves (Greene & Komeili, 2012; Van Teeseling *et al*, 2013).

Compartmentalization can be a true advantage for troublesome metabolic pathways that involve toxic or volatile molecules. As presented before, catabolic BMC original functions are to sequester toxic and volatile aldehyde species, thus protecting the cells (Penrod & Roth, 2006; Cai *et al*, 2009). Of note, the physical barrier is also effective in the other direction, protecting sequestered enzymes.

Enzymes encapsulated in virus-like particles had a longer half-life, at 25°C, as well as at increasing temperatures than similar but free enzymes (Das *et al*, 2020). They also resisted to greater solvent and chaotropic agent concentrations over time. Besides, the fluorescent mNeonGreen protein fused with a degradation peptide was shielded against proteolysis upon loading within encapsulin cages (Lau *et al*, 2018), showing that protein shells could protect cargo enzymes from cytosolic proteases.

Enzyme compartmentalization can also be performed with lipid vesicles (Yoshimoto *et al*, 2008; Chaize *et al*, 2004). However, this raises several challenges among which the difficulty to produce homogeneous vesicles (*i.e*. vesicles vary in size and can be more than bilayered). The lipidic nature of the vesicles also raises the problem of substrates and products diffusion while protein-based compartments usually bear pores of different diameters (5Å for the encapsulins, 7-14Å for BMCs, 2-10nm for phage capsids) (Chaize *et al*, 2004). Besides, unlike protein-based compartments which can self-assemble *in vivo* or *in vitro*, the lipidic vesicles require external handling to produce liposome-encapsulated enzymes, limiting their use as large-scale biomolecule factories.

BMCs offer a wide range of organizational possibilities: compartmentalization if one works at the shell level (creation of minimal BMC shell), protein scaffolds if one takes advantage of the shell subunit property to self-assemble as macrostructures or LLPS granules if one exploits intrinsically-disordered proteins such as CsoS2 to condensate enzymes.

They are gaining more and more interests, recently, in the synthetic biology field as a tool to control the enzyme spatial organization and thus improve catalysis efficiency (Lee *et al*, 2018b; Lawrence *et al*, 2014; Schmidt-Dannert *et al*, 2018).

3.1.2. Repurposing BMC metabolic functions

In the last decade, hijacking the BMC natural metabolic functions for the production of desired biomolecules has been marked by several successes (Li *et al*, 2020b; Lawrence *et al*, 2014; Liang *et al*, 2017). These exploits implied to have a deep understanding of the BMC biology and to tackle different issues prior to that. The first one was to express and reconstruct a BMC shell in a common expression host. Then, it required to be able to address heterologous enzymes to the interior of these structures. Finally, so that catalysis continues as long as the substrate is furnished, exchanges between the BMC lumen and the medium/cytosol should be ensured and the eventual cofactors regenerated.

BMC shell reconstruction

Heterologous BMC shell subunits can be expressed together in *E. coli* and assemble into BMC empty shells (Lassila *et al*, 2014; Parsons *et al*, 2010a). By recombinantly expressing the full set of α-CBX shell proteins (CsoS1A/B/C, CsoS4A/B and CsoS1D) in *E. coli*, Bonacci *et al* were able to observe α-CBX with a regular polyhedral shape in TEM (Bonacci *et al*, 2012). However, all subunits were not essential and minimal BMCs, with a reduced set of shell subunits, could also be reconstructed (figure 49A), such as minimal GRM2 (CmcC+CmcD), HO BMC (BMC-H+BMC-T_1+BMC-P) or β-CBX (CcmK1/2+CcmO+CcmL) (Kalnins *et al*, 2020; Hagen *et al*, 2018b; Sutter *et al*, 2019).

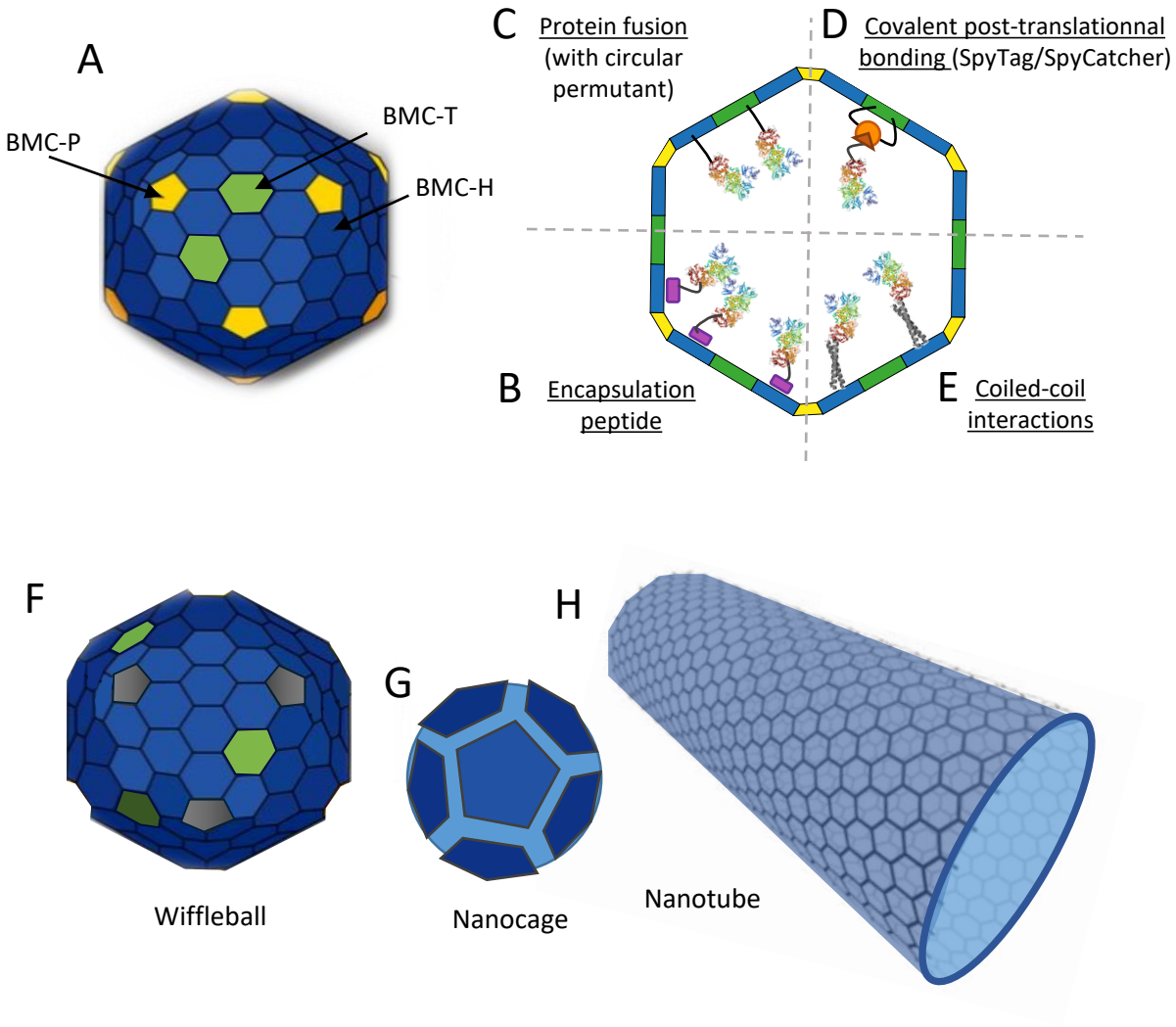


Figure 49. BMC shell and shell subunit engineering.
A. Minimal BMC shell recombinantly expressed and reconstructed in common expression strains. **B-E.** Methods to address enzymes to the BMC lumen. **F-H.** Possible macrostructures formed with BMC shell subunits. The wiffleball can be created by expressing BMC-T_1 and the unique BMC-H from *Haliangium ochraceum* in *E.coli*. A nanocage composed of 12 pentameric units can be formed by a circular permutant of PduA while the wild-type form of PduA can self-assemble as nanotubes. The minimal BMC is an illustration adapted from (Sutter *et al*, 2022).

Addressing enzymes to recombinant BMC shells

One way to encapsulate enzymes within BMCs can be by fusing the desired protein to natural encapsulation peptides (figure 49B), as it was done for a peroxidase reporter protein with the EP of PduP which successfully addressed the protein to recombinant PDUs (Lawrence *et al*, 2014).

BMC-shell 3D structures were resolved and allowed to determine that shell subunits were *a priori* all oriented with their concave face outward. Yet, the protomer N- and C-termini are generally located on this face. In that manner, if an enzyme is addressed to the BMC by directly fusing it to the N- or C-terminus of one of the shell subunits, this would immobilize it on the outer shell (equally for bait/prey tagging strategies). Exception of this protein terminus orientation is found in the circular permutants, like EutS or PduU, where the N- and C-termini localize to the convex face (Tanaka *et al*, 2010; Crowley *et al*, 2008).

Thus, to circumvent the problem of protein terminus orientation, circular permutants of PduA or of HO BMC-H were designed and led to luminal loading by coiled-coil interactions or by means of fused cargo, respectively (figure 49C) (Lee *et al*, 2018a; Ferlez *et al*, 2019b).

Another option to address cargo specifically to the BMC lumen was to use shell subunits with internal bait/prey tags (figures 49D & E). For instance, insertion of a SpyTag or SnoopTag sequence was performed between the second α-helix and the fourth β-strand of HO BMC-T_1 which gave rise to a trimer with bait tags protruding on the convex side (Hagen *et al*, 2018a; Kirst *et al*, 2022). Expression of SpyCatcher- or SnoopCatcher-tagged fluorescent proteins along with other shell subunits led to cargo loading inside the BMC lumen.

Several studies have proved successful in encapsulating heterologous enzymatic cargos into BMC through EP utilization, protein fusion or bait/prey association and achieved a proof of concept that BMCs can be used as production factories for a high diversity of molecules. Li *et al* demonstrated a 4-fold increase in aerobic H_2 production by BMC encapsulation of a ferredoxin-fused hydrogenase A along with the ferredoxin oxydo-reductase which catalyses the electron transfer from NADPH to the ferredoxin compared to free enzymes (Li *et al*, 2020b). Besides, the hydrogenase which is oxygen-sensitive was partially protected from aerobiosis inside the BMC while free enzymes completely lost their activity after 24h.

Similarly, polyphosphate production was shielded against competitive cytosolic phosphatases through encapsulation (Liang *et al*, 2017).

Another team repurposed PDU shell to encapsulate the pyruvate decarboxylase and alcohol dehydrogenase and improved ethanol yield by 20-fold compared to the free enzyme couple (Lawrence *et al*, 2014).

Alternative structures with BMC shell subunits

Manipulation of the BMC shell subunits can give rise to alternative compartments. Kirst *et al* implemented the pathway for pyruvate production within structures they denominated as wiffleballs (Kirst *et al*, 2022). These peculiar particles could be obtained by omitting the expression of a BMC-P and conducted to 6nm holes in place of the vertices (figure 49F). This assembly was only possible with HO BMCs that do not require pentamers to form closed structures, compared to elongated structures that were depicted in absence of pentamers for other BMCs (Cameron *et al*, 2013; Parsons *et al*, 2010a).

Besides, with the aim to design a novel BMC-H with different geometrical properties (bending and torsion), a second circular permutant of PduA was created and its crystallographic structure revealed a peculiar association mode (Jorda *et al*, 2016). Surprisingly, permuted PduA was no longer oligomerizing as a hexamer but instead formed a pentamer that further assembled into a 13nm nanocage (figure 49G).

Control of the shell permeability

In order to repurpose BMC metabolic functions, a critical parameter to control and potentially to engineer is the shell permeability to specific substrates and products. Shell pores should allow new substrate entry and product exit from the BMC lumen.

One way to modify the shell permeability could lie in the transfer of pore residues from one BMC-H of known properties to the BMC-H composing the minimal BMC shell (Cai *et al*, 2015b) or to create hybrid BMC shells resulting from a mixing of different shell subunits arising from distinct BMC types. Cai *et al* reported that hybrid BMCs could be obtained by recombinantly expressing CsoS1A from the α-CBX inside the β-CBX-endowed *Synechococcus elongatus 7942* (Cai *et al*, 2015b). However, the shell integrity was impacted by this mixing as depicted by the high-CO_2-requiring phenotype of the mutant strain.

Several PduA pore mutants were investigated and were shown to impact the shell permeability to the different molecules processed in the PDU. When the Lys37, placed on PduA hexamer concave face, lining the central pore, was mutated into a Glu, propionaldehyde, propionate and 1-propanol (the intermediate and products of the PDU) were accumulating in smaller quantity in the medium than with wild-type PduA PDU (Slininger Lee *et al*, 2017). This was accompanied by an increase in biomass, suggesting that these molecules were better retained in the BMC and could serve for cell growth rather than been excreted. In parallel, Chowdhury *et al* showed that PduA Ser40-pore mutant could change the size of the pore. Indeed, the Ser40His mutant shrank the pore diameter from 5,6 to 4,3Å while the Ser40Gln mutant completely occluded it (0,5Å) (Chowdhury *et al*, 2015).

With all the shell subunit diversity that exists, and once the substrate/product permeability of each will be better characterized, it should be easy to select the best suited subunit combination for desired pathway reconstruction inside a hybrid BMC.

3.1.3. BMC-H engineering

The fact that subunit concave faces are oriented outward the BMC shell might be seen as an opportunity rather than an issue. Indeed, in the context of a protein scaffold, such property could be used to immobilize an enzymatic set onto shell subunits. As we saw in the part 1, section 4.2, BMC-H has the characteristic to self-assemble and form highly organized and packed macrostructures like nanotubes or fibres or sheet (Pang *et al*, 2014; Noël *et al*, 2016; Pitts *et al*, 2012; Young *et al*, 2017). These structures are an interesting scaffold onto which a metabolic pathway could be grafted.

PduA nanotubes were hijacked to serve as a platform for ethanol production (figure 49H) (Lee *et al*, 2018b). Thanks to hetero-dimeric coiled-coil interactions, a pyruvate decarboxylase and an alcohol dehydrogenase were both addressed to coil-fused PduA. Although PduA coil seemed to affect nanotube length and packing (*i.e*. nanotubes were shorter and randomly arrayed in the cell cytosol in TEM), nanotubes still formed and enzyme scaffolding increased *in vivo* ethanol yield by 220%.

On the other hand, the canonical BMC-H EutM was also engineered to be the scaffold of several enzymatic pathways. To recall, *Clostridium* CD1918, an EutM homolog, was previously shown to form sheets that wrapped on themselves and formed Swiss-roll structures *in vivo* (Pitts *et al*, 2012) while other EutM would rather form fibres (Schmidt-Dannert *et al*, 2018). By addressing a SpyTag-GFP cargo to SpyCatcher-EutM fibres, 2 different teams obtained fluorescently-tagged structures *in vitro* but that had lost their fibre organization for the benefit of sheet formation (Zhang *et al*, 2018; Schmidt-Dannert *et al*, 2018). Surprisingly, the BMC-H resumed to fibre formation when a set of enzymes were immobilized onto the scaffold (Zhang *et al*, 2018). The enzymes were the alcohol and amine dehydrogenases which catalyse the conversion of 2-hexanol to 2-amino-hexane. Thanks to enzyme scaffolding, the pathway was improved by practically 2-fold.

EutM macrostructures have also been the object of a third attempt involving a chain immobilization of enzymes producing tagatose from lactose (Liu *et al*, 2022). In this system, EutM bore the SpyTag and recruited the arabinose isomerase that were tagged with both a SpyCatcher and a SnoopTag. Alternatively, the SnoopTag bound to its SnoopCatcher counterpart linked to the β-galactosidase. This chained-bait/prey-addressing to EutM scaffold increased slightly the tagatose produced by 1,34-fold and also improved enzyme stability over time.

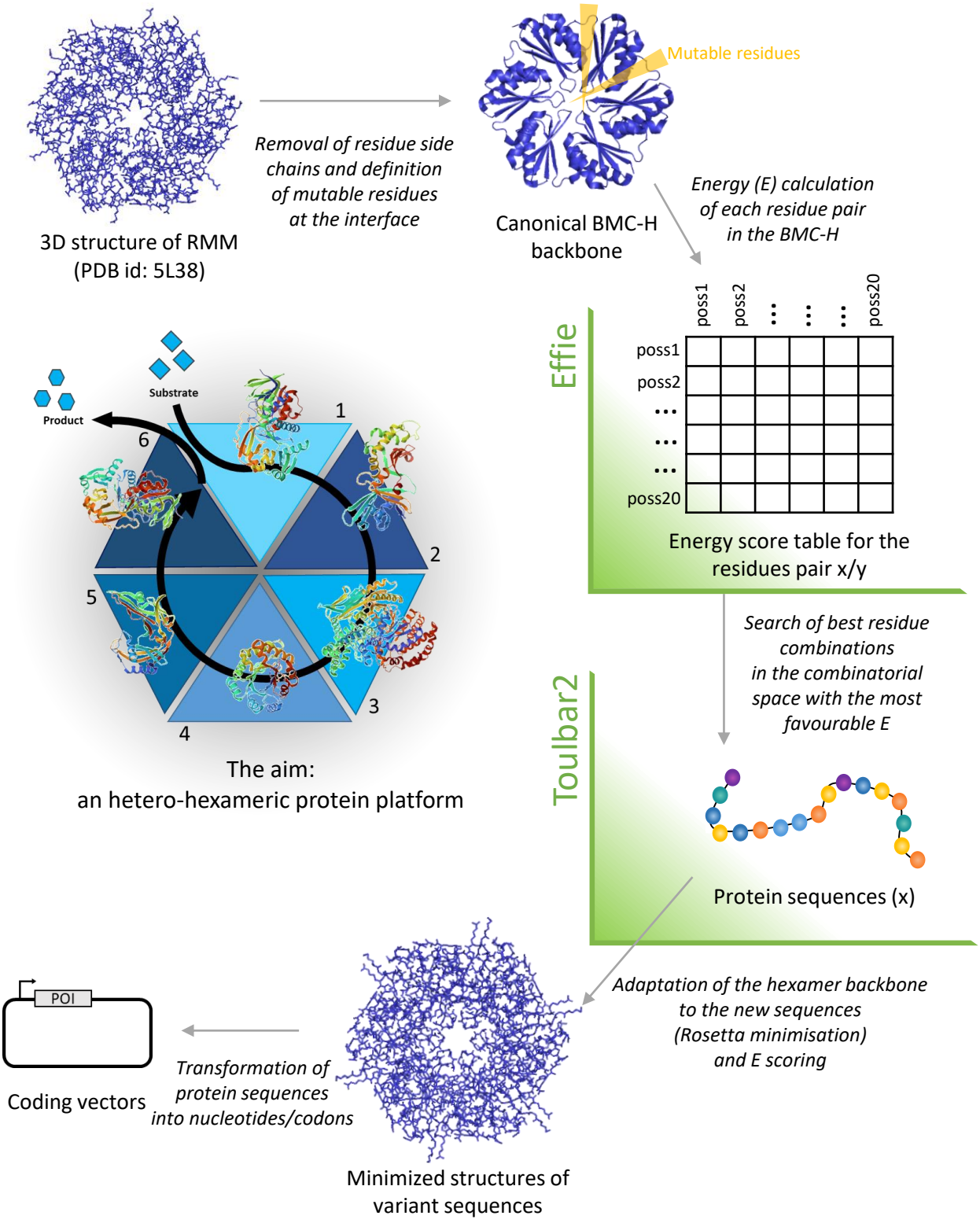


Figure 50. Steps of BMC-H variant semi-rational design.

In a more ambitious attempt, BMC-H naturally forming nanotubes were chosen to serve as nanowires for electron transport (Huang *et al*, 2020). The protein sequences of RMM and a Tyr42Ala HO BMC-H mutant were mutated to include heme-binding motifs (short peptide sequences recognized by the cytochrome C maturation system which covalently attaches a heme onto each motif). Unfortunately, these modifications abrogated the nanotube formation and, when heme-bound Tyr42Ala HO BMC-H was implemented with minimal HO BMC shell, the resulting shells appeared frequently as broken.

3.2. Design of BMC-H variants assisted by artificial intelligence

To recall, the second axis of my thesis was to create a protein platform on the basis of a hetero-hexamer. Ideally, this platform would be constituted by 6 different BMC-H and the place of each BMC-H could be controlled with precision. This would require specific intra-hexamer interfaces that allow one given BMC-H pair to interact through their interface A but preclude any other cross-interactions. Each of the 5 remaining intra-hexamer interfaces should share the same particularity.

Such hexameric platform would be of great interest in the synthetic biology field where it could serve as a production unit. Indeed, an enzymatic pathway could be implemented on the platform. By fusing one enzyme per BMC-H, this would enable the creation of a platform with up to 6 different enzymes which would benefit from a specific spatial proximity and organization in term of catalysis efficiency. Besides, this platform could be integrated within a minimal BMC shell to encompass troublesome metabolic pathways.

For the elaboration of the hexameric protein platform, we worked in close collaboration with a computer science team from the Mathématiques et Informatique Appliquées de Toulouse unit (Thomas Schiex, Marianne Defresne, Samuel Buchet and Simon de Givry) and with a modelling team from Toulouse Biotechnology Institute (Sophie Barbe, Delphine Desseaux and Jérémy Esqué). Together, they developed a 2-artificial intelligence (AI) system that generates protein sequences supposed to adopt a particular fold (and hold a function of interest for the case of an enzyme). Here, we aimed at the pfam00936 structural domain to recreate a BMC-H and expand the diversity at our disposal so that we can build the platform.

The system is composed of the energy function familiarly introduced as Effie module (Effie), created by Marianne Defresne (Defresne *et al*, 2023), and of the Toulouse Barcelona solver 2 (Toulbar2) developed in part by the team of Thomas Schiex and Simon de Givry (Allouche *et al*, 2015). Effie is a deep-learned AI which was trained on the protein data bank (PDB) to calculate the energy of

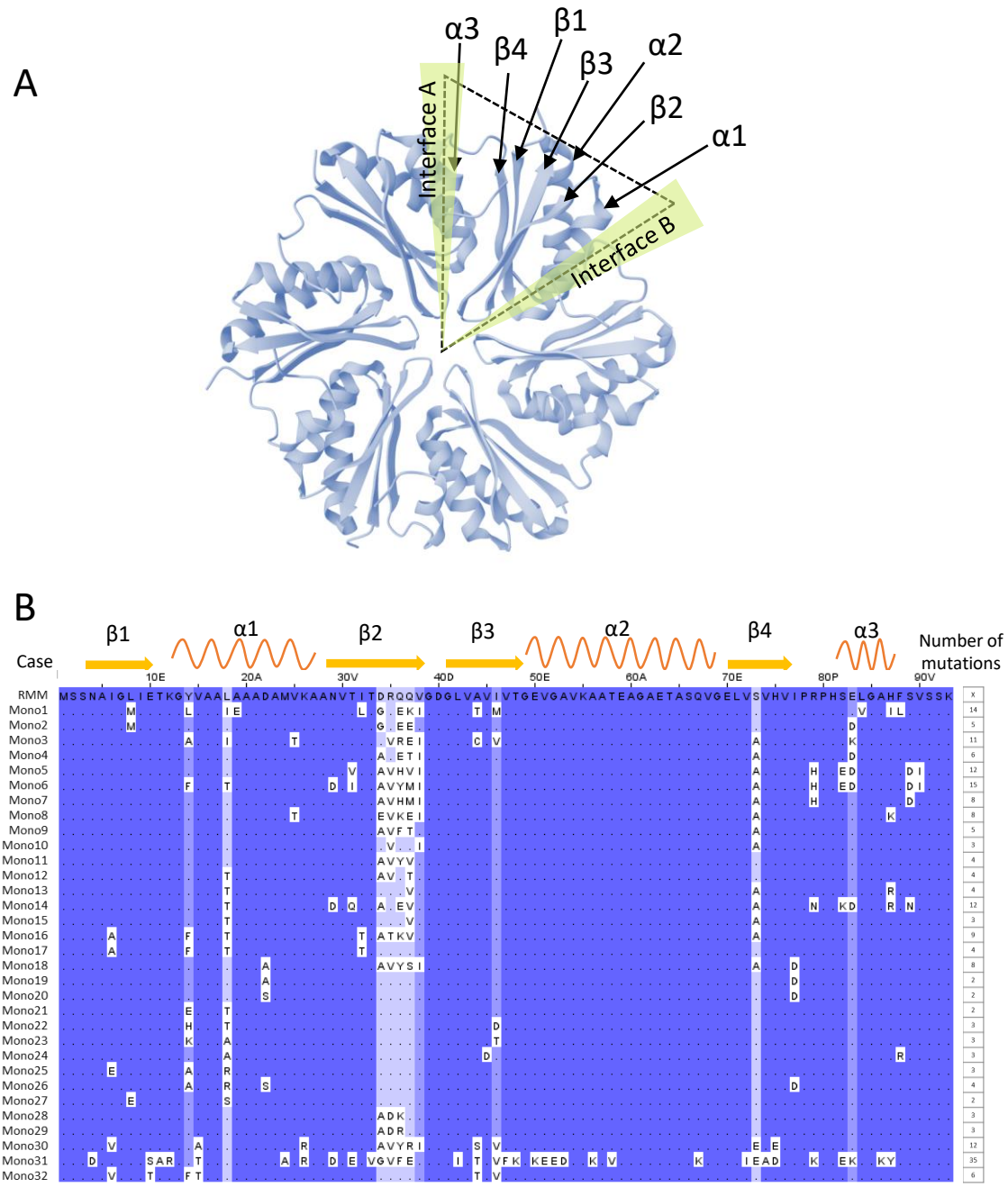


Figure 51. Alignment of semi-rationally-designed Mono sequences.
A. Ribbon representation of the 3D structure of RMM hexamer (5L38). **B.** Alignment of the Mono1 to 32 on the wild-type RMM sequence. Secondary structural elements of the canonical BMC-H fold (pfam00936) are shown on top with α-helices in orange and β-strands in yellow. The column on the right of the alignment gives the count of mutations with regard to the wild-type RMM.

a protein 3D structure on the basis of geometrical constraints (as AF2 and in contrast to Rosetta algorithm that takes into account physical forces such as Van der Waals interactions or the Coulomb law) (Smith & Meiler, 2020). On the other hand, Toulbar2 is an automated-reasoning AI that resolves optimization problem according to criteria defined by the user such as residue composition, positive state (oligomerization as a hexamer, pfam00936 fold) or negative state (discrimination of lower oligomerization state).

Briefly, a hexamer backbone is taken as the starting mesh (figure 50). Here, the design is performed starting from our model BMC-H RMM backbone (93 residues in total). The residues belonging to the intra-hexamer interfaces A and B (approximately 35 residues) are annotated as mutable (figure 51A). Thus, the combinatorial space for the AI system to explore was 20^n possible protein sequences with 20 the number of essential amino acids and n the number of mutable residues. The annotated backbone is then submitted to Effie which calculates energy scores for each residue pair present in the BMC-H protomer (that the residues be adjacent or distant in the 3D structure). As a final readout, Effie produces a score table on which Toulbar2 is launched to search for optimal sequences that would best fit the entry backbone structure, *i.e.* which have the best energy score sum. Finally, the sequences obtained are filtered by hand by the modelling team (energy minimization with Rosetta to adjust the initial backbone structure to the proposed sequences and minimized structures scored again with Effie) and transferred to our team for experimental testing.

3.3. Ability of the 2-AI system to design BMC-H variants

In order to test the modelling method ability to create new BMC-H interfaces, a series of 32 semi-rationally-designed BMC-H was designed by the modelling team by taking, as described above, the RMM sequence and mutating selected positions using the 2-AI system. Extreme cases had between 2 and 35 mutated residues, spread on both BMC-H interface, compared to RMM or, on average, variant sequences bore 6,8 mutations (figure 51B). The goal was that the variants recapitulate typical BMC-H characteristics: correct expression and solubility, oligomerization as homo-hexamers and formation of macrostructures. Of note, BMC-H self-assembly into macrostructures was not analysed but, as peripheral residues were not mutated, it should, in theory, be conserved.

Variants assembling as homo-hexamers

Each variant protein sequences was reverse-translated into DNA sequences and ordered integrated within pET29b vectors (Kan^R) which allowed the expression of a His_6-tagged form under the

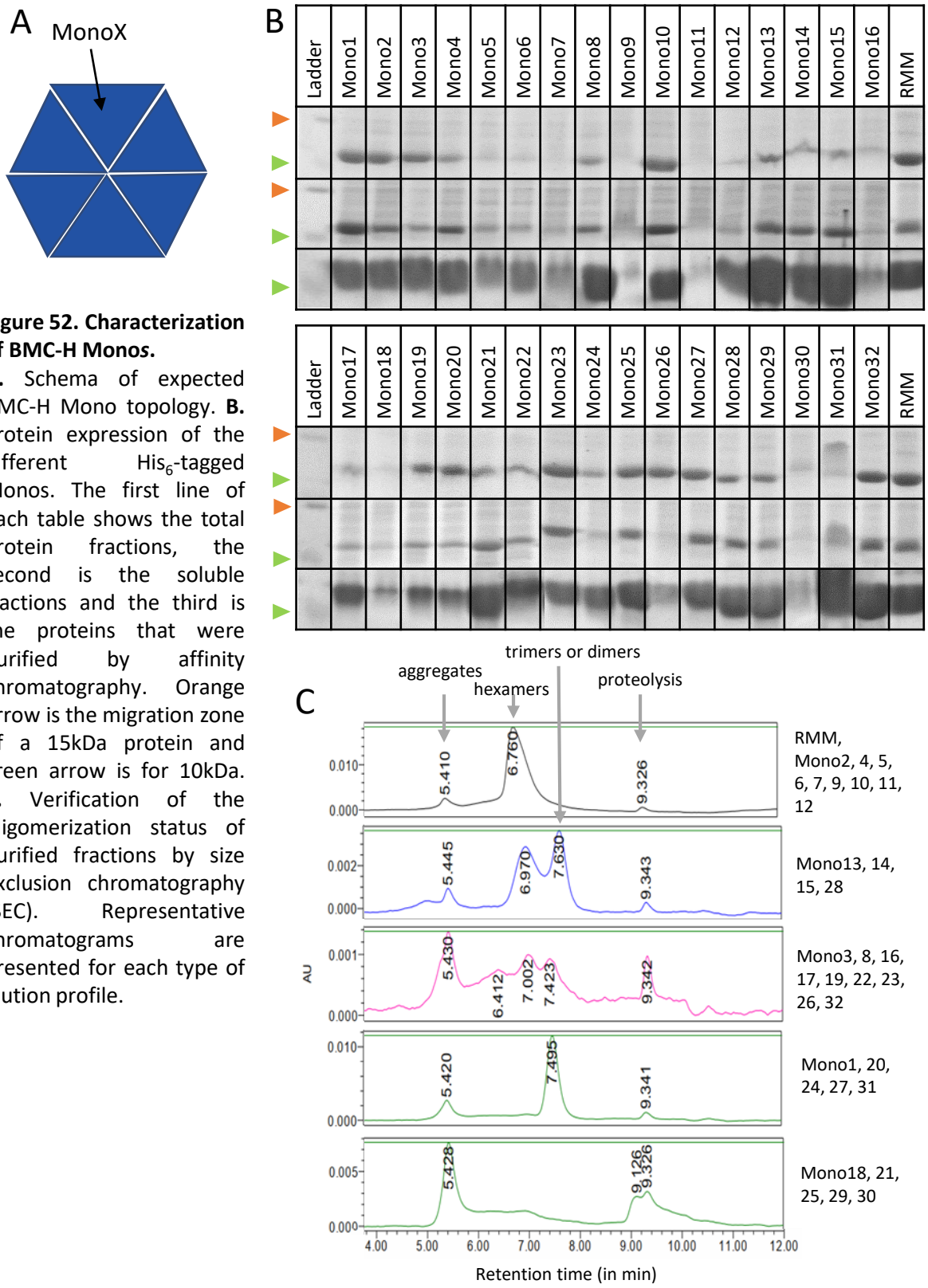


Figure 52. Characterization of BMC-H Monos.

A. Schema of expected BMC-H Mono topology. **B.** Protein expression of the different His_6-tagged Monos. The first line of each table shows the total protein fractions, the second is the soluble fractions and the third is the proteins that were purified by affinity chromatography. Orange arrow is the migration zone of a 15kDa protein and green arrow is for 10kDa. **C.** Verification of the oligomerization status of purified fractions by size exclusion chromatography (SEC). Representative chromatograms are presented for each type of elution profile.

control of a T7p. They were denominated as Mono1 to Mono32 (standing for mono BMC-H forming a hexamer and to distinguish them from BMC-H series created in the next sections).

To determine whether these variants were well-expressed, BL21(DE3) cells were transformed with each vector and induced overnight at standard conditions. A SDS-PAGE was then performed with the total protein fractions (figure 52B).

Different groups emerged with varied expression level. The Mono1 to 4, 8, 10, 13 to 15, 19 to 29 and Mono32 were highly expressed while the Mono5, 6, 12, 16 to 18, 30 and 31 were visible but depicted a fainter expression pattern. On the contrary, the Mono7, 9 and 11 seemed absent.

The analysis of proteins remaining soluble after centrifugation followed the same expression trend, with some exceptions (figure 52B). Surprisingly, Mono7 was present although the protein band remained faint. The profile of the Mono9 was also visible but appeared as a smear, suggesting protein instability or proteolysis, as did the Mono31 (the sequence that had 35 mutations). Two of the well-expressed cases were no longer visible on the soluble fraction gel, which depicted aggregated/insoluble proteins (Mono26 and 30). Alternatively, this observation could inform on a propensity to form assemblies. As I could not arbitrate between both possibilities, I decided to keep all Monos for subsequent protein purification and oligomerization status study.

Soluble fractions were subjected to purification on TALON columns in order to purify His_6-tagged proteins. Another SDS-PAGE gel was performed on eluted proteins (figure 52B). As I suspected the colorant used in the SDS-PAGE (Instant blue) to differ in staining depending on the protein residue composition, I switched for a Coomassie blue R250 staining. Protein concentration was measured in parallel at 280nm.

After purification, all the Monos were present, although with varied level. The protein concentrations ranged from 0,3mg.mL^{-1} to 4,2mg.mL^{-1}. Surprisingly, Mono7, 9 and 11 had a rather important protein concentration (1,2; 2,2 and 2,9mg.mL^{-1} respectively), contrasting with the faint bands observed in the total and soluble fraction gel profiles. Besides, Mono18 was at 2,9mg.mL^{-1} while Mono30 was at 1,2mg.mL^{-1}, yet their expression appeared lower in the purified protein gel than the Mono3 which concentration was of 1,2mg.mL^{-1}.

Together, these data finally confirmed that all Monos were expressed in our hands. They also evidenced that protein staining by classical colorants varied according to residue composition. Thus, Mono expression level should not be compared exclusively through SDS-PAGE.

To verify that the variants were still able to assemble as homo-hexamers, previously TALON-purified proteins were injected and eluted in a SEC column, after an overnight dialysis. Their migration profiles were compared to the wt-RMM (figure 52C). Protein standards of known molecular weights

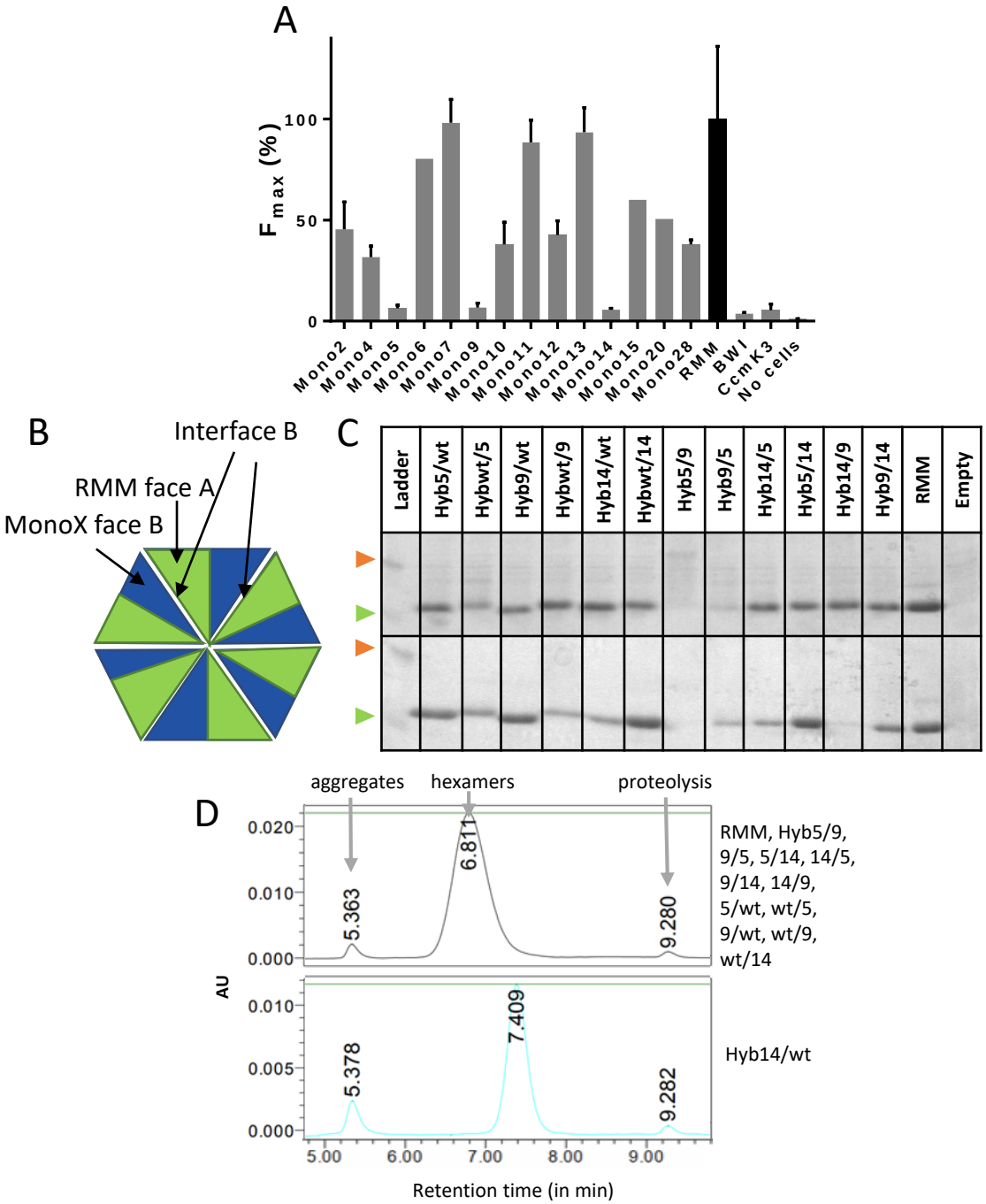


Figure 53. Investigation of BMC-H variant interface specificity.
A. tGFP assay of the cross-interaction between wild-type RMM and the different Monos. Maximal fluorescence (F_{max}) values are given as the percentage of the RMM homo-pair F_{max}. **B.** Schema of the hexamer formed by hybrid BMC-H (Hyb). An Hybrid is composed of wt-RMM face A and MonoX face B. Denomination is then Hybwt/X and for the inverse Hybrid with MonoX face A and wt-RMM face B, it would be HybX/wt. **C.** Protein expression of the different His_6-tagged hybrid BMC-H (Hyb) after 16h of IPTG induction. The top raw shows the total protein fractions, the bottom one corresponds to the proteins that were purified by affinity chromatography. Orange arrows denote the migration zone of a 15kDa protein and green arrows for 10kDa. **D.** Verification of the oligomerization status of the Hybrids by size exclusion chromatography after protein purification on TALON columns. Representative SEC elution profiles are presented for each behaviour group.

were also injected in order to calculate the equivalence between the retention time of a specie and its molecular weight (in kDa).

RMM eluted after 6,8min which corresponded approximately to the retention time of a hexamer (calculated MW ≈ 78kDa). As for the different Monos, they had diverse behaviours. One group, the most interesting one, was composed of Monos that were still forming homo-hexamers as the major specie: Mono2, 4 to 7 and 9 to 12. By contrast, no hexameric association was evidenced for Mono18, 21, 25, 29 and 30 for which potential aggregates and proteolysed peptides were observed (retention time of 5,4min and longer than 9min, respectively). Assembly intermediates, with a retention time approaching the values that would be expected for a BMC-H dimer or trimer, were monitored for the Mono1, 20, 24, 27 and 31. Mixed intermediate and hexameric species were present for the Mono13 to 15 and Mono28. Finally, the Mono3, 8, 16, 17, 19, 22, 23, 26 and 32 were composed of instable species as denoted by the large massif encompassing the region from 5,4 to 7,4min.

In summary, 2 groups of Monos proved to be of interests: the groups for which hexamer formation was clearly evidenced. Of note, although these cases seemed to associate as homo-hexamers, nothing guaranteed that the semi-rationally-designed sequences adopted the typical BMC-H fold. Further studies such as X-ray crystallography should be performed on these Monos to determine their exact 3D structure.

Probing the interface specificity of semi-rationally-designed variants

In a second phase, I wanted to know whether the newly designed interfaces were orthogonal. The Monos were designed starting from the RMM backbone. In that manner, there might be 2 possibilities: either the mutations implemented on the variants had profoundly changed the interfaces, preventing the Monos to interact with the wt-RMM, or not and the Monos were still able to form a hetero-hexamer with RMM. In the first case, this would imply the creation of an exclusive interface, *i.e.* a BMC-H able to interact with itself but not with another BMC-H homolog (here RMM), a characteristic that would be necessary for the elaboration of the hexameric platform.

The potential compatibility was assessed in tGFP by combining individual Mono with the wt-RMM (figure 53A). The Monos were tagged with the GFP10 and RMM with the GFP11. Of note, only the Monos for which homo-hexamer formation was visualized in SEC were screened.

The Mono6, 7, 11 and 13 had a F_{max} value similar to the RMM homo-pair, indicating that these Monos were interacting with the wild-type form. A second group composed of the Mono2, 4, 10, 12, 15, 20 and 28, although affected (around 40 to 50% of RMM F_{max} value), were still able to cross-interact with RMM. Remarkably, the Mono5, 9 and 14 had signals comparable to the negative control threshold. Yet, these cases corresponded to variants that were purified with good yields in the previous experiments (figure 52B).

Accordingly, overall data indicated that mutations present on these Mono interfaces had probably impeded the cross-interaction with RMM.

The Mono5 and 14 bear 12 mutations while the Mono9 has 5. For the latter, mutations localized exclusively on the β-strand 2 that contacts α-helix 1 and β-strand 2 on the adjacent monomer, in the interface A (figure 51). Mutations of the Mono5 and 14 were situated mainly on the β-strand 2 and the small α-helix 3 but they differed in nature. The mutations of the Mono5 increased the hydrophobicity of the β-strand 2 compared to the wt-RMM, alike the ones of the Mono9, and modified the α-helix 3 (which also interacts with the α-helix 1 and β-strand 2 of adjacent monomer, in the interface A) with negatively charged residues. On the contrary, the Mono14 had more negatively-charged residues on the β-strand 2 while α-helix 3 was switched to bear positively-charged residues (figure 51). These charged residues on the Mono5 and 14 might have created electrostatic repulsion of the wt-RMM. On the contrary, switching the DRQQ motif on the β-strand 2 for hydrophobic residues might have abolished interactions with counterpart β-strand 2 DRQQ motifs on both A and B interfaces.

Overall, these data suggested that the 2-AI system succeeded in creating variants which recapitulated natural BMC-H characteristics. Besides, tGFP assay permitted to determine that some of the variants lost the ability to cross-interact with the wt-RMM on which the modelling process was based. Hence, exclusive interfaces seemed to have been created by implemented mutations.

However, BMC-H have 2 inequivalent interfaces A and B (figure 51A). In the interface A lie the small α-helix 3, the β-strand 4 and the C-terminal half of the β-strand 2. The opposite interface B is composed of the α-helix 1 and the N-terminal half of the β-strand 2. With present results, it was not possible to ascertain whether the 2 interfaces were simultaneously exclusive or on the contrary the structural incompatibility was deriving from modifications on only one of them. Indeed, one face unable to cross-interact with a BMC-H homolog might be sufficient to prevent fluorescence in the tGFP screens. As BMC-H interfaces are mainly composed of hydrophobic patches, the protein would aggregate and be subsequently degraded if any of the 2 interfaces remained exposed to the aqueous milieu, no matter if the second interface was already involved in an interaction with the homolog. Yet, this was a crucial information for the continuation of this program.

Independent probing of each interface specifity with hybrid variants

In order to identify which interfaces were orthogonal or whether the 2 were participating in the specificity, hybrid variants were created based on the Monos that did not cross-interact with RMM in the tGFP assay. Hybrids were made of the Mono face B and RMM face A (figure 53B). In that manner, the same hybrid repeated 6 times within the hexamer and led to 6 repeated identical interfaces (former Mono interface B), thus allowing to study one interface kind separately. Inverse cases were

also constructed: hybrids with RMM face B and Mono face A, giving rise to 6 repeated interfaces that recomposed the former Mono interface A. Besides, hybrids composed of mixed Monos were constructed: Hybrid 5/9, for instance, resulting from Mono5 face A plus Mono9 face B, and conversely Hybrid9/5. His_6-tagged Hybrids were expressed from a pET29b in BL21(DE3) cells and total protein fractions were analysed by SDS-PAGE (figure 53C). In parallel, the proteins were purified by affinity chromatography (TALON columns) and their concentration measured at 280nm.

All Hybrids appeared over-expressed except the Hyb5/9 and Hyb9/5 which bands were visible in SDS-PAGE but in a much lower intensity. In comparison, protein concentration of these cases was more elevated than depicted on the gel. Indeed, the Hyb5/9 and Hyb14/9 were at 0,9mg.mL^{-1} and Hyb9/5 had a concentration similar to Hyb14/5 (0,3mg.mL^{-1}). This differences between the gel profiles and the protein concentrations could be indicative of aggregated material or higher-ordered assemblies.

Previously purified Hybrid proteins were analysed on a SDS-PAGE gel (figure 53C). Hybrid expression pattern was more consistent with measured protein concentrations with the exception of the Hyb5/9 and Hyb14/9 that were practically absent from the gel despite their high concentration. For the Hyb5/9 that was also barely present in the total protein fraction gel, this hinted again to a problem with the Instant blue protein staining. By contrast, absence of the Hyb14/9 could be explained by 2 phenomena: either the Hyb14/9 was self-assembling into macrostructures that remained with the pellet upon centrifugation or it was aggregated/insoluble, then only visible in the total fraction, hinting to non-interacting faces.

To determine which possibility was the correct and further characterize the Hybrids that were correctly expressed, purified hybrids were subjected to a SEC (figure 53D).

Unexpectedly, all Hybrids were migrating with the same retention time as the wt-RMM (6,8min; calculated MW ≈ 81kDa). Thus, they were oligomerizing as homo-hexamers. The only exception was the Hyb14/wt that had a retention time of 7,4min, the estimated time of a BMC-H dimer or trimer (calculated MW ≈ 39kDa).

By studying the variant interfaces A and B separately, I did not find clear proofs of sufficient specificity. Indeed, hybrid variants depicting only one kind of interface were still able to form homo-hexamers. Even the Hyb14/wt seemed to associate though not as a hexamer as it would be expected. This contrasted with the Mono/RMM tGFP assay in which the Mono5, 9 and 14 could not cross-interact anymore with the wt-RMM. However, this showed that, most likely, the number and type of mutations introduced for this study did not impacted enough each interface as to prevent interactions with RMM nor to create specific interfaces that selectively allow interaction with the Mono itself. Probably, mutations on residues involved in both interfaces acted in synergy to impede cross-interaction with RMM in the Monos.

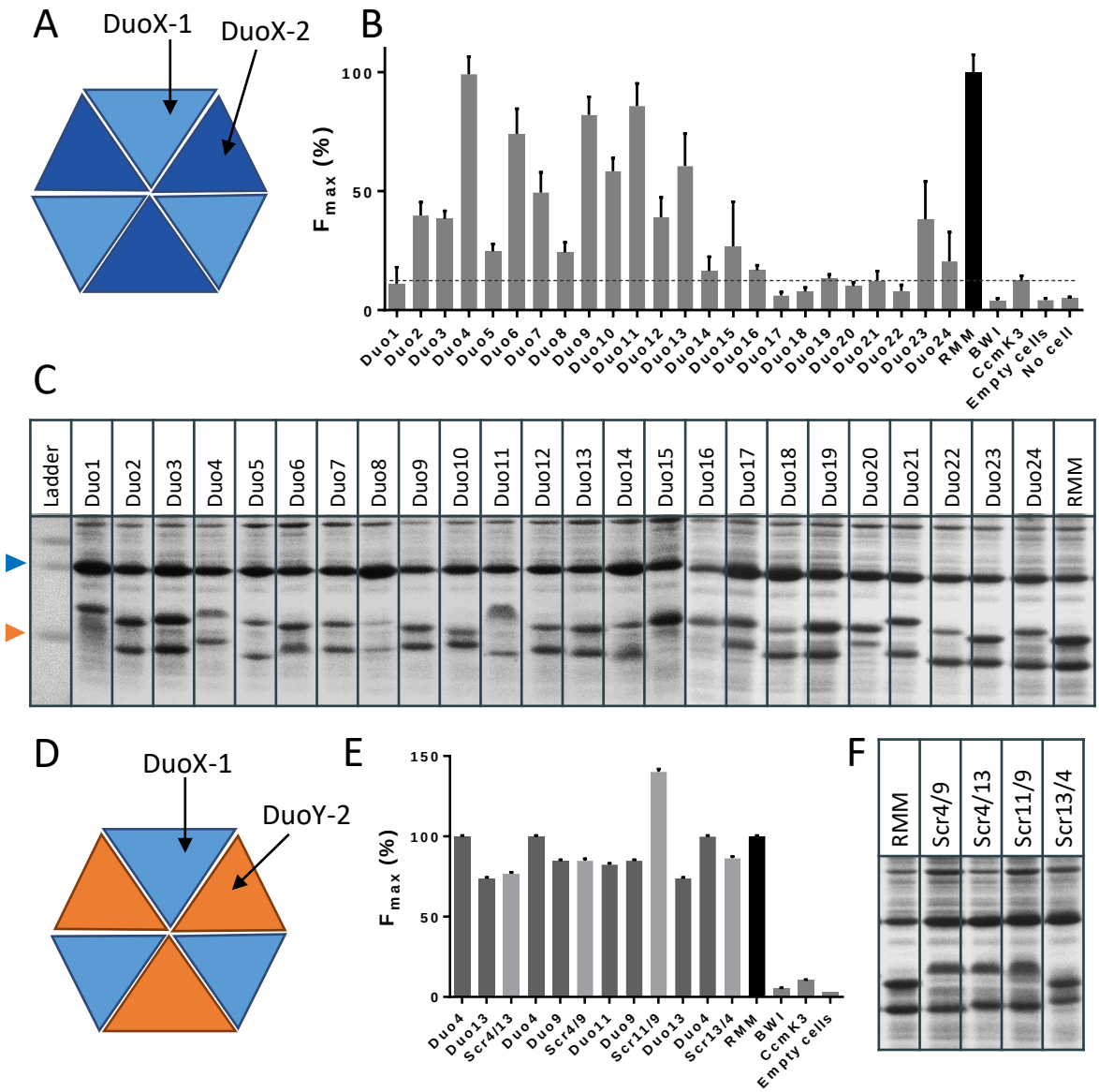


Figure 54. Characterization of BMC-H Duos.

A. Schema of the hexamer formed by a BMC-H DuoX. A Duo is composed of 2 BMC-H variants, DuoX-1 and DuoX-2. **B.** Hetero-hexamer formation by the BMC-H Duos. The Duos were assayed in tGFP for their ability to form hetero-hexamers. Maximal fluorescence (F_{max}) values are given as the percentage of the RMM homo-pair F_{max}. **C.** BMC-H Duo expression verification. Cell expressing the GFP-tagged forms of the Duos were induced with 10µM IPTG for 16h. Total protein fractions were recovered and analysed on SDS-PAGE 18%. The blue arrow indicates the migration zone of a 25kDa protein while the orange arrow is for a 15kDa protein. **D.** Schema of the hexamer that might be formed in the scramble Duo (Scr). One Scramble is composed of the POI1-GFP10 from DuoX (DuoX-1) and of the POI2-GFP11 coming from the DuoY (DuoY-2). **E.** Scramble Duo were assayed in tGFP and compared with the original Duos. F_{max} values are given as the percentage of the RMM homo-pair F_{max}. **F.** Scramble Duo expression verification. Total fractions were analysed as in the panel C.

To sum up, albeit the 2-AI approach succeeded in designing new variants, carrying a considerable number of mutations with regard to the template sequence, it did not created specific interfaces that would permit to control consecutive BMC-H organization within the hetero-hexamer.

3.4. Hetero-hexameric platform composed by a BMC-H duo or trio

Given the results obtained with the Monos and Hybrids in the search for orthogonal interfaces, the 2-AI system parameters were modified. We now targeted combinations of 2 or 3 monomers (referred to as Duo and Trio BMC-H) that would associate as AB or ABC dimers or trimers which would further oligomerize, giving rise to ABABAB or ABCABC hetero-hexamers. This implied the implementation of negative design constraints to the system to rule out AAAAAA or BBBBBB or CCCCCC homo-hexamer formation or any other combinations not obeying the defined order and number of each BMC-H within the hexamer.

Essentially, new search criteria were affected to Toulbar2: besides finding an hetero-hexamer ABABAB or ABCABC (figures 54A & 55A) with the most favourable global Effie energy score, it should ensure a less favourable or even unfavourable (less negative or even positive) score for the negative designs above mentioned. A series of BMC-H Duos and Trios (2 or 3 different BMC-H that work together to form hetero-hexamers) were obtained. The characterization of each series will be presented separately in the next sections.

Hetero-hexamers composed by a duo of variants

In order to increase the diversity of the BMC-H variants that would be proposed by Toulbar2, several runs were performed starting from different backbones. Besides the original RMM, the Mono2, 4, 5, 9, 12 and 28 were selected on the basis that they could form homo-hexamers. In that respect, 14 duos were designed (Duo1 to 14; supp figure 2). Also, for the modelling team to compare their 2-AI system with classical algorithm used for *de novo* protein design, 14 additional duos were conceived with ProteinMPNN (Duo15 to 24) (Dauparas *et al*, 2022). The Duos were constructed directly in tGFP bicistronic vectors and transformed in BL21(DE3) cells. Fluorescence apparition was then monitored for 16h, after a 10µM IPTG induction (figure 54B).

Some Duos were found clearly positive for an interaction with a GFP signal similar or slightly lower than the RMM reference homo-pair (Duo4, 6, 9 and 11). The Duo7, 10 and 13 had a F_{max} value that was between 50 to 60% of the RMM value, suggesting also an interaction. On the other hand, several Duos had a fluorescence signal that were under or near the CcmK3 negative threshold. These Duos were the

Duo1, 14 and 16 to 22. The PPI status of the last group of Duos was quite uncertain as their fluorescence ranged from 40% to 20% of the RMM F_{max} value.

To explain such discrepancies in GFP fluorescence, the protein expression was verified in SDS-PAGE (figure 54C).

Practically all variants of each Duo were correctly expressed, being present in the total protein fractions at comparable levels than the wt-RMM. The only exception were for the GFP11-tagged variants of the Duo1, 15 and 16. Indeed, only the GFP10-tagged variants, which are always the upper of both POI bands on the gel, were visible. For the Duo1, the GFP11-tagged variant could have been subject to proteolysis as revealed by the smear on the gel. Surprisingly, while the Duo4 fluorescence was as high as the one from RMM, the protein level of each BMC-H composing it was significantly lower than the RMM homo-pair. This might reflect the propensity of RMM hexamer to self-assemble and form nanotubes, *i.e.* a pool of hexamers inaccessible for the GFP reconstitution, as concluded in the first chapter of this thesis. Then, probably the Duo4 was not involved in such inter-hexamer interactions.

Thus, the lack of fluorescence in the tGFP assay were due to the absence of one of the tGFP partners for the Duo1, 15 and 16. Furthermore, the weak protein expression for the Duo8 correlated well with its low F_{max} value. For the other cases that had a low fluorescence, this might be owed to a negative PPI as both partners were visible on the SDS-PAGE gel. The majority of Duos that were positive for heteromer formation were the couples designed by the 2-AI system. Of note, only 1 out of the 14 Duos proposed by ProteinMPNN was found positive in tGFP (Duo23). This showed the superiority of the 2-AI system over ProteinMPNN algorithm in designing BMC-H variants. Besides, this highlighted the correctness of Effie and Toulbar2 to predict heteromer formation from a BMC-H couple.

Then, the specificity of the Duo interfaces were probed. The POI1-GFP10 from DuoX and the POI2-GFP11 from a DuoY were scrambled on the same tGFP vector: (1) POI1 from the Duo4 and POI2 from the Duo13 or (2) the opposite or (3) POI1 from the Duo4 and POI2 from the Duo9 or (4) POI1 from the Duo11 and POI2 from the Duo9, named Scr4/13, 13/4, 49 or 11/9, respectively (figure 54D). Of note, these cases were selected for the test because their original Duos were positive in tGFP and that, by contrast, their scramble versions were predicted to be non-interacting. Indeed, the Scr4/13, 13/4, 49 and 11/9 combinations were modelled as a hexamer by the computational design team of TBI and the energy scores of their hetero-hexamer structures were unfavourable. Their cross-interaction was assayed in tGFP and in parallel, protein expression was analysed.

Unexpectedly, all the Scrambles had a F_{max} value equivalent to the original Duos (figure 54E). The Scr119 fluorescence was even significantly higher than the Duo9 or 11 (1,7-fold higher). However, only

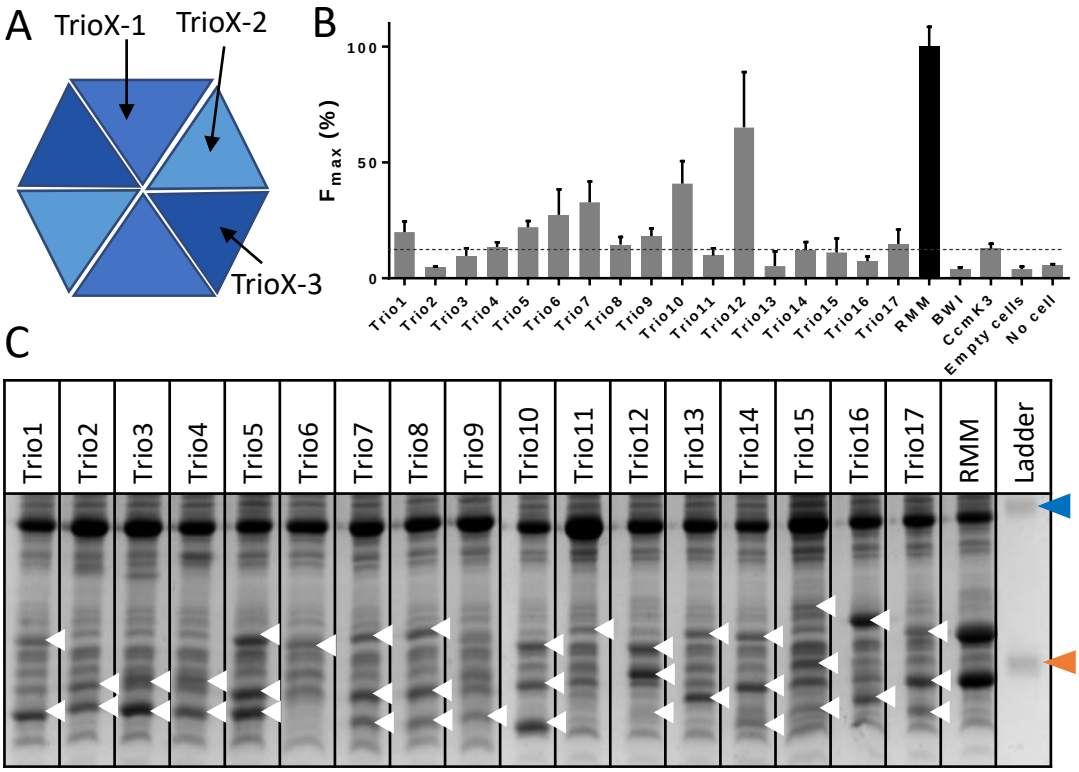


Figure 55. Characterization of BMC-H Trio*s*.

A. Schema of the hexamer formed by a BMC-H TrioX. **B.** Hetero-hexamer formation by the BMC-H Trios observed in tGFP. Maximal fluorescence (F_{max}) values are given as the percentage of the RMM homo-pair F_{max}. **C.** BMC-H Trio expression verification. Total protein fractions were recovered and analysed after a 16h induction with 10µM IPTG. Whites arrows point to POI bands. The blue arrow indicates the migration zone of a 25kDa protein while the orange arrow is for a 15kDa protein.

the expression of the POIs from the Duo4 seemed increased in the Scramble cases in which it was included, compared to the original Duos (figures 54C & F).

In the modelling process, the negative design constraints did not take into account to discriminate hetero-hexamer formation with POIs from other Duos. Anyhow, cross-interactions between variants from different Duos were observed, indicating that interfaces were very promiscuous and could accommodate a certain number of homologs.

A general conclusion for hetero-hexameric platform design is given at the end of the next section.

Increasing hetero-hexameric platform complexity with a trio of variants

In parallel of the Duo study, a series of 17 variant trios were proposed by the modelling team. The idea was to have a hetero-hexameric platform with 3 different BMC-H to be able to immobilize a more complex metabolic pathway. Toulbar2/Effie proposed 6 Trios (Trio1 to 6; supp figure 3). For comparison, ProteinMPNN was interrogated again and proposed 11 Trios (Trio7 to 17). The Trios were constructed on tGFP vectors as follow. The POI1 and 2 were tagged with the GFP10 and 11, respectively, while the POI3 was tagged in C-terminus with a Flag peptide. They were placed under the control of the same T7p which gave rise to a tricistronic mRNA upon transcription. The His_6-tagged GFP1-9 remained independently transcribed from the same vector. POI genetic order in the tricistron followed POI numbering. Hetero-hexamer formation was probed in a tGFP assay and compared to the RMM homo-pair (figure 55B). Of note, the tGFP results would only evidence hetero-hexamer deriving from interactions between the POI1 and 2 as only these 2 carried GFP tags, allowing GFP reconstitution.

Contrasting with the Duo study, the large majority of the Trios had a very low fluorescence, under or near the negative threshold (Trio2, 3, 4, 8, 11 and 13 to 17). Only the Trio12 reached a F_{max} value neighbouring 70% of RMM fluorescence, indicating heteromer formation. The Trio7 and 10 reached 40% of RMM signal while a few others were slightly fluorescent (Trio1, 5, 6 and 9).

Total protein expression was monitored by SDS-PAGE to explain such low fluorescence signals (figure 55C).

Surprisingly, the 3 POI concomitant expression was verified for few Trios, not necessarily the ones that showed fluorescence in the tGFP assay: Trio5, 7, 8, 10, 12, 14, 15, 17. Besides, expression pattern was unequal between POIs of the same Trio. Of note, the third POI was generally less expressed than its counterparts. This fact was probably due, in part, to the tricistron genetic organization where the POI3 is in the last position. Indeed, genetic organization was shown to impact protein level with pole position sequence becoming the more abundant protein and inversely for subsequent sequences (Gerngross *et al*, 2022; Lim *et al*, 2011).

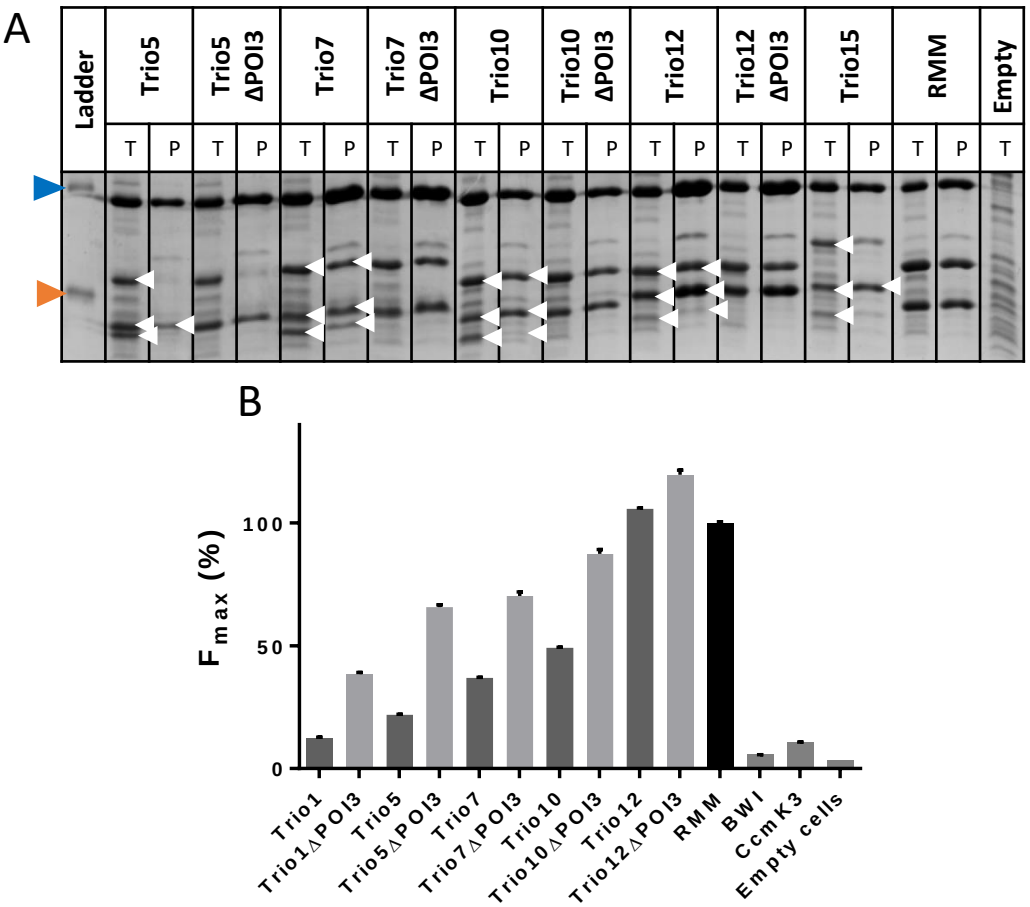


Figure 56. Investigation of interface specificity in BMC-H Trios.
A. Verification of the presence of each POI within formed hetero-hexamers. Proteins were recovered after a 200µM-IPTG induction for 4h (T) and purified by TALON affinity chromatography (P). Thus, all proteins that were complexed with the His_6-tagged GFP1-9 were co-purified. Also, the POI3 sequences were retrieved from the Trio vectors. Proteins deriving from these constructs were processed in the same fashion. Whites arrows point to POI bands. The blue arrow indicates the migration zone of a 25kDa protein while the orange arrow is for a 15kDa protein. **B.** Trios lacking the POI3 were assayed in tGFP and compared to the original Trios. F_{max} values are given as the percentage of the RMM homo-pair F_{max}.

In order to determine whether the POI3 was also included within the POI1/2 heteromers, the Trios in which all 3 variants were well expressed and produced a GFP signal were further studied. Cells expressing the Trio5, 7, 10 and 12 (plus Trio15 as a fluorescence-negative case) were induced with 200µM IPTG for 4h and subjected to protein extraction without detergent. A purification by affinity chromatography (TALON column to retain His_6-tagged protein) was performed on the soluble proteins to recover all the proteins that bound directly or indirectly to the His_6-tagged GFP1-9 (GFP10 or 11-tagged variants and any variant that would interact with them inside a hexamer). Afterwards, purified proteins were analysed by SDS-PAGE and compared with the total fractions (figure 56A).

The 3 variants were observed in the total fractions of all the Trios. Of note, the POIs 1 and 2 were present in all the total fraction samples collected from the cells expressing the Trio tGFP construct missing the POI3 (TrioXΔPOI3). This allowed to spot precisely the band corresponding to the POI3 in complete Trios as the POI3 was globally less expressed than its partners. In the purified fraction of the Trio7, all 3 variants were clearly evidenced. The Trio10 and 12 seemed to follow that trend although the third POI profile was more sparse than the Trio7 and than in the total fractions. As for the Trio5 and 15, at least the POI1-GFP10 was missing after purification.

Globally, these data showed that heteromers could form with a trio of variants. However, expression pattern on the SDS-PAGE gel depicted a disproportion between the different constituants of a Trio in the total and, more particularly, in the purified fractions. Even though explained by the tricistronic genetic organization, this fact implied that the ratio of each variant within the heteromer was not equal. Thus, if the variants were not in stoichiometric proportions, the expected hetero-hexamer ABCABC was unlikely.

Still, Trio7, 10 and 12 were of greatest interest as they displayed a potential to form a complex hetero-hexameric platform. BMC-H disproportional ratio could be advantageous for a metabolic pathway involving enzymes with different catalytic rates. Indeed, limiting-rate enzymes could be grafted on the most abundant BMC-H while enzymes with the better rate could be fused to the third and last BMC-H of the operon, equilibrating the global pathway rate.

Regarding the comparison between the 2-AI system and ProteinMPNN to design hetero-hexamers composed of 3 different BMC-H, ProteinMPNN was found to work better, contrasting with the higher performance of Toulbar2/Effie for the design of Duos. Indeed, all Trios that obtained the best results in expression, interactions and heteromer formation (Trio7, 10 and 12) were designed by ProteinMPNN modelling. This highlighted the necessity to improve the 2-AI system to make it more suitable for the design of more complex hetero-hexamers.

Finally, to uncover whether semi-rationally-designed variants within the effective Trios displayed specific interfaces, the Flag-tagged POI3 was removed from the most interesting tGFP Trio constructs (Trio1, 5, 7, 10, 12ΔPOI3). Fluorescence of these new vectors was monitored and compared to the original Trios coding for all 3 POIs (figure 56B).

Removing the POI3 did not abolish the fluorescence signal as expected. On the contrary, each TrioΔPOI3 demonstrated an increase in fluorescence compared to the original Trios. This showed that heteromer could still associate even though one of the partners was absent, thus, indicating that the interfaces of the variants were not orthogonal because the same face could interact with the POI3 as well as with the other POI of the Trio.

Increased fluorescence could be explained by an increased protein expression. Indeed, producing 3 proteins is more resource-demanding than producing 2. In that respect, for the same amount of resources, more proteins could have been produced for the TrioΔPOI3. To ascertain the cause of the increased fluorescence, total protein expression was studied as well as protein presence after purification as described earlier (figure 56A). Of note, the original Trios and TrioΔPOI3 were processed on the same gel to facilitate comparison.

Surprisingly, the amount of proteins produced in all the TrioΔPOI3 were similar to those of the original Trios. Thus, the increased fluorescence observed was not due to an increase in protein content. Without the POI3, ratio imbalance within the heteromer no longer existed and each variant appeared to be in equivalent proportions. Yet, stoichiometry is important for the GFP reconstitution: a 1/1 ratio between POIs would be required to obtain maximal fluorescence signal (up to 3 GFP reconstituted per hexamer). Then, increased fluorescence in the Trios lacking the POI3 could be the results of a greater number of reconstituted GFP per hetero-hexamer. If one looked at this increased fluorescence in the other sense, it implied that POI3 was present in the heteromer as it generated a fluorescence decreased due to fewer GFP reconstituted per hexamer.

The variants of the Trios were under the control of the same promoter and transcribed as a single tricistronic mRNA, mimicking a natural operon. Yet, a majority of Trios were not positive in the tGFP assay. BMC-H gene order in the *pdu* operon was shown to be crucial for BMC shell assembly (Parsons *et al*, 2010a; Chowdhury *et al*, 2016). Indeed, when *pduA* was moved from the beginning to the end of the operon, aberrant shells formed within cells, reminiscent of *pduA* deleted strains. Also, PduJ could only complement a *pduA* deletion if *pduJ* was placed in the first position of the operon. More generally, genetic order plays a role in protein complex assembly (Wells *et al*, 2016).

In the context of a BMC-H hexamer, one could suppose that the first translated protein of the operon could act as a nucleating support for next protein. One thing that should be considered to

increase the rate of success of hetero-hexamer formation for a given Trio would be to test the 6 possible BMC-H genetic orders (POI1/2/3, POI1/3/2, POI2/1/3, POI2/3/1, POI3/1/2 or POI3/2/1) and thus determine which variant best serves as a hetero-hexamer-nucleating centre.

To conclude on this section, hetero-hexamers were designed thanks to a 2-AI system (Effie and Toulbar2) that had to cope with different constraints: the discrimination of hetero-hexamers that would not respect the desired internal organization ABABAB or ABCABC and of homo-hexamers. The system was asked to proposed two independent series: one of BMC-H Duos and one of BMC-H Trios, all supposed to recompose hetero-hexamers.

My results revealed that several Duos and Trios were successfully associating as heteromers. However, many questions remain before envisioning to use them as a hexameric platform. Did the BMC-H Duos and Trios that were positive in tGFP oligomerize as hexamers? Do monomers A, B and C were in stoichiometric proportions within the heteromer? Was there an alternation between monomers A, B and C, respecting the positive design constraint?

As it was performed for the Mono characterization, the Duo and Trio oligomeric state could be determined by SEC in the next future. Monomer stoichiometry inside the heteromer could be analysed by native mass spectrometry on purified heteromers. As for the internal organization, the key technique to tackle the question would normally be X-ray crystal structure solving. However, past hetero-hexamer studies have proved to be problematic, resulting in only homo-hexameric units in the crystal (Garcia-Alles *et al*, 2019; Sommer *et al*, 2019). Yet, both BMC-H were present in the original protein mixture. This phenomenon could depict that hetero-hexamers are less stable than homo-hexamers, in the sense that they may constitute more dynamic complexes, able to switch one BMC-H subunit for another. Alternatively, hetero-hexamers could be symmetry breakers, forming angled, twisted or screwed hexamers that would prevent tight and stable packaging. As crystallisation would select the more stable species, only homo-hexamers would be represented.

On the other hand, hetero-hexamers might not form perfectly symmetric units. Indeed, BMC-H ratio and organization could vary within the hexamer. Besides, these hexamers could have up to 6 different orientations in the crystal lattice (considering that hexamers assembled as sheets as they did *in vitro* or *in cellulo*). Then, it is also possible that hetero-hexamers were not observed in both studies because X-ray crystallographic results obtained were an averaging of all hexamers in which, probably, one BMC-H homolog was preponderant. To circumvent this issue, crystallographic studies could be performed with the monomers A, B and C composing the hetero-hexamers bearing significant differences in their periphery to be distinguishable (bearing non-canonical amino acid or residues modifiable by click chemistry or different tags…).

Overall, BMC-H interfaces were shown to be very promiscuous. This went against the interaction specificity that we aimed for and would preclude the utilization of a combination of BMC-H Duo or Trio platform concomitantly to immobilize enzymes separately. It also raised the question on whether homo-hexamers AAAAAA, BBBBBB or CCCCCC were completely impeded or a minority still formed. The question could not be answered by simply expressing individual variant and analysing hexamer formation by SEC. Indeed, the possibility for homo-hexamer to associate would not rule out that, in presence of all BMC-H from a Duo or Trio, hetero-hexamers are favoured over the homo species. Mass spectrometry could be envisioned in order to determine the ratio between each species.

In the longer term, BMC-H Duos could be used for scaffolding a 2-enzyme pathway while up to 3 different enzymes could implemented on BMC-H Trios. This could be performed through post-translational peptide covalent bonding between SpyTag/SpyCatcher for instance (Zakeri *et al*, 2012). Indeed, the SpyTag is a very small peptide (13 residues) that would not perturb enzyme folding nor activity while SpyCatcher, which is bigger in size (138 residues), could be grafted on BMC-H. This system would allow separate expression of the enzymes and scaffold which is not achievable with protein fusion and has 2 advantages: (1) the hetero-hexameric platform can be produced and formed freely from bulky enzymes and (2) the platform would be adaptable to any pathway, one would just need to change the vector coding for the enzymatic set. A diversity of equivalent spontaneous-bond-forming tags would be required to address specifically each enzyme to adequate BMC-H and ensure precise enzymatic organization on the hexamer (Keeble *et al*, 2022).

Conclusions and perspectives

Part 3. Conclusions and perspectives

This PhD thesis was the fruit of 3 years of extended study on the hexameric shell proteins of the BMCs. It has both biological interests as BMC-H cross-interactions have been thoroughly examined with BMC-H homologs from a 3-BMC-coding organism, *Kpe 342*, and biotechnological stakes as it aimed at creating a protein platform, on the basis of a hetero-hexamer, for synthetic biology.

The tripartite GFP set-up

In order to fulfil these objectives, a PPI study tool was adapted to the specific case of BMC-H. Indeed, BMC-H are proteins with extended hydrophobic patches on their monomer/monomer interfaces (intra-hexamer interfaces), thus they cannot be on their own, with cytosol-exposed interfaces, for very long before aggregation and subsequent degradation. Besides, BMC-H have the particularity to self-assemble into macrostructures such as nanotubes, sheets or Swiss-rolls (figures 18A & 19A, B, C). In that manner, the tool set-up should permit to monitor specifically intra- and not inter-hexamer assembly (*i.e*. hexamer formation and not macrostructure formation). To do so, the tGFP was selected and the assay was adapted at best to the BMC-H case. Varying the length of the linker connecting the POIs and the GFP tags along with implementing a SUMO protein domain on our model BMC-H RMM permitted to evidence that the tGFP signal was mostly arising from intra-hexamer associations (figures 27, 28 & 29). Then, the test could be further adapted to the BMC-H. By comparing different set-ups, it was determined that coding the POIs to be tested on a single vector, under the control of a unique inducible T7p (independent from the GFP1-9 ORF) was the most suited. Also, the length of the linker impacted the assay results, with longer linkers leading to increased GFP reconstitution (figure 27). Thus long linkers (Lk30/27) were picked.

The tGFP set-up was put under the test with combinations of CcmK1, CcmK2, CcmK3 and CcmK4 BMC-H homologs that were shown to form hetero-hexamers with different techniques (Garcia-Alles *et al*, 2019; Sommer *et al*, 2019). The tGFP assay corroborated previous data of our team: of all combinations, only CcmK1/CcmK2 and CcmK3/CcmK4 couples were able to cross-interact and

associate as hetero-hexamers (figure 36C). This test also evidenced that GFP tag orientation (in the N- or C-terminus of the POIs) could affect the POI interactions (figures 36A & C). Although a general preference for C-terminus tagging was observed, some BMC-H preferred a N-terminal tag such as CcmK4 or CsoS1A. No prediction of tag orientation preference could be drawn from the 3D structures. Determination of such parameter could only be made through trial and error method.

Furthermore, to know whether the tGFP assay could also be intended for the study of BMC shell components from other structural families, CcmL, CsoS4B (BMC-P) and CsoS1D (BMC-T) homo-pairs were constructed in tGFP vectors and assayed. Homo-oligomer formation was successfully monitored which depicted the possibility to use the tGFP for other components of the BMC shell (figure 36D).

During the tGFP adaptation process, I observed unexpected phenomena sharing the same cause. Firstly, positive PPI couples coded from 2 separate vectors were found negative in the tGFP assay due to an absence of the POI coded alongside the GFP1-9 (figure 31). Secondly, negative PPI couples were practically always absent from SDS-PAGE gels when all the tGFP partners were coded on the same vector (figure 32B). However, when the GFP1-9 sequence was removed from screened constructs and proteins were analysed again in SDS-PAGE, negative PPI couple proteins were clearly visible.

Globally, these data suggested that, when the GFP1-9 was present, it induced the degradation of the non-interacting partners, a phenomenon that I termed the GFP1-9 pull-down. The pull-down effect could be seen both as a positive or negative phenomenon. Indeed, on one side, it constituted a clearing system that prevented unproductive couples from accumulating in the cytosol and inducing a false positive signal due to POI random encounters. On the other side, this effect impeded direct POI expression verification from the tGFP vector. Thus, unless individual POIs were expressed from a separate vector than the GFP1-9, it was impossible to conclude on whether a lack of GFP signal was due to a negative PPI couple or to the absence of expression of one or both of the POI partners.

This pull-down effect seemed to arise from the GFP1-9 insolubility. Indeed, in the GFP1-9, the β-barrel strands and central helix constituting the fluorophore are exposed to the aqueous medium, which would destabilize the protein as the uncovered region is enriched in hydrophobic residues. Partial reconstitution with either the GFP10 or 11 appeared insufficient as it led to POI pull-down and degradation. This phenomenon could only be thwarted by the full GFP reconstitution.

Presently, our team is working on the elaboration of a decoy peptide that would mimic the GFP10/11 and stabilize the GFP1-9 until a full GFP reconstitution happens. This peptide should have a lesser affinity for the GFP1-9 than the GFP10/11 so that, in the context of a positive PPI, the GFP10/11 binding prevails over the decoy peptide. Also, it should not produce fluorescence upon binding with the GFP1-9, nor in the case of partial reconstitution with either the POI-GFP10 or POI-GFP11. Utilization

of such decoy peptide would prevent the GFP1-9 pull-down effect and allow POI expression verification.

Klebsiella pneumoniae BMC-H interactome

Thanks to the adaptation of the tGFP set-up, a BMC-H pair library assay could be envisioned. The choice of the screened BMC-H was not random. We decided to focus on the BMC-H coded by a single organism: *Klebsiella pneumonia 342*. Indeed, *Kpe 342* is among a tiny group of bacteria which possesses multiple BMC *loci*. Here, *Kpe 342* codes for 3 different BMC types and a diversity of 11 BMC-H. Three BMC-H were from the EUT1 (EutK/M/S), 4 from the GRM2 (CmcA/B/C/E) and 4 from the PDU1A (PduA/J/K/U).

tGFP screens on the library evidenced, amongst others, a high proportion of cross-interacting BMC-H, showing that hetero-hexamer formation would be a conserved trend between BMC types (figure 45). Of note, the assays were performed on BMC-H pair recombinantly expressed in *E. coli*. It remains to be demonstrated that the same associations happen in the natural host, *Kpe 342*. If hetero-hexamers were to be formed, one could ask what would be the utility of such hybrid hexamers? In order to answer this, it would require, in first instance, to decipher the precise role of each BMC-H homo-hexamer. Indeed, several BMC-H homologs exist per BMC *locus* and yet, little is known about the exact functions that plays each in the BMC shell. Are they just structural subunits? Which one acts as a channel for BMC input and output? For which substrate and/or product? Do they constitute binding domain for proteins controlling the BMC dynamics within the cell? Evolution tends to eliminate redundant proteins and/or functions unless they bring extra functions or play essential roles in the organism that could not be lost upon gene inactivation. In that manner, would the different homologs arising from the same BMC type be endowed with specific functions each?

While canonical BMC-H appear quite redundant when one looks at their global 3D structure (except for their central pore nature that could indicate different molecule specificity), I supposed that circular permutants and extended BMC-H would bring functional diversification. For instance, EutS and PduU cavity on the concave face (not the β-barrel protruding on the convex face, contrary to what was proposed in (Jorda *et al*, 2015)) could be the binding domain for PduV (figure 46B) or other cargo proteins that would then be associated to the outer shell rather than being luminal. In Huseby *et al* EUT model, the triad EutA/B/C was proposed to localize on the shell and inject AA into the shell through the pores (Huseby & Roth, 2013). It might be that they do so by docking on EutS cavity.

Besides, EutK C-terminal DNA binding domain-like extension could link the BMC to the nucleoid, allowing its positioning along the cell longitudinal axis, while PduK cysteine-rich extension could be a binding domain for a [Fe-S] cluster, able to catalyse redox reactions (figure 39).

Determining the functions of each BMC-H homo-hexamer is critical to have a better understanding of BMC biology but also to be able to engineer functional minimal BMCs that could serve as mini-reactors for the production of biomolecules. Here, the term 'functional' includes a BMC with optimal shell architecture and integrity (non-disrupted shells, with a cargo encapsulation-efficient size and adequate substrate/product permeability) and even distribution along the cell longitudinal axis (polar aggregates are not desirable for efficient catalysis and cell viability).

Then, in a second step, it would be possible to determine whether hetero-hexamers serve alternative functions in the shell. Are they granting the BMC with the ability to respond to environment changes? Are they modifying the BMC shell permeability or/and integrity for subsequent BMC final disassembly and degradation? Many possibilities exist and need to be dissected. Either way, the fact that hetero-hexamer formation would be a feature common to all BMC studied up to now, including the β-CBX, suggested that hetero-hexamers hold important functions.

The tGFP assay revealed that even BMC-H coming from different BMC types were able to cross-interact (figure 47). This was very surprising as it would mean that in multiple BMC-coding organisms, upon concomitant BMC expression, the shell subunits could mix together. Hence, hybrid hetero-hexamers would form hybrid BMCs. This raised the question of shell integrity and of BMC metabolic functioning in such structures. Would they be impaired ? Would they still be working?

Nevertheless, it should be noted that our test was performed on recombinantly expressed BMC-H pairs, without the whole natural genetic context. Thus, other molecular effectors were missing and could not fulfil potential regulation that they would in the origin organism. For instance, PocR, the PDU positive regulator was shown to inhibit the *eut* expression, preventing hybrid PDU/EUT assembly (Sturms *et al*, 2015). It might also be that the genetic organization of the BMC *loci* dictates oligomer content or that adjacent subunits (BMC-P or BMC-T or chaperons) drive non-hybrid BMC formation. More probably, if 2 BMC-H were in open competition for hetero-hexamer formation with another BMC-H, the BMC-H from the same BMC would be preferred over the BMC-H arising from a different BMC type as their translation would happen in a closer spatial and temporal proximity like they emerge from the same polycistronic mRNA.

In vivo tests have been undertaken in order to determine whether the *eut1*, *cut2* and *pdu1a* operons could be expressed concomitantly in *Kpe 342*. They are presently still undergoing but should bring some first answer elements as to the possibility of hybrid hexamers and BMC assemblies.

Elaboration of the hetero-hexameric platform thanks to computational design

Besides contributing to our knowledge on BMC-H biology and interactions, this PhD thesis aimed at the creation of a platform that would be designed from a hetero-hexamer and that could serve to immobilize diverse enzymatic pathways (to spatially organize and increase their catalysis efficiency). This part was undertaken in close collaboration with 2 teams who created a 2-AI system specialized in protein design (figure 50). This system proposed a first series of protein sequences that were well-expressed and formed homo-hexamers (figure 52), which showed that it was able to design BMC-H variants.

However, among the cases that recapitulated BMC-H behaviour, some were still able to interact with the original RMM BMC-H (figure 53A). This showed that the new intra-hexamer interfaces were promiscuous which would not allow specific monomer/monomer associations upon Mono mixing nor a precise hetero-hexamer organization. Indeed, the initial modelling process lacked negative criteria discriminating unwanted associations. Besides, it created BMC-H that could form homo-hexamers which would not be the objective in ulterior steps towards the hetero-hexameric platform elaboration, on the contrary.

Then, the modelling process was refined to include negative states in the design constraints, *i.e*. disfavouring oligomers such as homo-hexamers or hetero-hexamers not obeying the targeted organizations (ABABAB or ABCABC). In that manner, a series of Duos as well as Trios of variants were conceived. In spite of the fact that sequences cumulated up to 35 mutations compared to the wt-RMM sequence, several cases were found to associate as heteromers in both series, which confirmed the fitness of the 2-AI system for the platform development (figures 54 & 55). However, many questions remain on the nature of the oligomers formed. For instance, were they hexamers? What was the precise ratio between the different variants within the heteromer? Did they respect the constraint organization ABABAB or ABCABC? More globally, are the variants still adopting a pfam00936 fold or did their 3D structure change with incorporated mutations? Thorough 3D structure studies would be required to dispel any uncertainties and determine the variant spatial organization within the heteromer.

Another question not addressed during the course of my thesis, yet very interesting, was whether the variants (in the homo- as in the hetero-hexamer series) recapitulated natural BMC-H propensity to self-assemble as macrostructures. Indeed, those hetero-hexameric platforms could be of greater benefit for the synthetic biology field if they were prone to self-assembly. In that manner, instead of individual, diffusing platforms, one could create protein scaffolds (adopting a nanotube or sheet architecture) to further increase the catalytic efficiency of the pathway grafted on them. Besides, through concomitant expression of a minimal set of shell subunits, integration of such hetero-

hexameric platforms in a BMC shell could be envisioned for more troublesome metabolic pathways involving toxic or volatile molecules or rate-limiting enzymes.

In parallel to tackling these important questions, a proof-of-concept should be performed to demonstrate the advantages of using an hetero-hexameric platform to immobilize and organize precisely different enzymes over freely diffusing enzymes. Several pathways involved in the synthesis of biomolecules of interest could be put under the test such as pathways producing ethanol or the bioplastic precursor 3-hydroxy-propionate or the insulin secretion-inducer sitgaliptin (Sierra-Ibarra *et al*, 2022; Rathnasingh *et al*, 2012; Khobragade *et al*, 2021).

Material and methods

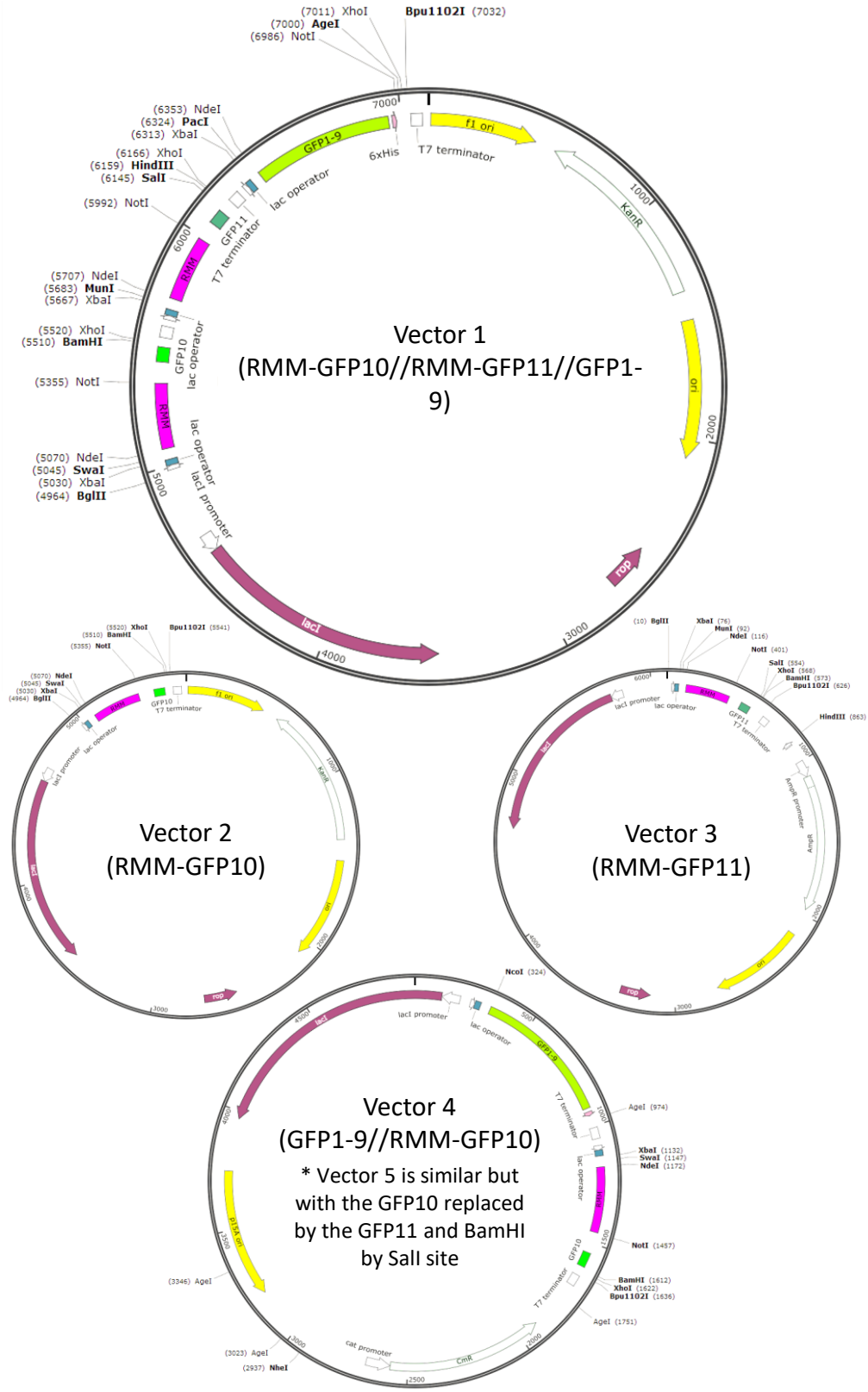


Figure 57. Vector general organization.

Part 4. Material and methods

4.1. Bacterial strains and media

Chemically competent *E. coli* TOP10 (C404010, ThermoFisher) were used as a general cloning strain and BL21(DE3) (C2527H, NEB) as an expression strain. For the construction of the BMC-H pair library in tGFP vectors, T7 express competent cells (C2566I, NEB) were used instead, as both cloning and expression strain.

Cells were cultured in lysogeny broth (LB) supplemented with due antibiotics: 40µg.mL^{-1} of kanamycin (Kan) for pET26b and pET29b utilization, 100µg.mL^{-1} of ampicillin (Amp) for pET15b and 40µg.mL^{-1} of chloramphenicol (Cm) for pACYC. For strategies combining 2 vectors, cultures were performed with either 25/25µg.mL^{-1} of Kan/Cm for pET26b/pACYC combination, 50/25µg.mL^{-1} of Amp/Cm for pET15b/pACYC, 50/25µg.mL^{-1} of Amp/Kan for pET15b/pET26b.

4.2. General procedures

Enzymatic digestion

Typically, enzymatic digestions were performed with FastDigest enzymes (ThermoFisher) according to the next proportions: 1µL of FastDigest Green buffer 10X plus 0,2µL of enzyme 1, 0,2µL of enzyme 2, 4-6µL of sample and water sq. 10µL. If the digestion included 3 different enzymes, volumes were decreased to 0,15µL each. Digestions were carried out at 37°C for 1h and followed by heat-inactivation 10min at 80°C.

Resulting open vector or fragment were purified on agarose gel (0,6% for the open vector; 1,2% for the fragment), unless otherwise stated, stained with SYBR Safe (S33102, ThermoFisher) and gel-extracted using the Monarch DNA Gel Extraction Kit (T1020L, NEB) and the manufacturer's protocol (typical elution volume was 15µL).

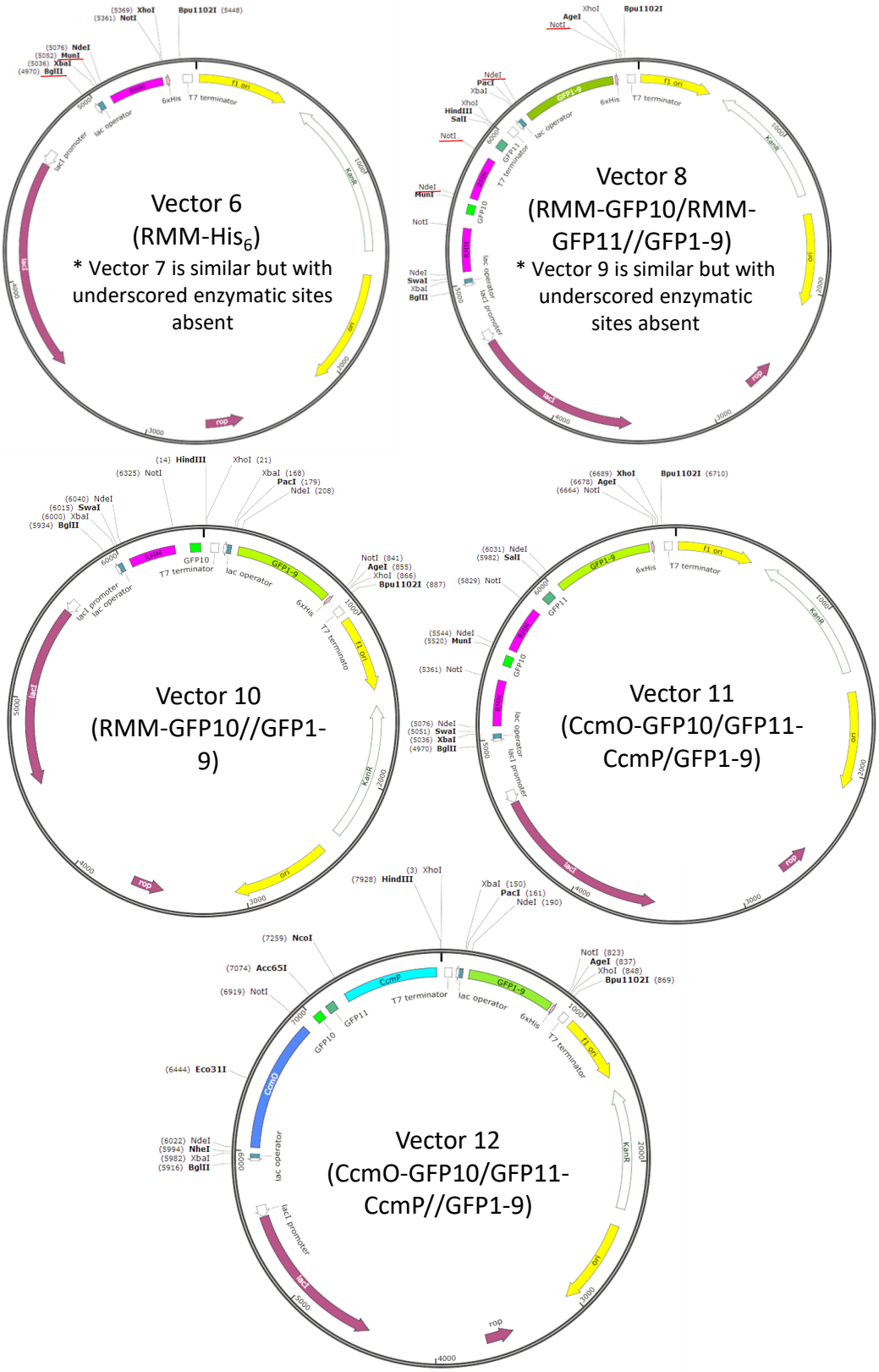


Figure 57. Vector general organization (continuation).

Ligation or circularization

Ligation was performed with: 2µL of open vector, 2µL of fragment, 0,5µL of T4 buffer 10X, 0,25µL of ATP at 10mM and 0,25µL of T4 DNA ligase (EL0014, ThermoFisher), incubated at 28°C for 1h. For simple circularization, the fragment volume was replaced by water.

Gibson assembly

Assembly was performed with the NEBuilder Hifi DNA Assembly Master Mix (E2621, NEB) with the following adapted protocol: 1,7µL of fragment 1 plus 1,7µL of fragment 2 (each at 5ng.µL^{-1}), 1µL of enzymatically opened and unpurified vector at 50ng.µL^{-1} and 4,4µL of master mix. The assembly mix was incubated at 50°C for 30min to 1h.

Competent cell preparation

Competent cells were prepared from a 100mL LB culture without antibiotics. Of note, in the 2-vector strategy, cells were transformed with a first plasmid before competence treatment (in presence of due antibiotic) and subsequent transformation with the second plasmid. When cells reached an OD_{600nm} of 0,5, they were pelleted at 6000g before medium removal and treatment with 20mL of 100mM $CaCl_2$, for 30min on ice. Cells were then centrifuged 5min at 6000g. The $CaCl_2$ solution was removed and cells were resuspended in 4mL of 100mM $CaCl_2$ plus 15% glycerol. Finally, they were aliquoted and conserved at -80°C upon utilization or used immediately for transformation.

Alternatively, competent cells could be prepared with a rubidium chloride treatment ($RbCl_2$) to reach a better transformation efficiency, notably for the T7 express used for preliminary tests (BMC-H pair library construction). Then, instead of $CaCl_2$, similar volume of the following solution was added after medium removal: $RbCl_2$ 100mM, $CaCl_2$ 10mM, $MnCl_2$ 50mM, potassium acetate 30mM plus glycerol 15% at pH 5,8. Finally, cells were conserved in 4mL of $RbCl_2$ 10mM, $CaCl_{2\ 75}$mM, MOPS 10mM plus glycerol 15% at pH 6,5.

Cell transformation

Competent cells were thawed on ice. Typical volume of cells used for transformation ranged from 10 to 50µL and DNA from 1µL (around 100ng) for purified plasmids to 2,5µL (around 10ng) for Gibson assembly or ligation reactions. After a 42°C water-bath heat shock of 30s, cells were placed back on ice and allowed to cool down for 2min. LB or super optimal medium with catabolic repressor (SOC) was added, typically 300 to 500µL, and cells were incubated at 37°C under shaking, for 45min when the plasmid bore a Kan or Cm resistance cassette or 15min if it bore a Amp^R. After incubation, transformed cells were pelleted 5min at 6000g, the medium removed. Cells were resuspended in 50µL

Table 1. General primer sequences.

Primers	Sequence	Purpose
P170	5′ CGGCGTAGAGGATCGAG	To sequence POI
P179	5′ GGTAAACAGTTCTTCGCCTTTGC	
P310	5′ CGTCCGGCGTAGAGGATCGAGATCT	To amplify Gibson fragment 1 with N-terminal GFP tag
P311	5′ CTTAAAGTATACTAGTCTGTACAGAGGAGGCTC	
P312	5′ CTTAAAGTATACTAGTCTGTACAGAGGACGCTCAT	To amplify Gibson fragment 1 with C-terminal GFP tag (with P310)
P320	5′ TACAGACTAGTATACTTTAAGAAGGAG	To amplify Gibson fragment 2
P321	5′ TTGCTCACGAGTTAACTCGAGAAAGCTTCTAGTCTG	

Table 2. POI sequences 1.

Case	Sequence (Start codon to last codon)	Origin
RMM	cataTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTG CAGATGCTATGGTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCT TAGTGGCAGTGATCGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCT GAAACTGCGTCGCAGGTTGGCGAGCTGGTTAGCGTGCATGTTATCCCACGTCCCCATTCGGAA CTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCggccgc	*M. smegmatis MC2 155*
SUMO-RMM	catATGGGCGATTCAGAAGTGAACCAGGAGGCGAAACCAGAAGTTAAGCCGGAGGTGAAGCCG GAGACCCACATCAATCTAAAAGTAAGCGACGGCTCGTCGGAGATTTTCTTTAAGATTAAGAAAA CAACCCCCCTGCGGCGTCTTATGGAGGCGTTTGCGAAGCGCCAAGGCAAGGAAATGGACTCAC TTCGTTTCCTGTACGATGGTATTCGGATTCAGGCCGACCAGACACCGGAGGATTTGGATATGGA GGATAATGATATCATCGAGGCGCATCGTGAGCAGATTGGATCCATGAGTAGTAACGCGATTGG TTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTATGGTAAAAGCTGC AAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGGG TGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCG AGCTGGTTAGCGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAG CTCAAAAGGTGCggccgc	
BWI	catATGGGATCTCTGCGTCAGTGCTCCGGTAAACAAGAATGGCCAGAGCTCGTTGGAGAGAGA GGGTCCAAGGCTGCCAAGATCATCGAAAACGAGAACGAAGACGTGCGAGCTATCGTCTTGCCT GAGGGTAGCGCGGTGCCTAGAGACCTCCGATGTGACCGTGTGTGGGTTTTCGTAGACGAGCG AGGAGTTGTTGTTGATACTCCTGTTGTTATGGGTGCggccgc	*Fagopyrum esculentum*
CcmK3	catATGGCACAAGCGGTGGGAGTGATTCAAACCTTGGGCTTTCCGAGCGTGTTAGCGGCGGCG GATGCGATGCTAAAAGGGGGCCGGGTGACGCTGGTGTATTATGACCTGGCTGAACGAGGCAA CTTTGTAGTAGCAATCCGAGGTCCCGTATCAGAGGTTAACCTTTCGATGAAGATGGGATTAGCA GCGGTAAACGAGTCCGTCATGGGAGGTGAAATCGTTAGCCATTATATTGTGCCGAACCCGCCC GAAAATGTGCTGGCGGTTCTGCCAGTGGAGTATACCGAAAAGGTTGCTCGTTTCCGGACGGGT GCggccgc	*Synechocystis PCC6803*
sm-RMM	catATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTG CAGATGCTATGGTAGATGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCT TAGTGGCAGTGATCGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCT GAAACTGCGTCGCAGGTTGGCGAGCTGGTTAGCGTGCATGTTATCCCACGTCCCCATTCGGAA CTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCggccgc	
dm-RMM	catATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTG CAGATGCTATGGTAAAAGCTGCAGATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCT TAGTGGCAGTGATCGTAACGGGTGAGGTTGGGGATGTAAAAGCTGCCACTGAAGCAGGCGCT GAAACTGCGTCGCAGGTTGGCGAGCTGGTTAGCGTGCATGTTATCCCACGTCCCCATTCGGAA CTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCggccgc	
tm-RMM	catATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTG CAGATGCTATGGTAGATGCTGCAGATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCT TAGTGGCAGTGATCGTAACGGGTGAGGTTGGGGATGTAAAAGCTGCCACTGAAGCAGGCGCT GAAACTGCGTCGCAGGTTGGCGAGCTGGTTAGCGTGCATGTTATCCCACGTCCCCATTCGGAA CTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCggccgc	

of fresh medium and spread on LB agar (liquid LB plus agar 15g.L^{-1}) petri dish with appropriate antibiotic(s).

Plasmid purification and verification

After cell transformation, that it be the TOP10 or T7 express strain, with constructed plasmids, several clones were picked per case (typically 2-3) and cultured overnight (ON) or overday in 4mL of SOC plus antibiotic(s) at 37°C. Cell pellets were then recovered and subjected to plasmid purification thanks to the EZ-10 Spin Column Plasmid DNA Min-preps Kit (BI-BS614, Euromedex) and following manufacturer's protocol. Purified plasmid were verified by enzymatic digestion (with enzyme combination depending on the construct). Finally the ones with a correct size were sequenced (at Eurofins) with primer P170 and reverse sequenced with P179 (Table 1, IDT) when the open reading frame (ORF) to cover was longer than 900 nucleotides.

4.3. Plasmid constructs

General plasmid construction methods started from an independent transcript-coding tGFP vector (vector 1), a POI-GFP10-coding pET26b (vector 2), a POI-GFP11-coding pET15b (vector 3), a GFP1-9/POI-GFP10-coding pACYC (vector 4), a GFP1-9/POI-GFP11-coding pACYC (vector 5). Full plasmid sequences are provided in the supplementary table 1 with RMM as an example (see also vector maps in figure 57).

Vectors coding for His_6-tagged POIs

His_6-tagged form of RMM and its mutants (sm-RMM, dm-RMM and tm-RMM; table 2) were already available in the lab, in a pET26b (vector 6). SUMO-RMM-His_6 (table 2), the 32 His_6-tagged Monos (supp table 2) and the 12 Hybrids (supp table 3) were ordered in a pET29b bearing a Kan^R cassette (vector 7, Twist Bioscience).

SUMO-His_6 was constructed from the SUMO-RMM-His_6 vector, digested by BamHI and recircularized.

Construction of the different tGFP vectors

RMM and its mutants, CcmK3 and BWI (table 2) were already available in the lab, in an independent transcript tGFP pET26b (vector 1), and as individual POIs in a POI-GFP10-coding pET26b (vector 2), a POI-GFP11-coding pET15b (vector 3), a GFP1-9/POI-GFP10-coding pACYC (vector 4), and a GFP1-9/POI-GFP11-coding pACYC (vector 5).

Table 3. CcmK and K1coil/E1coil pair sequences.

The CcmK homologs are from *Synechocystis PCC6803*. K1 and E1coil are *de novo* designed proteins. ORF are in bold with the GFP10 writen in light green and the GFP11 in dark green. SwaI site is in orange, NdeI in purple, NotI in blue and SalI in red. The asterisk before CcmK4 indicated the N-terminal orientation of the GFP10 tag. CcmK1/CcmK2 and *CcmK4/CcmK3 are provided as they were ordered and as examples of the possible genetic organizations depending on tag orientation. All subsequent POI sequences are only given between NdeI and NotI sites.

Case	Sequence
CcmK1/ CcmK2	atttAAATACTTTAAGAAGGAGATATACAt**ATGAGCATCGCTGTAGGTATGATCGAAACTCTGGGGTTTCCGGCTGT TGTGGAAGCAGCCGATAGCATGGTAAAAGCGGCGCGCGTGACCTTAGTGGGCTATGAAAAGATTGGCAGCGGT CGTGTCACCGTTATTGTTCGCGGGGATGTCAGCGAGGTGCAAGCGTCAGTGACGGCGGGTATCGAAAATATCCGT CGTGTAAACGGTGGAGAAGTACTGTCAAACCATATCATCGCACGCCCACATGAAAATCTGGAGTATGTTTTACCG ATTCGCTATACGGAAGCTGTGGAGCAGTTTCGTGGT**GCGGCCGC**ATCAGAAGGAGGCGGTAGCGGGGGCCCTGG TTCGGGAGGGGAAGGTTCTGCTGGGGGAGGGAGCGCTGGCGGGGGGTCT**GATTTACCAGACGATCATTACCTGA GCACACAAACGATCCTTTCGAAAGACCTGAACGCAAGCTGATAA**Ggatccttacaattgttatagaaggagatatacat**ATG AGTATCGCTGTGGGTATGATCGAAACACGCGGGTTTCCAGCGGTTGTGGAGGCGGCGGATTCAATGGTAAAAGC AGCGCGCGTTACCTTAGTGGGCTATGAAAAGATTGGCAGCGGTCGTGTAACCGTTATTGTGCGTGGGGATGTTA GCGAAGTCCAGGCAAGCGTCAGCGCCGGCATCGAGGCGGCAAATCGTGTGAATGGTGGGGAAGTACTGTCAAC GCATATCATCGCACGCCCACATGAAAATCTGGAGTATGTTTTACCGATCCGTTATACg GGTGCGGCCGCAGGTAGC GGTGGCAGTCCGGGTGGTGGCAGCGGTGGCAGCGGCAGCAGCGCAAGCGGTGGCAGCACCAGC**GAAAAACGCG ATCACATGGTGCTGCTGGAATATGTGACCGCGGCGGGCATTACCGATGCGAGCTAATGA**caagtatgtcgac
*CcmK4/ CcmK3	atttAAATACTTTAAGAAGGAGATATACAt**ATGGATTTACCAGACGATCATTACCTGAGCACACAAACGATCCTTTCG AAAGACCTGAACGGTGgGtcCGgcTCAGAAGGAGGCGGTAGCGGGGGCCCTGGTTCGGGAGGGGAAGGTTCTGC TGGGGGAGGGAGCGCTaGCGGcGGGTCCGCCCAGAGCGCCGTGGGCAGCATTGAAACCATTGGCTTTCCGGGCA TTCTTGCCGCCGCGGATGCGATGGTAAAAGCTGGTCGCATTACCATTGTGGGCTATATTCGTGCGGGCTCTGCGCG CTTTACGCTGAACATTCGTGGGGATGTGCAGGAAGTTAAAACGGCGATGGCTGCGGGCATCGATGCCATCAACC GTACAGAAGGAGCCGATGTGAAAACCTGGGTCATTATTCCGCGCCCACATGAAAATGTCGTTGCGGTTCTGCCGA TCGATTTTAGCTGATAAGgatcccttacaattgttatagaaggagatatacat**ATG**CCCCAAGCGGTGGGAGTGATTCAAACCT TGGGCTTTCCGAGCGTGTTAGCGGCGGCGGATGCGATGCTAAAAGGGGGCCGGGTGACGCTGGTGTATTATGAC CTGGCTGAACGAGGCAACTTTGTAGTAGCAATCCGAGGTCCCGTATCAGAGGTTAACCTTTCGATGAAGATGGG ATTAGCAGCGGTAAACGAGTCCGTCATGGGAGGTGAAATCGTTAGCCATTATATTGTGCCGAACCCGCCCGAAA ATGTGCTGGCGGTTCTGCCAGTGGAGTATACgGGTGCGGCCGCAGGTAGCGGTGGCAGTCCGGGTGGTGGCAGC GGTGGCAGCGGCAGCAGCGCAAGCGGTGGCAGCACCAGC**GAAAAACGCGATCACATGGTGCTGCTGGAATATG TGACCGCGGCGGGCATTACCGATGCGAGCTAATGA**caagtatgtcgac
CcmK1	CAt**ATGAGCATCGCTGTAGGTATGATCGAAACTCTGGGGTTTCCGGCTGTTGTGGAAGCAGCCGATAGCATGGTA AAAGCGGCGCGCGTGACCTTAGTGGGCTATGAAAAGATTGGCAGCGGTCGTGTCACCGTTATTGTTCGCGGGGA TGTCAGCGAGGTGCAAGCGTCAGTGACGGCGGGTATCGAAAATATCCGTCGTGTAAACGGTGGAGAAGTACTGT CAAACCATATCATCGCACGCCCACATGAAAATCTGGAGTATGTTTTACCGATTCGCTATACGGAAGCTGTGGAGC AGTTTCGTGGT**GCGGCCGC
CcmK2	cat**ATGAGTATCGCTGTGGGTATGATCGAAACACGCGGGTTTCCAGCGGTTGTGGAGGCGGCGGATTCAATGGTA AAAGCAGCGCGCGTTACCTTAGTGGGCTATGAAAAGATTGGCAGCGGTCGTGTAACCGTTATTGTGCGTGGGGA TGTTAGCGAAGTCCAGGCAAGCGTCAGCGCCGGCATCGAGGCGGCAAATCGTGTGAATGGTGGGGAAGTACTGT CAACGCATATCATCGCACGCCCACATGAAAATCTGGAGTATGTTTTACCGATCCGTTATACgGGT**GCGGCCGC
CcmK3	cat**ATGCCCCAAGCGGTGGGAGTGATTCAAACCTTGGGCTTTCCGAGCGTGTTAGCGGCGGCGGATGCGATGCTA AAAGGGGGCCGGGTGACGCTGGTGTATTATGACCTGGCTGAACGAGGCAACTTTGTAGTAGCAATCCGAGGTCC CGTATCAGAGGTTAACCTTTCGATGAAGATGGGATTAGCAGCGGTAAACGAGTCCGTCATGGGAGGTGAAATCG TTAGCCATTATATTGTGCCGAACCCGCCCGAAAATGTGCTGGCGGTTCTGCCAGTGGAGTATACgGGT**GCGGCCGC
CcmK4	cat**ATGTCCGCCCAGAGCGCCGTGGGCAGCATTGAAACCATTGGCTTTCCGGGCATTCTTGCCGCCGCGGATGCGAT GGTAAAAGCTGGTCGCATTACCATTGTGGGCTATATTCGTGCGGGCTCTGCGCGCTTTACGCTGAACATTCGTGGG GATGTGCAGGAAGTTAAAACGGCGATGGCTGCGGGCATCGATGCCATCAACCGTACAGAAGGAGCCGATGTGA AAACCTGGGTCATTATTCCGCGCCCACATGAAAATGTCGTTGCGGTTCTGCCGATCGATTTTAGCGGT**GCGGCCGC
K1 coil	CAT**ATGAGCAAAGTATCCGCTTTAAAGGAAAACGTTTCTGCTCTCAAAGAGAAGGTCAGTGCTCTGACCGAAAA AGTGTCAGCCTTGAAGGAAAAAGTATCAGCACTTAAAGAAGGT**GCGGCCGC
E1coil	CAT**ATGTCCAAAGTTTCCGCTTTAGAGAATGAAGTTTCTGCTCTCGAAAAAGAGGTCAGTGTCCTGGAAAAAGA GGTGTCAGCCTTGGAAAAGGAAGTACGTGCACTTGAGAAGGGT**GCGGCCGC

The CcmK series and the K1coil/E1coil or K1coil/K1coil were ordered as GFP-tagged POI pairs (table 3, Twist Bioscience). They were assembled in the bicistronic vector (vector 8, supp table 1) through SwaI/SalI digestion, fragment and open vector purification and ligation.

The Mono/wt-RMM tGFP vectors were created by transferring individual Mono from the His_6-tagged form vectors 7 into the vector 9 (bicistronic vector 8 (see *construction of bicistronic vector* section) on which NdeI and NotI sites are no longer repeated but only present in the POI1 ORF) through NdeI/NotI digestion and ligation.

Next POIs were ordered as 2 independent GFP10/11-alternatively-tagged fragments: Im9, Im2, E9*, Smt3, CobT, VHH, CutA, PIH1D1-N, SUMO-RMM with either a Lk30/27 or Lk1/1 linker (supp table 4), all the Duos (supp table 5, Twist Bioscience), EutM, PduA, CsoS1A, HO BMC-H, N-terminally tagged CcmK4 and RMM, CcmL, CsoS4B and CsoS1D (table 4, IDT). As for the Trios, they were ordered as 3 independent fragments (POI1-GFP10, POI2-GFP11 and POI3-Flag, supp table 6, Twist Bioscience).

These fragments incorporated 15 to 30 nucleotide-long homology regions with the Gibson assembly receptor vector (vector 9, supp table 1) and/or with adjacent fragments, depending on the fragment genetic order. Receptor vector was opened by NdeI/Acc65I/SalI digestion and assembly was performed as mentioned in the *Gibson assembly* section with unpurified vector.

Im9 and K1coil were constructed as individual GFP-tagged POIs by transferring them to both vectors 2 or 4 after a NdeI/NotI digestion, fragment and open vector purification and ligation. The same strategy was used to transfer E9* and E1coil to both vectors 3 or 5. POI-GFP10/GFP1-9-coding pET26b (vector 10) were obtained by a MunI/SalI digestion on the bicistronic vectors 8 followed by recircularization of the blunt ends generated by the Klenow fragment (same protocol as in *linker legth*).

Reduction of the linker length on the independent transcript vector

Linker length was modified by PCR with the separate vectors from the 2-vector strategy as template (vectors 2 and 3). Different primer combinations were used depending on the aimed length (table 5, IDT). Each was phosphorylated independently prior to the PCR: 2µL of buffer A 10X, 2µL of ATP at 10mM, 2µL of primer at 100µM, 13,5µL of water plus 0,5µL of T4 polynucleotide kinase (EK0031, ThermoFisher) incubated for 30min at 37°C. The PCR mix was the following: 0,6µL of template at $10ng.\mu L^{-1}$ plus 2µL of reverse/forward primer mix at 10µM, 0,5µL of dNTP at 10mM, 13µL of water, 4µL of supplied 10X buffer and 0,2µL of Phusion polymerase (F530S, ThermoFisher). The PCR program was composed of 34 cycles of 30s at 98°C then 30s at 58-62°C and 2min at 72°C. A final extension step at 72°C was performed for 10min.

Table 4. BMC-H, BMC-P and BMC-T sequences.
DNA fragments were ordered as POI1-GFP10 or POI2-GFP11. They followed CutA fragment organization (Table 6) but their sequence are provided here only between NdeI (in purple) and NotI sites (in blue).

Case	Sequence	Origin
EutM	CATATGGAGGCCCTGGGAATGATCGAAACTCGCGGGCTGGTCGCC CTCATTGAGGCCTCAGACGCGATGGTAAAAGCAGCGCGGGTGAAG CTGGTTGGCGTTAAACAAATTGGTGGTGGTCTCTGCACAGCGATGG TACGTGGAGATGTAGCCGCATGCAAGGCGGCCACCGACGCGGGG GCGGCAGCGGCACAGCGGATTGGGGAATTAGTGAGCGTGCATGTT ATTCCACGCCCTCATGGTGACCTGGAGGAAGTGTTTCCAATCGGTCT GAAGGGCGATTCCAGCAATCTGGGTGCGGCCGC	*E. coli*
PduA	CATATGCAGCAAGAAGCACTGGGAATGGTAGAAACTAAAGGGCTG ACAGCGGCCATCGAGGCAGCAGATGCTATGGTAAAGAGCGCAAAT GTTATGCTAGTGGGCTATGAAAAGATTGGCAGCGGTCTCGTGACTG TGATCGTACGTGGAGATGTAGGCGCAGTGAAGGCGGCCACCGACG CGGGGGCGGCAGCCGCACGTAATGTTGGTGAAGTAAAAGCTGTGC ATGTGATCCCTCGTCCTCATACGGATGTGGAAAAGATTCTGCCGAAG GGTATCAGCCAGGGTGCGGCCGC	*S. enterica*
CsoS1A	CATATGGCCGACGTGACCGGGATCGCACTGGGAATGATCGAGACT CGTGGACTGGTCCCGGCTATCGAGGCCGCAGATGCAATGACAAAA GCTGCGGAGGTGCGTCTGGTAGGCCGGCAGTTTGTGGGCGGTGGT TATGTCACGGTTTTAGTGCGGGGTGAGACCGGGGCGGTAAACGCA GCGGTCCGTGCTGGTGCAGATGCTTGCGAACGGGTGGGTGATGGG CTGGTAGCGGCACATATCATCGCTCGTGTCCATTCTGAAGTTGAAAA CATTCTGCCGAAGGCGCCACAGGGTGCGGCCGC	*H. neapolitanus*
HO BMC-H	CATATGGCTGACGCACTGGGAATGATCGAAGTACGTGGATTCGTCG GGATGGTAGAGGCCGCCGACGCGATGGTAAAAGCCGCGAAGGTG GAGCTGATTGGCTATGAAAAGACCGGCGGCGGCTATGTGACGGCG GTTGTGCGAGGTGACGTTGCTGCCGTAAAAGCTGCAACTGAAGCAG GCCAACGCGCGGCGGAGCGCGTTGGCGAGGTGGTGGCGGTGCAT GTGATCCCACGTCCTCATGTCAACGTTGATGCCGCGTTGCCGCTTGG CCGGACCCCCGGTATGGACAAATCAGCGGGTGCGGCCGC	*H. ochraceum*
CsoS4B	CATATGGAAGTTATGCGTGTTCGTAGCGATCTGATTGCAACCCGTCG TATTCCGGGTCTCAAAAACATTAGCCTGCGTGTTATGGAAGATGCAA CCGGCAAAGTTAGCGTTGCATGTGATCCGATTGGTGTTCCGGAAGG TTGTTGGGTTTTTACCATTAGCGGTAGCGCAGCACGTTTTGGTGTTG GTGATTTTGAAATTCTGACCGATCTGACCATTGGTGGCATTATTGATC ATTGGGTTACAGGTGCGGCCGC	*H. neapolitanus*
CcmL	CATATGCAGTTAGCGAAAGTTCTGGGAACGGTCGTTTCTACGTCAAA GACGCCTAACCTTACGGGAGTCAAGTTACTACTGGTACAGTTCCTAG ATACGAAAGGTCAGCCGCTGGAGCGTTATGAAGTCGCGGGTGATG TAGTTGGCGCGGGCTTGAACGAATGGGTCCTGGTGGCCCGCGGTA GCGCGGCGCGCAAGGAACGTGGTAACGGTGATCGCCCACTGGATG CGATGGTAGTCGGTATCATCGATACAGTGAATGTTGCAAGCGGGAG CCTTTACAATAAAAGGGACGATGGGCGGGGTGCGGCCGC	*Synechocystis PCC6803*
CsoS1D	CATATGAACAACATTGATTTGAGAGTTTACTCTTTCATTGACTCTTTGC AACCACAATTAGCCTCTTACTTGGCTACTTCTTCTCAAGGTTTCTTGCC AGTTCCAGGTGACGCTTGTTTGTGGATTGAAGTTGCTCCAGGTATGG CTGTTCACAGATTGTCTGATATTGCTTTGAAGGCTACCAACGTTCGGT TAGGTGAACAAGTTGTTGAAAGAGCTTTCGGATCTATGGAAATTCAC TACAGAAACCAATCTGACGTCTTGGCTTCTGGTGAGGCCGTTTTGAG AGAAATCAACCATGCTCAAGAAGATAGATTACCATGTAGAATCGCTT GGAAGGAGATCATCAGAGCTATTACTCCAGATCATGCCACCTTGATT AACAGACAATTAAGAAAGGGTTCCATGTTATTGCCTGGTAAATCAAT GTTCATCTTGGAGACCGAACCAGCTGGTTACATTGTTCAAGCTGCCA ACGAAGCCGAAAAAGCTGCACATGTTACTTTGATCGATGTTAGAGCC TTTGGTAACTTCGGTAGATTGACTATGATGGGTTCTGAAGCTGAAAC TGAAGAAGCTATGAGAGCTGCTGAGGCAACTATTGCCTCCATTAATG CTAGAGCAAGAAGAGCTGAAGGTTTTGGTGCGGCCGC	*H. neapolitanus*

Template plasmids were digested by DpnI (typically 0,3µL for 20µL of PCR) for 30min at 37°C. Then, the enzyme was heat-inactivated 10min at 80°C before purification of the PCR products and circularization. Modified ORFs were transferred on the independent transcript vector 1 through enzymatic digestion by SwaI/BamHI (for the ORF1) or by MunI/SalI (for the ORF2), fragment purification and ligation.

Construction of bicistronic and tricistronic vectors

Bicistronic vectors 8 were prepared from independent transcript vectors 1 by BamHI/MunI digestion to remove T7 terminator and promoter sequences between the first and second ORF. Overhangs were completed by Klenow fragment: 1µL digestion, 1µL buffer 10X, 0,5µL dNTP at 10mM, 7,3µL water and 0,2µL Klenow (EP0051, ThermoFisher) for 30min at 37°C. Plasmids were then circularized. A similar procedure was applied to prepare tricistronic vectors (vector 11) from bicistronic ones, using SalI/PacI digestion to remove T7 terminator and promoter between the second and third ORFs.

Constitutive promoter implementation on bicistronic vectors

A total of 16 constitutive promoters were selected from the iGEM part repertoire. They were ordered as forward and reverse oligonucleotides (table 6, IDT), annealed together through a temperature gradient from 95°C to 30°C (5°C steps of 30s) and finally phosphorylated with a T4 polynucleotide kinase (same protocol as in *linker length*). To replace the T7p of the GFP-tagged RMM by CPs, the bicistronic vector 8 was digested with BglII/SwaI, purified and ligated with the oligonucleotides. In a second phase, the same protocol was applied to BWI and CcmK3 bicistronic vectors using only the six selected promoters BBa_J23103, BBa_J23105, BBa_J23106, BBa_J23109, BBa_J23110 and BBa_J23115.

Construction of Klebsiella BMC-H pair library

The library was created with the assistance of the Toulouse White Biotechnology (TWB) strain engineering platform. This included robotic preparation of the different vectors by a 2-step Gibson assembly, T7 express transformation, fluorescent clone screening, plasmid purification and sequencing.

Prior to that, the GFP-tagged POIs were amplified by PCR from separate vectors 2 containing *Kpe 342* BMC-H sequences (supp table 7), with primers P310/311, P310/312 or P320/321 depending on the aimed fragment (same protocol as in *linker length* but in a volume of 50µL, an annealing temperature of 55°C and elongation step of 30s; table 1). PCR fragments were purified using the Monarch DNA Gel Extraction Kit, although they were not run on gel but directly mixed with 50µL of

Table 5. Primers used to reduce the linker length.

Purpose	Primer sequence
Lk1-GFP10	5′ GATTTACCAGACGATCATTACCTGAG
	5′ ACCTTTTGAGCTAACGCTAAAATGTG
Lk4-GFP10	5′ GATTTACCAGACGATCATTACCTGAG
	5′ TGCGGCCGCACCTTTTGA
Lk8-GFP10	5′ TCTGATTTACCAGACGATCATTACC
	5′ TCCTTCTGATGCGGCCG
Lk12-GFP10	5′ TCTGATTTACCAGACGATCATTACC
	5′ CCCGCTACCGCCTCCTTC
Lk18-GFP10	5′ TCTGATTTACCAGACGATCATTACC
	5′ CCCTCCCGAACCAGGGC
Lk24-GFP10	5′ TCTGATTTACCAGACGATCATTACC
	5′ TCCCCCAGCAGAACCTTCC
Lk1-GFP11	5′ GAAAAACGCGATCACATGGTGCT
	5′ ACCTTTTGAGCTAACGCTAAAATGTG
Lk4-GFP11	5′ GAAAAACGCGATCACATGGTGCT
	5′ TGCGGCCGCACCTTTTGA
Lk9-GFP11	5′ ACCAGCGAAAAACGCGATC
	5′ ACCGCTGCCTGCGGC
Lk13-GFP11	5′ ACCAGCGAAAAACGCGATC
	5′ GCCCGGGCTGCCACCG
Lk19-GFP11	5′ GGTAGTTCTGGCACCAGCGAAAAACGCG
	5′ ACCAGAGCTACCGCTGCCACCGCTGCCT

water and 200µL of binding buffer and loaded on the columns. Washing and elution steps followed the manufacturer's protocol. The different fragments and the receptor vector 12, opened by a BglII/Acc65I/HindIII digestion and unpurified, were provided to the platform at a concentration of 5ng.µL^{-1} and 50ng.µL^{-1}, respectively.

Next stages were performed in TWB facilities, by robotic means. Briefly, a 2-steps Gibson assembly was performed: first step with the open vector plus the fragment 2 (GFP11-tagged POI) 15min at 50°C and second step with the addition of the fragment 1 (GFP10-tagged POI) and equivalent volume of master mix, 45min at 50°C. Fragment and open vector quantities were the same as indicated in the *Gibson assembly* general procedure. T7 express cells were transformed with the Gibson product (typically 10µL of cells with 2,5µL of Gibson product). After robotic plating on LB agar with 40µg.mL^{-1} Kan, cells were stored for 2 to 3 days at 4°C in order to allow fluorescence development. Fluorescent clones were screened on a QPix 460 (Molecular Devices) and 3 clones were picked per case for subsequent culture. A portion of the culture was used to prepare glycerol stocks of each clone while the rest was subjected to plasmid purification. Sequencing was performed as indicated in the section *plasmid purification and verification*.

Globally, over 484 tGFP vectors, 21 were built by me in the preliminary tests (Gibson assembly setup and screen development) while 74 failed in robotics and required to be built manually *a posteriori*, following the same protocol. The only change in the construction protocol was that fluorescent clones were screened with a blue light transilluminator (λ_{exc} = 470nm) and an orange filter.

4.4. Tripartite GFP assay

Precultures of several clones for each case (2 or 3) were grown ON to reach saturation, at 37°C, under shaking, in 200µL of LB medium with due antibiotic(s). Next day, 2µL of the precultures were seeded in 200µL of LB with antibiotic(s), supplemented with 10µM IPTG. The culture was performed on a 96-well black plate with glass flat bottom (655892, Greiner), in the CLARIOstar Plus (BMG Labtech) which permitted cell incubation at 37°C and shaking at 300rpm while acquiring the OD at 600nm and the fluorescence (λ_{exc} = 470 ± 15nm λ_{em} = 515 ± 20nm) every 10min for 16h.

Of note, for the tGFP assay performed on the CPs with delayed GFP1-9 production, the acquisition was temporarily stopped at 4, 6 and 8h of culture to add IPTG to the medium. Fluorescence curves were then recomposed by putting the different segments of acquisition successively, each spaced by a 10min gap (time to perform the induction and resume the acquisition).

Table 6. Oligonucleotides used to reconstitute the constitutive promoters.
§ The primers bear a point mutation compared to original iGEM sequence. Mutation is indicated in red.

Promoter	DNA sequences
BBa_I14034	5' GATCTCGACATTATTGCAATTAATAAACAACTAACGGACAATTCTACCTAACAAATTT
	5' AAATTTGTTAGGTAGAATTGTCCGTTAGTTGTTTATTAATTGCAATAATGTCGA
BBa_I14018	5' GATCTCGATGTAAGTTTATACATAGGCGAGTACTCTGTTATGGAATTT
	5' AAATTCCATAACAGAGTACTCGCCTATGTATAAACTTACATCGA
BBa_J23103	5' GATCTCGACTGATAGCTAGCTCAGTCCTAGGGATTATGCTAGCAATTT
	5' AAATTGCTAGCATAATCCCTAGGACTGAGCTAGCTATCAGTCGA
BBa_J23105	5' GATCTCGATTTACGGCTAGCTCAGTCCTAGGTACTATGCTAGCAATTT
	5' AAATTGCTAGCATAGTACCTAGGACTGAGCTAGCCGTAAATCGA
BBa_J23106	5' GATCTCGATTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCAATTT
	5' AAATTGCTAGCACTATACCTAGGACTGAGCTAGCCGTAAATCGA
BBa_J23109	5' GATCTCGATTTACAGCTAGCTCAGTCCTAGGGACTGTGCTAGCAATTT
	5' AAATTGCTAGCACAGTCCCTAGGACTGAGCTAGCTGTAAATCGA
BBa_J23110	5' GATCTCGATTTACGGCTAGCTCAGTCCTAGGTACAATGCTAGCAATTT
	5' AAATTGCTAGCATTGTACCTAGGACTGAGCTAGCCGTAAATCGA
BBa_J23113	5' GATCTCGACTGATGGCTAGCTCAGTCCTAGGGATTATGCTAGCAATTT
	5' AAATTGCTAGCATAATCCCTAGGACTGAGCTAGCCATCAGTCGA
BBa_J23114	5' GATCTCGATTTATGGCTAGCTCAGTCCTAGGTACAATGCTAGCAATTT
	5' AAATTGCTAGCATTGTACCTAGGACTGAGCTAGCCATAAATCGA
BBa_J23115§	5' GATCTCGATTTATAGCTAGCTCAGTCCTTGGTACAATGCTAGCAATTT
	5' AAATTGCTAGCATTGTACCAAGGACTGAGCTAGCTATAAATCGA
BBa_J23116	5' GATCTCGATTGACAGCTAGCTCAGTCCTAGGGACTATGCTAGCAATTT
	5' AAATTGCTAGCATAGTCCCTAGGACTGAGCTAGCTGTCAATCGA
BBa_J23117	5' GATCTCGATTGACAGCTAGCTCAGTCCTAGGGATTGTGCTAGCAATTT
	5' AAATTGCTAGCACAATCCCTAGGACTGAGCTAGCTGTCAATCGA
BBa_J48104	5' GATCTCGATAATCAGTATGACGAATACTTAAAATCGTCATACTTATTTAATTT
	5' AAATTAAATAAGTATGACGATTTTAAGTATTCGTCATACTGATTATCGA
BBa_K137029	5' GATCTCGATTTAATTATATATATATATATATATAATGGAAGCGTTTTAATTT
	5' AAATTAAAACGCTTCCATTATATATATATATATATATAATTAAATCGA
BBa_K137085	5' GATCTCGATTGACAATATATATATATATATAATGCTAGCAATTT
	5' AAATTGCTAGCATTATATATATATATATATTGTCAATCGA
BBa_S03331§	5' GATCTCGATTGACAAGCATTTCCTCAGCTCCGTAAACTAATTT
	5' AAATTAGTTTACGGAGCTGAGGAAATGCTTGTCAATCGA

Data were then processed by GraphPad Prism 6: fluorescence and occasionally growth curves obtained were fitted to a sigmoidal function of equation:

$$y = F_{basal} + \frac{F_{max} - F_{basal}}{1 + 10^{\log(\text{half } F_{max} - x) \times \text{slope factor}}}$$

The F_{basal} is the fluorescence signal at time 0 and the Fmax is the value when fluorescence signal get to a plateau. For incorrect automatic fits (generally for low signal cases), Fmax values were manually reported. Fmax values were normalized by the model RMM homo-pair measured in the same assay (the exact RMM case is stated in each graph). In that manner, fluorescence values that varied between all experiments could be compared. Values reported in the tGFP graphs are the mean values ± standard deviations obtained after a minimum of 2 independent experiments, each one including 2 or 3 clones.

4.5. Analysis of protein expression and solubility

After transformation of chemically-competent BL21(DE3) or T7 express cells with the different constructs cloned in pET15b, pET26b, pET29b or pACYC vectors, single clones were cultured ON, until saturation, in LB supplemented with appropriate antibiotic(s), depending on the plasmid(s). Then, precultures were seeded in the same medium with a 100-fold dilution (2 to 10mL, depending on the experiment). Cells were induced with 10µM from the beginning and cultures were incubated 16h at 37°C and a 200rpm shaking, unless otherwise indicated.

For a simple SDS-PAGE analysis, cells were processed as follow. However, if proteins were to be purified in the last step, please refer to the *protein purification* section for cell processing. Cells were recovered by a 5min centrifugation at 6000g and 4°C. After discarding the supernatant, the pellet was lysed using Bugbuster Extraction Reagent (70923, Merck) supplemented with 25µg.mL^{-1} of lysozyme (L-6876, Sigma), 1mM of EDTA, 268U of Benzonase Nuclease (70746-4, Merck) and 1mM of protease-inhibitor phenylmethylsulfonyl fluoride (in isopropanol; P-7626, Sigma). Typically, 100µL of lysis solution was added for a 2mL-culture pellet. After incubation at room temperature (RT) for 10min, samples were placed on ice. Aliquots (40 □L) of total protein fraction were withdrawn, mixed with an equal volume of 2X loading dye before denaturation at 95°C for 10min. The remaining volume was centrifuged at 16000g for 10min, at 4°C. Supernatant aliquots were collected to prepare the soluble fractions in the same manner.

After a 10min denaturation at 95°C, samples (1 to 3µL) were analysed in SDS-PAGE gels of 15 or 18% concentration and a 1:29 reticulation. Gel staining was either performed ON with Coomassie

Brilliant Blue R-250 (161-0436, Bio-Rad) or for 30min with Instant Blue Coomassie Protein Stain (ab119211, Abcam), both under shaking and at RT.

4.6. Protein purification by affinity chromatography

Cells were recovered as described in the previous section, resuspended in Solution A (300mM NaCl, 10mM sodium phosphate, 10mM imidazole of pH 8,2) supplemented with 25ng.mL^{-1} of lysozyme after medium removal and placed on ice. Typically, 1mL of Solution A was added for a 10mL-culture pellet. Three to four 30s sonication cycles at 40% amplitude (SO-VCX130 sonicator equipped with a 630-0422 probe, Sonics) were applied on each sample, spaced by at least 1min on ice. Total and soluble protein fraction aliquots were prepared as described above. Remaining soluble fractions (500µL) were loaded on Co^{2+} affinity chromatography columns (TALON) in a 96-well plate format (VS-HT08CC02, VivaScience), previously equilibrated 3 times with 4°C pre-cooled Solution A (typically 400µL for each wash). The plate was centrifuged for 5min at 1500g and 4°C. The flowthroughs were discarded and columns were washed 3 times with cold Solution A. Finally, elution was performed with cold Solution A containing a total of 300mM of imidazole, of pH 7,8 (typically 300µL). EDTA was added to the purified fractions to a final concentration of 1mM, immediately after elution. Purified proteins aliquots of 40µL were collected, mixed with 2X loading dye and denature as before, 10min at 95°C.

4.7. Size-exclusion high-pressure liquid chromatography

Previously purified proteins (100µL) were dialysed ON at 4°C in Pur-A-Lyzer dialysis columns that had a 7kDa cut-off (69562, ThermoFisher), against Buffer B composed of 10mM Tris, 200mM NaCl and 1mM EDTA, of pH 8 (typically 300-fold volume exchange).

Dialyzed samples were then loaded on a size-exclusion high pressure liquid chromatography column (SEC2000, Beckman). Classical volumes loaded were 40µL. Migration was carried out with Buffer B, at a flow rate of 1mL.min^{-1} and retention times were monitored at 280nm. Several standards were injected separately: Rnase A (13,7kDa), conalbumin (75kDa), ferritin (440kDa) and dextran (2MDa). Through plotting the log of their molecular weight in function of their retention time, a linear curve of equation $y = ax + b$ was obtained and allowed to calculate an estimation of the molecular weight of eluted proteins from their retention time.

4.8. Transmission electron microscopy

After ON IPTG induction as indicated for expression experiments, 8mL cell cultures were pre-fixed with an equivalent volume of fixative solution (5% glutaraldehyde and 4% paraformaldehyde in 100mM cacodylate buffer of pH 7,2). After 15min at RT, the cells were pelleted by a 5min centrifugation at 6000g and resuspended in 2,5% glutaraldehyde and 2% PFA in the same cacodylate buffer (typically 1mL). The cells were incubated for 1h45 at RT and subsequently washed 3 times with cacodylate buffer. A post-fixation was performed with 1% osmium tetroxide in the cacodylate buffer for 1h at RT. The cells were washed again 3 times with the cacodylate buffer and inlayed in 2% low-melting point agarose before uranyl acetate 1% treatment for 1h at RT. Samples were dehydrated using an ethanol gradient: incubation in ethanol 25, 50, 70 and 90% for 15min, plus 3 additional 30min steps in ethanol 100%. They were then transferred in Epon resin baths (Embed 812, EMS) of increasing concentration (25, 50, 75% Epon in ethanol for 1h at RT and twice 2h in 100% Epon at 37°C). Finally, they were embedded in Epon resin by a 48h polymerization at 60°C.

Sections of 80nm of thickness were prepared with the Ultramicrotome UCT (Leica), mounted onto formvar/carbon-coated copper grids of 200-mesh and stained with Uranyless (EM-grade.com) and Reynolds lead citrate 3% (EM-grade.com). TEM acquisitions were made using a JEOL JEM-1400 at a 80kV voltage and a digital camera Gatan Orius.

4.9. Sequence alignments

Protein sequences were uploaded and aligned thanks to Clustal Omega website (https://www.ebi.ac.uk/Tools/msa/clustalo/). Clustal Omega also provided a protein phylogenetic tree for the same alignment.

4.10. AlphaFold2 structural predictions

Query sequences were fed in ColabFold v1.5.2-patch: AlphaFold2 using MMseqs2, following the instructions provided online (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb), without structural template. Five best ranked models were recovered. Structural images and analysis were performed on Pymol.

Others

Abbreviations

1,2-PD: 1,2-propanediol
3-PGA: 3-phosphoglycerate
AA: acetaldehyde
AAU: aminoacetone utilization BMC
AF2: Alphafold2
AI: artificial intelligence
Amp: ampicillin
avGFP: *Aequorea victoria* GFP
BMC: bacterial microcompartment
BMC-H: bacterial microcompartment hexamer-forming protomer
BMC-T: bacterial microcompartment trimer-forming protomer
BMC-P: bacterial microcompartment pentamer-forming protomer
BWI: buckwheat trypsin inhibitor
CA: carbonic anhydrase
CBX: carboxysome
C/C: C-terminal GFP10 and GFP11
CCM: carbon concentrating mechanism
Cm: chloramphenicol
CoA: coenzyme A
CP: constitutive promoter
cryoEM: cryo-electron microscopy
cut: choline utilization operon
E9: colicin endonuclease 9
EA: ethanolamine
Effie: energy function familiarly introduced as Effie
eGFP: enhanced green fluorescent protein
EP: encapsulation peptide
EUT: ethanolamine utilization BMC
F_{max}: maximal fluorescence
FRET: Förster resonance energy transfer
frGFP: folding-reporter GFP
GFP1-10 OPT: optimized GFP1-10
GRE: glycyl radical enzyme
GRM: glycyl radical enzyme-associated BMC
HS-AFM: high-speed atomic force microscopy
Im9: immunity protein 9
Kan: kanamycin
Kpe: *Klebsiella pneumoniae*
LB: lysogeny broth
Lk: linker
LLPS: liquid-liquid phase separation
McdA: maintenance of carboxysome distribution A
mRNA: messenger ribonucleic acid
MW: molecular weight
OD: optical density
ON: overnight
opv: open receptor vector
ORF: open reading frame
ORI: origin of replication
PCA: protein complementation assay
PCR: polymerase chain reaction
PDB: protein data bank
PDU: propanediol utilization BMC
POI: protein of interest
PPI: protein-protein interactions
PVM: *Planctomycete* and *Verrucomicrobia* BMC
RMM: *Rhodococcus* and *Mycobacterium* Microcompartment BMC-H
RMSD: root mean square deviation
RT: room temperature
RuBisCO: ribulose-1,5-bisphosphate carboxylase/oxygenase enzyme
RuBP: ribulose biphosphate
SEC: size-exclusion chromatography
sfGFP: superfolder GFP
SOC: super optimal medium with catabolic repressor
SUMO: small ubiquitin-related modifier
T7p: T7 promoter
TALE: transcription activator-like effector protein
TCA: tricarboxylic acid cycle
TEM: transmission electron microscopy
tGFP: tripartite GFP
TMA: trimethylamine
Toulbar2: Toulouse Barcelona solver 2
TWB: Toulouse White Biotechnology
Y2H: yeast two-hybrid
wt-RMM: wild-type RMM
ZFP: zinc finger protein

References

Aalbers FS & Fraaije MW (2017) Coupled reactions by coupled enzymes: alcohol to lactone cascade with alcohol dehydrogenase–cyclohexanone monooxygenase fusions. *Appl Microbiol Biotechnol* 101: 7557–7565

Allouche D, de Givry S, Katsirelos G, Schiex T & Zytnicki M (2015) Anytime Hybrid Best-First Search with Tree Decomposition for Weighted CSP. In *Principles and Practice of Constraint Programming* pp 12–29. Springer, Cham

Araiza-Olivera D, Chiquete-Felix N, Rosas-Lemus M, Sampedro JG, Peña A, Mujica A & Uribe-Carvajal S (2013) A glycolytic metabolon in *S accharomyces cerevisiae* is stabilized by F -actin. *The FEBS Journal* 280: 3887–3905

Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, Saito R, Ara T, Nakahigashi K, Huang H-C, Hirai A, *et al* (2006) Large-scale identification of protein–protein interaction of Escherichia coli K-12. *Genome Res* 16: 686–691

Asija K, Sutter M & Kerfeld CA (2021) A Survey of Bacterial Microcompartment Distribution in the Human Microbiome. *Front Microbiol* 12: 669024

Aussignargues C, Paasch BC, Gonzalez-Esquer R, Erbilgin O & Kerfeld CA (2015) Bacterial microcompartment assembly: The key role of encapsulation peptides. *Communicative & Integrative Biology* 8: e1039755

Axen SD, Erbilgin O & Kerfeld CA (2014) A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. *PLoS Comput Biol* 10: e1003898

Azaldegui CA, Vecchiarelli AG & Biteen JS (2021) The emergence of phase separation as an organizing principle in bacteria. *Biophysical Journal* 120: 1123–1138

Badger MR (2003) CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *Journal of Experimental Botany* 54: 609–622

Bagley ST (1985) Habitat Association of Klebsiella Species. *Infection Control & Hospital Epidemiology* 6: 52–58

Behrenfeld MJ, Randerson JT, McClain CR, Feldman GC, Los SO, Tucker CJ, Falkowski PG, Field CB, Frouin R, Esaias WE, *et al* (2001) Biospheric Primary Production During an ENSO Transition. *Science* 291: 2594–2597

Bertolini M, Fenzl K, Kats I, Wruck F, Tippmann F, Schmitt J, Auburger JJ, Tans S, Bukau B & Kramer G (2021) Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. *Science* 371: 57–64

Blaszczak E, Lazarewicz N, Sudevan A, Wysocki R & Rabut G (2021) Protein-fragment complementation assays for large-scale analysis of protein–protein interactions. *Biochemical Society Transactions* 49: 1337–1348

Bobik TA, Ailion M & Roth JR (1992) A single regulatory gene integrates control of vitamin B12 synthesis and propanediol degradation. *J Bacteriol* 174: 2253–2266

Bobik TA, Havemann GD, Busch RJ, Williams DS & Aldrich HC (1999) The Propanediol Utilization (*pdu*) Operon of *Salmonella enterica* Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B $_{12}$ -Dependent 1,2-Propanediol Degradation. *J Bacteriol* 181: 5967–5975

Bobik TA, Lehman BP & Yeates TO (2015) Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways: Bacterial microcompartments. *Molecular Microbiology* 98: 193–207

Böhm J, Thavaraja R, Giehler S & Nalaskowski MM (2017) A set of enhanced green fluorescent protein concatemers for quantitative determination of nuclear localization signal strength. *Analytical Biochemistry* 533: 48–55

Bonacci W, Teng PK, Afonso B, Niederholtmeyer H, Grob P, Silver PA & Savage DF (2012) Modularity of a carbon-fixing protein organelle. *Proceedings of the National Academy of Sciences* 109: 478–483

Bouin A, Zhang C, Lindley ND, Truan G & Lautier T (2023) Exploring linker's sequence diversity to fuse carotene cyclase and hydroxylase for zeaxanthin biosynthesis. *Metabolic Engineering Communications* 16: e00222

Bradley-Clarke J, Gu S, Rose R-S, Warren MJ & Pickersgill RW (2022) Enzyme encapsulation peptides bind to the groove between tessellating subunits of the bacterial microcompartment shell Biochemistry

Cabantous S, Nguyen HB, Pedelacq J-D, Koraïchi F, Chaudhary A, Ganguly K, Lockard MA, Favre G, Terwilliger TC & Waldo GS (2013) A New Protein-Protein Interaction Sensor Based on Tripartite Split-GFP Association. *Sci Rep* 3: 2854

Cabantous S, Terwilliger TC & Waldo GS (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* 23: 102–107

Cai F, Dou Z, Bernstein S, Leverenz R, Williams E, Heinhorst S, Shively J, Cannon G & Kerfeld C (2015a) Advances in Understanding Carboxysome Assembly in Prochlorococcus and Synechococcus Implicate CsoS2 as a Critical Component. *Life* 5: 1141–1171

Cai F, Menon BB, Cannon GC, Curry KJ, Shively JM & Heinhorst S (2009) The Pentameric Vertex Proteins Are Necessary for the Icosahedral Carboxysome Shell to Function as a CO2 Leakage Barrier. *PLoS ONE* 4: e7521

Cai F, Sutter M, Bernstein SL, Kinney JN & Kerfeld CA (2015b) Engineering Bacterial Microcompartment Shells: Chimeric Shell Proteins and Chimeric Carboxysome Shells. *ACS Synth Biol* 4: 444–453

Cai F, Sutter M, Cameron JC, Stanley DN, Kinney JN & Kerfeld CA (2013) The Structure of CcmP, a Tandem Bacterial Microcompartment Domain Protein from the β-Carboxysome, Forms a Subcompartment Within a Microcompartment. *Journal of Biological Chemistry* 288: 16055–16063

Cameron JC, Wilson SC, Bernstein SL & Kerfeld CA (2013) Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway. *Cell* 155: 1131–1140

Castellana M, Wilson MZ, Xu Y, Joshi P, Cristea IM, Rabinowitz JD, Gitai Z & Wingreen NS (2014) Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat Biotechnol* 32: 1011–1018

Cesle EE, Filimonenko A, Tars K & Kalnins G (2021) Variety of size and form of GRM2 bacterial microcompartment particles. *Protein Science* 30: 1035–1043

Chaize B, Colletier J-P, Winterhalter M & Fournier D (2004) Encapsulation of Enzymes in Liposomes: High Encapsulation Efficiency and Control of Substrate Permeability. *Artificial Cells, Blood Substitutes, and Biotechnology* 32: 67–75

Chang G & Chang J (1975) Evidence for the B12-dependent enzyme ethanolamine deaminase in Salmonella. *Nature* 254

Chelius MK & Triplett EW (2000) Immunolocalization of Dinitrogenase Reductase Produced by *Klebsiella pneumoniae* in Association with *Zea mays* L. *Appl Environ Microbiol* 66: 783–787

Chen L & Hatti-Kaul R (2017) Exploring Lactobacillus reuteri DSM20016 as a biocatalyst for transformation of longer chain 1,2-diols: Limits with microcompartment. *PLoS ONE* 12: e0185734

Choi SG, Olivet J, Cassonnet P, Vidalain P-O, Luck K, Lambourne L, Spirohn K, Lemmens I, Dos Santos M, Demeret C, *et al* (2019) Maximizing binary interactome mapping with a minimal number of assays. *Nat Commun* 10: 3907

Choudhary S, Quin MB, Sanders MA, Johnson ET & Schmidt-Dannert C (2012) Engineered Protein Nano-Compartments for Targeted Enzyme Localization. *PLoS ONE* 7: e33342

Chowdhury C, Chun S, Pang A, Sawaya MR, Sinha S, Yeates TO & Bobik TA (2015) Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc Natl Acad Sci USA* 112: 2990–2995

Chowdhury C, Chun S, Sawaya MR, Yeates TO & Bobik TA (2016) The function of the PduJ microcompartment shell protein is determined by the genomic position of its encoding gene. *Molecular Microbiology* 101: 770–783

Chowdhury C, Sinha S, Chun S, Yeates TO & Bobik TA (2014) Diverse Bacterial Microcompartment Organelles. *Microbiol Mol Biol Rev* 78: 438–468

Chrétien A-È, Gagnon-Arsenault I, Dubé AK, Barbeau X, Després PC, Lamothe C, Dion-Côté A-M, Lagüe P & Landry CR (2018) Extended Linkers Improve the Detection of Protein-protein Interactions (PPIs) by Dihydrofolate Reductase Protein-fragment Complementation Assay (DHFR PCA) in Living Cells. *Molecular & Cellular Proteomics : MCP* 17: 373

Conrado RJ, Wu GC, Boock JT, Xu H, Chen SY, Lebar T, Turnšek J, Tomšič N, Avbelj M, Gaber R, *et al* (2012) DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Research* 40: 1879–1889

Cormack BP, Valdivia RH & Falkow S (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 173: 33–38

Craciun S & Balskus EP (2012) Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA* 109: 21307–21312

Crowley CS, Cascio D, Sawaya MR, Kopstein JS, Bobik TA & Yeates TO (2010) Structural Insight into the Mechanisms of Transport across the Salmonella enterica Pdu Microcompartment Shell. *The Journal of Biological Chemistry* 285: 37838

Crowley CS, Sawaya MR, Bobik TA & Yeates TO (2008) Structure of the PduU Shell Protein from the Pdu Microcompartment of Salmonella. *Structure* 16: 1324–1332

Cui Y, Gao C, Zhao Q & Jiang L (2016) Using Fluorescent Protein Fusions to Study Protein Subcellular Localization and Dynamics in Plant Cells. In *High-Resolution Imaging of Cellular Proteins*, Schwartzbach SD Skalli O & Schikorski T (eds) pp 113–123. New York, NY: Springer New York

Dai W, Chen M, Myers C, Ludtke SJ, Pettitt BM, King JA, Schmid MF & Chiu W (2018) Visualizing Individual RuBisCO and Its Assembly into Carboxysomes in Marine Cyanobacteria by Cryo-Electron Tomography. *Journal of Molecular Biology* 430: 4156–4167

Dank A, Zeng Z, Boeren S, Notebaart RA, Smid EJ & Abee T (2021) Bacterial Microcompartment-Dependent 1,2-Propanediol Utilization of Propionibacterium freudenreichii. *Front Microbiol* 12: 679827

Das S, Zhao L, Elofson K & Finn MG (2020) Enzyme Stabilization by Virus-Like Particles. *Biochemistry* 59: 2870–2881

Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF, Wicky BIM, Courbet A, Haas RJ de, Bethel N, *et al* (2022) Robust deep learning–based protein sequence design using ProteinMPNN. *Science*

Defresne M, Barbe S & Schiex T (2023) Scalable Coupling of Deep Learning with Logical Reasoning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence* pp 3615–3623. Macau, SAR China: International Joint Conferences on Artificial Intelligence Organization

Delebecque CJ, Lindner AB, Silver PA & Aldaye FA (2011) Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science* 333: 470–474

Delmas J, Gibold L, Faïs T, Batista S, Leremboure M, Sinel C, Vazeille E, Cattoir V, Buisson A, Barnich N, *et al* (2019) Metabolic adaptation of adherent-invasive Escherichia coli to exposure to bile salts. *Sci Rep* 9: 2175

Drews G & Niklowitz W (1956) Cytology of Cyanophycea. Centroplasm and granular inclusions of Phormidium uncinatum. *Archiv fur Mikrobiologie* 24

Faber F, Thiennimitr P, Spiga L, Byndloss MX, Litvak Y, Lawhon S, Andrews-Polymenis HL, Winter SE & Bäumler AJ (2017) Respiration of Microbiota-Derived 1,2-propanediol Drives Salmonella Expansion during Colitis. *PLOS Pathogens* 13: e1006129

Fan C & Bobik TA (2011) The N-Terminal Region of the Medium Subunit (PduD) Packages Adenosylcobalamin-Dependent Diol Dehydratase (PduCDE) into the Pdu Microcompartment. *Journal of Bacteriology* 193: 5623–5628

Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, Jefferson RE, Yeates TO & Bobik TA (2010) Short N-terminal sequences package proteins into bacterial microcompartments. *Proc Natl Acad Sci USA* 107: 7509–7514

Fan C, Cheng S, Sinha S & Bobik TA (2012a) Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc Natl Acad Sci USA* 109: 14995–15000

Fan C, Cheng S, Sinha S & Bobik TA (2012b) Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc Natl Acad Sci USA* 109: 14995–15000

Farewell A & Neidhardt FC (1998) Effect of Temperature on In Vivo Protein Synthetic Capacity in Escherichia coli. *Journal of Bacteriology*

Faulkner M, Zhao L-S, Barrett S & Liu L-N (2019) Self-Assembly Stability and Variability of Bacterial Microcompartment Shell Proteins in Response to the Environmental Change. *Nanoscale Res Lett* 14: 54

Ferlez B, Sutter M & Kerfeld CA (2019a) Glycyl Radical Enzyme-Associated Microcompartments: Redox-Replete Bacterial Organelles. *mBio* 10: e02327-18

Ferlez B, Sutter M & Kerfeld CA (2019b) A designed bacterial microcompartment shell with tunable composition and precision cargo loading. *Metabolic Engineering* 54: 286–291

Fields S & Song O (1989) A novel genetic system to detect protein–protein interactions. *Nature* 340: 245–246

Forouhar F, Kuzin A, Seetharaman J, Lee I, Zhou W, Abashidze M, Chen Y, Yong W, Janjua H, Fang Y, *et al* (2007) Functional insights from structural genomics. *J Struct Funct Genomics* 8: 37–44

Fouts DE, Tyler HL, DeBoy RT, Daugherty S, Ren Q, Badger JH, Durkin AS, Huot H, Shrivastava S, Kothari S, *et al* (2008) Complete Genome Sequence of the N2-Fixing Broad Host Range Endophyte Klebsiella pneumoniae 342 and Virulence Predictions Verified in Mice. *PLoS Genet* 4: e1000141

Fox KA, Ramesh A, Stearns JE, Bourgogne A, Reyes-Jara A, Winkler WC & Garsin DA (2009) Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in *Enterococcus faecalis*. *Proc Natl Acad Sci USA* 106: 4435–4440

Francis DM & Page R (2010) Strategies to Optimize Protein Expression in E. coli. *Current Protocols in Protein Science* 61: 5241

Friedman H, Lu P & Rich A (1971) Temperature control of initiation of protein synthesis in Escherichia coli. *Journal of Molecular Biology* 61: 105–121

Garcia-Alles LF, Fuentes-Cabrera M, Truan G & Reguera D (2023) Inferring assembly-curving trends of bacterial micro-compartment shell hexamers from crystal structure arrangements. *PLoS Comput Biol* 19: e1011038

Garcia-Alles LF, Lesniewska E, Root K, Aubry N, Pocholle N, Mendoza CI, Bourillot E, Barylyuk K, Pompon D, Zenobi R, *et al* (2017) Spontaneous non-canonical assembly of CcmK hexameric components from β-carboxysome shells of cyanobacteria. *PLoS ONE* 12: e0185109

Garcia-Alles LF, Root K, Maveyraud L, Aubry N, Lesniewska E, Mourey L, Zenobi R & Truan G (2019) Occurrence and stability of hetero-hexamer associations formed by β-carboxysome CcmK shell components. *PLoS ONE* 14: e0223877

Garinot-Schneider C, Pommer AJ, Moore GR, Kleanthous C & James R (1996) Identification of Putative Active-site Residues in the DNase Domain of Colicin E9 by Random Mutagenesis. *Journal of Molecular Biology* 260: 731–742

Gerngross D, Beerenwinkel N & Panke S (2022) Systematic investigation of synthetic operon designs enables prediction and control of expression levels of multiple proteins Synthetic Biology

Greber BJ, Sutter M & Kerfeld CA (2019) The Plasticity of Molecular Interactions Governs Bacterial Microcompartment Shell Assembly. *Structure* 27: 749-763.e4

Greene SE & Komeili A (2012) Biogenesis and subcellular organization of the magnetosome organelles of magnetotactic bacteria. *Current Opinion in Cell Biology* 24: 490–495

Guo H, Ryan JC, Song X, Mallet A, Zhang M, Pabst V, Decrulle AL, Ejsmont P, Wintermute EH & Lindner AB (2022) Spatial engineering of E. coli with addressable phase-separated RNAs. *Cell* 185: 3823-3837.e23

Hagen A, Sutter M, Sloan N & Kerfeld CA (2018a) Programmed loading and rapid purification of engineered bacterial microcompartment shells. *Nat Commun* 9: 2881

Hagen AR, Plegaria JS, Sloan N, Ferlez B, Aussignargues C, Burton R & Kerfeld CA (2018b) In vitro assembly of diverse bacterial microcompartment shell architectures. *Nano Lett* 18: 7030–7037

Havemann GD, Sampson EM & Bobik TA (2002) PduA Is a Shell Protein of Polyhedral Organelles Involved in Coenzyme B $_{12}$ -Dependent Degradation of 1,2-Propanediol in *Salmonella enterica* Serovar Typhimurium LT2. *J Bacteriol* 184: 1253–1261

Heim R, Cubitt AB & Tsien RY (1995) Improved green fluorescence. *Nature* 373: 663–664

Herring TI, Harris TN, Chowdhury C, Mohanty SK & Bobik TA (2018) A Bacterial Microcompartment Is Used for Choline Fermentation by Escherichia coli 536. *J Bacteriol* 200

Hill NC, Tay JW, Altus S, Bortz DM & Cameron JC (2020) Life cycle of a cyanobacterial carboxysome. *Sci Adv* 6: eaba1269

Ho TQ (2002) Compatible bacterial plasmids are targeted to independent cellular locations in Escherichia coli. *The EMBO Journal* 21: 1864–1872

Huang J, Ferlez BH, Young EJ, Kerfeld CA, Kramer DM & Ducat DC (2020) Functionalization of Bacterial Microcompartment Shell Proteins With Covalently Attached Heme. *Front Bioeng Biotechnol* 7: 432

Huseby DL & Roth JR (2013) Evidence that a Metabolic Microcompartment Contains and Recycles Private Cofactor Pools. *J Bacteriol* 195: 2864–2879

Iancu CV, Ding HJ, Morris DM, Dias DP, Gonzales AD, Martino A & Jensen GJ (2007) The Structure of Isolated Synechococcus Strain WH8102 Carboxysomes as Revealed by Electron Cryotomography. *Journal of molecular biology* 372: 764

Iancu CV, Morris DM, Dou Z, Heinhorst S, Cannon GC & Jensen GJ (2010) Organization, Structure, and Assembly of α-Carboxysomes Determined by Electron Cryotomography of Intact Cells. *Journal of Molecular Biology* 396: 105–117

Iniguez AL, Dong Y & Triplett EW (2004) Nitrogen Fixation in Wheat Provided by *Klebsiella pneumoniae* 342. *MPMI* 17: 1078–1085

Jakobson CM, Kim EY, Slininger MF, Chien A & Tullman-Ercek D (2015) Localization of Proteins to the 1,2-Propanediol Utilization Microcompartment by Non-native Signal Sequences Is Mediated by a Common Hydrophobic Motif *. *Journal of Biological Chemistry* 290: 24519–24533

Jakobson CM, Tullman-Ercek D, Slininger MF & Mangan NM (2017) A systems-level model reveals that 1,2-propanediol utilization microcompartments enhance pathway flux through intermediate sequestration. *PLOS Computational Biology* 13: e1005525

Jorda J, Leibly DJ, Thompson MC & Yeates TO (2016) Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein. *Chem Commun* 52: 5041–5044

Jorda J, Liu Y, Bobik TA & Yeates TO (2015) Exploring Bacterial Organelle Interactomes: A Model of the Protein-Protein Interaction Network in the Pdu Microcompartment. *PLoS Comput Biol* 11: e1004067

Juodeikis R, Lee MJ, Mayer M, Mantell J, Brown IR, Verkade P, Woolfson DN, Prentice MB, Frank S & Warren MJ (2020) Effect of metabolosome encapsulation peptides on enzyme activity, coaggregation, incorporation, and bacterial microcompartment formation. *MicrobiologyOpen* 9: e1010

Kalnins G, Cesle E-E, Jansons J, Liepins J, Filimonenko A & Tars K (2020) Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles. *Nat Commun* 11: 388

Kalnins G, Kuka J, Grinberga S, Makrecka-Kuka M, Liepinsh E, Dambrova M & Tars K (2015) Structure and Function of CutC Choline Lyase from Human Microbiota Bacterium Klebsiella pneumoniae. *Journal of Biological Chemistry* 290: 21732–21740

Kaneko Y, Danev R, Nagayama K & Nakamoto H (2006) Intact Carboxysomes in a Cyanobacterial Cell Visualized by Hilbert Differential Contrast Transmission Electron Microscopy. *Journal of Bacteriology*

Kaval K, Singh K, Cruz M, DebRoy S, Winkler W, Murray B & Garsin D (2018) Loss of Ethanolamine Utilization in Enterococcus faecalis Increases Gastrointestinal Tract Colonization. *mBio* 9

Keeble AH, Yadav VK, Ferla MP, Bauer CC, Chuntharpursat-Bon E, Huang J, Bon RS & Howarth M (2022) DogCatcher allows loop-friendly protein-protein ligation. *Cell Chemical Biology* 29: 339-350.e10

Kennedy NW, Mills CE, Abrahamson CH, Archer AG, Shirman S, Jewett MC, Mangan NM & Tullman-Ercek D (2022) Linking the Salmonella enterica 1,2-Propanediol Utilization Bacterial Microcompartment Shell to the Enzymatic Core via the Shell Protein PduB. *J Bacteriol* 204: e00576-21

Kerfeld C & Erbilgin O (2015) Bacterial microcompartments and the modular construction of microbial metabolism. *Trends in microbiology* 23

Kerfeld CA & Melnicki MR (2016) Assembly, function and evolution of cyanobacterial carboxysomes. *Current Opinion in Plant Biology* 31: 66–75

Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M & Yeates TO (2005) Protein Structures Forming the Shell of Primitive Bacterial Organelles. *Science* 309: 936–938

Khobragade TP, Sarak S, Pagar AD, Jeon H & Giri P (2021) Synthesis of Sitagliptin Intermediate by a Multi-Enzymatic Cascade System Using Lipase and Transaminase With Benzylamine as an Amino Donor. *Front Bioeng Biotechnol* 9: 757062

Kinney JN, Salmeen A, Cai F & Kerfeld CA (2012) Elucidating Essential Role of Conserved Carboxysomal Protein CcmN Reveals Common Feature of Bacterial Microcompartment Assembly. *Journal of Biological Chemistry* 287: 17729–17736

Kirst H, Ferlez BH, Lindner SN, Cotton CAR, Bar-Even A & Kerfeld CA (2022) Toward a glycyl radical enzyme containing synthetic bacterial microcompartment to produce pyruvate from formate and acetate. *Proc Natl Acad Sci USA* 119: e2116871119

Klein MG, Zwart P, Bagby SC, Cai F, Chisholm SW, Heinhorst S, Cannon GC & Kerfeld CA (2009) Identification and Structural Analysis of a Novel Carboxysome Shell Protein with Implications for Metabolite Transport. *Journal of Molecular Biology* 392: 319–333

Kofoid E, Rappleye C, Stojiljkovic I & Roth J (1999) The 17-Gene Ethanolamine (*eut*) Operon of *Salmonella typhimurium* Encodes Five Homologues of Carboxysome Shell Proteins. *J Bacteriol* 181: 5317–5329

Kornprobst M, Turk M, Kellner N, Cheng J, Flemming D, Koš-Braun I, Koš M, Thoms M, Berninghausen O, Beckmann R, *et al* (2016) Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome. *Cell* 166: 380–393

Küffner AM, Prodan M, Zuccarini R, Capasso Palmiero U, Faltova L & Arosio P (2020) Acceleration of an Enzymatic Reaction in Liquid Phase Separated Compartments Based on Intrinsically Disordered Protein Domains. *ChemSystemsChem* 2: e2000001

Kumar G & Sinha S (2022) Phase Separation of Shell Protein and Enzyme: An Insight into the Biogenesis of a Prokaryotic Metabolosome Biophysics

Lamed R, Setter E & Bayer EA (1983) Characterization of a cellulose-binding, cellulase-containing complex in Clostridium thermocellum. *Journal of Bacteriology* 156: 828

Lassila JK, Bernstein SL, Kinney JN, Axen SD & Kerfeld CA (2014) Assembly of Robust Bacterial Microcompartment Shells Using Building Blocks from an Organelle of Unknown Function. *Journal of Molecular Biology* 426: 2217–2228

Lau YH, Giessen TW, Altenburg WJ & Silver PA (2018) Prokaryotic nanocompartments form synthetic organelles in a eukaryote. *Nat Commun* 9: 1311

Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, Xue W-F, Rowe ML, Mulvihill DP, Prentice MB, Howard MJ, *et al* (2014) Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synth Biol* 3: 454–465

Lee MJ, Mantell J, Brown IR, Fletcher JM, Verkade P, Pickersgill RW, Woolfson DN, Frank S & Warren MJ (2018a) De novo targeting to the cytoplasmic and luminal side of bacterial microcompartments. *Nat Commun* 9: 3413

Lee MJ, Mantell J, Hodgson L, Alibhai D, Fletcher JM, Brown IR, Frank S, Xue W-F, Verkade P, Woolfson DN, *et al* (2018b) Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. *Nat Chem Biol* 14: 142–147

Li C, Zhang R, Wang J, Wilson LM & Yan Y (2020a) Protein Engineering for Improving and Diversifying Natural Product Biosynthesis. *Trends in Biotechnology* 38: 729–744

Li T, Jiang Q, Huang J, Aitchison CM, Huang F, Yang M, Dykes GF, He H-L, Wang Q, Sprick RS, *et al* (2020b) Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nat Commun* 11: 5448

Liang M, Frank S, Lünsdorf H, Warren MJ & Prentice MB (2017) Bacterial microcompartment-directed polyphosphate kinase promotes stable polyphosphate accumulation in E. coli. *Biotechnology Journal* 12: 1600415

Lim HN, Lee Y & Hussein R (2011) Fundamental relationship between operon organization and gene expression. *Proc Natl Acad Sci USA* 108: 10626–10631

Lin T, Scott BL, Hoppe AD & Chakravarty S (2018) FRETting about the affinity of bimolecular protein–protein interactions. *Protein Science* 27: 1850–1856

Liu W, Jiang C, Zhang Y, Zhu L, Jiang L & Huang H (2022) Self-assembling protein scaffold-mediated enzymes' immobilization enhances in vitro d-tagatose production from lactose. *Food Bioengineering* 1: 47–57

Liu Y & Dai M (2020) Trimethylamine N-Oxide Generated by the Gut Microbiota Is Associated with Vascular Inflammation: New Insights into Atherosclerosis. *Mediators of Inflammation* 2020: 1–15

Long BM, Badger MR, Whitney SM & Price GD (2007) Analysis of Carboxysomes from Synechococcus PCC7942 Reveals Multiple Rubisco Complexes with Carboxysomal Proteins CcmM and CcaA *. *Journal of Biological Chemistry* 282: 29323–29335

Long BM, Tucker L, Badger MR & Price GD (2010) Functional Cyanobacterial β-Carboxysomes Have an Absolute Requirement for Both Long and Short Forms of the CcmM Protein. *Plant Physiol* 153: 285–293

MacCready JS, Basalla JL & Vecchiarelli AG (2020) Origin and Evolution of Carboxysome Positioning Systems in Cyanobacteria. *Molecular Biology and Evolution* 37: 1434–1451

MacCready JS, Hakim P, Young EJ, Hu L, Liu J, Osteryoung KW, Vecchiarelli AG & Ducat DC (2018) Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *eLife* 7: e39723

MacCready JS, Tran L, Basalla JL, Hakim P & Vecchiarelli AG (2021) The McdAB system positions α-carboxysomes in proteobacteria. *Mol Microbiol* 116: 277–297

Machleidt T, Woodroofe CC, Schwinn MK, Méndez J, Robers MB, Zimmerman K, Otto P, Daniels DL, Kirkland TA & Wood KV (2015) NanoBRET—A Novel BRET Platform for the Analysis of Protein–Protein Interactions. *ACS Publications*

Mahinthichaichan P, Morris DM, Wang Y, Jensen GJ & Tajkhorshid E (2018) Selective Permeability of Carboxysome Shell Pores to Anionic Molecules. *J Phys Chem B* 122: 9110–9118

Mahl MC, Wilson PW, Fife MA & Ewing WH (1965) Nitrogen Fixation by Members of the Tribe *Klebsielleae*. *J Bacteriol* 89: 1482–1487

Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD & Butt TR (2004) SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Func Genom* 5: 75–86

Mallette E & Kimber MS (2017) A Complete Structural Inventory of the Mycobacterial Microcompartment Shell Proteins Constrains Models of Global Architecture and Transport. *Journal of Biological Chemistry* 292: 1197–1210

Martínez-del Campo A, Bodea S, Hamer HA, Marks JA, Haiser HJ, Turnbaugh PJ & Balskus EP (2015) Characterization and Detection of a Widely Distributed Gene Cluster That Predicts Anaerobic Choline Utilization by Human Gut Bacteria. *mBio* 6: e00042-15

Matz MV, Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML & Lukyanov SA (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol* 17: 969–973

McKay RML, Gibbs SP & Espie GS (1993) Effect of dissolved inorganic carbon on the expression of carboxysomes, localization of Rubisco and the mode of inorganic carbon transport in cells of the cyanobacterium Synechococcus UTEX 625. *Arch Microbiol* 159: 21–29

Mellin JR, Tiensuu T, Bécavin C, Gouin E, Johansson J & Cossart P (2013) A riboswitch-regulated antisense RNA in *Listeria monocytogenes*. *Proceedings of the National Academy of Sciences* 110: 13132–13137

Moore TC & Escalante-Semerena JC (2016) The EutQ and EutP proteins are novel acetate kinases involved in ethanolamine catabolism: physiological implications for the function of the ethanolamine metabolosome in Salmonella enterica. *Molecular Microbiology* 99: 497–511

Morise H, Shimomura O, Johnson FH & Winant J (1974) Intermolecular Energy Transfer in the Bioluminescent System of Aequoreai.

Navon-Venezia S, Kondratyeva K & Carattoli A (2017) Klebsiella pneumoniae: a major worldwide source and shuttle for antibiotic resistance. *FEMS Microbiology Reviews* 41: 252–275

Nawrocki KL, Wetzel D, Jones JB, Woods EC & McBride SM (2018) Ethanolamine is a Valuable Nutrient Source that Impacts Clostridium difficile Pathogenesis. *Environmental microbiology* 20: 1419

Ni T, Jiang Q, Ng PC, Shen J, Dou H, Zhu Y, Radecke J, Dykes GF, Huang F, Liu L-N, *et al* (2023) Intrinsically disordered CsoS2 acts as a general molecular thread for α-carboxysome shell assembly. *Nat Commun* 14: 5512

Nikles D, Vana K, Gauczynski S, Knetsch H, Ludewigs H & Weiss S (2008) Subcellular localization of prion proteins and the 37 kDa/67 kDa laminin receptor fused to fluorescent proteins. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1782: 335–340

Noël CR, Cai F & Kerfeld CA (2016) Purification and Characterization of Protein Nanotubes Assembled from a Single Bacterial Microcompartment Shell Subunit. *Adv Mater Interfaces* 3: 1500295

Ochoa JM, Mijares O, Acosta AA, Escoto X, Leon-Rivera N, Marshall JD, Sawaya MR & Yeates TO (2021) Structural characterization of hexameric shell proteins from two types of choline-utilization bacterial microcompartments. *Acta Crystallogr F Struct Biol Commun* 77: 275–285

Ochoa JM, Nguyen VN, Nie M, Sawaya MR, Bobik TA & Yeates TO (2020) Symmetry breaking and structural polymorphism in a bacterial microcompartment shell protein for choline utilization. *Protein Science* 29: 2201–2212

Oltrogge LM, Chaijarasphong T, Chen AW, Bolin ER, Marqusee S & Savage DF (2020) Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nat Struct Mol Biol* 27: 281–287

Ormö M, Cubitt AB, Kallio K, Gross LA, Tsien RY & Remington SJ (1996) Crystal Structure of the ***Aequorea victoria*** Green Fluorescent Protein. *Science* 273: 1392–1395

Palacios S, Starai VJ & Escalante-Semerena JC (2003) Propionyl Coenzyme A Is a Common Intermediate in the 1,2-Propanediol and Propionate Catabolic Pathways Needed for Expression of the prpBCDE Operon during Growth of Salmonella enterica on 1,2-Propanediol. *Journal of Bacteriology*

Pang A, Frank S, Brown I, Warren MJ & Pickersgill RW (2014) Structural Insights into Higher Order Assembly and Function of the Bacterial Microcompartment Protein PduA. *Journal of Biological Chemistry* 289: 22377–22384

Pang A, Liang M, Prentice MB P & Pickersgill R (2012) Substrate channels revealed in the trimeric Lactobacillus reuteri bacterial microcompartment shell protein PduB. *Acta crystallographica Section D, Biological crystallography* 68

Pang A, Warren MJ & Pickersgill RW (2011) Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site. *Acta Crystallogr D Biol Crystallogr* 67: 91–96

Papa MFD & Perego M (2008) Ethanolamine Activates a Sensor Histidine Kinase Regulating Its Utilization in Enterococcus faecalis. *Journal of Bacteriology*

Park KS, Son RG, Kim SH, Abdelhamid MAA & Pack SP (2022) Soluble preparation and characterization of tripartite split GFP for In Vitro reconstitution applications. *Biochemical Engineering Journal* 187: 108643

Parsons J, Dinesh S, Deery E, Leech H, Brindley A, Warren M & Prentice M (2008) Biochemical and Structural Insights into Bacterial Organelle Form and Biogenesis. *Journal of Biological Chemistry* 283: 14366–14375

Parsons JB, Frank S, Bhella D, Liang M, Prentice MB, Mulvihill DP & Warren MJ (2010a) Synthesis of Empty Bacterial Microcompartments, Directed Organelle Protein Incorporation, and Evidence of Filament-Associated Organelle Movement. *Molecular Cell* 38: 305–315

Parsons JB, Lawrence AD, McLean KJ, Munro AW, Rigby SEJ & Warren MJ (2010b) Characterisation of PduS, the pdu Metabolosome Corrin Reductase, and Evidence of Substructural Organisation within the Bacterial Microcompartment. *PLoS ONE* 5: e14009

Pedelacq J-D & Cabantous S (2019) Development and Applications of Superfolder and Split Fluorescent Protein Detection Systems in Biology. *IJMS* 20: 3479

Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC & Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* 24: 79–88

Peeples W & Rosen MK (2020) Phase Separation Can Increase Enzyme Activity by Concentration and Molecular Organization Biochemistry

Penrod JT & Roth JR (2006) Conserving a volatile metabolite: a role for carboxysome-like organelles in Salmonella enterica. *Journal of Bacteriology* 188: 2865–2874

Pitts AC, Tuck LR, Faulds-Pain A, Lewis RJ & Marles-Wright J (2012) Structural Insight into the Clostridium difficile Ethanolamine Utilisation Microcompartment. *PLoS ONE* 7: e48360

Prentice MB (2021) Bacterial microcompartments and their role in pathogenicity. *Current Opinion in Microbiology* 63: 19–28

Rae BD, Long BM, Badger MR & Price GD (2012) Structural Determinants of the Outer Shell of β-Carboxysomes in Synechococcus elongatus PCC 7942: Roles for CcmK2, K3-K4, CcmO, and CcmL. *PLoS ONE* 7: e43871

Rae BD, Long BM, Badger MR & Price GD (2013) Functions, Compositions, and Evolution of the Two Types of Carboxysomes: Polyhedral Microcompartments That Facilitate CO_2 Fixation in Cyanobacteria and Some Proteobacteria. *Microbiol Mol Biol Rev* 77: 357–379

Rathnasingh C, Raj SM, Lee Y, Catherine C, Ashok S & Park S (2012) Production of 3-hydroxypropionic acid via malonyl-CoA pathway using recombinant Escherichia coli strains. *Journal of Biotechnology* 157: 633–640

Rondon MR, Horswill AR & Escalante-Semerena JC (1995) DNA polymerase I function is required for the utilization of ethanolamine, 1,2-propanediol, and propionate by Salmonella typhimurium LT2. *Journal of Bacteriology* 177: 7119

Roof D & Roth J (1992) Autogenous regulation of ethanolamine utilization by a transcriptional activator of the eut operon in Salmonella typhimurium. *Journal of bacteriology* 174

Rowley CA, Anderson CJ & Kendall MM (2018) Ethanolamine influences human commensal Escherichia coli growth, gene expression and competition with enterohemorrhagic E. coli O157:H7. *mBio* 9: e01429-18

Ruiz DM, Turowski VR & Murakami MT (2016) Effects of the linker region on the structure and function of modular GH5 cellulases. *Sci Rep* 6: 28504

Sagermann M, Ohtaki A & Nikolakakis K (2009) Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment. *Proc Natl Acad Sci USA* 106: 8883–8887

Samborska B & Kimber MS (2012) A Dodecameric CcmK2 Structure Suggests β-Carboxysomal Shell Facets Have a Double-Layered Organization. *Structure* 20: 1353–1362

Sampson EM & Bobik TA (2008) Microcompartments for B $_{12}$ -Dependent 1,2-Propanediol Degradation Provide Protection from DNA and Cellular Damage by a Reactive Metabolic Intermediate. *J Bacteriol* 190: 2966–2971

Savage DF, Afonso B, Chen AH & Silver PA (2010) Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery. *Science* 327: 1258–1261

Scarlett F & Turner J (1976) Microbial metabolism of amino alcohols. Ethanolamine catabolism mediated by coenzyme B12-dependent ethanolamine ammonia-lyase in Escherichia coli and Klebsiella aerogenes. *Journal of general microbiology* 95

Schmidt-Dannert S, Zhang G, Johnston T, Quin MB & Schmidt-Dannert C (2018) Building a toolbox of protein scaffolds for future immobilization of biocatalysts. *Appl Microbiol Biotechnol* 102: 8373–8388

Shieh Y-W, Minguez P, Bork P, Auburger JJ, Guilbride DL, Kramer G & Bukau B (2015a) Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science* 350: 678–680

Shieh Y-W, Minguez P, Bork P, Auburger JJ, Guilbride DL, Kramer G & Bukau B (2015b) Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science* 350: 678–680

Shimomura O, Johnson FH & Saiga Y (1962) Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology* 59: 223–239

Shively J, Ball F & Kline B (1973) Electron microscopy of the carboxysomes (polyhedral bodies) of Thiobacillus neapolitanus. *Journal of bacteriology* 116

Sierra-Ibarra E, Alcaraz-Cienfuegos J, Vargas-Tah A, Rosas-Aburto A, Valdivia-López Á, Hernández-Luna MG, Vivaldo-Lima E & Martinez A (2022) Ethanol production by *Escherichia coli* from detoxified lignocellulosic teak wood hydrolysates with high concentration of phenolic compounds. *Journal of Industrial Microbiology and Biotechnology* 49: kuab077

Sinha S, Cheng S, Sung YW, McNamara DE, Sawaya MR, Yeates TO & Bobik TA (2014) Alanine Scanning Mutagenesis Identifies an Asparagine–Arginine–Lysine Triad Essential to Assembly of the Shell of the Pdu Microcompartment. *Journal of Molecular Biology* 426: 2328–2345

Siu K-H, Chen RP, Sun Q, Chen L, Tsai S-L & Chen W (2015) Synthetic scaffolds for pathway enhancement. *Current Opinion in Biotechnology* 36: 98–106

Slininger Lee MF, Jakobson CM & Tullman-Ercek D (2017) Evidence for Improved Encapsulated Pathway Behavior in a Bacterial Microcompartment through Shell Protein Engineering. *ACS Synth Biol* 6: 1880–1891

Smith S (1994) The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB j* 8: 1248–1259

Smith ST & Meiler J (2020) Assessing multiple score functions in Rosetta for drug discovery. *PLOS ONE* 15: e0240450

Sommer M, Sutter M, Gupta S, Kirst H, Turmo A, Lechno-Yossef S, Burton RL, Saechao C, Sloan NB, Cheng X, *et al* (2019) Heterohexamers Formed by CcmK3 and CcmK4 Increase the Complexity of Beta Carboxysome Shells. *Plant Physiol* 179: 156–167

Staib L & Fuchs TM (2015) Regulation of fucose and 1,2-propanediol utilization by Salmonella enterica serovar Typhimurium. *Front Microbiol* 6

Sturms R, Streauslin NA, Cheng S & Bobik TA (2015) In Salmonella enterica, Ethanolamine Utilization Is Repressed by 1,2-Propanediol To Prevent Detrimental Mixing of Components of Two Different Bacterial Microcompartments. *J Bacteriol* 197: 2412–2421

Sun Y, Casella S, Fang Y, Huang F, Faulkner M, Barrett S & Liu L-N (2016) Light Modulates the Biosynthesis and Organization of Cyanobacterial Carbon Fixation Machinery through Photosynthetic Electron Flow. *Plant Physiol* 171: 530–541

Sun Y, Harman VM, Johnson JR, Brownridge PJ, Chen T, Dykes GF, Lin Y, Beynon RJ & Liu L-N (2022) Decoding the Absolute Stoichiometric Composition and Structural Plasticity of α-Carboxysomes. *mBio*

Sutter M, Faulkner M, Aussignargues C, Paasch BC, Barrett S, Kerfeld CA & Liu L-N (2016) Visualization of Bacterial Microcompartment Facet Assembly Using High-Speed Atomic Force Microscopy. *Nano Lett* 16: 1590–1595

Sutter M, Greber B, Aussignargues C & Kerfeld CA (2017) Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* 356: 1293–1297

Sutter M & Kerfeld C (2022) BMC Caller: a webtool to identify and analyze bacterial microcompartment types in sequence data. *Biology direct* 17

Sutter M, Laughlin TG, Sloan NB, Serwas D, Davies KM & Kerfeld CA (2019) Structure of a Synthetic *β* -Carboxysome Shell. *Plant Physiol* 181: 1050–1058

Sutter M, Melnicki MR, Schulz F, Woyke T & Kerfeld CA (2021) A catalog of the diversity and ubiquity of bacterial microcompartments. *Nat Commun* 12: 3809

Sweetlove LJ & Fernie AR (2018) The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation. *Nat Commun* 9: 2136

Takenoya M, Nikolakakis K & Sagermann M (2010) Crystallographic Insights into the Pore Structures and Mechanisms of the EutL and EutM Shell Proteins of the Ethanolamine-Utilizing Microcompartment of Escherichia coli. *Journal of Bacteriology* 192: 6056

Tan YQ, Ali S, Xue B, Teo WZ, Ling LH, Go MK, Lv H, Robinson RC, Narita A & Yew WS (2021) Structure of a Minimal α-Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules*

Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC & Yeates TO (2008) Atomic-Level Models of the Bacterial Carboxysome Shell. *Science* 319: 1083–1086

Tanaka S, Sawaya MR & Yeates TO (2010) Structure and Mechanisms of a Protein-Based Organelle in *Escherichia coli*. *Science* 327: 81–84

Tcherkez GGB, Farquhar GD & Andrews TJ (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci USA* 103: 7246–7251

Thompson MC, Cascio D, Leibly DJ & Yeates TO (2015) An allosteric model for control of pore opening by substrate binding in the E UT L microcompartment shell protein. *Protein Science* 24: 956–975

Trettel DS, Resager W, Ueberheide BM, Jenkins CC & Winkler WC (2022) Chemical probing provides insight into the native assembly state of a bacterial microcompartment. *Structure* 30: 537-550.e5

Tripet B, Yu L, Bautista D, Wong W, Irvin R & Hodges R (1997) Engineering a de novo designed coiled-coil heterodimerization domain for the rapid detection, purification and characterization of recombinantly expressed peptides and proteins. *Protein engineering* 10

Tsai Y, Sawaya M & Yeates T (2009) Analysis of lattice-translocation disorder in the layered hexagonal structure of carboxysome shell protein CsoS1C. *Acta crystallographica Section D, Biological crystallography* 65

Tsai Y, Sawaya MR, Cannon GC, Cai F, Williams EB, Heinhorst S, Kerfeld CA & Yeates TO (2007) Structural Analysis of CsoS1A and the Protein Shell of the Halothiobacillus neapolitanus Carboxysome. *PLoS Biol* 5: e144

Uddin I, Frank S, Warren MJ & Pickersgill RW (2018) A Generic Self-Assembly Process in Microcompartments and Synthetic Protein Nanotubes. *Small* 14: 1704020

Van Teeseling MCF, Neumann S & Van Niftrik L (2013) The Anammoxosome Organelle Is Crucial for the Energy Metabolism of Anaerobic Ammonium Oxidizing Bacteria. *Microb Physiol* 23: 104–117

Wagner AF, Frey M, Neugebauer FA, Schäfer W & Knappe J (1992) The free radical in pyruvate formate-lyase is located on glycine-734. *Proc Natl Acad Sci USA* 89: 996–1000

Waldo GS, Standish BM, Berendzen J & Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* 17: 691–695

Wang B, Zhang L, Dai T, Qin Z, Lu H, Zhang L & Zhou F (2021) Liquid–liquid phase separation in human health and diseases. *Sig Transduct Target Ther* 6: 290

Wang L, Zhao F, Li M, Zhang H, Gao Y, Cao P, Pan X, Wang Z & Chang W (2011) Conformational Changes of rBTI from Buckwheat upon Binding to Trypsin: Implications for the Role of the P8′ Residue in the Potato Inhibitor I Family. *PLoS ONE* 6: e20950

Wells JN, Bergendahl LT & Marsh JA (2016) Operon Gene Order Is Optimized for Ordered Protein Complex Assembly. *Cell Reports* 14: 679–685

Wheatley NM, Gidaniyan SD, Liu Y, Cascio D & Yeates TO (2013) Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins: Pentameric Structure of a Shell Protein from a Grp Microcompartment. *Protein Science* 22: 660–665

Wiryaman T & Toor N (2022) Recent advances in the structural biology of encapsulin bacterial nanocompartments. *Journal of Structural Biology: X* 6: 100062

Xu Q, Alahuhta M, Hewitt P, Sarai NS, Wei H, Hengge NN, Mittal A, Himmel ME & Bomble YJ (2021) Self-Assembling Metabolon Enables the Cell Free Conversion of Glycerol to 1,3-Propanediol. *Front Energy Res* 9: 680313

Yang M, Simpson DM, Wenner N, Brownridge P, Harman VM, Hinton JCD, Beynon RJ & Liu L-N (2020) Decoding the stoichiometric composition and organisation of bacterial metabolosomes. *Nat Commun* 11: 1976

Yang M, Wenner N, Dykes GF, Li Y, Zhu X, Sun Y, Huang F, Hinton JCD & Liu L-N (2022) Biogenesis of a bacterial metabolosome for propanediol utilization. *Nat Commun* 13: 2920

Yoshimoto M, Sato M, Yoshimoto N & Nakao K (2008) Liposomal Encapsulation of Yeast Alcohol Dehydrogenase with Cofactor for Stabilization of the Enzyme Structure and Activity. *Biotechnology Progress* 24: 576–582

You C & Zhang Y-HP (2013) Self-Assembly of Synthetic Metabolons through Synthetic Protein Scaffolds: One-Step Purification, Co-immobilization, and Substrate Channeling. *ACS Synth Biol* 2: 102–110

Young EJ, Burton R, Mahalik JP, Sumpter BG, Fuentes-Cabrera M, Kerfeld CA & Ducat DC (2017) Engineering the Bacterial Microcompartment Domain for Molecular Scaffolding Applications. *Front Microbiol* 8: 1441

Yu K, Liu C, Kim B-G & Lee D-Y (2015) Synthetic fusion protein design and applications. *Biotechnology advances* 33

Zakeri B, Fierer JO, Celik E, Chittock EC, Schwarz-Linek U, Moy VT & Howarth M (2012) Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc Natl Acad Sci USA* 109

Zang K, Wang H, Hartl FU & Hayer-Hartl M (2021) Scaffolding protein CcmM directs multiprotein phase separation in β-carboxysome biogenesis. *Nat Struct Mol Biol* 28: 909–922

Zarzycki J, Erbilgin O & Kerfeld CA (2015) Bioinformatic Characterization of Glycyl Radical Enzyme-Associated Bacterial Microcompartments. *Appl Environ Microbiol* 81: 8315–8329

Zhang G, Quin MB & Schmidt-Dannert C (2018) Self-Assembling Protein Scaffold System for Easy in Vitro Coimmobilization of Biocatalytic Cascade Enzymes. *ACS Catal* 8: 5611–5620

Zhu X-G, Long SP & Ort DR (2010) Improving Photosynthetic Efficiency for Greater Yield. *Annu Rev Plant Biol* 61: 235–261

Supplements

Supplementary table 1. Full sequences of the different vectors used (typical constructs with RMM).

Vector 1

Independent transcripts in pET26b (Kan^R)

RMM-GFP10//RMM-GFP11//GFP1-9

tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttctt
cccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattaggg
tgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctat
ctcggtctattcttttgatttataagggatttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
caggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaa
tgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggt
atcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgaga
atggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcga
gacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattc
ttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagcca
gtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattg
cccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacacccccttgtat
tactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctt
ttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcaga
gcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgc
cagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcga
acgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacag
gagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcgga
gcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcc
tttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcgg
tatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacac
ccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatca
ccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaat
gtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgaga
ggaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactca
gggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttc
cagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctgctaac
cagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgttt
ggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaa
tgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggtt
gaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctg
cattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggcc
ctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatccc
actaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcatt
cagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagac
gcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatact
gttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccac
tgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgc
cgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagct
ccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataa
cgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatg
cgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtccccggccacgggg
gcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcg
ccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaa
ataagatttaaatactttaagaaggagatatacatatgagtagtaacgcgattggtttaattgaaacgaaaggatacgtcgccgcactggctgctgcagatgctatggtaaaa
gctgcaaatgtgaccatcaccgaccggcagcaggttggcgatggcttagtggcagtgatcgtaacgggtgaggttggggccgtaaaagctgccactgaagcaggcgctgaa

actgcgtcgcaggttggcgagctggttagcgtgcatgttatcccacgtccccattcggaactcggcgcacattttagcgttagctcaaaaggtgcggccgcatcagaaggaggc
ggtagcgggggccctggttcgggaggggaaggttctgctgggggagggagcgctggcgggggtctgatttaccagacgatcattacctgagcacacaaacgatcctttcga
aagacctgaacgcaagctgataaggatccacttctcgagttaactcgtgagcaataactagcataacccctttggggcctctaaacgggtcttgaggggttttttgctgaaagta
cacggccgcataatcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataattttacaattgtttaagaaggagatatacatatgagt
agtaacgcgattggtttaattgaaacgaaaggatacgtcgccgcactggctgctgcagatgctatggtaaaagctgcaaatgtgaccatcaccgaccggcagcaggttggcg
atggcttagtggcagtgatcgtaacgggtgaggttggggccgtaaaagctgccactgaagcaggcgctgaaactgcgtcgcaggttggcgagctggttagcgtgcatgttatc
ccacgtccccattcggaactcggcgcacattttagcgttagctcaaaaggtgcggccgcaggcagcggtggcagcccgggcggcggcagcggcggcagcggcagcagcgc
gagcggcggcagcaccagcgaaaaacgcgatcacatggtgctgctggaatatgtgaccgcggcgggcattaccgatgcgagctaatgacaagtatgtcgactcctaggaaa
gctttctcgagttaactcgtgagcaataactagcataacccctttggggcctctaaacgggtcttgaggggttttttgctgaaagtacacggccgcataatcgaaattaatacgac
tcactataggggaattgtgagcggataacaattcccctctagaattaattaagtttaactttaagaaggagatatacatatgcgcaaaggcgaagaactgtttaccggcgtggt
gccgattctgattgaactggatggcgatgtgaacggccataaatttttgtgcgcggcgaaggcgaaggcgatgcgaccattggcaaactgagcctgaaatttatttgcaccac
cggcaaactgccggtgccgtggccgaccctggtgaccaccctgacctatggcgtgcagtgctttagccgctatccggatcacatgaaacgccatgatttttttaaaagcgcgat
gccggaaggctatgtgcaggaacgcaccatttatttaaagatgatggcacctataaaacccgcgcggaagtgaaatttgaaggcgataccctggtgaaccgcattgaactg
aaaggcattgattttaaagaagatggcaacattctgggccataaactggaatataactttaacagccataaagtgtatattaccgcggataaacagaacaacggcattaaag
cgaactttaccattcgccataacgtggaagatggcagcgtgcagctggcggatcattatcagcagaacaccccgattggcgatggcccggtgctgctgccggataacggcag
ctctggtgcacatcaccatcaccatcattaagcggccgcacttgttaccggtcacctctcgagaaaacgcgtcgagagctgagcaataactagcataacccctttggggcctcta
aacgggtcttgaggggttttttgctgaaaggaggaactatatccggat

Vector 2

Separate vector for the 2-vector strategy in pET26b (Kan^R)

RMM-GFP10

tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttctt
cccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattaggg
tgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctat
ctcggtctattcttttgatttataagggatttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
caggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaa
tgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggt
atcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgaga
atggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcga
gacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattc
ttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagcca
gtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattg
cccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacacccttgtat
tactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctt
ttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcaga
gcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgc
cagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcga
acgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacag
gagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcgga
gcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcc
tttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcgg
tatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacac
ccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatca
ccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaat
gtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgaga
ggaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactca
gggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttc
cagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctgctaac
cagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgttt
ggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaa
tgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggtt
gaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctg
cattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggcc
ctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatccc
actaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcatt

cagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagac
gcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatact
gttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccac
tgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgc
cgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagct
ccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataa
cgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatg
cgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtccccggccacgggg
gcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcg
ccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaa
ataagatttaaatactttaagaaggagatatacatatgagtagtaacgcgattggtttaattgaaacgaaaggatacgtcgccgcactggctgctgcagatgctatggtaaaa
gctgcaaatgtgaccatcaccgaccggcagcaggttggcgatggcttagtggcagtgatcgtaacgggtgaggttggggccgtaaaagctgccactgaagcaggcgctgaa
actgcgtcgcaggttggcgagctggttagcgtgcatgttatcccacgtccccattcggaactcggcgcacattttagcgttagctcaaaaggtgcggccgcatcagaaggaggc
ggtagcgggggccctggttcgggaggggaaggttctgctgggggagggagcgctggcgggggtctgatttaccagacgatcattacctgagcacacaaacgatcctttcga
aagacctgaacgcaagctgataaggatccacttctcgagaaaacgcgtcgagagctgagcaataactagcataacccctttggggcctctaaacgggtcttgaggggttttttg
ctgaaaggaggaactatatccggat

Vector 3

Separate vector for the 2-vector strategy in pET15b (Amp^{R})

RMM-GFP11

agaggatcgagatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttacaattgtttaagaaggagatata
catatgagtagtaacgcgattggtttaattgaaacgaaaggatacgtcgccgcactggctgctgcagatgctatggtaaaagctgcaaatgtgaccatcaccgaccggcagca
ggttggcgatggcttagtggcagtgatcgtaacgggtgaggttggggccgtaaaagctgccactgaagcaggcgctgaaactgcgtcgcaggttggcgagctggttagcgtg
catgttatcccacgtccccattcggaactcggcgcacattttagcgttagctcaaaaggtgcggccgcaggcagcggtggcagcccgggcggcggcagcggcggcagcggca
gcagcgcgagcggcggcagcaccagcgaaaaacgcgatcacatggtgctgctggaatatgtgaccgcggcgggcattaccgatgcgagctaatgacaagtatgtcgactcc
taggactcgaggatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccctttggggcctctaaacgggtcttgag
gggttttttgctgaaaggaggaactatatccggatatcccgcaagaggcccggcagtaccggcataaccaagcctatgcctacagcatccagggtgacggtgccgaggatga
cgatgagcgcattgttagatttcatacacggtgcctgactgcgttagcaatttaactgtgataaactaccgcattaaagcttatcgatgataagctgtcaaacatgagaattcttg
aagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataatgtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgttt
atttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttat
tccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaa
cagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtgttgacgccgggcaagagcaa
ctcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataacc
atgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaa
ccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttccc
ggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggt
ctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctga
gataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttg
ataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgctt
gcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtcc
ttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtctt
accgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgag
atacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagctt
ccaggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagc
aacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgct
cgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatgg
tgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacg
cgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagc
tgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgg
gccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacggg
ttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgtt
aatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaa
accgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagc
ctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggccaggacccaacgctgcccgagatgcgccgcgtgcggctgctggagatggcggacgcgatggata
tgttctgccaagggttggtttgcgcattcacagttctccgcaagaattgattggctccaattcttggagtggtgaatccgttagcgaggtgccgccggcttccattcaggtcgaggt

ggcccggctccatgcaccgcgacgcaacgcggggaggcagacaaggtatagggcggcgcctacaatccatgccaacccgttccatgtgctcgccgaggcggcataaatcgc
cgtgacgatcagcggtccagtgatcgaagttaggctggtaagagccgcgagcgatccttgaagctgtccctgatggtcgtcatctacctgcctggacagcatggcctgcaacg
cgggcatcccgatgccgccggaagcgagaagaatcataatggggaaggccatccagcctcgcgtcgcgaacgccagcaagacgtagcccagcgcgtcggccgccatgccg
gcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcg
cgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgcc
ccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccg
ctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgg
gcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatat
aacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaa
ccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtg
agatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtc
gcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctgg
tcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctgg
cacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgt
tgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataag
agacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcg
atggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaagg
agatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcgg
cgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgt

Vector 4

Separate vector for the 2-vector strategy in pACYC (Cm^R)

GFP1-9//RMM-GFP10

ccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcac
attcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcatta
ggaaattaatacgactcactataggggaattgtgagcggataacaattcccctgtagaaataattttgtttaactttaataaggagatataccatggcgcgcaaaggcgaaga
actgtttaccggcgtggtgccgattctgattgaactggatggcgatgtgaacggccataaatttttttgtgcgcggcgaaggcgaaggcgatgcgaccattggcaaactgagcct
gaaatttatttgcaccaccggcaaactgccggtgccgtggccgaccctggtgaccaccctgacctatggcgtgcagtgctttagccgctatccggatcacatgaaacgccatga
ttttttttaaaagcgcgatgccggaaggctatgtgcaggaacgcaccatttatttttaaagatgatggcacctataaaacccgcgcggaagtgaaatttgaaggcgataccctggt
gaaccgcattgaactgaaaggcattgattttaaagaagatggcaacattctgggccataaactggaatataactttaacagccataaagtgtatattaccgcggataaacaga
acaacggcattaaagcgaactttaccattcgccataacgtggaagatggcagcgtgcagctggcggatcattatcagcagaacaccccgattggcgatggcccggtgctgctg
ccggataacggcagctctggtgcacatcaccatcaccatcattaagcgtcggcacttgttaccggtcacctctggagttaactcgtgagcaataactagcataaccccttgggg
cctctaaacgggtcttgaggggttttttgctgaaagtacacggccgcataatcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaata
agatttaaatactttaagaaggagatatacatatgagtagtaacgcgattggtttaattgaaacgaaaggatacgtcgccgcactggctgctgcagatgctatggtaaaagctg
caaatgtgaccatcaccgaccggcagcaggttggcgatggcttagtggcagtgatcgtaacgggtgaggttggggccgtaaaagctgccactgaagcaggcgctgaaactg
cgtcgcaggttggcgagctggttagcgtgcatgttatcccacgtccccattcggaactcggcgcacattttagcgttagctcaaaaggtgcggccgcatcagaaggaggcggta
gcgggggccctggttcgggaggggaaggttctgctgggggagggagcgctggcgggggtctgatttaccagacgatcattacctgagcacacaaacgatcctttcgaaag
acctgaacgcaagctgataaggatccacttctcgagttaactagctgagcaataactagcataacccttggggcctctaaacgggtcttgaggggttttttgctgaaacctca
ggcatttgagaagcacacggtcacactgcttccggtagtcaataaaccggtaaaccagcaatagacataagcggctatttaacgaccctgccctgaaccgacgaccgggtcg
aatttgctttcgaatttctgccattcatccgcttattatcacttattcaggcgtagcaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgccctgccactc
atcgcagtactgttgtaattcattaagcattctgccgacatggaagccatcacagacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcgccttgcgtataatat
ttgcccatagtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcacccagggattggctgagacgaaaaacatattctcaataa
acccttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgttt
cagtttgctcatggaaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtctttcattgccatacggaactccggatgagcattcatcaggcgggcaaga
atgtgaataaaggccggataaaacttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaactgactgaaatgc
ctcaaaatgttctttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgataactcaaaaaatacgccc
ggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacaccaggatt
tatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtggc
ttctgtttctatcagctgtccctcctgttcagctactgacggggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggca
ctgatgagggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactcgct
acgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgagagggccgcggcaaa
gccgtttttccataggctccgccccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactataaagataccaggcgtttccccctggcgg
ctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaag
ctggactgtatgcacgaacccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccac
tggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcctccaagccagttacctcggttcaaagagtt

ggtagctcagagaaccttcgaaaaaccgccctgcaaggcggtttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagat
aaaatatttctagttttcagtgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgttaattctcatgttagtcatgccccgcgcccaccggaaggag
ctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgt
cgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttctttttcaccagtgagacgggcaacagctgattgcccttc
accgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggta
tcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaac
gatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagc
cagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggag
aaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaat
gatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcg
agatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaa
tgtaattcagctccgccatcgccgctt

Vector 5

Separate vector for the 2-vector strategy in pACYC (Cm^R)

GFP1-9//RMM-GFP11

ccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcac
attcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcatta
ggaaattaatacgactcactataggggaattgtgagcggataacaattcccctgtagaaataatttttgtttaactttaataaggagatataccatggcgcgcaaaggcgaaga
actgtttaccggcgtggtgccgattctgattgaactggatggcgatgtgaacggccataaatttttttgtgcgcggcgaaggcgaaggcgatgcgaccattggcaaactgagcct
gaaatttatttgcaccaccggcaaactgccggtgccgtggccgaccctggtgaccaccctgacctatggcgtgcagtgctttagccgctatccggatcacatgaaacgccatga
ttttttttaaaagcgcgatgccggaaggctatgtgcaggaacgcaccatttattttaaagatgatggcacctataaaacccgcgcggaagtgaaatttgaaggcgataccctggt
gaaccgcattgaactgaaaggcattgattttaaagaagatggcaacattctgggccataaactggaatataacttttaacagccataaagtgtatattaccgcggataaacaga
acaacggcattaaagcgaactttaccattcgccataacgtggaagatggcagcgtgcagctggcggatcattatcagcagaacaccccgattggcgatggcccggtgctgctg
ccggataacggcagctctggtgcacatcaccatcaccatcattaagcgtcggcacttgttaccggtcacctctggagttaactcgtgagcaataactagcataacccccttgggg
cctctaaacgggtcttgaggggttttttgctgaaagtacacggccgcataatcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaata
attttacaattgtttaagaaggagatatacatatgagtagtaacgcgattggtttaattgaaacgaaaggatacgtcgccgcactggctgctgcagatgctatggtaaaagctg
caaatgtgaccatcaccgaccggcagcaggttggcgatggcttagtggcagtgatcgtaacgggtgaggttggggccgtaaaagctgccactgaagcaggcgctgaaactg
cgtcgcaggttggcgagctggttagcgtgcatgttatcccacgtccccattcggaactcggcgcacattttagcgttagctcaaaaggtgcggccgcaggcagcggtggcagcc
cgggcggcggcagcggcggcagcggcagcagcgcgagcggcggcagcaccagcgaaaaacgcgatcacatggtgctgctggaatatgtgaccgcggcgggcattaccga
tgcgagctaatgacaagtatgtcgactcctaggaaagctttctcgagttaactagctgagcaataactagcataacccccttggggcctctaaacgggtcttgaggggttttttgc
tgaaacctcaggcatttgagaagcacacggtcacactgcttccggtagtcaataaaccggtaaaccagcaatagacataagcggctatttaacgaccctgccctgaaccgac
gaccgggtcgaatttgctttcgaatttctgccattcatccgcttattatcacttattcaggcgtagcaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccg
ccctgccactcatcgcagtactgttgtaattcattaagcattctgccgacatggaagccatcacagacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcgcctt
gcgtataatatttgcccatagtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcacccagggattggctgagacgaaaaacat
attctcaataaacccttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgat
gaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtctttcattgccatacggaactccggatgagcattcatcagg
cgggcaagaatgtgaataaaggccggataaaacttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaactg
actgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgataactcaaa
aaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacaggga
caccaggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtg
ctccagtggcttctgtttctatcagctgtccctcctgttcagctactgacggggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttac
tatgttggcactgatgagggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctca
ctgactcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgagagggc
cgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactataaagataccaggcgtttc
ccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagtt
cgctccaagctggactgtatgcacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggc
agcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcctccaagccagttacctcggtt
caaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatcttat
taatcagataaaatatttctagttttcagtgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgccccgcgcccacc
ggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgg
gaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttctttttcaccagtgagacgggcaacagctg
attgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagct
gtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgc
agtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatg

ccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtct
tcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcg
gatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttg
atcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgc
ggttgggaatgtaattcagctccgccatcgccgctt

Vector 6

Individual vector for His_6-tagged form expression in pET26b

RMM-His_6

ccggattggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggatttttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacacccccttgtattactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtatttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgctt

Vector 7

Individual vector for His_6-tagged form expression in pET29b

RMM-His_6

agatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatatacatatgagtagtaacgcgattggtttaattgaaacgaaaggatacgtcgccgcactggctgctgcagatgctatggtaaaagctgcaaatgtgaccatcaccgaccggcagcaggttggcgatggcttagtggcagtgatcgtaacgggtgaggttggggccgtaaaagctgccactgaagcaggcgctgaaactgcgtcgcaggttggcgagctggttagcgtgcatgttatcccacgtccccattcggaactcggcgcacattttagcgttagctcaaaaggtgcggccgcattggagcatcaccatcatcaccactaagacctcgagcaccaccaccacca

ccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggt
tttttgctgaaaggaggaactatatccggattggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgcc
agcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacg
gcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactctt
gttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc
gaattttaacaaaatattaacgcttacaatttaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaatt
aattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccga
ggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatc
accatgagtgacgactgaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaa
ccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcat
caacaatattttcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatg
gtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttccc
atacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccattgttggaatttaatcgcggcctagagcaagacgtttccc
gttgaatatggctcataacacccttgtattactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtag
aaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaac
tctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcg
ctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggg
ggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacag
gtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtc
gattttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatc
ccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctg
atgcggtattttctccttacgcatctgtgcggtatttcacaccgcaatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacg
tgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagct
gcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtcca
gctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcat
gggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatg
cggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataa
tggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcg
ctcgcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgccggcg
ataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgc
tccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccg
cgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgcttt
ccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttttcttttcaccagtgagacgggca
acagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataac
atgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaacca
gcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgaga
tatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcg
taccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtca
tccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggca
cccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttg
tgccacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagag
acaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgat
ggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggag
atggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcg
atataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcg

Vector 8

Bicistronic transcript and independent GFP1-9 in pET26b (Kan^R)

RMM-GFP10/RMM-GFP11//GFP1-9

ggattggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttttcgctt
tcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgatta
gggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccc
tatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaa
tttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatca
aatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctg

gtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgag
aatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcg
agacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatatt
cttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagcc
agtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgatt
gcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacacccttgta
ttactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcct
ttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcag
agcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctg
ccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcg
aacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaaca
ggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcgg
agcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgc
ctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcg
gtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgac
acccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcat
caccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgtta
atgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacga
gagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcact
cagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgt
ttccagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctgcta
accagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacg
tttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaa
aatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgg
gttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccag
ctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctg
gccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgta
tcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccc
tcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacg
cagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaata
atactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcag
cccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagattta
atcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaatt
cagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcg
tataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcc
cttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtccccggccc
acggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacct
gtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccct
ctagaaataagatttaaatactttaagaaggagatatacatatgagtagtaacgcgattggtttaattgaaacgaaaggatacgtcgccgcactggctgctgcagatgctatgg
taaaagctgcaaatgtgaccatcaccgaccggcagcaggttggcgatggcttagtggcagtgatcgtaacgggtgaggttggggccgtaaaagctgccactgaagcaggcg
ctgaaactgcgtcgcaggttggcgagctggttagcgtgcatgttatcccacgtccccattcggaactcggcgcacattttagcgttagctcaaaaggtgcggccgcatcagaag
gaggcggtagcgggggccctggttcgggaggggaaggttctgctgggggagggagcgctggcgggggtctgatttaccagacgatcattacctgagcacacaaacgatcc
tttcgaaagacctgaacgcaagctgataaggatcaattgtttaagaaggagatatacatatgagtagtaacgcgattggtttaattgaaacgaaaggatacgtcgccgcactg
gctgctgcagatgctatggtaaaagctgcaaatgtgaccatcaccgaccggcagcaggttggcgatggcttagtggcagtgatcgtaacgggtgaggttggggccgtaaaag
ctgccactgaagcaggcgctgaaactgcgtcgcaggttggcgagctggttagcgtgcatgttatcccacgtccccattcggaactcggcgcacattttagcgttagctcaaaag
gtgcggccgcaggcagcggtggcagcccgggcggcggcagcggcggcagcggcagcagcgcgagcggcggcagcaccagcgaaaaacgcgatcacatggtgctgctgg
aatatgtgaccgcggcgggcattaccgatgcgagctaatgacaagtatgtcgactcctaggaaagctttctcgagttaactcgtgagcaataactagcataaccccttggggc
ctctaaacgggtcttgaggggttttttgctgaaagtacacggccgcataatcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaattaat
taagtttaactttaagaaggagatatacatatgcgcaaaggcgaagaactgtttaccggcgtggtgccgattctgattgaactggatggcgatgtgaacggccataaatttttttg
tgcgcggcgaaggcgaaggcgatgcgaccattggcaaactgagcctgaaatttatttgcaccaccggcaaactgccggtgccgtggccgaccctggtgaccaccctgaccta
tggcgtgcagtgctttagccgctatccggatcacatgaaacgccatgatttttttaaaagcgcgatgccggaaggctatgtgcaggaacgcaccatttatttaaagatgatggc
acctataaaacccgcgcggaagtgaaatttgaaggcgataccctggtgaaccgcattgaactgaaaggcattgattttaaagaagatggcaacattctgggccataaactgg
aatataactttaacagccataaagtgtatattaccgcggataaacagaacaacggcattaaagcgaactttaccattcgccataacgtggaagatggcagcgtgcagctggc
ggatcattatcagcagaacaccccgattggcgatggcccggtgctgctgccggataacggcagctctggtgcacatcaccatcaccatcattaagcggccgcacttgttaccg
gtcacctctcgagaaaacgcgtcgagagctgagcaataactagcataacccctggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatcc

Vector 9

Bicistronic transcript and independent GFP1-9 in pET26b (Kan^R) with NdeI and NotI sites flanking the POI-GFP11 and GFP1-9 ORF retrieved

RMM-GFP10/RMM-GFP11//GFP1-9

ggattggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacaccccttgtattactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgaccacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgttccagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataagatttaaatactttaagaaggagatatacatatgagtagtaacgcgattggtttaattgaaacgaaaggatacgtcgccgcactggctgctgcagatgctatggtaaaagctgcaaatgtgaccatcaccgaccggcagcaggttggcgatggcttagtggcagtgatcgtaacgggtgaggttggggccgtaaa

agctgccactgaagcaggcgctgaaactgcgtcgcaggttggcgagctggttagcgtgcatgttatcccacgtccccattcggaactcggcgcacattttagcgttagctcaaa
aggatccgcaggcagcggtggaagtccgggtggcggttcaggcggtagcggcagctctgcgagcggcggcagcaccagcgaaaaacgcgatcacatggtgctgctggaat
atgtgaccgcggcgggcattaccgatgcgagctaatgacaagtatgtcgactcctaggaaagctttctcgagttaactcgtgagcaataactagcataacccttggggcctct
aaacgggtcttgaggggttttttgctgaaagtacacggccgcataatcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaattaattaa
gtttaactttaagaaggagatatacctatgcgcaaaggcgaagaactgtttaccggcgtggtgccgattctgattgaactggatggcgatgtgaacggccataaatttttgtgc
gcggcgaaggcgaaggcgatgcgaccattggcaaactgagcctgaaatttatttgcaccaccggcaaactgccggtgccgtggccgaccctggtgaccaccctgacctatgg
cgtgcagtgctttagccgctatccggatcacatgaaacgccatgatttttttaaaagcgcgatgccggaaggctatgtgcaggaacgcaccatttatttttaaagatgatggcacc
tataaaacccgcgcggaagtgaaatttgaaggcgataccctggtgaaccgcattgaactgaaaggcattgattttaaagaagatggcaacattctgggccataaactggaat
ataactttaacagccataaagtgtatattaccgcggataaacagaacaacggcattaaagcgaactttaccattcgccataacgtggaagatggcagcgtgcagctggcgga
tcattatcagcagaacaccccgattggcgatggcccggtgctgctgccggataacggcagctctggtgcacatcaccatcaccatcattaagcggcagcacttgttaccggtca
cctctcgagaaaacgcgtcgagagctgagcaataactagcataacccctttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatcc

Vector 10

Separate vector for the 2-vector strategy in pET26b (Kan^{R})

RMM-GFP10//GFP1-9

tcgactcctaggaaagctttctcgagttaactcgtgagcaataactagcataacccctttggggcctctaaacgggtcttgaggggttttttgctgaaagtacacggccgcataat
cgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaattaattaagtttaactttaagaaggagatatacatatgcgcaaaggcgaagaac
tgtttaccggcgtggtgccgattctgattgaactggatggcgatgtgaacggccataaatttttgtgcgcggcgaaggcgaaggcgatgcgaccattggcaaactgagcctga
aatttatttgcaccaccggcaaactgccggtgccgtggccgaccctggtgaccaccctgacctatggcgtgcagtgctttagccgctatccggatcacatgaaacgccatgatttt
ttttaaaagcgcgatgccggaaggctatgtgcaggaacgcaccatttatttttaaagatgatggcacctataaaacccgcgcggaagtgaaatttgaaggcgataccctggtga
accgcattgaactgaaaggcattgattttaaagaagatggcaacattctgggccataaactggaatataacttttaacagccataaagtgtatattaccgcggataaacagaac
aacggcattaaagcgaactttaccattcgccataacgtggaagatggcagcgtgcagctggcggatcattatcagcagaacaccccgattggcgatggcccggtgctgctgcc
ggataacggcagctctggtgcacatcaccatcaccatcattaagcggccgcacttgttaccggtcacctctcgagaaaacgcgtcgagagctgagcaataactagcataaccc
cttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtg
gttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggg
ctcccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacg
ttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaa
aaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttcta
aatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagc
cgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccct
cgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattac
gctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaa
tgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatg
catcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatg
tttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttgg
aatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacacccttgtattactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgt
gagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagc
ggtggtttgtttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccact
tcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttacc
ggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcg
ccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagt
cctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttg
ctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcga
gtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccg
catagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatc
cgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaa
gcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggt
cactgatgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactgg
aacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtag
ccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggt
cgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcac
gatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccg
aataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaag
acagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagct
aacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggc

gccagggtggttttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcga
aaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcg
cgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgcctt
cccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggt
gacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacatta
gtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacg
ccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaac
gccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctgg
cctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgcta
tcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgag
caccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagt
ggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctc
gatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataagatttaaatactttaagaaggagatatacatatgagtagtaac
gcgattggtttaattgaaacgaaaggatacgtcgccgcactggctgctgcagatgctatggtaaaagctgcaaatgtgaccatcaccgaccggcagcaggttggcgatggctt
agtggcagtgatcgtaacgggtgaggttggggccgtaaaagctgccactgaagcaggcgctgaaactgcgtcgcaggttggcgagctggttagcgtgcatgttatcccacgtc
cccattcggaactcggcgcacattttagcgttagctcaaaaggtgcggccgcatcagaaggaggcggtagcgggggccctggttcgggaggggaaggttctgctgggggag
ggagcgctggcgggggtctgatttaccagacgatcattacctgagcacacaaacgatcctttcgaaagacctgaacgcaagctgataaggatcaatt

Vector 11

Tricistronic transcript in pET26b (Kan^{R})

RMM-GFP10/RMM-GFP11/GFP1-9

ccggattggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcg
ctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgat
tagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaac
cctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttac
aatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcat
caaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcc
tggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtg
agaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctga
gcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcagga
tattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtca
gccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctg
attgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacacccctt
gtattactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgaga
tccttttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcag
cagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggct
gctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttgg
agcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcgg
aacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggg
gcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtatt
accgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatct
gtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccc
cgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccg
tcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagc
gttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaa
cgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatc
actcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccg
cgtttccagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctg
ctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaa
acgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgcc
gaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgac
tgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgc
cagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgc
ctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtc
gtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatg

ccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccaga
cgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaa
taatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatc
agcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagat
ttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgta
attcagctccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgaca
tcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctc
tcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtccccg
gccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgca
cctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcc
cctctagaaataagatttaaatactttaagaaggagatatacatatgagtagtaacgcgattggtttaattgaaacgaaaggatacgtcgccgcactggctgctgcagatgcta
tggtaaaagctgcaaatgtgaccatcaccgaccggcagcaggttggcgatggcttagtggcagtgatcgtaacgggtgaggttggggccgtaaaagctgccactgaagcag
gcgctgaaactgcgtcgcaggttggcgagctggttagcgtgcatgttatcccacgtccccattcggaactcggcgcacattttagcgttagctcaaaaggtgcggccgcatcag
aaggaggcggtagcgggggccctggttcgggaggggaaggttctgctgggggagggagcgctggcgggggtctgatttaccagacgatcattacctgagcacacaaacg
atcctttcgaaagacctgaacgcaagctgataaggatcaattgtttaagaaggagatatacatatgagtagtaacgcgattggtttaattgaaacgaaaggatacgtcgccgc
actggctgctgcagatgctatggtaaaagctgcaaatgtgaccatcaccgaccggcagcaggttggcgatggcttagtggcagtgatcgtaacgggtgaggttggggccgta
aaagctgccactgaagcaggcgctgaaactgcgtcgcaggttggcgagctggttagcgtgcatgttatcccacgtccccattcggaactcggcgcacattttagcgttagctca
aaaggtgcggccgcaggcagcggtggcagcccgggcggcggcagcggcggcagcggcagcagcgcgagcggcggcagcaccagcgaaaaacgcgatcacatggtgctg
ctggaatatgtgaccgcggcgggcattaccgatgcgagctaatgacaagtatgtcgactcctaggaaagctattaagtttaactttaagaaggagatatacatatgcgcaaag
gcgaagaactgtttaccggcgtggtgccgattctgattgaactggatggcgatgtgaacggccataaatttttgtgcgcggcgaaggcgaaggcgatgcgaccattggcaaa
ctgagcctgaaatttatttgcaccaccggcaaactgccggtgccgtggccgaccctggtgaccaccctgacctatggcgtgcagtgctttagccgctatccggatcacatgaaa
cgccatgatttttttaaaagcgcgatgccggaaggctatgtgcaggaacgcaccatttatttaaagatgatggcacctataaaacccgcgcggaagtgaaatttgaaggcgat
accctggtgaaccgcattgaactgaaaggcattgattttaaagaagatggcaacattctgggccataaactggaatataactttaacagccataaagtgtatattaccgcggat
aaacagaacaacggcattaaagcgaactttaccattcgccataacgtggaagatggcagcgtgcagctggcggatcattatcagcagaacaccccgattggcgatggcccg
gtgctgctgccggataacggcagctctggtgcacatcaccatcaccatcattaagcggccgcacttgttaccggtcacctctcgagaaaacgcgtcgagagctgagcaataac
tagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatat

Vector 12

Receptor vector for Gibson assembly with *Kpe 342* BMC-H

CcmO-GFP10/GFP11-CcmP//GFP1-9

ttctcgagttaactcgtgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaagtacacggccgcataatcgaaattaatacgactca
ctataggggaattgtgagcggataacaattcccctctagaattaattaagtttaactttaagaaggagatatacatatgcgcaaaggcgaagaactgtttaccggcgtggtgcc
gattctgattgaactggatggcgatgtgaacggccataaatttttgtgcgcggcgaaggcgaaggcgatgcgaccattggcaaactgagcctgaaatttatttgcaccaccgg
caaactgccggtgccgtggccgaccctggtgaccaccctgacctatggcgtgcagtgctttagccgctatccggatcacatgaaacgccatgatttttttaaaagcgcgatgcc
ggaaggctatgtgcaggaacgcaccatttatttaaagatgatggcacctataaaacccgcgcggaagtgaaatttgaaggcgataccctggtgaaccgcattgaactgaaa
ggcattgattttaaagaagatggcaacattctgggccataaactggaatataactttaacagccataaagtgtatattaccgcggataaacagaacaacggcattaaagcgaa
ctttaccattcgccataacgtggaagatggcagcgtgcagctggcggatcattatcagcagaacaccccgattggcgatggcccggtgctgctgccggataacggcagctctg
gtgcacatcaccatcaccatcattaagcggccgcacttgttaccggtcacctctcgagaaaacgcgtcgagagctgagcaataactagcataaccccttggggcctctaaacg
ggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgac
cgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgat
ttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttcttta
atagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaaca
aaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtat
ccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggag
aaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatc
aagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactc
gcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaaca
ctgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgtttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggat
aaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggc
gcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctaga
gcaagacgtttcccgttgaatatggctcataacacccttgtattactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagc
gtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatc
aagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcacc
gcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtc
gggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaaggga

gaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacc
tctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgt
tctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagc
ggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatac
actccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtg
accgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcc
tgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaag
ggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaa
ctggcggtatggatgcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgc
agatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagc
agtcgcttcacgttcgctcgcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggg
gccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacagg
ccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcga
cgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgc
gctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttttcttttca
ccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggtt
aacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccat
ctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctga
atttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatg
ctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacag
caatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcga
caccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactg
tttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggа
aacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaag
gttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaat
ggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttcccca
tcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacg
actcactataggggaattgtgagcggataacaattcccctctagaaataaggctagcatactttaagaaggagatatacatatgcctacgagcccaactatgacaagcgtccc
gattgcccgttcaccgcgaccgtcttatcagcaaattaaccagcatcagccgtcggatagcgcgttaggcttagtgtctacccgatcgttcccagcaatcgtagggacagcagat
atgatgctaaaatcgagccaggtgaccttagtgggttatgaaaaaatcgggagcggctattgcacggcggtggtgcggggcaaggtggcggatgtgcgtcttgccgtagaag
agggcgctcgtacagcggagcagttcggccaactggttagcaaattagtgatcccgcggccgatgcccaatcttcaggctgtgttcccaatcgggagccatttagtggagctgg
cacagcaacagcggggctacagccggctgtctaaccgctcgattggcttactggagacccgtggctttccggcaatggttggcgcggcggatgcgatgttaaaatcggcggat
gttcagctcgcgtcgtatgaaatcatcggcgatggtttatgcacggcgattgttcgtgggaccgtcgcgaacgttgcgatggctatcgaggtgggtatgcaagaagccgaacga
atcggtgaactacatgcagtaatgatcatcccacgcttactggaggatttagaacataccttccggtcgccacctattggcttgatgaaaatgaaccactgccaatgctactgc
caaaccaggtgcgtgaaaaacaacgccagctggttgcgttaccggagctggagaaagctgtggttccacagaggcaggctaaacccctgcccctgcaagaaaagaccgaa
gccccactggtcctggagaaagaggcggagaaacccattgttgaagtcctgggtccggagattgatggtgcggccgcatcagaaggaggcggtagcgggggccctggttcg
ggaggggaaggttctgctgggggagggagcgctggcgggggtctgatttaccagacgatcattacctgagcacacaaacgatcctttcgaaagacctgaacgcaagctaa
tgaggtacctacctaggttatagaaggagatatacaaatggaaaaacgcgatcacatggtgctgctggaatatgtgaccgcggcgggcattaccgatgcgagcggtgggtcc
ggctcagaaggaggcggtagcgggggccctggttcgggaggggaaggttctgctggtggagggagcgcaagcggcgccatgggtattgagctgcgaagctacgtgtactta
gatagtttacagtcacagcacgcggcatatatcggtacggtggcgtcgggttttcttccgcttccgggggattgcagcttgtgggtagaagtgagcccaggtattgaaattaacc
gcattaccgacatcgctcttaaagctgcggtcgttcggccgggtgtgttatttgttgaacggctgtatggcttactggagattcatgcgagcaaccagggtgaagtacgtgctgc
gggccaagcgattctggcctatattggtgcgaaggcgagcgattgcatcaagccgaaggtggtgagcagccagatcatccgtaatatcgatgcctatcagacgcagttgatta
accgtaatcgccggggccacatgctactggctggccagacgctgtttgtgttagaagttcagccagcggcgtatgcgagcttggcggcaaacgaggcggagaaatcagcgtc
gatcaacattttgcaggtcagctcgatcggttcttttgggcgtttgtatttaggcggtgaagagcgcgatattaaggcgggggcgcgggcggcaatcgctgcgatcgagaacgc
cccaggtaaagttccgaccctggagggcaaaaacgaaggttaatgagcctcctccgtacgtaggaaagct

Supplementary table 2. Mono BMC-H sequences.

Case	Sequence (NdeI to NotI sites)
Mono1	CATATGAGTAGTAACGCGATTGGTatgATTGAAACGAAAGGActgGTCGCCGCAatcgagGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGACCctgACCggcCGGgaaaaaatcGGCGATGGCTTAGTGacgGTGatgGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTAGCGT GCATGTTATCCCACGTCCCCATTCGGAAgtaGGCGCAattctgAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono2	CATATGAGTAGTAACGCGATTGGTatgATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCggcCGGgaagaaGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGG GTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTA GCGTGCATGTTATCCCACGTCCCCATTCGgatCTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono3	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGAgcgGTCGCCGCAatcGCTGCTGCAGATGCTATG actAAAGCTGCAAATGTGACCATCACCGAtgtacgtgaaatcGGCGATGGCTTAGTGtgtGTGgtgGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccGTGC ATGTTATCCCACGTCCCCATTCGaaaCTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono4	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcgCGGgaaaccatcGGCGATGGCTTAGTGGCAGTGATCGTAACGGG TGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgcc GTGCATGTTATCCCACGTCCCCATTCGgatCTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono5	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGgtgATCACCgcggtacatgtgatcGGCGATGGCTTAGTGGCAGTGATCGTAACGGGTG AGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccGT GCATGTTATCCCAcacCCCCATgaggatCTCGGCGCACATTTTgacatcAGCTCAAAAGGTGCGGCCGC
Mono6	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGAttcGTCGCCGCAaccGCTGCTGCAGATGCTATG GTAAAAGCTGCAgacGTGatcATCACCgcggtatatatgatcGGCGATGGCTTAGTGGCAGTGATCGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccGTGC ATGTTATCCCAcacCCCCATgaggatCTCGGCGCACATTTTgacatcAGCTCAAAAGGTGCGGCCGC
Mono7	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcggtacacatgatcGGCGATGGCTTAGTGGCAGTGATCGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccG TGCATGTTATCCCAcacCCCCATTCGGAACTCGGCGCACATTTTgacGTTAGCTCAAAAGGTGCGGCCGC
Mono8	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GactAAAGCTGCAAATGTGACCATCACCgaggtaaaagaaatcGGCGATGGCTTAGTGGCAGTGATCGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccG TGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCAaaaTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono9	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcggtatttaccGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccG TGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono10	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCGACgtaCAGCAGatcGGCGATGGCTTAGTGGCAGTGATCGTAACGG GTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgc cGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono11	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcggtatatgtgGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTAGC GTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono12	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCAaccGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcggtaCAGaccGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGGG TGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTAG CGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC

Supplementary table 2. Mono BMC-H sequences (continuation).

Case	Sequence (NdeI to NotI sites)
Mono13	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCAaccGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGgtgGTTGGCGATGGCTTAGTGGCAGTGATCGTAACG GGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTT gccGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCAcgtTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono14	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCAaccGCTGCTGCAGATGCTAT GGTAAAAGCTGCAgacGTGcaaATCACCgcgCGGgaagtgGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGGG TGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgcc GTGCATGTTATCCCAaatCCCCATaaggatCTCGGCGCAcgtTTTaatGTTAGCTCAAAAGGTGCGGCCGC
Mono15	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCAaccGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGgtgGTTGGCGATGGCTTAGTGGCAGTGATCGTAACG GGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTT gccGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono16	CATATGAGTAGTAACGCGgcgGGTTTAATTGAAACGAAAGGAttcGTCGCCGCAaccGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCaccACCgcgactaaagtgGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGGGTG AGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccGT GCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono17	CATATGAGTAGTAACGCGgcgGGTTTAATTGAAACGAAAGGAttcGTCGCCGCAaccGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCaccACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGG GTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTA GCGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono18	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAgcgGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcggtatatagcatcGGCGATGGCTTAGTGGCAGTGATCGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccG TGCATGTTgacCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono19	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAgcgGCTAT GGTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACG GGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTT AGCGTGCATGTTgacCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono 20	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAagcGCTAT GGTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACG GGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTT AGCGTGCATGTTgacCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono21	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGAgagGTCGCCGCAaccGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACG GGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTT AGCGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono22	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGAcacGTCGCCGCAaccGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGgacGTAACGG GTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTA GCGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono23	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGAaagGTCGCCGCAgcgGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGaccGTAACG GGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTT AGCGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono24	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCAgcgGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAgacATCGTAACG GGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTT AGCGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATcgtAGCGTTAGCTCAAAAGGTGCGGCCGC

Supplementary table 2. Mono BMC-H sequences (continuation).

Case	Sequence (NdeI to NotI sites)
Mono25	CATATGAGTAGTAACGCGgagGGTTTAATTGAAACGAAAGGAgcgGTCGCCGCACgtGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGG GTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTA GCGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono26	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGAgcgGTCGCCGCAcgtGCTGCTGCAagcGCTATG GTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGG GTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTA GCGTGCATGTTgacCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono27	CATATGAGTAGTAACGCGATTGGTgagATTGAAACGAAAGGATACGTCGCCGCAagcGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACG GGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTT AGCGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono28	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcggacaaaCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGG GTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTA GCGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono29	CATATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcggaccgtCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGGG TGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTAG CGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono30	CATATGAGTAGTAACGCGgtgGGTTTAATTGAAACGAAAGGATACgcaGCCGCACTGGCTGCTGCAGATGCTAT GGTAcgtGCTGCAAATGTGACCATCACCgcggtatatcgtatcGGCGATGGCTTAGTGtctGTGgtgGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgagGTGg aaGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono31	CATATGAGTAGTgatGCGATTGGTTTAATTagcgcacgtGGATACactGCCGCACTGGCTGCTGCAGATGCTgcgGT AcgtGCTGCAgacGTGgagATCgtgggcgtatttgaaGTTGGCGATGGCatcGTGactGTggtgttcaagGGTaaggaggaag acgTAAAAaagGCCgtcGAAGCAGGCGCTGAAACTGCGTCGaaaGTTGGCGAGCTGatcgaggcggatGTTATCCCA aagCCCCATgagaaaCTCGGCaagtatTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Mono32	CATATGAGTAGTAACGCGgtgGGTTTAATTaccACGAAAGGAttcactGCCGCACTGGCTGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGacaGTGgtgGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTAGC GTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC

Supplementary table 3. Hybrid BMC-H sequences.

Case	Sequence (NdeI to NotI sites)
Hybrid5wt	cataTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACtgGCTGCTGCAGATGCTATG GTAAAAGCTGCAaatGTGgtgATCACCgacgtacaggtgGTTGGCGATggtctcGTGGCAGTGATCGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTagcGTGC ATGTTATCCCAcgtCCCCATtcggaaCTCGGCGCAcatTTTagcGTTAGCTCAAAAGGTGCGGCCGC
Hybridwt5	cataTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGaccATCACCgcgcggcatcagatcGGCGATggtctcGTGGCAGTGATCGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccGTG CATGTTATCCCAcacCCCCATgaggatCTCGGCGCACATTTTgacatcAGCTCAAAAGGTGCGGCCGC
Hybrid9wt	CAtaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCGACgtaCAGaccGTTGGCGATggtctcGTGGCAGTGATCGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTAGC GTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Hybridwt9	CAtaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcgCGGtttCAGGTTGGCGATggtctcGTGGCAGTGATCGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccG TGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Hybrid14wt	CAtaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCAaccGCTGCTGCAGATGCTAT GGTAAAAGCTGCAgacGTGcaaATCACCGACCGGCAGgtgGTTGGCGATggtctcGTGGCAGTGATCGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTAGC GTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Hybridwt14	CAtaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcgCGGgaaCAGGTTGGCGATggtctcGTGGCAGTGATCGTAACGGG TGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgcc GTGCATGTTATCCCAaatCCCCATaaggatCTCGGCGCAcgtTTTaatGTTAGCTCAAAAGGTGCGGCCGC
Hybrid59	CAtaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcggtatttgtgGTTGGCGATggtctcGTGGCAGTGATCGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccGTG CATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Hybrid95	CAtaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcggtacataccatcGGCGATggtctcGTGGCAGTGATCGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccGTG CATGTTATCCCAcacCCCCATgaggatCTCGGCGCACATTTTgacatcAGCTCAAAAGGTGCGGCCGC
Hybrid514	CAtaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGgtgATCACCgcggtagaagtgGTTGGCGATggtctcGTGGCAGTGATCGTAACGGGTG AGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccGT GCATGTTATCCCAaatCCCCATaaggatCTCGGCGCAcgtTTTaatGTTAGCTCAAAAGGTGCGGCCGC
Hybrid145	CAtaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCAaccGCTGCTGCAGATGCTAT GGTAAAAGCTGCAgacGTGcaaATCACCgcgCGGcatgtgatcGGCGATggtctcGTGGCAGTGATCGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccGTG CATGTTATCCCAcacCCCCATgaggatCTCGGCGCACATTTTgacatcAGCTCAAAAGGTGCGGCCGC
Hybrid914	CAtaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTAT GGTAAAAGCTGCAAATGTGACCATCACCgcggtagaaaccGTTGGCGATggtctcGTGGCAGTGATCGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccG TGCATGTTATCCCAaatCCCCATaaggatCTCGGCGCAcgtTTTaatGTTAGCTCAAAAGGTGCGGCCGC
Hybrid149	CAtaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCCGCAaccGCTGCTGCAGATGCTAT GGTAAAAGCTGCAgacGTGcaaATCACCgcgCGGtttgtgGTTGGCGATggtctcGTGGCAGTGATCGTAACGGGTG AGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgccGT GCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC

Supplementary table 4. Control POI sequences.

DNA fragments were ordered as POI1-GFP10 or POI2-GFP11. Homology region with the open vector or adjacent fragment (in turquoise) permitted assembly by Gibson. CutA-GFP10 and CutA-GFP11 are given as examples of GFP10/11-tagged fragment organization. All subsequent POI sequences are only given between NdeI (in purple) and NotI sites (in blue). ORFs are in bold letter, the GFP10 sequence in light green and GFP11 in dark green. MunI site is in orange and SalI in red.

Case	Sequence	Origin
CutA-GFP10	taagaaggagatatacatatg**gaagaggtcgtgctgatcacggtgccgagcgaggaggtggcgcgtaccatcgccaa ggccctggtggaggagcgcttggccgcctgcgtgaacatcgtccccggcctgacctccatctaccgctggcaggggga ggtggtggaagaccaggagctgctgttgctggtcaagaccaccacccacgccttccctaagctgaaggaacgtgtcaag gccctccacccctacaccgtgcccgagatcgtggccctgcccatcgccgaggggaaccgtgagtacctggactggctt cgtgagaacacgggaggt**gcggccgcatcagaaggaggcggtagcgggggccctggttcgggtggggaaggttctgct ggggagggagcgctggcgggggtct**gatttaccagacgatcattacctgagcacacaaacgatcctttcgaaagacc tgaacgcaagctgataaggatcctgacaattgttatagaaggagata	
CutA-GFP11	ttatagaaggagatatacatatg**gaagaggtcgtgctgatcacggtgccgagcgaggaggtggcgcgtaccatcgcc aaggccctggtggaggagcgcttggccgcctgcgtgaacatcgtccccggcctgacctccatctaccgctggcagggg gaggtggtggaagaccaggagctgctgttgctggtcaagaccaccacccacgccttccctaagctgaaggaacgtgtca aggccctccacccctacaccgtgcccgagatcgtggccctgcccatcgccgaggggaaccgtgagtacctggactggc ttcgtgagaacacgggaggt**gcggccgcaggtagcggtggctctccgggtggcggcagtggtggcagcggcagcagcg cgagcggcggcagcaccagc**gaaaaacgcgatcacatggtgctgctggaatatgtgaccgcggcgggcattaccgatg cgagctaatgacaagtatgtcgactcctaggaaagctcc	
Im9	catATG**GAACTGAAGCATAGCATTAGTGATTATACAGAAGCTGAATTTTTACAACTTGTAA CAACAATTTGTAATGCGGACACTTCCAGTGAAGAAGAACTGGTTAAATTGGTTACACACT TTGAGGAAATGACTGAGCACCCTAGTGGTAGTGATTTAATATATTACCCAAAAGAAGGTG ATGATGACTCACCTTCAGGTATTGTAAACACAGTAAAACAATGGCGAGCCGCTAACGGTA AGTCAGGATTTAAACAG**ggcgcggccgc	*E. coli*
Im2	catATG**GAACTGAAACATAGTATTAGTGATTATACCGAGGCTGAATTTCTGGAGTTTGTAA AAAAAATATGTAGAGCTGAAGGTGCTACTGAAGAGGATGACAATAAATTAGTGAGAGA GTTTGAGCGATTAACTGAGCACCCAGATGGTTCTGATCTGATTTATTATCCTCGCGATGACA GGGAAGATAGTCCTGAAGGGATTGTCAAGGAAATTAAAGAATGGCGAGCTGCTAACGG TAAGTCAGGATTTAAACAG**ggcgcggccgc	*E. coli*
E9 colicin His575Ala	catATG**GAGAGTAAACGGAATAAGCCAGGGAAGGCGACAGGTAAAGGTAAACCAGTTG GTGATAAATGGCTGGATGATGCAGGTAAAGATTCAGGAGCGCCAATCCCAGATCGCATT GCTGATAAGTTGCGTGATAAAGAGTTCAAAAGTTTCGACGATTTTCGGAAGGCTGTATGG GAAGAGGTGTCGAAAGATCCTGAGCTGAGCAAAAACTTGAACCCAAGCAATAAGTCCA GTGTTTCAAAAGGTTATTCTCCGTTTACTCCAAAGAATCAACAGGTCGGAGGGAGAAAAG TCTATGAACTTCATCATGACAAGCCAATTAGTCAAGGTGGTGAGGTTTATGACATGGATA ATATCCGAGTGACTACACCTAAGCGAGCGATCGATATTCACCGAGGTAAG**ggcgcggccgc	*E. coli*
Smt3 (SUMO)	catATG**TCCGATAGCGAAGTGAACCAGGAGGCGAAACCAGAAGTAAAACCCGAAGTAA AACCCGAAACACACATTAATTTGAAAGTAAGCGACGGCTCGAGCGAGATTTTTTTTAAGA TTAAGAAAACGACCCCACTGCGGCGTCTGATGGAGGCCTTTGCCAAACGTCAGGGTAAA GAGATGGACAGCTTGCGTTTCCTGTACGATGGTATCCGTATCCAGGCTGATCAGACGCCG GAGGATCTGGATATGGAGGATAATGACATTATCGAAGCACATCGTGAACAAATCGGG**gg cgcggccgc	*S. cerevisiae*
CobT	catATG**CGTATTACAACCAAGGTTGGTGACAAAGGCTCGACACGCCTGTTTGGTGGGGAG GAAGTCTGGAAAGATTCCCCAATCATTGAGGCAAACGGCACCCTGGATGAACTCACGAG TTTTATTGGGGAAGCCAAGCACTACGTTGACGAGGAGATGAAAGGGATCCTGGAGGAAA TTCAAAACGACATTTACAAGATCATGGGGGAAATTGGCAGTAAGGGTAAGATCGAAGGC ATCAGTGAGGAGCGTATCAAGTGGCTGGAAGGGCTGATTTCTCGCTATGAAGAAATGGTC AATCTGAAGTCTTTTGTACTGCCAGGGGGTACTCTGGAAAGTGCTAAGCTGGATGTATGCC GTACCATTGCTCGCCGTGCCGAACGCAAGGTTGCTACAGTATTACGTGAATTTGGTATCGG TAAGGAGGCGCTGGTTTACTTGAATCGGCTGAGTGATCTGCTGTTCTTGCTGGCACGCGTT ATTGAAATCGAAAAGAACAAACTGAAGGAGGTCCGTTCA**ggcgcggccgc	*Pyrococcus horikoshii OT3*

Supplementary table 4. Control POI sequences (continuation).
DNA fragments were ordered as POI1-GFP10 or POI2-GFP11. Homology regions with the open vector or adjacent fragment (in turquoise) permitted assembly by Gibson. CutA-GFP10 and CutA-GFP11 are given as examples of GFP10/11-tagged fragment organization. All subsequent POI sequences are only given between NdeI (in purple) and NotI sites (in blue). ORFs are in bold letter, the GFP10 sequence in light green and GFP11 in dark green. MunI site is in orange and SalI in red.

Case	Sequence	Origin
CutA-GFP10	taagaaggagatatacatatg**gaagaggtcgtgctgatcacggtgccgagcgaggaggtggcgcgtaccatcgccaaggccctggtggaggagcgcttggccgcctgcgtgaacatcgtccccggcctgacctccatctaccgctggcaggggga ggtggtggaagaccaggagctgctgttgctggtcaagaccaccacccacgccttccctaagctgaaggaacgtgtcaag gccctccacccctacaccgtgcccgagatcgtggccctgcccatcgccgaggggaaccgtgagtacctggactggctt cgtgagaacacgggaggt**gcggccgc**atcagaaggaggcggtagcgggggccctggttcgggtggggaaggttctgct gggggagggagcgctggcgggggtct**gatttaccagacgatcattacctgagcacacaaacgatcctttcgaaagacc tgaacgcaagctgataa**ggatcctga**caattg**ttatagaaggagata	
CutA-GFP11	ttatagaaggagatatacatatg**gaagaggtcgtgctgatcacggtgccgagcgaggaggtggcgcgtaccatcgcc aaggccctggtggaggagcgcttggccgcctgcgtgaacatcgtccccggcctgacctccatctaccgctggcagggg gaggtggtggaagaccaggagctgctgttgctggtcaagaccaccacccacgccttccctaagctgaaggaacgtgtca aggccctccacccctacaccgtgcccgagatcgtggccctgcccatcgccgaggggaaccgtgagtacctggactggc ttcgtgagaacacgggaggt**gcggccgc**aggtagcggtggctctccgggtggcggcagtggtggcagcggcagcagcg cgagcggcggcagcaccagc**gaaaaacgcgatcacatggtgctgctggaatatgtgaccgcggcgggcattaccgatg cgagctaatga**caagtat**gtcgac**tcctaggaaagctcc	
Im9	cat**ATG**GAACTGAAGCATAGCATTAGTGATTATACAGAAGCTGAATTTTTACAACTTGTAA CAACAATTTGTAATGCGGACACTTCCAGTGAAGAAGAACTGGTTAAATTGGTTACACACT TTGAGGAAATGACTGAGCACCCTAGTGGTAGTGATTTAATATATTACCCAAAAGAAGGTG ATGATGACTCACCTTCAGGTATTGTAAACACAGTAAAACAATGGCGAGCCGCTAACGGTA AGTCAGGATTTAAACAG**ggcgcggccgc**	*E. coli*
Im2	cat**ATG**GAACTGAAACATAGTATTAGTGATTATACCGAGGCTGAATTTCTGGAGTTTGTAA AAAAAATATGTAGAGCTGAAGGTGCTACTGAAGAGGATGACAATAAATTAGTGAGAGA GTTTGAGCGATTAACTGAGCACCCAGATGGTTCTGATCTGATTTATTATCCTCGCGATGACA GGGAAGATAGTCCTGAAGGGATTGTCAAGGAAATTAAAGAATGGCGAGCTGCTAACGG TAAGTCAGGATTTAAACAG**gcgcggccgc**	*E. coli*
E9 colicin His575Ala	cat**ATG**GAGAGTAAACGGAATAAGCCAGGGAAGGCGACAGGTAAAGGTAAACCAGTTG GTGATAAATGGCTGGATGATGCAGGTAAAGATTCAGGAGCGCCAATCCCAGATCGCATT GCTGATAAGTTGCGTGATAAAGAGTTCAAAAGTTTCGACGATTTTCGGAAGGCTGTATGG GAAGAGGTGTCGAAAGATCCTGAGCTGAGCAAAAACTTGAACCCAAGCAATAAGTCCA GTGTTTCAAAAGGTTATTCTCCGTTTACTCCAAAGAATCAACAGGTCGGAGGGAGAAAAG TCTATGAACTTCATCATGACAAGCCAATTAGTCAAGGTGGTGAGGTTTATGACATGGATA ATATCCGAGTGACTACACCTAAGCGAGCGATCGATATTCACCGAGGTAAG**ggcgcggccgc**	*E. coli*
Smt3 (SUMO)	cat**ATG**TCCGATAGCGAAGTGAACCAGGAGGCGAAACCAGAAGTAAAACCCGAAGTAA AACCCGAAACACACATTAATTTGAAAGTAAGCGACGGCTCGAGCGAGATTTTTTTTAAGA TTAAGAAAACGACCCCACTGCGGCGTCTGATGGAGGCCTTTGCCAAACGTCAGGGTAAA GAGATGGACAGCTTGCGTTTCCTGTACGATGGTATCCGTATCCAGGCTGATCAGACGCCG GAGGATCTGGATATGGAGGATAATGACATTATCGAAGCACATCGTGAACAAATCGGG**gg cgcggccgc**	*S. cerevisiae*
CobT	cat**ATG**CGTATTACAACCAAGGTTGGTGACAAAGGCTCGACACGCCTGTTTGGTGGGGAG GAAGTCTGGAAAGATTCCCCAATCATTGAGGCAAACGGCACCCTGGATGAACTCACGAG TTTTATTGGGGAAGCCAAGCACTACGTTGACGAGGAGATGAAAGGGATCCTGGAGGAAA TTCAAAACGACATTTACAAGATCATGGGGGAAATTGGCAGTAAGGGTAAGATCGAAGGC ATCAGTGAGGAGCGTATCAAGTGGCTGGAAGGGCTGATTTCTCGCTATGAAGAAATGGTC AATCTGAAGTCTTTTGTACTGCCAGGGGGTACTCTGGAAAGTGCTAAGCTGGATGTATGCC GTACCATTGCTCGCCGTGCCGAACGCAAGGTTGCTACAGTATTACGTGAATTTGGTATCGG TAAGGAGGCGCTGGTTTACTTGAATCGGCTGAGTGATCTGCTGTTCTTGCTGGCACGCGTT ATTGAAATCGAAAAGAACAAACTGAAGGAGGTCCGTTCA**ggcgcggccgc**	*Pyrococcus horikoshii OT3*
VHH	cat**ATG**GCAGATGTGCAGCTGCAGGAGTCTGGGGGAGGCTCGGTGCAGGCTGGAGGGTC TCTGAGACTCTCCTGTACAGCCTCTGAATATACTTATAGTGACCTCTGCATGGGCTGGTACC GCCAGGCTCCAGGGCAGGAGCGTGAGGGGGTCGCAGCTATTAGCCGTGCTGGTACTAGC ACATACTACGTCGACTCCGTGAAGGGCCGATTCACCATCTCCCAGGACAACGCCAAGAAC ACGGTGTATCTGCAAATGAACAGCCTGAAACCTGAGGACACGGCCATCTATTACTGTGCA GCAGATGAGGGGCAGGGGTGTGACGCATACCCAAGCGACTATATTCGGATGGCCGGCA ATGGGTATAACTACTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCA**ggcgcggccgc**	*Camelus dromedarius*

Supplementary table 4. Control POI sequences (continuation).
Unlike CutA and PIH1D1-N, SUMO-RMM full fragment sequences are provided with homology regions in turquoise.

Case	Sequence	Origin
CutA	catATGGAAGAGGTCGTGCTGATCACGGTGCCGAGCGAGGAGGTGGCGCGTACCATCGC CAAGGCCCTGGTGGAGGAGCGCTTGGCCGCCTGCGTGAACATCGTCCCCGGCCTGACCTC CATCTACCGCTGGCAGGGGGAGGTGGTGGAAGACCAGGAGCTGCTGTTGCTGGTCAAGA CCACCACCCACGCCTTCCCTAAGCTGAAGGAACGTGTCAAGGCCCTCCACCCCTACACCGT GCCCGAGATCGTGGCCCTGCCCATCGCCGAGGGGAACCGTGAGTACCTGGACTGGCTTCG TGAGAACACGGGAggcgcggccgc	*Thermus thermophiles HB8*
PIH1D1-N	catATGGCGGCGCATAGCGCGGCGCTGGAAGTGCTGTTTCAAGGCCCGGGTCAGCCGGGC TTTTGCATTAAAACCAACAGCAGCGAAGGCAAAGTGTTTATTAACATTTGCCATAGCCCG AGCATTCCGCCGCCGGCGGATGTGACCGAAGAAGAACTGCTGCAGATGCTGGAAGAAG ATCAAGCGGGCTTTCGCATTCCGATGAGCCTGGGCGAACCGCATGCGGAACTGGATGCG AAAGGCCAAGGCTGCACCGCGTATGATGTGGCGGTGAATAGTGATTTTATCGCCGCATG CAGAATAGCGATTTTCTGCGCGAACTGGTGATTACCATTGCGCGCGAAGGCCTGGAAGAT AAATATAACCTGCAGCTGAACCCGGAATGGCGCATGATGAAAAACCGCCCGTTTATGGG CAGCATTggcgcggccgc	*Homo sapiens*
SUMO-RMM-Lk30-GFP10	taagaaggagatatacatATGGGCGATTCAGAAGTGAACCAGGAGGCGAAACCAGAAGTTA AGCCGGAGGTGAAGCCGGAGACCCACATCAATCTAAAAGTAAGCGACGGCTCGTCGGA GATTTTCTTTAAGATTAAGAAAACAACCCCCCTGCGGCGTCTTATGGAGGCGTTTGCGAAG CGCCAAGGCAAGGAAATGGACTCACTTCGTTTCCTGTACGATGGTATTCGGATTCAGGCC GACCAGACACCGGAGGATTTGGATATGGAGGATAATGATATCATCGAGGCGCATCGTGA GCAGATTGGATCCATGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGC CGCACTGGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCATCACCGACCGGCA GCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGGGTGAGGTTGGGGCCGTAAAAG CTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTAGCGTGCATG TTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGC CGCATCAGAAGGAGGCGGTAGCGGGGGCCCTGGTTCGGGAGGGGAAGGTTCTGCTGGG GGAGGGAGCGCTGGCGGGGGGTCTGATTTACCAGACGATCATTACCTGAGCACACAAAC GATCCTTTCGAAAGACCTGAACGCAAGCTGATAAGgatcAATTGttatagaaggagata	
SUMO-RMM-Lk27-GFP11	ttatagaaggagataCAATTGTTTAAGAAGGAGATATACCATGGGCGATTCAGAAGTGAACCA GGAGGCGAAACCAGAAGTTAAGCCGGAGGTGAAGCCGGAGACCCACATCAATCTAAAA GTAAGCGACGGCTCGTCGGAGATTTTCTTTAAGATTAAGAAAACAACCCCCCTGCGGCGT CTTATGGAGGCGTTTGCGAAGCGCCAAGGCAAGGAAATGGACTCACTTCGTTTCCTGTAC GATGGTATTCGGATTCAGGCCGACCAGACACCGGAGGATTTGGATATGGAGGATAATGA TATCATCGAGGCGCATCGTGAGCAGATTGGATCCATGAGTAGTAACGCGATTGGTTTAATT GAAACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTATGGTAAAAGCTGCAAA TGTGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGG GTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTT GGCGAGCTGGTTAGCGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTA GCGTTAGCTCAAAAGGATCCGCAGGCAGCGGTGGAAGTCCGGGTGGCGGTTCAGGCGGT AGCGGCAGCTCTGCGAGCGGCGGCAGCACCAGCGAAAAACGCGATCACATGGTGCTGCT GGAATATGTGACCGCGGCGGGCATTACCGATGCGAGCTAATGACAAGTATGTCGACtccta ggaaagctcc	
SUMO-RMM-Lk1-GFP10	taagaaggagatataCCATGGGCGATTCAGAAGTGAACCAGGAGGCGAAACCAGAAGTTAA GCCGGAGGTGAAGCCGGAGACCCACATCAATCTAAAAGTAAGCGACGGCTCGTCGGAG ATTTTCTTTAAGATTAAGAAAACAACCCCCCTGCGGCGTCTTATGGAGGCGTTTGCGAAGC GCCAAGGCAAGGAAATGGACTCACTTCGTTTCCTGTACGATGGTATTCGGATTCAGGCCG ACCAGACACCGGAGGATTTGGATATGGAGGATAATGATATCATCGAGGCGCATCGTGAG CAGATTGGtTCCaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGATACGTCGCC GCACTGGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCATCACCGACCGGCA GCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGGGTGAGGTTGGGGCCGTAAAAG CTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTAGCGTGCATG TTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGATTTA CCAGACGATCATTACCTGAGCACACAAACGATCCTTTCGAAAGACCTGAACGCAAGCTG ATAAGgatcaattgttatagaaggagata	
SUMO-RMM-Lk1-GFP11	ttatagaaggagatacaattgtttaagaaggagatatagCCATGGGCGATTCAGAAGTGAACCAGG AGGCGAAACCAGAAGTTAAGCCGGAGGTGAAGCCGGAGACCCACATCAATCTAAAAGT AAGCGACGGCTCGTCGGAGATTTTCTTTAAGATTAAGAAAACAACCCCCCTGCGGCGTCTT ATGGAGGCGTTTGCGAAGCGCCAAGGCAAGGAAATGGACTCACTTCGTTTCCTGTACGAT GGTATTCGGATTCAGGCCGACCAGACACCGGAGGATTTGGATATGGAGGATAATGATAT CATCGAGGCGCATCGTGAGCAGATTGGATCCaTGAGTAGTAACGCGATTGGTTTAATTGA AACGAAAGGATACGTCGCCGCACTGGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATG TGACCATCACCGACCGGCAGCAGGTTGGCGATGGCTTAGTGGCAGTGATCGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGG CGAGCTGGTTAGCGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCACATTTTAGC GTTAGCTCAAAAGGTGAAAAACGCGATCACATGGTGCTgctggaatatgtgaccgcggcgggcatt accgatgcgagctaatgacaagtatgtcgactcctaggaaagctcc	

Supplementary table 5. BMC-H Duo sequences.

DNA fragments were ordered as POI1-GFP10 or POI2-GFP11. Homology regions with the open vector or adjacent fragment (in turquoise) permitted assembly by Gibson. Duo1-1 and Duo1-2 are given as examples of GFP10/11-tagged fragment organization. All subsequent POI sequences are only given between NdeI (in purple) and NotI sites (in blue). Of note, the DuoX-1 is coded along the GFP10 while the DuoX-2 is with the GFP11. ORFs are in bold letter, the GFP10 sequence in light green and GFP11 in dark green. MunI site is in dark yellow and SalI in red.

Case	Sequence
Duo1-1	agatttAAAtactttaagaaggagatatacataTGAGTAGTAACGCGATTGGTgctATTcagACGAAAGGAaccgggGC CGCAatcGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCctgACCagcgctgataccacaGGCGATGG CaatGTGgtcGTGtacGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTG CGTCGCAGgacGGCGAGCTGGTTgcgGTGtatGTTaccCCACGTCCCCATTCGGAACTCGGCGCAaaacgtAGCG TTAGCTCAAAAGGTGCGGCCGCATCAGAAGGAGGCGGTAGCGGGGGCCCTGGTTCGGGAGGGGAAGGT TCTGCTGGGGGAGGGAGCGCTGGCGGGGGGTCTGATTTACCAGACGATCATTACCTGAGCACACAAACG ATCCTTTCGAAAGACCTGAACGCAAGCTGATAAggatcaattgtttaa
Duo1-2	taaggatcaattgtttaagaaggagatatacataTGAGTAGTAACGCGATTGGTatcATTattACGAAAGGAaccGTCG CCGCAgatGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCgacACCatcatagataccacaGGCGATG GCaatGTGttgGTGctgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACT GCGTCGCAGGTTGGCGAGCTGttaatcGTGattGTTctgCCACGTCCCCATTCGGAACTCGGCGCAgtgTTTAGC GTTAGCTCAAAAGGTGCGGCCGCAGGCAGCGGTGGCAGCCCGGGCGGCGGCAGCGGCGGCAGCGGCAG CAGCGCGAGCGGCGGCAGCACCAGCGAAAAACGCGATCACATGGTGCTGCTGGAATATGTGACCGCGGC GGGCATTACCGATGCGAGCTAATGAcaagtatgtcgactcctaggaaagctttt
Duo2-1	cataTGAGTAGTAACGCGATTGGTgggATTcagACGAAAGGAttcgctGCCGCACTGGCTGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGACCgtgACCgcggtagtgaccacaGGCGATGGCgagGTGaagGTGtacGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttaggtGTG ggcGTTATCCCACGTCCCCATTCGGAACTCGGCGCAattcgtAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo2-2	cataTGAGTAGTAACGCGATTGGTgtaATTaccACGAAAGGAttcatcGCCGCAgatGCTGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGACCccgACCGACcttgtgaccacaGGCGATGGCgagGTGttgGTGctgGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGttaGGCGAGCTGttaacgGTGgt gGTTctgCCACGTCCCCATTCGGAACTCGGCGCAgcgTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo3-1	cataTGAGTAGTAACGCGaagGGTgctATTcagACGAAAGGAtgggggGCCGCAattatcGCTGCAGATGCTATGat cAAAGCTGCAAATGTGACCctgACCagcgctaaaaccacaGGCggcGGCaatGTGGCAGTGtacGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGttaGGCGAGCTGGTTgcgGTGt atGTTttcCCACGTCCCggcTCGaaccacGGCGCAaaacgtAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo3-2	cataTGAGTAGTAACGCGATTGGTatcATTattACGAAAGGAtggGTCGCCGCAgatGCTGCTGCAGATGCTATGg agAAAGCTGCAAATGTGACCgacACCGACataaaaaccacaGGCggcGGCaatGTGttgGTGctgGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttaatcGTG ggcGTTgtgCCACGTCCCtggTCGGAACTCGGCGCAgtgTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo4-1	cataTGAGTAGTAACGCGATTGGTgtaATTattACGAAAGGAttcGTCGCCGCAgtgGCTGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGgtgctgACCagcgtatataacacaGGCGATGGCcaaGTGttgGTGctgGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttattcGTGat tGTTttcCCAcacCCCCATgaggatCTCGGCGCAggcgcggacatcAGCTCAAAAGGTGCGGCCGC
Duo4-2	cataTGAGTAGTAACGCGATTGGTTTAATTgcgACGAAAGGAttcgggGCCGCACTGGCTGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGgtgggcACCccgctttataacacaGGCGATGGCcaaGTGgtcGTGttcGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGgacGGCGAGCTGGTTgcgGTGt atGTTctgCCAcacCCCCATgaggatCTCGGCGCAgtgctggacatcAGCTCAAAAGGTGCGGCCGC
Duo5-1	cataTGAGTAGTAACGCGgtgGGTgggATTcagACGAAAGGAgcggggGCCGCAatcggcGCTGCAGATGCTATGttg AAAGCTGCAAATGTGgtgctgACCagcgctgaagtgacaGGCgcgGGCgagGTGgtcGTGtacGTAACGGGTGAGGTT GGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGgggGGCGAGCTGttagcgGTGgcgGT TATCCCAcacCCCctggagattttcGGCGCAaaccgtgacatcAGCTCAAAAGGTGCGGCCGC
Duo5-2	cataTGAGTAGTAACGCGctgGGTatcATTattACGAAAGGAgcgGTCGCCGCAgatgtcGCTGCAGATGCTATGgg gAAAGCTGCAAATGTGgtgggcACCagcactgaagtgacaGGCgcgGGCgagGTGttgGTGctgGTAACGGGTGAGGT TGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttaaatGTGctg GTTttcCCAcacCCCCATgagcagCTCGGCGCAgtgTTTgacatcAGCTCAAAAGGTGCGGCCGC

Supplementary table 5. BMC-H Duo sequences (continuation).

Case	Sequence
Duo6-1	cataTGAGTAGTAACGCGATTGGTTTAATTagcACGAAAGGAttcgggGCCGCACTGGCTGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGACCctgACCagcggatttaacacaGGCGATGGCaatGTGGCAGTGttcGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGgccGGCGAGCTGttagcgGTG CATGTTctgCCACGTCCCCATTCGGAACTCGGCGCAaaactgAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo6-2	cataTGAGTAGTAACGCGATTGGTgtaATTgtgACGAAAGGAttcactGCCGCAaccGCTGCTGCAGATGCTATGGT AAAAGCTGCAAATGTGACCATCACCagcgtatttaacacaGGCGATGGCaatGTGttgGTGctgGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGtacttaGTGa ttGTTatgCCACGTCCCCATTCGGAACTCGGCGCAattTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo7-1	cataTGAGTAGTAACGCGATTGGTgggATTGAAACGAAAGGAgcggggGCCGCAatcGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCctgACCGACataaccaacacaGGCGATGGCatgGTGGCAGTGtacGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGgccGGCGAGCTGatcgcgG TGgcgGTTttcCCACGTCCCCATTCGGAACTCGGCGCAacccgtAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo7-2	cataTGAGTAGTAACGCGATTGGTatcATTaccACGAAAGGAgcgGTCGCCGCAgatGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCccgACCgcgactaccaacacaGGCGATGGCatgGTGttgGTGctgGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGatcaatGT GattGTTATCCCACGTCCCCATTCGGAACTCGGCGCAaaaTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo8-1	cataTGAGTAGTAACGCGgtgGGTgggATTGAAACGAAAGGAgcggggGCCGCAatcGCTGCTGCAGATGCTATGtt gAAAGCTGCAAATGTGACCctgACCGACataaccaacacaGGCggcGGCatgGTGGCAGTGtacGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGgccGGCGAGCTGatcgcgGTG gcgGTTATCCCACGTCCCgcgTCGattttcGGCGCAacccgtAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo8-2	cataTGAGTAGTAACGCGATTGGTatcATTgtgACGAAAGGAgcgGTCGCCGCAgatgtcGCTGCAGATGCTATGGT AAAAGCTGCAAATGTGACCccgACCgcgactaccaacacaGGCggcGGCatgGTGggtGTGctgGTAACGGGTGAGG TTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGatcaatGTGat tGTTATCCCACGTCCCCATTCGGAACTCGGCGCAaaaTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo9-1	cataTGAGTAGTAACGCGATTGGTTTAATTaccACGAAAGGAaccgggGCCGCACTGGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCgtgACCagcataaaaagcagcGGCGATGGCaatGTGacgGTGttcGTAACGGGTG AGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGatcGGCGAGCTGttagcgGT GctgGTTATCCCACGTCCCCATTCGgatCTCGGCGCAgttctgAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo9-2	cataTGAGTAGTAACGCGATTGGTgtaATTgtgACGAAAGGAaccactGCCGCAgtgGCTGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGACCctgACCagctacaaaagcagcGGCGATGGCaatGTGttgGTGaccGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGgccGGCGAGCTGatcgttGTGt atGTTaacCCACGTCCCCATTCGgatCTCGGCGCAaaagcgAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo10-1	cataTGAGTAGTAACGCGgtgGGTTTAATTaccACGAAAGGAattgggGCCGCACTGgacGCTGCAGATGCTATGtt gAAAGCTGCAAATGTGACCATCACCagcatactgagctgtGGCggcGGCatgtgtacgGTGttcGTAACGGGTGAGGT TGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGatcGGCGAGCTGttagcgGTGctgG TTctgCCACGTCCCagcTCGaccCTCGGCGCAgttctgAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo10-2	cataTGAGTAGTAACGCGATTGGTgtaATTgtgACGAAAGGAattactGCCGCAgtgGCTGCTGCAGATGCTATGac tAAAGCTGCAAATGTGACCctgACCagcttcctgagctgtGGCggcGGCatgtgtttgGTGaccGTAACGGGTGAGGTT GGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGgccGGCGAGCTGttagttGTGagcGT TcgtCCACGTCCCattTCGcagCTCGGCGCAaaagcgAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo11-1	cataTGAGTAGTAACGCGATTGGTatcATTattACGAAAGGAttcGTCGCCGCAcatGCTGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGACCggcACCatcataatgacctgtGGCGATGGCaatGTGttgGTGATCGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGgcctacGTGg gcGTTctgCCACGTCCCCATTCGgatCTCGGCGCAgttctgAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo11-2	cataTGAGTAGTAACGCGATTGGTgctATTgcgACGAAAGGAttcgggGCCGCAatcGCTGCTGCAGATGCTATGGT AAAAGCTGCAAATGTGACCctgACCgcgttcatgacctgtGGCGATGGCaatGTGgtcGTGtacGTAACGGGTGAGG TTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGtggGGCGAGCTGGTTgcgGTGtat GTTaccCCACGTCCCCATTCGgatCTCGGCGCAtttgatAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo12-1	cataTGAGTAGTAACGCGgtgGGTTTAATTaccACGAAAGGAttcgggGCCGCACTGggcGCTGCAGATGCTATGGT AAAAGCTGCAAATGTGACCATCACCagcataaaaaaacacaGGCaacGGCaatGTGacgGTGttcGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGgccGGCGAGCTGttagcgGTG gcgGTTctgCCACGTCCCCATTCGgatCTCGGCGCAgttctgAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo12-2	cataTGAGTAGTAACGCGATTGGTgtaATTgtgACGAAAGGAttcactGCCGCAgcggtcGCTGCAGATGCTATGact AAAGCTGCAAATGTGACCctgACCagcgctaaaaacacaGGCaacGGCaatGTGGCAGTGctgGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttaatcGTGc tgGTTttcCCACGTCCCCATTCGcagCTCGGCGCAaaagcgAGCGTTAGCTCAAAAGGTGCGGCCGC

Supplementary table 5. BMC-H Duo sequences (continuation).

Case	Sequence
Duo13-1	cataTGAGTAGTAACGCGATTGGTTTAATTaccACGAAAGGAaccgggGCCGCACTGGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCgtgACCgtgataaccaaatgtGGCGATGGCtcaGTGaatGTGttcGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGatcGGCGAGCTGgccgcgGTG ctgGTTATCCCACGTCCCCATTCGGAACTCGGCGCAattctgAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo13-2	cataTGAGTAGTAACGCGATTGGTgtaATTattACGAAAGGAaccactGCCGCAaccGCTGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGACCctgACCagcataaccaaatgtGGCGATGGCtcaGTGttgGTGctgGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGgccGGCGAGCTGatcgttGTGta tGTTagcCCACGTCCCCATTCGGAACTCGGCGCAgcggcgAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo14-1	cataTGAGTAGTAACGCGATTGGTgggATTcagACGAAAGGAggcgggGCCGCAatcGCTGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGACCgtgACCggccttcgtaccacaGGCGATGGCgagGTggtcGTGtacGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttagcgGTGg gcGTTtacCCACGTCCCCATTCGGAACTCGGCGCActgcgtAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo14-2	cataTGAGTAGTAACGCGATTGGTatcATTattACGAAAGGAggcGTCGCCGCAgatGCTGCTGCAGATGCTATGG TAAAAGCTGCAAATGTGACCggcACCagccttcgtaccacaGGCGATGGCgagGTgttgGTGctgGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGttaGGCGAGCTGttagttGTGat tGTTctgCCACGTCCCCATTCGGAACTCGGCGCAgcgTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo15-1	cataTGAGTAGTAACGCGATTGGTatcATTGAAACGAAAGGAgtgactGCCGCAaccGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCcaaACCGACttccgtagcgacGGCGATGGCtcaGTGttgGTGctgGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGaagGGCGAGCTGttaatcGTG cgtGTTATCCCACGTCCCCATTCGGAACTCGGCGCAgcgTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo15-2	cataTGAGTAGTAACGCGATTGGTcaaATTaaaACGAAAGGAatggctGCCGCAatcGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCgagACCgcggtacgtagcgacGGCGATGGCtcaGTGGCAGTGttcGTAACGGGTG AGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGgggGGCGAGCTGcaagagGT GaccGTTgacCCACGTCCCCATTCGGAACTCGGCGCAaaatggAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo16-1	cataTGAGTAGTAACGCGATTGGTatcATTGAAACGAAAGGAgacactGCCGCAaacGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCATCACCggcCGGCAGCAGagcGGCGATGGCcaaGTGacgGTGctgGTAACGGG TGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGatgacg GTGaccGTTaaaCCACGTCCCCATTCGGAACTCGGCGCAgtggtgAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo16-2	cataTGAGTAGTAACGCGATTGGTcaaATTGAAACGAAAGGAtgggctGCCGCAatcGCTGCTGCAGATGCTATG GTAAAAGCTGCAAATGTGACCATCACCaatcttCAGCAGagcGGCGATGGCcaaGTGcgcGTGaatGTAACGGGT GAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttacaa GTGcagGTTgtgCCACGTCCCCATTCGGAACTCGGCGCAgcgTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo17-1	cataTGAGTAGTAACGCGgtgGGTatcATTGAAACGAAAGGAaacGTCGCCGCAgatttcGCTGCAGATGCTATGtt gAAAGCTGCAAATGTGACCATCACCGACgtaCAGcgtagcGGCaacGGCtcagacacgGTGATCGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGcgcGGCGAGCTGttaacgGTG accGTTccgCCACGTCCCtggTCGgcgataGGCGCAtgcTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo17-2	cataTGAGTAGTAACGCGgtgGGTcgcATTGAAACGAAAGGAttcgctGCCGCAatgatgGCTGCAGATGCTATGta cAAAGCTGCAAATGTGACCagcACCagcgtaCAGcgtagcGGCaacGGCtcaGTGacgGTGgtgGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttctgtGTGgg cGTTgagCCACGTCCCCATTCGaaaCTCGGCGCAaaacgtAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo18-1	cataTGAGTAGTAACGCGgtgGGTgggATTcagACGctgGGAgaggggGCCGCAaccGCTGCTGCAGATGCTATGGT AcagGCTGCAAATGTGaagaccACCGACatgaaagataatGGCaacGGCcacGTGacgGTGATCGTAACGGGTGA GGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGgccGGCGAGCTGatgcaaGTG gcgGTTATCCCACGTCCCaacatggatCTCGGCGCAtatTTTgccgcgAGCTCAAAAGGTGCGGCCGC
Duo18-2	cataTGAGTAGTAACGCGATTGGTatcATTGAAACGAAAGGAttcGTCGCCGCAatgtgtGCTGCAGATGCTATGG TAgatGCTGCAAATGTGaagctgACCgcggtaaaagataatGGCaacGGCcacGTGttgGTGcgtGTAACGGGTGAG GTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGcgcgagGTGc tgGTTATCCCACGTCCCctggactatCTCGGCGCAcgtcagggtatcAGCTCAAAAGGTGCGGCCGC
Duo19-1	cataTGAGTAGTAACGCGATTGGTgggATTgcgACGatgcggaacactGCCGCACTGaagGCTttaaacGCTATGGTA gcgGCTGCAAATGTGACCATCACCagcatagatcgtgacGGCGATagcgggagcacgGTGtggGTAACGGGTGAGGTT GGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGcgcggtGTGggcG TTctgCCACGTCCCaacgaggcgCTCGGCGCAtttTTTaagGTTAGCTCAAAAGGTGCGGCCGC
Duo19-2	cataTGAGTAGTAACGCGATTGGTgctATTaccACGagcGGAgcgGTCGCCGCAgtgatgGCTggcGATGCTATGGT AaccGCTGCAAATGTGACCatgACCaactgggatcgtgacGGCGATagcgggGTGacgGTGctgGTAACGGGTGAGGT TGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGatgGGCGAGCTGcaagagGTGttt GTTgagCCACGTCCCaccTCGaccCTCGGCGCAgcggcgtggGTTAGCTCAAAAGGTGCGGCCGC

Supplementary table 5. BMC-H Duo sequences (continuation).

Case	Sequence
Duo20-1	cataTGAGTAGTAACGCGatgGGTgggATTGAAACGctgactttcgctGCCGCAattatgGCTgaggcgGCTATGGTAgcgGCTGCAAATGTGgtgATCACCgcggtactgaaccaaGGCGATgcggacGTGaagGTGtggGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGaagcgcGTGgatGTTgagCCACGTCCCgataatgatCTCGGCGCAgtgTTTgacttcAGCTCAAAAGGTGCGGCCGC
Duo20-2	cataTGAGTAGTAACGCGgtgGGTatcATTGAAACGAAAGGAgagGTCGCCGCAgataagGCTGCAccgGCTATGGTAcgtGCTGCAAATGTGctgttcACCgcgaagctgaaccaaGGCGATgcggacacatgtGTGATCGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGcgcacgGTGaccGTTgcgCCACGTCCCattTCGccgataGGCGCAaaagtgtgtatcAGCTCAAAAGGTGCGGCCGC
Duo21-1	cataTGAGTAGTAACGCGcgtGGTgggATTcagACGtatagtgtgagcGCCGCACTGGCTGCTGCAaccGCTATGGTAAAAGCTGCAAATGTGACCctgACCgcggtacatgatagcGGCaacgcggagcacacgGTGgcgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGcgccaaGTGgcgGTTaacCCACGTCCCCATactctggtaGGCGCAaacattcgcctgAGCTCAAAAGGTGCGGCCGC
Duo21-2	cataTGAGTAGTAACGCGATTGGTagtATTaccACGtggGGAttcGTCGCCGCAtacgagGCTgagGATGCTATGGTAcgtGCTGCAAATGTGgcgccgACCgcgcttcatgatagcGGCaacgcggagcaatgtGTGctgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGatcacgGTGgtgGTTgagCCACGTCCCgcgTCGagcCTCGGCGCAgcgTTTgactacAGCTCAAAAGGTGCGGCCGC
Duo22-1	cataTGAGTAGTAACGCGgtgGGTgggATTtttACGAAAGGATACagcGCCGCACTGggcGCTGCAGATGCTATGgtAAAGCTGCAAATGTGACCATCACCGACCGGCAGCAGgacGGCgaaGGCTTAGTGtctGTGaaaGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttacaaGTGggcGTTaacCCACGTCCCgatTCGcgtaatGGCGCAgtTctgAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo22-2	cataTGAGTAGTAACGCGATTGGTgtaATTGAAACGAAAGGAgacttgGCCGCAgaatggGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCccgACCagctacCAGCAGgacGGCgaaGGCTTAGTGacgGTGctgGTAACGGGTGAGGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttagagGTGtttGTTATCCCACGTCCCgatTCGgtgataGGCGCAcgtTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo23-1	cataTGAGTAGTAACGCGATTGGTatcATTGAAACGAAAGGAattGTCGCCGCAatCGCTGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCATCACCgcgactcgtaacgacGGCGATGGCcggGTGttgGTGcgtGTAACGGGTGAGGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttaaatGTGggcGTTATCCCACGTCCCgatTCGaccCTCGGCGCActgctgtcCGTTAGCTCAAAAGGTGCGGCCGC
Duo23-2	cataTGAGTAGTAACGCGgtgGGTatgATTcagACGAAAGGAgaggggGCCGCAgtggtcGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCctgACCcacgtacgtaacgacGGCGATGGCcggGTGacgGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGtggGGCGAGCTGttacaaGTGCATGTTATCCCACGTCCCCATTCGgatCTCGGCGCAacctggAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo24-1	cataTGAGTAGTAACGCGATTGGTTTAATTaccACGAAAGGAgcgGTCGCCGCAatgttcGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCccgACCagcCGGCAGagcacaGGCGATGGCatggacacgGTGttcGTAACGGGTGAGGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGatcGGCGAGCTGttagagGTGgcgGTTaacCCACGTCCCaacTCcagcCTCGGCGCAcgttggAGCGTTAGCTCAAAAGGTGCGGCCGC
Duo24-2	cataTGAGTAGTAACGCGtgtGGTgctATTcagACGAAAGGAccgactGCCGCAgtgatgGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCctgACCGACgtaCAGagcacaGGCGATGGCatgGTggtcGTGATCGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGatcgagGTGggcGTTctgCCACGTCCCaacTCGaaatcaGGCGCAatttggAGCGTTAGCTCAAAAGGTGCGGCCGC

Supplementary table 6. BMC-H Trio sequences.

DNA fragments were ordered as POI1-GFP10, POI2-GFP11 or POI3-Flag. Homology regions with the open vector or adjacent fragment (in turquoise) permitted assembly by Gibson. Trio1-1, 1-2 and 1-3 are given as examples of GFP10/11-tagged or Flag-tagged fragment organization. All subsequent POI sequences are only given between NdeI (in purple) and NotI sites (in blue) or MunI (in dark yellow) or SalI (in red). Of note, the TrioX-1 is coded along the GFP10 (light green) while the TrioX-2 is with the GFP11 (dark green) and TrioX-3 is with the Flag (light purple).

Case	Sequence
Trio1-1	agatttAAAtactttaagaaggagatatacataTGAGTAGTAACGCGATTGGTgacATTaccACG AAAGGAttcgggGCCGCACTGGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGA CCctgACCagcgcttataccacaGGCGATGGCaatGTGacgGTGgtgGTAACGGGTGAGGTTG GGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGtacGGCGAGC TGttagcgGTGCATGTTATCCCACGTCCCCATTCGGAACTCGGCGCActggatAGCGTTAGC TCAAAAGGTGCGGCCGCATCAGAAGGAGGCGGTAGCGGGGGCCCTGGTTCGGGAGG GGAAGGTTCTGCTGGGGGAGGGAGCGCTGGCGGGGGGTCTGATTTACCAGACGATC ATTACCTGAGCACACAAACGATCCTTTCGAAAGACCTGAACGCAAGCTGATAAggatc aattgtttaa
Trio1-2	taaggatcaattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTgtaATTattACG AAAGGAttcGTCGCCGCAgtgGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGA CCATCACCagcatatataccacaGGCGATGGCaatGTGggtGTGctgGTAACGGGTGAGGTTG GGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAG CTGttaatcGTGattGTTctgCCACGTCCCCATTCGGAACTCGGCGCAgtgTTTAGCGTTAGC TCAAAAGGTGCGGCCGCAGGCAGCGGTGGCAGCCCGGGCGGCGGCAGCGGCGGCA GCGGCAGCAGCGCGAGCGGCGGCAGCACCAGCGAAAAACGCGATCACATGGTGCT GCTGGAATATGTGACCGCGGCGGGCATTACCGATGCGAGCTAATGACaagtatggatcc
Trio1-3	GCGAGCTAATGACaagtatggatcccAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGG TTTAATTGAAACGAAAGGAttcactGCCGCAcgtGCTGCTGCAGATGCTATGGTAAAAG CTGCAAATGTGACCaagACCagcCGGtataccacaGGCGATGGCaatGTGGCAGTGttcGTA ACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTC GCAGGTTGGCGAGCTGttagcgGTGtatGTTctgCCACGTCCCCATTCGGAACTCGGCGCA aaagtgAGCGTTAGCTCAAAAGGTTCTAGCGAAAATCTGTACTTCCAGAGTAGTGCggcc GCAGATTACAAAGATGACGATGATAAGTGAgtcgactcctaggaaagctttt
Trio2-1	cataTGAGTAGTAACGCGATTGGTgacATTaccACGAAAGGAttcGTCGCCGCAgatGCTGC TGCAGATGCTATGGTAAAAGCTGCAAATGTGACCggcACCccgatatataccacaGGCGATG GCatgGTGgagGTGgagGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAG GCGCTGAAACTGCGTCGCAGgccGGCGAGCTGatcgcgGTGCATGTTgacCCACGTCCCCA TTCGGAACTCGGCGCAgtggatAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio2-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTtggATTcagACGAAAGGAt tcgggGCCGCAatCGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCctgACCagc gcttataccacaGGCGATGGCatgGTggtcGTGtacGTAACGGGTGAGGTTGGGGCCGTAAA AGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGttaGGCGAGCTGgccgcgGTGctg GTTaccCCACGTCCCCATTCGGAACTCGGCGCAaaacgtAGCGTTAGCTCAAAAGGTGCGG CCGC
Trio2-3	ggatcccAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTatcATTattACGAAAGGAt tcactGCCGCAcgtGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCaagACCagc CGGtataccacaGGCGATGGCatgGTGttgGTGctgGTAACGGGTGAGGTTGGGGCCGTAA AAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttaatcGTGg gcGTTATCCCACGTCCCCATTCGGAACTCGGCGCAgtgTTTAGCGTTAGCTCAAAAGGTTCT AGCGAAAATCTGTACTTCCAGAGTAGTGCggccGC

Supplementary table 6. BMC-H Trio sequences (continuation).

Case	Sequence
Trio3-1	cataTGAGTAGTAACGCGATTGGTgacATTaccACGAAAGGATACGTCGCCGCAgatGCTGC TGCAGATGCTATGGTAAAAGCTGCAAATGTGACCggcACCatcatagaaaccGTTGGCGATG GCcaaGTGgagGTGgagGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCA GGCGCTGAAACTGCGTCGCAGgccGGCGAGCTGatcgcgGTGCATGTTgacCCACGTCCCC ATTCGGAACTCGGCGCAgtggatAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio3-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTtggATTcagACGAAAGGA TACgggGCCGCAatCGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCctgACCaa ggctgaaaccGTTGGCGATGGCcaaGTGgtcGTGtacGTAACGGGTGAGGTTGGGGCCGTA AAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGttaGGCGAGCTGgccccaGTG ctgGTTaccCCACGTCCCCATTCGGAACTCGGCGCAaaacgtAGCGTTAGCTCAAAAGGTGC GGCCGC
Trio3-3	ggatccCAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTatcATTattACGAAAGGAT ACactGCCGCAcgtGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCaagACCagc CGGgaaaccGTTGGCGATGGCcaaGTGttgGTGctgGTAACGGGTGAGGTTGGGGCCGTA AAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttaatcGTG ggcGTTATCCCACGTCCCCATTCGGAACTCGGCGCAgtgTTTAGCGTTAGCTCAAAAGGTTC TAGCGAAAATCTGTACTTCCAGAGTAGTGCggccGC
Trio4-1	cataTGAGTAGTAACGCGATTGGTgacATTaccACGAAAGGAgacGTCGCCGCAgatGCTGC TGCAGATGCTATGGTAAAAGCTGCAAATGTGACCggcACCatcatagataaaggtGGCGATG GCttcGTGgagGTGgagGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAG GCGCTGAAACTGCGTCGCAGgccGGCGAGCTGGTTgcgGTGCATGTTgacCCACGTCCCCA TTCGGAACTCGGCGCAgtggatAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio4-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTtggATTcagACGAAAGGA gacgggGCCGCAatCGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCctgACCca agctgataaaggtGGCGATGGCttcGTGgtcGTGtacGTAACGGGTGAGGTTGGGGCCGTAA AAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGttaGGCGAGCTGgccccaGTGct gGTTaccCCACGTCCCCATTCGGAACTCGGCGCAaaacgtAGCGTTAGCTCAAAAGGTGCG GCCGC
Trio4-3	ggatccCAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTatcATTattACGAAAGGAg acactGCCGCAcgtGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCaagACCagc CGGgataaaggtGGCGATGGCttcGTGttgGTGctgGTAACGGGTGAGGTTGGGGCCGTAAA AGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttaatcGTGgg cGTTATCCCACGTCCCCATTCGGAACTCGGCGCAgtgTTTAGCGTTAGCTCAAAAGGTTCTA GCGAAAATCTGTACTTCCAGAGTAGTGCggccGC
Trio5-1	cataTGAGTAGTAACGCGATTGGTagtATTaccACGAAAGGAattatcGCCGCAgtgGCTGCT GCAGATGCTATGGTAAAAGCTGCAAATGTGACCctgACCGACcttcgtgaaagcGGCGATGG CTTAGTGaagGTGtacGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAG GCGCTGAAACTGCGTCGCAGttaGGCGAGCTGGTTAGCGTGgtgGTTaccCCACGTCCCCA TTCGGAACTCGGCGCAaaccgtAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio5-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTgggATTGAAACGAAAGGA attgggGCCGCACTGGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCgtgACCgc ggtacgtgaaagcGGCGATGGCTTAGTGGCAGTGttcGTAACGGGTGAGGTTGGGGCCGTA AAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTggtGTG ggcGTTATCCCACGTCCCCATTCGGAACTCGGCGCAatCgtgAGCGTTAGCTCAAAAGGTGC GGCCGC
Trio5-3	ggatccCAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTgtaATTattACGAAAGGAa ttgacGCCGCAgatGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCggcACCGAC actcgtgaaagcGGCGATGGCTTAGTGgagGTGctgGTAACGGGTGAGGTTGGGGCCGTAA AAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGagcGGCGAGCTGGTTgttGTGtt tGTTgtgCCACGTCCCCATTCGGAACTCGGCGCAgcgTTTAGCGTTAGCTCAAAAGGTTCTA GCGAAAATCTGTACTTCCAGAGTAGTGCggccGC

Supplementary table 6. BMC-H Trio sequences (continuation).

Case	Sequence
Trio6-1	cataTGAGTAGTAACGCGATTGGTatcATTGAAACGaacGGAtgtGTCGCCGCAatcgtcGCTGCAgcgGCTATGGTAgaaGCTGCAAATGTGcaaATCACCGACgtacgtaacaataatGATaattgttgtGCAGTGttcGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttaaatGTGagcGTTcgtCCACGTCCCCAcatgGAACTCGGCGCAattTTTccaatcAGCTCAAAAGGTGCGGCCGC
Trio6-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTTTAATTGAAACGaacGGAtgtccaGCCGCAaacGCTGCTttaaacGCTATGGTAcgtGCTGCAAATGTGcaaccgACCgcgatacgtaacaataatGATaattgttgtacgGTGATCGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttagagGTGgcgGTTATCCCACGTCCCcggctaaaCTCGGCGCAaaaTTTccaatcAGCTCAAAAGGTGCGGCCGC
Trio6-3	ggatccAGAAGGAGATATAtCTaTGAGTAGTAACGCGgtgGGTgagATTaccACGaacGGAtgtgggGCCGCAgtggacGCTGCAGATGCTATGGTAgaaGCTGCAAATGTGatcctgACCgcggctcgtaacaataatGATaattgttgtacgGTGaccGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGacgGGCGAGCTGttagcgGTGattGTTATCCCACGTCCCgatatgcgtgctGGCGCAgtggcgccagcgAGCTCAAAAGGTTCTAGCGAAAATCTGTACTTCCAGAGTAGTGCggccGC
Trio7-1	cataTGAGTAGTAACGCGATTGGTatgATTcagACGAAAGGAttcgggGCCGCACTGGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCATCACCaatCGGgaaaccagcGGCGATGGCTTAGTGacgGTGATCGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGaagcgcGTGgatGTTATCCCACGTCCCCATTCGGAACTCGGCGCAacctggAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio7-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTatcATTGAAACGAAAGGAttcactGCCGCAgtgGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCATCACCaatgtagaaaccagcGGCGATGGCTTAGTGacgGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgagGTGaccGTTATCCCACGTCCCCATTCGGAACTCGGCGCAgtgTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio7-3	ggatccAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTatgATTGAAACGAAAGGAatgGTCGCCGCAatCGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCATCACCgcgactgaaaccagcGGCGATGGCTTAGTGGCAGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGcgccaaGTGgcgGTTATCCCACGTCCCCATTCGGAACTCGGCGCAaaaattAGCGTTAGCTCAAAAGGTTCTAGCGAAAATCTGTACTTCCAGAGTAGTGCggccGC
Trio8-1	cataTGAGTAGTAACGCGATTGGTgtaATTGAAACGAAAGGAattGTCGCCGCAatCGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCctgACCgcgCGGcgtagcagcGGCGATGGCgctGTgatgGTgaccGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttagagGTGaccGTTATCCCACGTCCCCATTCGGAACTCGGCGCAaaaTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio8-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTgggATTGAAACGAAAGGAttcgggGCCGCAgcgGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCctgACCagccttCAGagcagcGGCGATGGCgctGTGacgGTGctgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttagcgGTGgatGTTATCCCACGTCCCCATTCGGAACTCGGCGCActgTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio8-3	ggatccAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTatcATTggcACGAAAGGATACgctGCCGCACTGGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCATCACCaatgtaCAGagcagcGGCGATGGCgctGTGacgGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGtggGGCGAGCTGaaggagGTGaccGTTATCCCACGTCCCCATTCGGAACTCGGCGCAgtTctgAGCGTTAGCTCAAAAGGTTCTAGCGAAAATCTGTACTTCCAGAGTAGTGCggccGC

Supplementary table 6. BMC-H Trio sequences (continuation).

Case	Sequence
Trio9-1	cataTGAGTAGTAACGCGATTGGTTTAATTcagACGAAAGGAaacgggGCCGCAatCGCTGC TGCAGATGCTATGGTAAAAGCTGCAAATGTGACCATCACCGACgtagatCAGgacGGCGAT GGCtggGTGacgGTGcacGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCA GGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGatcgagGTGCATGTTccgCCACGTCCCC ATTCGGAACTCGGCGCAgtgTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio9-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTatcATTcagACGAAAGGA ttcactGCCGCAatCGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCctgACCagc atagatCAGgacGGCGATGGCtggGTGacgGTGcgtGTAACGGGTGAGGTTGGGGCCGTAA AAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGaaggagGTG gtgGTTATCCCACGTCCCCATTCGGAACTCGGCGCAcgttatAGCGTTAGCTCAAAAGGTGC GGCCGC
Trio9-3	ggatccAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTatgATTGAAACGAAAGGA TACGTCGCCGCACTGGCTGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCctgACCg cgCGGgatCAGgacGGCGATGGCtggGTGacgGTGgtgGTAACGGGTGAGGTTGGGGCCGT AAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGgccgagGT GgcgGTTATCCCACGTCCCCATTCGGAACTCGGCGCAtttattAGCGTTAGCTCAAAAGGTTC TAGCGAAAATCTGTACTTCCAGAGTAGTGCggccGC
Trio10-1	cataTGAGTAGTAACGCGgtgGGTatgATTGAAACGAAAGGATACgggGCCGCAatCgtcGCT GCAGATGCTATGatcAAAGCTGCAAATGTGACCctgACCGACgtacataacgccGGCaacGG CttcGTGacgGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGG CGCTGAAACTGCGTCGCAGGTTGGCGAGCTGtaccaaGTGtatGTTATCCCACGTCCCCAT TCGgatatgGGCGCAgcgTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio10-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGG AatttacGCCGCAgtgGCTGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCctgACCacc gtacataacGTTGGCaacGGCttcGTGatgGTGATCGTAACGGGTGAGGTTGGGGCCGTAA AAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttaacgGTGg cgGTTATCCCACGTCCCaaTTCGagcCTCGGCGCAaaacatAGCGTTAGCTCAAAAGGTGC GGCCGC
Trio10-3	ggatccAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTatgATTGAAACGAAAGGA attGTCGCCGCAtttGCTGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCATCACCcaa gtacataacgccGGCaacGGCttcatcGCAGTGttcGTAACGGGTGAGGTTGGGGCCGTAAA AGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGatcgcgGTGgg cGTTctgCCACGTCCCCATTCGggcCTCGGCGCAggctggAGCGTTAGCTCAAAAGGTTCTAG CGAAAATCTGTACTTCCAGAGTAGTGCggccGC
Trio11-1	cataTGAGTAGTAACGCGATTGGTatcATTGAAACGAAAGGAaacactGCCGCAatCgtcGC TGCAGATGCTATGttgAAAGCTGCAAATGTGACCctgACCagcCGGatgagcGTTGGCGATG GCtacGTGGCAGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCA GGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgagGTGaccGTTATCCCACGTCCC CATTCGagcCTCGGCGCAgtTctgAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio11-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGG ActgGTCGCCGCAatttacGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCATCACCaa tgtaatgagcGTTGGCGATGGCaagatcacgGTGaatGTAACGGGTGAGGTTGGGGCCGTAA AAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttagagGTGg cgGTTATCCCACGTCCCCATTCGagctcaGGCGCActgtatAGCGTTAGCTCAAAAGGTGCG GCCGC
Trio11-3	ggatccAGAAGGAGATATAtCTaTGAGTAGTAACGCGgtgGGTatgATTcagACGAAAGGAt tcgggGCCGCAgtgGCTGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCATCACCGACc ttatgagcGTTGGCGATGGCtacGTGacgGTGctgGTAACGGGTGAGGTTGGGGCCGTAAAA GCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttacaaGTGggc GTTctgCCACGTCCCCATTCGcgtCTCGGCGCAatgtggAGCGTTAGCTCAAAAGGTTCTAGC GAAAATCTGTACTTCCAGAGTAGTGCggccGC

Supplementary table 6. BMC-H Trio sequences (continuation).

Case	Sequence
Trio12-1	cataTGAGTAGTAACGCGATTGGTTTAATTGAAACGAAAGGAttcgggGCCGCAatCGCTGC TGCAGATGCTATGttgAAAGCTGCAAATGTGACCctgACCGACgtaattgaagacGGCGATGG CcacGTGacgGTGATCGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAG GCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttacaaGTGgcgGTTATCCCACGTCCCCA TTCGggcCTCGGCGCAgcgtggAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio12-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTgtaATTaccACGAAAGGA ttctgtGCCGCAatgatgGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCATCACCaatC GGgaagaacgcGGCGATGGCcacatcacgGTGATCGTAACGGGTGAGGTTGGGGCCGTAA AAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgagGTGt ttGTTATCCCACGTCCCCATTCGagcCTCGGCGCAaaaattAGCGTTAGCTCAAAAGGTGCG GCCGC
Trio12-3	ggatccCAGAAGGAGATATAtCTaTGAGTAGTAACGCGgtgGGTtgtATTcagACGAAAGGAa ccGTCGCCGCAatttgtGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCctgACCgcggt agaagaagacGGCGATGGCcacGTGacgGTGATCGTAACGGGTGAGGTTGGGGCCGTAAA AGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttagagGTGgc gGTTtgtCCACGTCCCCATTCGaacCTCGGCGCAaccTTTAGCGTTAGCTCAAAAGGTTCTAG CGAAAATCTGTACTTCCAGAGTAGTGCggccGC
Trio13-1	cataTGAGTAGTAACGCGATTGGTTTAATTcagACGgaaGGAggcGTCGCCGCACTGGCTGC TtctGATGCTATGGTAAAAGCTGCAAATGTGACCATCACCGACactCAGgatgacGGCaacgc gcacGTGacgGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGG CGCTGAAACTGCGTCGCAGatcGGCGAGCTGttagagGTGgcgGTTATCCCACGTCCCgatc ctaccCTCGGCGCAcagTTTcgcatcAGCTCAAAAGGTGCGGCCGC
Trio13-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTatgATTcatACGcagagtgt ggggGCCGCAatCGCTGCTGCAaacGCTATGGTAgaaGCTGCAAATGTGgagctgACCgcggt aCAGgatgacGGCaacgcgcaccgcacgGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAAG CTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGaaggagGTGCAT GTTATCCCACGTCCCCATcaaagcataGGCGCAagctggaatGTTAGCTCAAAAGGTGCGGC CGC
Trio13-3	ggatccCAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTatcATTaccACGAAAGGAg cgGTCGCCGCAattatgGCTGCAccgGCTATGGTAAAAGCTGCAAATGTGACCATCACCgcgg taCAGgatgacGGCaacgcgcacGTGacgGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAA GCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTgagGTGacc GTTATCCCACGTCCCCATcctcgtCTCGGCGCAgcgtatccaatcAGCTCAAAAGGTTCTAGCG AAAATCTGTACTTCCAGAGTAGTGCggccGC
Trio14-1	cataTGAGTAGTAACGCGATTGGTatgATTGAAACGAAAGGAttcgctGCCGCAatCGCTGCT ggcGATGCTATGGTAAAAGCTGCAAATGTGatcctgACCgcggctCAGaacgacGGCaacGGC atgGTGatgGTGttcGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGC GCTGAAACTGCGTCGCAGGTTGGCGAGCTGttaacgGTGggcGTTATCCCACGTCCCgatcc taacCTCGGCGCAcgtTTTaagGTTAGCTCAAAAGGTGCGGCCGC
Trio14-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTatcATTGAAACGAAAGGA aacGTCGCCGCAattttcGCTGCAGATGCTATGGTAcagGCTGCAAATGTGagcctgACCagc ataCAGaacgacGGCaacGGCatgGTGacgGTGATCGTAACGGGTGAGGTTGGGGCCGTA AAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGtggGGCGAGCTGGTTcaaGTG accGTTATCCCACGTCCCCATactagcCTCGGCGCAgtgTTTcgcGTTAGCTCAAAAGGTGCG GCCGC
Trio14-3	ggatccCAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTgctATTGAAACGAAAcaca ttactGCCGCAgtgtacGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGACCATCACCGAC gtaCAGaacgacGGCaacGGCatgGTGacgGTGgtgGTAACGGGTGAGGTTGGGGCCGTAA AAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGatgGGCGAGCTGGTTcaaGTGg gcGTTATCCCACGTCCCCATcctagcCTCGGCGCAggcgtgaatGTTAGCTCAAAAGGTTCTAG CGAAAATCTGTACTTCCAGAGTAGTGCggccGC

Supplementary table 6. BMC-H Trio sequences (continuation).

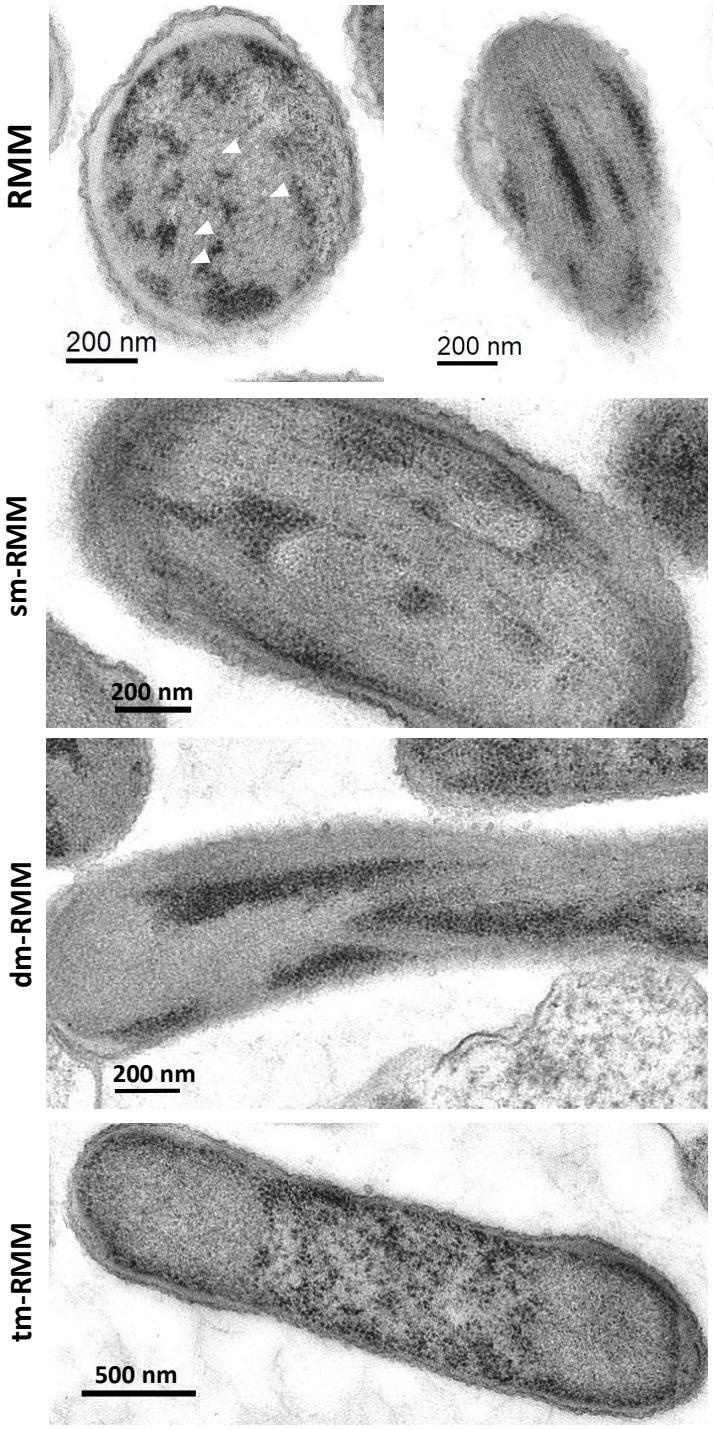
Case	Sequence
Trio15-1	cataTGAGTAGTAACGCGATTGGTTTAATTGAAACGggcgaggtggggGCCGCAatgGCTGCTtctGATGCTATGGTAAAAGCTGCAAATGTGgacctgACCGACactcataacgccGGCgaaGGCggGTGtctGTGagcGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGaagcaaGTGgcgGTTATCCCACGTCCCCATgaggatCTCGGCGCAgtgTTTttaGTTAGCTCAAAAGGTGCGGCCGC
Trio15-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGgtgGGTgtaATTaccACGcgtGGAgcgGTCGCCGCAattcgcGCTGCAGATGCTATGGTAcgtGCTGCAAATGTGagcctgACCggcatacataacgccGGCgaaGGCggggcgacgGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGaagacgGTGtttGTTcacCCACGTCCCCATcacaaagtaGGCGCAacctggaatgcgAGCTCAAAAGGTGCGGCCGC
Trio15-3	ggatccCAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTatgATTGAAACGcagGGATACGTCGCCGCAgtgtgtGCTGCAGATGCTATGGTAAAAGCTGCAAATGTGcgtctgACCgcgatgcataacgccGGCgaaGGCgggGTGacgGTGctgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttagagGTGgatGTTATCCCACGTCCCCATgctGAACTCGGCGCAcaggtgaatatcAGCTCAAAAGGTTCTAGCGAAAATCTGTACTTCCAGAGTAGTGCgccGC
Trio16-1	cataTGAGTAGTAACGCGATTGGTactATTGAAACGAAAGGAaactgtGCCGCAtatttcGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCATCACCagcCGGcatCAGgacGGCaacGGCgagGTGcaaGTGaccGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGatgGGCGAGCTGttacacGTGaccGTTATCCCACGTCCCCATTCGggcCTCGGCGCACATTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio16-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGgtgGGTtgtATTcagACGAAAGGAaccGTCGCCGCAattatgGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCctgACCagcatgcatCAGgacGGCaacGGCgagcgcacgGTGaccGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGGTTcaaGTGggcGTTATCCCACGTCCCCATTCGgatCTCGGCGCAgaaTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio16-3	ggatccCAGAAGGAGATATAtCTaTGAGTAGTAACGCGATTGGTTTAATTcagACGAAAGGATACagcGCCGCACTGGCTGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCcaaACCGACatgcatCAGgacGGCaacGGCgagGTGacgGTGctgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttagagGTGagcGTTATCCCACGTCCCCATTCGagcCTCGGCGCAtttgtgAGCGTTAGCTCAAAAGGTTCTAGCGAAAATCTGTACTTCCAGAGTAGTGCgccGC
Trio17-1	cataTGAGTAGTAACGCGgtgGGTatcATTGAAACGAAAGGAgtggggGCCGCAatttgtGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCATCACCgcgaagCAGgaaagcGGCgaaGGCatgGTGGCAGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttagagGTGgtgGTTATCCCACGTCCCaacTCGgatCTCGGCGCACATgtgAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio17-2	caattgtttaagaaggagatatagctaTGAGTAGTAACGCGATTGGTatcATTGAAACGAAAGGAgtgGTCGCCGCAaactggGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCctgACCgcgataCAGgaaagcGGCgaaGGCatgGTGttcGTGATCGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttacaaGTGgcgGTTATCCCACGTCCCCATTCGGAACTCGGCGCAcagTTTAGCGTTAGCTCAAAAGGTGCGGCCGC
Trio17-3	ggatccCAGAAGGAGATATAtCTaTGAGTAGTAACGCGgtgGGTgggATTGAAACGAAAGGAattgggGCCGCAatCGCTGCTGCAGATGCTATGttgAAAGCTGCAAATGTGACCctgACCGACgtaCAGgaaagcGGCgaaGGCatgGTGGCAGTGgtgGTAACGGGTGAGGTTGGGGCCGTAAAAGCTGCCACTGAAGCAGGCGCTGAAACTGCGTCGCAGGTTGGCGAGCTGttacaaGTGggcGTTaacCCACGTCCCCATTCGaacaagGGCGCAggcctgAGCGTTAGCTCAAAAGGTTCTAGCGAAAATCTGTACTTCCAGAGTAGTGCgccGC

Supplementary table 7. *Klebsiella pneumoniae 342* BMC-H sequences.
EutK either C- or N-terminally tagged with the GFP10 (light green) or 11 (dark green) is presented as an example of the different fragments obtained after amplification from individual POI-coding pET26b (vector 2 type) with primers in Table 1. Homology regions with the open vector or adjacent fragment for Gibson assembly are in italics. All subsequent POI coding sequences are given between NcoI (in pink) and BamHI sites (in blue). BglII site is in orange and HindIII in purple.

Case	Sequence
EutK-GFP10	*cgtccggcgtagaggatcg* agatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctCTAG AATACAGACTAGTATACTTTAAGAAGGAGATATACCATGGCAATTAACGCCTTAGGCTTACTGGAGGTGGAAGGTATG GTAGCTGCCGTTGATGCCGCGGATGCAATGCTAAAAGCCGCGAACGTTCGCTTGCTGTCACACGAAGTATTAGATCC GGGACGTCTAACCTTAGTGGTGGAAGGCGATTTAGCGGCGTGCCGCGCGGCGTTAGATGCGGGTTCAGTGGCAGC GGAGCGTACAGGCTGCGTGATTAGCCGGCGCGAGATTGGCCGTCCGGAAGAAGATACCCAGTGGCTGATTGGCGG CTTTCAGCCGCCGCCGCCAGCCCCACTGCCCCCGGCTGATCCAGCGTCAAGCGAGGCGTTACTGACGCTGCTTGCGA GCGTTCGCCAGGGCATGACGGCGGGCGAAGTGGCGGCGCATTTTGCGTGGCCGCTGGATAAAGCGCGCCAGGCGC TGGATCAGTTGTTTTCAGCAGGCACCTTACGCAAGCGCAGCAGCCGGTATCGTCTAAAAAATCCCGGatccGGtGGCT CCTCAGAAGGAGGCGGTAGCGGGGGCCCTGGTTCGGGAGGGGAAGGTTCTGCTGGGGGAGGGAGCGCTGGCGG GGGGTCTGATTTACCAGACGATCATTACCTGAGCACACAAACGATCCTTTCGAAAGACCTGAACGCAAGCTAATGAgc Gtcctctg*tacagactagtATACTTTAAG*
GFP10-EutK	*cgtccggcgtagaggatcg* agatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctCTAG AATACAGACTAGTATACTTTAAGAAGGAGATATACAAATGGATTTACCAGACGATCATTACCTGAGCACACAAACGATC CTTTCGAAAGACCTGAACGGTGGGTCCGGCTCAGAAGGAGGCGGTAGCGGGGGCCCTGGTTCGGGAGGGGAAGG TTCTGCTGGTGGAGGGAGCGCAAGCGGCGCCATGGCAATTAACGCCTTAGGCTTACTGGAGGTGGAAGGTATGGTA GCTGCCGTTGATGCCGCGGATGCAATGCTAAAAGCCGCGAACGTTCGCTTGCTGTCACACGAAGTATTAGATCCGGG ACGTCTAACCTTAGTGGTGGAAGGCGATTTAGCGGCGTGCCGCGCGGCGTTAGATGCGGGTTCAGTGGCAGCGGAG CGTACAGGCTGCGTGATTAGCCGGCGCGAGATTGGCCGTCCGGAAGAAGATACCCAGTGGCTGATTGGCGGCTTTC AGCCGCCGCCGCCAGCCCCACTGCCCCCGGCTGATCCAGCGTCAAGCGAGGCGTTACTGACGCTGCTTGCGAGCGT TCGCCAGGGCATGACGGCGGGCGAAGTGGCGGCGCATTTTGCGTGGCCGCTGGATAAAGCGCGCCAGGCGCTGGA TCAGTTGTTTTCAGCAGGCACCTTACGCAAGCGCAGCAGCCGGTATCGTCTAAAAAATCCCGGatccTAATGAgcctcct ctg*tacagactagtATACTTTAAG*
EutK-GFP11	*tacagactagtATACTTTAAG* AAGGAGATATACCATGGCAATTAACGCCTTAGGCTTACTGGAGGTGGAAGGTATGGTA GCTGCCGTTGATGCCGCGGATGCAATGCTAAAAGCCGCGAACGTTCGCTTGCTGTCACACGAAGTATTAGATCCGGG ACGTCTAACCTTAGTGGTGGAAGGCGATTTAGCGGCGTGCCGCGCGGCGTTAGATGCGGGTTCAGTGGCAGCGGAG CGTACAGGCTGCGTGATTAGCCGGCGCGAGATTGGCCGTCCGGAAGAAGATACCCAGTGGCTGATTGGCGGCTTTC AGCCGCCGCCGCCAGCCCCACTGCCCCCGGCTGATCCAGCGTCAAGCGAGGCGTTACTGACGCTGCTTGCGAGCGT TCGCCAGGGCATGACGGCGGGCGAAGTGGCGGCGCATTTTGCGTGGCCGCTGGATAAAGCGCGCCAGGCGCTGGA TCAGTTGTTTTCAGCAGGCACCTTACGCAAGCGCAGCAGCCGGTATCGTCTAAAAAATCCCGGatccGGtGGCTCCTCA GAAGGAGGCGGTAGCGGGGGCCCTGGTTCGGGAGGGGAAGGTTCTGCTGGGGGAGGGAGCGCTGGCGGGGGGT CTACCAGCGAAAAACGCGATCACATGGTGCTGCTGGAATATGTGACCGCGGCGGGCATTACCGATGCGAGCTAATGA GCGTCCTCTGTAcagactagaagctt*tctcgagttaactcgtgagcaa*
GFP11-EutK	*tacagactagtATACTTTAAG* AAGGAGATATACAAATGGAAAAACGCGATCACATGGTGCTGCTGGAATATGTGACCGC GGCGGGCATTACCGATGCGAGCGGTGGGTCCGGCTCAGAAGGAGGCGGTAGCGGGGGCCCTGGTTCGGGAGGG GAAGGTTCTGCTGGTGGAGGGAGCGCAAGCGGCGCCATGGCAATTAACGCCTTAGGCTTACTGGAGGTGGAAGGTA TGGTAGCTGCCGTTGATGCCGCGGATGCAATGCTAAAAGCCGCGAACGTTCGCTTGCTGTCACACGAAGTATTAGATC CGGGACGTCTAACCTTAGTGGTGGAAGGCGATTTAGCGGCGTGCCGCGCGGCGTTAGATGCGGGTTCAGTGGCAGC GGAGCGTACAGGCTGCGTGATTAGCCGGCGCGAGATTGGCCGTCCGGAAGAAGATACCCAGTGGCTGATTGGCGG CTTTCAGCCGCCGCCGCCAGCCCCACTGCCCCCGGCTGATCCAGCGTCAAGCGAGGCGTTACTGACGCTGCTTGCGA GCGTTCGCCAGGGCATGACGGCGGGCGAAGTGGCGGCGCATTTTGCGTGGCCGCTGGATAAAGCGCGCCAGGCGC TGGATCAGTTGTTTTCAGCAGGCACCTTACGCAAGCGCAGCAGCCGGTATCGTCTAAAAAATCCCGGatccTAATGAG CCTCCTCTGTAcagactagaagctt*tctcgagttaactcgtgagcaa*
EutM	ccATGGAGGCTCTAGGGATGATTGAAACGCGCGGTCTGGTAGCCTTAATCGAAGCCAGCGATGCAATGGTAAAAGCC GCGCGCGTGAAATTAGTGGGCGTGAAACAGATTGGCGGCGGCCTGGTGACCGCGATGGTGCGCGGCGATGTGGCG GCGTGCAAAGCGGCCACGGATGCGGGCGCCGCGGCGGCGCAGCGGATCGGTGAACTTGTTAGTGTGCATGTGATT CCGCGTCCGCACGGCGATCTGGAAGAAGTATTTCCGATCAGCTTTAAAGGGGATAGCAACATTGGatcc
EutS	ccATGGATAAAGAGCGCATTATCCAGGAGTTTGTTCCGGGGAAACAGGTTACGCTGGCACATCTGATTGCGCATCCGG GTGCCGAATTAGCGAAAAAGATTGGCGTGCCCGAATCGGGCGCGATTGGCATCATGACATTAACGCCGGGGGAAAC TGCGATGATTGCGGGCGATCTGGCGATGAAAGCTGCCGATGTTCATATCGGCTTTTTAGATCGGTTTAGCGGCGCGCT GGTGATTTATGGCCCGGTTGGCGCCGTGGAAGAAGCGCTGCTGCAGACCATCGGCGGCTTAGGCCGGCTGCTAAAC TACACCCTTTGTGAGCTGACAAAATCAGGatcc

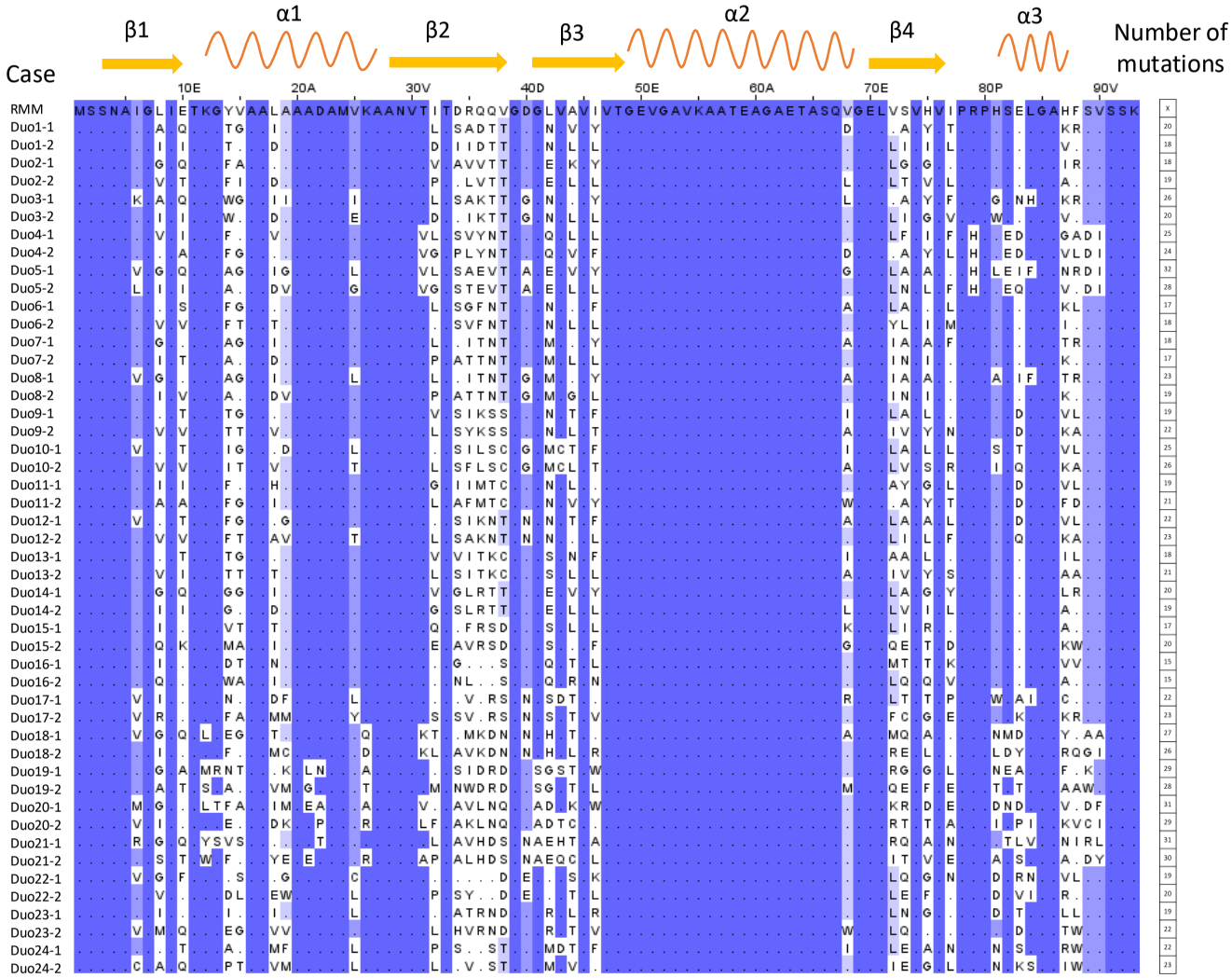
Supplementary table 7. *Klebsiella pneumoniae 342* BMC-H sequences (continuation).

Case	Sequence
CmcA	ccATGGGAGACGCCTTAGGCTTAATTGAAACCAAGGGCTTAGTGGCCTGTATCGAAGCTGCTGACGCGATGTGCAAA GCGGCGAACGTGGAACTGATTGGCTATGAAAACGTGGGCAGCGGCCTGGTGACCGTGATGGTGAAAGGTGATGTTG GCGCGGTGAAAGCTGCTGTGGATAGCGGTGTTGAAAGCGCGCAGCGCATTGGCGAAGTGGTGACCAGCCTGGTGA TCGCGCGTCCTCATAATGATATCAATAAAATCGTAATCAAACACAAGGCGGGatcc
CmcB	ccATGGGAGATGCTTTAGGCTTAATTGAAACCAAAGGGCTGGTGGCGTGCATCGAAGCGGCCGATGCCATGTGTAAA GCTGCGAACGTTGAACTTATTGGCTATGAAAACGTGGGCAGCGGCCTGGTGACCGCGATGGTGAAAGGCGATGTTG GCGCGGTGAAAGCTGCAGTGGATAGCGGCGTGGAAAGCGCGCAGCGCATTGGCGAAGTGGTGACCAGCCTGGTGA TTGCGCGCCCGCACAACGATATCAATAAAATCGTCTCACACTACAAAATCGCAGATGGatcc
CmcC	ccATGGCAAAGGAAGCTCTGGGATTAATTGAAACGAAAGGGCTGGTCGCCTGTATCGAAGCTGCGGATGCAATGTGT AAAGCGGCGAACGTGGAATTAATTGGCTATGAAAATGTTGGCAGCGGTTTAGTGACGGCGATGGTTAAAGGGGATGT GGGTGCCGTGAACGCCGCAGTGGATAGCGGCGTGGAAGCGGCGAAACGCATTGGTGAAGTTGTGAGCAGCCGGGT GATCGCACGCCCACATAACGATATTGAAAAAATCGCTGCACAGCACAAAGCAGGatcc
CmcE	ccATGGCAAAATCACTAGGGGTAATCGAGACGCGGGGTTGGGTAGCGGCGATTCAGGCTGTTGATGCAGCCTGCAA AGCTGCAGGTGTTACCTGCATTGGCTATCGTAAACCAGGCAGCGGTCTGGTCAGCGTGTGTTTTGAAGGTGAAATCA GCGCCATTCATACCGCGATTGAACGCGGCGTGGCGGTGGCGGGCGCGGAACATACCGTGAAATCGCTGGTGATTGC GCGCCCGGAAAGATGTGTGGTTGAAGCCCTGTCGAACCTGAAAGGTAACCCGCCGCCGGCGAAAAAAGCAGCGGA GCCGGTTGTGATTGCGGCGCCGGAGCCGATCGTGCCACCGGCCGCGCCAAACGAAACCGAAGATAAACACCCGGCT CTGAAGAAAGGAAAAAAGTCAGGatcc
PduA	ccATGGCACAACAAGAGGCTCTAGGAATGGTAGAAACGAAAGGACTGACAGCAGCTATCGAGGCTGCTGACGCTATG GTAAAGAGCGCTAATGTCCTTTTAGTGGGTTACGAACGAATTGGTAGCGGCCTGGTGACCGTGATTGTGCGCGGCGA TGTTGGCGCCGTGAAGGCGGCCACCGACGCGGGCGCGGCGGCGGCGCGCCATGTTGGCGAAGTCAAAGCTGTGC ATGTAATTCCACGCCCACATACCGATGTCGAAAAGATTTTACCGAAGGGCAATTCGCAAGGatcc
PduJ	ccATGGCAAATAATGCTTTGGGCTTAGTGGAAACCAAAGGGTTAGTGGGTGCGATCGAAGCTGCAGACGCGATGGTC AAAAGCGCCAACGTTCAGTTAATTGGCTATGAAAAGATCGGTAGCGGCCTGATTACCGTTATGGTCCGTGGGGATGTC GGTGCCGTGAAGGCGGCCGTGGATGCGGGCAGCGCCGCCGCGAGCGTGGTGGGTGAAGTAAAATCAAGCCATGT GATTCCGCGCCCGCATAGCGATGTGGAAGCGATTCTGCCAAAATCAGTTGGatcc
PduK	ccATGGCAAAGCAATCACTAGGTCTCTTAGAAGTGAGCGGCCTGGCGCTGGCGATTACCTGCGCGGATGCGATGGCG AAAGCGGCGGCGATCACCCTGTTAGCGCTGGAAAAAACGAACGGCAGCGGCTGGATGGTAGTCAAAATCGTCGGTG ATGTTGCGAGCGTGCAGGCGGCGGTGATGACGGGCGCGGAATTGGCCGATCGTCAGCAGGGCCTGGTTGCCCAGA AAGTAATCGCCCGTCCAGGCGCGGGCCTGCTGCCGGCACGGGTGGAGGCACCCTCGCCCGCTCCCGACGCAGCCTT AGAAGAGGAAAATGCCACGATCATGGACGAGCGGCGTGACCCAGCAGATACACTGCCCCGCCCAGCGGAACAGGT GACCTGCAACCTGTGCCTGGACCCGCACTGTCCCCGGCAAAAAGGTGAACCACGCAGCCAGTGCCTGCATGCCGGT AAACGAGGCGACGCCGGatcc
PduU	ccATGGAACCCCAGACGCCAACCGAACGTATGATTCAGGAATATGTGCCGGGCAAACAAGTGACCTTAGCGCATCTG ATCGCCAATCCGGGTAAAGACCTGTTTAAAAAATTAGGCCTGCCGGATGCGGTGAGCGCGATTGGCATTTTAACCATC ACTCCGTCTGAAGCAAGCATTATCGCGTGCGATATCGCGACGAAATCTGGTGCGGTGGAAATCGGCTTTCTGGACCG CTTTACCGGCGCGGTGGTGCTGACGGGCGATGTTAGCGCCGTTGAATACGCCTTGCGTCAAGTTACACGTACACTGG GTGAACTGATGCGTTTTACCGCGTGCCCGATTACCCGCACCGGatcc

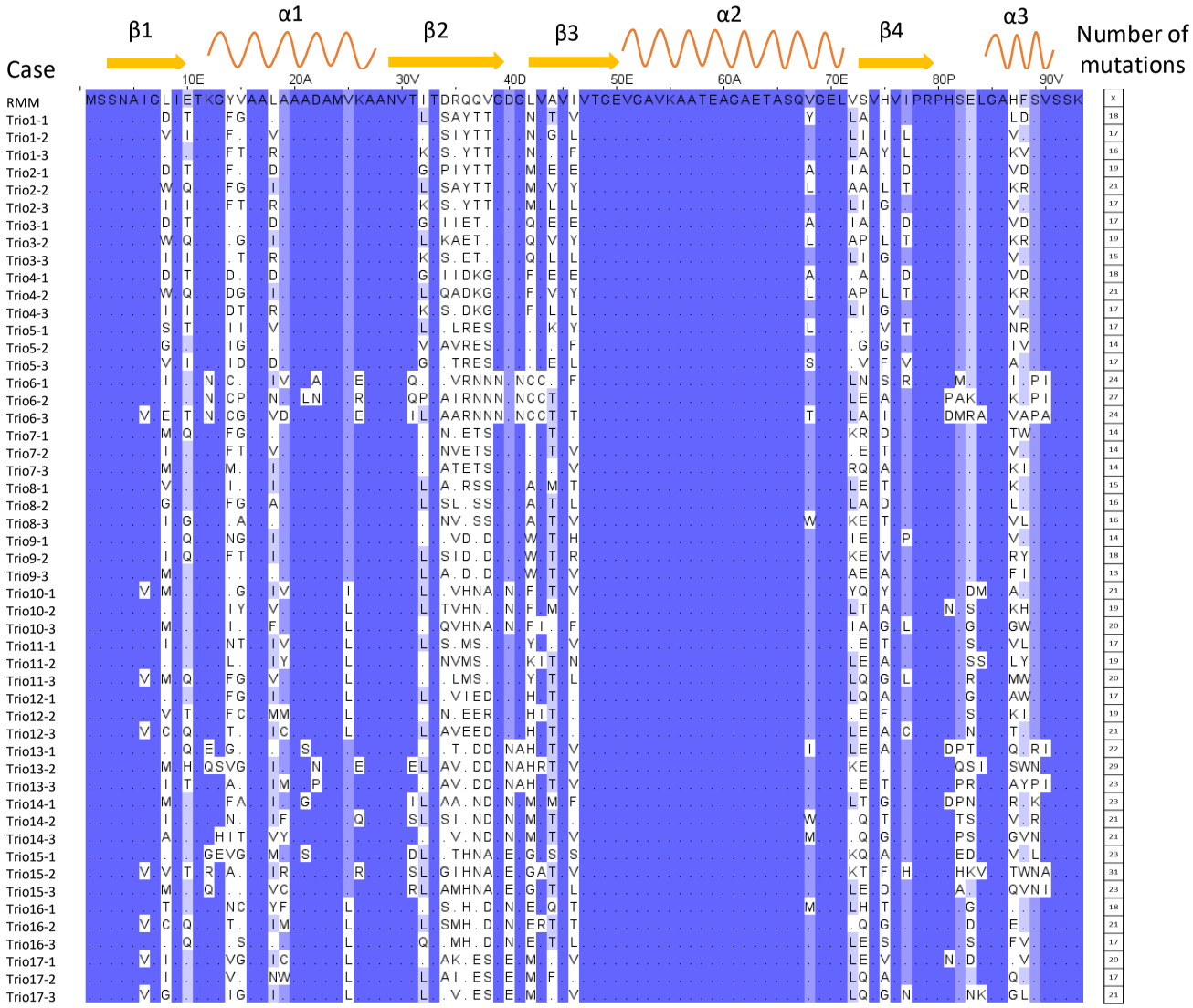


Supplementary figure 1. Macrostructure formation with the GFP-tagged version of RMM or its peripheral mutants.

TEM observations of *E. coli* cells overexpressing GFP10/GFP11-tagged wild-type RMM or RMM^{K26D} (sm-RMM), $RMM^{N29D,A53D}$ (dm-RMM), $RMM^{K26D,N29D,A53D}$ (tm-RMM) homo-pairs along with the GFP1-9. White arrows point at nanotube bundles in cell transversal view.



Supplementary figure 2. Sequence alignment between RMM and the variant Duos.



Supplementary figure 3. Sequence alignment between RMM and the variant Trios.

Résumé

Les microcompartiments bactériens (BMC) sont des structures polyédriques présentes dans de nombreux procaryotes. Ils sont composés d'une coque protéique semi-perméable renfermant un ensemble d'enzymes dont la nature définira la spécialisation du BMC dans une voie métabolique donnée. En effet, il existe divers BMCs parmi les phyla procaryotes (Axen *et al*, 2014) dont les plus connus sont le carboxysome et les microcompartiments utilisant l'éthanolamine ou le propanediol (CBX, EUT ou PDU respectivement).

Généralement, les protéines impliquées dans la formation et la fonction des BMCs sont codées par des gènes organisés en un même opéron. Toutefois, une partie de ces gènes peut être parfois retrouvée dans des *loci* chromosomiques indépendants comme pour le β-CBX. Les protéines de la coque se distinguent en 3 groupes : les pentamères retrouvés aux sommets des coques polyédriques, les trimères et les hexamères, constituant les faces et les arêtes des polyèdres. Plusieurs homologues de chaque sous-unités sont généralement présents dans une même bactérie (Axen *et al*, 2014). Pour les sous-unités qui s'associent en hexamère (BMC-H), les plus abondantes dans les coques, une moyenne de 3,5 exemplaires par opéron est retrouvée.

Depuis les premières études structurales des composants des coques, il était assumé que seuls des homo-hexamères se formaient (Kerfeld *et al*, 2005; Pitts *et al*, 2012; Mallette & Kimber, 2017; Tsai *et al*, 2007). Mais récemment, deux équipes, dont la nôtre, ont démontré la formation d'hétéro-hexamères impliquant différents homologues : les couples CcmK1/CcmK2 et CcmK3/CcmK4 du β-CBX (Garcia-Alles *et al*, 2019; Sommer *et al*, 2019). En effet, les BMC-H présentent une forte homologie de séquence (Sutter *et al*, 2017), notamment aux interfaces entre monomères. Toutefois, bien que certains organismes aient dans leur génome plusieurs opérons codant divers BMCs (Sutter *et al*, 2021) et soient, en théorie, capables de les exprimer simultanément, il y a à ce jour un manque d'informations concernant la possibilité d'interaction entre monomères de BMC de différents types au sein d'un même organisme. Toutefois, il est à noter que la formation de telles structures hybrides pourrait impacter l'intégrité des coques ainsi que les fonctions métaboliques des BMCs comme cela a été montré lorsque EutL ou EutS étaient co-exprimés avec les protéines du PDU (Sturms *et al*, 2015).

L'objectif de cette thèse fut double : (1) étudier la possibilité de cross-interactions entre homologues au sein de BMCs autres que le β-CBX ainsi qu'entre homologues venant de BMCs de différents types et (2) élaborer une plateforme protéique sur la base d'un hétéro-hexamère où chaque monomère aurait une place définie. Il faudrait pour cela maîtriser l'emplacement relatif des monomères grâce à des interfaces intra-hexamère qui reconnaitraient spécifiquement une paire

donnée de monomères (monomère A côté interface A et monomère B côté interface B), tout en évitant les interactions avec les 4 autres monomères du même hexamère et avec soi-même. A terme, cette plateforme permettrait un contrôle spatial à l'échelle nanométrique de modules que l'on grefferait sur chaque monomère, comme par exemple des enzymes pour améliorer l'efficacité catalytique de la voie métabolique qu'elles catalysent.

Ce travail de thèse s'est donc articulé en 3 parties. Dans le premier chapitre, j'ai choisi d'utiliser la tripartite GFP (tGFP) (Cabantous *et al*, 2013) et de l'adapter au cas des BMC-H. En effet, en partageant la GFP en 3 parties (la GFP1-9, la GFP10 et la GFP11, avec le numéro indiquant les brins β de la GFP inclus dans chaque partie) et en reliant les deux plus petits fragments GFP10 et GFP11 à 2 protéines d'intérêt (POI), il est possible de déterminer si ces 2 POIs interagissent ensemble ou non en suivant l'apparition d'un signal fluorescent. Une interaction entre les 2 POIs induit le rapprochement des étiquettes GFP10 et 11, ce qui favorise la reconstitution de la GFP entière et sa fluorescence. Ainsi, avec la tGFP comme technique d'étude d'interactions protéine/protéine (PPI), j'ai pu déterminer qu'un codage des POIs sur un même plasmide, dans un même cadre ouvert de lecture conduisant à un ARN messager bicistronique (avec la GFP1-9 transcrite sur un ARNm indépendant) était préférable. A noter que ce même agencement était aussi adapté à l'étude des interactions des autres composants de la coque des BMCs. Grâce à cela, j'ai pu valider les données obtenues précédemment avec les couples CcmK1/CcmK2 et CcmK3/CcmK4 (Garcia-Alles *et al*, 2019).

Dans le deuxième chapitre, la tGFP a été utilisée pour sonder toutes les cross-interactions possibles entre BMC-H issus de *Klebsiella pneumoniae 342* (*Kpe 342*). *Kpe 342* est une bactérie qui a dans son génome 3 opérons codant pour le PDU, l'EUT et un autre microcompartiment métabolisant la choline appelé GRM2. Au total, 11 homologues BMC-H sont présents chez *Kpe 342*. Une librairie de paires de BMC-H a donc été construite et testée en tGFP et j'ai pu montrer que la formation d'hétéro-hexamères était un phénomène communs aux 3 BMCs étudiés. De plus, des cross-interactions entre BMC-H issus de types de BMC différents ont été mises en évidence, posant la question de la relevance biologique de ces hétéro-hexamères hybrides. Il est probable que des systèmes de régulation existent afin de prévenir l'assemblage d'hexamères et de BMCs hybrides non-fonctionnels, comme cela a été montré chez *Salmonella enterica* entre l'*eut* et le *pdu* (Sturms *et al*, 2015). De plus amples études *in vivo*, chez *Kpe 342*, seront nécessaires pour éclaircir ce point.

Finalement, le troisième et dernier chapitre visait à établir les premières bases d'un projet à plus long-terme, l'élaboration d'une plateforme protéique hétéro-hexamérique. A ce titre, un système composé de 2 intelligences artificielles (IA ; Effie et Toulbar2), spécialisé dans le design *de novo* de protéines et créé par 2 équipes de collaborateurs en design computationnel, a été utilisé pour concevoir de nouvelles séquences adoptant le même repliement que des BMC-H. Le but, ici, était

d'augmenter la diversité de bio-briques disponibles pour l'élaboration de la plateforme hexamérique et de mieux appréhender leur assemblage. Ainsi, une première série de séquences conservant les attributs d'un BMC-H naturel (expression, solubilité, hexamérisation) a été designée.

Dans l'étape suivante, 2 plateformes différentes ont été visées : la première composée d'un couple de 2 monomères A et B (Duo) qui s'organiseraient en hétéro-hexamère ABABAB tandis que la deuxième serait composée de 3 monomères A, B et C (Trio) qui formeraient un hétéro-hexamère ABCABC. Le processus de design a donc été affiné pour inclure des contraintes d'états négatifs. Ces états négatifs comprenaient des oligomères défavorisés comme les homo-hexamères ou tous les hétéro-hexamères en respectant pas l'alternation de monomères visée. Le système à 2 IA a montré une grande fiabilité pour designer des Duos formant des plateformes hétéro-hexamériques stables. Ces plateformes rendraient possible l'immobilisation d'une voie métabolique composée de 2 enzymes différentes. Toutefois, il était moins adapté pour proposer des hétéro-hexamères à 3 BMC-H (Trio). que Protein MPNN (Dauparas *et al*, 2022). Ainsi, le système à 2 IA devra être amélioré afin de permettre le design de plateformes hétéro-hexamériques plus complexes.

Pour la suite de ce projet, 3 pistes seront à approfondir. La première sera de confirmer l'organisation ABABAB ou ABCABC au sein des hétéro-hexamères. La seconde serait de faire la preuve de concept que des enzymes peuvent être immobilisées et organisées avec précision sur ces plateformes. La dernière, mais non des moindres, serait de savoir si les BMC-H designés *de novo* conservent la capacité des BMC-H naturels à s'auto-assembler pour former des macrostructures (nanotubes) ou les facettes d'un BMC. En effet, inclure les plateformes designées dans des échafaudages protéiques ou au sein d'une coque de BMC pourraient permettre d'augmenter d'autant plus l'efficacité de catalyse de la voie qu'elles portent ou d'envisager l'immobilisation de voies métaboliques normalement problématiques (incluant un intermédiaire toxique ou volatile notamment).

Abréviations

BMC microcompartiment bactérien, BMC-H : monomère formant des sous-unités de la coque de BMC hexamériques, CBX : carboxysome, EUT : éthanolamine utilisation BMC, IA : intelligence artificielle, *Kpe 342* : *Klebsiella pneumoniae 342*, PDU : propanediol utilisation BMC, POI : protéines d'intérêt, PPI : interactions protéine/protéine, tGFP : tripartite GFP.

Abstract

Bacterial microcompartments (BMC) are protein structures, naturally found in some bacteria in which they act as bioreactor and process specific substrates. For instance, depending on the BMC type, the enzymatic set they encapsulate can fixate atmospheric CO_2 or catabolize the ethanolamine, 1,2-propanediol or the choline. The BMC shell is polyhedral and is composed of 3 different subunits, including the BMC-H, a protomer associating as an hexamer which are the main and the most diverse shell subunits, in terms of number of homologs within a single BMC operon. Indeed, genomic surveys indicate an average of 3,5 BMC-H homologs per operon, with some organisms like *Clostridium saccharolyticum WM1* coding for up to 15 BMC-H split between 3 BMC types.

Although it has long been thought that only homo-hexamers existed, it was recently evidenced that hetero-hexamer formation occurred between BMC-H homologs in 2 different β-carboxysome-expressing bacteria. Indeed, numerous BMC-H homologs share a high sequence identity, notably at the intra-hexamer interfaces. Besides paving the way for possible hetero-hexamer formation beyond the β-carboxysome, inside organisms equipped with one BMC type, these recent studies raise the question of possible cross-interactions between BMC-H coming from multiple BMC types.

One objective during my PhD thesis was to examine the occurrence of hetero-hexamers in nature. To this end, the tripartite GFP was adapted to study protein-protein interactions among BMC-H and implemented on the case study of *Klebsiella pneumonia 342* BMC-H. Of note, this organism is very interesting because it has in its genome 3 BMC loci, comprising a total of 11 BMC-H homologs. Then, besides allowing to determine whether hetero-hexamers do form aside from the β-CBX, in 3 other BMC types, their study would also bring some answer elements to the question of the cross-interactions between BMC-H arising from different BMC types.

A novel method to enhance a pathway catalytic efficiency (other than by classical enzymatic engineering) is gaining more and more interests nowadays: enzyme spatial organization. The idea is that, by putting in close proximity or in an arranged fashion the enzymes from a metabolic pathway, one could increase the efficiency of the pathway, through substrate channelling between the different enzymes, for instance, or enzyme clusterisation.

The majority of hexamers formed by the BMC-H have the intrinsic property to self-assemble and form higher-ordered macrostructures (nanotubes, Swiss-rolls, 2D sheets) when recombinantly expressed alone in *E. coli*. This peculiarity has already been exploited in multiple studies to create a protein scaffold for the immobilization of enzymes. In these proof-of-concepts, a sole BMC-H was used to build the scaffold, which would only permit to immobilized different enzymes in a random fashion.

Here, we propose to go further with the idea of spatial organization and aimed to elaborate a protein platform starting from an hetero-hexamer. This hetero-hexamer would be composed by 2 up to 6 different BMC-H with each BMC-H constituting an anchoring point for a future enzymatic domain. With such platform, the spatial organization of the enzymes would be more finely controlled which would further enhance the catalysis efficiency of a metabolic pathway.

To meet this goal, *de novo* designed BMC-H were created by 2 collaborator teams of computational design. I studied them and searched for BMC-H couples that would depict orthogonal intra-hexamer interfaces. Indeed, to be able to control precisely the organization onto the platform, this would require to ensure a specific BMC-H order within the hetero-hexamer and thus, tightly control which BMC-H is adjacent to which one and prevent any other association.